MICHAEL W. KLYMKOWSKY & MELANIE M. COOPER

BIOFUNDAMENTALS

Contribution:

OER

Open Educational Resources
UKM Literasi Informasi & Perpustakaan Unsyiah

biofundamentals

A two semester introduction to the core concepts of evolutionary, molecular, cellular & genetical systems
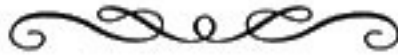
Michael W. Klymkowsky & Melanie M. Cooper

Molecular, Cellular & Developmental Biology, University of Colorado Boulder

Chemistry, Michigan State University

note: one more complete proofreading will be carried out in the Summer of 2018 – comments welcome

You know how it is.
You pick up a book, flip to the dedication & find that, once again,
the author has dedicated a book to someone else & not to you.

Not this time.

Because we haven't yet met/have only a glancing acquaintance/are just crazy about each
other/haven't seen each other in much too long/are in some way related/will never meet,
but will, I trust, despite that, always think fondly of each other….

This one's for you.

for the explorer inside all of us

courtesy of Neil Gaiman

Preface: A biofundamentalist's approach to teaching & learning biology

Our goal is to present the key observations and unifying concepts upon which modern biology is based; it is certainly not to survey all of biology! Once understood, these key observations and unifying concepts should enable you to approach most any biological system or process, from the origin of disease to cooperation and kindness, from a scientific perspective.
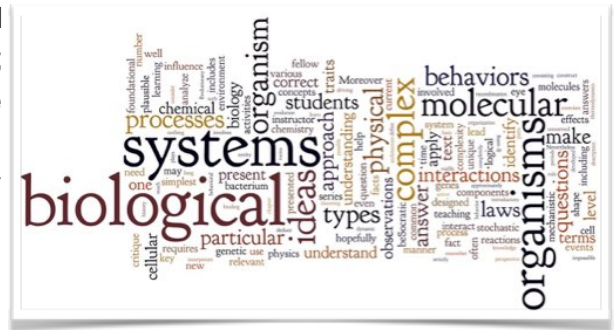
To understand biological systems we need to consider them from two complementary perspectives; how they came to be – the historic, that is, the evolutionary, and how they work – the physicochemical, that is, how their structures, traits, and behaviors are produced and maintained. We will also consider what it means to read and answer a question scientifically, how to draw meaningful conclusions from data, and to how to recognize when more (or better) data is needed.

We are biological entities, the products of evolutionary and developmental processes acting on inherited information stored in molecules and acting within a dynamic (cellular) chemical system. We live in complex and often unstable social arrangements with other humans and other organisms whose behaviors influence us in both subtle and profound ways. As we alter our environment we inevitably alter ourselves. Science is a coherent and communal strategy by which we seek to better understand the Universe and ourselves; how the physical world and its history shapes and constrains what is and what is not possible, and why. That said, science does not provide us with a prescription for how things should be. Science cannot tell us what is morally right or wrong, it can only attempt to explain what is and what might be. Our scientific understanding of almost every topic, and particularly the remarkably complex behaviors of biological systems, is incomplete. It is not even certain that the Universe is coherent. The difficulties in producing a single theory that encompasses both the behavior of the very large (gravity) and the very small (quantum mechanics) raises the possibility that a single theory of everything may not be possible or if possible, it may not be comprehensible to us.[1]

While science is a powerful strategy to understand and manipulate the world, it is certainly no guide to moral behavior. But its power can be seductive. Periodically a perspective (an ideology) known as scientism gains popularity in certain circles. Scientism holds that science provides a complete and exclusively valid description of the Universe, a picture that dictates how we should behave. We caution against this view, in part based on the lessons

> *Scientific knowledge is a body of knowledge of varying degrees of certainty-some most unsure, some nearly sure, but none absolutely certain … Now we scientists are used to this, and we take it for granted that it is perfectly consistent to be unsure, that it is possible to live and not know.*
> *- Richard Feynman.*
>
> *…it is always advisable to perceive clearly our ignorance.*
> *– Charles Darwin.*

of history and in part because it violates our own deeply held, some might say enlightenment, view that we are each unique individuals who are valuable in and of ourselves, deserving of respect. Individuals are not objects to be sacrificed to abstract ideals, that is, blown up or otherwise abused for ideological

---

[1] Physics's pangolin: Trying to resolve the stubborn paradoxes of their field, physicists craft ever more mind-boggling visions of reality: http://aeon.co/magazine/science/margaret-wertheim-the-limits-of-physics/ & Scientific method: Defend the integrity of physics: http://www.nature.com/news/scientific-method-defend-the-integrity-of-physics-1.16535

reasons, whether scientific, political, religious, or economic. A number of serious crimes committed against humanity as a whole and specific individuals have been justified based on purportedly established "facts" or beliefs that later turned out to be untrue, incomplete, tragically misapplied, more or less irrelevant, or simply cruel.[2] Crimes against people in the name of science are as unforgivable as crimes against people in the name of religious beliefs, political ideologies, or simple greed.

That said, scientific thinking and observations are indispensable if we want to distinguish established, empirically supported observations from frauds and fantasies. Such frauds and fantasies can often be harmful, such as the anti-vaccine campaigns that have led to an increase in childhood deaths and avoidable diseases.[3] When we want to cure diseases, reduce our impact on the environment, or generate useful tools we are best served by adopting a dispassionate, empirically-based scientific approach to inform, rather than dictate, our decisions. Scientific studies help us decide between the possible and the impossible and to assess the costs and benefits of various interventions. In this context it is worth noting the important difference between what has been established scientifically, what those conclusions imply, and how they interact with and influence other social, economic, political, and personal decisions.[4]

How biology differs from physics and chemistry

While it is true that biological systems, in other words organisms, obey the laws of physics and chemistry, they are not deducible from those laws, they are more than just highly complex chemical and physical systems. Why is that, you might ask? Because each organism is a unique entity, distinguishable from others by the genetic information it carries and, at the molecular and cellular levels, by the various stochastic events that have combined to influence its behavior. Even identical twins can be distinguished in terms of their molecular and behavioral details. Moreover, each organism has a unique history that runs back in time for an unbroken period of ~3,500,000,000 years, where the symbol "~" means "approximately". To understand an organism's current shape, internal workings, and various behaviors requires an appreciation of the general molecular, cellular, developmental, social, and ecological processes involved in producing these traits. Such mechanistic processes are themselves the product of what the molecular biologist François Jacob (1920-2013) referred to as evolutionary tinkering, that is, they reflect each organisms' unique evolutionary history (stretching back billions of years) as well as its current environment.[5]

Looking at the evidence, it is clear that no organism, including us, was designed *de novo*, from the Latin meaning, anew. Rather each organism is the product of continuous evolutionary processes that have been in play since the origin of life (~3.5 billion years ago). While a particular individual does not evolve, populations do over time, which means that evolution is simply a description of how populations change over time. The reason(s) for these changes can be grouped under the broad term of

---

[2] Walter Gratzer: The Undergrowth of Science

[3] How vaccine denialism in the West is causing measles outbreaks in Brazil & http://www.historyofvaccines.org/content/articles/history-anti-vaccination-movements

[4] What Daniel Sarewitz has terms trans-science: Saving science

[5] François Monod: Evolution and Tinkering & Tinkering: a conceptual and historical evaluation

evolutionary mechanisms; taken together these mechanisms lead to distinct populations of individuals adapted to particular life styles (ecological niches) through a combination of random (stochastic) and non-random events. These evolutionary mechanisms, which we will discuss in some detail, include the origin of mutations, that is, changes that alter the genetic material (double-stranded deoxyribonucleic acid or DNA) and the effects of these molecular variations, the genotype, on the shape or behavior of the organism, its phenotype. The genetic material is dynamic and subject to various forms of additions, deletions, and shuffling. The primary driver of changes in populations over time is known as "selection" and is essentially reproductive success. Various types of selection arise through internal processes and an organism's interactions with other organisms and changes in its environment. Because of these complex and interacting processes, one cannot readily deduce the details of a particular organism from physical first principles – and there are millions of different types of organisms. Take for example the vertebrate eye, which behaves completely in accord with physical laws, yet nevertheless displays idiosyncrasies associated with its evolutionary history. Such differences enable us to deduce that the details of the vertebrate eye arose independently from, for example, the eyes of mollusks, that is squid and octopi.[6] Evolutionary processes lead to the emergence of new traits and modified types of organisms while at the same time playing a conservative role, maintaining organisms against the effects of molecular level noise in the form of mutations.[7] The interactions between organisms and their environment produce evolutionary changes in often unpredictable ways. These processes can lead to the extinction of some lineages as well as the appearance of new versions of organisms within existing lineages. Evolutionary processes have produced the millions of different types of organisms currently in existence, in addition to the many more that are now extinct.

Another important difference between biological and physicochemical systems is that even the simplest of biological systems, organisms consisting of an individual cell (we will define what exactly a cell is in the next chapter), are more complex than the most complex non-biological physical system. A bacterium, one of the simplest types of organisms in terms of molecular components, typically contains more than ~3000 distinct genes, and hundreds to thousands of concurrent and interdependent chemical reactions, whose interactions influence which genes are active (active genes are often said to be "expressed") and which are silent (inactive or not expressed), the range of ecological and environmental interactions that occur between organisms, and how an individual bacterium responds to them. Often these processes are controlled by a small number (one to a few hundred) of a particular type of molecule; the small number of molecules involved inevitably results in noisy (stochastic) behaviors that are difficult or impossible to predict on the individual cellular level. We will consider the implications of such stochastic processes in various systems in detail.

Not withstanding their complexity, there are common themes within biological systems that we will return to over and over again to help make such systems intelligible. We will rely on the fact that we can understand how molecules interact (through collisions and binding interactions), how chemical reactions interact with one another (through reaction coupling), and how physical laws, in particular the laws of thermodynamics, constrain and shape biological behaviors.

---

[6] How the Eye Evolved

[7] From an evolutionary perspective, a mutation is be considered harmful if it negatively effects on organism's reproductive success; whether a mutation is harmful or beneficial is determined by the context in which it occurs (a point we will return to).

Your background and our (Socratic) teaching approach

While it is often the case that biology is taught early in a science sequence, this seems counterintuitive and silly to us, since biological systems and processes are much more complex that non-living physical or chemical systems (even though biological processes are based on, and constrained by, physical and chemical principles.) We recognize that it is unlikely that most students will enter the course completely comfortable with the relevant physical and chemical concepts and we have written the text presuming very little. Where references to physicochemical concepts are necessary, we have attempted to point them out explicitly and to address them at a level that we believe should be adequate for you to be able to deal productively with the ideas presented. That said, your responsibility as a learner is to speak up if you do not think you understand an idea or the significance of a particular observation. Given that biology students are a large fraction of the cliental of introductory physics and chemistry courses, one can only hope that over time these courses will evolve to better help life science students learn what they need to know.[8] We suggest that students interested in learning more about the physical and chemical concepts that underlie biological systems might want to read Einstein and Infeld's "The Evolution of Physics"[9] and our own "Chemistry, Life, the Universe, and Everything."[10]
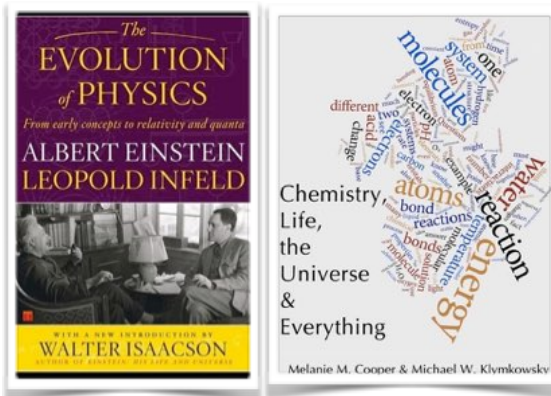
The complexity of biological systems can be daunting. All too often biology has been presented as a series of vocabulary terms, with little attention paid to its underlying conceptual (sense-making) foundations. This emphasis on memorization can be off-putting and, in fact, is not particularly valuable in helping you, the learner, develop a working understanding of biological systems. Our driving premise is that while biological systems are complex, both historically and mechanistically, there are a small set of foundational observations and general concepts that apply to all biological systems.[11] Their complexity and the incompleteness of our understanding often make an unambiguously accurate answer to biological questions difficult. Nevertheless, it is possible to approach biological questions in an informed, data-based (empirical), and logical manner. In general, we are less concerned with whether you can remember or reproduce the "correct" answer to a particular question and more interested in your ability to identify the facts, observations, and over-arching concepts relevant to a question and to then construct a plausible, logical, observation-based, and internally consistent response. More often than not, such a response will be the correct one.

Going beyond memorization means that you need to apply your understanding of key facts and overarching ideas to particular situations. This will require that you develop, through practice, the ability

---

[8] Physics for (molecular) biology students.

[9] Einstein and Infeld's The evolution of physics

[10] CLUE: Chemistry, Life, the Universe & Everything

[11] Klymkowsky: Thinking about the **conceptual** foundations of the biological sciences.

to analyze a biological situation, to identify what factors are critical, recognize those that are secondary or irrelevant, and then apply your understanding to make predictions or critique conclusions. To this end we will repeatedly ask you to dissect various situations in order to reach your own conclusions or solutions. To give you opportunities to practice, each section of the book includes a number of questions to answer and ponder. You should be able to generate plausible answers to these questions, answers that we will ask you to present to, and analyze with, your instructor and fellow

> *We think the way we do because Socrates thought the way he did.*
> *- Bettany Hughes*

students. Where you do not understand how to approach a question you should storm, in a civil and polite manner, into class and be able to articulate exactly why you are confused, something that can take some serious introspection. You will need to actively search for, and if you cannot find it, demand help in developing a viable approach that enables you to answer these questions or to explain in clear detail why the questions make no sense to you. As part of this process, we use the web-based interactive reading tool, Nota Bene (nb); we use your responses to frame in class discussions. We also use interactive beSocratic activities, accessible through web links[12]; these activities are designed to help you develop your ability to construct models and explanations of important phenomena. In many cases, you will receive feedback within the context of the activity. That said, there is no substitute for engaging in discussions with other students and your instructors. Ideas that you find obscure or that make no sense to you need to be addressed directly, do not let them go unchallenged! Learning to critique or question an explanation will help you identify what is relevant, irrelevant, conceptually correct, or logically absurd in your and your fellow students' thinking, so that by the time we reach the end of the course, you will have learned something substantial about biological systems (including yourself). One mark of an educated person is that they are an accurate BS detector.[13]

Learning how to explain, critique, and argue scientifically: We have noticed that students often have a difficult time generating a scientifically reasonable and plausible explanation of a biological process, or in explaining the reasoning behind their answers or their choices on multiple choice type tests. To this end we will spend time during the course, rather than in the book, to help you practice organizing your thoughts and generating an explanation, argument, or critique based on explicitly stated assumptions and logic. To learn how to write effectively it is important to practice and get feedback, have people ask you to justify your assumptions and your logic, and for you to feel willing to revise your thinking as necessary. This reflects the fact that such "hard thinking" and clear speaking (writing) are not natural, but needs to be learned, nurtured, and mastered.[14]

When you are answering a question we suggest that you write out your answer and then read it back to yourself, preferably aloud.[15] Reading your own writing helps you recognize awkwardly phrased or illogical constructions that you might miss when you read silently (in your head).[16] In part this is due

---

[12] beSocratic can currently be accessed here: *http://beSocratic.com*; Nota bene is found here: http://nb.mit.edu/welcome

[13] Issac Newton and BullSh*t detector A Guide to Being Less Wrong

[14] Review of "Thinking fast and slow"

[15] NYT: The Benefits of Talking to Yourself

[16] Reading aloud: http://writingcenter.unc.edu/handouts/reading-aloud/

to the fact that different parts of the brain are involved in hearing.[17] Many computer operating systems and word processing programs have text-to-speech capabilities and they can be quite useful in clarifying your writing.

What we are not "covering": One important point to the text and the course is that we aim at both an engaging narrative and a concerted effort to avoid unnecessary distractions. Why? Because it has been found that while experts can focus on the key aspects of a problem or system, novices (such as students in an introductory biology class) tend to take everything equally seriously – which can be distracting. We have tried to avoid introducing information or concepts that we will not be using again. Core ideas will be returned to repeatedly. Details will be avoided unless they are critical – as a example, there are many proteins involved in DNA replication, but a key fact is that polymerases work in one direction only, an important fact that one needs to remember (as you will see when we get to it). If you think we have introduced a distraction, please let us know.

Revisions to the text: Because this is an introductory course and because the ideas and observations presented are well established, we expect no need for dramatic revisions of content. The one exception is in the systems and genetics sections; the advent of cheap genomic sequencing has created a flood of new data and observations that we have incorporated, where appropriate, into a new set of chapters together with a review of key terms and concepts. We think of biofundamentals as a two semester course, so the molecular ideas introduced in part 1 should be reviewed and reinforced at the start of part 2. While most of the underlying ideas are well established, these new tools can help to better present these ideas and enable you to work with real data.

At the same time, we have much to learn about how to best help students master and apply complex biological ideas, so we are using student responses from the on-line interactive Nota Bene readings, from beSocratic activities, and from classroom interactions to identify effective activities and to fix ineffective ones.[18] New "editions" will incorporate these insights. You should check the "version date" at the bottom of each page to insure you have the latest version. Your observations, criticisms, and suggestions are greatly appreciated, feel free to express yourself (to us).

A note on footnotes: The authors have an inordinate fondness for footnotes. We do not expect you, the student or the more casual reader, to read them or the follow the links within them, but they enable us to indulge our interests in various topics. Please be careful to avoid getting lost in the footnotes–that may well be a mistake, a needless distraction, or an extremely interesting diversion.

---

[17] Speech and the Brain: http://webspace.ship.edu/cgboer/speechbrain.html

[18] The Design and Transformation of Biofundamentals: A Nonsurvey Introductory Evolutionary and Molecular Biology Course

PART I - Foundations

In which we consider evolutionary mechanisms, the physicochemical properties of organisms and the capture of energy, the nature of genetic information, how it is encoded, replicated, and read out, and how proteins work and interact.

Chapter 1: Understanding (biological) science & thinking scientifically

In which we consider what makes science a distinct, productive, and progressive way of understanding how the universe works. Science enables us to identify what is possible and plausible and what is impossible or irrelevant. We consider the "rules" that distinguish a scientific approach to a particular problem from a non-scientific one.

A major feature of science, and one that distinguishes it from many other human activities, is its essential reliance upon shareable experiences rather than personal revelations. Thomas Paine (1737-1809), one of the intellectual parents of the American Revolution, made this point explicitly in his book The Age of Reason (↓).[19] In science, we do not accept that an observation or a conclusion is true just because another person claims it to be. We do not accept the validity of revelation or what we might term "personal empiricism." What is critical is that, based on our description of a phenomenon, an observation, or an experiment, others should, in practice (if they have the resources and opportunity) be able to repeat the observation or the experiment. Science is based on social, that is, shared, knowledge rather than revealed truth.

> *Revelation is necessarily limited to the first communication – after that it is only an account of something which that person says was a revelation made to him; and though he may find himself obliged to believe it, it can not be incumbent on me to believe it in the same manner; for it was not a revelation made to ME, and I have only his word for it that it was made to him.*
> *– Thomas Paine, The Age of Reason.*

As an example consider sunlight. It was originally held that white light was "pure" and that somehow, when light passed through a prism, the various colors of the spectrum, the colors we see in a rainbow, were created *de novo*. In 1665, Isaac Newton (1642–1727) performed a series of experiments that he interpreted as demonstrating that white light was not pure, but in fact was composed of light of many different colors.[20] This conclusion was based on a number of experimental observations. First, he noted that sunlight passed through a prism generated a spectrum of light of many colors. He then used a lens to focus the spectrum emerging from one prism so that passed through a second prism (Part A→); a beam of white light emerged from the second prism. He went on to show that the light emerging from the prism 1 lens prism 2 combination behaved the same as the original beam of white light; he passed it through a third prism, which again produced a spectrum. In the second type of experiment (Part B→), Newton used a screen with a hole in it, an aperture, and showed

Gravity explains the motions of the planets, but it cannot explain who sets the planets in motion.
- Isaac Newton

that light of a particular color was not altered when it passed through a second prism - no new colors were produced. Based on these observations, Newton concluded that white light was not what it appeared to be – that is, a simple pure substance – but rather was composed, unexpectedly, of light of

---

[19] The Age of Reason: http://www.ushistory.org/paine/reason/singlehtml.htm

[20] Newton's Prism Experiments: http://micro.magnet.fsu.edu/primer/java/scienceopticsu/newton/ & http://youtu.be/R8VL4xm_3wk

many distinct "pure" colors. The spectrum was produced because the different colors of light were "bent" or refracted by the prism to different extents. Why this occurred was not clear at the time nor was it clear what, exactly, light is. Newton's experiments left these questions unresolved. This is typical: scientific answers are often extremely specific, elucidating a particular phenomenon, rather than providing a universal explanation of reality.

Two basic features make Newton's observations and conclusions scientific. The first is reproducibility. Based on his description of his experiment others could reproduce, confirm, and extend his observations. If you have access to glass prisms and lenses, you can repeat Newton's experiments yourself and you will come to the same empirical conclusions; that is, you will observe the same phenomena that Newton did.[21] In 1800, William Herschel (1738-1822) did just that. He used Newton's experimental approach and discovered infrared (beyond red) light. While infrared light is invisible to us, its presence can be revealed by the fact that when absorbed by an object, say by a thermometer or a hand, it leads to an increase in the temperature of the object.[22] In 1801, inspired by Herschel's discovery, Johann Ritter (1776 –1810) used the ability of light to initiate the chemical reaction:

silver chloride + light → silver + chlorine

to reveal the existence of another type of light, which he called "chemical light" and which we call ultraviolet light.[23] Subsequent researchers established that visible light is just a small portion of a much wider spectrum of "electromagnetic radiation" that ranges from X-rays to radio waves. Studies on how light interacts with matter have led to a wide range of technologies, from X-ray imaging to an understanding of the history of the Universe. All these findings emerge, rather unexpectedly, from attempts to understand the rainbow.

The second scientific aspect of Newton's work was his clear articulation of the meaning and implications of his observations, the logic and limitations of his conclusions. These led to explicit predictions, such as that a particular color will prove to be homogenous, that is, not composed of other types of light, which he then confirmed. His view was that the different types of light, which we see as different colors, differ in the way they interact with matter. One way these differences are revealed is the extent to which the different colors of light are bent when they enter a prism. Newton used some of these ideas when he chose to use mirrors rather than lenses to build his reflecting (Newtonian) telescope. His design avoided the color distortions that arise when light passes through simple lenses.

The features of Newton's approach make science, as a social and progressive enterprise, possible. We can reproduce an observation or experiment, and follow the investigator's explicit thinking. We can identify unappreciated factors that can influence the results observed and identify inconsistencies in logic and explore implications that may influence how various scientific disciplines interact with one another. Science rests on the premise that there is a world outside ourselves, that this world is real and constrains what is possible and what is not possible – it rules out "magical thinking", and so can be upsetting to some.

---

[21] Infrared astronomy

[22] There are some animals that can see infrared light: see link & link

[23] Ritter discovers ultraviolet light

The interconnectedness (self-consistency) of science

At one point in time the study of biology, chemistry, physics, geology, and astronomy appeared to be distinct, but each has implications for the others; they all deal with the same real world. In particular, it is clear that biological systems obey the laws and rules established by physics and chemistry. As we will see, it was once thought that there were aspects of biological systems that somehow transcended physics and chemistry, a point of view known as vitalism. If vitalism had proven to be correct, it would have forced a major revision of chemistry and physics. As an analogy, the world of science is like an extremely complex crossword puzzle, where the answer to one question must be compatible with the answers to all of the others.[24] Alternatively, certain questions, and their answers, once thought to be meaningful can come to be seen as irrelevant or meaningless. For example, how many angels can dance on the head of a pin is no longer considered a scientific question.

What has transpired over the years is that biological processes ranging from the metabolic to the conscious have been found to be consistent with physicochemical principles. What makes biological processes different is that they are the product of evolutionary processes influenced by historical events that stretch back in an uninterrupted "chain of being" over billions of years. Moreover, biological systems in general are composed of many types of molecules, cells, and organisms that interact in complex ways. All this means is that while biological systems obey physicochemical rules, their behavior cannot be predicted based on these rules. It may well be that life, as it exists on Earth, is unique. The only way we will know otherwise is if we discover life on other planets, solar systems, galaxies, and universes, if such things exist, all seriously non-trivial but exciting possibilities.

At the same time, it is possible that studies of biological phenomena could lead to a serious rethinking of physicochemical principles. There are, in fact, research efforts into proving that phenomena such as extrasensory perception, the continuing existence of the mind/soul after death, and the ability to see the future or remember the (long distant) past are real. At present, these all represent various forms of pseudoscience, and most likely, are forms of self-delusion and wishful thinking, but they would produce a scientific revolution if they could be shown to be real, that is, if they were reproducible and based on discernible mechanisms with explicit implications and testable predictions. This emphasizes a key feature of scientific explanations: they must produce logically consistent, explicit, testable, and potentially falsifiable predictions. Ideas that can explain any possible observation or are based on untestable assumptions, something that some would argue is the case for a number of religions or string theory in physics, are no longer science, whether or not they are "true" in some other sense.[25]

---

[24] This analogy is taken from a talk by Alan Sokal: http://youtu.be/kuKmMyhnG94; graphic from http://scienceblogs.com/principles/2013/10/09/quantum-crosswords-my-tednyc-talk/

[25] The lecture by Alan Sokal is worth listening to [here]: also see Farewell to Reality, Not even Wrong & Wronger than Wrong

Models, hypotheses, and theories

Scientific models are used in various ways. There are explanatory models that capture a certain approach to a system as well as exploratory and predictive models that are used to test ideas. Predictive models are commonly known as hypotheses. Such models are valuable in that they serve as a way to clearly articulate one's assumptions and their implications. They form the logical basis for generating testable predictions about the phenomena they purport to explain. As scientific models become more sophisticated, their predictions can be expected to become more and more accurate or apply to areas that previous forms of the model could not handle. Let us assume that two models are equally good at explaining a particular observation. How would we decide between them? One way is the rule of thumb known as Occam's Razor, also known as the Principle of Parsimony, named after the medieval philosopher William of Occam (1287–1347). This rule states that all other things being equal, the simplest explanation is to be preferred. This is not to imply that an accurate scientific explanation will be simple, or that simple explanations are correct, only that to be useful, a scientific model should not be more complex than necessary. Consider two models for a particular phenomenon, one that involves angels and the other that does not. We need not seriously consider the model that invokes angels unless we can accurately monitor the presence of angels and if so, whether they are actively involved in the process to be explained. Why? Because angels, if they exist, imply more complex factors than does a simple natural explanation. For example, we would have to explain what angels are made of, their origins, and how they intervene in, or interact with the physical world, that is, how they make matter do things. Do they obey the laws of thermodynamics? What determines when and where they intervene? Are their interventions consistent or capricious? Assuming that an alternative, angel-free model is as or more accurate at describing the phenomena and making verifiable predictions, the scientific choice would be the angel-free model. Parsimony (an extreme unwillingness to spend money or use resources) has the practical effect that it lets us restrict our thinking to the minimal model that is needed to explain specific phenomena. The surprising result, illustrated by a talk by Murray Gell-Mann [26], is that simple, albeit often counter-intuitive rules can explain much of the Universe with remarkable precision. A model that fails to accurately describe and predict the observable world must be missing something and is either partially or completely wrong.

Scientific models are continually being modified, expanded, or replaced in order to explain more and more phenomena more and more accurately. It is an implicit assumption of science that the Universe can be understood in scientific terms, and this presumption has been repeatedly confirmed but has by no means been proven. A model that has been repeatedly confirmed and covers many different observations is known as a theory – at least this is the meaning of the word in a rigorous scientific context. It is worth noting that the word theory is often misused, even by scientists who might be expected to know better. If there are multiple "theories" to explain a particular phenomenon, it is more correct to say that i) these are not actually theories, in the scientific sense, but rather working models or speculations, and that ii) one or more, and perhaps all of these models are incorrect or incomplete. A scientific theory is a very special set of ideas that explains, in a logically consistent, empirically supported, and predictive manner a broad range of phenomena. Moreover, it has been

---

[26] Beauty, truth and ... physics?

tested repeatedly by a number of critical and objective people – that is, people who have no vested interest in the outcome – and found to provide accurate descriptions of the phenomenon it purports to explain. It is not idle speculation. If you are curious, you might count how many times the word theory is misused, at least in the scientific sense, in your various classes and day to day experiences.

That said, theories are not static. New or more accurate observations that a theory cannot explain will inevitably drive the theory's revision or replacement. When this occurs, the new theory explains the new observations as well as everything explained by the older theory. Consider for example, gravity. Isaac Newton's law of gravity describes how objects behave; it is possible to make extremely accurate predictions of how objects behave using its rules. However, Newton did not really have a theory of gravity, that is, a naturalistic explanation for why gravity exists and why it behaves the way it does. He relied, in fact, on a supernatural explanation.[27] When it was shown that Newton's law of gravity failed in specific situations, such as when an object is in close proximity to a massive object like the sun, new rules and explanations were needed. Albert Einstein's Theory of General Relativity not only more accurately predicts the behavior of these systems, but also provides a naturalistic explanation for the origin of the gravitational force.[28] It has also made predictions about future observations, such as gravity waves, that have subsequently been confirmed.[29] So is general relativity true? Not necessarily, which is why scientists continue to test its predictions in increasingly extreme situations and to higher and higher degrees of accuracy.

Knowing what you know: constructing models, answers, explanations & critiques

How do we know what we know? This is a central question in philosophy and is equally relevant to teaching and learning. There is plenty of evidence that people consistently over-estimate their own skills, including what they believe they have learned in a class.[30] There is, however, a well-established approach to evaluating one's, and other's, understanding, namely the Socratic dialog. In a Socratic dialog, with an engaged and critical person, we can discover our assumptions and consider whether they are valid. We use Socratic dialog when we ask you about your answers to questions and when you consider the statements of others: is the application of scientific concepts and relevant observations relevant and logical? Are unspoken assumptions in play? You should be ready to discuss, Socratically, the answers to the "questions to answer and ponder" found throughout the book.

To answer and explain, it is important to be clear that you understand exactly what it is that the question you are being asked wants to know, or what you need to explain. The ability to read a question, accurately decode what it is asking, and to then compose a coherent and evidence-based response requires basic literacy.[31] While it may be difficult or awkward to ask for clarifications of a question, that is, what exactly is the question about, within the context of an exam, it should be your

---

[27] Want to read an interesting biography of Newton, check out "Isaac Newton" by James Gleick

[28] A good video on General Relativity [here]

[29] Physicists find another gravitational wave to suggest that Einstein was right

[30] The Kruger & Dunning effect: Unskilled and Unaware

[31] Norris & Phillips. 2003. How literacy in its fundamental sense is central to scientific literacy

first response within a particular reading or during class. Feel free to ask your own clarifying questions.[32] We will ask you to frame your question in the context of what you think the question is asking and why, exactly, you find it unclear or confusing. In a testing scenario, this can also be a useful strategy, restate what you think the question is asking and then answer that question. By using the nota bene interactive reading system, you can ask other students what they think a question is about, or you can help explain it to others. If they are equally confused, or unhelpful, ask the instructor; typically we will share both your question and our response with the entire class, since it is very likely that you are not the only person who wants or needs clarification.

Once you understand what the question wants to know, you can begin to construct your response. You need to identify what facts and general rules apply to the particular question; these will be used in the construction of your answer. As an example, consider this question: "Based on the accumulation of an isotope that is known to be generated only by radioactive decay, a geologist claims a particular rock is ~2 billion years old, while a creationist claims that a fossil within the rock is ~6000 years old. Why can't both be correct?" To answer such a question, we begin by clearly articulating to ourselves what the question is based on. Geologists date rocks, typically igneous (originally molten, often volcano-derived) based on assumptions about the rock's stability and composition. Many observations indicate that the rate and products of the radioactive decay of a particular isotope are constant and universal; they are not influenced by other factors. Assuming that the rock used to assign a date is stable, that is, no atoms enter or leave it, then the ratio of the original isotope and the isotope produced by its decay serves as an atomic clock, providing an estimate of the age of the rock, that is the time since its formation. Fossils are found in sedimentary rocks, but not volcanic ones, since the heat associated with volcanic rocks destroys organic remains. Sedimentary rocks are difficult to date accurately, since they are derived, through processes of erosion, from other older rocks. The geologist dates the fossil containing rock based on the age of the surrounding rock layers. It is less clear what scientific ideas the creationist uses to date rocks and the fossils within them. Since there is no evidence that rates of radioactive decay have changed over the history of the Universe, and assuming no other natural processes are at play (and it is hard to imagine what they might be, in any case), the creationist is most likely to be incorrect – their assumptions implicitly contradict well established knowledge from physics, chemistry, and geology.

As you can see, answering a question can be a complex process – constructing an answer can rely on a number of assumptions that need to be recognized and stated explicitly. In the case of dating a fossil, you would consider the observed rate of radioactive decay, the method used to date sedimentary (and igneous) rocks, and the mechanism(s) by which fossils are generated. Our answer needs to identify the assumptions we are making. The complexity of explaining why correct answers are correct is one of the reasons that we often ask you to explain why wrong answers, in multiple-choice type questions, are wrong or irrelevant. Typically a wrong answer is wrong for a single incorrect assumption or, if correct, is irrelevant to the question at hand.

A similar situation applies when explaining something to someone, you need to identify the various ideas, and the observations upon which those ideas are based, that the person you are talking to will

---

32 The answers can often be surprising. see McClymers & Knowles.Ersatz Learning, Inauthentic Testing

need to know to be able to understand your explanation – you probably should also determine whether they understand what you think they understand. As an example, consider the short video interview with the physicist Richard Feynman (1918-1988) [video link]; in it he describes what it takes to explain magnetic attraction.[33] At the same time, you will need to be prepared to explain those ideas – the person you are explaining something to can be expected to ask you to justify your assumptions, clarify your logic, and defend your conclusions. You are taking part in a Socratic dialog. The same applies when you are in class listening to an explanation from an instructor; do their assumptions make sense? Are they telling you all you need to know to be able to understand their explanation? Similarly, when you are listening to someone else's explanation, you need to consider whether the evidence they are using is correct and relevant, do their conclusions follow logically. In a scientific discussion, are the methods they are using capable of generating the data upon which their argument rests?

It can be helpful to study with a group of people who are comfortable questioning, and explaining to each other. But we often find ourselves called upon to learn materials on our own. You can improve this process by developing your own "inner Socrates", a voice that helps refine your thinking by asking "am I answering the question I am being asked? have I identified the key ideas and observations needed to answer the question? Are there other observations or concepts that need to be considered? Are other, simpler explanations possible?"

Questions to answer:
1. How would you use Occam's razor to distinguish between two equally accurate models?
2. What does it mean when there are two theoretical explanations for the same phenomena? How might you resolve this situation?
3. Outline your approach to deciding whether a particular idea, model, or hypothesis is scientific.

Science is social

The social nature of science is something that we want to stress yet again. While science is often portrayed as an activity carried out by isolated individuals, the image of the isolated mad scientist comes to mind (→), in fact science is an extremely social activity. It works only because it involves and depends upon an interactive community who keep each other, in the long run, honest.[34] Scientists present their observations, hypotheses, and conclusions in the form of scientific papers, where their relevance and accuracy can be evaluated, more or less dispassionately, by others.

Over the long term, this process leads to an evidence-based consensus. Certain ideas and observations are so well established that they can be reasonably accepted as universally valid, whereas others are extremely unlikely to be true, such as perpetual motion machines or "intelligent design creationism." These are ideas that can be safely ignored. As we see it, modern biology is based on a small set of theories: these include the Physicochemical Theory of Life, the Cell Theory, and the

---

[33] Feynman & magnets

[34] A good introduction of how science can be perverted is "The undergrowth of Science" by Walter Gatzer.

Theory of Evolution, to which we will return to in detail.[35] That said, as scientists we keep our minds open to exceptions and work to understand them and their implications. The openness of science means that a single person, taking a new observation or idea seriously, can challenge and change accepted scientific understanding. That is not to say that it is easy to change the way scientists think. Most theories are based on large bodies of evidence and have been confirmed on multiple occasions using multiple methods. It generally turns out that most "revolutionary" observations are either mistaken, misinterpreted, or can be explained within the context of established theories. It is, however, worth keeping in mind that it is not at all clear that all phenomena can be put into a single "theory of everything." For example, it has certainly proven difficult to reconcile quantum mechanics with the general theory of relativity.

A final point, mentioned before, is that the sciences are not independent of one another. Ideas about the behaviors of biological systems cannot contradict well established observations and theories in chemistry or physics. If they did, one or the other would have to be modified. For example, there is substantial evidence for the dating of rocks based on the behavior of radioactive isotopes. There are also well established patterns of where rock layers of specific ages are found. When we consider the dating of fossils, we use rules and evidence established by geologists. We cannot change the age we assign to a fossil, making it inconsistent with the rocks that surround it, without challenging our understanding of the atomic nature of matter, the quantum mechanical principles involved in isotope stability, or a range of geological mechanisms. A classic example of this situation arose when the physicist William Thompson (1824-1907), also known as Lord Kelvin, estimated the age of the Earth to be between ~20 to 100 million years, based on the assumption that the Earth was once completely molten together with the known rate of heat dissipation of such a molten object.[36] This was a time-span that seemed too short for a number of geological and evolutionary processes, and greatly troubled Darwin. Somebody was wrong, or better put, their understanding was incomplete. The answer was with the assumptions that Kelvin made (one reason to closely examine the assumptions upon which ideas are based); his calculations ignored the effects of radioactive decay, not surprising since radioactivity had yet to be discovered. Including the heat released by radioactive decay in their calculations led to an increase in the estimated age of the earth by more than ten to one hundred fold, to ~5 billion years, an age compatible with both evolutionary and geological processes.

Teaching and learning science

An important point to appreciate about science is that because of the communal way that it works, understanding builds by integrating new observations and ideas into a network of others' ideas and observations. Following this discipline, science often arrives at conclusions that can be strange, counterintuitive, and sometimes disconcerting but that are nevertheless logically unavoidable. While it is now accepted that the Earth rotates around its axis and revolves around the sun, which is itself moving around the center of the Milky Way galaxy, and that the Universe as a whole is expanding at what appears to be an ever increasing rate, none of these facts are immediately obvious and relatively few people who believe or accept them would be able to explain how we have come to know that these

---

[35] Thinking about the conceptual foundations of the biological sciences: http://www.ncbi.nlm.nih.gov/pubmed/21123685

[36] An interesting book on this topic is "Discarded Science: Ideas That Seemed Good at the Time" by Paul Barnett

ideas accurately reflect the way the universe is organized. At the same time, when these ideas were first being developed they conflicted with the assumption that the Earth was stationary, which, of course it appears to be, and that it is located at the center of a static Universe, which seems quite reasonable. Scientists' new conclusions about the Earth's actual position in the Universe were seen as a threat to the sociopolitical order. A number of people were persecuted for holding "heretical" views on the topic. Most famously, the mystic Giordano Bruno (1548 –1600) was burnt at the stake for holding these and other ideas, some of which are similar to those currently being proposed by modern physicists. Galileo Galilei (1564–1642), known as the father of modern physics, was arrested in 1633, tried by the Roman Catholic Inquisition, forced to publicly recant his views on the relative position of the Sun and Earth, and spent the rest of his life under house arrest.[37] In 1616 the Church placed Galileo's book, which held that the sun was the center of the solar system, on the list of forbidden books – it remained there until 1835.

The idea that we are standing on the surface of a planet that is rotating at ~1000 miles an hour and flying through space at ~67,000 miles per hour is difficult to reconcile with our everyday experience, yet science continues to generate even weirder ideas. Based on observations and logic, it appears that the Universe arose from "nothing" ~13.8 billion years ago.[38] Current thinking suggests that the Universe will continue to expand forever at an increasingly rapid rate. Einstein's theory of general relativity implies that matter distorts space-time, which is really one rather than two discrete entities, and that this distortion produces the attraction of gravity and leads to black holes. A range of biological observations indicate that all organisms are derived from a single type of ancestral cell that arose from non-living material between 3.5 to 3.8 billion years ago. There appears to be an uninterrupted link between that cell and every cell in your body, and to the cells within every other living organism. You yourself are a staggeringly complex collection of cells. Your brain and its associated sensory organs, which act together to generate consciousness and self-consciousness, contains ~86 billion (10^9) neurons as well as a similar number of non-neuronal (glial) cells. These cells are connected to one another through ~1.5×10^{14} connections, known as synapses.[39] How exactly such a system produces thoughts, ideas, dreams, feelings, and self-awareness remains obscure, but it appears that these are all emergent behaviors that arise from this staggeringly complex natural system. Scientific ideas, however weird, arise from the interactions between the physical world, our brains, and the social system of science that tests ideas based on their ability to explain and predict the behavior of the observable universe.

Understanding scientific ideas

One of the difficulties in understanding scientific ideas and their implications is that these ideas build upon a wide range of observations and are intertwined with one another. One cannot really understand biological systems without understanding the behavior of chemical reaction systems, which in turn requires an understanding of molecules, which rests upon an understanding of how atoms (matter) and energy behave and interact. It is our working premise that to understand a topic, or a discipline, it is necessary to know the key observations and common rules upon which basic

---

[37]The History, Philosophy, and Impact of the Index of Prohibited Books: http://www.unc.edu/~dusto/dusto_prague_paper.pdf

[38] The Origin Of The Universe: From Nothing Everything?

[39] Are There Really as Many Neurons in the Human Brain as Stars in the Milky Way? & http://onlinelibrary.wiley.com/doi/10.1002/cne.21974/abstract

conclusions and working concepts are based. To test one's understanding of a system, you need to be able to construct plausible claims for how, and why the system behaves the way it does, and how various perturbations can be expected to influence it; your analysis needs to be based on facts, observations, or explicit presumptions that logically support your claim. You also need to present your model to others, knowledgeable in the topic, to get their feedback, to answer rather than ignore or disparage their questions, and address their criticisms and concerns.[40] Sometimes you will be wrong because your knowledge of the facts is incomplete or inaccurate, your understanding or application of general principles is incorrect, or your logic is faulty. It is important to appreciate that generating coherent scientific explanations and arguments takes time. We hope to help you learn how to understand biological systems and processes through useful coaching and practice. In the context of various questions, we, and your fellow students, will attempt to identify when you produce a coherent critique, explanation or prediction, and where you fall short. Our goal is to help you learn how to think accurately and Socratically about biological systems.

Distinguishing the scientific from the trans-scientific

When we consider various personal and public policy decisions, including the ramifications of global warming, and what to do about it, the genetic engineering of human embryos and other organisms, and more generally the use of genetic data in medicine and society, as well as the costs and benefits of various science-informed decisions, we are often told that science has reached a consensus, but what exactly does that mean? By consensus, we mean the common conclusions accepted by scientists working in the field, conclusions supported by available evidence – what we might term "working knowledge". But evidence is rarely complete; for example, measurements can always be more accurate. In addition, when approaching a system scientifically, it is often necessary to make simplifying assumptions. This simplifying assumptions make the system tractable, they make it possible to make the kinds of unambiguous predictions upon which science is based. But when we want to act on scientific conclusions on complex systems such as the human brain and body, Earth's climate, or the response of individuals to specific medical treatments, we find that outcomes are less predictable. How a particular person responds to a particular drug is influenced by many factors, not all of which are perfectly defined. The limits of our understanding mean that interventions have side-effects, both desirable and undesirable. Only treatments that do nothing, homeopathy comes to mind, have no effects[41] and may leave a serious condition untreated.[42] The risks of taking a drug, getting a vaccination, undergoing a surgery, opening or closing nuclear (or coal-based) power plants are outweighed by their benefits, but knowing **exactly** what the costs and benefits are may be impossible.

Moreover, such a cost-benefit analysis, when applied to political, social, or economic decisions, involves non-scientific factors? Consider, for example, the interconnected issues of increasing population, poverty, industrialization, and the ecological impacts of humans. One can argue, rather

---

[40] This is exact opposite of the alt-fact environment that appears to be all the rage these days.

[41] Because homeopathic remedies are in most cases water or other inert chemicals. As we go along, given what we know about the movement of molecules and their constant collisions, you can probably explain why, for homeopathy to work, many laws of physics and chemistry would have to be broken.

[42] The case of Steve Jobs and his pancreatic cancer is a case in point. see link

convincingly, that bringing basic human rights and autonomy, together with access to contraception, to women will help control human population growth – it has already led to reduced populations (fewer children per person) in much of the world.[43] At the same time, the idea of female autonomy can be deeply troubling (divisive) in certain cultures. There are potential economic effects, such as the extent to which women enter the work-force, and how that might impact cultural dynamics and stability. What is, exactly, the cost of female autonomy in terms of social cohesion and conflict? on personal happiness and political stability? While sensible answers may rely on input from the sciences, they are not scientific questions, they are trans-scientific. Similarly, in the context of evolutionary processes, every adaptation involves an inherent cost-benefit calculation, a design trade-off, opportunity's gained and curtailed, with the final decision based on reproductive success.[44] There are no perfect solutions, just compromises that work more or less well. When we think about biological systems and processes, we need to keep this trade-off / cost-benefit calculation in mind.

Questions to answer:

4. A news story reports that spirit forces influence the weather. Produce a set of questions whose answers would enable you to decide whether the report was scientifically plausible.
5. If "science" concludes that free will is an illusion, would you be wise or silly to start behaving like a machine?
6. How would you describe the major differences between scientific thinking in physics and biology?

Questions to ponder

- Is attaining "truth" and developing a theory of everything the goal of science?
- How should we, as a society, deal with the tentative nature of scientific knowledge?
- What distinguishes scientific from trans-scientific conclusions?
- What factors determined how people and governments should act in the face of scientific evidence?

---

[43] Hans Rosling: Don't Panic – The Facts About Population

[44] Weinstein. Evolutionary trade-offs as a central organizing principle in biology

Chapter 2: Life and its origins

In which we consider what biology is all about, namely the study of organisms and their diversity. We will discover that organisms are built of one or more, sometimes many (millions to billions) cells. Social processes are involved in multicellular organisms and when single-celled organisms act in a coordinated manner. We consider plausible models for the origins of organisms, their basic properties, and their relationships to one another.

Biology is the science of organisms, how organisms function, behave, interact, vary genetically from one another, adapt, and, as populations, evolve over time. As we will see, organisms are discrete, highly organized, bounded but open, non-equilibrium, physicochemical systems. Now that is a lot of words, so the question is what do they all mean? How is a rock different from a mushroom that looks like a rock? What is genetic variation and how does it influence the properties and behavior of an organism? What exactly is a bounded, non-equilibrium system? The answers are not simple; they assume a working knowledge of core concepts and observations. For example, to understand what it means to be a "bounded, non-equilibrium system" you need to understand basic thermodynamics, a topic that we will address in some detail in Chapter 5. For the moment, when we talk about a non-equilibrium system, we mean a system that can do various forms of work. Of course we then need to define what we mean by work. For simplicity, we will start by defining work as some outcome that takes the input of energy to achieve. In the context of biological systems, work ranges from generating and maintaining molecular gradients and driving a range of unfavorable, that is energy-requiring reactions, such as the synthesis of a wide range of biomolecules, including nucleic acids, proteins, lipids, and carbohydrates, required for growth, reproduction, movement, and so on.

We will focus on what is known as Gibbs free energy, which is energy available to make things happen, that is, to do work. When a system is at equilibrium its free energy is 0, which means that no macroscopic (visible) or net changes are possible. While static at the macroscopic level, at the molecular level there is constant movement and change because, at all temperatures above absolute zero, molecular systems have kinetic energy which manifests as movement and vibrations. Organisms maintain their non-equilibrium state, that is, their Gibbs free energy is much greater than zero, by importing energy in various forms from the external world. Organisms are different from other non-equilibrium systems in that they contain a genetic, heritable component. While other types of non-equilibrium systems occur – hurricanes and tornados are non-equilibrium systems – they differ from organisms in that they are transient. They arise *de novo*, they do not have "parents", and when they dissipate they leave no offspring, no baby hurricanes. In contrast, each organism alive today arose from one or more pre-existing organisms, its parent(s), and each organism, with some exceptions, has the ability to produce offspring. As we will see, the available evidence indicates that each and every organism, past, present, and future, has, or will have, an uninterrupted history stretching back billions of years. This is a remarkable conclusion, given the obvious fragility of life, and makes organisms unique among physicochemical systems.

Biology has only a few over arching theories. One of these, the Cell Theory of Life, explains the historic continuity of organisms, while the Theory of Evolution by Natural Selection (and other processes), explains both the diversity of organisms and how populations of organisms change over time. Finally, the Physicochemical Theory of Life explains how it is that organisms can display their remarkable properties without violating the laws that govern all physical and chemical systems.[45]

What is life, exactly?

Clearly, if we are going to talk about biology, and organisms and cells and such, we have to define exactly what we mean by life. This raises a problem peculiar to biology as a science. We cannot define life generically because we know of only one type of life. While you might think that we know of many different types of life, from mushrooms to whales, from humans to the bacterial communities growing on the surfaces of your teeth (that is what dental plaque is, after all), we will discover that the closer we look the more these different "types of life" are in fact all versions of a common underlying motif, they represent versions of a single type of life. Based on their common chemistry, molecular composition, cellular structure, and the way that they encode, read, and use hereditary information in the form of molecules of deoxyribonucleic acid (DNA), all topics we will consider in depth as we go on, there is no reasonable doubt that all organisms are related to one another, they are descended from a common ancestor. We do not know whether this type of life is the only type of life possible or whether radically different forms of life exist elsewhere in the universe or even on Earth, in as yet to be recognized forms.

We cannot currently answer the question of whether the origin of life is a simple, likely, and predictable event given the conditions that existed on the early Earth when life first arose, or whether it is an extremely rare and unlikely event. In the absence of empirical data, one can question whether scientists are acting scientifically, or more as lobbyists for their own pet projects, when they talk about doing astrobiology or speculating on when and where we will discover alien life forms. That said, asking seemingly silly questions, provided that empirically-based answers can be generated, is a critical driver of scientific progress. Consider, for example, current searches for life on Earth, almost all of which are based on what we already know about life on Earth. Specifically, most of the methods used rely on the fact that all known organisms use DNA to encode their genetic information. These methods would not be expected to recognize dramatically different types of life, if they exist. They would not detect organisms that used a non-DNA-based mechanism to encode genetic information. If we could generate living systems *de novo* in the laboratory we would have a better understanding of what functions are necessary for life and how to look for possible "non-standard" organisms using more appropriate methods. New methods might even lead to the discovery of alternative forms of life right here on Earth, assuming they exist.[46] That said, until someone manages to create or identify such non-standard forms of life, it seems reasonable to concentrate on the characteristics of life as we know them.

So, let us start again in trying to produce a good definition, or given the fact that we know only of one version of life, a useful description of what we mean by life. First, the core units of life are organisms, which are individual living objects. From a structural and thermodynamic perspective, each organism is a bounded, non-equilibrium system that persists over time and, from a practical point of

---

[45] Thinking about the [conceptual foundations of the biological sciences](#)

[46]The [possibility of alternative microbial life on Earth](#)   [Signatures of a shadow biosphere](#)   [Life on Earth but not as we know it](#)

view, can produce one or more copies of itself. Even though organisms are composed of one or more cells, it is the organism that is the basic unit of life. It is the organism that produces new organisms.[47]

Why the requirement for and emphasis on reproduction? This is basically a pragmatic criterion. Assume that a non-reproducing form of life was possible. Any such system runs the risk of death, or perhaps better put, extinction, by accident. Over time, the probability of death for any individual approaches one – that is, certainty ($\rightarrow$).[48] In contrast, a system that can reproduce makes multiple copies of itself and so minimizes, although by no means eliminates, the chance of extinction by accident, that is, the death of all of their descendants. We see the value of this strategy when we consider the history of life. Even though there have been a number of mass extinction events over the course of life's history,[49] organisms descended from a single common ancestor that appeared billions of years ago continue to survive and flourish.
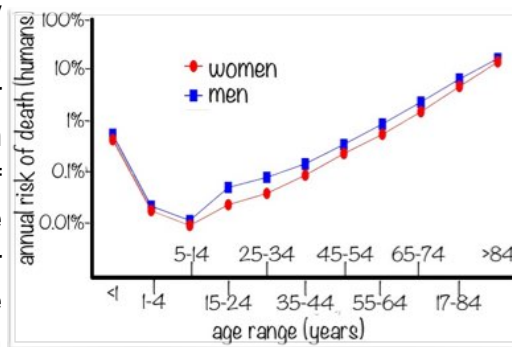
So what does the open nature of biological systems mean? Basically, organisms need to be able to import, in a controlled manner, energy and matter from outside of themselves and to export waste products into their environment.[50] This implies that there is a distinct boundary between the organism and the rest of the world. All organisms have such a barrier (boundary) layer, as we will see later on. The basic barrier of organisms appears to be a homologous structure–that is, it was present in and inherited from their common ancestor. As we will see, the importation of energy, specifically energy that can be used to drive various cellular processes, is what enables the organism to maintain its non-equilibrium state and its dynamic structure, and to grow and reproduce. The boundary must be able to retain the valuable molecules generated, while at the same time allow waste products to leave. This ability to selectively import matter and export waste enables the organism to grow and to reproduce. While we assume that you have at least a basic understanding of the laws of thermodynamics, we will review the central ideas in Chapter 5.

We find evidence of the non-equilibrium nature of organisms most obviously in their ability of move, but it is important for all aspects of the living state. In particular, organisms use energy, captured from their environment, to drive a wide range of thermodynamically unfavorable chemical reactions. These reactions are driven by coupling them to thermodynamically favorable reactions. An organism that reaches thermodynamic equilibrium is dead.

There are examples of non-living, non-equilibrium systems that can "self-organize"; these appear de novo. Hurricanes and tornados form spontaneously and then disperse. They use energy from their environment, which is then released back into the environment, a process associated with increased entropy. These non-living systems differ from organisms in that they do not produce offspring - they are

---

[47] In Chapter 4, we will consider how multicellular and social organisms come to be.

[48] Image modified from "risk of death" graph: http://www.medicine.ox.ac.uk/bandolier/booth/Risk/dyingage.html

[49] Mass extinction events

[50] In fact, this is how they manage to organize themselves, by exporting entropy. So be careful when people (or companies) claim to have a zero-waste policy, which is an impossibility according to the laws of thermodynamics that all systems obey.

the result of specific atmospheric conditions. They are individual entities, unrelated to one another; they do not and cannot evolve. Tornados and hurricanes that formed billions or millions of years ago would, if we could observe them, be similar to those that form today. Since we understand, more or less, the conditions that produce tornados and hurricanes, we can predict, fairly reliably, the conditions that will lead to their appearance and how they will behave once formed. In contrast, organisms present in the past were different from those that are alive today. The further into the past we go, the more different they appear. Some ancient organisms became extinct, some gave rise to the ancestors of current organisms. In contrast, modern tornados and hurricanes originate anew, they are not derived from parental storms.

Questions to answer:

7. How might you decide whether a particular object is alive or not?
8. Using the graph on risk of death as a function of age in humans, provide a plausible explanation for the shape of the graph. In your answer, consider that factors that influence the various regions of the curve and how they might be different for different types of organisms?
9. Consider the effects of population size on the curve; would the graph be exactly the same if you collected data from different populations?

Questions to ponder:

- Should be the points in the graph be connected?

The cell theory and the continuity of life

Toward the end of the 1800's, observations using microscopes revealed that all organisms examined contained structurally similar "cells." Based on such observations, a rather sweeping conclusion, the Cell Theory, was formulated by naturalists. The Cell Theory has two distinct parts. The first is the prediction that every organism is composed of one or more, and in some cases millions to billions, of cells together with non-cellular products produced by cells, such as bone, hair, scales, and slime. The cells that the Cell Theory deals with are membrane-bounded, open, non-equilibrium physicochemical systems, a definition much like that for life itself. Over the course of all these observations (up to the present day) there is no evidence that modern cells can be formed from non-cellular materials. Therefore the second part of the Cell Theory is that cells arise only from pre-existing cells. The implication is that organisms, and the cells that they are composed of, arise in this way and no other. The Cell Theory does not say anything about how life on Earth originated.

We now know, and will consider in greater detail as we proceed, that in addition to their basic non-equilibrium nature, cells also contain a unique material that encodes hereditary information in a physical and relatively stable form, namely molecules of double-stranded deoxyribonucleic acid (DNA). Based on a large body of data, the Cell Theory implies that all organisms currently in existence, and the cells that compose them, are related through an unbroken series of cell division events that stretch back in time. Other studies, based on the information present in DNA molecules, as well as careful comparisons of how cells are constructed at the molecular level, suggests that there was a single common ancestor for all life and that this organism lived between ~3.5 to ~3.8 billion years ago. This is a remarkable conclusion, given the fragility of life. It implies that each cell in every currently living organism, including you, has an uninterrupted multibillion year old history. What the Cell Theory does

not address is the processes that led to the origin of the ancestral organism (cell).

The earliest events in the origin of life, that is, exactly how the first cells were formed and what they looked like, are unknown although there is plenty of speculation to go around. Our confusion arises in large measure from the fact that the available evidence indicates that all organisms that have ever lived on Earth share a single common ancestor, and that that ancestor, likely to be a singled-cell organism, was quite complex. We will discuss how we come to these conclusions, and their implications, later on in this chapter. One rather weird point to keep in mind is that the "birth" of a new cell is a continuous process by which one cell becomes two. Each cell is defined, in part, by the presence of a distinct surface barrier, known as the cell or plasma membrane. The new cell is formed when that original membrane pinches off to form two distinct cells (→). The important point here is that there is no discontinuity, the new cell does not spring into existence but rather emerges from the preexisting cell. This continuity, from cell to cell, extends back in time for billions of years. We often define the start of a new life with the completion of cell division, or in the case of sexually reproducing organisms, including humans, the fusion of an egg cell and a sperm cell. But again there is no discontinuity, both egg cell and sperm cell are derived from other cells and when they fuse, the result is a new hybrid cell. In the modern world, all cells, and the organisms they form, emerge from pre-existing cells and inherit from those cells both their cellular structure, the basis for the non-equilibrium living system, and their genetic material, their DNA. When we talk about cellular or organismic structures, we are in fact talking about information present in the living structure, information that is lost if the cell/organism dies. The information stored in DNA molecules, known as an organism's genotype, is more stable than the organism itself; it can survive the death of the organism, at least for a while. In fact, information-containing DNA molecules can move between unrelated cells or from the environment into a cell, a process known as horizontal gene transfer, which we will consider in detail later on. In fact DNA is being explored as a high-density, high-stability data storage system, outside of organisms.[51]

The organization of organisms

Some organisms consist of a single cell, while others are composed of many cells, often many distinct types of cells. These cells vary in a number of ways and can be extremely specialized, particularly within the context of multicellular organisms, yet they are all clearly related to one another, sharing many molecular and structural details. So why do we consider the organism rather than the cell to be the basic unit of life? The distinction may seem trivial or arbitrary, but it is not. It is a matter of reality versus abstraction. It is organisms, whether single- or multi-cellular, that produce new organisms. As we will discuss in detail when we consider the origins of multicellular organisms, a cell within a multicellular organism normally cannot survive outside the organism nor can it produce a new organism – it depends upon cooperation with the other cells of the organism. In fact, each multicellular organism

---

[51] A DNA-Based Archival Storage System

is an example of a cooperative, highly integrated social system. The cells of a typical multicellular organism are part of a social system in which most cells have given up their ability to reproduce a new organism; their future depends upon the reproductive success of the organism of which they are a part. It is the organism's success in generating new organisms that underlies evolution's selective mechanisms. Within the organism, the cells that give rise to the next generation of organisms are known as germ cells, those that do not, that is, the cells that die when the organism dies, are known as somatic cells.[52] All organisms in the modern world and, apparently the last ~3.5-3.8 billion years, arise from a pre-existing organism or, in the case of sexually reproducing organisms, from the cooperation of two organisms, an example of social evolution that we will consider in greater detail in Chapter 4. We will also see that breakdowns in such social systems can lead to the death of the organism or the disruption of the social system. Cancer is the most obvious example of an anti-social behavior; in evolutionary terms, it can, initially, be rewarded (more copies of the cancerous cell are produced) but ultimately leads to the extinction of the cancer, and often the death of the organism within which the cancer occurs.[53] This is because evolutionary mechanisms are not driven by long term outcomes, but only immediate cost-benefit "calculations", revealed in terms of reproductive success.

Spontaneous generation and the origin of life

The ubiquity of organisms raises obvious questions: how did life start and what led to all these different types of organisms? At one point, people believed that these two questions had a single answer, but we now recognize that they are really two quite distinct questions and their answers involve distinct mechanisms. An early view, held by those who thought about such things, was that supernatural processes were necessary to produced life in general and human beings in particular. The articulation of the Cell Theory and the Theory of Evolution by Natural Selection, which we will discuss in detail in the next chapter, together with an accumulation of data enables us to conclude, quite persuasively, that life had a single successful origin, that only natural processes were involved, and that various (again natural) processes generated the diversity of life.

But how did life itself originate? It was once widely accepted that various types of organisms, such as flies, frogs, and even mice, could arise spontaneously, from non-living matter.[54] Flies, for example, were thought to appear from rotting flesh and mice from wheat. If true, on-going spontaneous generation would have profound implications for our understanding of biological systems. For example, if spontaneous generation based on natural processes was common, there must be a rather simple process at work, a process that presumably can produce remarkably complex outcomes. In contrast, all bets are off if the process is supernatural. If each organism arose independently, we might expect that at the molecular level details of each would be unique, since they presumably arose independently from different stuff and under different conditions compared to other organisms. However, we know this is

----

[52] If we use words that we do not define and that you do not understand, look them up or ask your instructor!

[53] Cancer cells as sociopaths: cancer's cheating ways Recently the situation has gotten more complex with the recognition of transmissible cancers and http://www.ncbi.nlm.nih.gov/pubmed/19956175

[54] Farley. The spontaneous generation controversy (1700-1860): The origin of parasitic worms. and The spontaneous generation controversy (1859-1880): British and German reactions to the problem of abiogenesis.

not the case, since all organisms are clearly related and can be traced back to a single ancestor, a conclusion to which we return, repeatedly.

A key event in the conceptual development of modern biology was the publication in 1668 of Francesco Redi's (1626–1697) paper entitled "Experiments on the Generation of Insects". His hypothesis (informed guess) was that spontaneous generation did not occur.[55] He thought that the organisms that appeared had developed from "seeds" deposited by adults, an idea that led to a number of predictions. One was that if adult flies were kept away from rotting meat maggots, the larval form of flies, would never appear no matter how long one waited. Similarly, the type of organism that appeared would depend not on the type of rotting meat, but rather on the type of adult fly that had access to the meat. To test his hypothesis Redi set up two sets of flasks both of which contained meat. One set of flasks was exposed directly to the

> *He who experiments increases knowledge. He who merely speculates piles error upon error.*
> *- Arabic epigraph quoted by Francisco Redi.*

air and so to flies, the other was sealed with paper or cloth. Maggots appeared only in the flasks open to the air. Redi concluded that organisms as complex as insects, and too large to pass through the cloth, could arise only from other insects, or rather eggs laid by those insects – that life was continuous, that is, life came from life.

The invention of the light microscope and its use to look at biological materials by Antony van Leeuwenhoek (1632-1723) and Robert Hooke (1635-1703) led to the discovery of a completely new and totally unexpected world of organisms, known as microbes or microscopic organisms. We now know these as the bacteria, archaea, and a range of unicellular photosynthetic and non-photosynthetic eukaryotes.[56] Although it was relatively easy to generate compelling evidence that macroscopic (that is, big) organisms, such as flies, mice, and people could not arise spontaneously, it seemed plausible that microscopic and presumably much simpler organisms could form spontaneously.

The discovery of microbes led a number of scientists to explore their origin and reproduction. Lazzaro Spallazani (1729-1799) showed that after a broth was boiled it remained sterile, that is, without life, as long as it was isolated from contact with fresh air. He concluded that microbes, like larger organisms, could not arise spontaneously but were descended from other microbes, many of which were floating in the air. Think about possible criticisms to this experiment – perhaps you can come up with ones that we do not mention!

One obvious criticism was that it could be that boiling the broth destroyed one or more key components that were necessary for the spontaneous formation of life. Alternatively, perhaps fresh air was the "vital" ingredient. In either case, boiling and isolation would have produced an artifact that obscured rather than revealed the true process. In 1862 (after Charles Darwin had published On the Origin of Species in 1859), Louis Pasteur (1822-1895) carried out a particularly convincing set of experiments to address both of these concerns. He sterilized broths by boiling them in special "swan-necked" flasks. What was unique about his experimental design was the shape of the flask neck; it allowed air but not air-borne microorganisms to reach the broth. Microbes in the air were trapped in the bended region of the flask's neck (↓). This design enabled Pasteur to address a criticism of previous

---

[55] see Richard Feynman's description of the role of guessing in the scientific process

[56] see the wikipedia article on protists

experiments, namely that access to air was necessary for spontaneous generation to occur. He found that the liquid, even with access to air, remained sterile for months. However, when the neck of the flask was broken the broth was quickly overrun with microbial growth. He interpreted this observation to indicate that air, by itself, was not necessary for spontaneous generation, but rather was normally contaminated by microbes. On the other hand, the fact that the broth could support microbial growth after the neck was broken served as what is known as a "positive control" experiment; it indicated that the heating of the broth had not destroyed some vital element needed for standard growth to occur. We carry out positive control experiments to test our assumptions; for example, if we are using a drug in a study, we first need to establish that the sample of the drug we are using is active. In Pasteur's experiment, if the boiled broth could not support growth (after the flask was broken) we would not expect it to support spontaneous generation, and so the experiment would be meaningless. We will return to the description of a "negative control" experiment later.[57]
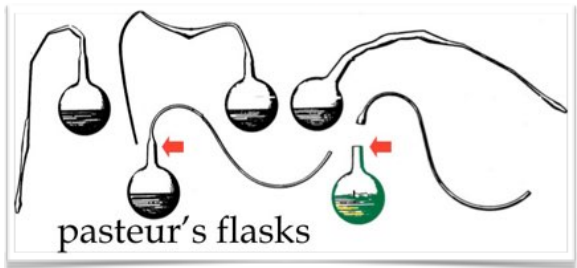
Of course, not all, in fact, probably not any experiment is perfect, nor does it have to be for science to work. For example, how would one argue against the objection that the process of spontaneous generation normally takes tens to thousands, or millions, of years to occur? If true, this objection would invalidate Pasteur's conclusions. Clearly an experiment to address that particular objection has its own practical issues. Nevertheless, the results of various experiments on spontaneous generation have led to the conclusion that neither microscopic nor macroscopic organisms can arise spontaneously, at least not in the modern world. The problem, at least in this form, became uninteresting to working scientists.

So what explains the absence of spontaneous generation in the modern world, or in a world in which life (organisms) already exist? Consider the fact that living systems are complex chemical reaction networks. In the modern world, there are many organisms around, essentially everywhere, who are actively eating complex molecules to maintain their non-equilibrium state, to grow and to reproduce. If life were to arise by a spontaneous but natural process, it is possible that it could take thousands to hundreds of millions of years to occur. We can put some limits on the minimum time it could take from geological data using the time from when the Earth's surface solidified from its early molten state to the first fossil evidence for life, about 100 to 500 million years. Given the tendency of organisms to eat one another, one might argue (as Darwin did →) that once organisms had appeared in a particular environment they would suppress any subsequent spontaneous generation event – they would have eaten the molecules needed for the process to occur. But, as we will see, evolutionary processes have led to the presence of organisms essentially everywhere on Earth that life can survive – there are basically no

It is often said that all the conditions for the first production of living organisms are now present. But if (and oh! what a big if!) we could conceive in some warm little pond, with all sorts of ammonia and phosphoric salts, light, heat, electricity, etc. present, that a proteine compound was formed, ready to undergo still more complex changes, at the present day such matter would be instantly devoured or absorbed, which would not have been the case before living creatures were formed. - Charles Darwin (1887).

---

[57] Wikipedia on control experiments and observations

welcoming and sterile, that is, life-less places left within the modern world. Here we see the importance of history. According to the current scientific view, life could arise *de novo* only in the absence of life; once life had arisen, the conditions had changed. The presence of life is expected to suppress the origin of new forms of life. Once life was present, only its descendants could survive. In such a system, history matters.

The death of vitalism

Naturalists originally thought that life itself was a type of supernatural process, too complex to obey or be understood through the laws of chemistry and physics.[58] In this vitalistic view, organisms were thought to obey different laws from those acting in the non-living world. For example, it was assumed that molecules found only in living organisms, known as organic molecules, could not be synthesized outside of an organism; they <u>had</u> to be made by a living organism. In 1828, Friedrich Wöhler (1800–1882) challenged this view by synthesizing urea in the laboratory. Urea is a simple organic molecule, $O=C(NH_2)_2$ found naturally in the waste derived from living organisms. Urine contains lots of urea. Wöhler's *in vitro* or "in glass", as opposed to *in vivo* or "in life", synthesis of urea was simple. In an attempt to synthesize ammonium cyanate (NH_4NCO), he mixed the inorganic compounds ammonium chloride (NH_4Cl) and silver cyanate ($AgNCO$). Analysis of the products of this reaction revealed the presence of urea. What actually happened was this reaction:

$$AgNCO + NH_4Cl \rightarrow NH_4NCO + AgCl \rightarrow O=C(NH_2)_2 + AgCl.$$
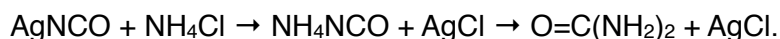
Please do not memorize this reaction! What is important here is to recognize that this is a chemical reaction between two compounds that are not derived from living systems. The point here is that the urea derived from this "inorganic" reaction is identical to the naturally occurring urea found in urine.

While simple, Wöhler's *in vitro* synthesis of urea had a profound impact on the way scientists viewed so called organic processes. It suggested that there was nothing supernatural involved, the synthesis of urea was a standard chemical process. Based on this and similar observations on the *in vitro* synthesis of other, more complex organic compounds, we (that is, scientists) are now comfortable with the idea that all molecules found within cells can, in theory at least, be synthesized outside of cells, using appropriate procedures. This is not to say that all such molecules have been synthesized *in vitro*; it means that we assume that given enough effort they could be. Organic chemistry has been transformed from the study of molecules found in organisms to the study of molecules containing carbon atoms. A huge amount of time and money is devoted to the industrial synthesis of a broad range of organic molecules that are used for purposes as diverse as pharmaceuticals to the synthesis of polymers to energy transfer and storage.

Questions to answer:
10. What does the result of a positive control experiment tell you?
11. Why did the discovery of bacteria reopen the debate on spontaneous generation?
12. Explain how Wöhler's synthesis of urea transformed thinking about organic molecules.
13. What types of evidence would support the view that the origin of life (or consciousness) requires supernatural intervention?

---

[58] In a sense this is true since many physicists at least do not seem to understand biology.

Questions to ponder:
- Is the assumption of spontaneous generation inherently unscientific? Explain your reasoning.
- Can you imagine an observation that would lead scientists to reject the naturalistic perspective?

Thinking about life's origins

There are at least three possible approaches to the study of life's origins. A religious (i.e., non-scientific) approach would likely postulate that life was created by a supernatural being. Different religious traditions differ as to the details of this event, but since the process is supernatural it cannot, by definition, be studied scientifically. Nevertheless, intelligent design creationists often claim that we can identify those aspects of life that could not possibly have been produced by natural processes, by which they mean various evolutionary and molecular mechanisms; we will discuss these processes throughout the book, and more specifically in the next chapter. It is important to consider whether these claims would, if true, force us to abandon a scientific approach to the world around us in general, and the origin and evolution of life in particular. Given the previously noted interconnectedness of the sciences, one might well ask whether a supernatural (intelligent design) biology would not also call into question the validity of all scientific disciplines. For example the dating of fossils is based on geological and astrophysical (cosmological) evidence for the age of the Earth and the Universe, which themselves are based on physical and chemical observations and principles. A truly non-scientific biology would be incompatible with a scientific physics and chemistry. The lesson of history, however, is different. Predictions as to what is beyond the ability of science to explain have routinely been found to be wrong, often only a few years after such predictions were made! This speaks to the power of science and the technologies based on science; for example, would an intelligent design creationist try to synthesize human proteins in bacteria, something now done routinely to make a range of drugs, such as insulin?[59] Would they predict that genetic modifications could make it possible to transplant pig hearts (and other organs) in the people?[60]

Another type of explanation for the appearance of life on Earth, termed panspermia, assumes that advanced aliens brought (or left) life on Earth. Perhaps we owe our origins to casually discarded litter from these alien visitors. Unfortunately, the principles of general relativity, one of the best confirmed of all scientific theories, limit the speed of travel. Given the size of the Universe, travelers from beyond the solar system seem unlikely, if not totally impossible. More to the point panspermia postpones but does not answer the question of how life began. Our alien visitors must have come from somewhere and panspermia does not explain where they came from. Given our current models for the history of the Universe, understanding the origin of alien life is really no simpler than understanding the origin of life on Earth. On the other hand, if there is life on other planets or the moons in our solar system, and we can retrieve and analyze it, it would be extremely informative, particularly if it were found that this extra-terrestrial life originated independently from life on Earth, rather than being transferred from Earth through various astronomical impact events.[61]

---

[59] Making human insulin in bacteria

[60] New life for pig-to-human transplants

[61] Top 5 Bets for Extraterrestrial Life in the Solar System: http://www.wired.com/wiredscience/2009/01/et-life/

Experimental studies on the origins of life

One strategy to understanding how life might have arisen naturally involves experiments to generate plausible precursors of living systems in the laboratory. The experimental studies carried out by Stanley Miller (1930-2007) and Harold Urey (1893-1981) were an early and influential example of this approach.[62] These scientists made an educated, although now apparently incorrect, guess as to the composition of Earth's early atmosphere. They assumed the presence of oceans and lightning. They set up an apparatus to mimic these conditions and then passed electrical sparks through their experimental atmosphere. After days they found that a complex mix of compounds had formed; included in this mix were many of the amino acids found in modern organisms, as well as lots of other organic molecules. Similar experiments have been repeated with other combinations of compounds, more likely to represent the environment of early Earth, with similar results: various biologically important organic molecules accumulate rapidly.[63] Quite complex organic molecules have been detected in interstellar dust clouds, and certain types of meteorites have been found to contain complex organic molecules. Between ~4.1 through ~3.9 billion years ago, a time known as the period of the heavy bombardment, meteorite impacts with the Earth could have supplied substantial amounts of organic molecules.[64] It therefore appears likely that early Earth was rich in organic molecules, which are, remember, carbon containing rather than life-derived molecules, the building blocks of life.

Given that the potential building blocks for life were present, the question becomes what set of conditions were necessary and what steps led to the formation of the first living systems? Assuming that these early systems were relatively simple compared to modern organisms, or the common ancestor of life for that matter, we hypothesize that the earliest proto-biotic systems were molecular communities of chemical reactions isolated in some way from the rest of the outside world. This isolation or selective boundary was necessary to keep the system from dissolving away (dissipating). One possible model is that such systems were originally tightly associated with the surface of specific minerals and that these mineral surfaces served as catalysts, speeding up important reactions; we will return to the role of catalysts in biological systems later on. Over time, these pre-living systems acquired more sophisticated boundary structures (membranes) and were able to exist free of the mineral surface, perhaps taking small pieces of the mineral with them.[65]

The generation of an isolated but open system, which we might call a protocell, was a critical step in the origin of life. Such an isolated system has important properties that are likely to have facilitated the further development of life. For example, because of the membrane boundary, changes that occur within one such structure will not be shared with neighboring systems. Rather, they would accumulate in, and favor the survival of, one system over its neighbors. Such systems could also reproduce in a crude way by fragmentation. If changes within one such system improved its stability, its ability to accumulate resources, or its ability to survive and reproduce, that system, and its progeny, would be

---

[62] The Miller-Urey experiment & wikipedia: http://en.wikipedia.org/wiki/Miller–Urey_experiment

[63] A reassessment of prebiotic organic synthesis in neutral planetary atmospheres:

[64] A time-line of life's evolution: http://exploringorigins.org/timeline.html

[65] Mineral Surfaces, Geochemical Complexities, and the Origins of Life

likely to become more common. As these changes accumulate and are passed from parent to offspring, the organisms will inevitably evolve, as we will see in detail in the next chapter.

As in living systems today, the earliest steps in the formation of the first organisms required a source of energy to maintain the non-equilibrium living state. There are really two choices for the source of this energy, either light (electromagnetic radiation from the sun) or thermodynamically unstable chemicals present in the environment. There have been a number of plausible scenarios, based on various observations, for the steps leading to life. For example, a recent study based on the analysis of the genes, and the proteins that they encode, found in modern organisms, suggests that the last universal common ancestor (LUCA) arose in association with hydrothermal vents and derived energy from thermodynamically favorable chemical reactions.[66] But whether this reflects LUCA or an ancestor of LUCA that became adapted to living in association with hydrothermal vents is difficult, and perhaps impossible to resolve unambiguously, particularly since LUCA lived ~3.4-3.8 billion years ago and cannot be studied directly.

Mapping the history of life on earth

Assuming, as seems likely, that life arose spontaneously, we can look at what we know about the fossil record to better understand the diversification of life and life's impact on the Earth. This is probably best done by starting with what we know about where the Universe and Earth came from. The current scientific model for the origin of the universe is known as the "Big Bang" (also known as the "primeval atom" or the "cosmic egg"), an idea originally proposed by the priest, physicist and astronomer Georges Lemaître (1894-1966).[67] The Big Bang model arose from efforts to answer the question of whether the fuzzy nebulae identified by astronomers were located within or outside of our galaxy. This required some way to determine how far these nebulae were from Earth. Edwin Hubble (1889-1953) and his co-workers were the first to realize that nebulae were in fact galaxies in their own right, each very much like our own Milky Way, and that each is composed of many billions of stars. This was a surprising result. It made Earth, sitting on the edge of one (the Milky Way) among many, many galaxies seem less important – a change in cosmological perspective similar to that associated with the idea that the Sun, rather than Earth, was the center of the solar system and the Universe.

To measure the movement of galaxies with respect to Earth, Hubble and colleagues combined two types of observations. The first of these allowed them to estimate the distance from the Earth to various galaxies and the second used measurements of the Doppler shift of the light from stars within distant galaxies. The Doppler shift is the effect on the wavelength of sound or light of an object's velocity relative to an observer. In the case of light emitted from an object moving toward an observer, the wavelength will be shortened, that is, shifted to the blue end of the spectrum. Light emitted from an object moving away from the observer will be lengthened, that is, shifted to the red end of the spectrum. Based on the observed Doppler shifts of light coming from stars in galaxies and the observation that the further a galaxy appears to be from Earth, the greater that shift is toward the red, Hubble concluded that galaxies, outside of our local group, were all moving away from one another. Running time

---

[66] Meet LUCA, the Ancestor of All Living Things: http://www.nytimes.com/2016/07/26/science/last-universal-ancestor.html?_r=1

[67] Georges Lemaître: http://www.physicsoftheuniverse.com/scientists_lemaitre.html

backward, he concluded that at one point in the past, all of the matter and energy in the universe must have been concentrated in a single point.[68] A prediction of this Big Bang model is that the Universe is ~13.8 +/- 0.2 billion (10^9) years old. This is a length of time well beyond human comprehension; it is sometimes referred to as deep time – you can get some perspective on deep time using the "Here is Today" website (http://hereistoday.com). Other types of data have been used to arrive at an estimated age of Earth and the other planets in the solar system as ~4.5 x 10^9 years.

After Earth formed, it was bombarded by extraterrestrial materials, including comets and asteroids. This bombardment began to subside around ~3.9 billion years ago and reached its current level by ~3.5 billion years ago.[69] It is not clear whether life arose multiple times and was repeatedly destroyed during the early history of Earth (4.5 to 3.6 billion years ago) or if the origin of life was a one-time event, taking hundreds of millions of years before it succeeded, after which it managed to survive and expand to the present day.

Fossil evidence for the history of life on earth

The earliest period in Earth's history is known as the Hadean, after Hades, the Greek god of the dead. The Hadean is defined as the period between the origin of the Earth up to the first appearance of life. Fossils provide our only direct evidence for when life appeared on Earth. They are found in sedimentary rock, which is rock formed when fine particles of mud, sand, or dust entomb an organism before it can be eaten by other organisms. Hunters of fossils (paleontologists) do not search for fossils randomly but use geological information to identify outcroppings of sedimentary rocks of the specific age they are interested in, in order to direct their explorations.

Early in the history of geology, before Charles Darwin and Alfred Wallace proposed the modern theory of evolution, geologists recognized that fossils of specific types were associated with rocks of specific ages. This correlation was so robust that rocks could be accurately dated based on the types of fossils they contained. At the same time, particularly in a world that contains young earth creationists who claim that Earth was formed less than ~10,000 years ago, it is worth remembering both the interconnectedness of the sciences and that geologists do not rely solely on fossils to date rocks. This is in part because many types of rocks do not contain fossils. The non-fossil approach to dating rocks is based on the physics of isotope stability and the chemistry of atomic interactions. It uses the radioactive decay of elements with isotopes with long half-lives, such as ^{235}Ur (uranium) which decays into ^{207}Pb (lead) with a half-life of ~704 million years and ^{238}Ur which decays into ^{206}Pb with a half-life of ~4.47 billion years. Since these two Pb isotopes appear to be formed exclusively through the decay of ^{235}Ur, the ratios of Ur and Pb isotopes can be used to estimate the age of a rock, assuming that it originally contained only Ur, and no Pb.

In order to use isotope abundance to accurately date rocks, it is critical that all of the atoms in a mineral measured originated there and stayed there, that is, that none were washed into or out of the rock. Since Ur and Pb have different chemical properties, this can be difficult to establish in some types of minerals. That said, with care, and using rocks that contain chemically inert minerals, like zircons, the isotope ratio method can be used to measure the age of rocks to an accuracy of ~1% or better. Such

---

[68] The origin of the universe and the primeval atom

[69] The violent environment of the origin of life

age estimates, together with other types of evidence, support James Hutton's (1726-1797) famous dictum that Earth is ancient, with "no vestige of a beginning, no prospect of an end."[70] We know now, however, that this statement is not true; while very old, Earth had a beginning, it coalesced around ~5 billion years ago, and it will disappear when the sun expands and engulfs it in about ~5.5 billion years from now.[71]

Now, back to fossils. There are many types of fossils. Chemical fossils are molecules that, as far as we know, are naturally produced only through biological processes.[72] Their presence in ancient rock implies that living organisms were present at the time the rock formed. Chemical fossils first appear in rocks that are between ~3.8 to ~3.5 x 10^9 years old. What makes chemical fossils problematic is that there may be non-biological but currently undiscovered or unrecognized mechanisms that could have produced them, so we have to be cautious in our conclusions.

Moving from the molecular to the physical, there are what are known as trace fossils. These can be subtle or obvious. Organisms can settle on mud or sand and make impressions. Burrowing and slithering animals make tunnels or disrupt surface layers. Leaves and immotile organisms can leave impressions. Walking animals can leave footprints in sand, mud, or ash. How does this occur? If the ground is covered, compressed, and converted to rock, these various types of impressions can become fossils. Later erosion can then reveal these fossils. For example, if you live near Morrison, Colorado, you can visit the rock outcrop known as Dinosaur Ridge and see trace fossil dinosaur footprints; there may be similar examples near where you live.

We can learn a lot from trace fossils, they can reveal the general shape of an organism or its ability to move or to move in a particular way. To move, an organism must have some kind of muscle or alternative mobility system and probably some kind of nervous system that can integrate information and produce coordinated movements. Movement also suggests that the organisms that made the trace had something like a head and a tail. Tunneling organisms are likely to have had a mouth to ingest sediment, much like today's earthworms - they were predators, eating the microbes they found in mud.

In addition to trace fossils, there are also the type of fossils that most people think about, which are known as structural fossils, namely the mineralized remains of the hard parts of organisms such as teeth, scales, shells, or bones. As organisms developed hard parts, fossilization, particularly of organisms living in environments where they could be buried within sediment before being dismembered and destroyed by predators or microbes, became more likely.

Unfortunately for us (as scientists), many and perhaps most types of organisms leave no trace when they die, in part because they live in places where fossilization is rare or unlikely. Animals that live in woodlands, for example, rarely leave fossils. The absence of fossils for a particular type of organism does not imply that these types of organisms do not have a long history, rather it means that the conditions where they lived and died or their body structure is not conducive to fossilization. Many types of living organisms have no fossil record at all, even though, as we will see, there is molecular evidence that they arose tens to hundreds of millions of years ago.

---

[70] Changing Views of the History of the Earth

[71] How the sun will die

[72] Although as Wohler pointed out, they can be generated in the laboratory.

Life's impact on the earth

Based on fossil evidence, the current model for life on Earth is that for a period of ~2 x 10^9 (billion) years after the appearance of LUCA, the only forms of life on Earth were microscopic. Today, there are three families of organisms, which we describe briefly here and in more detail later on: the bacteria, the archaea, and the eukaryotes. While the exact nature of LUCA is unclear, it is likely that it was single celled and relatively simple in general organization (→) consisting of a boundary membrane, controlling the movement of molecules into and out of the cell, a cytoplasm, in which various biosynthetic reactions took place, and molecules of the genetic material, DNA, located within the cytoplasm. Both bacteria and archaea have this same basic type of cellular organization, although they differ in a range of molecular details, although not in basic molecular mechanisms.[73] As we will discuss later, eukaryotes are more complex structurally; they contain internal membrane systems and their genetic material is located within a double membrane compartment (the nucleus) located within the cytoplasm. Movement between nuclear interior and cytoplasm is facilitated by complex molecular machines, known as nuclear pores. How the nucleus came to be remains (not surprisingly) unclear, but it is possible that the proto-eukaryote (that is, with a nucleus) arose through a fusion event that involved both bacterial and archaeal ancestors.[74] Alternatively, it might be directly descended from LUCA – the problem is that we do not have direct evidence as to the details of LUCA's structure, just inferences (informed guesses). It is clear, however, that the formation of eukaryotes involved a symbiotic event (discussed in more detail in Chapter 5) in which an α-proteobacterium (a type of bacteria) was engulfed, but not digested, by the protoeukaryote (→). This "endogenous bacterium" became the eukaryotic mitochondria. Essentially all eukaryotes (the protozoa, fungi, animals, and plants) have mitochondria, apparently descended from this event. Later in the history of life, a second endosymbiotic event occurred in which the mitochondrial containing eukaryote engulfed but did not digest a second type of bacteria, a photosynthetic cyanobacterium, leading to the algae and the plants.
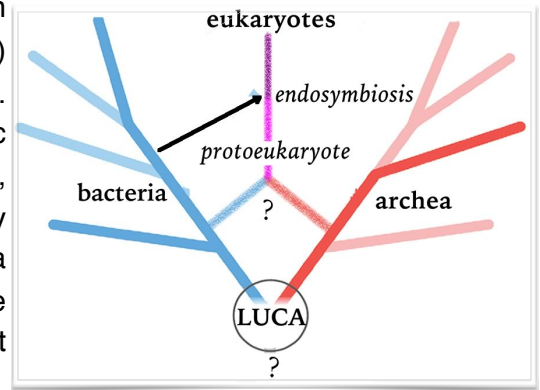
While the earliest organisms probably used energy released in the course of chemical reactions to maintain their structural integrity and to grow, relatively soon bacterial-type organisms appeared that could capture the energy in light and use it to drive various thermodynamically unfavorable reactions. A major class of such reactions involves combining CO_2 (carbon dioxide), H_2O (water), and other molecules to form carbohydrates (sugars) and other important biological molecules, such as lipids, proteins, and nucleic acids. At some point during the early history of life on Earth, organisms appeared that released molecular oxygen (O_2) as a waste product of light-driven reactions, known generically as

---

[73] see the Common Ancestor of Archaea and Eukarya

[74] Origin of eukaryotes & The common ancestor of archaea and eukarya was not an archaeon

oxygenic photosynthesis. These oxygen-releasing organisms became so numerous that they began to change Earth's surface chemistry - they represent the first life-driven ecological catastrophe.

The level of atmospheric O_2 represents a balance between its production, primarily by organisms carrying out oxygenic photosynthesis, and its removal through various chemical reactions. Early on as O_2 appeared, it reacted with iron to form deposits of water-insoluble Fe (III) oxide (Fe_2O_3) – that is, rust. This rust reaction removed large amounts of O_2 from the atmosphere, keeping levels of free O_2 low. The rusting of iron in the oceans is thought to be largely responsible for the massive banded iron deposits found around the world.[75] O_2 also reacts with organic matter, as in the burning of wood, so when large amounts of organic matter are buried before they can react, as occurs with the formation of coal, more O_2 accumulates in the atmosphere. Although O_2 was probably being generated and released earlier, by ~2 billion years ago, atmospheric O_2 had appeared in detectable amounts and by ~850 million years ago O_2 had risen to significant levels (→). Atmospheric O_2 levels have changed significantly since then, based on the relative rates of its synthesis and destruction. Around ~300 million years ago, atmospheric O_2 levels reached ~35%, almost twice the current level. It has been suggested that these high levels of atmospheric O_2 made the evolution of giant insects possible.[76]
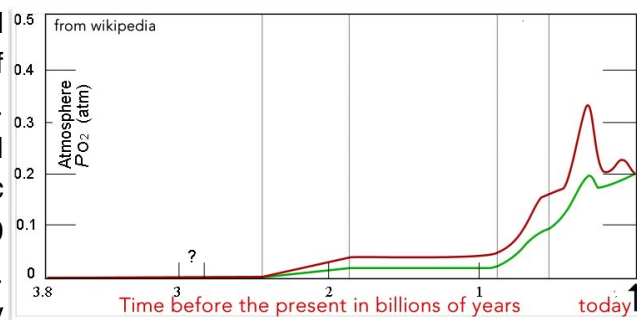
Although we tend to think of O_2 as a natural and benign substance, it is in fact highly reactive and potentially toxic; its production and accumulation posed serious challenges and unique opportunities to, organisms. As we will see later on O_2 can be "detoxified" through reactions that lead to the formation of water; this type of thermodynamically favorable reaction appears to have been co-opted for a wide range of biological purposes. For example, through coupled reactions O_2 can be used to capture the maximum amount of energy from the breakdown of complex molecules (food), leading to the generation of CO_2 and H_2O, both of which are very stable.

Around the time that O_2 levels were first rising, that is ~10^9 years ago, the first trace fossil burrows appeared in the fossil record. These were likely to have been produced by simple worm-like, macroscopic multicellular organisms, known as metazoans, that is, multi-cellular animals, capable of moving along and through the mud on the ocean floor. About 0.6×10^9 years ago, new and more complex structural fossils began to appear in the fossil record. The first of these to appear in the fossil record are the so-called Ediacaran organisms (→), named after the geological formation in which their fossils were first found.[77] Current hypotheses suggest they were immotile, like modern sponges but flatter; it remains unclear how or if they are related to later animals. Since the fossil record does not contain all organisms, we are left to speculate on what earlier metazoans looked like. By the beginning of the Cambrian age (~545×10^6 years ago), a wide variety of

---
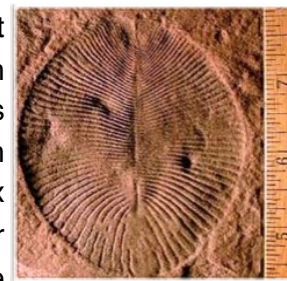
[75] Paleoecological Significance of the Banded Iron-Formation: http://econgeol.geoscienceworld.org/content/68/7/1135.abstract

[76] see Geological history of oxygen & Atmospheric oxygen and giant Paleozoic insects

[77] http://en.wikipedia.org/wiki/Ediacara_biota

organisms had appeared within the fossil record, many clearly related to modern animals. Molecular level data suggest that their ancestors originated more than ~30 million years earlier. These Cambrian organisms show a range of body types. Most significantly, many were armored. Since building armor involves expending energy to synthesize these components, the presence of armor suggests the presence of predators, and a need for a defensive response.

Viruses: Now, before we leave this chapter you might well ask, have we forgotten viruses? Well, no - viruses are often a critical component of an ecosystem and an organism's susceptibility or resistance to viral infection is often an important evolutionary factor, but viruses are different from organisms in that they are non-metabolic. That means they do not carry out reactions and cannot replicate on their own, they replicate only within living cells. Basically they are not alive, so even though they are extremely important, we will discuss viruses only occasionally and in quite specific contexts. That said, the recent discovery of giant viruses, such as Mimivirus, suggests that something interesting is going on.[78]

Questions to answer

14. In 1961 Frank Drake, a radio astronomer, proposed an equation to estimate the number of technological civilizations that exist within the observable Universe (N).[79]
    The equation is $N = R^* \times f_p \times n_e \times f_l \times f_i \times f_c \times L$ where:
    R^* = The rate of formation of stars suitable for the development of intelligent life.
    f_o = The fraction of those stars with planetary systems.
    n_e = The number planets, per solar system, with an environment suitable for life.
    f_l = The fraction of suitable planets on which life actually appears.
    f_i = The fraction of life-bearing planets on which intelligent life emerges.
    f_c = The fraction of civilization that develop a technology that releases detectable signs of their existence into space.
    L = The length of time such civilizations release detectable signals into space.
    Identify those parts of the Drake equation that can and those that cannot be established (at present) empirically. Is the Drake equation scientific, or does it just look "sciency"; explain your reasoning.

15. What factors would influence the probability that a particular type of organism will be fossilized?
16. What factors might drive the appearance of teeth, bones, shells, muscles, nervous systems, and eyes?
17. What factors, biological and geological, determine atmospheric O_2 levels?
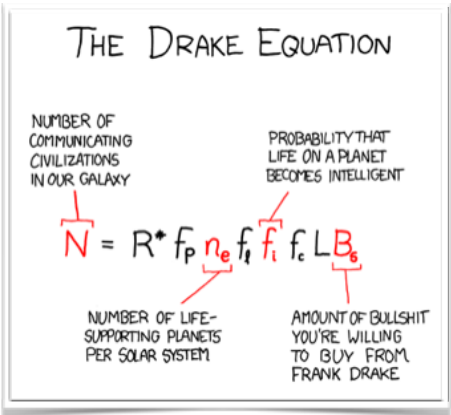
Questions to ponder

- Can the origin of life be studied scientifically, and if so, how?
- If we assume that spontaneous generation occurred in the distant past, why is it not occurring today? How could you tell if it were?

---

[78] http://www.giantvirus.org/intro.html

[79] The Drake equation: http://www.seti.org/drakeequation and cartoon: http://xkcd.com/384/

Chapter 3: Evolutionary mechanisms and the diversity of life

In which we consider the rather exuberant diversity of organisms and how they came to be. To understand these processes involves introducing core evolutionary mechanisms, both adaptive (natural selection) and non-adaptive (drift and bottlenecks).

In medieval Europe there was a tradition of books known as bestiaries; these were illustrated catalogs of real and imagined organisms; often each particular organism was associated with a moral lesson. "Male lions were seen as worthy reflections of God the Father, for example, while the dragon was understood as a representative of Satan on earth."[80] One can see these books as an early version of a natural theology, that is, an attempt to gain an understanding of the supernatural through the study of natural objects.[81] In this case, the presumption was that each type of organism was created for a particular purpose, and that often this purpose was to provide people with a moral lesson. This way of thinking grew more and more problematic as more and more different types of organisms were recognized, many of which had no obvious significance to humans. Currently, scientists have identified approximately ~1,500,000 different species of plants, animals, and microbes. The actual number of different types of organisms, referred to as species, may be as high as ~10,000,000.[82] These numbers refer, of course, to the species that currently exist, but we know from the fossil record that many now extinct species existed in the past. So the obvious question is, why are there so many different types of organisms?[83] Do they represent multiple independent creation events, and if so, how many such events have occurred? Given how different types of organisms look and behave, it seems possible that trees, mushrooms, spiders, whales, and humans represent distinct lineages and separate creation events.

As the actual diversity of organisms was discovered, a number of observations served to undermine the early concept that organisms were created to serve humanity. The first of these was the fact that a number of organisms had very little obvious importance to the human condition. While particularly obvious in the case of extinct organisms, this extended to a range of newly discovered organisms; panda bears, potatoes, and maize come to mind. At the same time students of nature, known generically as naturalists, discovered many different types of upsetting and cruel behaviors within the natural world. Consider the fungus *Ophiocordyceps unilateralis*, which infects the ant *Camponotus leonardi*. The fungus takes control of the ant's behavior, causing infected ants to migrate to positions that favor fungal growth before killing the infected ant. Similarly, the nematode worm *Myrmeconema neotropicum* infects the ant *Cephalotes atratus*, leading to dramatic changes in the infected ant's morphology and behavior. The infected ant's abdomen turns red and is held raised up, which makes it resemble a fruit and increases the likelihood of the infected ant being eaten by birds

---

[80] Northumberland Bestiary

[81] What Is Natural Theology?

[82] How many species are there on Earth and in the ocean?

[83] As a technical point, which we will return to, we will refer to each distinct type of organism as a specie

(→). The birds transport the worms, which survive in their digestive systems until they are excreted; they are then eaten by, and infect new ants to complete the worm's life cycle.[84] Perhaps the most famous example of this type of behavior occurs in wasps of the family *Ichneumonidae*. Female wasps deposit their fertilized eggs into the bodies of various types of caterpillars. The wasp eggs hatch out and produce larvae which then feed on the living caterpillar, consuming it from the inside out. Charles Darwin, in a letter to the American naturalist Asa Gray, remarked "There seems to me too much misery in the world. I cannot persuade myself that a beneficent & omnipotent God would have designedly created the *Ichneumonidae* with the express intention of their feeding within the living bodies of caterpillars, or that a cat should play with mice." Rather than presume that a supernatural creator was responsible for such apparently cruel behaviors, Darwin and others sought alternative, morally neutral naturalistic processes that could both generate biological diversity and explain biological behaviors.

As the diversity of organisms became increasingly apparent and difficult to ignore, another broad and inescapable conclusion began to emerge from anatomical studies: many different organisms displayed remarkable structural similarities. For example, as naturalists characterized various types of animals, they found that they either had an internal skeleton (the vertebrates) or did not (the invertebrates). Comparative studies revealed that there were often many similarities between quite different types of organisms. A classic work, published in 1555, compared the skeletons of a human and a bird, both vertebrates.[85] While many bones have different shapes and relative sizes, what was most striking is how many bones are at least superficially similar between the two organisms (→). This type of "comparative anatomy" revealed many similarities between apparently unrelated organisms. For example, the skeleton of the dugong, a large aquatic mammal, appears quite similar to that of the European mole, a small terrestrial mammal that tunnels underground on land. In fact, there are general skeletal similarities between all vertebrates. The closer we look, the more similarities we find. These similarities run deeper than the anatomical, as we will find, they extend to the cellular and molecular levels as well. So the scientific question is, what explains such similarities? Why build an organism that walks, runs, and climbs, such as humans, with a skeleton similar to that of a organism that flies (birds), swims (dugongs), or tunnels (moles). Are these anatomical similarities just flukes or do they imply something deeper about how organisms were initially formed?

Homology: A Philosophical and Biological Perspective
Olivier Rieppel

External form and skeleton of the pectoral fin of the Dugong (*Halicore indicus*).
External form and skeleton of the fore-limb of the Mole (*Talpa Europea*).

---
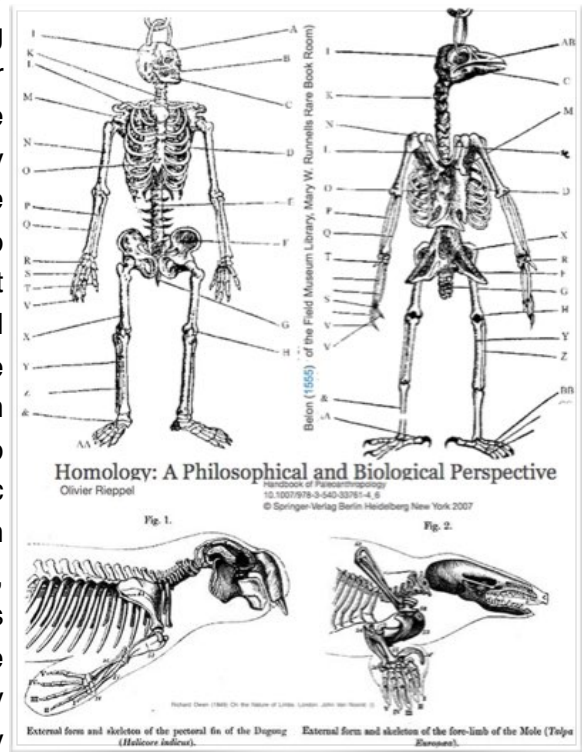
[84] The Life of a Dead Ant: The Expression of an Adaptive Extended Phenotype

[85] Belon (1555) L'Histoire de la Nature des Oyseaux. Paris, Guillaume Cavellat

Organizing organisms, hierarchically

Carl Linnaeus (1707-1778) was the pioneer in taking the similarities between different types of organisms seriously. Based on such similarities (as well as differences), he developed a system to classify organisms in a coherent and hierarchical manner. Each organism had a unique place in this scheme, a unique set of coordinates.[86] What was, and occasionally still is, the controversial aspect of such a classification system is in how to decide which traits should be considered significant and which are superficial or unimportant, at least for the purposes of classification. Linnaeus had no real idea for how to explain why organisms could even be classified in a hierarchical manner.

This might be a good place to reconsider the importance of guesses, hypotheses, models, and theories in biology, and science in general. Linnaeus noticed the apparent similarities between organisms and used it to generate his classification scheme, but he had no explanation for why such similarities should exist in the first place, very much like Newton's law of gravitation did not explain why there was gravity, just how it behaved. So what are the features of a scientific (predictive) model? Such a model has to suggest observations or predict outcomes that have not yet been observed. It is the validity of these predictions that enables us to identify useful models. A model that makes no empirically testable predictions is not useful scientifically. In this light, Linnaeus's scheme was not scientific, just descriptive. The value of a scientific model, that is, a model that makes explicit predictions, even if they prove to be wrong, is that it enables us to refine, or force us to abandon, our model. A scientific model that, through its various predictions and their confirmation, refutation, or revision, has been found to accurately explain a particular phenomenon, if it explains enough, becomes a theory. We assume that the way the model works is the way the world works. This enables us to distinguish between a law and a theory. A law describes what we see but not why we see it. A theory provides the explanation for why the law works.[87]

Back to Linnaeus, whose classification system placed organisms of a particular type together into a species. This, of course, raises a number of interesting questions - how different do two organisms have to be to no longer fall into the same species. How do we make such a decision? As we will see, each organism is unique genetically (its genotype) as well as in its various observable traits: its phenotype. If we look at organisms that appear similar, do we place larger individuals (of the same age) into a different species than smaller ones? The situation is even more complex when we think about models of reproduction. Some organisms can reproduce, that is, produce offspring, by themselves; such organisms can be either asexual or self-fertilizing, often called hermaphroditic - a distinction that we will return to later. Other types of organisms are sexual, individuals need to cooperate with another of the same type to produce offspring. Here we find a reasonably common, but not universal, situation known as sexual dimorphism, individuals of the two sexes can appear dramatically different from one

---

[86] Each organism can be identified by a species, within a genus, within a family, within an order, within a class, within a phylum, within a Kingdom.

[87] If we go back, Newton's law of gravity explained how objects behaved gravitationally, but it not why. In contrast, Einstein's theory of general relativity explained why there was gravity, and predicted behaviors that were not predicted by Newton's law.

another (left ♂ & right ♀ →).[88] It is often the case that organisms of the same type but different sexes, different developmental stages, and even growing under different conditions can have different phenotypes. It therefore requires careful study to recognize and characterize a particular type of organism.

Of course, what originally counted as a discrete type of organism, a particular species, was based on Linnaeus's or some other naturalists' judgement as an observer and classifier; it depended on which particular traits were assumed to be significant and useful to distinguish organisms of one species from those of another, perhaps quite, similar appearing species. The choice of these key traits is subject to debate. Based on the perceived importance and presence of particular traits, organisms could be split into two or more types (species), or two types originally considered separate could be reclassified into a single species.

As we will see, the individual organisms that make up a species are not identical but share many traits. As noted above, in organisms that reproduce sexually, there are sometimes dramatic differences between males and females of the same species; these differences can be so dramatic that without further evidence, it can be difficult to tell whether two animals are members of the same or different species. In this light the primary criteria for determining whether sexually reproducing organisms are members of the same or different species is whether they can and do successfully interbreed with one another in the wild. This criterion, reproductive compatibility, can be used to determine species distinctions on a more empirical basis, but it is not useful with asexual species, such as most microbes. An asexual organism is essentially a clone and species distinctions have to be based on other criteria, which we will return to later when we discuss genes and genomes. Within a species, there are sometimes regional differences that are distinct enough to be recognizable. Where this is the case, these groups are known as populations, races, or subspecies. While distinguishable, the organism in these groups retain the ability to interbreed and so are members of a single species. As an example tigers are *Panthera tigris*, while Siberian tigers are known as *Panthera tigris sumatrae*.

After defining species, Linnaeus next grouped species that displayed similar traits into a larger group, known as a genus. While a species can be considered a natural, interbreeding population, a genus is a more artificial group. Which species are placed together within a particular genus depends on the common traits deemed important or significant by the person doing the classifying. This can lead to conflicts between researchers that can be resolved by the collection of more comparative data.

In the Linnaean classification scheme, each organism has a unique name, which consists of its genus and species names. The accepted usage is to write the name in italics with the genus name capitalized, for example, *Homo sapiens*. Following on this pattern, one or more genera are placed into larger, more inclusive groups (the next larger group is known as a "family"), and these groups, in turn, are placed into larger groups. The end result of this process is the rather surprising observation that all organisms fall into a small number of "supergroups" or phyla. We will not worry about the traditional group names,

---

[88] Sexual dimorphism & sexual dimorphism in spiders

because in most cases they really do not help in our understanding of basic biology. Perhaps most surprising of all, all organisms and all phyla – all of the organisms on Earth – can be placed into a single unified phylogenetic "tree" or perhaps better put, bush – they are all connected. That this should be the case is by no means obvious. This type of analysis could have produced multiple, disconnected classification schemes, but it did not.

Natural and un-natural groups

It is worth reiterating that while a species, particularly in sexually reproducing species, can be seen as a natural group, the higher levels of classification may or may not reflect biologically significant information. Such higher-level classification is an artifact of the human need to make sense of the world; it also has the practical value of organizing information, much like the way books are organized in a library. We can be sure that we are reading the same book, and studying the same organism!

Genera and other higher-level classifications are based on a decision to consider one or more traits as more important than others. The assignment of a particular value to a trait can seem arbitrary. Let us consider, for example, the genus *Canis*, which includes wolves and coyotes and the genus *Vulpes*, which includes foxes. The distinction between these two groups is based on smaller size and flatter skulls in *Vulpes* compared to *Canis*. Now let us examine the genus *Felis*, the common house cat, and the genus *Panthera*, which includes tigers, lions, jaguars and leopards. These two genera are distinguished by cranial features and the fact that *Panthera*, but not *Felix* have the ability to roar. So what do we make of these distinctions, are they really sufficient to justify distinct groups, or should *Canis* and *Vuples* (and *Felix* and *Panthera*) be merged together? Are the differences between these groups biologically meaningful? They are in the sense that they recognize similarities and differences between organisms, but these similarities and differences may be ambiguous. Such ambiguity is illustrated by the fact that the higher order classification of an organism can change: organisms originally placed in one genus can become a separate genus within a family, the next more inclusive grouping, and vice versa, or a species can be moved from one genera to another. Consider the types of organisms commonly known as bears. There are a number of different types of bear-like organisms, a fact that Linnaeus's classification scheme acknowledged. Looking at all bear-like organisms we recognize eight types.[89] We currently consider four of these, the brown bear (*Ursus arctos*), the Asiatic black bear (*Ursus thibetanus*), the American bear (*Ursus americanus*), and the polar bear (*Ursus maritimus*) to be significantly more similar to one another, based on the presence of various traits, than they are to other types of bears. We therefore placed them in their own genus, *Ursus*. We have placed each of the other types of bear-like organisms, the spectacled bear (*Tremarctos ornatus*), the sloth bear (*Melurus ursinus*), the sun bear (*Helarctos mayalanus*), and the giant panda (*Ailuropoda melanoleuca*) in their own separate genera, because scientists consider these species more different from one another than are the members of the genus *Ursus*. The problem here is how big do these differences have to be to warrant a new genus? Hopefully, it is obvious to you that there are parts of any classification system that are subject to argument and others that are unambiguous.

---

[89] http://en.wikipedia.org/wiki/List_of_bears

Evolution: making theoretic sense of Linnaean classification

So where does that leave us? Here the theory of evolution together with the cell theory, that is, the continuity of life, come together. We work on the assumption that the more closely related, evolutionarily, two species are, the more traits they will share and that the development of a new, biologically significant trait is what distinguishes one group from another. Traits that underlie a rational classification scheme are known as synapomorphies, a technical term; basically these are traits that appeared in one or the other branch point of a family tree and serve to define that branch point, such that an organism on one branch represent an evolutionary lineage, and so are part of a "natural" group, more closely related to one another and distinct from those on the other branch, which is less closely related (→). The organisms within each branch are placed in a common Linnaean group. Going back further in time, the two groups, share a common ancestor, and are part of a larger, more inclusive Linnaean group. The continuous (unbroken) ancestral relationships between all organisms provides a reason for why organisms can be arranged into a hierarchical classification scheme.
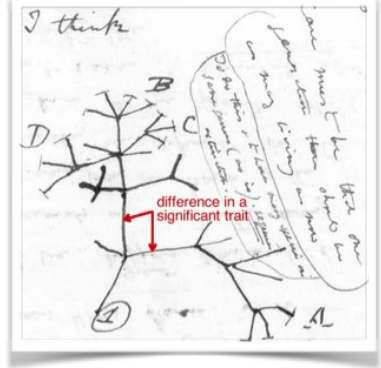
So a remaining question is, how do we determine ancestry when the ancestors lived, thousands, millions, or billions of years in the past. Since we cannot travel back in time, we have to deduce relationships from comparative studies of living and fossilized organisms. Here the biologist Willi Hennig (1913-1976) played a key role.[90] He established rules for using shared, empirically measurable traits to reconstruct ancestral relationships, such that each group should have a single common ancestor (or ancestral population). As we will discover later on, one of the traits now commonly used in modern studies is gene (DNA) sequence and genomic organization data, although even here there are plenty of situations where ambiguities remain, due to the very long times that separate ancestors from present day organisms.

Fossils and family relationships: introducing cladistics (briefly)

As mentioned previously, we continue to discover new fossils, new organisms, and, as we will see, new genes. In most cases, fossils appear to represent organisms that lived many millions to hundreds of millions of years ago but which are now extinct. We can expect that there are dramatic differences between the ability of different types of organisms to become fossilized.[91] Perhaps the easiest organisms to fossilize are those with internal or external skeletons, yet it is estimated that between 85 to 97% of such organisms are not represented in the fossil record. A number of studies indicate that many other types of organisms have left no fossils whatsoever[92] and that the number of organisms at the genus level that have been preserved as fossils may be less, often much less than ~5%.[93] For

---

[90] A description of Willi Hennig's impact on taxonomy

[91] Your inner fish video

[92] The incompleteness of the fossil record

[93] Absolute measures of the completeness of the fossil record

some categories of modern organisms, such as the wide range of microbes, essentially no informative fossils exist at all.

Once scientists recognized that fossils provide evidence for extinct organisms, the obvious question was, do extinct organisms fit into the same classification scheme as do living organisms or do they form their own groups or even their own separate trees, which could provide evidence for multiple independent origins of life and multiple distinct common ancestors? This can be a difficult question to answer, since many fossils are only fragments of the intact organism. The fragmentary nature of the fossil record can lead to ambiguities. Nevertheless, the conclusion that has emerged upon careful characterization is that we can place essentially all fossilized organisms within the same cladistic classification scheme developed for modern organisms. That said where organisms, such as the Ediacarian organisms, should be placed remains ambiguous.[94] The presumption is, however, that if we had samples of Ediacarian organisms for molecular (DNA) analyses, we could quickly resolve this question, and we would find that they fall nicely into the same classification scheme as all other organisms do (a topic we will return to).[95] A similar example are the dinosaurs, which while extinct, are clearly descended from a specific type of reptile that also gave rise to modern birds, while mammals are more closely related to a second, now extinct group, known as the "mammal-like reptiles."

In rare cases, particularly relevant to human evolution, DNA sequence data can be recovered from bones. For example, it is possible to extract and analyze DNA from the bones of Neanderthals and Denisovian-type humanoids; both types of human-like organisms went extinct ~30,000 years ago. DNA sequence information has been used to clarify the relationship between Neanderthals, Denisovians, and modern humans, *Homo sapiens*.[96] In fact, such data provides compelling evidence for limited interbreeding between these groups and has led for calls to reclassify Neanderthals and Denisovians as subspecies of Homo sapiens.[97]

Questions to answer:

18. Explain how extinct species fit into the same classification scheme as used for living (observable) organisms.
19. Why are differences between organisms significantly less informative in determining phylogenetic relationships than similarities?
20. What factors would influence your decision as to whether a trait found in two different organisms was present in their common ancestor?
21. You discover life on a planet orbiting another star in another galaxy; would you expect such organisms to fit into the Linnaean classification system?

Questions to ponder:

- What observations would you consider to decide whether Neanderthals and Denisovians were distinct species from *H. sapiens*?
- Was sex with a Neanderthal immoral?

---

[94] Doser, 2015. The advent of animals: The view from the Ediacaran

[95] On the eve of animal radiation: phylogeny, ecology and evolution of the Ediacara biota

[96] Paleogenomics of archaic hominins:http://www.ncbi.nlm.nih.gov/pubmed/22192823

[97] Humans mated with Neandertals much earlier and more frequently than thought & The downside of sex with Neanderthals

The theory of evolution and the organization of life

Why exactly is it that birds, whales, and humans share common features, such as the organization of their skeletons, similarities that led Linnaeus to classify them together as vertebrates? Why are there extinct organisms, known only from their fossils, but which nevertheless share many common features with living organisms? And most importantly, why are there so many different types of organisms? Charles Darwin (1809-1882) and Alfred Wallace (1823–1913) proposed a model, described in great detail in Darwin's book The Theory of Evolution by Natural Selection,

The main unifying idea in biology is Darwin's theory of evolution through natural selection.
– John Maynard Smith

originally published in 1858, that answered these and a number of other questions.

As we will see, evolutionary theory is based on a series of direct observations of the natural world and their logical implications. Evolutionary theory explains why similar organisms share similar traits and why we can easily place them into a nested classification system. Organisms are similar because they are related to one another – they share common ancestors.[98] Moreover, we can infer that the more characters two species share the more recently they shared a common ancestor. We can even begin to make plausible, empirical and testable deductions about what those common ancestors looked like. As an example, we can predict that the common ancestor of all terrestrial vertebrates will resemble a fish

with leg-like limbs - and we can predict the number and shape of the bones found in those limbs. Scientists have discovered fossils of such an organism, *Tiktaalik* (←).[99] Its discovery is one more example of the fact that since its original introduction, and well before the mechanisms of heredity and any understanding of the molecular nature of organisms were resolved, evolutionary

Tiktaalik roseae, an extinct organism that lived ~375 million years ago, is likely to be similar to the common ancestor of all terrestrial vertebrates (from "Your inner fish" by Neil Shubin)

theory explained what was observed, made testable predictions about what would be found, and has been supported by what has, in fact, been found. In the case of particularly fast growing organisms, and very strong selections pressures (such as the presence of an antibiotic), we can observe even evolutionary processes over the course of days, weeks, and months – that is, in real time.[100]

Evolution theory's core concepts

So what are the facts and inferences upon which the Theory of Evolution is based? Two of its foundational observations are deeply interrelated and based on empirical observations associated with plant and animal breeding and the characteristics of natural populations. The first is the fact that whatever type of organism we examine, if we look carefully enough, making accurate measurements of visible and behavioral traits, which is known as its phenotype, we find that individuals vary with respect

---

[98] As we will discover, there are organisms can appear similar that are not closely related; this is due to what is known as convergent evolution. That said, such organisms share a common ancestor, although it existed further back in time.

[99] Meet Tiktaalik roseae: An Extraordinary Fossil Fish A similar situation applies to the terrestrial ancestors of whales

[100] Visualizing evolution as it happens

to one another. More to the point, plant and animal breeders recognized that the offspring of controlled matings between individuals often displayed phenotypes similar to those of their parents, indicating that phenotypic (observable) traits can be inherited. Over many generations, domestic animal and plant breeders used what is now known as artificial selection to generate the range of domesticated plants and animals with highly exaggerated phenotypes. For example, beginning ~10,000 years ago plant breeders in Mesoamerica developed modern corn (maize) by the selective breeding of variants of the grass teosinte (→).[101] Current evidence supports the idea that all of the various breeds of dogs, from the tiny to the rather gigantic, appear to be derived from a common ancestor that lived between ~19,000 to 32,000 years ago. Although it is certainly true that new evidence could be discovered that might change our estimates of where and when this common ancestor(s) lived.[102] In all cases, the crafting of domesticated organisms followed the same pattern.

not to scale (bigger!)
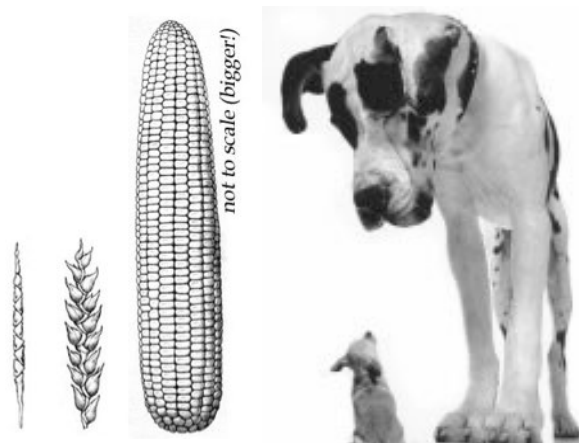
In artificial, that is, human-determined selection, those organisms with desirable (or desired) traits were selected for breeding with one another. Organisms that did not have these traits were discarded and not permitted to breed. This process of artificial selection, carried out over hundreds to thousands of generations, led to organisms that display distinct or exaggerated forms of the selected trait. What is crucial to understand is that this strategy could work only if different versions of the trait were present in the original selected population and at least a part of this phenotypic variation was due to genetic, that is heritable, factors. Originally, what these genetic heritable factors were was unclear. We refer to them as the organism's genotype, even though early plant and animal breeders would never have used that term.

The power of selection is based on the assumption that different organisms have different genotypes and that different genotypes produce different phenotypes. But the source of genotypic differences was not known to early plant and animal breeders. Were these differences imprinted on the organism in some way based on its experiences or were they the result of environmental factors? Was the genotype stable or could it be modified by experience? How were genotypic factors passed from generation to generation? And how, exactly, did a particular genotype produce or influence a specific phenotypic trait. As we will see this last question still remains poorly resolved for many phenotypes.

So what do we mean by genetic factors?

Here the answer is empirical. Traditional plant and animal breeders had come to recognize that offspring tended to display the same or similar traits as their parents. Such observations led them to assume that there was some factor within the parents that was expressed within the offspring and could, in turn, be passed

---

[101] Molecular Evidence and the Evolution of Maize

[102] From wild animals to domestic pets, an evolutionary view of domestication

from the offspring to their own offspring. A classic example is the Hapsburg lip (→), which was passed through a European ruling family for generations.[103] In the case of artificial selection, an important point to keep in mind is that the various types of domesticated organisms produced are often dependent for their survival on their human creators, much like European royal families. Human protection relieves them of the constraints they would experience in the wild. Because of this dependence, artificial selection can produce quite exaggerated and, in the absence of human intervention, highly deleterious traits. Just look at domesticated chickens and turkeys, which, while not completely flightless, can fly only short distances and so are extremely vulnerable to predators. Neither modern corn (*Zea mays*) or chihuahuas, one of the smallest breeds of dog, developed by Mesoamerican breeders, would be expected to survive for long in the wild.[104]

Limits on populations

It is an empirically demonstrable fact that all types of organisms, as opposed to specific individuals, are capable of producing many more than one copy of themselves. Consider, as an example, a breeding pair of elephants or a single asexually reproducing bacterium. Let us further assume that there are no limits to their reproduction, that is, that once born, the offspring will reproduce periodically over the course of their lifespan. By the end of 500 years, a single pair of elephants could

> *A single cell of the bacterium E. coli would, under ideal circumstances, divide every twenty minutes. That is not particularly disturbing until you think about it, but the fact is that bacteria multiply geometrically: one becomes two, two become four, four become eight, and so on. In this way it can be shown that in a single day, one cell of E. coli could produce a super-colony equal in size and weight to the entire planet Earth.*
> *- Michael Crichton (1969) The Andromeda Strain*

theoretically produce ~15,000,000 living descendants.[105] Clearly if these 15,000,000 elephants paired up to form 7,500,000 breeding pairs, within another 500 years (1000 years altogether) there could be as many as $7.5 \times 10^6 \times 1.5 \times 10^7$ or 1.125×10^{14} elephants. Assuming that each adult elephant weighs ~6000 kilograms, which is the average between larger males and smaller females (an example of sexual dimorphism), the end result would be ~6.75×10^{18} kilograms of elephant. Allowed to continue unchecked, within a few thousand years a single pair of elephants could produce a mass of elephants larger than the mass of the Earth, an absurd conclusion. Clearly we must have left something out of our calculations! As another example, let us turn to a solitary, asexual bacterium, which needs no mate to reproduce. Let us assume that this is a photosynthetic bacterium that relies on sunlight and simple compounds, such as water, carbon dioxide, a nitrogen source, and some minerals, to grow. A bacterium is much smaller than an elephant but it can produce new bacteria at a much faster rate. Under optimal conditions our bacterium might divide once every ~20 minutes, or even faster, and would, within approximately a day, produce a mass of bacteria greater than that of Earth as a whole. Again, we are clearly making at least one mistake in our logic.

Elephants and bacteria are not the only types of organism on the Earth. In fact every known type of organism can produce many more offspring than are needed to replace themselves before they die.

---

[103] 'Imperial Stigmata!' The Habsburg Lip, A Grotesque 'Mark' Of Royalty Through The Centuries!: & Genes and Queens

[104] How DNA sequence divides chihuahua and great dane

[105] Darwin's elephants

This trait is known as superfecundity. But unlimited growth does not and cannot happen for very long - other factors act to constrain it. In fact, if you were to monitor population numbers, you would find that the numbers of most organisms in a particular environment tend to fluctuate around a so-called steady state level. By steady state we mean that the number of objects added to the system equals the number removed, so that the overall number, over time, remains constant, or nearly so. As an example, in a steady state population animals are continually being born and are dying, but the total number of organisms remains roughly constant. If a population is growing is size, the birth rate exceeds the death rate.

So what balances the effects of superfecundity, what limits population growth? The obvious answer to this question is the fact that the resources needed for growth are limited and there are limited places for organisms to live. Thomas Malthus (1766-1834) was the first to clearly articulate the role of limited resources as a constraint on population. His was a purely logical argument. Competition between increasing numbers of organisms for a limited supply of resources would necessarily limit the number of organisms. Malthus painted a rather gloomy picture of organisms struggling with one another for access to these resources, with many living in an organismal version of extreme poverty, starving to death because they could not out-compete others for the food or spaces they needed to survive and reproduce. One point that Malthus ignored, or more likely was ignorant of, is that organisms rarely behave in this way. It is common to find various types of behaviors that limit the direct struggle for resources. For example, in some organisms, an adult has to establish, and defend, a territory before it can successfully reproduce.[106] The end result of this type of behavior is to stabilize the population around a steady state level, which is a function of both environmental and behavioral constraints.

An organism's environment includes all factors that influence the organism. Environmental factors include changes in climate, as well as changes in the presence or absence of other organisms. For example, if one organism depends in important ways upon another, the extinction of the first will necessarily influence the survival of the second.[107] Similarly, the introduction of a new type of organism or a new trait, such as oxygen-generating photosynthesis, in an established environment can disrupt existing interactions and conditions. When the environment changes, the existing steady state population level may be unsustainable or many of the different types of organisms present may not be viable. If the climate gets drier or wetter, colder or hotter, if yearly temperatures reach greater extremes, or if new organisms, including as an example, new disease-causing pathogens enter an area, the average population density may change or in some cases, if the environmental change is drastic enough, may drop to zero, in other words certain populations could go extinct. Environmental conditions and changes will influence the sustainable steady-state population level of an organism (something to think about in the context of global warming, whatever its cause).

An immediate example of this type of behavior involves the human population. Once constrained by disease, war, and periodic famine, the introduction of better public health and sanitation measures such as clean water and a more secure food supply, have led to reductions in infant mortality that has resulted in the growth of the human population. Now, in many countries, populations appear to be heading to a new steady state level, although exactly what that final population total level will be is

---

[106] Territorial Defense, Territory Size, and Population Regulation

[107] Why the Avocado Should Have Gone the Way of the Dodo & Neotropical Anachronisms: The Fruits the Gomphotheres Ate

unclear.[108] Various models have been developed based on different levels of average fertility (↓). In a number of countries, the birth rate has already fallen into the low fertility domain, although that is no guarantee that it will stay there![109] In this low fertility domain (ignoring immigration), a country's population actually decreases over time, since the number of children born is not equal to the number of people dying. This itself can generate social stresses. Decreases in birth rate per woman correlate with reductions in infant mortality, generally due to vaccination, improved nutrition, and hygiene, and increases in the educational level and the reproductive self-determination, that is, the emancipation of women. Where women have the right to control their reproductive behavior, the birth rate tends to be lower. Clearly changes in the environment, and here we include the sociopolitical environment, can dramatically influence behavior and serve to limit reproductive rates and population levels.

High fertility (2.5 children per woman)
Medium fertility (2.0 children per woman)
Low fertility (1.6 children per woman)
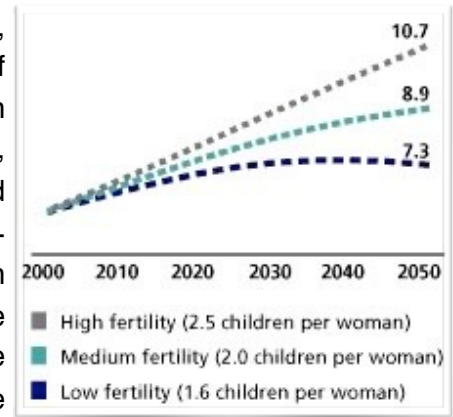
The conceptual leap made by Darwin and Wallace

Darwin and Wallace recognized the implications and significance of these key biological facts: the hereditable nature of variation between organisms, the ability of organisms to reproduce many more offspring than are needed to replace themselves, and the constraints on population size due to limited environmental resources. Based on these facts, they drew a logical implication, namely that individuals would differ in their reproductive success – that is, different individuals would leave behind different numbers of descendants. Over time, we would expect that the phenotypic variations associated with greater reproductive success, and the genotypes underlying these phenotypic differences, will increase in frequency within the population; over time they would replace those organisms with less reproductively successful phenotypes. Darwin termed this process natural selection, in analogy to the process of artificial selection practiced by plant and animal breeders. As we will see, natural selection is one of the major drivers of biological evolution.

Just to be clear, however, reproductive success is more subtle than survival of the fittest. First and foremost, from the perspective of future generations, surviving alone does not matter much if the organism fails to produce offspring. An organism's impact on future generations will depend not on how long it lives but on how many fertile offspring it generates, a definition of success different from the standard English (American) definition. An organism that can produce many reproductively successful offspring at an early age will have more of an impact on subsequent generations than an organism that lives an extremely long time but has few offspring. Again, there is a subtle point here. It is not simply the number of offspring that matter but the relative number of reproductively successful offspring produced.

If we think about the factors that influence reproductive success, we can classify them into a number of distinct types. For example, organisms that reproduce sexually need access to mates, and must be able to deal successfully with the stresses associated with normal existence and reproduction. This includes the ability to obtain adequate nutrition and to avoid premature death from predators and

---

[108] Global population growth & The Joy of Stats

[109] Hans Rosling: Religions and babies

pathogens. Similarly, organisms can cooperate (help) each other, and through such cooperation increase the odds that their offspring will survive, compare to solitary organisms. Both individual and social traits are part of the organism's phenotype, which is what natural selection acts on. It is worth remembering, however, that not all traits are independent of one another. Often the mechanism (and genotype) involved in producing one trait influences other traits – traits are often interdependent and sometimes incompatible, after all they are aspects of a single organism. There are also non-genetic sources of variation. For example, there are molecular fluctuations that occur at the cellular level; these can lead genotypically identical cells to display different behaviors, that is, different phenotypes. Environmental factors and stresses also influence the growth, health, and behavior of organisms. These are generally termed physiological adaptations. An organism's genotype influences how it responds phenotypically to environmental factors, so the relationship between phenotype, genotype, and the organism's environment is complex.

Mutations and the origins of genotype-based variation

So now the question arises, what is the origin of genetic, that is, inheritable variation? How do genotypes change? As a simple and not completely incorrect analogy, we can think of an organism's genotype as a book of instructions. This book is also known as its genome; do not worry if this seems too simple, we will add needed complexities as we go along. An organism's genome is no ordinary book. For simplicity we can think of it as a single unbroken string of characters. In humans, this string is approximately 3.2 billion (~3,200,000,000) characters or letters long and most types of cells in your body contain two very similar, but not identical copies of this book. In case you are wondering, a character corresponds to a base pair within a DNA molecule, which we will consider in detail in Chapter 7. Within this string of characters there are regions that look like words and sentences, that is, regions that look like they have meaning. There are also long regions that appear to be meaningless. To continue our analogy, a few critical changes to the words in a sentence can change the meaning of a story, sometimes subtly, sometimes dramatically, and sometimes a change will lead to a story that makes no sense at all.

At this point we will define the meaningful regions, the words and sentences, as corresponding to genes and the other sequences as intragenic regions, that is, spaces between genes. We estimate that humans have ~25,000 genes; we will return to a molecular level discussion of genes and how they work in Chapters 7 through 9. As we continue to learn more about the molecular biology of organisms, our understanding of both genes and intragenic regions will become more sophisticated. Regions that originally appeared meaningless can be found to influence the meaning of the genome. Many regions of the genome are unique, they occur only once within the string of characters. Others are repeated, sometimes hundreds to thousands of times. When we compare the genotypes of individuals of the same type of organism, we find that they differ at a number of places. For example, over ~55,000,000 variations have been found between all human genomes examined to date, and more are likely to be identified. When present within a population of organisms, these genotypic differences are known as polymorphisms, from the Latin meaning multiple forms. Polymorphisms are the basis for DNA-based forensic identification tests. One thing to note, however, is that only a small number of these variations are present within any one individual, and considering the size of the human genome, most people differ from one another at less than 1 to 4 letters out of every 1000. That amounts to between 3 to 12

million letter differences between two unrelated individuals. Most of these differences are single characters, but there can be changes that involve moving regions from one place to another, or the deletion or duplication of specific regions.

In sexually reproducing organisms, like humans, there are typically two copies of this book in most types of cells of the body, one derived from each of the organism's parents. Organisms (and cells) with two genomic "books" are known as diploid. When a sexual organism reproduces, it produces reproductive cells, known as gametes: sometimes these are the same size. When gametes differ in size, the smaller one is known as a sperm and the larger is known as an egg. Each gamete contains one copy of its own unique version of the genomic book and is said to be haploid. This haploid genome is produced through a complex process known as meiosis (considered in detail in Chapter 11). Meiosis leads to a shuffling of the organism's original parental genomes. When the haploid sperm and haploid egg cells fuse, a new and unique (diploid) organism is formed with its own unique pair of genomic books. The situation is rather different in asexual organisms, and we will discuss the implications later on when we consider horizontal gene transfer.

The origins of polymorphisms

So what produces the genomic variations between individuals found within a population? Are these processes still continuing to produce genotypic and phenotypic variations or have they ended? First, as we have alluded to, and will return to again and again, the sequence of letters in an organism's genome corresponds to the sequence of characters in DNA molecules. A DNA molecule in water (and over ~70% of a typical cell is water) is thermodynamically unstable and can undergo various types of reactions that lead to changes in the sequences of characters within the molecule.[110] In addition, we are continually bombarded by radiation that can damage DNA.[111] Mutagenic radiation, that is, the types of radiation capable of damaging the genome, comes from various sources, including cosmic rays that originate from outside of the solar system, UV light from the sun, the decay of naturally occurring radioactive isotopes found in rocks and soil, including radon, and the ingestion of naturally occurring isotopes, such as potassium-40. DNA molecules can absorb such radiation, which can lead to chemical changes, that is, mutations. Many but not all of these changes can be identified and repaired by cellular repair systems, which we will consider, albeit only briefly, later in the book.

The second, and major source of change to the genome involves the process of DNA replication itself. DNA replication happens every time a cell divides and while remarkably accurate it is not perfect. Copying creates mistakes. In humans, it appears that replication creates one error for every ~100,000,000 (10^8) characters copied. The proof-reading and error repair systems correct ~99% of these errors, leading to an overall error rate during replication of 1 in 10^{10} bases replicated. Since a single human cell contains ~6,400,000,000 (> 6 billion) bases of DNA sequence, that means that less than one new mutation is introduced per cell division cycle. Given the number of generations (cell division cycles) from fertilized egg to sexually active adult, that ends up producing ~100-200 new

---

[110] Instability and decay of the primary structure of DNA & DNA has a 521-year half-life:

[111] Although not not to worry, the radiation energy associated with cell phones, bluetooth, and various wifi devices is too low to damage DNA.

mutations (changes) added to an individual's genome per generation.[112] These mutations can have a wide range of effects, complicated by the fact that essentially all of the various aspects of an organism's phenotype are determined by the action of hundreds to thousands of genes working in a complex network. And here we introduce our last new terms for a while; when a mutation leads to change in a gene, it creates a new version of that gene, which is known as an allele of the gene. When a mutation changes the DNA's sequence, whether or not it is part of a gene, it creates what is known as a sequence polymorphism or simply a polymorphism (a different DNA sequence). Once an allele or polymorphism has been generated, it is stable - it can be inherited from a parent and passed on to an offspring. Through the various processes associated with reproduction, which we will consider in detail later on, each organism carries its own distinctive set of alleles and its own unique set of polymorphisms. Taken together these genotypic differences, that is, differences in alleles and polymorphisms, produce different phenotypes. The DNA tests used to determine paternity and forensic identity work because they use the unique polymorphisms and alleles present within an individual's genome as a type of bar code for that person. We will return to and hopefully further clarify the significance of alleles and polymorphisms when we consider DNA in greater detail later on.

Two points are worth noting about genomic changes or mutations. First, whether produced by mistakes in replication or chemical or photochemical reactions, it appears that these changes occur randomly within the genome. With a few notable and highly specific exceptions there are no known mechanisms by which the environment (or the organism) can specify where a mutation will occur. The second point is that a mutation may or may not influence an organism's phenotype. The effects of a mutation will depend on a number of factors, including exactly where the mutation is in the genome, its specific nature, the role of the mutated gene within the organism, the rest of the genome (the organism's genotype), and the environment in which the organism finds itself. We will consider the factors that influence gene and genome dynamics when we consider the behavior of DNA later on.

Questions to answer:
 22. Explain why superfecundity is required for evolution to occur.
 23. Why is the presence of genetically inheritable variation essential for any evolutionary model?

Questions to ponder:
- What advantages might be associated with self-imposed controls on mating?
- How could behaviors that limit an individual's ability to reproduce arise?

Genotype-phenotype relationships: discrete and continuous traits

When we think about genetic polymorphisms and alleles, it is tempting to assume simple relationships. In some ways, this is a residue from the way you may have been introduced to genetics. Perhaps you already know about Gregor Mendel (1822-1884) and his peas. He identified distinct alleles of particular genes that were responsible for distinct phenotypes - yellow versus green peas, wrinkled versus smooth peas, tall versus short plants, etc. Other common examples might be the alleles associated with sickle cell anemia (and increased resistance to malarial infection), cystic fibrosis, and the major blood types. Which alleles of the ABO gene you inherited determines whether you have O, A,

---

[112] Human mutation rate revealed

B or AB blood type. We will consider what genes are and how they work in greater detail later, but for now it is enough to know that the ABO gene encodes for a polypeptide; this polypeptide is a glycotransferase, that is, a protein catalyst (an enzyme) that adds a specific chemical group, a carbohydrate) to a protein. Differences in the DNA sequences of the A, B, and O alleles results in differences in the polypeptides they encode. The polypeptides encoded by the A and B alleles are active catalysts, but they differ in the reactions that they catalyze – different sugar groups are added by the A and B polypeptides; in contrast the O allele does not encode a functional glycotransferase. Remember you are diploid, so you have two copies of each gene, including the ABO gene, in your genome, one inherited from your mom and one from your dad. The two ABO alleles you inherited from your parents may be the same or different.[113] If they are A and B, the proteins on your red blood cells have both the A and B modifications, resulting in an AB blood type, If they are A and O or A and A, your red blood cells have only the A modification, if they are B and O or B and B, your red blood cells have only the B modification, and if you have O and O, no modification (of this type) occurs and you have an O blood type (→). These are examples of what are known as discrete traits; you are either A, B, AB, or O blood type – there are no intermediates. You cannot be 90% A and 10% B.[114] The situation when the presence of a particular allele uniquely determines a particular trait, as in the case of the ABO gene, is rare – most traits are genetically more complex.
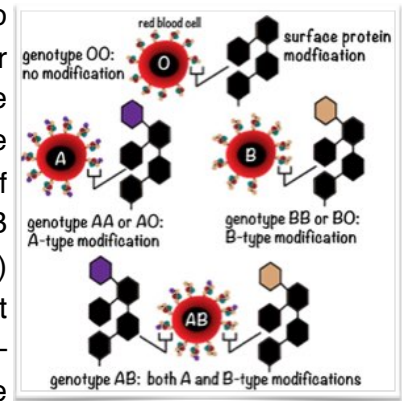
The vast majority of traits are continuous rather than discrete, they involve hundreds to thousands of genes (and their various alleles). For example, people come in a continuous range of heights, rather than in discrete sizes. If we look at the values of the trait within a population, that is, if we can associate a discrete number to the trait (which is not always possible), we find that each population can be characterized graphically by a distribution. For example, let us consider the distributions of weights in a group of 8440 adults in the USA (→). The top panel (A) presents a graph of the weights, along the horizontal or X-axis, versus the number of people with that weight along the vertical or Y-axis. We can define the "mean" or average of the population (x̄) as the sum of the individual values of a trait (in this case each person's weight) divided by the number of individuals measured, as defined by the equation:

$$\bar{x} = \frac{x_1 + x_2 + \cdots + x_n}{n}$$

In this particular case (data set), the mean weight of the population is ~180 pounds. It is common to recognize another characteristic of the population, the median. The median value is the point at which half of the individuals have a smaller value of the trait and half have a larger value. In this case, the

---
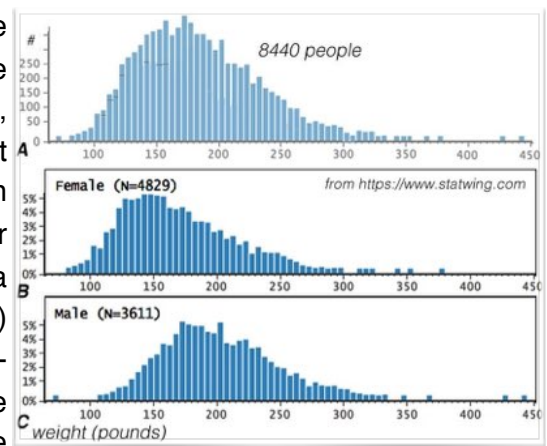
[113] There are a number of common alleles of the ABO gene present in the human population, the most common (by far) are the A, B, and O alleles: http://omim.org/entry/110300

[114] Human blood types have deep evolutionary roots

median is ~176. Because the mean does not equal the median, we say that the distribution is asymmetric, that is there are more people who are heavier than the mean value compared to those who are lighter. Another way to characterize the shape of the distribution is by what is known as its standard deviation, indicated by the Greek letter sigma (σ). There are different ways to calculate the standard deviation that reflect the shape of the population distribution, but for our purposes we will use a simple one, the so-called uncorrected sample standard deviation (→).[115] To calculate this value, you subtract the mean value for the population ($\bar{x}$) from the value for each individual (x_i); since x_i can be larger or smaller than the mean, this difference can be a positive or a negative number. We then take the square of the difference, which makes all values positive (hopefully this makes sense to you). We sum these squared differences together, divide that sum by the number of individuals in the population (N), and take the square root, which reverses the effects of our squaring x_i, to arrive at the standard deviation of the population. The smaller the standard deviation, the narrower the distribution - the more organisms in the population have a value similar to the mean. The larger σ is, the greater is the extent of the variation in the trait.

$$\sigma = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(x_i - \bar{x})^2}$$

So how do we determine whether a complex (that is, determined by many genes and their allelic variants) trait like weight, or any of a number of other non-discrete, continuously varying traits, is genetically determined? We could imagine, for example, that an organism's weight is simply a matter of how easy it was for it to get food. A standard approach to determine whether a trait has a genetic component is to ask whether there is a correlation between the phenotype in the parents (e.g. their heights) and the phenotypes of the offspring (its height). That such a correlation between parents and offspring exists for height is suggested by this graph (→). What we cannot determine from such data, however, is how many genes are involved in the genetic determination of a trait or how their effects are influenced by the environment and the offspring's specific history. As an example, "human height has been increasing during the 19th century when comprehensive records began to be kept. The mean height of Dutchmen, for example, increased from 165cm in 1860 to a current 184cm, a spectacular increase that probably reflects improvements in health care and diet", rather than changes in genes.[116] Geneticists currently estimate that allelic differences at more than ~50 genes make significant contributions to the determination of height, while allelic differences at hundreds of other genes have smaller effects that contribute to differences in height.[117] At the same time, specific alleles of certain genes can lead to extreme shortness or tallness. For example, mutations that inactivate or over-activate genes encoding factors required for growth can lead to dwarfism or gigantism.

---
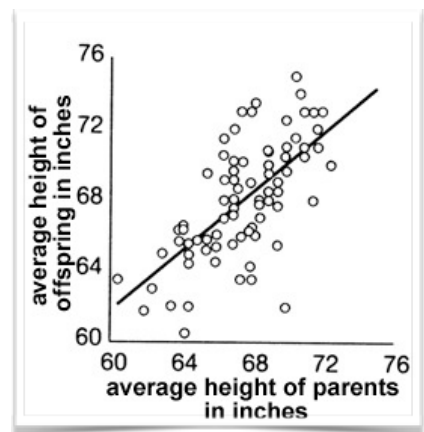
[115] wikipedia: standard deviation & http://www.mathsisfun.com/data/standard-deviation.html

[116] "From Galton to GWAS: quantitative genetics of human height": http://www.ncbi.nlm.nih.gov/pubmed/21429269

[117] Genetics of human height: http://www.ncbi.nlm.nih.gov/pubmed/19818695

On a didaskalogenic note[118], you may remember learning that alleles are often described as if they are either dominant or recessive (a topic we will return to in great depth). But the extent to which an allele is dominant or recessive often depends upon how well we define a particular trait and whether it can be influenced by other factors and other genes. These effects reveal themselves through the fact that people carrying the same alleles of a particular gene can display (or not display) the associated trait, which is known as penetrance, and they can vary in the strength of the trait, which is known as expressivity. Both the penetrance and expressivity of a trait can be influenced by the rest of the genome, that is, the presence or absence of particular alleles of other genes. Environmental factors can also have significant effects on the phenotype associated with a particular allele or genotype.

Variation, selection, and speciation

Combining genetic and associated phenotypic variation, superfecundity, and stable population size, Darwin and Wallace's breakthrough conclusion was that different members of the population would display differences in reproductive success. Some genotypes, and the alleles they contain, would become more common within subsequent generations because the individuals that contained them would reproduce more successfully. Other genotypes would become less common, or disappear altogether. The effects of specific alleles on an organism's reproductive success will, of course, be influenced by the rest of the organism's genotype, its structure and behaviors, both selectable traits (that is traits that influence reproductive success), and its environment. While some alleles can have a strong positive or negative impact on reproductive success, the effects of most alleles are subtle, assuming they produce any noticeable phenotypic effects at all. A strong positive effect will increase the frequency of the allele (and genotype) associated with it in future generations, while a strong negative effect can lead to the allele disappearing altogether. An allele that increases the probability of death before reproductive age is likely to be strongly selected against, whereas an allele that has only modest effects on the number of offspring an organism produces will be selected for, or against, more weakly.

What Darwin and Wallace did not know was that genetic information is stored in molecules of DNA, and that that information can be altered through a variety of mechanisms (mutations) that include sequence duplication, deletion, and recombination (shuffling). Moreover, because DNA molecules are relatively stable they can survive the death of the organism, be released into the environment, and (under certain conditions) be transferred into living organisms and become part of their genetic material. These are all features of the molecular nature of genetic information (genes) and how DNA is manipulated, that is, replicated, repaired, and used to express information within cells. Recognizing this fact led to what is known as the Modern Synthesis of evolutionary theory.[119] While the basic Darwinian rules are the same, the possible molecular complexities make evolutionary processes even more powerful. We will be considering these various molecular processes as we proceed.

---

[118] We call instruction/instructor-dependent thinking didaskalogenic:

[119] Modern synthesis in evolutionary biology

Questions to answer:

23. How would you explain the observation that the products of artificial selection are not generally competitive with "native" organisms?
24. What does the word correlation mean to you? what does it mean mathematically?
25. If an individual's height is determined by the genetics of their parents, then why don't all of the individual height measurements line on a straight line? Where does the scatter come from?
26. Consider a population and generate graphs that display the effects of larger and smaller standard deviations as well as median values that are higher or lower than the mean.

Types of simple selection

While it is something of an oversimplification, we begin with three basic types of selection: stabilizing (or conservative), directed, and disruptive. We will then introduce the complexities associated with the random aspects of reproduction and the linked nature of genes. We start with a population composed of individuals displaying genetic variation in a particular trait. The ongoing processes of mutation continually introduces new genotypes, and their associated effects on phenotype. The effects of mutations can range from the lethal, the organism that carries the mutation either dies or produces no offspring, to completely neutral – an organism that carries the mutation displays no obvious change in phenotype. A complicating factor, that we will consider in more detail later, is that the phenotypic effects of a particular mutation, leading to a mutant or alternative allele, often depend upon the rest of the genome - due to so called genetic background effects. At the same time, changes in the population and the general environment influence the predominant types of selection that occur over time, and different types of selection may well (and most certainly are) occurring for different traits.

For each type of selection, we will illustrate the effects as if they were acting along a single dimension, for example smaller to larger, stronger to weaker, lighter to darker, or slower to faster. In fact, most traits vary along a number of dimensions. For example, consider the trait of ear, paw, heart, or big toe shape. An appropriate type of graph would be a multi-dimensional surface, but that is harder to draw. It is also possible that a genotype that influences one trait may also influence another, apparently independent, trait. Also, for simplicity, we will start with populations whose distribution for a particular trait can be described by a simple and symmetrical curve, that is the mean and the median are the same. New variants, based on new mutations (new alleles and combinations of alleles), generally fall more or less randomly within this distribution. Under these conditions, for selection NOT to occur we would have to make an seriously unrealistic assumption, namely that an organism (or a pair of organisms, assuming that this is a sexually reproducing species) are all equally successful at surviving and producing offspring, something observably not true. Any time genetic variation influences reproductive success selection occurs, although the strength of selection (the average difference in the number of viable offspring produced) may vary dramatically between traits.

Stabilizing selection: Sometimes a population of organisms appears static for extended periods of time, that is, the mean and standard deviation of a trait are not changing. Does that mean that selection has stopped? Obviously we can turn this question around: if we assume that there is a population with a certain stable mean and standard deviation of a trait – what would happen over time if selection disappeared?

Let us assume we are dealing with an established population living in a stable environment. This is a real world population, where organisms are superfecund, that is, capable of reproducing more, and sometimes, many more organisms than are needed to replace them when they die and that these organisms mate randomly with one another. Now we have to consider the factors that lead to the original population distribution: why is the mean value of the trait the value that it is? What factors influence the observed standard deviation? Assuming that natural selection is active, it must be that organisms that display a value of the trait far from the mean are (on average) at a reproductive disadvantage compare to those with the mean value of the trait (→). We do not know why this is the case and don't really care at the moment. Now if selection, at least for this value of the trait is inactive, what happens? The organisms far from the mean are no longer at a reproductive disadvantage, so their numbers in the population will increase. The standard deviation will grow larger, until at the extreme, the distribution will be almost flat, characterized only by a maximum and a minimum value. New mutations and existing alleles that alter the trait will not be selected against, so they will increase in frequency. But in a real population, the mean and standard deviation associated with the trait remain constant, assuming that the environment is constant. We therefore predict "negative" selection against extreme values of the trait, which means that these individuals tend to produce fewer viable offspring than those with a value of the trait near the mean.[120] We can measure that degree of selection "pressure" by following the reproductive success of individuals with different values of the trait. We might predict that the more extreme the trait, that is, the further from the population mean, the greater its reproductive disadvantage (negative selection) will be, so that with each generation, the contribution of these outliers in the population is reduced. The distribution's mean will remain constant. The stronger the disadvantage, referred to as negative selective pressure, the outliers face, the narrower the distribution will be – that is, the smaller the standard deviation. In the end, the size of the standard deviation will reflect both the strength of selection against outliers and the rate at which new variations enters the population through mutation. Similarly, we might predict that where a trait's distribution is broad the impact of the trait on reproductive success will be relatively weak.
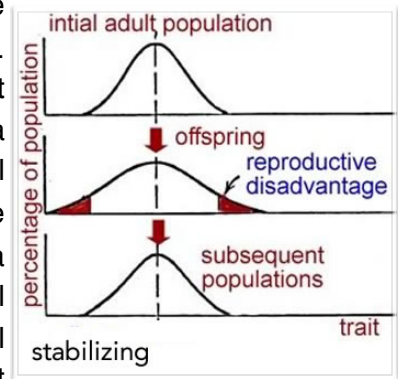
Directed selection: Imagine that the population's environment changes. It may now be the case that the phenotype of the mean is no longer the optimal phenotype in terms of reproductive success, the only factor that matters, evolutionarily; a smaller or a larger value may be more favorable. Under these conditions we would expect that, over time, the mean of the distribution would shift toward the phenotypic value associated with maximum reproductive success (→). Once reached, and assuming the environment stays constant, stabilizing selection again becomes the predominant process. One outcome to emerge from a changing environment leading to directed selection is that as the selected population's mean moves, it may well alter the environment of other organisms.

---
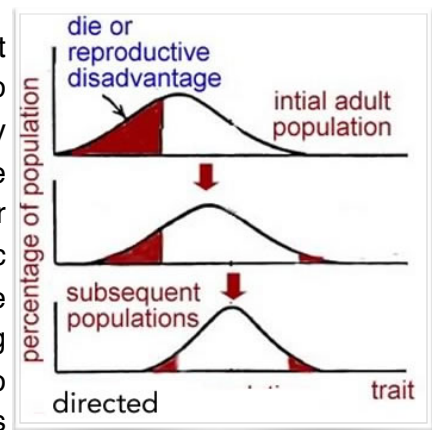
[120] By "viable" we mean offspring that live to reproduce, and that themselves reproduce successfully.

For directed selection to work, the environment must change at a rate and to an extent compatible with the changing mean phenotype of the population. Too big and/or too rapid a change and the reproductive success of all members of the population could be dramatically reduced. The ability of the population to change will depend upon the genetic variation already present within the population and the rate at which new mutations are produced, a relatively slow process.[121] In some cases, the change in the environment is so fast or so drastic and the associated impact on reproduction so severe that selection will fail to move the population and extinction will occur.

Disruptive selection: A third possibility is that a population of organisms find themselves in an environment in which traits at the extremes of the population's phenotypic distribution have a reproductive advantage over those around the mean. If we think about the trait distribution as a multidimensional surface, it is possible that in a particular environment, there will be multiple distinct strategies that lead to greater reproductive success compared to others. This leads to what is known as disruptive selection (→). In an asexually reproducing population, various lineages will be subject to selective pressure based on the environments they come to inhabit, and the likelihood that individuals move from environment to environment, or that the environment changes dramatically. The effect of disruptive selection in a sexually reproducing population will be opposed by the random mating between members of the population, which does not occur in asexual populations. But is random mating a good assumption? It could be that the different environments, which we will refer to as ecological niches, are physically distant from one another and that organisms do not travel far to find a mate. The population will split into subpopulations in the process of adapting to the two different niches. Over time, two species could emerge, since whom one chooses to mate with and the productivity of that mating, are themselves selectable traits. Disruptive selection will overtime lead to the generation of new species, and over long periods of time, the millions of existing species and the ever greater number of extinct species. The diversity of life was the observation that Darwin and Wallace originally set out to explain, and evolutionary processes provide a plausible mechanism.
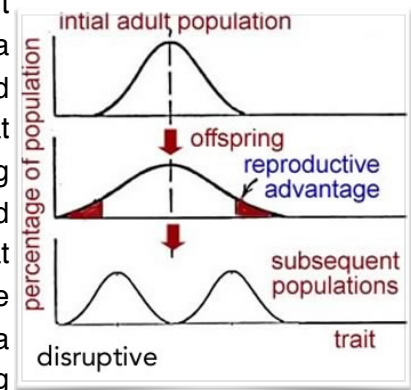
Questions to answer:
27. Why does variation never completely disappear even in the face of strong stabilizing selection?
28. Under what conditions would stabilizing selection be replaced by directed or disruptive selection?
29. By looking at a population, how might one estimate the strength of conservative selection with respect to a particular trait?

Questions to ponder:
- Why is it difficult to be sure you know why a particular allele or trait was selected?
- How might phenotypic variation influence the choice of a mate (during sexual reproduction)?

---

[121] As we will consider later when we consider these molecular processes, there are times when physiological stress can lead to increased global mutations rate. Mutation as a Stress Response and the Regulation of Evolvability

Considering stochastic processes

Biological systems are characterized by what are known as stochastic processes. We will find that stochastic processes play an important role in evolutionary mechanisms (population bottlenecks, founder effects, genetic drift, meiotic recombination - all discussed below) as well as molecular processes within cells and tissues (again discussed later on). You may not be familiar with the word stochastic, it is a word whose meaning is often confused with random. So, what exactly distinguishes a stochastic from a random process? A truly random process has no underlying natural cause and so is completely unpredictable. A miracle could be considered a random process. From a scientific perspective, one could argue that there are no truly random natural processes or events, no miracles. Our working hypothesis is that all natural events have identifiable and measurable causes. That said, that does not mean that every individual event can be predicted. Natural events can be unpredictable for one of two basic reasons: the event may be determined by theoretically unknowable or currently unknown factors, as in the case of the radioactive decay of atoms. Alternatively, the event may be the result of a large number of theoretically knowable events that are, for a variety of practical reasons, impossible to measure accurately. Such events are analogous to, or versions of, Brownian motion, a phenomena named after the Scottish botanist Robert Brown (1773-1858). In Brownian motion, small, but visible particles suspended in a solution (air or water) are found to move in a jerky and irregular manner (A→). Brownian motion arises because the visible particle is colliding with many invisible objects (molecules) present in the environment (air/water: B→).[122] The average energy transferred through these collisions reflects the temperature of the system. At higher temperatures the molecules have a higher average (mean) kinetic energy ($\frac{1}{2} mv^2$). During a particular time interval, the sum of all collisions can lead to an unbalanced force on the particle that causes it to move. A short time later the vector sum of these collision forces is likely to point in a different direction and the particle will now move in that direction. Collisions between molecules supply the energy to drive the dissociation of molecules from one another and supply the activation energy required for chemical reactions to proceed, topics that we will return to when we consider the thermodynamics of reaction systems (Chapter 5). At the individual event level, the system is unpredictable in practice (but not in theory) because there are so many molecules and collision events involved – for example, in water there are ~3×10^{22} water molecules per cubic centimeter, with the average water molecule traveling ~2.5×10^{-8} centimeters between collisions.[123] The end result is that the speed and direction of visible particle and invisible molecule movements are constantly changing.
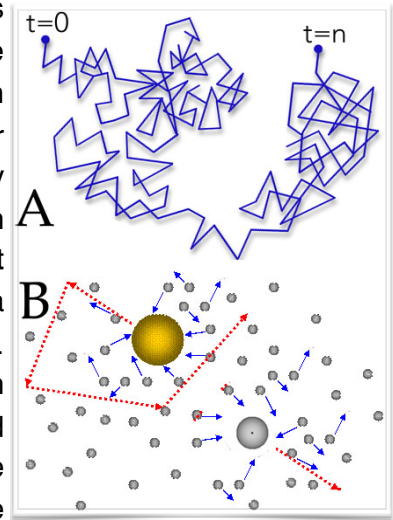
In classical (that is, pre-quantum mechanical) physics, it was assumed that if it were possible to know the velocity (speed and direction) of every molecule in the system, as well as the dynamics of the

---

[122] Albert Einstein: The Size and Existence of Atoms & Einstein and Brownian Motion

[123] The properties of water: http://galileo.phys.virginia.edu/classes/304/h2o.pdf

collisions, we could predict the future behavior of the system and the paths of Brownian movements.[124] But it turns out that the world does not behave that way. In fact, we cannot (even theoretically) achieve this level of accurate measurement; we are limited by what is known as the Heisenberg Uncertainty principle, which arises from the fact that matter is composed of objects with both wave- and particle-like properties, rather than simple billiard ball-like particles.[125]

So why isn't Brownian motion evidence of a random process? What makes it possible for it to be studied scientifically? The answer is based on the fact that when we look at many objects, the behavior of the population becomes predictable – this predictability implies an underlying cause. For example, consider measurements of a large number of particles undergoing Brownian movement. If we measure the distance between where they start (t=0) and where they end up (t=n) as a function of time (see A↑ above), we find that the average distance travelled (but not the direction of travel) is predictable and reflects the size of the particle, the nature of the system (water, air, etc), and its temperature. Its predictability indicates that Brownian motion is due to underlying (calculable) physical processes.

The situation is similar to that of rolling dice. While it is impossible to accurately predict the outcome of a single dice roll, as we increase the number of rolls (the population of rolls), we find increasingly predictable behavior, each of the six numbers (assuming that this is a fair cube dice) will appear 1/6th of the time. The larger the number of rolls, the more closely the number of each possible outcome will approach 1/6th of the total. While the outcome of any individual roll is unpredictable, the behavior of a population of rolls is predictable – a behavior known as the law of large numbers. A similar situation occurs with radioactive atoms; while it is impossible to predict when any particular atom will decay, we find that when we consider a large enough population we can accurately predict when any particular percentage of the original population will have decayed. Typically, the time it takes for 50% of the original atoms to decay is known as the "half-life" of the isotope and can be determined to very high accuracy.

In the case of rolling dice, and other similar (simple) stochastic processes, it is important, but hard to remember, that each individual event is independent, what happened in the past does not influence what will happen next. Forgetting this rule leads to what is known as the Gambler's Fallacy.[126] As an example, you roll a die eight times and get 2, 2, 5, 2, 2, 6, 2, 2. Assuming of course that this is a fair dice, what is the probability that the next roll will come up 2? No matter how many times a 2 came up in the past, the chance of rolling a 2 on next roll remains the same, 1/6.

A complexity that occurs within biological systems is that while a particular event can be stochastic, individually unpredictable, but well behaved in a large enough population, in the context of a cell or the organism, a single event can change the system leading to different long term outcomes. For example, a mutation can start the process of a cell becoming cancerous It is therefore possible, and perhaps likely, that if the history of the organism (or life) were to be "rerun" (a completely impossible situation), the outcomes would be different.

_____

[124] see Laplace's demon: https://en.wikipedia.org/wiki/Laplace's_demon

[125] Luckily for you, this is not a physics course, so the details of Heisenberg and his principle are sketched out in only a superficial way here. Need to know more, check out: What is the Heisenberg Uncertainty Principle?

[126] Gambler's Fallacy: https://en.wikipedia.org/wiki/Gambler's_fallacy

Questions to answer:

30. What types of behaviors define a stochastic event; what types of everyday stochastic events are you familiar with. How do you know that they are not random?
31. What types of events are not, in theory, study-able scientifically?

Population size, founder effects and population bottlenecks

When we think about evolutionary processes from a strictly selection-based perspective, we ignore important factors that can impact the evolution of a population. For example, what happens when a small number of organisms (derived from a much larger population) colonize a new environment? This is a situation that produces what is known as a founder effect. Something similar happens when a large population is dramatically reduced in size for any of a number of reasons, a situation known as a population bottleneck. In both founder effects and population bottlenecks, the small populations that result can have different allele frequencies than the original "parental" population and are more susceptible to the effects of stochastic, non-selective effects, a process known as genetic drift. Together founder effects, bottlenecks, and drift can produce populations with unique traits that are not directly due to the effects of natural selection. Since founder effects and population bottlenecks can occur a number of times during the course of a populations' evolution, it is a mistake to assume that all observed traits have positive effects on reproductive success. If we think of evolutionary change as reflecting the movement of a population through a fitness landscape–the combination of the various factors that influence reproductive success over time–then the isolation of small populations, and evolutionary change within them, can cause a random jump from one place in the landscape to another. Once in the new position, and as the population grows larger, new adaptations can be possible – selection again becomes the main, but not exclusive, driver of evolutionary change. Deleterious effects, that become frequent due to non-adaptive processes, can be ameliorated. A population invading a new environment will encounter a new set of organisms to compete and cooperate with. A catastrophic environmental change will change the selective landscape, removing competitors, predators, pathogens, and cooperators, favoring new adaptations and selecting against others that might have once been beneficial, in terms of reproductive success. One effect of the major extinction events that have occurred during the evolution of life on Earth is that they provide a new adaptive context, a different and less densely populated playing field with fewer direct competitors.[127] The expansion of various species of birds and mammals that followed the extinction of the dinosaurs is an example of one such opportunity, associated with changes in selective pressures.

Founder effects

What happens when a small subpopulation, a few individuals, becomes isolated, for whatever reason, from its parent population? The original (large) population will contain a number of genotypes and alleles. If this population is in a new environment it will be governed primarily by directed and conservative selection. We can characterize this parental population in terms of the frequencies of the various alleles present within it. For the moment, we will ignore the effects of new mutations, which will continue to arise within the population but at a slow rate. Now assume that a small group of organisms

---

[127] Big Five mass extinction events

comes to colonize a new, geographically separate environment such that it is reproductively isolated from its parental population – no individuals travel between the parent and the colonizing population.

The classic example of such a situation is the colonization of newly formed islands, but the same process applies more generally during various types of migrations. By chance, the frequency of alleles in a small isolated population is likely to be different from the allele frequencies found in the much larger parent population. Why is that? It is a question of the randomness of sampling of the original population. Consider, as an example, rolling a die. If rolled a large enough number of times, a fair six-sided (cubical) die will be produce the numbers 1, 2, 3, 4, 5, and 6 with equal probabilities. Each will appear 1/6th of the time. But imagine that the number of rolls is small. Would you expect to get each number appearing with equal probability? You can check your intuition using various on-line dice applets, but the answer is decidedly NO!!![128] See how many throws are required to arrive at an equal 1/6th probability distribution; the number is almost certainly much larger than you would guess.

We can apply this "law of large numbers" to populations using the following logic. First, we recognize that if we wanted to determine the exact frequency of each allele of a particular genetic locus or gene in a particular population at a particular time, it would require that we determine which allele(s) are present in each individual, BUT that is quite an intensive, expensive, and often impossible task. So we have to use some other method to estimate allele frequencies – we turn to "sampling". We examine a random set of individuals, a sample. If the number in the sample is small with respect to the total population size, we can expect significant differences in measured (sampled) and actual (total) population allele frequencies. These differences become smaller as the sample size increases. To provide a concrete example, consider a large population in which each individual carries one (and only one) of six alleles of a particular gene and that the percentage of each type is equal (1/6th). The selection of any one individual from this population is like a throw of the fair die; there is an equal 1/6th chance of selecting an individual with one of the six alleles. Since the parental population is large, the removal of one individual does not appreciably change the distribution of alleles remaining, so the selection of a second individual produces a result that is independent of the first, just like individual rolls of the die and are equally likely to result in a 1/6th chance to select any one of the six alleles. But producing a small subpopulation with 1/6th of each allele (or the same percentages of various alleles as are present in the parent population) is, like the die experiment above, extremely unlikely. The more genotypically complex the parent population, the more unlikely it is; imagine that the smaller colonizing population only has, for example, 3 members (three rolls of the die) – not all alleles present in the original population can possibly be represented. Similarly, the smaller the subpopulation the more likely that the new subpopulation will be genetically different from the original population. So when a small group from a parent population invades or migrates into a new environment, it is likely to have a different genotypic (allelic) profile compared to its parent population. This difference is not due to natural selection but rather to chance alone. Nevertheless, it will influence subsequent evolutionary events; the small subpopulation will likely respond in different ways to new mutations and environmental pressures based on which alleles are present. The situation will be further influenced if genetic factors impact migratory behavior or reproductive success in the new environment.

---

[128] Here is a reasonably good one: http://www.math.uah.edu/stat/apps/DiceExperiment.html

The human species appears to have emerged in Africa ~500,000 years ago.[129] The people living in Africa represent the parent population of *Homo sapiens* and genetic studies reveal that the African population displays a much greater genotypic complexity than do groups derived from it, that is, everyone else. What remains controversial is the extent to which migrating populations of humans in-bred with what are known as archaic humanoids (such as Neanderthals and the Denisovians), which diverged from our lineage (*Homo sapiens*) ~1.2 million years ago.[130]

Population bottlenecks

A population bottleneck is similar, but distinct in important ways from a founder effect. Population bottlenecks occur when some environmental change leads to the dramatic reduction in the size of a population. Catastrophic environmental changes, such as asteroid impacts, massive and prolonged volcanic eruptions associated with continental drift, or the introduction of a particularly deadly pathogen that kills a high percentage of the organisms that it infects, can all create population bottlenecks (↓). Who survives the bottleneck can be random, due only to luck, or based on genetic factors, for example, associated with disease resistance.
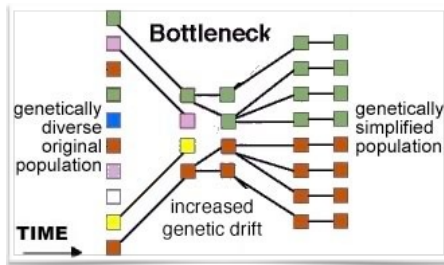
There is compelling evidence that such drastic environmental events are responsible for population bottlenecks so severe that they led to mass extinctions. The most catastrophic of these extinction events was the Permian extinction that occurred ~251 million years ago, during which it appears that ~95% of all marine species and ~75% of land species went extinct.[131] If most species were affected, we would not be surprised if the surviving populations experienced serious bottlenecks. The subsequent diversification of the surviving organisms, such as the Dinosauria, which includes the extinct dinosaurs and modern birds, and the Cynodontia, which includes the ancestors of modern mammals, including us, could be due in part to these bottleneck-associated effects, for example, through the removal of competing species or predators. An asteroid impact, known as the Cretaceous-Tertiary event, occurred ~65 million years ago; it contributed to the extinction of the dinosaurs and led to the diversification of mammals, which had first appeared in the fossil record ~100 million years earlier.

While surviving an asteroid impact, or other dramatic changes in climate may be random, in other cases who survives a bottleneck is not. Consider the effects of a severe drought or highly virulent bacterial or viral infection; the organisms that survive may have specific phenotypes, and associated genotypes, that significantly influence their chance of survival. In such a case, the effect of the bottleneck event would produce non-random changes in the distribution of genotypes (alleles) in the post-bottleneck population – these selective effects could continue to influence the population in various ways. For example, a trait positively associated with pathogen resistance may also have negative phenotypic effects. After the pathogen-driven bottleneck, mutations that mitigate the

---

[129] Although dating origins depends upon finding fossils: see Oldest Homo sapiens fossil claim rewrites our species' history

[130] Genetic Data and Fossil Evidence Tell Differing Tales of Human Origins

[131] The Permian extinction and the evolution of endothermy

resistance trait's negative effects, and may have their own effects, can have a selective advantage (that is, increase reproductive success). The end result is that traits that would not be selected in the absence of the pathogen, are selected and become common. In addition, the very occurrence of a rapid and extreme reduction in population size has its own effects. For example, it would be expected to increase the effects of genetic drift (see below) and could make finding a mate more (or less) difficult.

We can identify extreme population reduction events, such as founder effects and bottlenecks, by looking at the variation in genotypes, that is, the sequence of DNA molecules, particularly sequence changes not expected to influence phenotypes, mating preference, or reproductive success. These so-called neutral polymorphisms are expected to accumulate in the regions of the genome between genes (intragenic regions) at a constant rate over time (can you suggest why?) The rate of the accumulation of neutral polymorphisms serves as a type of population-based biological clock. Its rate can be estimated, at least roughly, by comparing the genotypes of individuals of different populations whose time of separation can be accurately estimated, assuming of course that there has been no significant migration between the populations.

Such studies of genomic sequence data, which we will return to later in greater detail, indicate that the human population arose in Africa ~500,000 years ago.[132] Before this, the ancestral population leading to modern humans (*Homo sapiens*) appears to have undergone a bottleneck around ~1.2 million years ago.[133] Once established, groups of modern humans migrated within and out of Africa (→), undergoing a series of founder effect events between ~45,000 to ~60,000 years ago. Groups (small populations) of humans migrated out of southern Africa into the Horn of Africa, then into the Arabian peninsula, and from there into Europe, Asia, Oceania, and finally into North America and then through central and South America. Comparing genotypes, that is, neutral polymorphisms, between isolated populations enables us to estimate that aboriginal Australians reached Australia ~45,000 years ago and that humans arrived in the Americas in multiple waves beginning ~16,000 years ago. The arrival of humans into a new environment has been linked to the extinction of a group of mammals known as the megafauna in those environments.[134] The presence of humans changed the environmental pressures on these organisms around the world.
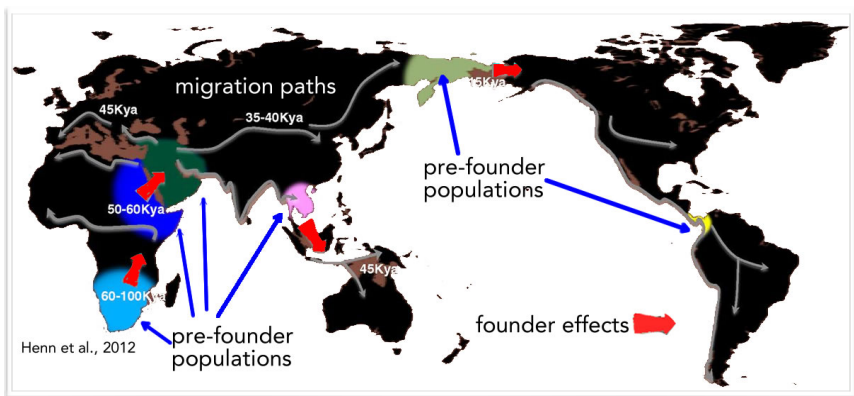
Henn et al., 2012

---

[132] The great human expansion

[133] Mobile elements reveal small population size in the ancient ancestors of Homo sapiens:

[134] http://australianmuseum.net.au/Megafauna-extinction-theories-patterns-of-extinction and an interesting video: http://youtu.be/8WZ5Q2JYbLY

Questions to answer:

32. What happens if a sample of a population is not random, how does that influence one's conclusions about the behavior of the larger population?
33. How does the extinction of one type of organism influence the evolution of others?
34. What factors make a bottleneck different from a founder effect?
35. How can a founder effect/bottleneck lead to deleterious alleles becoming more frequent in a population? How does that impact future evolution?

Questions to ponder:

- How is determining allele frequency in a population similar to and different from political polling?

Genetic drift

Genetic drift is a stochastic process that becomes important in small populations or over long periods of time. It leads to non-adaptive evolutionary phenomenon that explain a number of observations. Consider the observation that many primates are strictly dependent on the presence of vitamin C (ascorbic acid) in their diet. Primates are divided into two suborders, the Haplorhini, from the Greek meaning "dry noses", and the Strepsirrhini, meaning "wet noses". The Strepsirrhini include the lemurs and lorices, while the Haplorhini include the tarsiers and the anthropoids, monkeys, apes, and humans. The Haplorhini, but not the Strepsirrhini, all share a requirement for vitamin C in their diet. In vertebrates, vitamin C plays an essential role in the synthesis of collagen, a protein involved in the structural integrity of a wide range of tissues. In humans, the absence of dietary vitamin C leads to the disease scurvy, which according to Wikipedia, "often presents itself initially as symptoms of malaise and lethargy, followed by formation of spots on the skin, spongy gums, and bleeding from mucous membranes. Spots are most abundant on the thighs and legs, and a person with the ailment looks pale, feels depressed, and is partially immobilized. As scurvy advances, there can be open, suppurating wounds, loss of teeth, jaundice, fever, neuropathy, and death."[135]

The requirement for dietary vitamin C in the *Haplorhini* is due to a mutation in a gene, known as gulo1, which encodes the enzyme 1-gulono-gamma-lactone oxidase (Gulo1) that is required for the synthesis of vitamin C. One can show that the absence of a functional gulo1 gene is the root cause of vitamin C dependence in *Haplorhini* by putting a working copy of the gene, for example derived from a mouse, into human cells. The mouse-derived gulo1 allele, which encodes a functional form of the Gulo1 enzyme, "cures" the human cells' need for exogenous vitamin C. But, no matter how advantageous a working gulo1 allele might be, particularly for British sailors, who died in large numbers before a preventative treatment for scurvy was discovered[136], no new, functional gulo1 allele has appeared in the lineage leading to modern humans and other Haplorhini. Since mutation is a stochastic process, organisms do not always produce the genes or alleles they need or that might be beneficial. Alleles are selected from alleles already present in the population or that appear through *de novo* (new) mutation. In some cases there may be no plausible molecular pathway that can generate such an allele (or such a gene).

---

[135] An amazing fact is that it took the deaths of thousands of sailors to understand the nutritional role of vitamin C.

[136] http://mentalfloss.com/article/24149/how-scurvy-was-cured-then-cure-was-lost

The mutant gulo1 allele appears to have become fixed, that is the only gulo1 allele present in the ancestral population that gave rise to the Haplorrhini ~40 million years ago. So the question is, how did we (that is our ancestors) come to loose a functional version of such an important gene? It seems obvious that when the non-functional allele became fixed in that population, the inability to make vitamin C cannot have been strongly selected against, its loss would appear to have led to negative little or no effect on reproductive success pressure. We can imagine such an environment and associated behavior; namely, these organisms must have obtained sufficient vitamin C from their diet, so that the loss of their own ability to synthesize vitamin C had little negative effect on them.

So how were the functional alleles involved in vitamin C synthesis lost? In small populations, non-adaptive – that is, non-beneficial and even mildly deleterious genotypic changes and their associated traits can increase in frequency through a process known as genetic drift. In such populations, selection continues to be active, but it has significant effects only for traits and their associated alleles when the trait strongly influences reproductive success. While genetic drift occurs in asexual populations, it is due to random effects on organismic survival, which can, in practice be difficult to distinguish from selective effects. In contrast, drift is unavoidable in small populations of sexually reproducing organisms. This is because cells known as gametes are produced during the process of sexual reproduction (Chapter 4). While the cell that generates these gametes contains two copies of each gene, and each gene reflects one of the alleles present within the population, any particular gamete contains only a single allele of each gene. Two gametes then fuse to produce a new diploid organism. This process combines a number of chance events: including which allele is present in a particular gamete and which gametes fuse to produce a new organism. Moreover, not all gametes (something particularly true of sperm) become part of the next generation. In a small population, over a reasonably small number of generations, one or the other alleles at a particular genetic locus is likely to be lost simply by chance. In this figure (→), six different experimental outcomes (each line) are analyzed over the course of 100 generations. In each case, the population size is set to 50; at the start of the experiment half

the individuals have one allele and half have the other. While we are watching only one genetic locus, this same type of behavior impacts every gene for which multiple alleles exist. In one of these six populations, one allele has been lost (red dot), in the other (blue dot) the other allele is close to being lost. When a particular allele becomes the only allele within a population, it is said to have been fixed. Assume that the two alleles convey no selective advantage with respect to one another, can you predict what will happen if we let the experiment run through 10,000 generations? If you are feeling mathematically inclined, you can even calculate (or estimate) the effect of mild to moderate positive or negative selective pressures on allele frequencies and the probability that a particular allele will be lost or fixed.
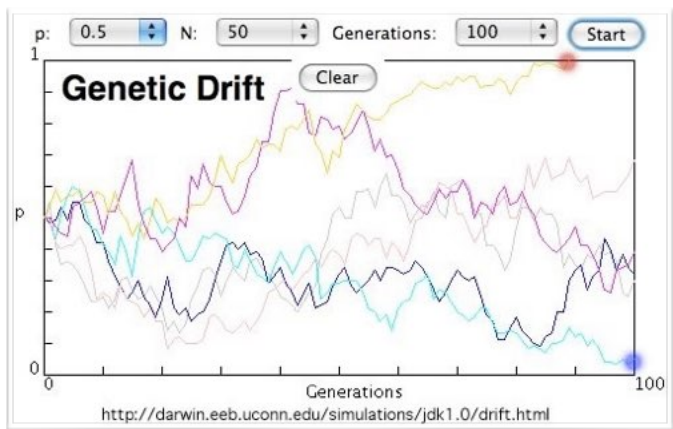
Since the rest of the organism's genotype can influence the phenotype associated with a particular allele, the presence or absence of various alleles within the population can influence the phenotypes observed (a topic we will return to in chapter 12). If an allele disappears because of genetic drift, future evolutionary changes may be constrained, or perhaps better put, redirected. At each point, the future directions open to evolutionary mechanisms depend in large measure on the alleles currently present in the population. Of course new alleles continue to arise by mutation, but they are originally infrequent, just one of each in the entire population, so unless they are strongly selected for (and even if they are selected for) they are likely to disappear from the population by genetic drift.[137] Drift can lead to some weird outcomes. For example, what happens if drift leads to the fixation of a mildly deleterious allele, let us call this allele BBY. Now the presence of BBY will change the selective landscape: mutations and or alleles that ameliorate the negative effects of BBY will increase reproductive success, selection pressures will select for those alleles. This can lead to evolution changing direction even if only subtly. With similar effects going on across the genome, one quickly begins to understand why evolution is something like a drunken walk across a selective landscape, with genetic drift, founder and bottleneck effects resulting in periodic staggers in random directions.

The use of pre-existing variation, rather than the idea that an organism invents variations in its genome as they are required, was a key point in Darwin's view of evolutionary processes. The organism cannot create the alleles it needs or "wants", nor are there any known processes that can produce specific alleles in order to produce specific phenotypes. Rather, the allelic variation generated by mutation, selection, and drift are all that evolutionary processes have to work with.[138] Only a rare mutation that recreates the lost allele can bring an allele back into the population once it has been lost. Founder and bottleneck effects, together with genetic drift combine to produce what are known as non-adaptive processes and make the history of a population a critical determinant of its future evolution.

Questions to answer:
36. How does natural selection influence the effects of genetic drift and vice versa?
37. Describe the relative effects of selection and drift following a bottleneck.
38. How is it that drift (the probability of allele loss) can be accurately quantified, but is unpredictable in any particular population?

Questions to ponder:
- Does passing through a bottleneck improve or hamper a population's chances for evolutionary success?

A reflection on the complexity of phenotypic traits

We can classify traits into three general types: adaptive, non-adaptive, and deleterious. Adaptive traits are those that, when present increase the organism's reproductive success. These are the traits we normally think about when we think about evolutionary processes. Non-adaptive traits are those generated by stochastic processes, like drift, founder effects, and bottlenecks. These traits

---

[137] If the population is small, instead of disappearing, any particular mutation (allele) could become fixed through genetic drift - use the genetic drift applet and look for examples where an allele almost disappears and then becomes fixed; it does happen.

[138] The exception involves the transfer of genes from organism to organism, a process known as horizontal gene transfer, which we will come to.

become established not because they improve reproductive success but simply because they happened to have become fixed within the population. If an allele is deleterious independent of its environment, it will be expected to rapidly disappear from the population, unless other factors are in play. Rare, strongly deleterious alleles are, most likely, the result of new mutations.

When we consider a deleterious allele we are always referring to its effects on reproductive success. An allele can harm the individual organism carrying it yet persist in the population because it improves reproductive success, that is, it leads to an increased number of viable offspring. Similarly, there are traits that can be seen as actively maladaptive, but which occur within the population because they are linked mechanistically to some other positively selected, adaptive trait. Many genes are involved in a number of distinct processes and their alleles can have multiple phenotypic effects. Such alleles are said to be pleiotropic, meaning they have multiple effects on an organism's phenotype. Not all of the pleiotropic effects of an allele are necessarily of the same type; some can be beneficial, others deleterious. As an example, a trait that dramatically increases the survival of the young, and so increases their potential reproductive success, but leads to senility and death in older adults could well be positively selected for. In this scenario, the senility trait is maladaptive but is not eliminated by selection because it is mechanistically associated with the highly adaptive juvenile survival trait. What is happening is a form of cost-benefit analysis; it the net benefit exceeds the cost, the trait (and the alleles responsible for generating it) will be positively selected. If the costs exceed the benefits, it will be selected against. It is worth noting that a trait that is advantageous in one environment may be disadvantageous in another, think the effects of diet on the effects of the gulo1 mutation. All of which is to say that when thinking about evolutionary mechanisms, do not assume that a particular trait exists independently of other traits, that it functions in the same way in all environments, or that the presence of a trait is evidence that it is beneficial.

Gene linkage: one more complication

So far, we have not worried overly much about the organization of genes in an organism. We also have not consider what, exactly a gene is. For now, let us just say that a gene is information encoded within a region of a molecule of DNA (deoxyribonucleic acid) and that multiple genes can be found within a single DNA molecule – we will consider specific aspects of genes below and then in much greater detail in chapter 7.

It could be that each gene behaves like an isolated object, but in fact that is not the case. We bring it up here because the way genes are organized can, in fact, influence evolutionary processes. In his original genetic analyses, Gregor Mendel (1822–1884) spent a fair amount of time looking for "well behaved" genes and alleles, those that displayed simple recessive and dominant behaviors and that acted as if they were independent from one another.[139] But it quickly became clear that these behaviors are not how most genes behave. In fact, genes often act as if they are linked together, because often they are; gene linkage arises from the organization of genes within chromosomes, that is individual DNA molecules. So what happens to linked genes when a particular allele of a particular gene is strongly selected for or against? That allele, together with whatever alleles are found in genes linked to

---

[139] Mendelian controversies: a botanical and historical review: http://www.amjbot.org/content/88/5/737.full

it, are also selected. We can think of this as a by-stander, or sometimes termed a "piggy-back" effect, where an allele's frequency in a population increases (or decreases) not because of its direct effects on reproductive success, but because of its location within the genome, its "linkage" to an allele that strongly influences selection.

As we will see later on in the course, linkage between alleles (or between genes) is not a permanent situation; there are processes (meiotic recombination) that can shuffle the alleles on a chromosome. The end result of such recombination events is that the further away two genes are from one another on a chromosome, the more likely it is that alleles of those genes will appear to be unlinked, that is, have independent effects on reproductive success. Over time, the effects of linkage will eventually be lost, but not necessarily before particular alleles have been fixed, and other alleles lost, within the population. For example, extremely strong selection for a particular allele of gene A can lead to the fixation of mildly deleterious alleles in closely linked (neighboring) genes.

At this point, let us clarify some terms related to genes. These terms arise from the history of biology in general, and genetics in particular. We now know that genetic information is stored in the sequence of double-stranded DNA molecules. A gene is the region of a DNA molecule that encodes a particular "gene product", either an RNA molecule or a polypeptide, together with regions of the DNA molecule required for the gene product to be "expressed", a term that captures the ability of the gene product to be made and used (that is, to impact the cell/organism within which the gene is located). Where and when a gene is expressed is regulated by networks of interacting molecules. All of the DNA molecules present in a cell are known collectively as the cell's genome. We refer to the position of a particular gene within the genome as a genetic locus (or the plural, loci). In Latin locus means 'place'; think location – a word derived from the same root. A particular genetic locus (gene) can be occupied by any of a number of distinct alleles (DNA sequences). There are various mechanisms that can duplicate, delete, insert, or move a region of DNA within the genome, creating (or eliminating) new genetic loci. The phenotype associated with an allele is influenced by position within a genetic locus, as well as the details of the rest of the genome.

It is worth noting that the combination of non-adaptive, non-selective processes can lead to the appearance and maintenance of mildly non-advantageous (deleterious) traits within a population. Similarly, a trait that increases reproductive success, by increasing the number of surviving offspring, may be associated with other not-so-beneficial, and sometime seriously detrimental (to individuals) effects. The key is to remember that evolutionary mechanisms do not necessarily result in what is best for an individual organism but what in the end enhances net reproductive success. Evolutionary processes do not select for particular genes or new versions of genes but rather for those combinations of alleles that optimize reproductive success.[140]

Of course, the situation gets more complicated when evolutionary mechanisms generate organisms, like humans, who think and feel and can actively object to the outcomes of evolutionary processes. From the point of view of self-conscious organisms, evolution can appear cruel, or at the very least totally apathetic to the desires and happiness of individuals. This was one reason that Darwin preferred impersonal (naturalistic) mechanisms over the idea of a God responsible for what can appear to be the gratuitously cruel aspects of their creation.

---

[140] There is complexity that we will get to associated with the fact that genes are molecules, and molecules can move between cells (a process known as horizontal (or lateral) gene transfer. Viruses also contain genes, which they can transfer from organism to organism. We will consider these topics later on.

Questions to answer:

39. How does the linkage of genes along a chromosome influence evolutionary processes?
40. What, exactly, is the difference between a gene and an allele? a gene and a chromosome? How many DNA molecules in a chromosome?
41. Consider this quote from Charles Darwin, "Natural selection will never produce in a being any structure more injurious than beneficial to that being, for natural selection acts solely by and for the good of each." How would you modify it in light of our modern understanding of evolutionary mechanisms?

Question to ponder:
- How does evolution's focus on reproductive success, and cost-benefit analysis, rather than individual well-being impact the view that the natural is inherently good?

Speciation & extinction

As we have noted, an important observation that any useful biological theory needs to explain is why, exactly, there are so many (millions) of different types of organisms currently present on Earth. The Theory of Evolution explains this observation through the process of speciation. The basic

So, naturalists observe, a flea has smaller fleas that on him prey; and these have smaller still to bite 'em; and so proceed ad infinitum. - Jonathan Swift

idea is that populations of organisms can split into distinct groups. Over time evolutionary mechanisms acting on these populations produce distinct types of organisms, that is, different species. At the same time, we know from the fossil record and from modern experiences, that types of organisms can disappear – they can become extinct. What leads to the formation of a new species or the disappearance of existing ones?

To answer these questions, we have to consider how populations behave. A population of a particular type of organism will typically inhabit a particular geographical region. The size of these regions can range from over an entire continent or more, to a small limited region, such as a single isolated lake. Moreover, when we consider organisms that reproduce in a sexual manner, which involves a degree of cooperation between individuals, we have to consider how far a particular organism (or its gametes) can travel. The range of some organisms is quite limited, whereas others can travel significant distances. Another factor to consider is how an organism makes its living - where does it get the food and space it needs to successfully reproduce? Together these are referred to as a specific species' (population's) ecological niche.

An organism's ecological niche, which is the result of its past evolutionary history, the past selection pressures acting within a particular environment, and its current behavior, combines all of these factors. In a stable environment, and a large enough population, reproductive success will reflect how effectively organisms' exploit their ecological niche. Over time, stabilizing selection will tend to optimize the organism's adaptation to its niche. At the same time, it is possible that different types of organisms will compete for similar resources, for a similar niche. This interspecies competition leads to a new form of selective pressure. If individuals of one population can exploit a different set of resources or the same resources (a different niche) differently, these organisms can minimize competition with other species and become more reproductively successful compared to individuals that continue to compete directly with other species. This can lead to a number of outcomes. In one case, one species becomes

much better than others at occupying a particular niche, driving the others to extinction. Alternatively, one species may find a way to occupy a new or related niche, and within that particular niche, it can more effectively compete, so that the two species come to occupy distinct niches. Finally, one of the species may be unable to reproduce successfully in the presence of the other and become (at least) locally extinct. These scenarios are captured by what is known as the competitive exclusion principle or Gause's Law, which states that two species cannot stably occupy the same ecological niche – over time either one will leave (or rather be forced out) of the niche, or will evolve to fill a different, often subtly different niche.[141] What is sometimes hard to appreciate is how specific a viable ecological niche can be. For example, consider the situations described by the evolutionary biologist Theodosius Dobzhansky (1900-1975): "*Some organisms are amazingly specialized. Perhaps the narrowest ecologic niche of all is that of a species of the fungus family Laboulbeniaceae, which grows exclusively on the rear portion of the elytra (the wing cover) of the beetle Aphenops cronei, which is found only in some limestone caves in southern France. Larvae of the fly Psilopa petrolei develop in seepages of crude oil in California oilfields; as far as is known they occur nowhere else.*"

While it is tempting to think of ecological niches in broad terms, the fact is that subtle environmental differences can favor specific traits and specific organisms. If an organism's range is large enough and each individual's range is limited, distinct traits can be prominent in different regions of the species' range. These different subpopulations, termed subspecies or races, reflect local adaptations. For example, it is thought that human populations migrating out of the equatorial regions of Africa were subject to differential selection based on exposure to sunlight, due in part to the roles of sunlight in the synthesis of vitamin D as well as damage to skin due to exposure (sun burn).[142] In their original ecological niche, the ancestors of humans were thought to hunt in the open savannah (rather than within forests), and so developed adaptations to control their body temperature - human nakedness is thought to be one such adaptation, although there may be aspects of sexual selection involved as well (discussed in the next chapter). Yet, the absence of a thick coat of hair also allowed direct exposure to the UV-light from the sun. While UV exposure is critical for the synthesis of vitamin D, too much exposure can lead to skin cancer. Dark skin pigmentation is thought to be an adaptive compromise. As human populations moved away from the equator, the dangers of UV exposure decreased while the need for vitamin D production remained. Under such conditions, allelic variation that favored lighter skin pigmentation, but retaining the ability to tan, at least to some extent, appears to have been selected (→). Genetic analyses of different populations have begun to reveal exactly which mutations, and the alleles they produced, occurred in different human populations as they migrated out of Africa and across the Earth. Of course, with humans the situation has an added level of complexity. For example, the (relatively recent) trait of wearing clothing directly impacts the pressure of "solar selection."

A number of different phenotypic variations can occur over the geographical range of a species. Differences in climatic conditions, pathogens, predators, and prey can all lead to local adaptations, like

---

[141]Competitive exclusion principle: https://en.wikipedia.org/wiki/Competitive_exclusion_principle

[142]Genetics of skin color: image sources: http://hmg.oxfordjournals.org/content/18/R1/R9.full

those associated with human skin color. For example, many species are not continuously fertile and only mate at specific times of the day or year. When the range of a species is large, organisms in geographically and climatically distinct regions may mate at somewhat different times. As long as there is sufficient migration of organisms between regions and the organisms continue to be able to interbreed and to produce fertile offspring, the population remains a single species.

Mechanisms of speciation

So now we consider the various mechanisms that can lead a species to give rise to one or more new species. Remembering that species, at least species that reproduce sexually, are defined by the fact that they can and do interbreed to produce fertile offspring, you might already be able to propose a few plausible scenarios. An important point is that the process of speciation is continuous, there is generally no magic moment when one species changes into another, rather a new species emerges over time from a pre-existing species, after which the two populations evolve independently.[143] The origin of species through evolutionary mechanisms is therefore formally analogous to the Cell Theory, where each cell is derived from a pre-existing cell – the difference is that the process of cell division result in a unambiguous benchmark in the history of a cell. The situation is more ambiguous in organisms that reproduce asexually, but we will ignore that for the moment. More generally, species are populations of organisms at a moment in time, they are connected to past species and can produce new species in the future (or go extinct).

Perhaps the simplest way that a new species can form is if the original population is physically divided into isolated subpopulations. This is termed allopatric speciation. By isolated, we mean that individuals of the two subpopulations no longer mingle with one another, they are restricted to specific geographical areas. That also means that they no longer interbreed with one another. If we assume that the environments inhabited by the subpopulations are distinct and that they represent distinct sets of occupied and available ecological niches, distinct climate and geographical features, and distinct predators, prey, and pathogens, then these isolated subpopulations will be subject to different selection pressures, different phenotypes, and the genotypes associated with them, will differ in their reproductive success in a particular environment. Assuming the physical separation between the populations is stable, and persists over a sufficient period of time, the populations will diverge. Both selective and non-selective processes drive this divergence, which will be influenced by what new mutations arise and give rise to alleles. The end result will be populations adapted to specific ecological niches, which may well be different from the niche of the parental population. For example, it is possible that while the parental population was a generalist, occupying a broad range of ecological niches, the subpopulations may be specialized to a specific niche. Consider the situation with various finches (honeycreepers) found in the Hawai'ian islands.[144] Derived from an ancestral founder population, these organisms have adapted to a number of highly specialized niches. These specializations give them a competitive edge with respect to one another in feeding off particular types of flowers. As they

_____

[143] An interesting exception occurs in some plants (which can self-fertilize), where there are instances new species formed in one generation due to changes in ploidy: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2442920/

[144] Hawaiian honeycreepers and their tangled evolutionary tree

specialize, however, they become more dependent upon the continued existence of their host flower or flower type (→). It is a little like a fungus that can only grow on one particular place on a particular type of beetle, as we discussed earlier. We begin to understand why the drive to occupy a particular ecological niche also leads to vulnerability, if the niche disappears for some reason, the species adapted to it may not be able to cope and effectively and competitively exploit the remaining niches, leading to its extinction.[145]
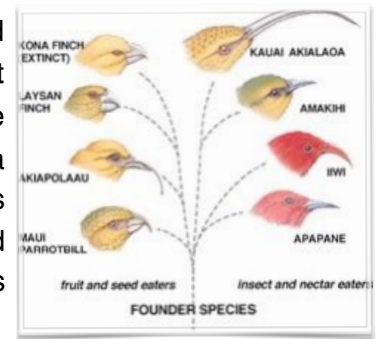
It is a sobering thought that current estimates are that greater that ~98% of all species that have or now live on Earth are extinct, presumably due in large measure in changes in, or the disappearance of, their niche. You might speculate (and provide a plausible argument to support your speculation) as to which of the honeycreepers illustrated above would be most likely to become extinct in response to environmental changes.[146] In a complementary way, the migration of organisms into a new environment can produce a range of effects as the competition for existing ecological niches get resolved.[147] If an organism influences its environment, the effects can be complex. As noted earlier, a profound and global example is provided by the appearance, early in the history of life on Earth, of photosynthetic organisms that released molecular oxygen (O_2) into the atmosphere as a waste product. Because of its chemical reactivity, the accumulation of molecular oxygen led to loss of some ecological niches and the creation of new ones. While dramatic, similar events occur on more modest levels all of the time. It turns out that extinction is a fact of life – at the same time, life has continued and diversified in an uninterrupted manner for over 3,500,000,000 years.

Gradual or sudden environmental changes, ranging from the activity of the sun, to the drift of continents and the impacts of meteors and comets, lead to the disappearance of existing ecological niches and appearance of new ones. For example, the collision of continents with one another leads to the formation of mountain ranges and regions of intense volcanic activity, both of which can influence climate and the connectedness of populations. There have been periods when Earth appears to have been completely or almost completely frozen over.[148] These geological processes continue to be active today, with the Atlantic ocean growing wider and the Pacific ocean shrinking, the splitting of Africa along the Great Rift Valley, and the continuing collision of India with the rest of Asia. As continents move and sea levels change, organisms that evolved on one continent may be able to migrate into another. All of these processes combine to lead to extinctions, which open ecological niches for new organisms, and so it goes.

At this point you should be able to appreciate the fact that evolution never actually stops. Aside from various environmental factors, each species is part of the environment of other species. Changes in one species can have dramatic impacts on others as the selective landscape changes. An obvious

---

[145] Here is a great video of organisms that have survived (often with human help) the extinction their partners: The Ghosts of Evolution: Nonsensical fruit, missing partners, and other ecological anachronisms

[146] The Perils of Picky Eating: Dietary Breadth Is Related to Extinction Risk in Insectivorous Bats

[147] Humans spread through South America like an invasive species

[148] One such "snowball Earth" period has been suggested to play an important role in the emergence of macroscopic multicellular life. see http://www.bbc.com/earth/story/20150112-did-snowball-earth-make-animals

example is the interrelationship between predators, pathogens, and prey. Which organisms survive to reproduce will be determined in large part by their ability to avoid predators or recover from infection. Certain traits may make the prey more or less likely to avoid, elude, repulse, discourage, or escape a predator's attack. As the prey population evolves in response to a specific predator or pathogen, these changes will impact the predator or pathogen, which will also have to adapt. This situation is often call the Red Queen hypothesis, and it has been invoked as a major driver for the evolution of sexual reproduction, which we will consider in greater detail as we go on.[149]

> *As the Red Queen said to Alice ... "Here, you see, it takes all the running you can do to keep in the same place"*
> *-Lewis Carroll, Though the Looking Glass*

Isolating mechanisms

Think about a population that is on its way to becoming specialized to fill a particular ecological niche. What is the effect of cross breeding with a population that is, perhaps, on an adaptive path to another adapting to another ecological niche? Most likely the offspring will be poorly adapted to either niche. This leads to a new selective pressure, selection against cross-breeding between individuals of the two populations. Even small changes in a particular trait or behavior can lead to significant changes in mating preferences and outcomes. Consider Darwin's finches or Hawai'ian honeycreepers. A major feature that distinguishes these various types of birds is the size and shapes of their beaks. These adaptations represent both the development of a behavior – that is the preference of birds to seek food from particular sources, for example, particular types of flowers or particular size seeds – and the traits needed to successfully harvest that food source, such as bill shape and size. Clearly the organism has to display the behavior, even if it is in a primitive form, that makes selection of the physical trait beneficial. This is a type of loop, where behavioral and physical traits are closely linked. You can ask yourself, could a long neck have evolved in a species that did not eat the leaves of trees?

Back to finches and honeycreepers. Mate selection in birds is often mediated by song, generally males sing and females respond (or not). As beak size and shape changes, the song produced also changes.[150] This change is, at least originally, an unselected trait that accompanies the change in beak shape, but it can become a selected trait if females recognize and respond to songs more like their own. This would lead to preferential mating between organisms with the same trait (beak shape). Over time, this preference could evolve into a stronger and stronger preference, until it becomes a reproductive barrier between organisms adapted to different ecological niches.[151] Similarly, imagine that the flowers a particular subpopulation feeds on open and close at different times of the day. This could influence when an organism that feeds on a particular type of flower is sexually receptive. You can probably generate your own scenarios in which one behavioral trait has an influence on reproductive preferences and success. If a population is isolated from others, such effects may develop but are irrelevant; they become important when two closely related but phenotypically distinct populations come back into contact. Now matings between individuals in two different populations, sometimes termed

---

[149] Running with the Red Queen: the role of biotic conflicts in evolution

[150] A good background article on Darin's finches and speciation is here: Sisyphean evolution

[151] Beaks, Adaptation, and Vocal Evolution in Darwin's Finches & Vocal mechanics in Darwin's finches: correlation of beak gape and song frequency

hybridization, can lead to offspring poorly adapted to either niche. This can create a selective pressure to minimize hybridization. Again, the reproductive isolation of two populations can arise spontaneously, such as when two populations mate at different times of the day or the year or respond to different behavioral queues, such as mating songs. Traits that enhance reproductive success by reducing the chance of detrimental hybridization will be preferentially chosen. The end result is what is known as reproductive isolation.[152] As reproductive isolation occurs, what was one species becomes two. A number of different mechanisms ranging from the behavioral to the structural and the molecular are involved in generating reproductive isolation. Behaviors may not be "attractive," genitalia may not fit together,[153] gametes might not fuse with one another, or embryos might not be viable - there are many possibilities.

Ring species

Ring species demonstrate a version of allopatric speciation. Imagine populations of the species A. Over the geographic range of A there exist a number of subpopulations. These subpopulations (A_1 to A_5) and (A_a to A_e) have limited regions of overlap with one another but where they overlap they interbreed successfully ($\rightarrow$). But populations A_5 and A_e no longer interbreed successfully – are these populations separate species? In this case, there is no unambiguous answer (and sometimes we have to get used to the idea of ambiguity, something that should be more widely appreciated). That said, it is likely that the link between the various populations will be broken and one or more species may arise in the future. Consider the black bear *Ursus americanus*. Originally distributed across all of North America, its distribution is now much more fragmented. Isolated populations are free to adapt to their own particular environments and migration between populations is limited. Clearly the environment in Florida is different from that in Mexico, Alaska, or Newfoundland. Different environments will favor different adaptations. If, over time, these populations were to come back into contact with one another, they might or might not be able to interbreed successfully - reproductive isolation may occur and one species may become many.
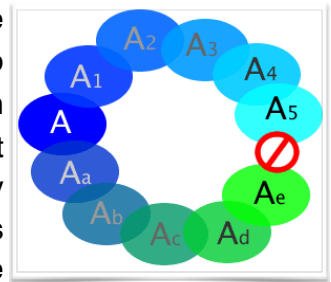
Sympatric speciation

While the logic and mechanisms of allopatric speciation are relatively easy to grasp (we hope), there is a second type of speciation, known as sympatric speciation, which was originally more controversial. It occurs when a single population of organisms splits into two reproductively isolated communities within the same physical region. How could this possibly occur? What stops (or inhibits) the distinct sub-populations from inbreeding and reversing the effects of selection and nascent speciation? Recently a number of plausible mechanisms have been identified. One involves host selection.[154] In host selection, animals (such as insects) that feed off a specific host may find

---

[152] Beak size matters for finches' song: http://news.nationalgeographic.com/news/2004/08/0827_040827_darwins_finch.html

[153] Causes and Consequences of Genital Evolution: http://icb.oxfordjournals.org/content/early/2016/09/13/icb.icw101.abstract

[154] Sympatric speciation by sexual selection & Sympatric speciation in phytophagous insects: moving beyond controversy?

themselves reproducing in distinct zones associated with their hosts. For example, organisms that prefer blueberries will mate in a different place, time of day, or time of year than those that prefer raspberries. There are blueberry- and raspberry-specific niches. Through a process of disruptive selection (see above), organisms that live primarily on a particular plant (or part of a plant) can be subject to different selective pressures, and reproductive isolation will enable the populations to more rapidly adapt. Mutations that reinforce an initial, perhaps weak, mating preference can lead to reproductive isolation - this is a simple form of sexual selection, which we will discuss soon.[155] One population has become two distinct, reproductively independent populations, one species has become two.

Questions to answer:
42. What is involved in establishing reproductive isolation between populations (species formation); what factors favor speciation?
43. How are sympatric and allopatric speciation the same and how do they differ?
44. Describe the (Darwinian) cycle of selection associated with the development of a trait, such as the extended neck of giraffes. Consider the feedback between behavior and anatomy.

Questions to ponder:
- How would you determine whether two species are part of the same genus?
- How might asexual organism be assigned to specific species?
- How might you decide whether an organism, identified through fossil evidence, was part of a new or a living species?

Signs of evolution: homology and convergence

When we compare two different types of organisms we often find traits that are similar. On the basis of evolutionary theory, these traits can arise through either of two processes: the trait could have been present in the ancestral population that gave rise to the two species or the two species could have developed their versions of the trait independently. In this latter case, the trait was not present in the last common ancestor shared by the organism. Where a trait was present in the ancestral species it is said to be a homologous trait. If the trait was not present in the ancestral species but appeared independently within the two lineages, it is known as an analogous trait that arose through evolutionary convergence.

For example, consider the trait of vitamin C dependence, found in Haplorrhini primates and discussed above. Based on a number of lines of evidence, we conclude that the ancestor of all Haplorrhini primates was vitamin C dependent and that vitamin C dependence in Haplorrhini primates is a homologous trait. On the other hand Guinea pigs (*Cavia porcellus*), which are in the order Rodentia, are also vitamin C dependent, but other rodents are not. It is estimated that the common ancestor of primates and rodents lived more than ~80 million years ago, that is, well before the common ancestor of the Halporrhini. Given that most rodentia are vitamin C independent, we can assume that the common ancestor of the rodent/primate lineages was itself vitamin C independent. We conclude that vitamin C dependence in Guinea pigs and Halporrhini are analogous traits, they arose as the result of independent events.

---

155 The sexual selection: http://www.youtube.com/watch?v=JakdRczkmNo

As we look at traits, we have to look carefully, structurally, and more and more frequently, molecularly, that is, directly at the genotype, to determine whether they are homologous or analogous - the result of evolutionary convergence or ancestry. Consider the flying vertebrates. The physics of flight, and many other behaviors that organisms perform, are constant. Organisms of similar size face the same aerodynamic and thermodynamic constraints. In general there are only a limited number of physically workable solutions to deal with these constraints. Under these conditions different populations that are in a position to exploit the benefits of flight will, through the process of variation and selection, end up with structurally similar solutions. This process is known as convergent evolution. Convergent evolution occurs when only certain solutions to a particular problem are evolutionarily accessible.
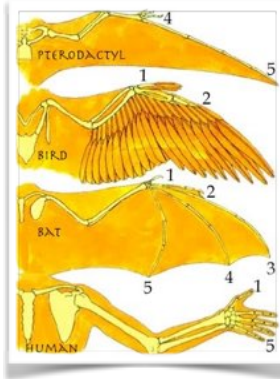
Consider the wing of a pterodactyl, which is an extinct flying reptile, a bird, and a bat, a flying mammal (←). These organisms are all tetrapod (four legged) vertebrates – their common ancestor had a structurally similar forelimb, so their forelimbs are clearly homologous. Therefore this evolutionary adaptation, using the forelimb for flight, began from a structurally similar starting point. But most tetrapod vertebrates do not fly, and forelimbs have become adapted to many different functions. An analysis of tetrapod vertebrate wings indicates that each took a distinctly different approach to generating wings. In the pterodactyl, the wing membrane is supported by the 5th finger of the forelimb, in the bird by the 2nd finger, and in the bat, by the 3rd, 4th and 5th fingers. The wings of pterodactyls, birds, and bats are clearly analogous structures, while their forelimbs are homologous.

As another example of evolutionary convergence consider teeth. The use of a dagger is an effective solution to the problem of killing another organism. Variations of this solution have been discovered or invented independently many times; morphologically similar dagger-like teeth have evolved independently, that is, from ancestors without such teeth, in a wide range of distinct lineages. Consider, for example, the placental mammal *Smilodon* and the marsupial mammal *Thyacosmilus* (→); both have similarly-shaped highly elongated canine teeth. Marsupial and placental mammals diverged from a common ancestor ~160 million years ago and this common ancestor, like most mammals, appears to have lacked such dagger-like teeth. While teeth are a homologous feature of *Smilodon* and *Thyacosmilus*, elongated dagger-like teeth are analogous structures that resulted from their convergent evolution.
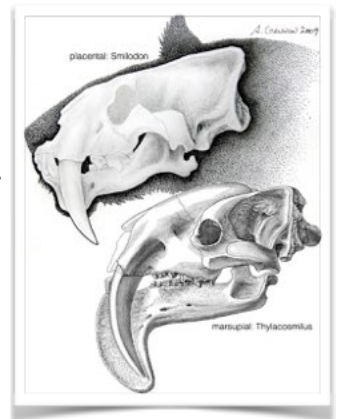
Losing traits

A major challenge when trying to determine a plausible relationship between organisms based on anatomy has been to distinguish homologous from convergent (analogous) traits. Homologous traits, known as synapomorphies, are the basis of placing organisms together within a common group. In contrast, convergent traits are independent solutions to a similar problem, and so are irrelevant when it comes to defining evolutionary relationships. It is, however, also true that evolution can lead to the loss of traits; this can confuse or complicate the positioning of an organism in a classification scheme. It is

worth noting that very often developing a particular trait, whether it is an enzyme or an eye, requires energy. If the trait does not contribute to an organism's reproductive success it will not be selected for; on the other hand, if it is expensive to build, but has no useful function, its loss may be selected for. As organisms adapt to a specific environment and lifestyle, traits once useful can become irrelevant or distracting, and may be lost. A classic example is the reduction of hind limbs during the evolution of whales [↓]. Another is the common loss of eyes often seen as populations adapt to environments in

which light is absent. The most dramatic cases of loss involves organisms that become obligate parasites of other organisms. In many cases, these parasitic organisms are completely dependent on their hosts for many essential functions, this allows them to become quite simplified even though they are in fact highly evolved. For example, they lose many genes as they become dependent upon the host. The loss of traits can itself be an adaptation if it provides an advantage to organisms living in a particular environment. This fact can make it difficult to determine whether an organism is primitive (that is, retains ancestral features) or highly evolved.

Questions to answer:
45. How would you decide whether vitamin C dependences in Haplorrhini and guinea pigs were independent events?
46. How would you decide whether a trait is primitive (ancestral) or specialized (derived)?

Signs of evolutionary history

Evolution is an ongoing experiment in which random mutations are selected based on the effects of the their resulting phenotypes on reproductive success. As we have discussed, various non-adaptive processes are also involved, which can impact evolutionary trajectories. The end result is that adaptations are based on past selective pressures and i) are rarely perfect and ii) may actually have become outdated, if the environment the organisms live in has changed. One might want to keep this in mind when one considers the differences associated with living in small groups in a pre-technological world on the African savannah and living in New York City. In any case, evolution is not a designed process that reflects a predetermined goal but involves responses to current constraints and opportunities - it is a type of tinkering in which selective and non-selective processes interact with pre-existing organismic behaviors and structures and is constrained by cost and benefits associated with various traits and their effects on reproductive success.[156] What evolution can produce depends on the alleles present in the population, or which can be generated by mutation, and the current form of the organism. Not all desirable phenotypes (that is, leading to improved reproductive success) may be accessible from a particular genotype, and even if they are, the cost of attaining a particular adaptation, no matter how desirable to an individual, may not be repaid by the reproductive advantage it provides within a population.

As an example, our ability to choke on food could be considered a serious design flaw, but it is the result of the evolutionary path that produced us (and other four-legged creatures), a path that led to the crossing of our upper airway (leading to the lungs) and our pharynx (leading to our gastrointestinal system). That is why food can lodge in the airway, causing choking or death. It is possible that the costs

---

[156] Evolutionary tinkering: http://virtuallaboratory.colorado.edu/Biofundamentals/lectureNotes/Readings/EvolutionTinkering.pdf

of a particular "imperfect" evolutionary design are offset by other advantages (↓). For example, the small but significant possibility of death by choking may, in an evolutionary sense, be worth the ability to make more complex sounds (speech) involved in social communication.[157]

Biro. (2011) The evolutionary reason for the need to secure the airway in anesthesiology. IMAJ 13: 5-8

In fish, the swim bladder (red) is connected to the gastrointestinal track (black) through a "T" junction. In different types of fish the swim bladder can be either dorsal (A) or ventral (B) to the esophagus.

In reptiles, birds, and mammals (including humans) the airway crosses the gastrointestinal track in an "X" junction.

In lungfish and amphibians, the proto-lungs/lungs are connected to the gastrointestinal track through a ventral "T" junction.

As a general rule, evolutionary processes generate structures and behaviors that are as good as they need to be for an organism to effectively exploit a specific set of environmental resources and behaviors, and to compete effectively with its neighbors, that is, to successfully occupy its niche. If being better than good enough does not enhance reproductive success, it will not be selected for, and variations in that direction will be lost, particularly if they come at the expense of other important processes or abilities.

In this context it is worth noting that we are always dealing with an organism throughout its life cycle. Different traits can have different reproductive values at different developmental stages. Being cute can have important survival benefits for a baby but be less useful in a corporate board room, although perhaps that is debatable. A trait that improves survival during early embryonic development or enhances reproductive success as a young adult can be selected for even, if it produces negative effects on older, post-reproductive individuals. Moreover, since the probability of being dead by accident or disease, and so no longer reproductively active, increases with age, selection for traits that benefit the old will inevitably be weaker than selection for traits that benefit the young, although this trend can be modified in organisms in which the presence of the old, for example, grandparents, positively influences the survival and reproductive success of the young, for example through teaching and babysitting. Of course survival and fertility curves can change in response to changing environmental factors, which alter selective pressures. In fact, lifespan itself is a selected trait, since it is the population not the individual that evolves.[158] In this light, while most large mammals have long lifespans, a number of large and complex invertebrates, such as squid, octopus, and cuttlefish have short lifespans.[159]

We see the evidence for various compromises involved in evolutionary processes all around us.[160] They explain the limitations of our senses, as well as our tendency to get backaches, need hip-

---
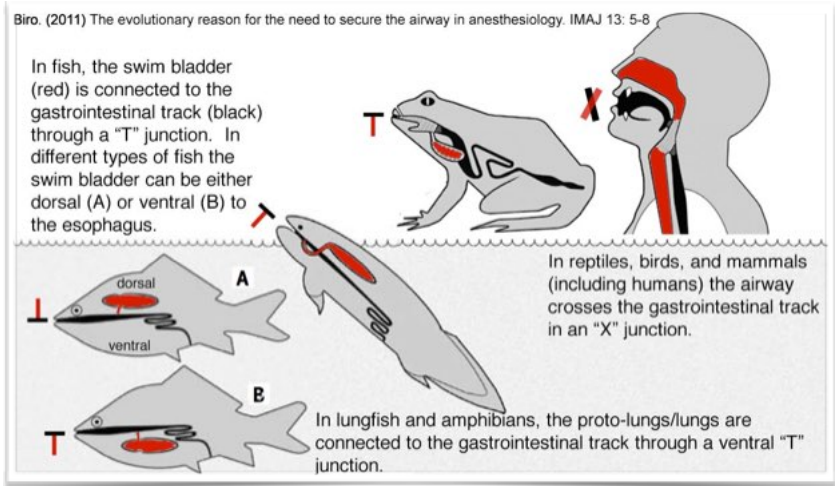
[157] How the Hyoid Bone Changed History: http://www.livescience.com/7468-hyoid-bone-changed-history.html

[158] Methusaleh's Zoo: how nature provides us with clues for extending human health span & Why Men Matter: Mating Patterns Drive Evolution of Human Lifespan

[159] As described in Peter Godfrey-Smith's Other Minds: The Octopus, the Sea, and the Deep Origins of Consciousness

[160] Wikipedia: Evidence of common descent

replacements,[161] and our susceptibility to diseases and aging.[162] For example, the design of our eyes leaves a blind spot in the retina. Complex eyes have arisen a number of times during the history of life, apparently independently, and not all have such a blind spot - a blind spot is not a necessary feature of a complex eye. We have adapted to this retinal blind spot through the use of saccadic eye movements because this is an evolutionarily easier fix to the problem than rebuilding the eye from scratch, which is likely to be impossible (evolutionarily). An "intelligently designed" human eye would presumably not have such an obvious design flaw, but given the evolutionary path that led to the vertebrate eye, it may simply have been impossible to "back up" and fix this flaw. More to the point, since the vertebrate eye works well, there is no apparent reward in terms in reproductive success associated with removing the blind spot. This is a general rule: current organisms work, at least in the environment that shaped their evolution. Over time, organisms that diverge from the current optimal, however imperfect, solution will be at a selective disadvantage. The current vertebrate eye is maintained by stabilizing selection. The eyes of different vertebrates differ in their acuity, basically how fine a pattern of objects they can resolve at what distance, and sensitivity, what levels and wavelengths of light they can perceive. Each species has eyes, and their connections to the brain, adapted for their specific ecological niche. For example, an eagle sees details at a distance four to five times are far as the typical human; why? because such visual acuity is useful in terms of the eagle's life-style, whereas such visual details would likely be a non-useful distraction for humans.[163]

Homologies provide evidence for a common ancestor

The more details two structures share, the more likely they are to be homologous. In the 21st century molecular methods, particularly complete genome (DNA) sequencing, have made it possible to treat gene sequences and genomic organization as traits that can be compared quantitatively. Detailed analyses of many different types of organisms reveals the presence of a common molecular signature that strongly suggests that all living organisms share a large numbers of homologies, which implies that they are closely related - that they share a common ancestor. These universal homologies range from the basic structure of cells to the molecular machinery involved in energy capture and transduction, information storage and utilization. All organisms

- use double-stranded DNA as their genetic material;
- use the same molecular systems to access the information stored in DNA;
- express that information initially in the form of RNA molecules;
- use a common genetic code, with a few variations, and messenger RNA to specify the sequence of polypeptides (proteins);
- use ribosomes to translate the information stored in messenger RNAs into polypeptides; and
- share common enzymatic (metabolic) pathways and structures (lipid-based boundary membranes).

---

[161] Hip pain may be 'hangover from evolution': http://www.bbc.com/news/health-38251031

[162] How Bipedalism Arose

[163] What If Humans Had Eagle Vision?

Anti-evolution arguments

The theory of evolution has been controversial since its inception largely because it deals with issues of human origins and behavior, our place in the Universe, and life and its meaning. Its implications can be quite disconcerting, but many observations support the fact that organisms on Earth are the product of evolutionary processes and these processes are consistent with what we know about how matter and energy behave. As we characterize the genomes of diverse organisms, we see evidence for the interrelationships, observations that non-scientific (creationist) models would never have predicted and do not explain. That evolutionary mechanisms have generated the diversity of life and that all organisms found on Earth share a common ancestor is as well-established as the atomic structure of matter, the movement of Earth around the Sun, and the solar system around the Milky Way galaxy. The implications of evolutionary processes remain controversial, but not evolution itself. We would argue that religions that deny the evolutionary relationships between organisms, and the role of evolutionary mechanisms in shaping organisms, including humans, run the risk of making themselves look ridiculous, at least in terms of data-based (scientific) discussions.[164]

> *Scientific knowledge is a body of knowledge of varying degrees of certainty-some most unsure, some nearly sure, but none absolutely certain … Now we scientists are used to this, and we take it for granted that it is perfectly consistent to be unsure, that it is possible to live and not know. - Richard Feynman.*
>
> *...it is always advisable to perceive clearly our ignorance.*
> *– Charles Darwin.*

Questions to answer:
47. Describe a scenario in which the loss of a trait or a gene is beneficial?
48. Explain why the loss of a trait or convergent evolution complicates lineage analysis?
49. Describe a scenario in which the simplification of a complex organism would be selected for?
50. Construct a diagram that shows the difference between homologous and analogous traits, and use it to explain the difference.

Questions to ponder:
- Describe testable predictions that emerge from "intelligent design creationism"?
- In what ways might organisms direct (or influence) their own evolution?
- If the environment were constant, would extinction or evolution occur?
- Should modern genetic engineering methods be used to fix evolutionary design flaws?

---

[164] Go ahead and "teach the controversy:" it is the best way to defend science.

Chapter 4: Social evolution and sexual selection

In which we consider how unicellular organisms evolved to cooperate with one another and how cooperation led to the evolution of multicellular organisms composed of distinct cell types. Similar evolutionary mechanisms have produced a range of cooperative (social) behaviors as well as opportunities for social cheating, and the need for organisms and societies to defend themselves against cheaters. One particularly important social behavior is sexual reproduction and we consider its effects on organisms and their evolution.

The naturalist Ernst Mayr (1904-2005) stressed the difference in thinking in biology compared to physics and chemistry. The history of an electron, an atom, or a molecule is totally irrelevant to its physical and chemical properties. Each carbon isotope atom, for example, is identical to all others - one could be replaced by another and you could never, in theory or in practice, be able to tell the difference. In contrast, each organism, how it is built, how it behaves, how it interacts with other organisms, and the future evolution of its descendants is the result of a continuous evolutionary process involving both adaptive (selective) and non-selective and non-adaptive processes stretching back ~3.5 billion years. This history encompasses an unimaginable number of individually unpredictable events (mutations, accidents, environmental disasters, isolated and merging populations). Because of its molecular and cellular complexity and distinct history, each organism is unique and distinguishable from all others.[165]

In biology, we normally talk about organisms, but this may be too simplistic. When does an organism begin? What are its boundaries? The answers can seem obvious, but then again, perhaps not. When a single-celled organism reproduces it goes through some form of cell division, and when division is complete, one of the two organisms present is considered a new organism and the other the old (preexisting) one, but often it is not clear which is which. In fact, both are old, both reflect a continuous history stretching back to the origin of life. When an organism reproduces sexually, the new organism arises from the fusion of two pre-existing cells and it itself produces cells that fuse to form the next generation. But if we trace the steps backward from any modern organism, we find no clear line between the different types (that is, species) of organisms. When did humans (*Homo sapiens*) appear from pre-humans, or modern birds from their dinosaurian progenitors? The answer is necessarily arbitrary, since cellular and organismic continuity is never interrupted - life does not start, stop, and start again, it continues until it stops in death. Because of superfecundity, selection, and speciation, it also generates branches.

In a similar manner, we typically define the boundaries of an organism in physical terms, but organisms interact with one another, often in remarkably close and complex ways. A dramatic example of this behavior are the eusocial organisms. While many of us are familiar with the social structure of ants and bees, fewer (we suspect) are aware of naked (*Heterocephalus glaber*) and Damaraland (*Cryptomys damarensis*) mole rats. In these organisms reproduction occurs at the group level; only

---

[165] While these events obey physical and chemical laws, in practice, the number of variables involved makes them unpredictable. At the same time, because they are based on natural processes, when we consider large numbers of such events, they become predictable. So while the mutation rate is predictable, which mutations occur in which organism is not.

select females, termed queens, because they are large, produce offspring. Most members of the group are effectively sterile female workers, along with a few males that inseminate the queen.[166] So what, exactly, is the organism, the social group or the individuals that make it up? From an evolutionary perspective, selection is occurring at a social level as well as the organismic level. Similarly, consider yourself and other multicellular organisms (animals and plants). Most of the cells in your body, known as somatic cells, do not directly contribute to the next generation, rather they cooperate to insure that a subset of cells, known as germ line cells, have a chance to form a new organism. In a real sense, the somatic cells are sacrificing themselves so that the germ line cells can produce a new organism. They are the sterile workers to the germ line's queen. Clearly the term "sacrifice" in the context of the somatic cells of a multicellular organism seems weird, since both germ line and somatic cells are necessary parts of a single organism. We might argue that it is the organism, rather than the cells that compose it, that is the biologically meaningful object. Similarly, in a eusocial organism, it is the social group that matters.

We find examples of social behavior at the level of unicellular organisms as well. For example, think about a unicellular organism that divides but in which the offspring of that division stick together. As this process continues, we get what we might term a colony. Is such a clump of cells one or many organisms? If all of the cells within the group can produce new cells, and so new colonies, we consider it a colony of organisms. So where does a colony of organisms turn into a colonial organism? The distinction is certainly not unambiguous, but we can adopt a set of guidelines or rules of thumb.[167] One criterion would be that a colony becomes an organism when it displays traits that are more than just sticking together or failure to separate, that is, when it acts more like an individual or a coordinated group. This involves the differentiation of cells, one from the other, so that certain cells within the group become specialized to carry out specific roles. Producing the next generation of organisms is one such specialized cellular role. Other cells may become specialized for feeding or defense, they act to support the process of reproduction, in part by enabling the resulting organism to occupy a particular ecological niche. The differentiation of cells from one another within a multicellular aggregate has moved a colony of organisms to a multicellular organism. What is tricky about this process is that originally reproductively competent cells have given up their ability to reproduce, and are now acting, in essence, to defend or support the cells that do reproduce. This is a social event and is similar (analogous) to the behavior of naked mole rats. Given that natural selection acts on reproductive success, one might expect that the evolution of this type of cellular and organismic behavior would be selected against or simply impossible to produce, yet multicellularity and social interactions have arisen independently many times during the history of life on earth.[168] Is this a violation of evolutionary theory or do we have to get a little more sophisticated in our thinking?

---

[166]An Introduction to Eusociality: http://www.nature.com/scitable/knowledge/library/an-introduction-to-eusociality-15788128

[167]A twelve-step program for evolving multicellularity and a division of labor

[168] The Origins of Multicellularity

Questions to answer:

51. What features (behaviors) are important when defining an organism? Does your definition include both uni- and multi-cellular organisms?
52. How would you characterize humans in terms of sociality?

Selecting social (cooperative) traits

So how does evolution produce multicellularity? To answer this question, we need to approach evolutionary processes more broadly. The first new idea we need to integrate into our theoretical framework is that of inclusive fitness, which is sometimes referred to as kin selection. For the moment, let us think about traits that favor the formation of a multicellular organism - later we will consider traits that have a favorable effect on other, related organisms, whether or not they directly benefit the cell or organism that expresses that trait. Finally, we will consider social situations in which behaviors have become fixed to various extents, and are extended to strangers; humans can, but do not always, display such behaviors. The importance of mutual aid in evolutionary thinking, that is the roles of cooperation, empathy, and altruism in social populations, was a point emphasize by the early evolutionary biologist (and anarchist) (Prince) Peter Kropotkin (1842–1921)(→).
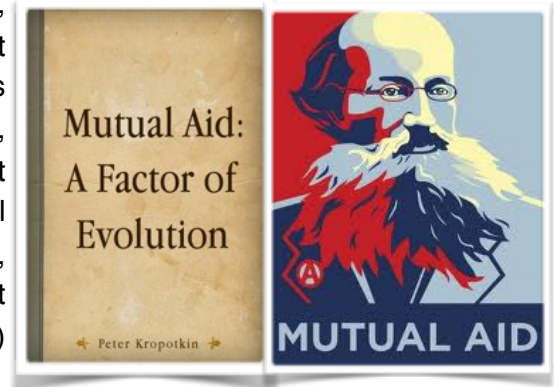
All traits can be considered from a cost-benefit perspective. There are costs (let us call that term "c") in terms of energy needed to produce a trait and risks associated with expressing the trait, and benefits ("b") in terms of the trait's effects on reproductive success. To be evolutionarily preferred, that is, selected, the benefit b must be greater than the cost c, that is $b > c$. Previously we had tacitly assumed that both cost and benefit applied to one and the same organism, but when we consider cooperative (social) behaviors and traits, this is not the case. We can therefore extend our thinking as follows: assume that an organism displays a trait. That trait has a cost to produce and yet may have little or no direct benefit to the organism that produces it; it may even harm it. Now let us assume further that this same trait benefits neighboring organisms, a situation similar to the fireman who risks his life to save an unrelated child in a burning building. How is it possible for a biological system (the fireman), the product of evolutionary processes, to display this type of self-sacrificing behavior? The answer is social systems.

Let us consider an example of this type of behavior, provided by social amoebae of the genus *Dictyostelium*.[169] These organisms have a complex life style that includes a stage in which unicellular amoeba-like organisms crawl around in the soil eating bacteria, growing, and dividing. In this phase of their life cycle, known as the vegetative cycle, the cells divide asexually (as if vegetables don't have sex, but we will come back to that!). If, or rather when, the environment turns hostile, the isolated amoeba sense this change and begin to secrete a small molecule that influences their own and their neighbor's behaviors. They begin to migrate toward one another, forming aggregates of thousands of

---

[169] Molecular phylogeny and evolution of morphology in the social amoebas & A Simple Mechanism for Complex Social Behavior. A nice video here: http://youtu.be/bkVhLJLG7ug

cells (→). Now something rather amazing happens: these aggregates begin to act as coordinated entities, they migrate around as multicellular "slugs" for a number of hours. Within the soil they respond to environmental signals, for example moving toward light, and then settle down and undergo a rather spectacular process of differentiation.[170] All through the cellular aggregation and slug migration stages, part of the social cycle, the original amoeboid cells remain distinct. Upon differentiation ~20% of the cells in the slug differentiate into stalk cells, which can no longer divide, in fact they die. Before they die the stalk cells act together, through changes in their shape, to lift the non-stalk cells above the soil, where they go on to form spores. The stalk cells sacrificed themselves so that they remaining cells can form spores. These spores are specialized cells that can survive harsh conditions; they can float in air and be transported by the wind and other mechanisms into new environments. Once these spore cells land in a new, and hopefully hospitable environment, they convert back into unicellular amoeba that begin to feed and reproduce vegetatively. The available evidence indicates that within the slug the "decision" on whether a cell will form a stalk or a spore cell is stochastic rather than pre-determined. What is important at this point is that the decision is not based on genetic (genotypic) differences between the cells within a slug - two genetically identical cells may both form spores, both stalk cells, or one might become a stalk and one a spore cell.[171] One of every five cells will become a stalk cell, but which cells become stalk cells is (normally) unpredictable, its unpredictability arises from molecular level stochastic processes.

Community behaviors & quorum sensing

Another type of community behavior active at the unicellular level involves what is known as quorum sensing. This is a process by which organisms can sense the density of other organisms in their immediate environment. Each individual secretes specific molecules that they also respond to; the organism's response to this molecule is dependent on the secreted molecule's concentration. The response is non-linear. As the concentration of the signaling molecule increases. There is a discrete concentration, known as the threshold concentration, below which the cells (or organisms) do not change their behavior in response to the secreted compound. When cells or organisms are present at a low density, the concentration of the signaling molecule never exceeds the threshold concentration. As the density of organisms per unit volume increases, however, the concentration of the substance exceeds the threshold concentration and interesting things start to happen; there are changes in behavior, often associated with changes in gene expression (we are getting to what that means).[172] We

---

[170] Behavior of cellular slime molds in the soil: http://www.mycologia.org/content/97/1/178.full

[171] This type of behavior occurs in a number of organisms, including the bacteria: see From cell differentiation to cell collectives: *Bacillus subtilis* uses division of labor to migrate: http://www.ncbi.nlm.nih.gov/pubmed/25894589

[172] Quorum sensing in bacteria: http://www.ncbi.nlm.nih.gov/pubmed/11544353

can think of this type of non-linear response as a strategy to avoid over-reacting to minor fluctuations in the environment. Only when the signal concentration gets high enough (exceeds a certain threshold) does the system respond; the threshold concentration is a function of the concentration of signaling molecules, their binding affinity to the receptor, and other factors that we will consider in greater detail when we consider molecular interactions.

A classic example of a number of cooperative and quorum sensing behaviors is provided by the light emitting marine bacteria *Vibro fischeri*. While there are many steps in the colonization process, and its regulation is complex, here we consider just a few to indicate how cooperative behaviors between the bacteria play a critical role (for more, see the extra chapter at the end of the book). For the colonization of the squid's light organs the *V. fisherei* bacteria must bind to a specific region of the juvenile squid. Given that bacteria are small, you can imagine that very little light would be emitted from a single bacterium. If there were only a small number of bacteria within the light organ, they would be unable to generate a useful level to light, while at the same time, they would be using energy (all costs, no benefit). To increase the numbers (and concentration) of bacteria, they begin to divide and as they divide, they sense the presence of their neighbors and begin to secrete molecules that form of gooey matrix - this leads to the formation of a specialized aggregate of cells, known as a biofilm. Within the biofilm, the bacteria acquire the ability to follow chemical signals produced by the squid's light organ cells. The bacteria swim, through a process known as chemotaxis, that is the ability to move toward (positive) or away from (negative) source of a chemical. In the case of the light organ, they move toward the secreted signal, thereby entering and colonizing the light organs.

The bacteria in the light organs emit light through a reaction system involving the molecule luciferin (→) and coupled chemical reactions involving the input of energy leading to the emission of light (energy) – we will consider in some detail the thermodynamics of such reactions in the next section of the course. The light emitting reaction is catalyzed (that is, sped up) by the protein enzyme luciferase. The luciferase protein is encoded by one of the bacteria's genes (its original role has been proposed to be in the "detoxification of deleterious oxygen derivatives".[173] The light emitting reaction is regulated so that it occurs only when the number of bacteria within a light organ is high enough to make the emission of light useful, which decreases the cost to benefit ratio.
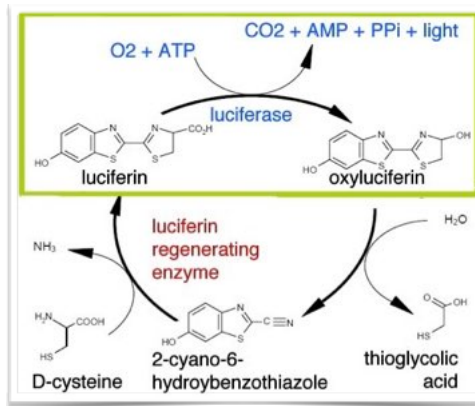
So how do the bacteria know that they are in the presence of sufficiently high concentration of neighbors? Here is where quorum sensing comes into play. A molecule secreted by the bacteria regulates the components of the light reaction. At high concentrations of bacteria, the concentration of the secreted molecule rises above a threshold, and the bacteria respond by turning on their light emitting systems - that is, they express the genes encoding the protein luciferase and the proteins involved in synthesizing luciferin.

---

[173] Experimental evidence for the physiological role of bacterial luciferase: http://www.ncbi.nlm.nih.gov/pubmed/14669913

Mechanistically similar systems are involved in a range of processes including the generation of toxins (virulence factors) and antibiotics directed against other types of organisms. These are produced when the density of the bacterium rises above a threshold concentration. This insures that when a biologically costly molecules are made (such as luciferase and luciferin), they are effective – that is, they are produced at a level high enough to carry out their intended roles. These high levels can only be attained through cooperative behaviors involving many individuals.

Questions to answer:
53. Why does a quorum signal need to be secreted (released) from the organism? What other components are necessary for such cooperative behavior.
54. Is a population of bacteria that display quorum sensing behavior a single organism, justify your answer.
55. Why is a non-linear response to a stimulus important in biological systems? How is it achieved?

Question to ponder:
- How it might it impact the social behavior of slime molds if the percentage of spore cells were 1% rather than 80%?

Active (altruistic) cell death and survivors

One type of behavior you might think would be impossible for evolutionary processes to produce would be the active, intentional or programmed death of a cell or an organism. Yet, such behaviors are surprisingly common in a wide range of systems.[174] The death and release of leaves from deciduous trees in the autumn is an example of a programmed cell death process known generically as apoptosis, from the Greek, meaning to fall off. The programmed cell death process amounts to cellular suicide. It plays important roles in the formation of various structures within multicellular organisms, such as the fingers of your hands, which would develop as paddles without it. Programmed cell death also plays a critical role in the development of the immune and nervous systems, topics well beyond the scope of this book, but extremely important.[175] The process of programmed cell death is distinct from accidental cell death, such as occurs when a splinter impales a cell or you burn your skin. Such accidental death leads to what is known as necrosis. In necrosis, cellular contents are spilled out of the dying cell. The release of cellular debris provokes various organismic defense systems to migrate into the damaged area and (primarily) fight off invading bacteria. The swelling and inflammation associated with injury is an indirect result of necrotic cell death. In contrast, apoptotic cell death occurs by a well-defined pathway and requires energy to carry out. Cell contents are retained during the process, and no inflammatory, immune system response is provoked. In general programmed cell death appears to play specific and important roles within the context of the organism.

Commitment to active cell death is very tightly controlled, although a detailed discussion of the molecular mechanisms involved in apoptosis is beyond our scope. Here we will consider active/ programmed cell death in the context of simpler systems, specifically those formed by unicellular organisms. In unicellular organisms, active cell death is a process triggered by environmental stresses together with quorum sensing. In this situation, a subset of the cells will stochastically "decide" to

---

[174] See On the paradigm of altruistic suicide in the unicellular world: http://www.ncbi.nlm.nih.gov/pubmed/20722725

[175] Apoptosis in the nervous system & Apoptosis in the immune system

undergo active cell death by activating a pathway that leads to the death of the cell. Now when one cell in a densely populated environment dies, its contents are released and can be used by the living cells that remain (→). These living cells gain a benefit, and we would predict that the increase in nutrients will increase their chances of their survival and successful reproduction. This strategy works because as the environment becomes hostile, not all cells die at the same time. From the point of quorum sensing and evolution, it makes no sense if an isolated cell dies through programmed cell death, since the release of nutrients would fail to benefit its (related) neighbors – instead of dying, better to change into what is known as a "persister"; in the persister state the bacterium stops growing and minimizes its use of (and need for) energy (→). In the persister state, the bacterium can survive until the antibiotic in its environment disappears. As we will see later on, these types of individualistic behaviors (programmed cell death or the adoption of a persister phenotype) can occur even in a group of genetically identical cells through the action of stochastic processes.
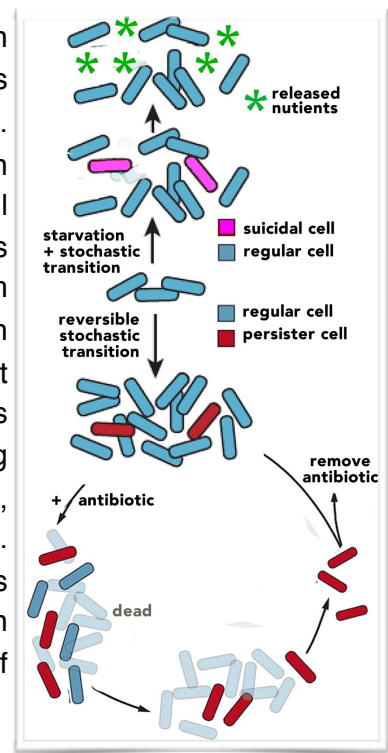
So how do cells kill themselves (on purpose)? Many use a similar strategy. They contain what is known as an addiction module, which consists of two genes - the first encodes a toxic molecule. The toxic molecule, which can kill the cell, is expressed (synthesized) continuously. Many distinct toxin molecules have been identified, so they appear to be analogous rather than homologous. Now you may well wonder how such a gene could exist, how does the cell survive in the presence of a gene that encodes a toxin. The answer is that the cell contains a second gene that encodes an anti-toxin molecule; the anti-toxin typically acts on the toxin and renders it inactive. Within the cell, the toxin-anti-toxin complex forms but does not harm the cell – the toxin molecule's activity is inhibited by its interactions with the the anti-toxin. So far, so good - but you might ask, what is the point - nothing interesting is going on! But the system has one more wrinkle. The toxin and anti-toxin molecules differ in an important way. The toxin molecule is relatively stable - once made it exists for a substantial period of time before it is degraded by other molecular systems within the cell. In terms we have discussed previously, it has a long half-life. In contrast, the anti-toxin molecule is unstable; it is rapidly degraded. The anti-toxin molecule can be maintained at a high enough level to inhibit the toxin only if new anti-toxin molecules are continually synthesized. In a sense the cell has become addicted to the toxin-anti-toxin module.

What happens if the cell is stressed, either by changes in its environment or perhaps infection by a virus? Often cellular activity, including the synthesis of cellular components, such as the anti-toxin, slows or stops. Now can you predict what happens? The level of the stable toxin molecule within the cell remains high, decreasing only slowly, remember, it has a long half-life, while the level of the unstable (short half-life) anti-toxin drops rapidly. As the level of the anti-toxin drops below the threshold level required to keep the toxin inactive, the now active toxin initiates the process of cell death, and the release of the dying cell's components into the environment.

In addition to the dying cell sharing its resources with its neighbors, active cell death can be used as a population-wide defense mechanism against viral infection. One of the key characteristics of viruses

is that they must replicate within a living cell. Once a virus enters a cell, it typically disassembles itself and sets out to reprogram the cell's biosynthetic machinery to generate new copies of the virus. During the period between viral disassembly and the appearance of newly synthesized viruses, the infectious virus disappears - it is said to be latent. If the cell kills itself before new viruses are synthesized, it also kills the infecting virus. By killing the virus (and itself) the infected cell acts to protect its neighbors from viral infection - this can be seen as the kind of altruistic, self-sacrificing behavior we have been considering.[176]

Inclusive fitness, kin and group selection, and social evolution

The question that troubled Darwin (and others) was, how can evolutionary processes produce this type of social, self-sacrificing behavior? Consider, for example, the behavior of bees. Worker bees, who are sterile females, "sacrificed themselves to protect their hives" even though they themselves do not reproduce, they are sterile.[177] Another example, taken from the work of R.A. Fisher (1890–1962), involved the evolution of noxious taste as a defense against predators. Assuming that the organisms eaten by predators do not directly benefit from this trait, after all, they have been eaten, how could the trait of "distastefulness" arise in the first place? If evolution via natural selection is about an individual's differential reproductive success, how are such traits possible? W.D. Hamilton (1936–2000) provided the formal answer, expressed in the equation $(r)(b) > c$, defined by Sewall Wright (1889–1988). As before (in our consideration of costs and benefits), "b" stands for the trait's benefit to the organism and others, "c" stands for the cost of the trait to the individual, while "r" indicates the extent to which two organisms within the population are related to one another, it is a measure of genetic similarity.

Let us think some more about what this means. How might active cell death in bacterial cells be beneficial evolutionarily? In this case, reproduction is asexual and the organism's (cell's) offspring, and likely neighbors, are likely to be closely related – sharing very similar genomes. They are clonally-related to one another in the same way that the cells of a multicellular organism, such as yourself, are derived from a single cell, the fertilized egg which, once formed, divides in an asexual manner. Aside from occasional mutations (changes in DNA), the cells in a clone and within an organism are genotypically identical, that is they have DNA molecules that are identical.[178] Their genotypic similarity arises from the molecular processes by which the genetic material (DNA) replicates and is delivered to the two daughter cells. We can characterize the degree of relationship or genotypic similarity through their r value, the coefficient of relationship. In two genetically identical organisms, r = 1. Two unrelated organisms, with minimum possible genotypic similarity would have an r very close to, but slightly larger than 0 (you should be able to explain why r, why very small, is not equal to 0).[179] Now let us return to

---

[176] The evolution of eusociality: http://www.ncbi.nlm.nih.gov/pubmed/20740005.1

[177] Dugatkin, L.A. 2007. Inclusive Fitness Theory from Darwin to Hamilton. http://www.genetics.org/content/176/3/1375.full

[178] There is an exception to this role involving a subset of the cells of the immune system, but it is not important here.

[179] We will consider the complicating effects of sexual reproduction (which is involved in the formation of the fertilized egg) later on. Suffice it to say, that you are not genetically identical to either of your parents or your own siblings (if you have any, and unless you are have an identical twin). As an approximation, you share ~50% of your genetic material with either of your parents and ~25% with your siblings.

our cost-benefit analysis of a trait's effect on reproductive success. As we introduced before, each trait has a cost of c to the organism that produces it, as well as a potential benefit of b in terms of reproductive success. Selection leads to a trait becoming prevalent (frequent or even fixed) within a population if b >> c. But this equation ignores the effects of a trait on other related and neighboring organisms. In this case, we have to consider the benefits accrued by these organisms as well. Let us call the benefit to the individual as a result of their cooperative/altruistic behavior = b_i and the benefit to others/neighbors = b_o. To generate our social equation, known as Hamilton's rule, we need to consider what is known as the inclusive fitness, namely the benefits provided to others as a function of their relationship to the cooperator. So b > c becomes $b_i + r \times b_o > c$. This leads to the conclusion that a trait can evolve if the cost to the cell or organism that displays it, in terms of metabolic, structural, or behavioral impact on its own reproductive ability, is offset by a sufficiently large increase in the reproductive success of individuals related to it. The tendency of an organism to sacrifice itself for others will increase, that is, be selected for, provided that the reproductive success of closely enough related organisms is increased sufficiently. We will see that we can apply this logic to a wide range of situations and it provides an evolutionary mechanism driving the appearance and preservation of various social behaviors. Given the clonal nature of many types of microbes, inclusive fitness can be particularly powerful in these organisms, although it is also significant in small populations of sexually reproducing organism.

That said, the situation can be rather more complex. Typically, to have a significant impact, inclusive fitness requires a close relationship to the recipient of the beneficial act. So how can we assess this relationship? How does one individual "know" (that is, how is its behavior influenced by the degree of relationship to others) that it is making a sacrifice for its relatives and not just a bunch of (semi-) complete strangers? As social groups get increasingly larger, identifying relatives becomes a more and more difficult task. One approach is to genetically link the social trait, the altruistic behavior, to a physically discernible trait, like smell or a visible structure or behavior. This is sometimes called a "green beard" trait. The likelihood that an organism will behave socially is, one way or the other, linked to the display of a recognizable trait, e.g. a green beard. The presumption is that it is difficult to lose the social cooperation trait without also loosing the green beard trait. The presence of the green beard trait indicates that an organism with the trait will cooperate. Assuming a close linkage between the two traits (social and visible), one can expect social behavior from an individual who displays the trait, even if they are only distantly related. In some cases, a trait may evolve to such a degree that it becomes part of an interconnected set of behaviors.

Once, for example, humans developed a brain sufficiently complex to do what it was originally selected for (assuming that it was brain complexity that was selected, something we might never know for sure), this complexity may have produced various unintended byproducts. Empathy, self-consciousness, and a tendency to neurosis may not be directly selected for but could be side effects of behavioral processes or tendencies that were. As a completely unsupported (but plausible) example, the development of good memory as an aid to hunting might leave us susceptible to nightmares. Assume, for the moment (since we are speculating here), that empathy and imagination are "unintended" products of selective processes. Once present, they themselves can alter future selection pressures and they might not be easy to evolve away from, particularly if they are mechanistically linked to a trait that is highly valued, that is, selected for. The effects of various genetic mutations on

personality and behavior strongly supports the idea that such traits have a basis in one's genotype. That said, this is a topic far beyond the scope of this book.

Group selection

A proposed alternative to inclusive fitness (sometimes known as kin selection) is the concept of group selection. In this type of evolutionary scenario, small groups of organisms of the same species are effectively acting as single (perhaps colonial) organisms. It is the reproductive success of the group, rather than the individuals within the group, compared to other groups of the organism that is the basis of selection. In certain situations, groups that display cooperative and altruistic traits have a selective advantage over groups that do not. Again, the mathematical analysis is similar, and it has been claimed that mathematically group and kin selection are equivalent, even though one occurs between population groups and the other within a population group.[180] The costs of a trait must be offset by the benefits, but now the key factor is membership in a particular group, and typically, members of a group tend to be more closely related to one another. The life cycle of the bacterium *Myxococcus xanthus* provides an example of this type of behavior. When environmental conditions are harsh, the cells aggregate into dense, 100 μm diameter, "fruiting bodies", each containing ~100,000 stress resistant spores. When the environment improves, and nutrients become available, the spores are released en mass and return to active life. They move and feed in a cooperative manner through the release of digestive enzymes, which because they are acting in a quorum mode, can reach high levels.[181] A well-coordinated group is expected to have a significant reproductive advantage over a more anarchic collection of individuals.

While their functional roles are clearly different, analogous types of behavior are seen in flocks of birds, schools (or shoals) of fish, swarms of bees, blooms of algae, and groups of slime mold cells (→).[182] Each of these examples represents a cooperative strategy by which organisms gain a reproductive advantage over those that do not display the cooperative behavior. While the original behavior is likely the result of kin selection, in the wild it is possible that different groups (communities) are in competition with one another, and the group(s) that produces the most offspring, that is, the most reproductively successful group will come to dominate.

---

180 Mathematics of kin- and group-selection: formally equivalent? http://www.ncbi.nlm.nih.gov/pubmed/19929970

181 Evolution of sensory complexity recorded in a myxobacterial genome: http://www.ncbi.nlm.nih.gov/pubmed/17015832

182 How Does Social Behavior Evolve?

Defense against social cheaters

Now an interesting question arises: within a social organization, such as a group of cooperating microbes or hunters,[183] we can expect that, through mutation and other behavioral mechanisms, cheaters will arise. What do we mean by a cheater? Imagine a bacterium within a swarm, a cell in an organism, or an animal in a social group that fails to obey the rules - it may benefit from the social organization without contributing to it. For example when an individual accepts help from others, but fails to help others in need. In the case of slime mold aggregates, imagine that a cell can avoid becoming a non-reproductive stalk cell, instead it always differentiates into a reproductively competent spore. Let us further assume that this trait has a genetic basis. What happens over time? One plausible scenario would be that this spore cell begins its own clone of migratory amoeba, but when conditions change so that aggregation and fruiting body formation occur, most of the cells avoid forming the stalk. We would predict that the resulting stalk would be short or non-existent and so would not be able to lift the spore forming region above the soil, reducing or eliminating the efficiency of dispersion. Different populations would differ based on the percentage of individuals with the cheater phenotype. If dispersion is important for long term reproductive success, there would be selection for populations with low levels of cheaters.

Now the question is, once a social behavior has evolved, under what conditions can evolutionary mechanisms maintain it. One approach is to link the ability to join a social group with various internal and external mechanisms. This makes cooperators recognizable and works to maintain a cooperative or altruistic trait even in the face of individual costs. There are a number of plausible mechanisms associated with specific social traits. This is, however, a topic that can be easily expanded into an entire course. We will focus on common strategies with occasional references to specific situations. To illustrate these mechanisms, we will use human tissues as an example. We can consider the multicellular organism as a social system. The cells that compose it have given up their ability to reproduce a new organism for the ability to enhance the reproductive success of the whole organism. In this context cancer, particularly early on-set and childhood cancers, are diseases that arise from mutations that lead to a loss of social control. Cells whose survival and reproduction is normally strictly controlled lose that control; they become "anti-social" and begin to divide in an uncontrolled and/or inappropriate manner, disrupting the normal organization of the tissue in which they find themselves, and can become malignant, which means that they can breakaway from their original location, migrate, and colonize other areas of the body, a process known as metastasis. The controlled growth of the primary tumor and these metastatic colonies leads eventually to the death of the organism as a whole.

When we think about maintaining a social behavior, we can think of two general mechanisms: intrinsic and extrinsic policing. For example, assume that a trait associated with the social behavior is also linked to or required for cellular survival. In this case, a mutation that leads to the loss of the social trait may lead to cell death (apoptosis). Consider this in the context of cancer. Normal cells can be considered to be addicted to normality. When their normality is disrupted they undergo apoptosis, a type of active cell death (see above). A cell carrying a mutation that enables it to grow in an

---

[183] An interesting read: The stag hunt and the evolution of social structure.

uncontrolled and inappropriate manner will likely kill itself before it can produce significant damage.[184] For a tumor to grow and progress, other mutations must somehow disrupt and inactivate the apoptotic process. The apoptotic process reflects an intrinsic-mode of social control. It is a little like the guilt experienced by (some) people when they break social rules or transgress social norms. The loss of social guilt or embarrassment is analogous to the inhibition of apoptosis in response to various cues associated with abnormal behavior.[185]

In humans, and in a number of other organisms, there is also an extrinsic social control system. This is analogous to the presence of external policeman (guilt and apoptosis are the internal policemen). Mutations associated with the loss of social integration – that is, the transformation of a cell to a cancerous state – can lead to changes in the character of the cell. Specialized cells of the immune system can recognize these changes and kill the mutant cell.[186] Of course, given that tumors occur and kill people, we can assume that there are mutations that enable tumor cells to avoid such immune system surveillance. As we will see, one part of the cancerous phenotype is often a loss of normal mutation and genome repair systems. In effect, the mutant cell has increased the number of mutations, and consequently, the genetic variation in the cancer cell population. While many of these variants are lethal, the overall effect is to increase the rate of cancer cell evolution. This leads to an evolutionary race. If the cancer is killed by intrinsic and extrinsic social control systems, no disease occurs. If, however, the cancer evolves so as to avoid death by these systems, the cancer will progress and spread. As we look at a range of social systems, from cooperating bacteria to complex societies, we see examples of intrinsic and extrinsic control.

Driving the evolutionary appearance of multicellular organisms

Now that we have some idea about cooperative behaviors and how evolutionary mechanisms can select and maintain them, we can begin to consider their roles in the evolution of multicellular organisms.[187] As we have mentioned there are a number of strategies that organisms take to exploit their environment. Most prokaryotes (bacteria and archaea) are unicellular, but some can grow to gigantic sizes. For example, the bacterium *Epulopiscium fishelsoni* inhabits the gut of the brown surgeonfish *Acanthurus nigrofuscus* and can grow to more than 600 μm in length. As we will see, the cells of the unicellular eukaryotic algae of the genus *Acetabularia* can be more than 10 cm in length. Additionally, a number of multicellular prokaryotes exhibit quite complex behaviors. A particularly interesting example is a species of bacteria that form multicellular colonial organisms that sense and migrate in response to magnetic fields.[188] Within the eukaryotes, there are unicellular species, although most are significantly larger than the unicellular prokaryotes, as well as a range of macroscopic and

_____

[184] Apoptosis in cancer: http://carcin.oxfordjournals.org/content/21/3/485.full

[185] In an age of rampant narcissism and social cheating – the importance of teaching social evolutionary mechanisms.

[186] Immune recognition of self in immunity against cancer & New generations of anti-cancer drugs work by reactivating immune survalience

[187] The evolutionary-developmental origins of multicellularity: http://www.amjbot.org/content/101/1/6.long

[188] A novel species of ellipsoidal multicellular magnetotactic prokaryotes from Lake Yuehu in China.

multicellular species, including those with which we are most likely to be familiar with, namely animals, plants, and fungi.

What drove the appearance of multicellular organisms? Scientists have proposed a number of theoretical and empirically supported models. Some have suggested that predation is an important driver, either enabling the organisms to become better (or more specific) predators themselves or to avoid predation. For example, it appears that the presence of a predator can lead to the evolution of multicellularity. As an example, when the unicellular algae *Chlorella vulgaris* (5 to 6 μm in diameter) was grown together with a unicellular predator *Ochromonas vallescia*, which typically engulfs its prey, it was found that over time, *Chlorella* formed multicellular colonies that *Ochromonas* could not ingest.[189]

At this point, however, what we have is more like a colony of organisms rather than a colonial organism or a true multicellular organism. The change from multi-individual colony to multicellular organism involves cellular specialization, so that different types of cells within the organism come to carry out different functions. The most dramatic specialization being that between the cells that generate the body of the organism, somatic cells, and those that give rise to the next generation of organisms, the germ cells. At the other extreme, instead of producing distinct types of specialized cells to carry out distinct functions, a number of unicellular eukaryotes, known as protists, have complex cells that display a number of highly specialized behaviors such as directed motility, predation, osmotic regulation, and digestion (→). But such specialization can be carried out much further in multicellular organisms, where there is a socially based division of labor. The stinging cells of jellyfish provide a classic example where highly specialized cells deliver poison to any organism that touches them through a harpoon-like mechanism
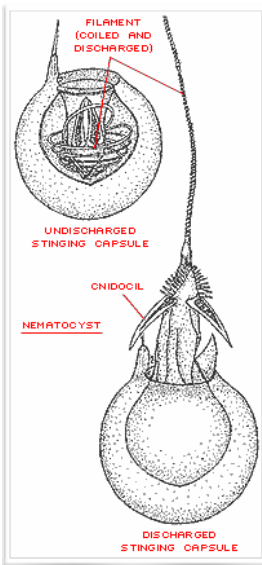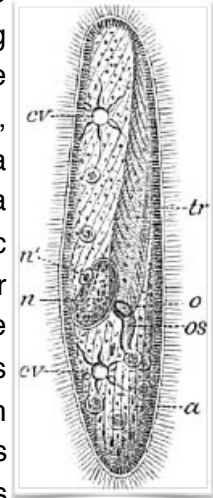
(←). The structural specialization of these cells makes processes such as cell division impossible and typically a stinging cell dies after it discharges; presumably, it is simpler to generate a new stinging cell than it is to reset a discharged cell. New cells are produced by a process known as cellular differentiation, which we will consider later (but only in passing). While we are used to thinking about individual organisms, the same logic can apply to groups of distinct organisms. The presence of cooperation extends beyond a single species, into ecological interactions in which organisms work together to various degrees to achieve that which would be much more difficult or impossible to achieve on their own (while maintain their ability to reproduce.

Based on the study of a range of organisms and their genetic information, we have begun to clarify the origins of multicellular organisms. Such studies indicate that multicellularity has arisen independently in a number of eukaryotic lineages. This strongly suggests that in a number of contexts, becoming multicellular is a successful way to establish an effective relationship with the environment.

---

[189] Phagotrophy by a flagellate selects for colonial prey: A possible origin of multicellularity

Questions to answer:

56. What type(s) of mutation would enable an organism to escape a cell death module?
57. What types of mechanisms enable organisms (cells) to recognize each other as cooperators?
58. What strategies can be used to defend against the effects of cheaters in a population?
59. How would these mechanisms apply to social interactions?
60. Make a model for the process that could lead to the evolution of social interactions.
61. What factors limit the complexity of a unicellular organism?
62. Is the schooling or herd behavior seen in various types of animals (such as fish and cows) a homologous or an analogous trait?

Questions to ponder:

- Why is r (the relationship between organisms) never 0.
- What are some of the advantages of multicellularity? What are the drawbacks? Why aren't all organisms unicellular or multicellular?

Origins and implications of sexual reproduction

One type of social interaction that we have mentioned in passing is sex. Sexual reproduction involves a cooperative interaction between organisms of different mating types, something unnecessary in asexual (clonal) reproduction. While we are used to two distinct sexes (male and female), this is not universal: many unicellular eukaryotes are characterized by an number of distinct "mating types". Typically, sexual reproduction involves the fusion of two specialized cells, known as gametes, of different mating types or sexes. Through mechanisms we will consider later, the outcome of sexual reproduction leads to increased genetic diversity among offspring.

So what are the common hallmarks of sexual reproduction? Let us return to the slime mold *Dictyostelium* as an exemplar. We have already considered its asexual life cycle, but *Dictyostelium* also has a sexual life cycle. Under specific conditions, two amoeboid cells of different mating types will fuse together (a version of sex) to form a single cell (→). The original cells are haploid, meaning that they each have a single copy of their genome. When two haploid cells fuse, the resulting cell has two copies of the genetic material and is referred to as diploid. This diploid cell can then go through a series of events, known collectively as meiosis (a topic that we will return to later on). The end result of meiosis results in the shuffling of genetic material and the production of four haploid cells. The critical point is that the genotypes of the haploid cells that emerge from meiosis are different from the haploid cells that originally fused together. Some organisms can spend a significant amount of time in the haploid state, while others spend most of their lives in the diploid state. You, for example, had a reasonably short haploid stage (as both an egg AND a sperm cell), and your diploid stage began when these two cells fused.
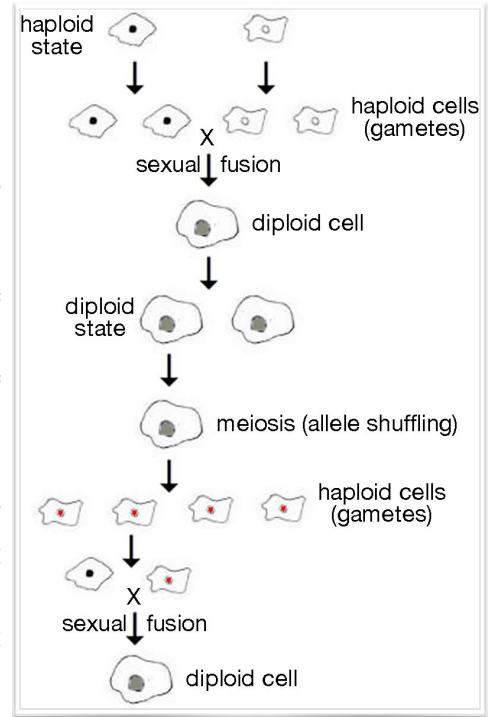
The oscillation between haploid and diploid states has some interesting implications. The first is that in the diploid state, there are (generally) two copies of each gene. The different versions of a gene are

known as alleles – these two copies can be identical or different. If they are the same, the cell/organisms is known as homozygous at that genetic locus (gene); if they are different, they are heterozygous for that gene. Alleles can have a range of phenotypic effects, from cellular lethality to more subtle effects due to differences in the activity, localization, stability, or amount of the gene product. In the diploid phase of the life cycle, the effects of a lethal or deleterious allele can be masked by the presence of the other, functional or wild type allele. Such masked alleles are commonly referred to as recessive, which we will return to in much greater detail later on. Where genes are used, that is, actively expressed and functionally important, in the haploid state, which is not always the case, the presence of a lethal allele can lead to the death of the haploid cell/organism. In this way, the presence of an extended haploid phase of an organisms' life cycle can lead to the elimination of such alleles from the population.

Sexual dimorphism

What, biologically, defines whether an organism is female or male, and why does it matter? The question is largely meaningless in unicellular organisms with multiple mating types. For example, the microbe *Tetrahymena* has seven different mating types, all of which appear morphologically identical. An individual *Tetrahymena* cell (organism) can mate with another single-celled individual of a different mating type but not with an individual of the same mating type as itself. Mating involves fusion and so the identity of the parents is lost; the four cells that are produced by the fused cell (through the process of meiosis) are of one or the other of the original mating types.

In multicellular organisms, the parents do not themselves fuse with one another. Rather they produce cells, known as gametes, that do. Also, instead of multiple mating types, there are usually only two, male and female. This, of course, leads to the question, how do we define male and female? The answer is superficially simple but its implications are profound. Which sex is which is defined by the relative size of the fusing cells the organisms produce. The larger fusing cell is termed the egg and an organism that produces eggs is termed a female. The smaller fusing cell, which is often motile (eggs are generally immotile), is termed a sperm and organisms that produce sperm are termed male. At this point, we should note the limits of these definitions. There are organisms that can change their sex, which is known as sequential hermaphroditism. For example, in a number of fish it is common for all individuals to originally develop as males; based on environmental cues, the largest of these males changes its sex to become female.[190] Alternatively, one organism can produce both eggs and sperm; such an organism is known as a simultaneous hermaphrodite.

The size difference between male and female gametes changes the reproductive stakes for the two sexes. Simply because of the larger size of the egg, the female invests more energy in its production (per egg) than a male invests in the production of each sperm cell. It is therefore relatively more important, from the perspective of reproductive success, that each egg produce a viable and fertile offspring. As the cost to the female of generating an egg, and in many organisms, the costs involved in rearing the newly formed offspring, increases, the more important the egg's reproductive success becomes. Because sperm are relatively cheap to produce, and because, in many species, males have

---

[190]Gender-bending fish: http://evolution.berkeley.edu/evolibrary/article/fishtree_07

little investment in rearing their offspring, the selection pressure associated with sperm production and sexual reproduction is often significantly less than that associated with producing an egg and rearing offspring. The end result is that there emerges a conflict of interest between females and males. This conflict of interest increases as the disparity in the relative investment per gamete or offspring increases.

This is the beginning of an evolutionary economics, cost-benefit analysis. First there is what is known as the two-fold cost of sex, which is associated with the fact that each individual asexual organism can, in theory at least, produce offspring but that two sexually reproducing individuals must cooperate to produce offspring and the resulting offspring are genetically distinct from either parent. Other, more specific factors influence an individual's reproductive costs. For example, the cost to a large female laying a small number of small eggs that develop independently is less than that of a small female laying a large number of large eggs. Similarly, the cost to an organism that feeds and defends its young for some period of time after they are born (that is, leave the body of the female) is larger than the cost to an organism that lays eggs and leaves them to fend for themselves. Similarly, the investment of a female that raises its young on its own is different from that of a male that simply supplies sperm and leaves. As you can imagine, there are many different reproductive strategies (many more than we can consider here), and they all have distinct bio-economic implications, benefits, and constraints. For example, a contributing factor in social evolution is that where raising offspring is particularly biologically expensive, cooperation between the sexes or within groups of organisms in child rearing (protection) can improve reproductive success and increase the return on the investment of the organisms involved. It is important to remember (and be able to apply in specific situations) that the reproductive costs and benefits, and so evolutionary interests, of the two sexes can diverge dramatically from one another, and that such divergence has evolutionary and behavioral implications.

Consider, for example, the situation in placental mammals, in which fertilization occurs within the female and relatively few new organisms are born from any one female. The female must commit resources to supporting the development and nurturing of the new organisms during the period from fertilization to birth. In addition, female mammals both protect their young and feed them with milk, generated using specialized mammary, that is, milk-secreting glands. Depending on the species, the young are born at various stages of development, from the active and frisky (such as goats)(→) to the relatively helpless (humans). During the period when the female feeds and protects its offspring, the female is more stressed and vulnerable than other times. Under specific conditions, cooperation with other females can occur (as often happens in pack animals) or with a specific male (typically the father) can greatly increase the rate of survival of both mother and offspring, as well as the reproductive success of the male. But consider this: how does a cooperating male know that the offspring he is helping to protect and nurture are his? Spending time protecting and gathering food for unrelated offspring is time and energy diverted from the male's search for a new mate; it might well reduce the male's overall reproductive success, and so is a behavior likely to be selected against. Carrying this logic out to its conclusion can lead to behaviors such as males guarding of females from interactions with other males.

As we look at the natural world, we see a wide range of sexual behaviors, from males who sexually monopolize multiple females (polygyny) to polyandry, where the female has multiple male "partners." In some situations, no pair bond forms between male and female, whereas in others male and female pairs are stable and (largely) exclusive. In some cases these pairs last for extremely long times; in others there is what has been called serial monogamy, pairs form for a while, break up, and new pairs form (this seems relatively common among performing arts celebrities). Sometimes females will mate with multiple males, a behavior that is thought to confuse males (they cannot know which offspring are theirs) and so reduces infanticide by males.[191] Female wild fowl (birds) can bias the success of a mating event in favor of dominant males by actively ejecting the sperm of subdominant males following mating with a more dominant male, a mating event likely to result in more robust offspring, that is, offspring more likely to survive and reproduce.[192] It should be noted that these are not conscious decisions on the part of the female but physiological responses to various cues.

It is common that while caring for their young, females are reproductively inactive. Where a male monopolizes a female, the arrival of a new male who displaces the previous male can lead to behaviors such as infanticide. By killing the young, fathered by another male, the female becomes reproductively active sooner, and so able to produce offspring related to the new male. There are situations, for example in some spiders, in which the male may risk, or even allow itself to be eaten during the course of sexual intercourse as a type of nuptial gift, which both blocks other males from mating with the female (who is, after all, busy eating) and increases the number of the male's offspring that result from the mating. This is an effective reproductive strategy for the male if its odds of mating with a female are low: better (evolutionarily) to mate (reproduce) and die than never to have mated (reproduced) at all. An interesting variation on this behavior is described in a paper by Albo et al.[193] Male *Pisaura mirablis* spiders offer females nuptial gifts, in part perhaps to avoid being eaten during intercourse. Of course, where there is a strategy, there are counter strategies. In some cases, instead of an insect wrapped in silk, the males offer a worthless gift, an inedible object wrapped in silk. Females cannot initially tell that the gift is worthless but quickly terminate mating if they discover that it is. This reduces the odds of a male's reproductive success. As deceptive male strategies become more common, females are likely to display counter strategies. For example, a number of female organisms store sperm from a mating and can eject that sperm and replace it with that of another male (or multiple males) obtained from subsequent mating events.[194] There is even evidence that in some organisms, such as the wild fowl *Gallus gallus*, females can bias against fertilization by certain males, a situation known as cryptic female choice, cryptic since it is not overtly visible in terms of who the female does or does not mate with.[195] And so it goes, each reproductive strategy leads, over time, to counter measures.[196] For example, in species in which a male guards a set of females (its harem), groups of males can work

---

191 Promiscuous females protect their offspring. http://www.ncbi.nlm.nih.gov/pubmed/16701243

192 Female feral fowl eject sperm of subdominant males: http://www.ncbi.nlm.nih.gov/pubmed/10866198

193 Worthless donations: male deception and female counter play in a nuptial gift-giving spider

194 Evolution: Sperm Ejection Near and Far: http://www.sciencedirect.com/science/article/pii/S0960982204004452

195 Cryptic female choice favors sperm from major histocompatibility complex-dissimilar males

196 Sperm Competition and the Evolution of Animal Mating Systems

together to distract the guarding male, allowing members of their group to mate with the females. These are only a few of the mating and reproductive strategies that exist.[197] Molecular studies that can distinguish an offspring's parents suggest that cheating by both males and females is not unknown even among highly monogamous species. The extent of cheating will, of course, depend on the stakes. The more negative the effects on reproductive success, the more evolutionary processes will select against it.

In humans, a female can have at most one pregnancy a year, while a totally irresponsible male could, in theory at least, make a rather large number of females pregnant during a similar time period. Moreover, the biological cost of generating offspring is substantially greater for the female, compared to the male.[198] There is a low but real danger of the death of the mother during pregnancy, whereas males are not so vulnerable, at least in this context. So, if the female is going to have offspring, it would be in her evolutionary interest that those offspring be as robust as possible, meaning that they are likely to survive and reproduce. How can the female influence that outcome? One approach is to control fertility, that is, the probability that a "reproductive encounter" results in pregnancy. This is accomplished physiologically, so that the odds of pregnancy increase when the female has enough resources to successfully carry the fetus to term. One might argue that the development of various forms of contraception are yet another facet of this type of behavior, but one in which females (and males) consciously control reproductive outcomes.

Sexual selection

As we have already noted, it is not uncommon to see morphological and behavioral differences between the sexes. Sometimes the sexual dimorphism and associated behavioral differences between the sexes are profound; they can even obscure the fact that the two sexes are actually members of the same species (→). In some cases, specific traits associated with one sex can appear to be maladaptive, that is, they might be expected to reduce rather than enhance an organism's reproductive potential.[199] The male peacock's tail, the gigantic antlers of male moose, or the bright body colors displayed by some male birds are classic examples. Darwin recognized the seriousness of this problem for evolutionary theory and addressed it in his book The Descent of Man and Selection in Relation to Sex (1871). Where the investment of the two sexes in successful reproduction is not the same, as is often the case, the two sexes may have different and potentially antagonistic reproductive strategies. Organisms of different sexes may be "looking" for different traits in their mates. In general, the larger parental investment in the production and rearing of offspring, the less random is mating and the more prominent are the effects of sexual selection, that is, the choice of

female

From "The Descent of Man" - C. Darwin

male

female

male

---
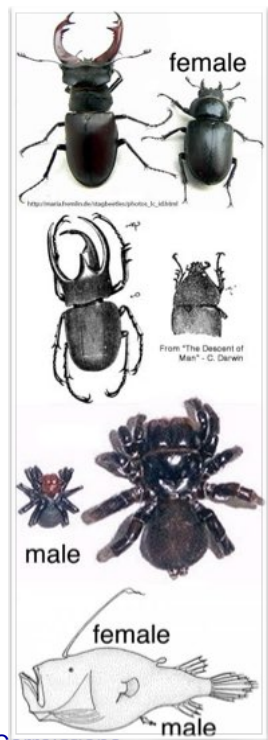
[197] The Evolution of Alternative Reproductive Strategies: Fitness Differential, Heritability, and Genetic Correlations

[198] 'Parental investment: http://www.anthro.utah.edu/PDFs/maynardsmith77parenting.pdf

[199] "Flaunting It' - Sexual Selection and the Art of Courtship: http://youtu.be/g3B8hS80k6A

who to mate with.[200] It is difficult not to place these behaviors in the context of conscious choices, (looking, wanting, etc.), but they appear to be the result of evolved (that is, selected) behaviors and do not imply self-conscious decision making or moral judgements. Presumably, they are the result of interactions between biological costs and benefits. In humans, how consciousness, self-conscious-ness, social organization, ideological and theo-political choices influence sexual behavior (and selection) is even more complex (and beyond our scope here).

Consider an example in which the female does not require help in raising offspring but in which the cost to the female is high. Selection would be expected to favor a behavior in which females mate preferentially with the most robust, but not necessarily the most cooperative or dependable males available. Females will select their mates based on male phenotype on the (quite reasonable) assumption that the most robust appearing male will be the most likely to produce the most robust offspring. In the context of this behavior, the reproductive success of a male would be enhanced if they could advertise their genetic robustness, generally through visible and unambiguous features. To be a true sign of the male's robustness, this advertisement needs to be difficult to fake and so accurately reflects the true state of the male.[201] For example consider scenarios involving territoriality. Individuals, typically males, establish and defend territories. Since there are a limited number of such territories and females only mate with males that have established and can defend such a territory, only the most robust males are reproductively successful. An alternative scenario involves males monopolizing females sexually. Because access to females is central to their reproductive success, males will interact with one another to establish a dominance hierarchy, typically in the form of one or more "alpha" males. Again, the most robust males are likely to emerge as alpha males, which in turn serves the reproductive interests of the females. This type of dominance behavior is difficult or impossible to fake. But, cooperation between non-alpha males can be used to thwart the alpha male's monopolization of females.

Now consider how strategies change if the odds of successful reproduction are significantly improved if the male can be counted on to help the female raise their joint offspring. In this situation, there is a significant reproductive advantage if females can accurately identify those males who will, in the future, display this type of reproductive loyalty.[202] Under these conditions (the shared rearing of offspring with a committed male) females will be competing with other females for access to such (perhaps rare) loyal males. Moreover, it is in the male's interest to cooperate with fertile females, and often females (but not human females) advertise their state of fertility, that is the probability that mating with them will produce offspring through external signals.

There are of course, alternative strategies. For example, groups of females, including sisters, mothers, daughters, aunts, and grandmothers can cooperate with one another, thereby reducing the importance of male cooperation. At the same time, there may be what could be termed selection conflicts. What happens if the most robust male is not the most committed male? A female could

---

[200] R. Trivers, Parent investment and Sexual selection : http://joelvelasco.net/teaching/3330/trivers72-parentalinvestment.pdf

[201] In Male Rhinoceros Beetle, Horn Size Signals Healthy Mate

[202] From an evolutionary standpoint what is the meaning of romantic love?

maximize their reproductive success by mating with a robust male and bonding with a committed male, who helps rear another male's offspring. Of course this is not in the committed male's reproductive interest. Now selection might favor male's that cooperate with one another to ward off robust but promiscuous and transient males. Since these loyal males already bond and cooperate with females, it may well be a simple matter for them to bond and cooperate with each other. In a semi-counter intuitive manner, the ability to bond with males could be selected for based on its effect on reproductive success with females. On the other hand, a male that commits himself to a cooperative (loyal and exclusive) arrangement with a female necessarily limits his interactions with other females. This implies that he will attempt to insure that the offspring he is raising are genetically related to him. Of course, another possibility is that a loyal male may be attractive to multiple females, who in turn compete for his attention and loyalty. Clearly the outcome of such interactions is influenced by how many females can the male effectively protect (that is, improve their reproductive success) as well as how significant to female reproductive success male cooperation actually is.

The situation quickly gets complex and many competing strategies are possible. Different species make different choices depending upon their evolutionary history and environmental constraints. As we noted above, secondary sexual characteristics, that is, traits that vary dramatically between the two sexes, serve to advertise various traits, including health, loyalty, robustness, and fertility. The size and symmetry of a beetle's or an elk's antlers communicate rather clearly their state of health.[203] The tail of the male peacock is a common example, a male either has a large, colorful and symmetrical tail, all signs of health or it does not – there is little room for ambiguity. These predictions have been confirmed experimentally in a number of systems; the robustness of offspring correlates with the robustness of the male, a win for evolutionary logic.[204]

It is critical that both females and males correctly read and/or respond to various traits, and this ability is likely to be selected for. For example, males that can read the traits of other males can determine whether they are likely to win a fight with that male; not

One of the most robust and reliable findings in the scientific literature on interpersonal attraction is the overwhelming role played by physical attractiveness in defining the ideal romantic partner (Hatfield & Sprecher, 1986; Jackson, 1992). Both men and women express marked preference for an attractive partner in a non- committed short-term (casual, one night stand) relationship (Buss & Schmitt, 1993).

For committed long-term relationships, females appear to be willing to relax their demand for a partner's attractiveness, especially for males with high social status or good financial prospects (for a review see Buss, 1999).

Males also look for various personality qualities (kindness, understanding, good parental skills) in their search for long-term mating partners, but unlike females, they assign disproportionately greater importance to attractiveness compared to other personal qualities (Buss, 1999).

The paramount importance of attractiveness in males' mate choices has been recently demonstrated by using the distinction between necessities (i.e., essential needs, such as food and shelter) and luxuries (i.e., objects that are sought after essential needs have been satisfied, such as a yacht or expensive car) made by economists.

Using this method, Li et al., (2002) reported that males treat female attractiveness as a necessity in romantic relationships; given a limited "mating budget," males allocate the largest proportion of their budget to physical attractiveness rather than to other attributes such as an exciting personality, liveliness, and sense of humor.
- from Mating strategies for young women by Devendra Singh (2004).

---

[203] Attractiveness of grasshopper songs correlates with their robustness against noise

[204] Paternal genetic contribution to offspring condition predicted by size of male secondary sexual character

being able to make such an accurate determination could result in crippling injuries. A trickier question is how does a female or a male determine whether a potential mate will be loyal? As with advertisements of overall robustness, we might expect that traits that are difficult or expensive to generate will play a key role. So how does one unambiguously signal one's propensity to loyalty and a willingness to cooperate? As noted above, one could use the size and value of nuptial gifts. The more valuable (that is, the more expensive and difficult the gift is to attain), the more loyal the recipient can expect the gift giver to be. On the other hand, once valuable gift-giving is established, one can expect the evolution of traits in which the cost of the gift given is reduced and by which the receiver tests the value of the gift, a behavior we might term rational skepticism, as opposed to naive gullibility.

This points out a general pattern. When it comes to sexual (and social) interactions, organisms have evolved to "know" the rules involved. If the signs an organism must make to another are expensive, there will be selective pressure to cheat. Cheating can be suppressed by making the sign difficult or impossible to fake, or by generating counter-strategies that can be used to identify fakes. These biological realities produce many behaviors, some of which are disconcerting. These include sexual cannibalism, male infanticide, and various forms of infidelity, mentioned above. What we have not considered as yet is the conflict between parents and offspring. Where the female makes a major and potentially debilitating investment in its offspring, there can be situations where continuing a pregnancy can threaten the survival of the mother. In such cases, spontaneous abortion could save the female, who can go on and mate again. In a number of organisms, spontaneous abortion occurs in response to signs of reproductive distress in the fetus. Of course, spontaneous abortion is not in the interest of the offspring and we can expect that mechanisms will exist to maintain pregnancy, even if it risks the life of the mother, in part because the fetus and the mother, while related are not identical; there can be a conflict of interest between the two.[205]

There are many variations of reproductive behavior to be found in the biological world and a full discussion is beyond our scope here, but it is a fascinating subject with often disconcerting implications for human behavior. Part of the complexity arises from the fact that the human brain (and the mind it generates) can respond with a wide range of individualistic behaviors, not all of which seem particularly rational. It may well be that many of these are emergent behaviors; behaviors that were not directly selected for but emerged in the course of the evolution of other traits, and that once present, play important roles in subsequent organismic behavior and evolution. Such emergent traits may be difficult or impossible to remove or modify, evolutionarily, if they are integral to the primary function of the trait.

Questions to answer

63. Explain how it is possible that individuals of different sexes can be in conflict, reproductively, and how do such differences impact sexual selection?
64. Explain how it is possible that a parent's interests can conflict with the interests of its offspring?
65. Why do the different sexes often display different traits?
66. If the two sexes appear phenotypically identical, what might you conclude (at least tentatively) about their reproductive behaviors?

---

[205] Maternal-Fetal Conflict: https://www2.aap.org/sections/bioethics/PDFs/Curriculum_Session14.pdf

Curbing runaway selection

Sexual selection can lead to what has been termed, but is not really, runaway selection. For example, the more prominent the peacock male's tail the more likely he will find a mate even though larger and larger tails also have significant negative effects. All of which is to say that there will be both positive and negative selection for tail size, which will be influenced by the overall probability that a particular male mates successfully. Selection does not ever really run away, but settles down when the positive benefit, in terms of sexual success, and the negative cost of a trait come to be roughly equal to each other. Sufficient numbers of male peacocks emerge as reproductively successful even if many males are handicapped by their tails and fall prey to predators. In part, this is due to the fact that, in peacocks, there is a reproductive skew for males, that is, a significant number of males in a population will never successfully mate and have offspring. In contrast, almost all females have offspring. For another example, consider the evolution of extremely large antlers associated with male dominance and mate accessibility, such as occurred in *Megaloceros giganteous* (→). These antlers can be expected to act to constrain the animal's ability to move through heavily wooded areas. In a stable environment, the costs of generating antlers and the benefits of effective sexual advertising would be expected to balance out; selection would produce an optimal solution. But if the environment changes, pre-existing behaviors and phenotypes could act to limit an organism's ability to adapt or to adapt fast enough to avoid extinction. In the end, as with all adaptations, there is a balance between costs and benefits, particularly within a changing environment.

Summary: Social and ecological interactions apply to all organisms, from bacteria to humans. They serve as a counter-balance to the common caricature of evolution as a ruthless and never ceasing competition between organisms. This hyper-competitive view, often known as the struggle for existence or Social Darwinism, was not supported by Darwin or by scientifically-established evolutionary mechanisms, but rather by a number of pundits who used it to justify various political (that is, inherently non-scientific) positions, particularly arguing against social programs that helped the poor (often characterized as the unfit) at the "expense" of the wealthy. Assuming that certain organisms were inherently less fit, and that they could be identified, this view of the world gave rise to Eugenics, the view that genetically inferior people should be killed, removed, or sterilized, before their "bad" traits overwhelmed a particular culture. Eugenics was a particularly influential ideology in the United States during the early part of the 20th century and inspired the genocidal programs of the Nazis in Germany. What is particularly odd about this evolutionary perspective is that it is actually anti-evolutionary, since if the unfit really were unfit, they could not possibly take over a population. In addition, it completely ignores the deeply social (cooperative) aspect of the human species.

Questions to answer

67. What does it mean to cheat, in terms of sexual selection - is the "cheating" organism actually being consciously deceptive?

68. What types of "cheating" behaviors do females use with males? What about males with females?

69. What are the costs involved when a male tries to monopolize multiple females? What are the advantages?
70. What limits runaway selection, or better, why is runaway selection impossible?

Questions to ponder
- Should human ethical or ideological beliefs and decisions be more important than evolutionary cost-benefit calculations?

Chapter 5: Molecular interactions, thermodynamics & reaction coupling

In which we drastically change gears and move from evolutionary mechanisms to the physicochemical properties of organisms that, of course, constrain evolutionary possibilities and biological behaviors. We consider how molecules interact and react with one another and how these interactions and reactions determine the properties of substances and systems, particularly the bounded, non-equilibrium system that is life.
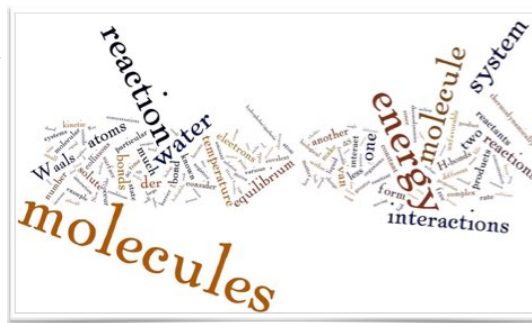
A very little thermodynamics

While the diversity of organisms and the unique properties of each individual organism are the products of evolutionary processes initiated billions of years ago, it is equally important to recognize that all biological systems and processes, from growth and cell division to thoughts and feelings, obey the rules of chemistry and physics, in particular the laws of thermodynamics. What makes biological systems unique is that, unlike simpler physicochemical systems that move toward thermodynamic equilibrium, organisms must maintain an uninterrupted non-equilibrium state in order to remain alive. While a chemical reaction system is easy to assemble *de novo*, every biological system has been running continuously for billions of years. So, before we continue we have to be clear about what it means and implies when we say that a system is at equilibrium versus being in a obligate non-equilibrium state, since a biological system at equilibrium is dead.

To understand the meaning of thermodynamic equilibrium we have to learn to see the world differently, and learn new meanings for a number of words. First we have to make clear the distinction between the macroscopic world that we perceive directly and the sub-microscopic, molecular world that we can understand based on scientific observations and conclusions - it is this molecular world that is particularly important in the context of biological systems. The macroscopic and the molecular worlds behave very differently - in particular, the molecular world often behaves stochastically. To illustrate this point we will use a simpler model that displays the basic behaviors that we want to consider but is not as complex as a biological system. In our case let us consider a small, well-insulated air-filled room in which there is a table upon which is resting a bar of gold – we use gold since it is chemically rather inert, that is, un-reactive. Iron bars, for example, could rust, which would complicate things. In our model the room is initially at a cosy 70 ºF (~21 ºC) and the gold bar is at 200ºC. What will happen as a function of time; try and generate a graph that describes how the system behaves.

Our first task is to define the system – that is, the part of the universe we are interested in. We could define the system as the gold bar or the room with the gold bar in it. Notice, we are not really concerned about how the system came to be the way it is - that is, its history. We could, if we wanted to, demonstrate convincingly that the system's history has no influence on its future behavior – this is a critical difference between biological and simple physicochemical systems. We are, however, concerned as to whether the system is open or closed, that is whether energy and matter can enter or leave the system. For now we will consider the room to be a closed system - no energy enters or leaves it.

Common sense tells us that energy will be transferred from the gold bar to the rest of the room and that the temperature of the gold bar will decrease over time, while the final temperature of the room + the gold bar will depend upon relative sizes of both. This energy transfer occurs primarily through molecular collisions between the molecules of the gold bar together with the molecules in the air and the table. The behavior of the system has a temporal direction. Why do you think that is? Why, exactly, doesn't the hot bar get hotter and the rest of the system, the room, get cooler? We will come back to this question shortly. What may not be quite as obvious is that the temperature of the room will increase slightly as the gold bar cools. Eventually the block of gold and the room will reach the same temperature; when that happens, the system will be said to be at equilibrium.

Remember we defined the system as closed; no matter or energy passes into or out of the room. Because it is a closed system, once the system reaches its final temperature no further macroscopic change will occur. This does not mean, however, that nothing is going on. If we could look at the molecular level we would see that molecules of air are moving, constantly colliding with one another and colliding with the particles within the bar, the table, and the walls of the room. The molecules within the bar and the table are also vibrating. The speeds of these molecular movements is a function of temperature, the higher or lower the temperature, the faster or slower these motions, on average, will be. Collisions between molecules can change the velocities of the colliding molecules. What would happen if there was no air in the room or if it were possible to suspend the gold bar in the center of the room, for example if the room were in outer space? Would this change your graph of system's behavior?

As we consider further on, all of the molecules in the system have kinetic energy, which is the energy of motion. Through their interactions (primarily collisions), the kinetic energy of any one particular molecule will be constantly changing. At the molecular level the system is dynamic, even though at the macroscopic level it is static. And this is what is important about a system at equilibrium: it is macroscopically static, there is no net change, even though at the molecular level there is still movement. The energy of two colliding molecules is the same after a collision as before, even though the energy may be distributed differently between the colliding molecules. In physical terms, the system as a whole cannot do anything. that is, it cannot do work - no macroscopic changes are possible. This is a weird idea, since (at the molecular level) things are still moving. So, as we return to living systems, which are clearly able to do lots of things, including moving macroscopically, growing, thinking, and such, it is clear that they cannot be at equilibrium. We will come back to this insight repeatedly.

We can ask, then, what is necessary to keep a system from reaching equilibrium? The most obvious answer (we believe) is that unlike our imaginary closed system, a non-equilibrium system must be open, that is, energy and matter must be able to enter and leave the system. An open system is no longer isolated from the rest of the universe, it is part of it. For example, we might imagine a system in which energy, in the form of radiation, can enter and leave our room. We could maintain a difference in the temperature between the bar and the room by illuminating the bar and removing heat from the room as a whole. A temperature difference between the bar and the room could then (in theory) produce what is known as a heat engine, which can do work (that is, produce macroscopic changes.) As long as we continue to heat the block and remove heat from the rest of the system, we can continue to do work, that is, macroscopically observable changes can happen.

Cryptobiosis: At this point, we have characterized organisms as dynamic, open, non-equilibrium systems. An apparent exception to the dynamic aspect of life are organisms that display a rather special phenotypic adaptation, known generically as cryptobiosis. Organisms, such as the tardigrad, or water bear (→), can be freeze-dried and persist in a state of suspended animation for decades. What is critical to note, however, is that when in this cryptobiotic state the organism is not at equilibrium, in much the same way that a piece of wood in air is not at equilibrium, but capable of reacting. The organism can be reanimated when returned to normal conditions.[206] Cryptobiosis is a genetically-based adaptation that takes energy to produce and energy is needed to emerge from stasis. While the behavior of tardigrads is extreme, many organisms display a range of adaptive behaviors that enable them to survive hostile environmental conditions.
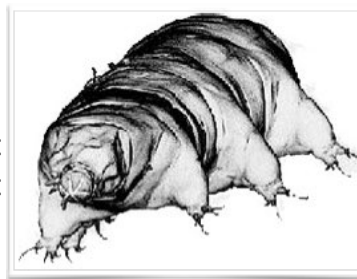
Reactions and energy: favorable and unfavorable, their dynamics and coupling

As we will see, biological systems are extremely complex; both their overall structural elements and many of their molecular components (including DNA and proteins) are the products of thermodynamically unfavorable processes and reactions. How do these reactions take place in living systems? The answer comes from the coupling of thermodynamically favorable reactions to thermodynamically unfavorable reactions. This is a type of work, although not in the standard macroscopic physics model of work (w) = force x distance. In the case of (chemical) reaction coupling, the work involved drives thermodynamically unfavorable reactions, typically the synthesis of large and complex molecules and macromolecules (that is, very large molecules). Here we will consider the thermodynamics of these processes.

Thermodynamics is at its core about energy and changes in energy. This leads to the non-trivial question, what is energy? Many have struggled to answer this question, and there is no simple satisfactory answer. Perhaps a way around it is to say that for every change, there is also an associated energy change. While it may appear that there are many types of energy (and you may have been taught this earlier) in fact there are only two forms of energy, kinetic and potential. For example, the energy associated with the movement and vibrations of objects with mass is kinetic energy. Potential energy is associated with a object's position in a field (electrical, magnetic, gravitational) and the particle's nature, its mass, electrical charge, "spin". All systems, whether they are macroscopic or microscopic can be characterized in terms of the sum of their kinetic and potential energies. But wait, you might say, what about the energy associated with electromagnetic radiation, the most familiar form is visible light. Electromagnetic radiation is a form of kinetic energy, energy that is transferred from place to place via photons. Finally, there is the counterintuitive idea that energy and matter, are interconvertible as described by the equation:

$$e \text{ (energy)} = m \text{ (mass)} \times c^2 \text{ (c = speed of light)}$$

but not to worry, such interconversion events are not directly relevant to biological systems.

---

[206] On dormancy strategies in tardigrades & Towards decrypting cryptobiosis

That said, it is clear that kinetic energy can be converted into potential energy and vice versa. To illustrate this principle, we can call on our day-to-day experiences. Forces (which mediate the transfer of energy) can be used to make something move. Imagine a system of a box sitting on a rough floor. You shove the box so that it moves (but do not continue to push it) – the box travels a short distance and then stops. By shoving the box you added (kinetic) energy to the system. The first law of thermodynamics (see below) states that the total energy in a system is constant. So the question is where has the energy gone when the box slows and stops moving? One answer might be that the energy was destroyed - but we know that that is not true. Careful observations lead us to conclude that the energy still exists but that it has been transformed and/or transferred somewhere else. Measurements can prove that the mass of the box has not changed. In fact, if we measured the temperature of both the box and the floor we would see that both have increased (by a very small amount). The friction generated by moving the box represents an increase in the movements of the molecules of the box and the floor over which the box moved. Through collisions and vibrations, this energy will, over time, be distributed throughout the system - the temperature of the system will increase (if only slightly). The presence of this thermal motion is revealed by what is known as Brownian motion. In 1905, Albert Einstein explained Brownian motion in terms of the existence, size, and movements of molecules (→).[207]
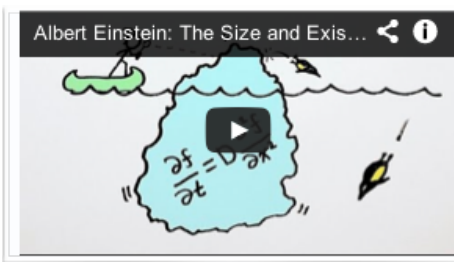
In the system we have been considering, the energy that was transferred to the box by pushing it has been spread throughout the system. While one can use the a directed push (input energy) to move something (to do work), the diffuse thermal energy cannot be used to do work. While the total amount of energy is conserved, its ability to do things has been decrease (almost abolished). This involves the concept of entropy, which we will turn to next.

Questions to answer:

71. How does energy move from molecule to molecule within a system?

72. What are the common components of a non-equilibrium system and explain why a dried out tardigrad is alive?

Thinking entropically (and thermodynamically)

We certainly are in no position to teach you (rigorously) the basics of physics, chemistry and chemical reactions, but we can provide a short refresher that focuses on the key points we will be using over and over again.[208] The first law of thermodynamics is that the total amount of energy within a closed system remains constant. The energy may be transformed from kinetic to potential (and vice versa) but in a closed system the total does not change. Again, we need to explicitly recognize the distinction between a particular system and the universe as a whole, although the universe as a whole is itself (apparently) a closed system. For any system we must define a system boundary; this can be a real boundary such as a container, or even an imaginary boundary. What is inside the boundary is part

---

[207] Albert Einstein: The Size and Existence of Atoms http://youtu.be/nrUBPO6zZ40

[208] Of course, we recommend a chemistry course sequence based on Cooper & Klymkowsky, 2014. Chemistry, Life, the Universe and Everything: here: http://clue.chemistry.msu.edu/

of the system, and the rest of the universe outside of the boundary layer is not. While we will consider the nature of the boundary of biological systems (cells) in greater molecular detail in the next chapter, we can anticipate that one of the boundary's key features is its selectivity in letting matter and energy pass into and out of the system, and what constraints it applies to those movements.

Assuming that you have been introduced to chemistry, you might recognize the Gibbs free energy equation: $\Delta G = \Delta H – T\Delta S$, where T is the temperature of the system.[209] From our biological perspective, we can think of ΔH as the amount of heat energy transferred between the system and the surroundings during any change, and ΔS as the change in a system factor known as entropy. Entropy is related to the ways that energy and matter can be arranged, and the more possible ways, the greater the entropy. In the earlier example of the gold bar in the isolated room, energy is transferred between the bar and the room until the two are at equal temperatures; over time, the bar and the room come equilibrium. The process does not run in reverse, the bar does not get hotter while the room cools. This is because transferring energy from hot to cold is statistically more probable (See CLUE for more details), and the factor that we use to characterize these probabilities is called entropy (S). Often entropy is used colloquially to describe random or disordered systems, or the state of systems, and it is true that a gas (which is more disordered) has more entropy than a liquid (which is less disordered), but technically the gas has more entropy because there are more possible arrangements for the gas particles and their associated energies.

For any change, the entropy of the universe always increases - which is usually stated as the Second Law of Thermodynamics. This Law has never been disproven. At this point you might be saying wait a minute, I know changes where the entropy decreases, and you would be right. For example, it is certainly possible to change a gas (high entropy) into a liquid (lower entropy), but the critical part here is that this system is not the universe. While the system may decrease in entropy, the universe still increases. This is because when gas → liquid energy most be removed and that energy is transferred to the surroundings, which increases the entropy of the surroundings (by making molecules move fast and vibrate). While the entropy of particular region may decrease, the total entropy of the universe increases.

It turns out that it is difficult to measure energy and entropy changes for the universe. Usually we can only do this for the system we are studying. Fortunately there is a way to account for the total entropy change during a process (or reaction) using the equation $\Delta G = \Delta H – T\Delta S$, which tells us about the change in energy (and therefore entropy) for a process within a system. When ΔG is < 0 we say the change is thermodynamically favorable, and can occur. Conversely when ΔG is > 0 we say the change is thermodynamically unfavorable, and will not occur. When ΔG for the system = 0 no observable (macroscopic) change will occur. The system is at equilibrium.

Every reaction is characterized by its equilibrium constant, K_{eq}, that is a function of both the reaction itself and the conditions under which the reaction is carried out. These conditions include parameters such as the initial state of the system, the concentrations of the reactants, and system temperature and pressure. In biological systems we generally ignore pressure, although pressure will be important for

---

[209] in the real world, the value of ΔG depends upon the concentrations of solute and solvent, but we will ignore that complexity for the moment.

organisms that live on the sea floor (or mountain tops).

The equilibrium constant (K_{eq}) for a reaction A + B ⇋ C + D is defined (→) as the $$K = \frac{[C][D]}{[A][B]}$$ concentrations of the products (C and D) divided by the concentrations of the reactants at equilibrium, where nothing macroscopic is happening. At equilibrium the concentrations are not changing (and that is why K is a constant). For a thermodynamically favorable reaction, that is one that favors the products, K will be greater, often much greater than one. The larger K_{eq} is, the more product and the less reactant there will be when the system reaches equilibrium. If the equilibrium constant is less than 1, then at equilibrium, the concentration of reactants will be greater than the concentration of products.

While the concentration of reactants and products of a reaction at equilibrium remains constant it is a mistake to think that the system is static. If we were to peer into (or imagine) the system at the molecular level we would find that, at equilibrium, reactants are continuing to form products and products are rearranging to form reactants at similar rates.[210] That means that the rate of the forward reaction is equal to the rate of the reverse reaction. If, at equilibrium, a reaction has gone almost to completion and $K_{eq} \gg 1$, there will be very little of the reactants left and lots of the products. Given that most reactions involve physical collisions between molecules, the changes in the frequency of productive collisions between reactants or products increases as their concentrations increase. Even improbable events can occur, albeit infrequently, if the probability of precursor events (collisions between particular molecules) is high enough.

Reaction rates

Knowing whether a reaction is thermodynamically favorable and its equilibrium constant does not tell us much (or really anything) about whether the reaction actually occurs to any significant extent under a particular set of conditions. To know the reaction's rate we need to know how the rate changes depending upon the time and the concentrations of reactants (or products) for the specific system with which we are dealing. Such reaction kinetics data tell us the rate at which the reaction actually occurs under a particular set of conditions. For example, consider a wooden log, which is composed mainly of the carbohydrate polymer cellulose $(CH_2O)_n$. In the presence of molecular oxygen (O_2) the reaction:

$$nO_2 + \text{wooden log } ((CH_2O)_n) ⇋ nCO_2 + nH_2O + heat$$
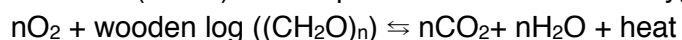
is extremely thermodynamically favorable, that is, it has a negative ΔG and a large equilibrium constant (once the reaction starts it goes completely to CO_2 and H_2O), yet the log is stable - trees do not burst into flames spontaneously. The question is, of course, why not? Or more generally why is the world so annoyingly complex?

The answer lies in the details of the reaction, how exactly the reactants are converted into the products. In the case of logs burning we have to apply a spark, perhaps a lightening strike. In general we have to supply some energy to get the reaction started. This is called the activation energy, and all

---

[210] This, of course, assumes that we have a closed system, that is, that neither the products or the reactants can leave the system, and that the volume of the system also remains constant. If the reactants can "leave the scene" of the reaction, then of course the back reaction, Products ⇋ Reactants, will be much less likely to occur.

reactions require some activation energy to get started (otherwise the world would just fall apart). For simplicity let us consider another non-chemical but rather widespread type of reaction. In this reaction system, there is a barrier between two compartments, specifically the barrier membrane that separates the inside from the outside of a cell. At this point, we do not need to consider the exact details of the barrier's structure (although we will in the next chapter). In our particular example, outside the cell the concentration of molecule A is high while inside the cell its concentration is low. We can write out this reaction equation as A $_{outside}$ ⇆ A $_{inside}$; perhaps you can make a prediction of the ΔG of this reaction and

what your prediction is based on (note that in this case at equilibrium K = [A $_{inside}$]/[A $_{outside}$] = 1 (because at equilibrium these concentrations will be equal. The reaction consists of moving A molecules across the barrier between the inside and the outside of the cell. In our example, the concentration of A outside the cell, written [A$_{outside}$], with the square brackets indicating concentration, is much greater than [A$_{inside}$]. At any moment in time, the number of collisions between A$_{outside}$ and the barrier will be much greater than the number of collisions between A$_{inside}$ and the barrier. Assuming that the probability of crossing the barrier is a function of the collision frequency, there will be a net movement of A$_{outside}$ to A$_{inside}$. The real question is how large the net flux (flux = movement of A in − movement of A out) is. The rate of movement will depend on the amount of energy a molecule needs to cross the barrier. We can represent this energy as the highest peak in a reaction graph; here we assume a simple process with a single peak, in the real world it can involve a number of sub-reactions and look more like a roller-coaster than a simple hill (→). In such a graph, we begin with the free energy of the reactants along the Y-axis, and plot the changing free energies of the various intermediates along the X-axis, leading to the free energy of the products. In our simplified view of the subject, the difference between the point with the highest free energy the transition state and the free energy of the reactants (G$_{reactants}$) is known as the activation energy and determines the rate limiting step in the reaction. For a thermodynamically favorable reaction the Gibbs free energy of the products

is smaller than the Gibbs free energy of the reactants: that is ΔG is negative (G$_{products}$ − G$_{reactants}$). Conversely, if ΔG is positive the reaction is unfavorable and will not be observed unless we do something about it.
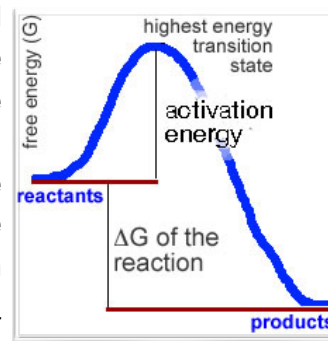
The reason why (most) thermodynamically favorable reactions do not occur immediately when reactants come into contact is that there has to be enough energy in the system to surmount the activation energy barrier (ΔG$_{activation}$). For example, for the reaction involving movement through a barrier, the movement of reactants (A $_{outside}$) to products (A $_{inside}$), the reactants must capture enough energy from their environment to traverse the barrier between outside and inside. In biological systems there are two major sources for this energy: light and collisions with other molecules. A molecule can absorb a photon (a particle of light) or energy can be transferred to a molecule from other molecules through a collision. In liquid water, molecules are moving; at room temperature they move on average at about 640 meters/second. That is not to say that all molecules are moving with the same speed. If we were to look at the population of molecules, we would find a distribution of speeds known as a Boltzmann (or Maxwell-Boltzmann) distribution (↓). As they collide with one another, the molecules exchange kinetic energy, and one molecule can emerge from a collision with much more energy than it

entered with.[211] Since reactions occur at temperatures well above absolute zero, there is plenty of energy available in the form of the kinetic energy of molecules. Occasionally a molecule with high
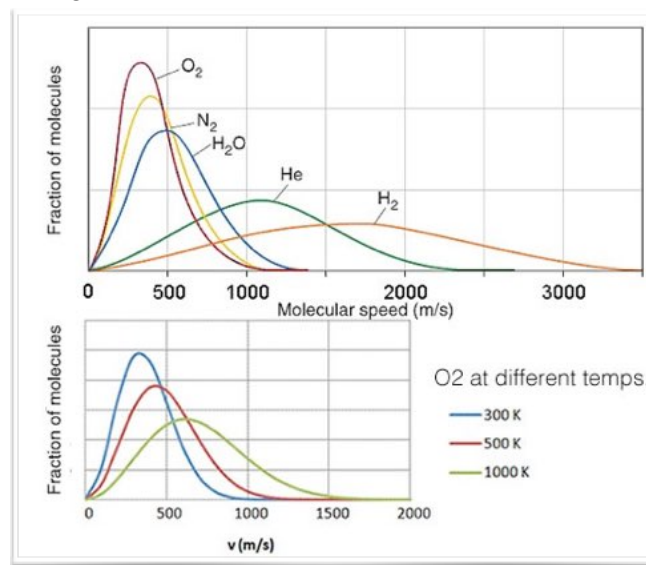
energy will emerge from a collision. If this molecule collides with a membrane it can cross the boundary layer, that is, move from outside to inside, or vice versa. If it does not have sufficient energy it will simply bounce back and collide with other molecules. It is this dynamic exchange of kinetic energy between molecules that drives the movement of molecules, as well as providing the activation energy to initiate the breaking of bonds associated with chemical reactions, the first step in many reactions).

In the case of our model barrier system, since the A molecules are the same whether inside or outside the cell, the difference in the free energies of the reactants and products reflects (primarily) the difference in their concentrations. A higher concentration correlates with higher free energy, remember, we are interested in the ΔG of the

$$A_{outside} \leftrightarrows A_{inside} \text{ reaction.}$$

Clearly the more molecules of A are present, the higher the actual Gibbs energy of A (another way to think of this is that the probability of collisions is higher when the concentration is higher). One point is worth emphasizing, it is possible for a reaction to have a large $\Delta G_{reaction}$ and either a large or small activation energy. So assuming that there is enough energy in the system, and the activation energy is small enough, the reaction can proceed rapidly, or at least at a noticeable rate. You should be able to predict what happens to the system as it moves toward equilibrium. On the other hand, if the activation energy is high enough, the $A_{outside} \leftrightarrows A_{inside}$ reaction will not occur to any significant extent. This is why trees (and humans) are safe, we do not undergo spontaneous combustion, even though we are composed of thermodynamically unstable molecules; there has to be an addition of energy to start any such reaction.

Questions to answer:
73. In the context of the $A_{outside} \leftrightarrows A_{inside}$ reaction, what does the reaction graph look like when $[A_{inside}] = [A_{outside}]$ or when $[A_{outside}] > [A_{inside}]$?
74. A reaction is at equilibrium; we increase the amount of reactant or product. What happens (over time) to the amounts of reactants and products?
75. What does reducing the activation energy of a reaction do to a system at equilibrium? What does it do to a system far from equilibrium?
76. Where does the energy come from to reach (and pass through) the transition state?

_____

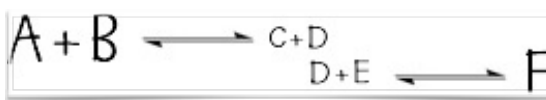[211] This behavior is illustrated here

Coupling reactions

There are large numbers of different types of reactions that occur within cells. As a rule of thumb, a reaction that produces smaller molecules from larger ones will be thermodynamically favored, while reactions that produce larger molecules from smaller ones will be unfavorable. Similarly a reaction that leads to a molecule moving from a region of higher concentration to a region of lower concentration will be thermodynamically favorable. So how exactly can we build the big molecules, such as DNA and proteins, and generate the concentration gradients upon which life depends?

As we noted before reactions can be placed into two groups, those that are thermodynamically favorable (negative ΔG, equilibrium constant greater, typically much greater, than 1) and those that are thermodynamically unfavorable (positive ΔG, equilibrium constant less, often much less than 1). Thermodynamically favored reactions are typically associated with the breakdown of various forms of food molecules and the release of energy, known generically as catabolism. Reactions that build up biomolecules, known generically as anabolism, are typically thermodynamically unfavorable. An organism's metabolism is the sum total of all of these various reactions.

An unfavorable reaction can occur when it is coupled to a thermodynamically favorable reaction. This requires that the two reactions share a common intermediate. In this example (→) the two reactions share the component "D". Let us assume that the upper reaction is unfavorable while the lower reaction is favorable. What happens? Let us further assume that both reactions are occurring at measurable rates, perhaps through the mediation of appropriate catalysts; a catalyst is a substance that lowers the activation energy of a reaction. Assume that E is present within the system. At the start of our analysis, the concentrations of A and B are high. We can then use Le Chatelier's principle to make our predictions.[212]

Let us illustrate how Le Chatelier's principle works. Assume for the moment that the reaction
$A + B \leftrightharpoons C + D$ has reached equilibrium, the rates of the forward reaction and the reverse reaction are equal. Now consider what happens to the reaction if, for example, we remove (somehow, do not worry about how) C from the system. Now the rate of the reverse reaction will slow down (because there is not much C to collide with D to initiate the reaction). This means that the rate of the forward reaction is now greater than the reverse reaction: the reaction is no longer at equilibrium. The reaction moves to the right even though that reaction is thermodynamically unfavorable. Similarly if we add some B, the rate of the forward reaction will increase and the reaction will move to the right to produce more products, until a new equilibrium position is established. In this case, the addition of B leads to the increased rate of production of C + D until their concentration reached a point where the rate of the C + D $\leftrightharpoons$ A + B reaction is equal to the A + B $\leftrightharpoons$ C + D reaction. This type of behavior arises directly from the fact that at equilibrium reaction systems are not static but dynamic (at the molecular level) – things are still occurring but at the same rate so that no net change occurs. When you add or take something away from the system, it becomes unbalanced, that is, it is no longer at equilibrium. Because the reactions are occurring at a measurable rate, the system will return to equilibrium over time. This

---

[212] http://en.wikipedia.org/wiki/Le_Chatelier's_principle

general idea is called Le Chatelier's principle, which states that if a change is made to a system at equilibrium, then the system will shift to counteract that change, basically because the number of productive collision events associated with one direction of the reaction will increase compared to those associated with the other direction.

So back to our system of coupled reactions. As the unfavorable A+B reaction occurs and approaches equilibrium it will produce a small amount of C+D. However, the D+E reaction is favorable, and as D is formed it will react with E to produce F, while at the same time removing D from the system. As D is removed, it influences the A+B reaction (because it makes the C+D "back reaction" less probable even though the A+B "forward reaction" continues.) The result is that more C and D will be produced. Assuming that a sufficient amounts of E is present, more D will be removed. The end result is that, even though it is energetically unfavorable, more and more C and D will be produced, while D will be used up to make F. It is the presence of the common component D and its use as a reactant in the D+E reaction that drives the synthesis of C from A and B, something that would normally not be expected to occur to any great extent. Imagine then, what happens if C is also a reactant in some other favorable reaction(s)? In this way reactions systems are linked together, and the biological system proceeds to use energy and matter from the outside world to produce the complex molecules needed for its maintenance, growth, and reproduction.

Questions to answer:
77. How does adding or removing components of the reaction system change the energy of the system?
78. How is LeChatelier's principle involved in reaction coupling?
79. When examining a reaction system, how would you go about deciding whether the system involved coupled reactions?
80. What does a catalyst do? Draw the effect of adding a catalyst in terms of effects on reaction graphs.
81. Assume that the reactions within a reaction system require catalysts to occur at reasonable rates; what happens within a reaction systems if the catalysts are missing or inactive?

Question to ponder:
- Why are catalysts required for life?

Inter- and Intra-molecular interactions

We have briefly (admittedly absurdly briefly) defined what energy is and begun to consider how it can be transferred in reaction systems. Now we need to consider what we mean by matter, which implies an understanding of the atomic organization of the molecules that compose matter. As you hopefully know by now, all matter is composed of atoms. The internal structure of atoms is the subject of quantum physics and we will not go into it in any depth. Suffice to say that each atom consists of a tiny positively charged nucleus and cloud of negatively charged electrons.[213] Typically atoms and molecules, which after all are collections of atoms, interact with one another through a number of different types of forces. The first are known as van der Waals interactions, which are mediated by London Dispersion Forces (LDF). These forces arise from the fact that the relatively light negatively-charged electrons are in continual movement, compared to the relatively massive and stationary

---

[213] Why don't electrons fall into the nucleus

positively-charged nuclei.[214] (→) Because charges on the protons and electrons are equal in magnitude the atom is electrically neutral, but because the electrons are moving, at any one moment, an observer outside of the atom or molecule will experience a small fluctuating electrical field.
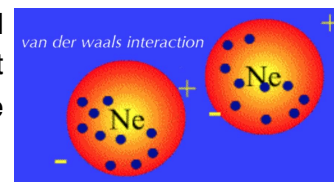
As two molecules approach one another, their fluctuating electric fields will interact, this interaction generates an attractive force, named after its discoverer Fritz Wolfgang London (1900–1954). This force varies as $\sim 1/R^6$ where R is the distance between the molecules; this relationship means that LDFs act only over very short distances, typically less than 1 nanometer (1 nm = 10^{-9} m). As a frame of reference, a carbon atom has a radius of ~0.07 nm. The magnitude of this attractive force reaches its maximum when the two molecules are separated by what is known as the sum of their van der Waals radii (the van der Waals radius of a carbon atom is ~0.17 nm (→). If they move closer than this distance, the attractive LDF is quickly overwhelmed by the rapidly increasing, and extremely strong repulsive force that arises from the electrostatic interactions between the negatively charged electrons of the two molecules, and the two positively charged nuclei.
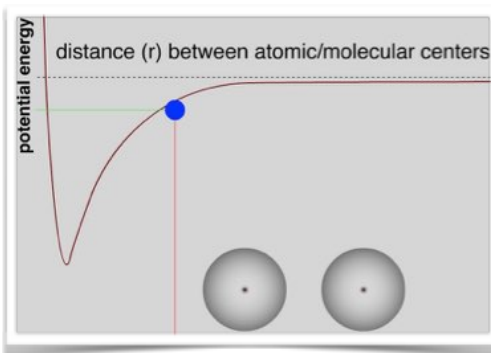
Each atom and molecule has its own characteristic van der Waals radius, although since most molecules are not spherical, it is better to refer to a molecule's van der Waals surface. This surface is the closest distance that two molecules can approach one another before repulsion kicks in and drives them back away from one another. It is common to see molecules displayed in terms of their van der Waals surfaces. Every molecule generates LDFs when it approaches another, so van der Waals interactions are universal.

The strength of the van der Waals interactions between molecules is determined primarily by their shapes. The greater the surface complementarity between two molecules, the stronger their interaction. Compare the interaction between two monoatomic Noble atoms, such as helium, neon or argon, and two molecules with more complex shapes (→). The two monoatomic particles interact via LDFs at a single point, so the strength of the interaction is minimal. On the other hand, the two more complex molecules interact over extended surfaces, so the LDFs between them are greater resulting a stronger van der Waals interaction.
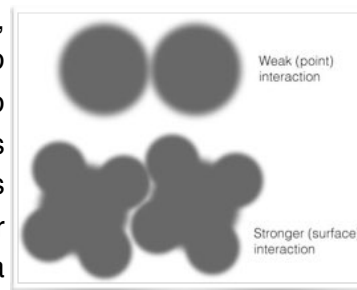
Covalent bonds

In the case of van der Waals interactions, the atoms and molecules involved retain their hold on their electrons, they remain distinct and discrete. There are cases, however, where atoms come to "share" each other's electrons. This sharing involves pairs of electrons, one from each atom. When electron pairs are shared, the atoms stop being distinct in that their shared electrons are no longer

---

[214] explored further at: http://virtuallaboratory.colorado.edu/LDF+binding-interactions/1.2-interactions-0.html

restricted to one or the other. In fact, since one electron cannot even in theory be distinguished from any other electron, they become a part of the molecule's electron system. This sharing of electrons produces what is known as a covalent bond. Covalent bonds are ~20 to 50 times stronger than van der Waals interactions. What exactly does that mean? Basically, it takes 20 to 50 times more energy to break a covalent bond compared to the energy needed to break a van der Waals interaction. While the bonded form of atoms in a molecule is always more stable than the unbounded form, it may not be stable enough to withstand the energy delivered by collisions with neighboring molecules. Different bonds between different atoms in different molecular contexts differ in terms of bond stability; the bond energy refers to the energy needed to break a particular bond. A molecule is stable if the bond energies associated with bonded atoms within the molecule are high enough to survive the energy delivered to the molecule through either collisions with neighboring molecules or the absorption of energy (light).

When atoms form a covalent bond, their individual van der Waals surfaces merge to produce a new molecular van der Waals surface. There are a number of ways to draw molecules, but the space-filling or van der Waals surface view is the most realistic (at least for our purposes). While realistic it can also be confusing, since it obscures the underlying molecular structure, that is, how the atoms in the molecule are linked together. This can be seen in this set of representations of the simple molecule 2-methylpropane (→).[215] As molecules become larger, as is the case with many biologically important molecules, it can become impossible to appreciate their underlying organization based on a van der Waals surface representation.
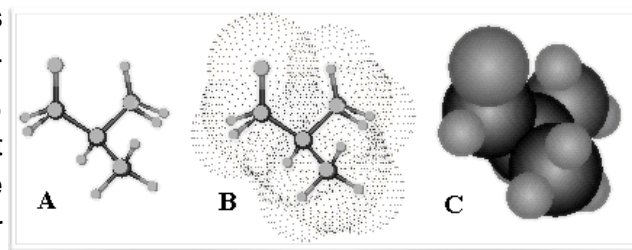
Because they form a new stable entity, it is not surprising (perhaps) that the properties of a molecule are quite distinct from, although certainly influenced by, the properties of the atoms from which they are composed. The shapes of molecules are determined by each atom's underlying quantum mechanical properties and, particularly as molecules get larger, as they so often do in biological systems, the interactions between different parts of the molecule with one another. Some atoms, common to biological systems, such as hydrogen (H), can form only a single covalent bond. Others can make two (oxygen (O) and sulfur (S)), three (nitrogen (N)), four (carbon (C)), or five (phosphorus (P)) bonds.

In addition to smaller molecules, biological systems contain a number of distinct types of extremely large molecules, composed of many thousands of atoms; these are known as macromolecules. Such macromolecules are not rigid; they can often fold back on themselves leading to intramolecular interactions. There are also interactions between molecules. The strength and specificity of these interactions can vary dramatically and even small changes in molecular structure, such as caused by mutations and associated allelic variations, can have dramatic effects on molecular shape and function.

Molecules and molecular interactions are dynamic. Collisions with other molecules can lead to parts of a molecule rotating with respect to one another around a single bond.[216] The presence of a double bond restricts these kinds of movements; rotation around a double bond requires what amounts to

---

[215] Explicit Concepts of Molecular Topology: http://www.chem.msu.ru/eng/misc/babaev/match/top/top02.htm

[216] This could be basis of a square dance like in class activity!

breaking and then reforming one of the bonds. In addition, and if you have mastered some chemistry you already know this, it is often incorrect to consider bonds as distinct entities, isolated from one another and their surroundings. Adjacent bonds can interact forming what are known as resonance structures that behave as mixtures of single and double bonds. Again this restricts free rotation around the bond axis and acts to constrain molecular geometry. As we will come to see, the peptide bond that occurs between a carbon (C) and a nitrogen (N) atom in a polypeptide chain, is an example of such a resonance structure. Similarly, the ring structures found in the various "bases" present in nucleic acids result in flat structures that pack one on top of another. These various geometric complexities combine to make predicting a molecule's three dimensional structure increasingly challenging as its size increases.

Bond stability and thermal motion (a non-biological moment)

Molecules do not exist out of context. In the real, or at least the biological world they do not sit alone in a vacuum. Most biologically-relevant molecular interactions occur in aqueous solution. That means, biological molecules are surrounded by other molecules, mostly water molecules. As you may already know from physics there is a lowest possible temperature, known as absolute zero (0 K, −273.15 ℃ , −459.67 °F). At this biologically irrelevant temperature, molecular movements are minimal but not, apparently, absent all together.[217]

When we think about a system, we inevitably think about its temperature. Temperature is a concept that makes sense only at the system level. Individual molecules do not have a temperature. The temperature of a system is a measure of the average kinetic energy of the molecules within it. The average kinetic energy is:

$$E_k = 1/2 \text{ (average mass) x (average velocity)}^2$$

It does not matter whether the system is composed of only a single type of molecule or many different types of molecules, at a particular temperature the average kinetic energy of all of the different molecules has one value. This is not to say that all molecules have the same kinetic energy, they certainly do not; each forms part of a distribution that is characterized by its average energy, this distribution is known as the Boltzmann or Maxwell-Boltzmann distribution (→). The higher the temperature, the more molecules will have a higher kinetic energy.
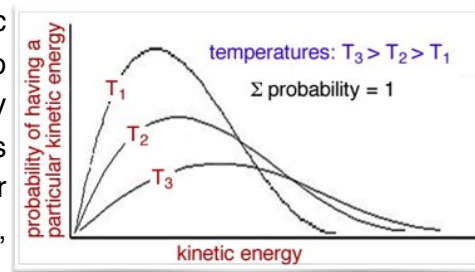
In a gas we can largely overlook the attractive intermolecular interactions between molecules because the average kinetic energies of the molecules of the system are sufficient to disrupt such intermolecular interactions - that is, after all, why they are a gas. As we cool the system, we remove energy from it, and the average kinetic energy of the molecules decreases. When the average kinetic energy gets low enough, the molecules will form a liquid. In a liquid, the movement of molecules is not sufficient to disrupt the interactions between them. This is a bit of a simplification, however. Better to think of it more realistically. Consider a closed box partially filled with a substance in a liquid state. What

---

[217] zero point energy (from wikipedia)

is going on? Assuming there are no changes in temperature over time, the system will be a equilibrium. What we will find, if we think about it, is that there is a reaction going on, that reaction is:

Molecule (gas) ⇋ Molecule (liquid).
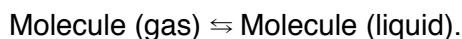
At a particular temperature, the liquid phase is favored, although there will be some molecules in the system's gaseous phase. The point is that at equilibrium, the number of molecules moving from liquid to gas will be equal to the number of molecules moving from the gas to the liquid phase. If we increase or decrease the temperature of the system (that is add or remove energy), we will alter this equilibrium state, that is, the relative amounts of molecules in the gaseous versus the liquid states will change. The equilibrium is dynamic, in that different molecules may be in gaseous or the liquid states, even though the level of molecules will be steady.

In a liquid, while molecules associate with one another, they can still move with respect to one another. That is why liquids can be poured, and why they assume the shape of the (solid) containers into which they are poured. This is in contrast to the container, whose shape is independent of what it contains. In a solid the molecules are tightly associated with one another and so do not translocate with respect to one another, although they can rotate and jiggle in various ways. Solids do not flow. The cell, or more specifically, the cytoplasm, acts primarily as a liquid. Most biological processes take place in the liquid phase: this has a number of implications. First molecules, even very large macromolecules, move with respect to one another. Driven by thermal motions, molecules will move in a Brownian manner, a behavior known as a random walk.

Thermal motion will influence whether and how molecules associate with one another. We can think about this process in the context of an ensemble of molecules, let us call them A and B; A and B interact to form a complex, AB. Assume that this complex is held together by van der Waals interactions. In an aqueous solution, the A:B complex is colliding with water molecules. These water molecules have various energies (from low to high), as described by the Boltzmann distribution. There is a probability that in any unit of time, one or more of these collisions will deliver energy greater than the interaction energy that holds A and B together; this will lead to the disassociation of the AB complex into separate A and B molecules. Assume we start with a population of 100% AB complexes, the time it takes for 50% of these molecules to dissociate into A and B is considered the "half-life" of the complex. We use the term half-life repeatedly to characterize the stability of a complex or macromolecule. Now here is the tricky part, much like the situation with radioactive decay, but subtly different. While we can confidently conclude that 50% of the AB complexes will have disassembled into A and B at the half-life time, we can not predict exactly which AB complexes will have disassembled and which will remain intact. Why? Because we cannot predict exactly which collisions will provide sufficient energy to disassociate a particular AB complex.[218] Dissociation is a stochastic process, and like all stochastic processes (such as genetic drift) is best understood in terms of probabilities.

Stochastic processes are particularly important within biological systems because, generally, cells are small and may contain only a relatively small number of molecules of a particular type. If, for example, the expression of a gene depends upon a protein binding to a specific site on a DNA molecule, and if there are relatively small numbers of that protein and usually only one or two copies of

---

[218] It should be noted that, in theory at least, we might be able to make this prediction if we mapped the movement of every water molecule. This is different from radioactive decay, where it is not even theoretically possible to predict the behavior of an individual radioactive atom.

the gene, that is, the DNA molecule, present in a cell, we will find that whether or not a copy of the protein is bound to a specific region of the DNA is a stochastic process.[219] If there are enough cells, then the group average may well be predictable, but the behavior of any one cell will not be.[220] In an individual cell, sometimes the protein will be bound and the gene will be expressed and sometimes not, all because of thermal motion and the small numbers of interacting components involved. This stochastic property of cells can play important roles in the control of cell and organismic behavior.[221] It can even transform a genetically identical population of organisms into subpopulations that display two or more distinct behaviors, a property with important implications, that we will return to.

Questions to answer:
77. How does temperature influence intermolecular interactions? How might changes in temperature influence molecular shape (particularly in a macromolecule)?
78. Why is the effect of temperature on covalent bond stability is not generally significant in biological systems?
79. In considering generating a graph that describes radioactive decay or the dissociation of a complex, like the AB complex discussed above, why does population size matter?

Questions to ponder:
- Why is the Boltzmann distribution asymmetric around the highest point.

Bond polarity, inter- and intramolecular interactions

So far, we have been considering covalent bonds in which the sharing of electrons between atoms is more or less equal, but that is not always the case. Because of their atomic structures, based on quantum mechanical principles, not to be discussed here, different atoms have different affinities for their own electrons. When an electron is removed or added to an atom (or molecule) that atom/molecule becomes an ion. Atoms of different elements differ in the amount of energy it takes to remove an electron from them; this is, in fact, the basis of the photoelectric effect explained by Albert Einstein, in another of his 1905 papers.[222] One way to characterize this property is through electronegativity. Each type of element has a characteristic electronegativity, a measure of how tightly it holds onto its electrons when it is bonded to another atom, an idea that you may have mastered in general chemistry. If the electronegativities of the two atoms in a bond are equal or similar, then the electrons are shared more or less equally between the two atoms and the bond is said to be non-polar, meaning without direction. There are no stable regions of net negative or positive charge on the surface of the resulting molecule. If the electronegativities of the two bonded atoms are unequal, however, then the electrons will be shared un-equally. On average, there will be more electrons more of the time around the more electronegative atom and fewer around the less electronegative atom. This leads to partially negatively and partially positively-charged regions to the bonded atoms – the bond has a direction. Charge separation produces an electrical field, known as a dipole. A bond between atoms of differing electronegativities is said to be polar.

---

[219] This is illustrated here and we will return to this type of behavior later on.

[220] Biology education in the light of single cell/molecule studies

[221] Single Cells, Multiple Fates, and Biological Non-determinism: https://www.ncbi.nlm.nih.gov/pubmed/27259209

[222]Albert Einstein: Why Light is Quantum: http://youtu.be/LWIi7NO1tbk

Atoms of O and N are more electronegative than C and H, and will sequester electrons when bonded to atoms of H and C. The O and N become partly negative and the C and H become partly positive. Because of the quantum mechanical organization of atoms, these partially negative regions are organized in a non-uniform manner (the atoms have regions with different partial charges), which we will return to. In contrast, there is no significant polarization of charge in bonds between C and H atoms, and such bonds are termed non-polar. The presence of polar bonds leads to the possibility of electrostatic interactions between molecules. Such interactions are stronger than van der Waals interactions but much weaker than covalent bonds; like covalent bonds they have a directionality to them – the three atoms involved have to be arranged more or less along a straight line. There is no similar geometric constraint on van der Waals interactions.

Since the intermolecular forces arising from polarized bonds often involve an H atom interacting with an O or an N atom, these have become known generically and perhaps unfortunately, as hydrogen or H-bonds (→). Why unfortunate? Because H atoms can take part in covalent bonds, but H-bonds are not covalent bonds, they are very much weaker. It takes much less energy to break an H-bond between molecules or between parts of (generally macro-) molecules that it does to break a covalent bond involving a H atom.
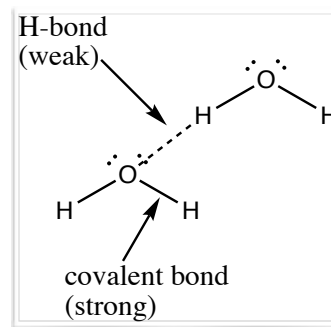
The implications of bond polarity

Melting and boiling points are important physical properties of molecules, although this applies primarily to small molecules and not macromolecules. Here we are considering a pure sample that contains extremely large numbers of the molecule in question. Let us start at a temperature at which the sample is liquid. The molecules are moving with respect to one another, there are interactions between the molecules, but they are transient - the molecules are constantly switching neighbors. As we increase the temperature of the system, the energetics of collisions are now such that all interactions between neighboring molecules are broken, and the molecules fly away from one another. If they happen to collide with one another, they do not adhere; the bond that might form is not strong enough to resist the kinetic energy delivered by collision with other molecules. The molecules are said to be a gaseous state and the transition from liquid to gas is the boiling point. Similarly, starting with a liquid, when we reduce the temperature, the interactions between molecules become longer lasting until a temperature is reached at which the energy transferred through collisions is no longer sufficient to disrupt the interactions between molecules.[223] As more and more molecules interact, the position of neighboring molecules becomes permanent - the liquid is transformed into a solid. While liquids flow and assume the shape of their containers, because neighboring molecules are free to move with respect to one another, solids maintain their shape–neighboring molecules stay put. The temperature at which a liquid changes to a solid is known as the melting point. These temperatures mark what are known as phase transitions: solid to liquid and liquid to gas.

---

[223] The nature of the geometric constrains on inter-molecular interactions will determine whether the solid is crystalline or amorphous. see: https://en.wikipedia.org/wiki/Crystal

At the macroscopic level, we see the rather dramatic effects of bond polarity on melting and boiling points by comparing molecules of similar size with and without polar bonds and the ability to form H-bonds (↓). For example, neither CH_4 (methane) and Ne (neon) contain polar bonds and cannot form intra-molecular H-bond-type electrostatic interactions. In contrast NH_3 (ammonia), H_2O (water), and FH (hydrogen fluoride) have three, two and one polar bonds, respectively, and can take part in one or more intra-molecular H-bond-type electrostatic interactions. All five compounds have the same number of electrons, ten. When we look at their melting and boiling temperatures, we see rather immediately how the presence of polar bonds influences these properties.

Compounds	CH_4	NH_3	OH_2	FH	Ne
molecular weight	16.04	17.02	18.02	20.01	20.18
bond electronegativity	0.45	0.94	1.34	1.88	N/A
# of electrons	10	10	10	10	10
# of bonds	4	3	2	1	0
melting point	-182°C	-77.7°C	0°C	-83°C	-248.6°C
boiling point	-161.5°C	-33.4°C	100°C	19.5°C	-246.1°C

In particular, water stands out as dramatically different from the rest of the molecules, with significantly higher (> 70ºC) melting and boiling points than its neighbors. So why is water different? Well, in addition to the presence of polar covalent bonds, we have to consider the molecule's shape. Each water molecule can take part in four hydrogen bonding interactions with neighboring molecules - it has two partially positive Hs and two partially negative sites on its O. These sites of potential H-bond-type electrostatic interactions are arranged in a nearly tetrahedral geometry (→). Because of this arrangement, each water molecule can interact through H-bond-type electrostatic interactions with four neighboring water molecules. To remove a molecule from its neighbors, four H-bond-type electrostatic interactions must be broken, which is relatively easy since they are each rather weak. In the liquid state, molecules jostle one another and change their H-bond-type electrostatic interaction partners constantly. Even if one interaction is broken, however, the water molecule is likely to remain linked to multiple neighbors via H-bond-type electrostatic interactions.
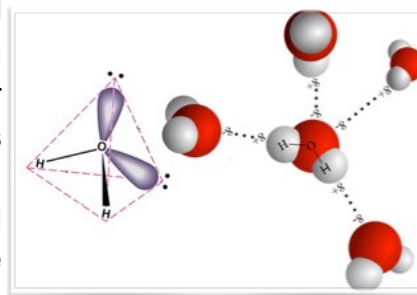
This molecular hand-holding leads to water's high melting and boiling points as well as its high surface tension. We can measure the strength of surface tension in various ways. The most obvious is the weight that the surface can support. Water's surface tension has to be dealt with by those organisms that interact with a liquid-gas interface. Some, like the water strider, use it to cruise along the surface of ponds. As the water strider (←) walks on the surface of the water, the molecules of its feet do not form H-bond-type electrostatic interactions with water molecules, they are said to be hydrophobic, although that is clearly a bad name - they are not afraid of water, rather they are simply apathetic to it. Hydrophobic molecules interact with other molecules, including water molecules, only through van der Waals interactions. Molecules that can make H-bonds (or other polar interactions) with water are termed hydrophilic. As molecules increase in size they can

have regions that are hydrophilic and regions that are hydrophobic (or hydroapathetic). Molecules that have distinct hydrophobic and hydrophilic regions are termed amphipathic and we will consider them in greater detail in the next chapter.

Interacting with water

We can get an idea of the hydrophilic, hydrophobic/hydroapathetic, and amphipathic nature of molecules through their behaviors when we try to dissolve them in water. Molecules like sugars (carbohydrates), alcohols, and most amino acids are primarily hydrophilic, they dissolve readily in water. Molecules like fats are highly hydrophobic (hydroapathetic), and they do not dissolve significantly in water. So why the difference? To answer this question we have to be clear what we mean when we say that a molecule is soluble in water. We will consider this from two perspectives. The first is what the solution looks like at the molecular level, the second is how the solution behaves over time. To begin we need to understand what water alone looks like. Because of its ability to make and donate multiple H-bond-type electrostatic interactions in a tetrahedral arrangement, water molecules form a dynamic three-dimensional intermolecular interaction network. In liquid water the H-bond-type electrostatic interactions between the molecules break and form rapidly.

To insert a molecule A, known as a solute, into this network you have to break some of the H-bond-type electrostatic interactions between the water molecules, known as the solvent. If the A molecules can make H-bond-type electrostatic interactions with water molecules, that is, if they are hydrophilic, then there is little net effect on the free energy of the system. Such a molecule is soluble in water. So what determines how soluble the solute is. As a first order estimate, each solute molecule will need to have at least one layer of water molecules around it, otherwise it will be forced to interact with other solute molecules. If the number of these interacting solute molecules is large enough, the solute will no longer be in solution. In some cases, aggregates of solute molecule can, because they are small enough, remain suspended in the solution. This is a situation known as a colloid. While a solution consists of individual solute molecules surrounded by solvent molecules, a colloid consists of aggregates of solute molecules in a solvent. We might predict that all other things being equal (a unrealistic assumption), the larger the solute molecule the lower its solubility. You might be able to generate a similar rule for the size of particles in a colloid.

Now we can turn to a conceptually trickier situation, the behavior of a hydrophobic solute molecule in water. Such a molecule cannot make H-bond-type electrostatic interactions with water molecules, so when it is inserted into water the total number of H-bond-type electrostatic interactions in the system decreases - the energy of the system increases (remember, bond forming lowers potential energy). However, it turns out that much of this "enthalpy" change, conventionally indicated as ΔH, is compensated for by van der Waals interactions (that is, non-H-bond-type electrostatic interactions) between the molecules. Generally, the net enthalpic effect is minimal. Something else must be going on to explain the insolubility of such molecules.

Turning to entropy

In a liquid, water molecules will typically be found in a state that maximizes the number of H-bond-type electrostatic interactions present. Because these interactions have a distinct, roughly tetrahedral geometry, their presence constrains the possible orientations of molecules with respect to one another. This constraint is captured when water freezes; it is the basis for ice crystal formation, why the density of water increases before freezing and decreases with freezing, and why ice floats in liquid water.[224] In the absence of a hydrophobic solute molecule there are many many equivalent ways that liquid water molecules can interact to produce these geometrically specified arrangements. But the presence of a solute molecule constrains the number of appropriate orientations of water molecules: a much smaller number of configurations result in maximizing H-bond formation between water molecules. The end result is that the water molecules become arranged in a limited number of ways around each solute molecule; they are in a more ordered, that is, in a more improbable state than they would be in the absence of solute. The end result is that there will be a decrease in entropy (indicated as ΔS), the measure of the probability of a state. ΔS will be negative compared to arrangement of water molecules in the absence of the solute.

How does this influence whether dissolving a molecule into water is thermodynamically favorable or unfavorable. It turns out that the interaction energy (ΔH) of placing most solutes into the solvent is near 0, so that it is the ΔS that makes the difference. Keeping in mind that $\Delta G = \Delta H - T\Delta S$, if ΔS is negative, then −TΔS will be positive. The ΔG of a thermodynamically favorable reaction is, by definition, negative. This implies that the reaction:

$$\text{water + solute} \leftrightharpoons \text{solution (water + solute)}$$

will be thermodynamically unfavorable; the reaction will move to the left. That is, if we start with a solution, it will separate so that the solute is removed from the water. How does this happen? The solute molecules aggregate with one another. This reduces their effects on water, and so the ΔS for aggregation is positive. If the solute is oil, and we mix it into water, the oil will separate from the water, driven by the increase in entropy associated with minimizing solute-water interactions. This same basic process has a critical influence on macromolecular structures.

Questions to answer:
80. Predict (and explain your prediction), the factors that influence the solubility of a molecule in water
81. Why does the separation of oil and water represent a more disordered state?
82. How would you explain to a "normal" person how it is possible for a water strider to walk on water; what concepts would you need to introduce them to?
83. Predict (and explain the basis of your prediction) the effects of H-bonding on a molecule's boiling point.

Questions to ponder:
- Given what you know about water, why is ice less dense than liquid water?

---

[224] Why does ice float in water? http://youtu.be/UukRgqzk-KE

Chapter 6: Membrane boundaries and capturing energy

In which we consider how the aqueous nature of biological systems drives the formation of lipid-based barrier membranes and how such membranes are used to capture and store energy from the environment and chemical reactions.   We consider how coupled reactions are used to drive macromolecular syntheses and growth, and how endosymbiotic events, involving the capture of aerobic and photosynthetic bacteria, played a critical role in the evolution of eukaryotic cells

Defining the cell's boundary

A necessary step in the origin of life was the generation of a discrete barrier, a boundary layer, that separates the living non-equilibrium reaction system from the rest of the universe. This original boundary layer, the structural ancestor of the plasma membrane of modern cells, serves to maintain the integrity of the living system and mediates the movement of materials and energy into and out of the cell. The plasma membrane of all cells, whether bacterial, archaeal or eukaryotic, appears to be a homologous structure derived from a precursor present in the last common ancestor of life. So what is the structure of this barrier (plasma) membrane? How is it built and how does it work?

When a new cell is formed its plasma membrane is derived from the plasma membrane of the progenitor cell. As the cell grows, new molecules are added into the membrane to enable it to increase its surface area. Biological membranes are composed of two general classes of molecules, proteins (which we will discuss in much greater detail in the next section of the course) and lipids. It is worth noting explicitly that, unlike a number of other types of molecules that we will be considering, such as proteins, nucleic acids, and carbohydrates, lipids are not a structurally coherent group, that is they do not have one particular basic structure. Structurally diverse molecules, such as cholesterol and phospholipids, are both considered lipids (→). While there is a relatively small set of common lipid types, there are many different lipids found in biological systems and the characterization of their structures and functions has led to a new area of specialization known as lipidomics.[225]

All lipids have two distinct domains: a hydrophilic domain (circled in red in this figure →) characterized by polar regions and one or more hydrophobic/hydroapathetic domains that are usually made up of C and H and are non-polar. Lipids are amphipathic. In aqueous solution, entropic effects will drive the hydrophobic/hydroapathetic parts of the lipid out of an aqueous solution. But in contrast to totally non-polar molecules, like oils, the hydrophobic/hydroapathetic part of the lipid is connected to a hydrophilic domain that is soluble in water. Lipid molecules deal with this dichotomy by

---

[225] O[n the future of "omics": lipidomics](#) & [Lipidomics: new tools and applications](#)
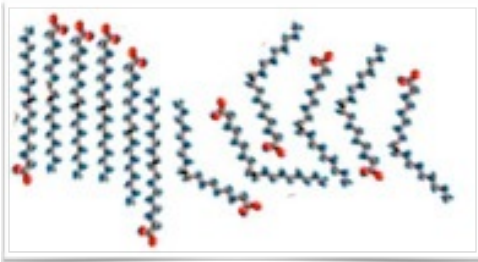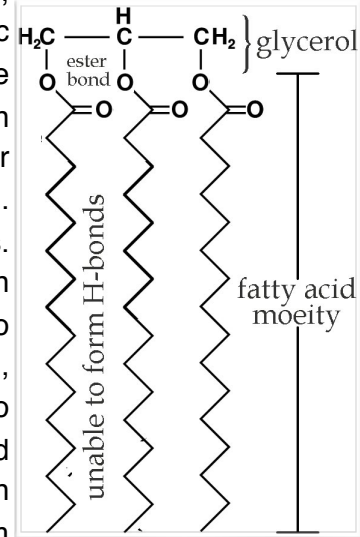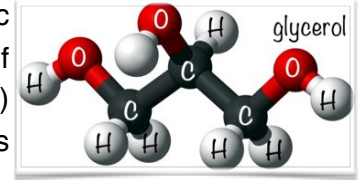
associating with other lipid molecules in multimolecular structures in which the interactions between the hydrophilic parts of the lipid molecule and water molecules are maximized and the interactions between the lipid's hydrophobic/hydroapathetic parts and water are minimized. Many different multi-molecular structures can be generated that fulfill these constraints (→). The structures that form depend upon the details of the system, including the shapes of the lipid molecules involved and the relative amounts of water and lipid present. In every case, the self-assembly of these structures leads to an increase in the total overall entropy of the system, a somewhat counterintuitive idea. For example, in a micelle the hydrophilic region is in contact with the water, while the hydrophobic regions are inside, away from direct contact with water. This leads to a more complete removal of the lipid's hydrophobic domain from contact with water than can be arrived at by a purely hydrophobic oil molecule, so unlike oil, lipids can form stable structures in solution. The diameter and shape of the micelle is determined by the size of its hydrophobic domain. As this domain gets longer, the center of the micelle becomes more crowded. Another type of organization that avoids "lipid-tail crowding" is known as a bilayer vesicle. Here there are two layers of lipid molecules, pointing in opposite directions. The inner layer surrounds a water-filled region, the lumen of the vesicle, while the outer layer interacts with the external environment. In contrast to the situation within a micelle, the geometry of a vesicle means that there is significantly less crowding as a function of lipid tail length. Crowding is further reduced as a vesicle increases in size to become a cellular membrane. Micelles and vesicles can form a colloid-like system with water, that is they exist as distinct structures that can remain suspended in a stable state. We can think of the third type of structure, the planar membrane, as simply an expansion of the vesicle to a larger and more irregular size. Now the inner layer faces the inner region of the cell (which is mostly water) and the opposite region faces the outside world, which again is mostly water. For the cell to grow, new lipids have to be inserted into both inner and outer layers of the membrane; how exactly this occurs typically involves interactions with proteins. For example, there are proteins that can move a lipid from the inner to the outer domain of a membrane; they flip the lipid between layers, and are known as flipases (when we consider proteins, perhaps you can generate a plausible mechanism for such an activity.)

A number of distinct mechanisms are used to insert molecules into membranes, but they all involve a pre-existing membrane – this is another aspect of the continuity of life. Totally new cellular membranes do not form, membranes are built on pre-existing membranes. For example, a vesicle, that is a spherical lipid bilayer, can fuse into or emerge from a planar (bilayer) membrane. These processes are typically driven by thermodynamically favorable reactions involving protein-based molecular machines. When the membrane involved is the plasma (boundary) membrane, these processes are known as exocytosis and endocytosis, respectively. These terms refer explicitly to the fate of the material within the vesicle. Exocytosis releases that material from the vesicle interior into the outside world, whereas endocytosis captures material from outside of the cell and brings it into the cell. Within a cell, vesicles can fuse with and emerge from one another.

As noted above, there are hundreds of different types of lipids, generated by a variety of biosynthetic pathways catalyzed by proteins encoded in the genetic material. We will not worry too much about all of these different types of lipids, but we will consider two generic classes, the glycerol-based lipids (→) and cholesterol, because considerations of their structures illustrates general ideas related to membrane behavior. In bacteria and eukaryotes, glycerol-based lipids are typically formed from the highly hydrophilic molecule glycerol combined with two or three fatty acid molecules (a three fatty acid chain molecule is shown →). Fatty acids contain a long chain hydrocarbon with a polar (carboxylic acid) head group. The molecular nature of these fatty acids influences the behavior of the membrane formed. Often these fatty acids have what are known as saturated hydrocarbon tails. A saturated hydrocarbon contains only single bonds between the carbon atoms of its tail domain. While these chains can bend and flex, they tend to adopt a more or less straight configuration. In this straight configuration, they pack closely with one another, which maximizes the lateral (side to side) van der Waals interactions between them. Because of the extended surface contact between the chains, lipids with saturated hydrocarbon chains are typically solid around room temperature (←). Solid means that the molecules rarely not tend to exchange positions with one another. On the other hand, there are cases where the hydrocarbon tails are "unsaturated", that is they contain double bonds (–C=C–) in them. These are typically more fluid and flexible because unsaturated hydrocarbon chains have permanent kinks due to the rigid nature and geometry of C=C bonds, so they cannot pack as regularly as saturated hydrocarbon chains. The less regular packing means that there is less interaction area between the molecules, which lowers the strength of the van der Waals interactions between them. Lower van der Waals interaction energy in turn, lowers the temperature at which these bilayers change from a solid, no movement of the lipids relative to each other within the plane of the membrane, to a liquid, with relatively free movements within the plane of the membrane. Recall that the strength of interactions between molecules determines how much energy is needed to overcome a particular type of interaction. Because these van der Waals intermolecular interactions are relatively weak, changes in environmental temperature influence the physical state of the membrane. The liquid-like state is often referred to as the fluid state. The membrane's state is important because it can influence the behavior and activity of the proteins embedded within it. If it is the membrane is in a solid state, proteins within the membrane will be immobile; if is in the liquid state, these proteins will move by diffusion, that is, by collision-driven movements within the plane of the membrane. In addition, since lipids and proteins are closely associated with one another in the membrane, the physical state of the membrane can influence the activity of embedded proteins, a topic to which we will return.
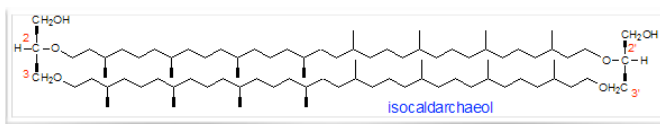
Cells can manipulate the solid-to-liquid transition temperature of their membrane by altering the membrane's lipid composition. Increasing the ratio of saturated to unsaturated chains can increase the

melting temperature. Controlling chain saturation involves altering the activities of the enzymes involved in various saturation/desaturation reactions. That these enzymes can be regulated implies a feedback mechanism, by which either temperature or membrane fluidity acts to regulate metabolic processes. This type of feed back mechanism is part of what is known as the homeostatic and adaptive system of the cell (and the organism) and is another topic we will return to as we proceed.

There are a number of differences between the lipids used in bacterial and eukaryotic organisms and archaea.[226] For example, instead of straight chained hydrocarbons, archaeal lipids are constructed of branched isoprene ($CH_2=C(CH_3)CH=CH_2$) polymers linked to the glycerol group through an ether, rather than an ester linkage (→). The bumpy and irregular shape of the isoprene groups (compared to the relatively smooth saturated hydrocarbon chains) means that archaeal membranes will tend to melt (go from solid to liquid) at lower temperatures.[227] At the same time the ether linkage is more stable (requires more energy to break) than the ester linkage. It remains unclear why the bacteria and the eukaryotes use straight chain hydrocarbon lipids, while the archaea use isoprene-based lipids. One speculation is that the archaea were originally adapted to live at higher temperatures, where the greater stability of the ether linkage would provide a critical advantage.
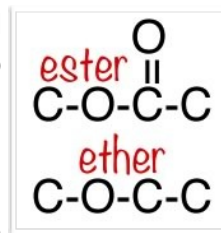
Some archaea and bacteria, known generically as thermophiles and hyper-thermophiles, live (happily, apparently) at temperatures up to 110 °C.[228] At the highest temperatures, thermal motion might be expected to disrupt the integrity of the membrane, allowing small charged molecules (ions) and other larger hydrophilic molecules to pass through the membrane.[229] Given the importance of membrane integrity, you may (perhaps) not be surprised to find "double-headed" lipids in such thermophilic organisms. These lipid molecules have two distinct hydrophilic glycerol moieties (←), one located at each end of the molecule; this enables them to span the membrane. The presumption is that such lipids act to stabilize the membrane against the disruptive effects of high temperatures.

The solid-fluid nature of biological membranes, as a function of temperature, is complicated by the presence of cholesterol and structurally similar lipids. For example, in eukaryotes the plasma membrane can contain as much as 50% cholesterol, in terms of the number of molecules present (→). Cholesterol has a short bulky hydrophobic domain that does not pack well with other lipids: a hydrocarbon chain lipid (left) and cholesterol (right). Its presence dramatically influences the solid-liquid behavior of the membrane. The diverse roles of lipids is a complex subject that goes beyond our scope here.
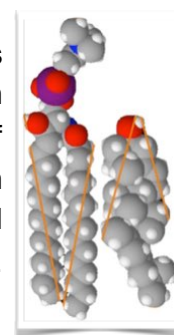
[226] A re-evaluation of the archaeal membrane lipid biosynthetic pathway

[227] The origin and evolution of Archaea: a state of the art

[228] You might want to consider how this is possible and under want physical conditions you might find these "thermophilic" archaea.

[229] Ion permeability of the cytoplasmic membrane limits the maximum growth temperature of bacteria and archaea

The origin of biological membranes

The cell membrane is composed of a number of different types of lipids. The hydrophobic "tails" of modern lipids range from 16 to 20 carbons in length. The earliest membranes, however, were likely to have been composed of similar molecules with shorter hydrophobic chains. Based on the properties of lipids, we can map out a plausible scenario for the appearance of membranes. Lipids with very short hydrophobic chains, from 2 to 4 carbons in length, can dissolve in water, can you explain why? As the lengths of the hydrophobic chains increases, the molecules begin to self-assemble into micelles. By the time the hydrophobic chains reach ~10 carbons in length, it becomes more difficult to fit the hydrocarbon chains into the interior of the micelle without making larger and larger spaces between the hydrophilic heads. Water molecules can begin to move through these spaces and interact with the hydrocarbon tails. At this point, the hydrocarbon-chain lipid molecules begin to associate into semi-stable bilayers (→). One interesting feature of bilayers is that the length of the hydrocarbon chain is no longer structurally limiting, in contrast to the situation in micelles. One problem, though, are the edges of the bilayer, where the hydrocarbon region of the lipid would come in contact with water, a thermodynamically unfavorable situation. This problem is avoided by linking edges of the bilayer to one another, forming a balloon-like structure. Such bilayers can capture regions of solvent, that is water and the solutes dissolved within it.
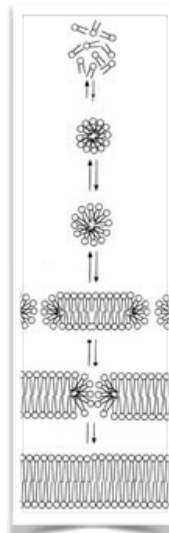
Bilayer stability increases further as hydrophobic chain length increases. At the same time, membrane permeability decreases. It is a reasonable assumption that the earliest biological systems used shorter chain lipids to build their "proto-membranes" and that these membranes were relatively leaky.[230] The appearance of more complex lipids, capable of forming more impermeable membranes, must therefore have depended upon the appearance of mechanisms that enabled hydrophilic molecules to pass through such membranes. The interdependence of change is known as co-evolution. Co-evolutionary processes were apparently common enough to make the establishment of living systems possible.

Questions to answer:
84. Draw diagrams to show how increasing the length of a lipid's hydrocarbon chains affects the structures that it can form and use your diagrams to explain how the effects at the hydrophobic edges of a lipid bilayer are minimized?
85. Some lipids have phosphate groups attached to the glycerol as well as fatty acids - explain how the presence of "phospho-lipids" will impact membrane structure and stability.
86. Make a set of general rules on the effects of size and composition on the ability of a molecule to pass through a membrane.

Questions to ponder:
- Why do fatty acid and isoprene lipids form similar bilayer structures?

---

230 http://astrobiology.arc.nasa.gov/workshops/1996/astrobiology/speakers/deamer/deamer_abstract.html

Transport across membranes

As we have said before (and will say again), the living cell is a historically continuous non-equilibrium system. To maintain its living state both energy and matter have to move into and out of the cell, which leads us to consider intracellular and extracellular environments and the membrane that separates them. The differences between the regions inside and outside of the plasma membrane are profound. Outside, even for cells within a multicellular organism, the environment is generally mostly water, with relatively few complex molecules. Inside the membrane-defined space is the cytoplasm, a highly concentrated (300 to 400 μg/ml) solution of proteins, nucleic acids, smaller molecules, and thousands of interconnected chemical reactions.[231] Cytoplasm (and the membrane around it) is inherited by each cell when it is formed, and represents an uninterrupted continuous system that first arose more than 3 billion years ago.

A lipid bilayer membrane poses an interesting barrier to the movement of molecules. First for larger molecules, particles or other organisms, it acts as a physical barrier. Typically when larger molecules, particles (viruses), and other organisms enter a cell, they are first engulfed by the membrane (process 1 known as endocytosis)(→).[232] A superficially similar process, running in "reverse" (process 3 known as exocytosis), is involved in moving molecules to the cell surface and releasing them into the extracellular space. Both endocytosis and exocytosis involve membrane vesicles emerging from or fusing into the plasma membrane. These processes leave the topology of the cell unaltered, in the sense that a molecule within a vesicle is still "outside" of the cell, or at least outside of the cytoplasm. These movements are driven by various molecular machines that we will consider rather briefly; they are typically considered in greater detail in subsequent courses on cell biology. We are left with the question of how molecules can enter or leave the cytoplasm, this involves passing directly through a membrane (process 2).
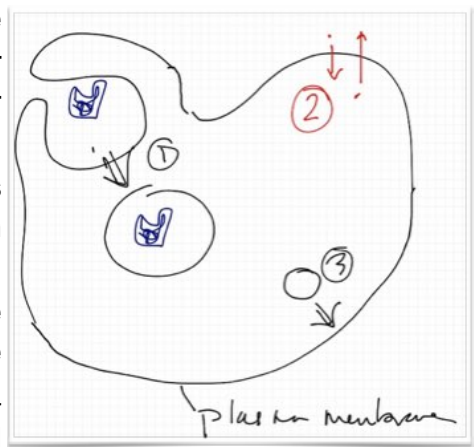
So the question is, how does the membrane "decide" which molecules to allow into and out of the cell. If we think about it, there are three possible general mechanisms (let us know if you can think of more). Molecules can move on their own through the membrane, some move passively across the membrane using specific "carriers" or "channels", while others are moved actively using some kind of "pump", an energy dependent process involving coupled reactions. Which types of carriers, channels, and pumps are present will determine what types of molecules move through the cell's membrane, as well as which directions they move, or rather the net flux of their movement. As we will see, in the vast majority of cases, these carriers, channels, and pumps are protein-based molecular machines, the

---

[231] A model of intracellular organization

[232] These processes, ranging from pinocytosis (cell drinking) to endocytosis (cell entry) and phagocytosis (cell eating) involve different molecular machines, beyond our scope here.

structure of which we will consider in greater detail later on. We can think of this molecular movement reaction generically as:

$$\text{Molecule}_{outside} \rightleftharpoons \text{Molecule}_{inside\ membrane} \rightleftharpoons \text{Molecule}_{inside\ cell}.$$
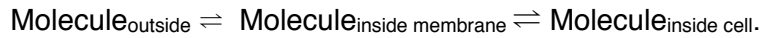
As with standard chemical reactions, movement through a membrane involves an activation energy, which amounts to the energy needed to pass through the membrane. So, you might well ask, why does the membrane, particularly the hydrophobic center of the membrane, pose a barrier to the movement of hydrophilic molecules. Here the answer involves the difference in the free energy of the moving molecule within an aq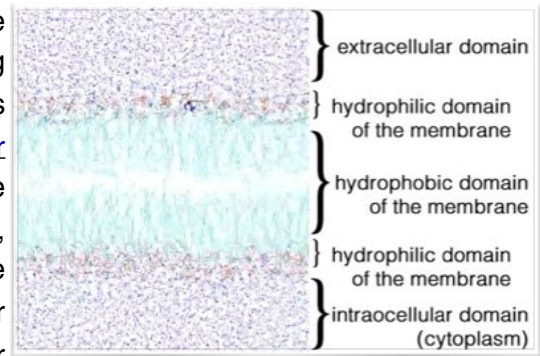ueous solution, including the hydrophilic surface region of the membrane, where H-bond type electrostatic interactions are common between molecules, and the hydrophobic/hydro-apathetic region of the membrane, where only van der Waals interactions are present. The situation is exacerbated for charged molecules, since water molecules are typically organized in a dynamic shell around each ion. We are considering molecules of one particular substance moving through the membrane and so the identity of the molecule does not change during the transport reaction. If the concentrations of the molecules are the same on both sides of the membrane, then their Gibbs free energies are also equal, the system will be in equilibrium with respect to this reaction. In this case, as in the case of chemical reactions, there will be no net flux of the molecule across the membrane, but molecules will be moving back and forth at an equal rate. The rate at which they move back and forth will depend on the size of the activation energy associated with moving across the membrane as well as the concentrations of the molecules.

To think about molecules crossing lipid membranes, let us begin with water itself, which is small and uncharged, although polarized. When a water molecule begins to leave the aqueous phase and enter the hydrophobic (central) region of the membrane, there are no H-bonds to take the place of those that are lost, no strong molecular handshakes; the result is that often the molecule is "pulled back" into the water phase ($\rightarrow$)(see video of a water molecule moving through a membrane). Nevertheless, there are so many molecules of water outside (and inside) the cell, and water molecules are so small, that once they enter the membrane, they can pass through it. The activation energy for the $\text{Water}_{outside} \rightleftharpoons \text{Water}_{inside}$ reaction is low enough that water can pass through a membrane (in both directions) at a reasonable rate.

- extracellular domain
- hydrophilic domain of the membrane
- hydrophobic domain of the membrane
- hydrophilic domain of the membrane
- intraocellular domain (cytoplasm)

Small non-polar molecules, such as O_2 and CO_2, can pass through a biological membrane relatively easily. There is more than enough energy available through collisions with other molecules (thermal motion) to provide them with the energy needed to overcome the activation energy involved in leaving the aqueous phase and passing between the molecular domains within the center of the membrane. However now we begin to see changes in the free energies of the molecules on the inside and outside of the cell. For example, in organisms that depend upon O_2 (obligate aerobes), the O_2 outside of the cell comes from the air; it is generated by plants that release O_2 as a waste product. Once O_2 enters the cell, it takes part in the reactions of respiration (we will get back to both processes further on in this chapter.) The result is that the concentration of O_2 outside the cell will be greater than the concentration of O_2 inside the cell. That means that the free energy of O_2 outside will be greater than the free energy

of O_2 inside. The reaction $O_2 \text{ outside} \rightleftharpoons O_2 \text{ inside}$ is now thermodynamically favorable and there will be a net flux of O_2 into the cell ($\rightarrow$). We can consider how a similar situation applies to water. The intracellular domain of a cell is a concentrated solution of proteins and other molecules. Typically, the concentration of water outside of the cell is greater than the concentration of water inside the cell. Our first order presumption is that the reaction: $H_2O \text{ outside} \rightleftharpoons H_2O \text{ inside}$ is favorable, so water will flow into the cell. The obvious question is, what happens over time? We will return to how cells (and organisms) answer this question shortly.
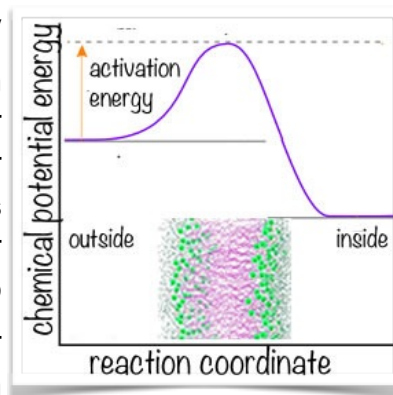
Instead of reactants and products we can plot the position of a molecule relative to the membrane. If a molecule is hydrophobic (non-polar) it will be more soluble in the membrane's central hydrophobic environment than it in the surrounding aqueous environment ($\rightarrow$)(video link). In contrast the situation will be distinctly different for hydrophilic molecules. By this point, we hope you will recognize that in a biologically unrealistic lipid-only membrane, the shape of this graph, and specifically the height of the activation energy peak will vary depending upon the characteristics of the molecule we are considering moving as well as the membrane itself. A totally hydrophobic molecule will accumulate within the membrane, and an activation energy would be associated with leaving the hydrophobic membrane, rather than passing through it.
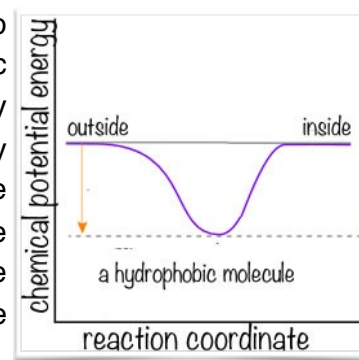
Questions to answer:
87. Consider the reaction diagram for flipping a lipid molecule's orientation by 180° perpendicular to the plane of the membrane: what energy barriers are associated with such a movement?
88. Draw a graph to show how the potential energy changes as an ion moves across a membrane.
89. What do you expect to happen to the O_2 gradient if an aerobic cell's ability to use O_2 is inhibited?

Channels and carriers

Beginning around the turn of the last century, a number of scientists began working to define the nature of the cellular boundary layer. In the 1930's it was noted that small, water soluble molecules entered cells faster than predicted based on the assumption that the membrane acts like a simple hydrophobic barrier. Collander et al., postulated that membranes were more than simple hydrophobic barriers, specifically that they contained features that enabled them to act as highly selective molecular sieves.[233] Most of these features are proteins (we are getting closer to a discussion of proteins) that can act as channels, carriers, and pores. If we think about crossing the membrane as a reaction, then the activation energy of this reaction can be quite high for highly hydrophilic and larger molecules, we will need a catalyst to reduce the activation energy so that the reaction can proceed at a reasonable rate. There are two generic types of membrane permeability catalysts: carriers and channels.

---

[233] Does Overton still rule? http://www.nature.com/ncb/journal/v1/n8/full/ncb1299_E201.html

Carrier proteins are membrane proteins that shuttle back and forth across the membrane. They bind to specific hydrophilic molecules when they are located in the hydrophilic region of the membrane, hold on to the bound molecule as they traverse the membrane's hydrophobic region, and then release their "cargo" when they again reach a hydrophilic region of the membrane. Both the m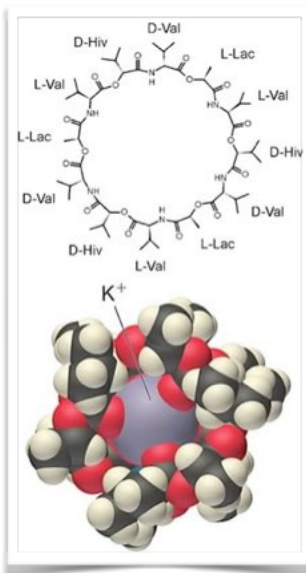ovements of carrier and cargo across the membrane, and the release of transported molecules, are stochastic and are driven by thermal motion (collisions with other molecules), so no other energy source is needed. We can write this class of reactions as:

$Molecule_{outside} + carrier_{empty} \rightleftharpoons carrier- Molecule_{outside} \rightleftharpoons carrier- Molecule_{inside} \rightleftharpoons Molecule_{inside} + carrier_{empty}$.
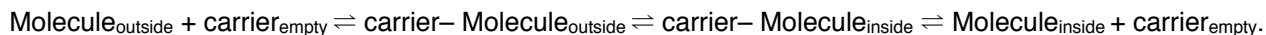
There are many different types of carrier molecules and each type of carrier has preferred cargo. Related molecules may be bound and transported, but with much less specificity and so at a much lower rate. Exactly which molecules a particular cell will allow to enter will be determined in part by which carrier protein genes it expresses. Mutations in a gene encoding a carrier can change (or abolish) the range of molecules that that carrier can transport across a membrane.

Non-protein carriers: An example of a membrane carrier is a class of antibiotics, known generically as ionophores, that carry ions across membranes. They kill cells by disrupting the normal ion balance across the cell's membrane and within the cytoplasm, which in turn disrupts normal metabolic activity.[234] One of these ionophore antibiotics is valinomycin ($\rightarrow$), a molecule made by *Streptomyces* type bacteria.[235] The valinomycin molecule has a hydrophobic periphery and a hydrophilic core. It binds K^+ ions ~10^5 times more effectively than it binds Na^+ ions. Together with the bound ion, the valinomycin molecule continually shuttles back and forth across the membrane. In the presence of a K^+ gradient, that is a higher concentration of K^+ on one side of the membrane compared to the other, K^+ will tend to bind to the valinomycin molecule, whereas on the side where $[K^+]$ is low, the K^+– valinomycin complex will dissociate (in response to collisions with other molecules), breaking the valinomycin–K^+ interaction, releasing K^+ into the cell.

Where there is a K^+ concentration gradient, the presence of valinomycin will produce a net flux of K^+ from the high to the low concentration sides of membrane, reducing and eventually eliminating the K^+ gradient. In the absence of specific K^+ channels and pumps, K^+ cannot pass through the membrane, the activation energy is too high. Again, to be clear, in the absence of a gradient, K^+ ions will move across the membrane (in the presence of the carrier), but there will be no net change in the concentration of K^+ ion inside the cell, no net flux. For the experimentally inclined, you might consider how you could prove that movements are occurring even in the absence of a gradient. In a similar manner, there are analogous carrier systems that move hydrophobic molecules through water.

Channel molecules sit within a membrane and contain an aqueous channel that spans the membrane's hydrophobic region. Hydrophilic molecules of particular sizes and shapes can pass

---

[234] There is little data in the literature on exactly which cellular processes are disrupted by which ionophore; in mammalian cells (as we will see) these molecules are by disrupting ion gradients in mitochondria and chloroplasts, apparently.

[235] Valinomycin: https://en.wikipedia.org/wiki/Valinomycin

through this aqueous channel and their movement involves a significantly lower activation energy than would be associated with moving through the lipid part of the membrane in the absence of the channel. Channels are generally highly selective in terms of which molecules will pass through them. For example, there are channels which will, on average, pass 10,000 K^+ ions for every one Na^+ ion.

Often the properties of these channels can be regulated, including through the binding of small molecules to the protein; they can exist in two or more distinct structural states. For example, in one state the channel can be open and allow particles to pass through or it can be closed, that is the channel can be turned on and off. Channels cannot, however, determine in which direction an ion will move - that is based on the gradients across the membrane.

Another method of channel control depends on the fact that channel proteins are embedded within a membrane and contain charged groups. As we will see cells can (and generally do) generate ion gradients, that, is a separation of charged species across their membranes. For example if the concentration of K^+ is higher on one side of the membrane, there will be an ion gradient where the ions will (if movement is possible) move from the region of higher to lower K^+ concentration.[236] In some cases, the generation of an ion gradient can, in turn, produce an electrical field across the plasma membrane. As these fields change, they can produce (induce) changes in channel structure that can switch the channel from open to closed and vice versa. Organisms typically have many genes that encode specific channel proteins that are involved in a range of processes from muscle contraction to thinking. As in the case of carriers, channels do not determine the direction of molecular motion. The net flux of movement is determined by the presence of molecular gradients, with the thermodynamic driver being entropic factors. That said, the actual movement of the molecules through the channel is driven by thermal motion.

Questions to answer:
90. What does it mean to move up (against) a concentration gradient? Is this a favorable or unfavorable event?
91. Where does the energy involved in moving molecules come from?
92. What happens to the movement of molecules through channels and transporters if we reverse the concentration gradients across a membrane?
93. Draw a diagram to show how K^+ ions are transported by an ionophore across a membrane. Draw a graph to show how the potential energy changes as the ion moves. Be sure to include the relative concentrations.

Generating gradients: using coupled reactions and pumps

Both carriers and channels allow the directional movement of molecules across a membrane, but there is a net directional flux only when a concentration gradient is present - that is if the concentration of the molecule is different on each side of the membrane. If a membrane contains active channels and carriers (as all biological membranes do), without the input of energy eventually concentration gradients across the membrane will disappear (disperse). The $[molecule]_{outside}$ will become equal to $[molecule]_{inside}$. Removing a concentration gradient across of cell's plasma membrane is a good way to kill the cell. When we look at cells we find lots of concentration gradients, which raises the question, what produces and then maintains these gradients.

---

[236] In fact this tendency for species to move from high to low concentration until the two concentrations are equal can be explained by the Second Law of Thermodynamics. Check with your chemistry instructor for more details

The common sense (or rather thermodynamically correct) answer is that there must be molecules (generally proteins) that can transport specific types of molecules across the membrane and against their concentration gradient. We will call these types of molecules pumps and write the reaction they are involved in as:

$$[Molecule]_{low\ concentration} + pump \longleftrightarrow [Molecule]_{high\ concentration} + pump$$
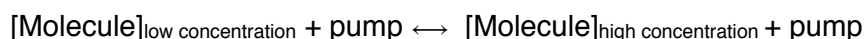
As you might suspect moving this reaction to the right is thermodynamically unfavorable; like a familiar macroscopic pump, it will require the input of energy to work. We will have to "plug in" our molecular pump into some source of energy to move a molecule against its concentration gradient. So, what energy sources are available to biological systems? Basically we have two choices: the system can use electromagnetic energy (light) or it can use chemical energy. In a light-driven pump, there is a system that captures (absorbs) light; the absorbance of light (energy) is coupled to the pumping system ($\rightarrow$). Where the pump is driven by a chemical reaction, a thermodynamically favorable reaction is often catalyzed by the pump, which also acts to facilitate the movement of one or more molecules against their membrane-associated concentration gradients.
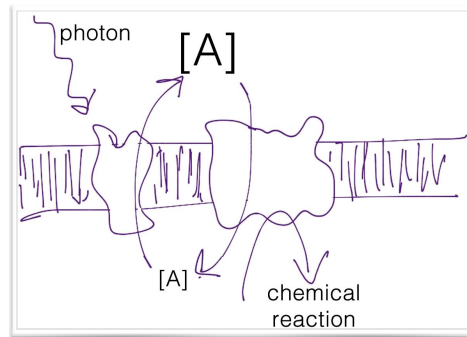
A number of chemical reactions can be used to drive such pumps and these pumps can drive various reactions (remember reactions can move in both directions). One of the most common reactions involves the movement of energetic electrons through a membrane-bound, protein-based "electron transport" system; this, in turn, leads to the creation of an H^+ based electrochemical gradient. The thermodynamically favorable movement of H^+ down such a concentration gradient is coupled to a reaction that leads to the synthesis of adenosine triphosphate (ATP) through reactions catalyzed by the membrane-bound ATP synthase enzyme:

$$H^+_{(extracellular)} \rightleftharpoons H^+_{(intracellular)}$$
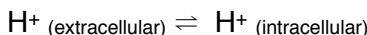
ATP synthase (membrane-localized catalyst)

$$H^+ + adenosine\ diphosphate\ (ADP) + phosphate \rightleftharpoons adenosine\ triphosphate\ (ATP) + H_2O$$
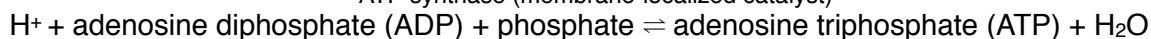
The reaction takes cytoplasmic ADP, phosphate and H^+ and releases ATP and water into the cytoplasm. The thermodynamically favorable movement of H^+ down its concentration gradient is coupled to the the thermodynamically unfavorable ATP synthesis reaction. The reaction can run in reverse, so that the thermodynamically favorable ATP hydrolysis reaction:

$$ATP + H_2O \rightleftharpoons ADP + phosphate + H^+$$

ATPase-driven pump (ATP synthase running backward)

$$H^+_{(intracellular)} \rightleftharpoons H^+_{(extracellular)}$$
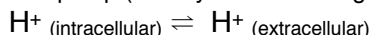
a reaction that results in the generation of a H^+ gradient across the membrane. So, find that the same membrane molecule, the ATP synthase/pump, makes it possible to use energy present in a chemical gradient (across a membrane) to drive ATP synthesis within the cell and it can enable ATP hydrolysis to generate a concentration gradient.

Simple Phototrophs

Phototrophs are organisms that capture particles of light (photons) and transform their electromagnetic energy into energy stored in unstable molecules, such as ATP and carbohydrates.

Phototrophs "eat" light. Light can be considered as both a wave and a particle (that is quantum physics for you) and the wavelength of a photon determines its color and the amount of energy it contains. Again, because of quantum mechanical considerations, a particular molecule can only absorb photons of specific wavelengths (energies). This property enables us to identify molecules at great distances based on the photons they absorb or emit. This idea is the basis of spectroscopy. Our atmosphere allows mainly visible light from the sun to reach the earth's surface, but most biological molecules do not absorb visible light very effectively if at all. To capture this energy, organisms have evolved the ability to synthesize molecules, known as pigments, that can capture (absorb) visible light, that organisms can then use. The colors we see for a typical pigment are the colors of the light that it does not absorb but rather that it reflects. For example chlorophyl appears green because light in the red and blue regions of the spectrum is absorbed and green light is reflected. The question we need to answer is, how does the organism use the electromagnetic energy that is absorbed?

One of the simplest examples of a phototrophic system, that is, a system that directly captures the energy of light and transforms it into the energy stored in a chemical system, is provided by the archaea *Halobacterium halobium*.[237] *Halobacteria* are extreme halophiles (salt-loving) organisms. They live in waters that contain up to 5M NaCl. *H. halobium* uses the membrane protein bacteriorhodopsin to capture light. Bacteriorhodopsin consists of two components, a polypeptide, known generically as an opsin, and a non-polypeptide prosthetic group, the pigment retinal, a molecule derived from vitamin A.[238] Together the two, opsin + retinal, form the functional bacteriorhodopsin protein.

Because its electrons are located in extended molecular orbitals with energy gaps between them that are of the same order as the energy of visible light, absorbing a photon of visible light moves an electron from a lower to a higher energy molecular orbital. Such extended molecular orbitals (highlighted in the figure →) are associated with molecular regions that are often drawn as containing alternating single and double bonds between carbons; these are known as conjugated π orbital systems. Conjugated π systems are responsible for the absorption of light by pigments such as chlorophyll and heme (the pigment that makes blood red). When a photon of light is absorbed by the retinal group, it undergoes a reaction that leads to a change in the pigment molecule's shape and composition, which in turn leads to a change in the structure of the polypeptide to which the retinal group is attached. This is called a photoisomerization reaction.
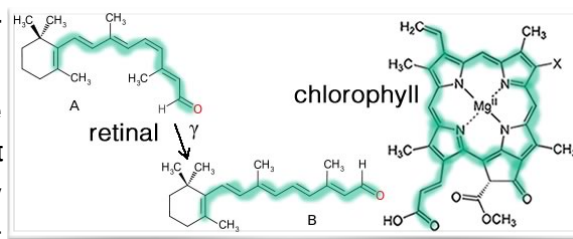
The bacteriorhodopsin protein is embedded within the plasma membrane where it associates with other bacteriorhodopsin proteins to form protein patches. These patches of membrane protein give the organisms their purple color and are known as purple membrane. When one of these bacteriorhodopsin proteins absorbs light, the change in the associated retinal group produces a light-induced change in protein structure that results in the movement of a H^+ ion from the inside to the outside of the cell (→).

---

[237] Gradients and reactions (short video)

[238] As we will return to later, proteins are functional entities, composed of polypeptides and prosthetic group. The prosthetic group is essential for normal protein function. The protein without the prosthetic group is known as the apoprotein.

The protein and its associate pigment then return to its original low energy (ground) state, that is, its state before it absorbed the photon of light. The return of bacteriorhodopsin to the ground state is NOT associated with the movement of a H^+ ion across the membrane. Because all of the bacteriorhodopsin molecules in the membrane are oriented with the same orientation, as light is absorbed all of the H^+ ions move in the same direction across the membrane, leading to the formation of a H^+ concentration gradient with $[H^+]_{outside} > [H^+]_{inside}$. This H^+ gradient is also associated with an electrical gradient because the movement of H^+ leads to more positive charge outside the cell. As light is absorbed the concentration of H^+ outside the cell increases and the concentration of H^+ inside the cell decreases. The question is – where are the moving H^+'s coming from? As you (perhaps) learned in chemistry, water undergoes a dissociation reaction (although this reaction is quite unfavorable):

$$H_2O \rightleftharpoons H^+ + OH^-$$

At pH, 7.0 water contains 10^{-7} moles of H^+ and it is these H^+ s that move.

As H^+s move across the membrane, they leave behind OH^- ions. The result is that the light driven movement of H^+ ions produces an electrical field, with excess + charges outside and excess – charges inside. As you know from your physics, positive and negative charges attract, but the intervening membrane stops them from reuniting. The result is the accumulation of positive charges on the outer surface of the membrane and negative charges on the inner surface. This charge separation produces an electric field across the membrane. Now, an H^+ ion outside of the cell will experience two distinct forces, those associated with the electric field and those arising from the concentration gradient. If there is a way across the membrane, the $[H^+]$ gradient will lead to the movement of H^+ ions back into the cell. Similarly the electrical field will also drive the positively charged H^+ back into the cell. The formation of the $[H^+]$ gradient basically generates a battery, a source of energy that the cell can use.

So how does the cell tap into this battery? The answer is through a second membrane protein, an enzyme known as the H^+–driven ATP synthase. H^+ ions move through the ATP synthase molecule in a thermodynamically favorable sequence of reactions. The ATP synthase couples this favorable movement to an unfavorable chemical reaction, a condensation reaction:

ATP synthase
$$H^+_{outside} + ADP + \text{inorganic phosphate (Pi)} + H^+ \rightleftharpoons ATP + H_2O + H^+_{inside}$$
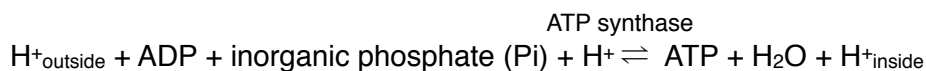
ATPase pump (ATP synthase running backward)

This reaction continues as long as light is absorbed and for a short time afterward. In the light, bacteriorhodopsin acts to generate a H^+ gradient. When the light goes off (that is, at night time) the H^+ gradient persists until H^+ ions have moved through the ATP synthase. ATP synthesis continues until the H^+ gradient no longer has the energy sufficient to drive the ATP synthesis reaction. The net result is that the cell uses light to generate ATP, which is stored for later use. ATP acts as a type of chemical battery, in contrast to the electrochemical battery of the H^+ gradient.
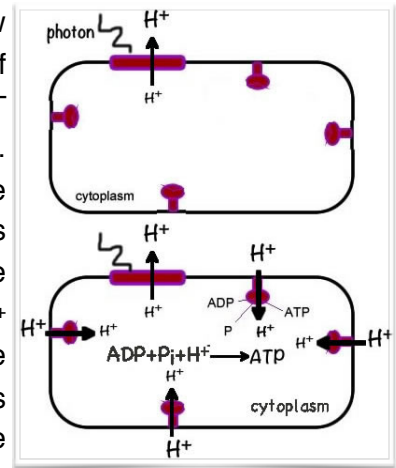
An interesting feature of the ATP synthase molecule (→) is that the H^+ ions move through it by hopping from one acidic amino acid to another in a thermodynamically favored sequence (video link). As the protons move, they change the interactions between parts of the ATP synthase, causing changes in shape, which in turn causes a region of the molecule to rotate. It rotates in one direction when it drives the synthesis of ATP; it rotates in the opposite direction to couple ATP hydrolysis to the pumping of H^+ ions against their concentration gradient. In this form it is better called an ATPase (or hydrolase) pump, involving the thermodynamically favorable reaction:

<div align="center">

ATPase pump

$ATP + H_2O + H^+_{inside} \rightleftharpoons H^+_{outside} + ADP + \text{inorganic phosphate (Pi)} + H^+$
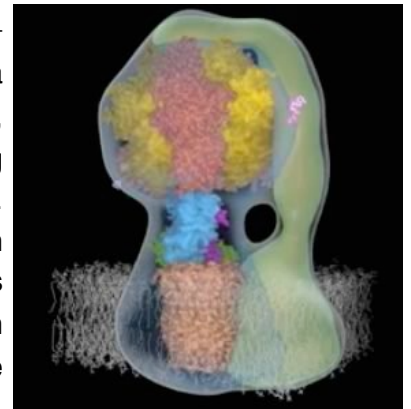
ATP synthase (ATPase pump running backward)
</div>

Because the enzyme rotates when it hydrolyzes ATP, it is rather easy to imagine how the energy released through this reaction could be coupled, through the use of an attached paddle-like extension, to cellular or fluid movement.

Questions to answer

94. Indicate in a diagram the direction of H^+ movement in a phototroph when exposed to light.
95. Why does the H^+ gradient across the membrane dissipate when the light goes off? What happens to the rate of ATP production? When does ATP production stop and why?
96. What limits the "size" of the H^+ gradient that bacteriorhodopsin can produce?
97. What is photoisomerization? Is this a reversible or an irreversible reaction?

Questions to ponder

- How might ATP hydrolysis lead to cell movement.
- What would happen if bacteriorhodopsin molecules were oriented randomly within the membrane

Chemo-osmosis (an overview)

One of the most surprising discoveries in biology was the wide spread, almost universal use of H^+–based electrochemical gradients to generate ATP. What was originally known as the chemiosmotic hypothesis was produced by the eccentric British scientist, Peter Mitchell (1920–1992).[239] Before the significance of H^+ membrane gradients was widely appreciated, Mitchell proposed that energy captured through the absorption of light (by phototrophs) or the breakdown of molecules into more stable molecules (by various types of chemotrophs) relied on the same basic (homologous, that is, evolutionarily-related) mechanism, namely the generation of H^+ gradients across membranes (the plasma membrane in prokaryotes and the internal membranes of mitochondria or chloroplasts (intracellular organelles, derived from bacteria – see below) in eukaryotes.

What makes us think that these processes might have a similar evolutionary root, that they are homologous? Basically, it is the observation that in both light- and chemical-based processes captured energy is transferred through the movement of electrons through a membrane-embedded "electron transport chain". An electron transport chain involves a series of membrane and associated proteins

---

[239] Chemo-osmosis and Peter Mitchell (wikipedia)

and a series of reduction-oxidation or redox reactions (see below) during which electrons move from a high energy donor to a lower energy acceptor. Some of the energy difference between the two is used to move H^+ ions across a membrane, generating a H^+ concentration gradient. Subsequently the thermodynamically favorable movement of H^+ down this concentration gradient (across the membrane) is used to drive ATP synthesis, a thermodynamically unfavorable reactions. ATP synthesis itself involves the rotating ATP synthase. The reaction can be written:

$$H^+_{outside} + ADP + Pi + H^+ \rightleftharpoons ATP + H_2O + H^+_{inside,}$$
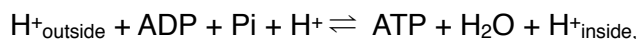
where "inside" and "outside" refer to compartments defined by the membrane containing the electron transport chain and the ATP synthase, with the ATP synthesis reaction occurring within the membrane-bound compartment. Again, this reaction can run backwards. When this occurs, the ATP synthase acts as an ATPase (ATP hydrolase) that can pump H^+ (or other molecules) against their concentration gradient. Such pumping ATPases establishes most biologically important ion gradients across membranes. In such a reaction:

ATP + H_2O + molecule in low concentration region $\rightleftharpoons$ ADP + Pi + molecule in high concentration region.

The most important difference between phototrophs and chemotrophs is, essentially, where do the high energy electrons come from - energized by absorption of light, or derived from unstable molecules.

Oxygenic photosynthesis

Compared to the salt loving archaea *Halobium* with its purple bacteriorhodopin-rich membranes, photosynthetic cyanobacteria (which are true bacteria), green algae, and higher plants (both eukaryotes) use more complex molecular systems through which to capture and utilize light. The photosynthetic systems of these organisms appear to be homologous, that is, derived from a common ancestor, a topic we will return to soon. For simplicity's sake we will describe the photosynthetic system of cyanobacterium; the system in eukaryotic algae and plants, while more complex, follows the same basic logic. At this point, we consider only one aspect of this photosynthetic system, known as the oxygenic or non-cyclic system (look to more advanced classes for more details.) The major pigment in this system, chlorophyll, is based on a complex molecule, a porphyrin (see above) and it is primarily these pigments that give plants their green color. As in the case of retinal, they absorb visible light due to the presence of a conjugated bonding structure (drawn as a series of alternating single and double) carbon-carbon bonds. Chlorophyll is synthesized by a conserved biosynthetic pathway, variants of this scheme are used to synthesize heme, which is found in the hemoglobin of animals and in the cytochromes, within the electron transport chain present in both plants and animals (which we will come to shortly), vitamin B_{12}, and other biologically important prosthetic (that is non-polypeptide) groups associated with proteins and required for their normal function.[240]

Chlorophyll molecules are organized into two distinct protein complexes that are embedded in membranes. These are known as the light harvesting and reaction center complexes. Light harvesting complexes (lhc) act as antennas to increase the amount of light the organism can capture. When a photon is absorbed, an electron is excited to a higher molecular orbital. An excited electron can be

_____

[240] Mosaic Origin of the Heme Biosynthesis Pathway in Photosynthetic Eukaryotes:

passed between components of the lhc and eventually to the reaction center ("rc") complex ($\rightarrow$). Light harvesting complexes are important because photosynthetic organisms often compete with one another for light; increasing the efficiency of the system through which an organism captures light can provide a selective (evolutionary) advantage.
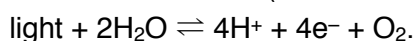

In the oxygenic, that is molecular oxygen (O_2) generating photosynthesis reaction system, high energy (excited) electrons are passed from the reaction center through a set of membrane proteins, the electron transport chain ("etc"). As an excited electron moves through the electron transport chain its energy is used to move H^+s from inside to outside of the cell. This is the same geometry of movement that we saw previously in the case of the purple membrane system. The end result is the generation of a H^+ based electrochemical gradient. As with purple bacteria, the energy stored in this H^+ gradient is used to drive the synthesis of ATP within the cell's cytoplasm, a coupled reaction catalyzed by the ATP synthase.
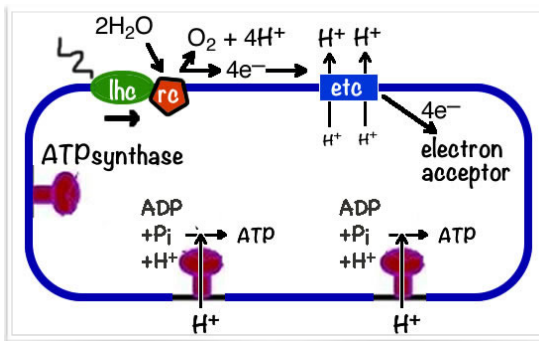
Now you might wonder, what happens to the originally excited electrons, and the energy that they carry. In what is known as the cyclic form of photosynthesis, low energy electrons from the electron transport chain are returned to the reaction center, where they return the pigments to their original (before they absorbed a photon) state. In contrast, in the non-cyclic process that we have been considering, electrons from the electron transport chain are delivered to an electron acceptor. Generally this involves the absorption of a second photon, a mechanistic detail that need not trouble us here. This is a general type of chemical reaction known as an reduction-oxidation (redox) reaction. Where an electron is within a molecule's electron orbital system influences the amount of energy present in the molecule: adding a negative charge (an electron) to a molecule can increase electron-electron repulsion and raise the molecule's its potential energy. When an electron is added to a molecule, that molecule is said to have been "reduced", and yes, it does seem weird that adding an electron "reduces" a molecule ($\rightarrow$). Generally, when an electron is removed, the molecule's energy is changed (decreased) and the molecule is said to have been "oxidized".[241] Since electrons, like energy, are neither created nor destroyed in biological systems, so the reduction of one molecule is

always coupled to the oxidation of another. In a system of redox reactions, electrons are removed from the reduced molecule are used to drive various types of thermodynamically unfavorable reactions, including the movement of H^+ across a membrane.
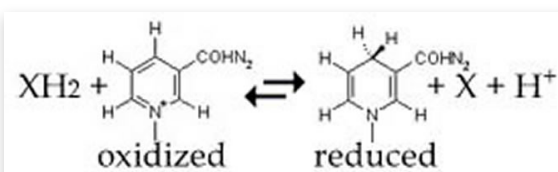
Again, the laws of conservation imply that when electrons leave the photosynthetic system (in the non-cyclic process) they must be replaced. So where might these electrons be coming from? Here we see what appears to be a major evolutionary breakthrough. During the photosynthetic process, the reaction center couples light absorption to the oxidation (removal of electrons) from water molecules:

$$\text{light} + 2H_2O \rightleftharpoons 4H^+ + 4e^- + O_2.$$

---

[241] you can review redox here: http://www.biologie.uni-hamburg.de/b-online/e18/18b.htm or in CLUE: http://tinyurl.com/ze5yqlx

The four electrons, derived from two molecules of water, pass to the reaction center, while the $4H^+$s contribute to the proton gradient across the membrane.[242] O_2 is a waste product of this reaction. Over millions of years, the photosynthetic release of O_2 changed the Earth's atmosphere from containing essentially 0% molecular oxygen to the current ~21% level at sea level. Because O_2 is highly reactive, this transformation is thought to have been a major driver of subsequent evolutionary change. However, there remain organisms that cannot use O_2 and cannot survive in its presence. They are known as obligate anaerobes, to distinguish them from organisms that normally grow in the absence of O_2 but that can survive in the presence of O_2, which are known as facultative anaerobes. In the past the level of atmospheric O_2 has changed dramatically; its level is based on how much O_2 is released into the atmosphere by oxygenic photosynthesis and how much is removed by various reactions, such as the decomposition of plant materials. When large amounts of plant materials are buried before they can decay, such as occurred with the formation of coal beds during the Carboniferous period, from ~360 to 299 million years ago, the level of atmospheric O_2 increased dramatically, up to an estimated ~35%. It is speculated that such high levels of atmospheric molecular oxygen made it possible for organisms without lungs (like insects) to grow to gigantic sizes.[243]

Chemotrophs

Organisms that are not phototrophic capture energy from other sources, specifically by transforming thermodynamically unstable molecules into more stable species. Such organisms are known generically as chemotrophs. They can be divided into various groups, depending upon the types of food molecules (energy sources) they use: these include organotrophs, which use carbon-containing molecules (you yourself are an organotroph) and lithotrophs or rock eaters, which use various inorganic molecules. In the case of organisms that can "eat" H_2, the electrons that result are delivered, along with accompanying H^+ ions, to CO_2 to form methane (CH_4) following the reaction:

$$CO_2 + 4H_2 \rightleftharpoons CH_4 + 2H_2O.$$
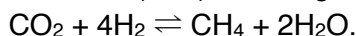
Such organisms are referred to as methanogens (methane-producers).[244] In the modern world methanogens (typically archaea) are found in environments with low levels of O_2, such as your gut. In many cases reactions of this type can occur only in the absence of O_2. In fact O_2 is so reactive, that it can be thought of as a poison for organisms that cannot actively "detoxify" it. When we think about the origins and subsequent evolution of life, we have to consider how organisms that originally arose in the absence of molecular O_2 adapted as significant levels of O_2 began to appear in their environment. It might be that modern obligate anaerobes might still have features common to the earliest organisms.

The amount of energy that an organism can capture is determined by the energy of the electrons that the electron acceptor(s) they employ can accept. If only electrons with high amounts of energy can be captured, which is often the case, then inevitably large amounts of energy are left behind, with the acceptor. On the other hand, the lower the amount of energy that an electron acceptor can accept, the more energy can be extracted and captured from the original "food" molecules and the less energy is

---

[242] Photosystem II and photosynthetic oxidation of water: an overview: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC1693055/

[243] When Giants Had Wings and 6 Legs: http://www.nytimes.com/2004/02/03/science/when-giants-had-wings-and-6-legs.html

[244] Lithotrophic (wikipedia)

left behind. Molecular oxygen is unique in its ability to accept low energy electrons ($\rightarrow$). For example, consider an organotroph that eats carbohydrates (molecules of the general composition $[C_6H_{10}O_5]_n$), a class of molecules that includes sugars, starches, and wood, through a process known as glycolysis, from the Greek words meaning sweet (glyco) and splitting (lysis). In the absence of O_2, that is under anaerobic conditions, the end product of the breakdown of a carbohydrate leaves ~94% of the theoretical amount of energy present in the original carbohydrate molecule in molecules that cannot be broken down further, at least by most organisms. These are molecules such as ethanol (C_2H_6O) and lactic acid ($CH_3CH(OH)CO_2H$). However, when O_2 is present, carbohydrates can be broken down more completely into CO_2 and H_2O, a process known as respiration. In such O_2 using (aerobic) organisms, the energy released by the formation of CO_2 and H_2O is transferred to (stored in) energetic electrons and used to generate a membrane-associated H^+ based electrochemical gradient that in turn drives ATP synthesis, through a membrane-based ATP synthase. In an environment that contains molecular oxygen, organisms that can use O_2 as an electron acceptor have a distinct advantage; instead of secreting energy rich molecules, like ethanol, they release the energy poor (stable) molecules CO_2 and H_2O.
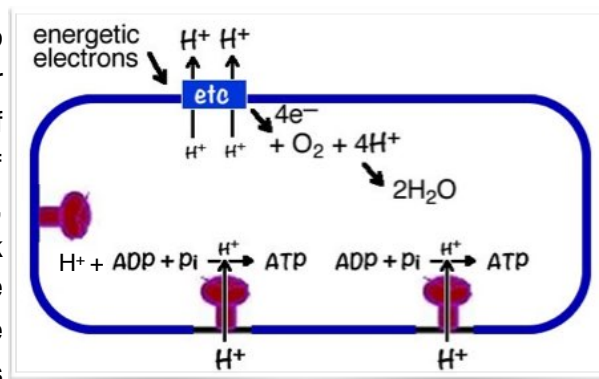
No matter how cells (and organisms) capture energy, to maintain themselves and to grow, they must make a wide array of various complex molecules. Understanding how these molecules are synthesized lies (traditionally) within the purview of biochemistry. That said, in each case, thermodynamically unstable molecules (like lipids, proteins, and nucleic acids) are built through series of coupled reactions that rely on energy capture from light or the break down of food molecules.

Questions to answer

98. How (do you suppose) does an electron move through an electron transport chain? Make a diagram and a graph that describes its energy as it moves through the chain.
99. In non-cyclic photosynthesis, where do electrons end up?
100. What would happen to an aerobic cell's ability to make ATP if it where exposed to an H^+ carrier or channel?
101. Why are oxidation and reduction always coupled?
102. Why are carbohydrates good for storing energy?

Questions to ponder

- Which do you think would have a greater evolutionary advantage, an organism growing aerobically or anaerobically? What factors would influence your answer?

Using the energy stored in membrane gradients

The energy captured by organisms is used to drive a number of processes in addition to synthesis reactions. For example, we have already seen that ATP synthases can act as pumps (ATP-driven transporters), coupling the favorable ATP hydrolysis reaction to the movement of molecules against their concentration gradients ($\downarrow$). The resulting gradient is a form of stored (potential) energy, energy that can be used to move other molecules, that is molecules that are not moved directly by a ATP-

driven transporter.[245] Such processes involve what is known as coupled transport.[246] They rely on membrane-bound proteins that enable a molecule to pass through a membrane, and so allow for a net flux down a concentration gradient. In contrast to simple carriers and channels, however, this thermodynamically favorable net flux down, that is, from high concentration to low concentration, is physically coupled to the movement of a second net flux against a gradient, that is from low to high concentration. When the two transported molecules move in the same direction, the transporter is known as a symporter; when they move in opposite directions, it is known as an antiporter. Which direction(s) the molecules move will be determined by the nature of the transporter and the relative sizes of the concentration gradients of the two types of molecules moved. There is no inherent directionality associated with the transporter itself - the net movement of molecules reflects the relative concentration gradients of the molecules that the transporter can productively bind. What is important here is that energy stored in the concentration gradient of one molecule can be used to drive the movement of a second type of molecule against its concentration gradient. In mammalian systems, it is common to have Na^+, K^+, and Ca^{2+} gradients across the plasma membrane, and these are used to transport molecules into and out of cells. Of course, the presence of these gradients implies that there are ion-specific pumps that couple an energetically favorable reaction, typically ATP hydrolysis, to an energetically unfavorable reaction, the movement of an ion against its concentration gradient. Without these pumps, and the chemical reactions that drive them, the membrane battery would quickly run down. Many of the immediate effects of death are due to the loss of membrane gradients and much of the energy needs of cells (and organisms) involves running such pumps.

Osmosis and living with and without a cell wall

Cells are packed full of molecules. These molecules take up space, space no longer occupied by water. The concentration of water outside of the cell $[H_2O]_{out}$ will necessarily be higher than the concentration of water inside the cell $[H_2O]_{in}$. This solvent concentration gradient leads to the net movement of water into the cells[247]. Such a movement of solvent is known generically as osmosis. Much of this movement occurs through the membrane, which is somewhat permeable to water (see above). A surprising finding, which won Peter Agre a share of the 2003 Noble prize in chemistry, was that the membrane also contains water channels, known as aquaporins.[248] Follow the video link (↓) to a molecular simulation of a water molecule (yellow) moving across a membrane, through an aquaporin

---

[245] Although we will not consider it hear, membrane gradients are also used to send signals throughout the nervous system.

[246] Structural features of the uniporter/symporter/antiporter superfamily

[247] One important note here is that if you learn about osmosis in chemistry classes you will almost certainly be taught that water moves from a region of low SOLUTE concentration to a region of high SOLUTE concentration. These two definitions mean the same thing but it is easy to get confused.

[248] Water Homeostasis: Evolutionary Medicine: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3540612/

protein. It turns out that the rate of osmotic movement of water is dramatically reduced in the absence of aquaporins. In addition to water, aquaporin-type proteins can facilitate the movement of other small uncharged molecules across cellular membranes.
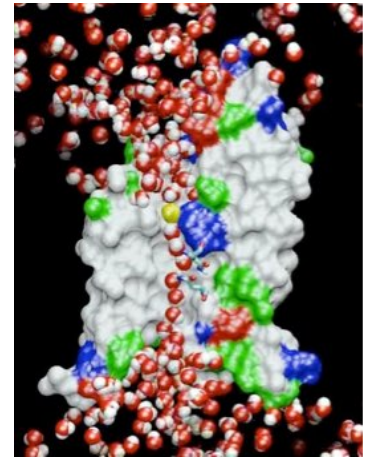
The difference or gradient in the concentrations of water across the cell membrane, together with the presence of aquaporins, leads to a system that is capable of doing work. The water gradient, can lift a fraction of the solution against the force of gravity, something involved in how plants stand up straight.[249] How is this possible? If we think of a particular molecule in solution, it will move around through collisions with its neighbors. These collisions drive the movement of particles randomly. But if there is a higher concentration of molecules on one side of a membrane compared to the other, then the random movement of molecules will lead to a net flux of molecules from the area of high concentration to that of low concentration, even though each molecule on its own moves randomly (or rather stochastically), that is, without a preferred direction [this video [250] is good at illustrating this behavior]. At steady state in a biological systems, the force generated by the net flux of water moving down its concentration gradient is balanced by forces acting in the other direction.

The water concentration gradient across the plasma membrane of most organisms leads to an influx of water into the cell. As water enters, the plasma membrane expands; you might want to think about how that occurs, in terms of membrane structure. If the influx of water continued unopposed, the membrane would eventually burst like an over-inflated balloon, killing the cell. One strategy to avoid this lethal outcome, adopted by a range of organisms, is to build a semi-rigid "cell wall" exterior to the plasma membrane (→). The synthesis of this cell wall is based on the controlled assembly of macromolecules secreted by the cell through the process of exocytosis (see above). As water passes through the plasma membrane and into the ce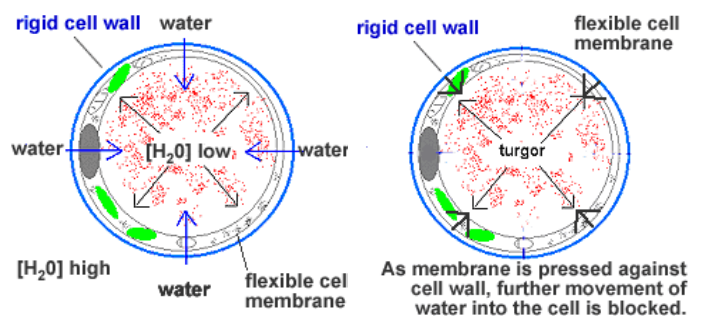ll (driven by osmosis), the plasma membrane is pressed up against the cell wall. The force exerted by the rigid cell wall on the membrane balances the force of water entering the cell. When the two forces are equal, the net influx of water into the cell stops. Conversely, if the $[H_2O]_{outside}$ decreases, this pressure is reduced, the membrane moves away from the cell wall and, because they are only semi-rigid, the walls flex. It is this behavior that causes plants to wilt when they do not get enough water. These are passive behaviors, based on the structure of the cell wall; they are built into the wall as it is assembled. Once the cell wall has been built, a cell with a cell wall does not need to expend energy to resist osmotic effects. Plants, fungi, bacteria and archaea all have cell walls. A number of antibiotics work by disrupting the assembly of bacterial cell walls. This leaves the bacteria osmotically sensitive, water enters these cells until they burst and die.

---

[249] Interested? check out the "water" virtual lab

[250] Water permeation through phospholipid membrane: http://youtu.be/ePGqRaQiBfc

Questions to answer:
103. Make a graph of the water concentration across a typical cellular membrane for an organism living in fresh water; explain what factors influenced your drawing.
104. How might cell wall-less organisms deal with challenges associated with the loss of a cell wall?
105. Plants and animals are both eukaryotes; how would you decide whether the common ancestor of the eukaryotes had a cell wall.
106. What are potential evolutionary benefits of losing a cell wall?
107. There is a concentration gradient of A across of membrane, but no net flux – what can we conclude?

Questions to to ponder:
- Why might an aquaporin channel not allow a Na+ ion to pass through it?

An evolutionary scenario for the origin of eukaryotic cells

When we think about how life arose, and what the first organisms looked like, we are moving into an area where data is fragmentary or unobtainable and speculation is rampant. These are also events that took place billions of years ago. But such obstacles do not mean we cannot draw interesting, albeit at best tentative conclusions – there is relevant data present in each organisms' genetic data (its genotype), the structure of its cells, and their ecological interactions. It is this type of data that can inform and constrain our various speculations.

Animal cells do not have a rigid cell wall; its absence allows them to be active predators, moving rapidly and engulfing their prey whole or in macroscopic bits through phagocytosis (see above). They use complex "cytoskeletal" and "cytomuscular" systems to drive these thermodynamically unfavorable behaviors (again, largely beyond our scope here)(→). Organisms with a rigid cell wall can't perform such functions. Given that bacteria and archaea have cell walls, it is possible that cell walls were present in the common ancestral organism. But this leads us to think more analytically about the nature of the earliest organisms and the path back to the common ancestor. A cell wall is a complex structure that would have had to be built through evolutionary processes before it would be useful. If we assume that the original organisms arose in an osmotically friendly, that is, non-challenging environment, then a cell wall could have been generated in steps, and once adequate it could enable the organisms that possessed it to invade new, more osmotically challenging (dilute) environments - like most environments today.
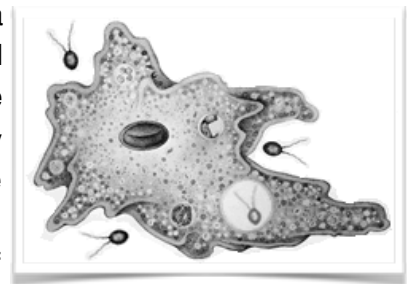
For example, one plausible scenario is that the ancestors of the bacteria and the archaea developed cell walls originally as a form of protection against predation, or as a way to explore osmotically challenging environments, environments of dilute salt solutions and high water concentration. So who were the predators? Where they the progenitors of the eukaryotes? If so, we might conclude that organisms in the eukaryotic lineage never had a cell wall, rather than that they had one once and subsequently lost it. In this scenario, the development of eukaryotic cell walls by fungi and plants represents an example of convergent evolution and that these structures are analogous (rather than homologous) to the cell walls of prokaryotes (bacteria and archaea).

But now a complexity arises, there are plenty of eukaryotic organisms, including microbes like the amoeba, that live in osmotically challenging environments. How do they deal with the movement of

water into their cells? How might they have followed their prey (bacteria and archaea) into the non-salty world? One approach is to actively pump the water that flows into them back out using an organelle known as a contractile vacuole. Water accumulates within the contractile vacuole, a membrane-bounded structure within the cell; as the water accumulates the contractile vacuole inflates. To expel the water, the vacuole connects with the plasma membrane and is squeezed out by the contraction of a cytomuscular system. This squirts the water out of the cell. The process of vacuole contraction is an active one, it involves work and requires energy. One might speculate that such as cytomuscular system was originally involved in predation in the salty world, that is, enabling the cell to move its membranes, to surround and engulf other organisms (phagocytosis). The resulting vacuole became specialized to aid in killing and digesting the engulfed prey. When digestion is complete, this micro-stomach can fuse with the plasma membrane to discharge the waste, using either a passive or an active contractile system. It turns out that the molecular systems involved in driving active membrane movement are related to the systems involved in dividing the eukaryotic cell into two during cell division; a distinctly different systems than is used by prokaryotes.[251] So which came first, different cell division mechanisms, which led to differences in membrane behavior, with one leading to a predatory active membrane and the other that led to a passive membrane, perhaps favoring the formation of a cell wall? At the same time, escape for predation and improved predation could be involved.

Making a complete eukaryote

Up to this point we have touched on only a few of the ways that prokaryotes (bacteria and archaea) differ from eukaryotes. The major ones include the fact that eukaryotes have their genetic material isolated from the cytoplasm by a complex double-layered membrane/pore system known as the nuclear envelope (which we will discuss briefly later on). Exactly how the nucleus came into being in the lineage leading to eukaryotes remains poorly defined, as is often the case in historical processes that occurred billions of years ago.[252] Another difference is the relative locations of chemo-osmotic/ photosynthetic systems in the two types of organisms. In prokaryotes, these systems (light absorbing systems, electron transport chains and ATP synthases) are located within the plasma membrane or within internal membrane vesicles derived from the plasma membrane. In contrast, in eukaryotes (plants, animals, fungi, protozoa, and other types of organisms) these structural components are not located on the plasma membrane, but rather within discrete and distinctive intracellular structures. In the case of the system associated with aerobic respiration, these systems are found in the inner membranes of a double-membrane bound cytoplasmic organelles known as a mitochondrion (plural: mitochondria). Photosynthetic eukaryotes (algae and plants) have a second type of cytoplasmic organelle, in addition to mitochondria, known as chloroplasts. Like mitochondria, chloroplasts are also characterized by the presence of a double membrane and an electron transport chain located within the inner membrane and membranes apparently derived from it. These are just the type of structures one might expect to see if a bacterial cell was engulfed by the ancestral pro-eukaryotic cell, with the host cell's membrane surrounding the engulfed cells plasma membrane (↓). A more detailed molecular analysis reveals that the mitochondrial and chloroplast electron transport systems, as well as the ATP synthase proteins,

---

[251] The cell cycle of archaea & Bacterial cell division

[252] Endosymbiotic theories for eukaryote origin

more closely resemble those found in two distinct types of bacteria, rather than in archaea. In fact, detailed analyses of the genes and proteins involved suggest that the electron transport/ATP synthesis systems of eukaryotic mitochondria are homologous to those of a γ-proteobacteria while the light harvesting/reaction center complexes, electron transport chains and ATP synthesis proteins of photosynthetic eukaryotes (algae and plants) appear to be homologous to those of a second type of bacteria, a photosynthetic cyanobacteria.[253] In contrast, many of the nuclear systems found in eukaryotes appear more similar to systems found in archaea. How do we make sense of these observations?
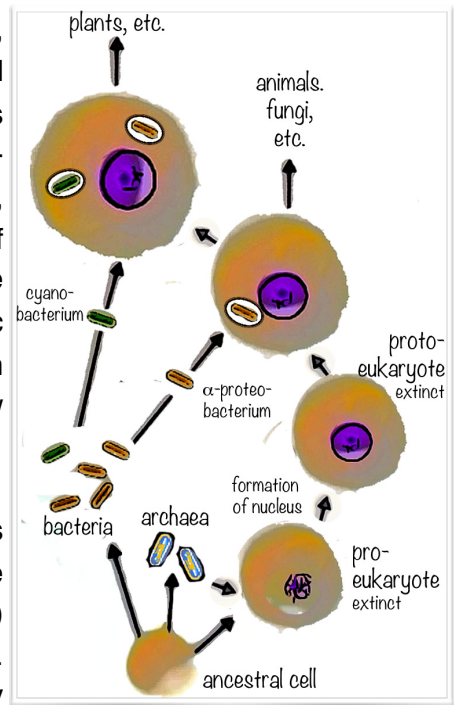
When a eukaryotic cell divides it must have also replicated its mitochondria and chloroplasts, otherwise they would eventually be lost through dilution. In 1883, Andreas Schimper (1856-1901) noticed that chloroplasts divided independently of their host cells. Building on Schimper's observation, Konstantin Merezhkovsky (1855-1921) proposed that chloroplasts were originally independent organisms and that plant cells were symbionts, essentially two independent organisms living together. In a similar vein, in 1925 Ivan Wallin (1883-1969) proposed that the mitochondria of eukaryotic cells were derived from bacteria. This "endosymbiotic hypothesis" for the origins of eukaryotic mitochondria and chloroplasts fell out of favor, in large part because the molecular methods needed to unambiguously resolve their implications were not available. A breakthrough came with the work of Lynn Margulis (1938-2011) and was further bolstered when it was found that both the mitochondrial and chloroplast protein synthesis machineries were sensitive to drugs that inhibited bacterial but not eukaryotic protein synthesis. In addition, it was discovered that mitochondria and chloroplasts contained circular DNA molecules organized in a manner similar to the DNA molecules found in bacteria (we will consider DNA and its organization soon).

All eukaryotes appear to have mitochondria. Suggestions that some eukaryotes, such as the human anaerobic parasites *Giardia intestinalis*, *Trichomonas vaginalis* and *Entamoeba histolytica* [254] do not failed to recognize cytoplasmic organelles, known as mitosomes, as degenerate, or more politely termed evolutionarily simplified mitochondria. Based on these and other data it now seems likely that all eukaryotes are derived from a last common (eukaryotic) ancestor (sometime referred to as LECA) that engulfed an aerobic α-proteobacteria-like bacterium. Instead of being killed and digested, these (or even one) of these bacteria survived within the pre-eukaryotic cell, replicated, and were distributed into the progeny cell when the parent cell divided. This process resulted in the engulfed bacterium becoming an endosymbiont, which over time became mitochondria. In the course of time, the original genome of the bacterium has been dramatically reduced in size, with many (but not all) genes transferred to the nucleus (we will consider the implications of this process later on). At the same time the engulfing cell became dependent upon the presence of the endosymbiont, initially to detoxify molecular oxygen, and then to utilize molecular oxygen as an electron acceptor so as to maximize the

---

253 The origin and early evolution of mitochondria: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC138944/

254 The mitosome, a novel organelle related to mitochondria in the amitochondrial parasite Entamoeba histolytica

energy that could be derived from the break down of complex molecules. All eukaryotes, including us), are descended from this mitochondria-containing eukaryotic ancestor, which has been estimated to have appeared ~2 billion years ago. The second endosymbiotic event in eukaryotic evolution occurred when a cyanobacteria-like bacterium formed a relationship with a mitochondria-containing eukaryote. This lineage gave rise to the glaucophytes, the red and the green algae. The green algae, in turn, gave rise to the plants.

As we look through modern organisms there are a number of examples of similar events, that is, one organism becoming inextricably linked to another through symbiotic processes. There are also examples of close couplings between organisms that are more akin to parasitism rather then a mutually beneficial interaction (symbiosis).[255] For example, a number of insects have intracellular bacterial parasites and some pathogens and parasites live inside human cells.[256] In some cases, even these parasites can have parasites. Consider the mealybug *Planococcus citri*, a multicellular eukaryote; this organism contains cells known as bacteriocytes (outlined in white →). Within these cells are *Tremblaya princeps* type β-proteobacteria (red). Surprisingly, within these Tremblaya live *Moranella endobia*-type γ-proteobacteria (green).[257] In another example, after the initial endosymbiotic event that formed the proto-algal cell, the ancestor of red and green algae and the plants, there have been endocytic events in which a eukaryotic cell has engulfed and formed an endosymbiotic relationship with eukaryotic green algal cells, to form a "secondary" endosymbiont. Similarly, secondary endosymbionts have been engulfed by yet another eukaryote, to form a tertiary endosymbiont.[258] The conclusion is that there are combinations of cells that can survive (and more importantly reproduce) better in a particular ecological niche than either could alone. In these phenomena we see the power of evolutionary processes to populate extremely obscure ecological niches in rather surprising ways.
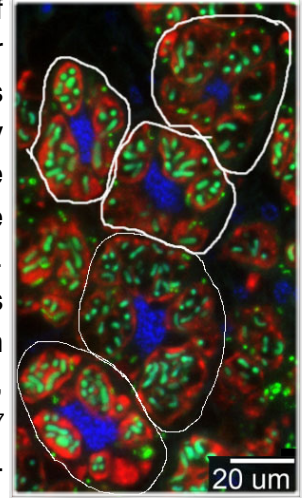
Questions to answer:
108. How would you define an osmotically friendly environment? what would be its limitations, evolutionarily?
109. Are the mitochondria of plants and animals homologous or analogous? How might you decide?
110. What advantage would the host get from early bacterial symbionts? Was there an advantage for the engulfed bacteria?
111. How would you distinguish a symbiotic from a parasitic relationship? is it always simple?

Questions to ponder:
- Why might a plant cell not notice the loss of its mitochondria? why do you think plants retain mitochondria?
- What evidence would lead you to suggest that there were multiple symbiotic events that gave rise to the mitochondria of different eukaryotes?

---

255 Mechanisms of cellular invasion by intracellular parasites: http://www.ncbi.nlm.nih.gov/pubmed/24221133

256 Intracellular protozoan parasites of humans: the role of molecular chaperones in development and pathogenesis.

257 Snug as a Bug in a Bug in a Bug & Mealybugs nested endosymbiosis

258 Photosynthetic eukaryotes unite: endosymbiosis connects the dots

Chapter 7: The molecular nature of the heredity material

In which we discover how the physical basis of genetic inheritance, DNA, was identified and learn about the factors that influence how DNA encodes genetic information, how that information is replicated and read into a useable form, how mutations occur and are often repaired, and how such an extravagantly long molecule is organized within such small cells.

One of the most amazing facts associated with Darwin and Wallace's original evolutionary model was their complete lack of a coherent or accurate understanding of genetic mechanisms. While it was clear, based on the experiences of plant and animal breeders, that organisms varied with respect to one another and that part of that variation was inherited by an organism's offspring, the mechanism(s) by which genetic information is stored and transmitted was not clear and at the time could not have been known, a situation that promotes groundless speculation. Nevertheless there were a number of hypotheses at the time, some of which relied on supernatural or metaphysical mechanisms.[259] For example, some thought that evolutionary variation was generated by an inner drive or logic within the organism or even the species - an idea known as orthogenesis. This had the comforting implication that evolutionary processes reflected some kind of over-arching design, that things were going somewhere, that there was an over-arching purpose to existence. Well before the modern theory of evolution was proposed in 1859, Jean-Baptiste Lamarck (1744–1829) suggested that inheritance somehow reflected the desires and experiences of the parent.[260] This would have predicted a type of "internally directed" evolution. In contrast Darwin's model, based on random variations in the genetic material, seemed more arbitrary and unsettling. It implied a lack of an over-arching purpose to life in general, and human existence in particular.

Another surprising realization is that modern genetics had its origins in the work of Gregor Mendel (1822–1884). He published his work on sexually reproducing peas in 1865, shortly after the introduction of the modern theory of evolution. Since Darwin published revised editions of "On the Origin of Species" through 1872, one might ask why did he not incorporate a Mendelian view of heredity into his theory? The simplest explanation would be that Darwin was unaware of Mendel's work – in fact, the implications of Mendel's work were largely ignored until the early years of the 20th century. So why were the significance and implications of Mendel's observations not immediately recognized? It turns out that Mendel's conclusions were quite specialized and could be attributed to design details of his experiments and his choice of organism. Mendel carefully selected discrete traits (phenotypes) displayed by the garden pea *Pisum sativum*: smooth versus wrinkled seeds, yellow versus green seeds, grey versus white seed coat, tall versus short plants, etc. In the plants he used, he found no intermediate versions of these traits. In addition, these traits were independent, the presence of one trait did not influence any of the other traits he was considering. Each was controlled, as we now know,

---

[259] [The eclipse of Darwin: wikipedia](#)

[260] It is perhaps worth reading Evolution in Four Dimensions (reviewed here: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC1265888/) which reflects on the factors that influence selection.

by variation at a single genetic locus (gene or position within the genome), with different genes "controlling" different traits, although as we will see, the connection between genetic information and trait is generally complex.[261] The vast majority of traits, however, do not behave in a simple Mendelian manner; most play a role in a number of different traits and a particular trait is generally controlled (and influenced) by many genes. Allelic variations in multiple genes, often referred to as the genetic background, interact in non-additive and not easily predictable ways. For example, the extent to which a trait is visible, even assuming the underlying genetic factor (allele) is present, can vary dramatically depending upon the rest of the organism's genotype, the genetic background. Finally, in an attempt to established the general validity of his conclusions Mendel was urged to examine the behavior of a number of other plants, including hawkweed. Unfortunately, hawkweed uses a specialized, asexual reproductive strategy, known as apomixis, which does not follow Mendel's laws.[262] This did not help reassure Mendel or others that his genetic laws were universal or useful. Subsequent work, published in 1900, led to the recognition of the general validity of Mendel's basic conclusions.[263]

Mendel deduced that there are stable hereditary "factors" – which became known as genes – and, as that genes are present as discrete objects within an organism. Each gene can exist in a number of different forms, known as alleles. In many cases specific alleles (versions of a gene) are associated with specific forms of a trait or the presence or absence of a trait. For example, in mammals, the ability to digest lactose depends upon whether you can make the enzyme lactase. The lactase enzyme is encoded by the LCT gene.[264] Lactase is made when the LCT gene is expressed. In most mammals, the LCT gene stops being expressed with age. In ~65% of human adults around the world, the expression of the LCT gene, and so lactase production, is off. In various sub-populations, however, the ability to digest lactose persists in adults – we refer to this trait as adult lactose tolerance. Molecular studies indicate that adult lactose tolerance has arisen independently in a number of human populations. One version of the adult lactose tolerance trait is based on which allele of the MCM6 gene you carry. The MCM6 allele that promotes lactose tolerance acts in a dominant manner to maintain the expression of the LCT gene into adulthood. As we proceed, we will consider the molecular level details involved in producing the lactose tolerance phenotype. Note, words like alleles, genomes and genotypes are terms you should be familiar with from our previous discussion of evolutionary mechanisms.

When a cell divides, all of its genes must be replicated so that each daughter cell receives a full set of genes, a genome. The exact set of alleles a cell inherits determines its genotype. Later it was recognized that sets of genes are linked together in a physical way, but that this linkage is not permanent - that is, processes exist that can shuffle linked genes, or rather the alleles of genes. In sexually reproducing organisms, such as the peas that Mendel originally worked with and most multicellular organisms, including humans, two copies of each gene are present in each somatic (body) cell. Such cells are said to be diploid. During sexual reproduction, specialized cells, known as germ

---

[261] Actually more complex that we can address here: see An expanded view of complex traits: from polygenic to omnigenic.

[262] Apomixis in hawkweed: Mendel's experimental nemesis: link

[263] Rediscovery of Mendel's work: link

[264] The Co-evolution of Genes and Culture: link

cells, are produced; these cells contain only a single copy of each gene and are referred to as haploid, although monoploid might be a better term. Two such haploid cells, known as gametes, fuse to form a new diploid organism. While gametes can be morphologically identical, in animals and plants, they are generally quite different in size and shape. The gametes of animals are known as sperm and egg, while in plants they are known as pollen and ovule. Generally an individual sexually reproducing organism produces only a single type of gamete, with the organism producing the morphologically larger gametes known as the female and the organism producing the smaller gametes are known as male. As we discussed earlier (Chapter 4), this difference in size has evolutionary (selective) implications. In any particular population there are typically a number of different alleles for each particular gene, and many thousands of different genes.[265] An important feature of sexual reproduction is that the new organism carries a unique combination of alleles inherited from its two parents. This increases the genetic variation within the population, which enables the population, as opposed to specific individuals, to deal with a range of environmental factors, including pathogens, predators, prey, and competitors. It leaves unresolved, however, exactly how genetic information is replicated and how new alleles form, how information is encoded, regulated, and utilized at the molecular, cellular, and organismic levels.

Questions to answer

112. Make a plausible diagram (based on what you know now) of the genetic system responsible for the ability to digest lactose in mammals; include how it is normally regulated over baby to adult time and where mutations could lead to adult lactose tolerance.
113. Under what conditions would be being tolerant as an adult be positively selected for; produce a model for why adult lactose tolerance not a universal trait of mammals?

Discovering how nucleic acids store genetic information

To follow the historical pathway that led to our understanding of how heredity works, we have to start back at the cell, the basic living unit. As it became more firmly established that all organisms are composed of one or more cells, and that all cells were derived from pre-existing cells, it became more and more likely that inheritance had to be a cellular phenomenon. As part of their studies, cytologists (students of the cell) began to catalog the common components of cells; because of resolution limits associated with available microscopes, these studies were restricted to larger eukaryotic cells. One such component of eukaryotic cells was the nucleus. At this point it is worth remembering that most cells do not contain pigments. Under these early (bright-field) microscopes, they appear clear, after all they are ~70% water. To be able to discern structural details cytologists had to stabilize the cell and to visualize its various components. As you might suspect, stabilizing the cell means killing it. To be observable, the cell had to be killed (known technically as "fixed") in such a way as to insure that its structure was preserved as close to the living state as possible. Originally, this process involved the use of chemicals, such as formaldehyde, that could cross-link various molecules together. Cross-linking stops these molecules from moving with respect to one another; it is not unlike boiling an egg. Alternatively, the cell could be treated with organic solvents such as alcohols; this leads to the local precipitation of the cell's water soluble components and solubilization of the lipids that form cellular membranes. As long as the methods used to visualize the fixed tissue were of low magnification and
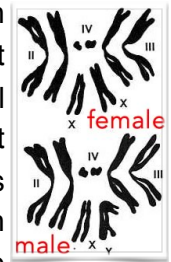
---

[265] You can get an idea of the alleles present in the human population by using the ExAC browser: link

resolution, the results obtained using chemical fixatives were generally acceptable. In more modern studies, using higher resolution optical methods[266] and electron microscopes, such crude fixation methods became unacceptable, and have been replaced by various alternatives, including rapid freezing. Even so it was hard to resolve the different subcomponents of the cell. To do this the fixed cells were treated with various dyes. Some dyes bind preferentially to molecules located within particular parts of the cell. The most dramatic of these cellular sub-regions was the nucleus, which could be readily identified because, due to its bulk chemical composition, it was stained very differently from the surrounding cytoplasm. One standard stain involves a mixture of hematoxylin (actually oxidized hematoxylin and aluminum ions) and eosin, which leaves the cytoplasm pink and the nucleus dark blue.[267] The nucleus was first described by Robert Brown (1773-1858), the person after which Brownian motion was named. The presence of a nucleus was characteristic of eukaryotic (true nucleus) organisms.[268] Prokaryotic cells (before a nucleus) are typically much smaller and originally it was impossible to determine whether they had a nucleus or not – they do not.

The careful examination of fixed and living cells revealed that the nucleus undergoes a dramatic reorganization during the process of cell division; it loses its typically roughly spherical shape, which was replaced by discrete stained strands, known as chromosomes (colored bodies). In 1887 Edouard van Beneden (1846-1910) reported that the number of chromosomes in a somatic (diploid) cell was constant for each species and that different species had different numbers of chromosomes (→). Within a particular species the individual chromosomes can be recognized

species	chromosome #
Ophioglossum reticulatum (a fern)	1260 (630 pairs)
Canis familiaris (dog)	78 (39 pairs)
Cavia cobaya (guinea pig)	60 (30 pairs)
Solanum tuberosom (potato)	48 (24 pairs)
Homo sapiens (humans)	46 (23 pairs)
Macaca mulatta (monkey)	42 (21 pairs)
Mus musculus (mouse)	40 (20 pairs)
Felis domesticus (house cat)	38 (19 pairs)
Saccharamyces cervisae (yeast)	32 (16 pairs)
Drosophila melanogaster (fruit fly)	8 (4 pairs)
Myrmecia pilosula (ant)	2 (1 pair)

based on their distinctive sizes and shapes. For example, in the somatic cells of the fruit fly *Drosophila melanogaster* there are two copies of each of 4 chromosomes (→). In 1902, Walter Sutton published his observation that chromosomes obey Mendel's rules of inheritance, that is that during the formation of the cells (gametes) that fuse during sexual reproduction, each cell received one and only one copy of each chromosome. This strongly suggested that Mendel's genetic factors were associated with chromosomes.[269] By this time, it was recognized that there were many more Mendelian factors than chromosomes, which implied that many factors must be present on a particular chromosome. These observations provided a physical explanation for the observation that many genetic traits did not behave independently but acted as if they were somehow linked together. The behavior of the nucleus, and the chromosomes that appeared to exist within it, mimicked the type of behavior that a genetic material would be expected to display.

---

266 Optical microscopy beyond the diffraction limit: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2645564/

267 The long history of hematoxylin: http://www.ncbi.nlm.nih.gov/pubmed/16195172

268 There are some eukaryotic cells, like human red blood cells, that do not have a nucleus, they are unable to divide.

269 http://www.nature.com/scitable/topicpage/developing-the-chromosome-theory-164

These cellular anatomy studies were followed by studies on the composition of the nucleus. As with many scientific studies, progress is often made when one has the right "model system" to work with. It turns out that some of the best systems for the isolation and analysis of the components of the nucleus were sperm and pus, isolated from discarded bandages from infected wounds (yuck). It was therefore assumed, quite reasonably, that components enriched in this material would likely be enriched in nuclear (genetic information containing) components. Using sperm and pus as a starting material Friedrich Miescher (1844–1895) was the first to isolate a phosphorus-rich compound, called nuclein.[270] At the time of its isolation there was no evidence linking nuclein to genetic inheritance. Later nuclein was resolved into an acidic component, deoxyribonucleic acid (DNA), and a basic component, primarily proteins known as histones. Because they have different properties (acidic DNA, basic histones), chemical "stains" that bind or react with specific types of molecules and absorb visible light, could be used to visualize the location of these molecules within cells using a light microscope. The nucleus stained for both highly acidic and basic components - which suggested that both nucleic acids and histones were localized to the nucleus, although what they were doing there was unclear.

Questions to answer

114. How was the nucleus first visualized? Why was this accomplished in eukaryotic, rather than prokaryotic cells?
115. Is there a correlation between the number of chromosomes and the complexity of an organism. Does that tell you anything useful about genes?

Questions to ponder

- How would you define a model system? What is it that makes model systems useful?
- In comparing organisms, what does complexity mean?

Locating hereditary material within the cell

Further evidence suggesting that hereditary information was localized in the nucleus emerged from transplantation experiments carried out in the 1930's by Joachim Hammerling (1901-1980); he used the giant unicellular green alga *Acetabularia acetabulum*, known as the mermaid's wineglass (→). Hammerling's experiments (video link) illustrate two important themes in the biological sciences. The idiosyncrasies of specific organisms can be exploited to carry out useful studies that are simply impossible to perform elsewhere. At the same time, the underlying evolutionary homology of organisms makes it possible to draw broadly relevant conclusions from studies on a particular organism, something unlikely to be true if each represented a unique creation event. In this case, Hammerling exploited three unique features of *Acetabularia*. The first is the fact that each individual is a single cell, with a single nucleus. Through microdissection, it is possible to isolate nuclear and anucleate (without a nucleus) regions of the organism. Second, these cells are very large (1 to 10 cm in height), which makes it possible to carry out various microsurgical operations. You can remove and transplant regions of one organism (cell) to another. Finally, different species of *Acetabularia* have mophologically distinct
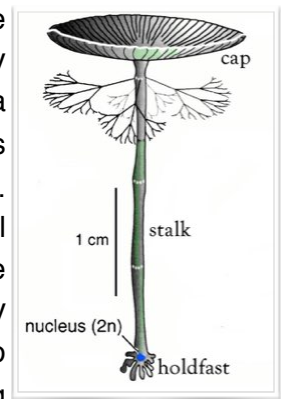
[270] Friedrich Miescher and the discovery of DNA: http://www.sciencedirect.com/science/article/pii/S0012160604008231

"caps" that regrow faithfully following amputation. In his experiments, he removed the head and stalk regions from one individual, leaving a "holdfast" region that was much smaller but, importantly, contained the nucleus. He then transplanted large regions of a anuclear stalk, derived from an organism of different species with a distinctively different cap morphology, onto the nucleus-containing holdfast region. When the cap regrew it had the morphology characteristic of the species that provided the nucleus - no matter that this region was much smaller than the transplanted, anucleate stalk region. The conclusion was that the information needed to determine the cap's morphology was located within the region of the cell that contained the nucleus, rather than dispersed throughout the cytoplasm. Its just a short step from these experimental results to the conjecture that all genetic information is located within the nucleus.

Identifying DNA as the genetic material

The exact location, and the molecular level mechanisms behind the storage and transmission of genetic information, still needed to be determined. Two kinds of experiment led to the realization that genetic information was stored in a chemically stable form. In one set of studies, H.J. Muller (1890–1967) found that exposing fruit flies to X-rays (a highly energetic form of light) generated mutations that could be passed from generation to generation. This suggested that genetic information was stored in a chemical form and that that information could be altered through interactions with radiation, which presumably altered the molecule(s) storing the information. Once altered, the information was again stable.

The second piece of experimental evidence supporting the idea that genetic information was encoded in a stable chemical form came from a series of experiments initiated in the 1920s by Fred Griffith (1879–1941). He was studying two strains of the bacterium *Streptococcus pneumoniae,* which causes bacterial pneumonia. When introduced into mice, the mice got sick and died. Griffith grew these bacteria in the laboratory. This is known as culturing the bacteria. We say that bacteria grown in culture have been grown *in vitro* or in glass (although in modern labs, they are generally grown in plastic), as opposed to *in vivo* or within a living animal. Following common methods, he grew bacteria on plates covered with solidified agar (a jello-like substance derived from salt water algae) containing various nutrients. Typically, a liquid culture of bacteria is diluted and spread on the agar surface of the plate. When diluted sufficiently, isolated individual bacteria come to rest on the agar surface. Individual bacteria bind to the plate independently of, and separated from, one another. Bacteria are asexual and so each bacterium can grow up into a colony, a clone of the original bacterium that landed on the plate. The disease-causing strain of *S. pneumoniae* grew up into smooth or S-type colonies, due to the fact that the bacteria secrete a slimy mucus-like substance. Griffith found that mice injected with S strain *S. pneumoniae* quickly sickened and died (→). However, if he killed the bacteria with heat before injection the mice did not get sick, indicating that it was the living bacteria that produced (or evoked) the disease symptoms rather than some heat-stable chemical toxin.
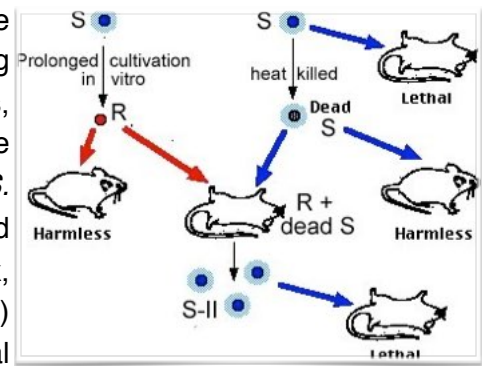
During extended cultivation *in vitro*, however, cultures of S strain bacteria sometimes gave rise to rough (R) colonies; R colonies were not smooth and shiny but rather rough in appearance. This appeared to be a genetic change since once isolated, R-type strains produced R-type colonies, a process that could be repeated many, many times. More importantly, mice injected with R strain *S. pneumoniae* did not get sick. A confusing complexity emerged however; mice co-injected with the living R strain of *S. pneumoniae* (which did not cause disease) and dead S strain *S. pneumoniae* (which also did not cause the disease) did, in fact, get sick and died! Griffith was able to isolate and culture *S. pneumoniae* from these dying mice and found that, when grown in vitro, they produced smooth colonies. He termed these S-II (smooth) strains. His hypothesis was that a stable chemical (that is, non-living) component derived from the dead S bacteria had "transformed" the avirulent (benign) R strain bacteria to produce a new virulent S-II strain of bacteri.[271] Unfortunately Fred Griffith died in 1941 during the Nazi- bombing of London, which put an abrupt end to his studies.[272]

In 1944 Griffith's studies were continued and extended by Oswald Avery (1877-1955), Colin McLeod (1909-1972) and Maclyn McCarty (1911-2005). They set out to use Griffith's assay to isolate what they termed the "transforming principle" responsible for turning R strains of *S. pneumoniae* into S strains. Their approach was to grow up large numbers of cells *in vitro*, and to then grind up these cells and isolate their various components, such as proteins, nucleic acids, carbohydrates, and lipids. They then digested these extracts with various enzymes (reaction specific catalysts) and ask whether the transforming principle remained intact.

Treating cellular extracts with proteases (which degrade proteins), lipases (which degrade lipids), or RNAases (which degrade RNAs) had no effect on the transforming principle. In contrast, treatment of the extracts with DNAases, enzymes that degrade DNA, destroyed the extracts transforming activity. Further support for the idea that the "transforming substance" was DNA was suggested by the fact that purified transforming substance had the physical properties of DNA; for example it absorbed light like DNA rather than protein (absorption spectra of DNA versus protein →). Subsequent studies confirmed this conclusion. Furthermore DNA isolated from R strain bacteria was not able to produce S-II strains from R strain bacteria, whereas DNA from S strain bacteria could transform R strains into S-II strains. They concluded that DNA derived from S cells contains the information required for the conversion – it is, or rather contains, a gene required for the S strain phenotype. This information had, presumably, been lost by mutation during the formation of R strains.
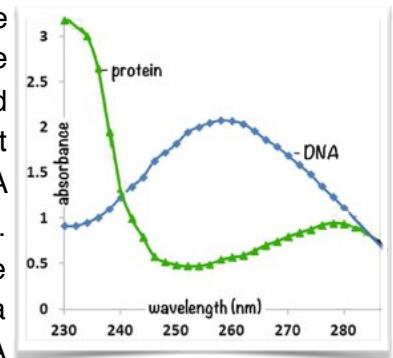
The basic phenomena exploited by Griffiths and Avery et al., known as transformation, is an example of horizontal gene transfer, which we will discuss in greater detail later on. It is the movement of genetic information from one organism to another. This is a distinctly different process than the movement of genetic information from a parent to an off-spring, which is known as vertical gene transfer. Horizontal gene transfer can occur between unrelated organisms, although it is most common

---

[271] link: Griffith's experiment

[272] And provides yet another good reason (as if we need more) to hold Nazis in contempt.

among prokaryotes. Various forms of horizontal gene transfer occur within the microbial world and allow genetic information to move between species. For example horizontal gene transfer is responsible for the rapid expansion of populations of antibiotic-resistant bacteria. Viruses use a highly specialized (and optimized) form of horizontal gene transfer, known as transduction.[273] An obvious question is, how is this possible? While we might readily accept that genetic information must be transferred from parent to offspring (we see the evidence for this process with our eyes in the form of family resemblances), the idea that genetic information can be transferred between different organisms that are not (apparently) related to one another is quite a bit more difficult to swallow. As we will see, horizontal gene transfer is possible primarily because all organisms share the same basic system for encoding, reading and replicating genetic information. The hereditary machinery is homologous among existing organisms.

Questions to answer

116. How would Hammerling's observations have been different if hereditary information was localized in the cytoplasm?
117. In Griffith's study, he found that dead smooth *S. pneumoniae* could transform living rough strains of *S. pneumoniae* when co-injected into a mouse. Would DNA from an unrelated species of bacteria give the same result? Explain your reasoning.
118. What caused the change from S to R strains in culture? Why is DNA from the R strain unable to produce S-II cells?
119. In the spectrometric analysis of DNA and protein, what is plotted on the X- and Y-axes?

Questions to ponder

- What is the difference between a strain and a species?
- How might horizontal gene transfer confuse molecular phylogenies (family trees)?
- How might a creationist explain horizontal gene transfer?

Unraveling Nucleic Acid Structure

Knowing that the genetic material was DNA was a tremendous break through, but it left a mystery - how was genetic information stored and replicated. Nucleic acids were thought to be aperiodic polymers, that is, molecules built from a defined set of subunits, known as monomers, but without a simple overall repeating pattern. The basic monomeric units of nucleic acids are known as nucleotides (→). A nucleotide consists of three distinct types of molecules joined together, a 5-carbon sugar (ribose or deoxyribose), a nitrogen-rich "base" that is either a purine (guanine (G) or adenine (A)) or a pyrimidine (cytosine (C), or thymine (T)) in DNA or uracil (U) instead of T in RNA, and a phosphate group. The carbon atoms of the sugar are numbered 1' to 5'. The nitrogenous base is attached to the 1' carbon and the phosphate is attached to the 5' carbon. The other functionally important group is a hydroxyl group attached to the 3' carbon of the ribose/deoxyribose moiety.[274] RNA differs from DNA in that there is a hydroxyl group attached to the 2' carbon of the ribose, this hydroxyl is absent in DNA, which is why it is "deoxy" ribonucleic acid! We take particular note of the 5' phosphate and 3' hydroxyl groups of the

BASE

Phosphate

deoxyribonucleotide
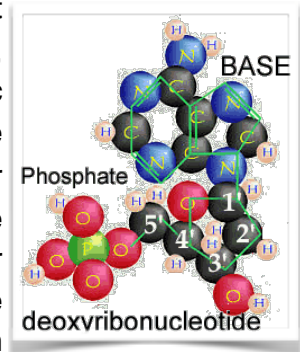
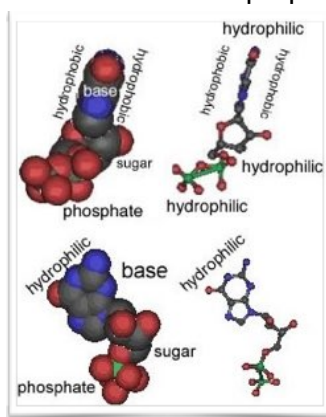[273] link:: Virus-like particles speed bacterial evolution

[274] "Moiety" defined

ribose/deoxyribose because they are directly involved in the linkage of nucleotide monomers together to form nucleic acid polymers.

Discovering the structure of DNA

A critical clue to understanding the structure of nucleic acids came from the work of Erwin Chargaff (1905–2002). When analyzing DNA from various sources, he found that the relative amounts of G, C, T and A nucleotides present varied between organisms but were the same (or very similar) for organisms of the same type or species. On the other hand, the ratios of A to T and of G to C were always equal to 1, no matter where the DNA came from. Knowing these rules, James Watson (1928–) and Francis Crick (1916–2004) built a model of DNA that fit what was known about the structure of nucleotides and structural data from Rosalind Franklin (1920–1958).[275] Franklin got her data by pulling DNA molecules into oriented strands, fibers of many molecules aligned parallel to one another. By passing a beam of X-rays through these fibers she was able to obtain a diffraction pattern.[276] This pattern is based on the structure of DNA molecules and defines key parameters that constrain any model of the molecule's structure. By making a model that was predicted to produce the observed X-ray data, Watson and Crick were able to make a number of conclusions about the structure of a DNA molecule.

To understand this process, let us consider the chemical nature of a nucleotide and a nucleotide polymer (a nucleic acid) such as DNA. First the nucleotide bases in DNA (A, G, C and T) have a number of similar properties. Each nucleotide has three hydrophilic regions: the negatively charged

phosphate group, a sugar which has a number of O–H groups, and a hydrophilic edge of the base (where the N–H and N groups lie)(←). While the phosphate and sugar are three-dimensional moieties, the bases are flat, the atoms in the rings are all in one plane. The upper and lower surfaces of the rings are hydrophobic (non-polar) while the edges have groups that can interact via hydrogen bonds. This means that the amphipathic factors that favor the assembly of lipids into bilayer membranes are also at play in nucleic acid structure. In their model Watson and Crick had the bases stacked on top of one another, hydrophobic surface next to hydrophobic surface, to reduce their interactions with water.
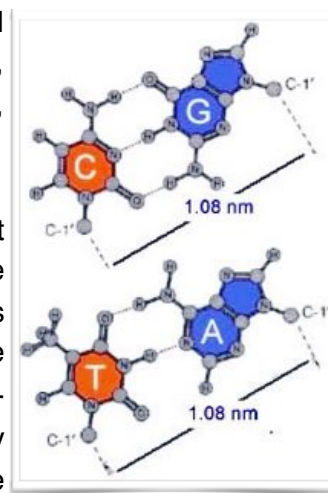
This left each base's hydrophilic edge, with -C=O and -N-H groups that can act as H-bond acceptors and donors, to be dealt with. How were these hydrophilic groups to be arranged? With the two polynucleotide strands arranged in opposite orientations, that is, anti-parallel to one another: one from 5' → 3' and the other 3' ← 5', the bases attached to the sugar-phosphate backbone could interact with one another in a highly specific way (→). An A can form two hydrogen bonding interactions with a T on the

1.08 nm

1.08 nm

---

[275] An interesting depiction of this process is provided by the movie "Life Story"

[276] Fiber diffraction

opposite (anti-parallel) strand, while a G could form three hydrogen bonding interactions with a C. A key feature of this arrangement is that the lengths of the A::T and G:::C base pairs are almost identical. The hydrophobic surfaces of the bases are stacked on top of each other, while the hydrophilic sugar and phosphate groups are in contact with the surrounding aqueous solution. The repulsion between negatively charged phosphate groups is neutralized (or shielded) by the presence of positively charged ions present in the solution from which the X-ray measurements were made. This model also provided a direct explanation for why Chargaff's rules were universal in double stranded DNA.

Each DNA polymer strand has a directionality to it, it runs from the 5' phosphate group of the ribose/deoxyribose at one end to the 3' hydroxyl group of the ribose/deoxyribose at the other. Each nucleotide monomer is connected to the next through a phosphodiester linkage (→) involving its 5' phosphate group attached to the 3' hydroxyl of the existing strand. In their final model Watson and Crick depicted what is now known as B-form DNA. This is the usual form of DNA in a cell. However, it is worth noting that under different salt conditions, DNA can form two other double helical forms, known as A and Z. While the A and B forms of DNA are "right-handed" helices, the Z-form of DNA is a left-handed helix (←). We will not concern ourselves with these other forms of DNA, leaving that to more advanced courses, but you can imagine that they might well influence the types of intermolecular interactions that occur between DNA and other molecules, particularly proteins.
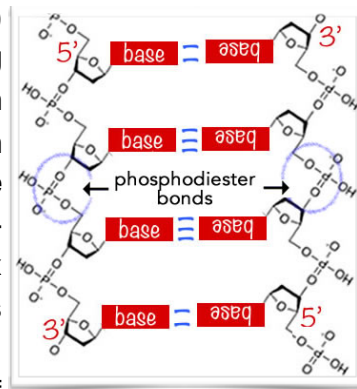
As soon as the Watson-Crick model of DNA structure was proposed its explanatory power was obvious. Because the A::T and G:::C base pairs are of the same length, the sequence of bases along the length of a DNA molecule (written, by convention in the 5' to 3' direction) has little effect on the overall three-dimensional structure of the molecule. That implies that essentially any sequence can be found, at least theoretically, in a DNA molecule. If information were encoded in the sequence of nucleotides along a DNA strand, any information could be placed there and that information would be as stable as the DNA molecule itself. This is similar to the storage of information in various modern computer memory devices, that is, any type of information can be stored, because storage does not involve any dramatic change in the basic structure of the storage material. The structure of a flash memory drive is not dramatically different whether in contains photos of your friends, a song, a video, or a textbook. At the same time, the double-stranded nature of the DNA molecule's structure and complementary nature of base pairing (A to T and G to C) suggested a simple model for DNA (and information) replication - that is, pull the two strands of the molecule apart and build new (anti-parallel) strands using the two original strands as templates. This model of DNA replication is facilitated by the fact that the two strands of the parental DNA molecule are held together by weak hydrogen bonding interactions; no covalent bonds are broken when the strands are separated from one another. In fact, at physiological temperatures DNA molecules often open up over short stretches and then close again, a process known as DNA breathing.[277] This makes the replication of the information stored in the molecule conceptually straightforward, even though the actual biochemical process is complex, in part

_____

[277] Dynamic approach to DNA breathing: http://www.ncbi.nlm.nih.gov/pubmed/23345902

because of the importance of accurate replication. The existing strands determine the sequence of nucleotides on the newly synthesized strands. The newly synthesized strand can, in turn, direct the synthesis of a second strand, identical to the original strand. Finally, the double stranded nature of the DNA molecule means that any information within the molecule is, in fact, stored in a redundant fashion. If one strand is damaged, that is its DNA sequence is lost or altered, the second undamaged strand can be used to repair that damage. A number of mutations in DNA are repaired using this type of mechanism (see below).

Questions to answer

120. How is a DNA molecule analogous to a lipid bilayer; draw a diagram that reveals the similarities and note the most important differences?
121. Which do you think is stronger (and why), a AT or a GC base pair?
122. Why is the ratio of A to T the same in all organisms?
123. Normally DNA exists inside of cells at physiological salt concentration (~140 mM KCl, 10 mM NaCl, 1 mM $MgCl_2$ and some minor ions). Predict what would happen (what is thermodynamically favorable) if you place DNA into distilled water (that is, in the absence of dissolved salts.)]
124. Consider a double stranded DNA molecule in which one strand has been broken and a regions nucleotides within in have been lost. Generate a model by which the molecular can be accurately repaired, that is, with any loss of information.

Questions to ponder

- Why does the ratio of A to G differ between organisms?
- You isolated DNA from an organism, and you find it fails to obey Chargaff's rule; what might you predict about the structure of its DNA?

DNA, sequences & information

We can now assume that somehow the sequence of nucleotides in a DNA molecule encodes information but exactly what kind(s) of information are stored in DNA? Early students of DNA could not read DNA sequences, as we can now, so they relied on various measurements to better understand the behavior of DNA molecules. For example, the way a double stranded DNA molecule interacts with light is different from how a single stranded DNA molecule interacts with light. Since the two strands of double stranded DNA molecules, often written dsDNA, are linked only by hydrogen bonds, increasing the temperature of the system will lead to their separation into two single stranded molecules (ssDNA) (left panel figure ↓ next page). ssDNA absorbs light at 260nm (in the ultraviolet range) more strongly than does dsDNA, so the absorbance of a DNA solution can be used to determine the relative amounts of single and double stranded DNA in a sample. What we find is that the temperature at which 50% of dsDNA molecules have separated into ssDNA molecules varies between organisms. This is not particularly surprising given Chargaff's observation that the ratio of AT to GC varies between various organisms and the fact that GC base pairs, mediated by three H-bonds, are more stable (take more energy to separate) than AT base pairs, which are held together by only two H-bonds. In fact, one can estimate the AT:GC ratio of a DNA molecule based on melting curves (middle pane ↓ next page).

It quickly became clear that things were more complex than previously expected. Here a technical point needs to be introduced. Because of the extreme length of the DNA molecules found in biological systems, it is almost impossible to isolate them intact. In the course of their purification, the molecules will be sheared into shorter pieces, typically thousands to tens of thousands of base pairs in length
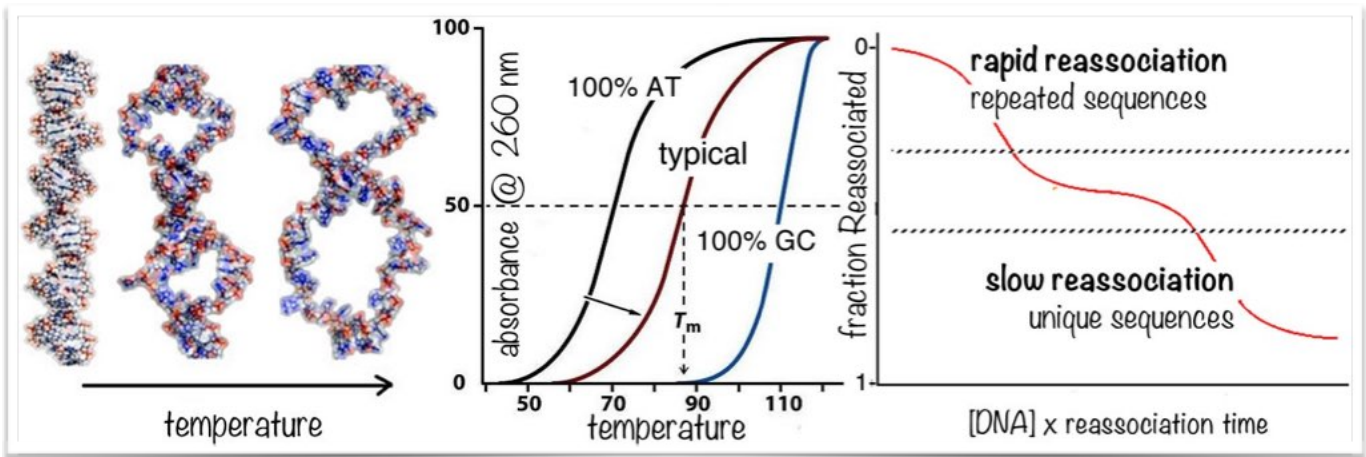
compared to the millions to hundreds of millions of base pairs in intact molecules. In another type of experiment, one can look at how fast ssDNAs (the result of a melting experiment) reforms dsDNA. The speed of these "reannealing reactions" depends on DNA concentration. When such experiments were carried out, it was found that there was a fast annealing population of DNA fragments and various slower annealing populations (right panel ↑). How to explain this result, was it a function of AT:GC ratio? Subsequent analyses revealed that it was due to the fact that within the DNA isolated from organisms, particularly eukaryotes, there were many (hundreds to thousands) of molecular regions that contained very similar nucleotide sequences. Because the single strands of these fragments can associate with one another, these sequences occurred in much higher effective concentrations compared to regions of the DNA with unique sequences. This type of analysis revealed that much of the genome of eukaryotes is composed of various families of repeated sequences and that regions of unique sequence amount to less than ~5% of the total genomic DNA. While a complete discussion of these repeated sequence elements is beyond our scope here, we can make a few points. As we will see, there are mechanisms that can move regions of a DNA molecule from one position to another within the genome, or that can generate a copy of a sDNA sequence and insert it into another position of the genome (leaving the original sequence behind). The end result is that the genome (the DNA molecules) of a cell/organism is dynamic, a fact with profound evolutionary implications.

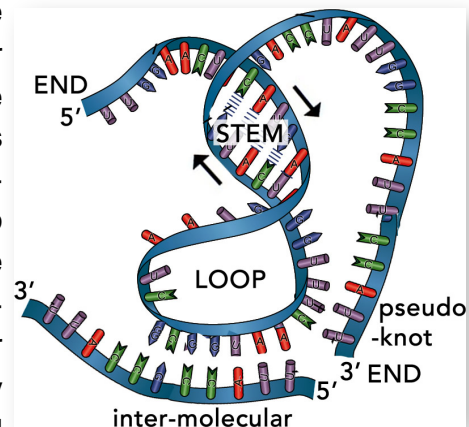**Discovering RNA: structure and some functions**

DNA is not the only nucleic acid found in cells. A second class of biological nucleic acid is known as ribonucleic acid (RNA.) RNA differs from DNA in that RNA contains i) the sugar ribose (with a hydroxyl group on the 2' C) rather than deoxyribose; ii) it contains the pyrimidine uracil instead of the pyrimidine thymine found in DNA (→); and iii) RNA is typically single rather than double stranded.[278] Nevertheless, RNA molecules can associate with an ssDNA molecule with a complementary nucleotide sequence. Instead of the A-T pairing in DNA we find A pairing with U instead. This change does not make any significant difference when the RNA strand interacts with DNA, since the number of hydrogen bonding

---
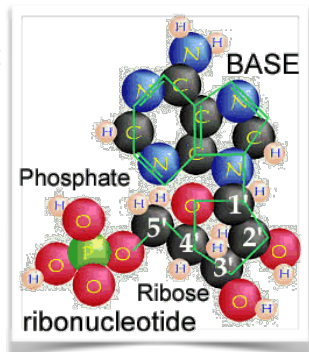
[278] The exception involves viruses, where double stranded RNA is found as the genetic material: https://www.ncbi.nlm.nih.gov/pubmed/15010213

interactions are the same.

When RNA is isolated from cells, the major population was found to reassociate with unique sequences within the DNA. As we will see later, this class of RNA, includes molecules, known as messenger or mRNAs, that carry information from DNA to the molecular machinery that mediates the synthesis of proteins (the ribosome). In addition to mRNAs there are a number of other types of RNAs in cells; in each case, their synthesis is directed by DNA-dependent RNA polymerases. These non-mRNAs include structural, catalytic, and regulatory RNAs. As you may already suspect, the same hydrophobic/hydrophilic/H-bond considerations that were relevant to DNA structure apply to RNA structure, but because RNA is generally single stranded, the structures found in RNA are different and more varied. A single-stranded RNA molecule can fold back on itself (intra-molecular interactions) to create double stranded regions (→); similarly distinct RNA molecules can interact through double-stranded regions (inter-molecular interactions). In both cases, and just as in DNA, these strands are anti-parallel to one another. This results in double-stranded regions ("stems") that end in single-stranded "loops" (or molecular ends). Regions within a stem, which can be as short as 1 base pair), that do not base pair will "bulge out". The end result is that RNA molecules can adopt a wide range of complex three-dimensional structures in solution.

Transfer RNAs (tRNAs)(→), an integral component of the protein synthesis system, are one well studied example of how intermolecular interactions within an RNA molecule can produce complex three-dimensional shapes that carry out specific molecular functions (described in greater detail in the next chapter).
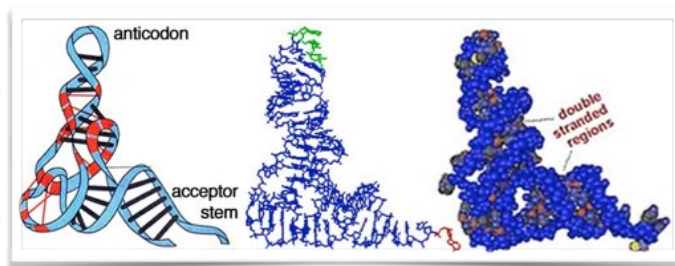
In addition to intra- and inter-molecule interactions involving RNA molecules, RNAs can also interact with proteins to form "riboprotein" complexes. For example, the recently described CRISPR-Cas9 system involves a double-stranded DNA endonuclease (an enzyme that generates the cleavage of both strands of a double-stranded DNA molecule) that is directed to specific DNA sequences through an associated RNA molecule (known as a guide RNA). Other RNA-protein complexes (to be considered in greater detail later on) are involved in the control of RNA synthesis and stability, among a number of other functions. The classic riboprotein is the ribosome, a macromolecular machine that mediates the synthesis of polypeptides. A ribosome is a complex of structural and catalytic RNAs (known as ribosomal or rRNAs) and ~50 to 80 proteins (polypeptides), depending upon whether you are prokaryotic or eukaryotic; altogether it has a molecular weight of ~3.2 x 10^6 daltons.

The ability of RNA to both encode information in its base sequence and to mediate catalysis through its three dimensional structure has led to the "RNA world" hypothesis that proposes that early in the evolution of life various proto-organisms relied on RNAs, or more likely simpler RNA-like molecules, rather than DNA and proteins, to store genetic information and to catalyze at least a subset of metabolic

reactions. Some modern day viruses use single or double-stranded RNAs as their genetic material. According to the RNA world hypothesis, it was only later in the history of life that organisms developed the more specialized DNA-based systems for genetic information storage and proteins for catalysis and other structural functions. While this idea is compelling, there is no reason to believe that simple polypeptides and other molecules were not also present and playing a critical role in the early stages of life's origins. At the same time, there are many unsolved issues associated with a simplistic RNA world view, the most important being the complexity of RNA itself, its abiogenic (that is, without life) synthesis, and the survival of nucleotide triphosphates in solution. Nevertheless, it is clear that catalytic and regulatory RNAs play a key role in modern cells and throughout their evolution. The catalytic activity of the ubiquitous ribosome, which is involved in protein synthesis in all known organisms, is based on a ribozyme, a RNA-based catalyst.

Questions to answer:

125. How would you calculate the probability that two DNA sequences (of length N) are identical by chance?
126. How does the annealing curve of genomic DNA change as the number of repeated sequence change from 0 to 1000?
127. Propose a plausible model for how a single-stranded RNA molecule could act as a catalyst; consider why double-stranded DNA is unlikely to act catalytically.

Question to ponder:

- What are the possible functions for the unique and repeated sequences of DNA in a genome.

DNA replication

Once it was proposed, the double-helical structure of DNA immediately suggested a simple mechanism for the accurate duplication of the information stored in DNA. Each strand contains all of the information necessary to specify the sequence of the complementary strand. The process begins when a dsDNA molecule opens to produce two single-stranded regions. Where DNA is naked, that is, not associated with other molecules (proteins), the opening of the two strands can occur easily, since the two strands are held together by weak H-bonding interactions. Normally, the single strands simply reassociate with one another. To replicate DNA the open region has to be stabilized and the catalytic machinery involved organized. We will consider how this is done only in general terms, in practice this is a complex and highly regulated process involving a number of components.

The first two issues we have to address in the context of DNA replication may seem arbitrary, but they turn out to be common (conserved) features of DNA synthesis. The enzymes (DNA-dependent, DNA polymerases) that catalyze the synthesis of new DNA strands cannot start the synthesis of a new polynucleotide strand on their own, they must add nucleotides onto the end of an pre-existing nucleic acid polymer, they depend on a "polynucleotide primer". In contrast, the catalysts that synthesize RNA (DNA-dependent, RNA polymerases) do not require a pre-existing nucleic acid strand, they can start the synthesis of a new RNA strand, based on complementary DNA sequence, *de novo*, that is without a polynucleotide primer. Both DNA and RNA polymerases link the 5' end of a nucleotide triphosphate molecule to the pre-existing 3' end of a nucleic acid molecule; the polymerization reaction is said to proceed in the 5' to 3' direction. As we will see later on, the molecules involved in DNA replication and RNA synthesis rely on signals within the DNA that are recognized by proteins and that determine where

and when nucleic acid replication occurs and where synthesis starts and stops. For now let us assume that some process has determined where DNA replication starts.
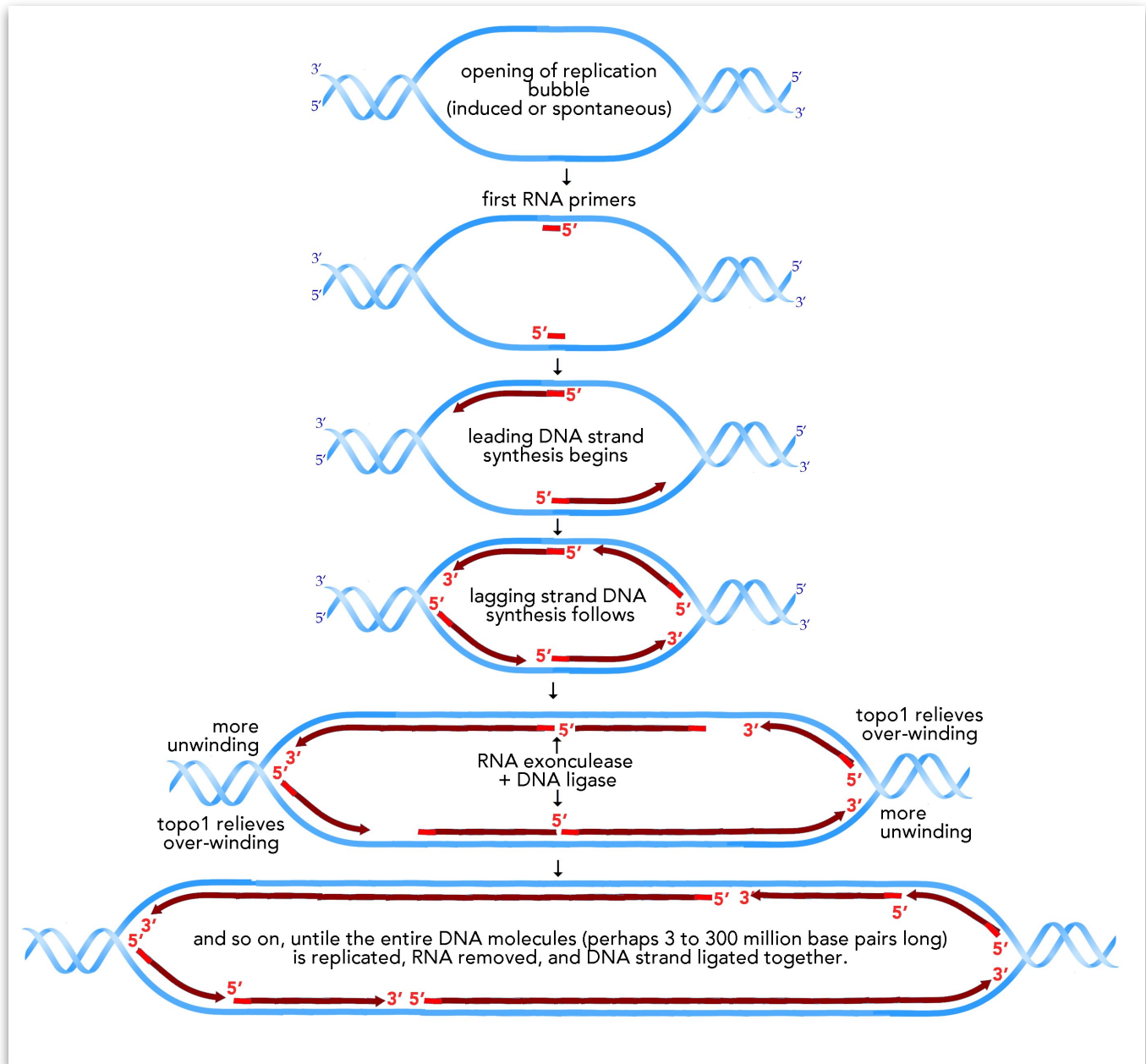
The first step in DNA replication (↑) is to locally open up the dsDNA molecule. A specialized RNA-dependent, RNA polymerase, known as primase, collides with, binds to, and synthesizes a short RNA molecule, known as a primer. Because the two strands of the DNA molecule point in opposite directions (they are anti-parallel), one primase complex associates with each of the now separated DNA strands; two RNA primers are generated, one on each strand. Once these RNA primers are in place, DNA-dependent, DNA polymerases replace the primase enzymes and begin to catalyze the deoxynucleotide-addition reaction; which nucleotide is added is determined by which nucleotide is present next in the existing DNA strand. The nucleotide addition reaction involves various nucleotides

colliding with the DNA-primer-polymerase complex; only the appropriate nucleotide, complementary to the nucleotide residue in the existing DNA strand is bound and used in the reaction.

Nucleotides exist in various phosphorylated forms within the cell, including nucleotide monophosphate (NMP), nucleotide diphosphate (NDP), and nucleotide triphosphate (NTP). To make the nucleic acid polymerization reaction thermodynamically favorable, the reaction uses the NTP form of the nucleotide monomers, generated through the reaction:

$$(5'P)NTP(3'OH) + (5'P)NTP(3'OH) + H_2O \longleftrightarrow (5'P)NTP\text{-}NMP(3'OH) + diphosphate.$$
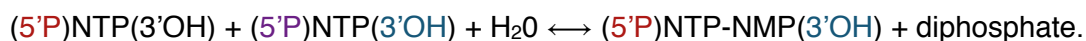
During the reaction the terminal diphosphate of the incoming NTP is released (a thermodynamically favorable reaction) and the nucleotide mono-phosphate is added to the existing polymer through the formation of a phosphodiester [-C-O-P-O-C] bond. This reaction creates a new 3' OH end for the polymer that can, in turn, react with another NTP. In theory, this process can continue until the newly synthesized strand reaches the end of the DNA molecule. The strand synthesized from the original primer is known as the "leading" strand. For the process to continue, however, the double stranded region of the original DNA will have to open up further, exposing (generating) more single-stranded DNA. Keep in mind that this process is moving, through independent complexes, in both directions along a DNA molecule. Because the polymerization reaction only proceeds by 3' addition, as new single stranded regions are opened new primers must be created by RNA primase and then extended by DNA polymerase; these are known as the lagging strands. While there are two leading strands leaving a particular DNA replication start site, there are a number of lagging strands involved. If you try drawing what this looks like, you will realize that i) this process is asymmetric in relation to the start site of replication; ii) the process generates RNA-DNA hybrid molecules; and iii) that eventually an extending DNA polymerase will run into the RNA primer part of an "upstream" molecule. However, keep in mind, RNA regions, derived from the primers, are not found in "mature" DNA molecules, so there must be a mechanism that removes them. As it turns out, the DNA polymerase complex, like a number of other enzyme systems, contains more than one catalytic activity (analogous to the ATP synthase and pump), something that we will return to further on. When the DNA polymerase complex reaches the upstream nucleic acid chain it runs into an RNA containing region; an RNA exonuclease activity associated with the DNA polymerase complex removes the RNA nucleotides and replaces them with DNA nucleotides using the existing DNA strand as the primer. Once the RNA portion is removed, a DNA ligase activity acts to join (generate a covalent phosphodiester bond between) the two DNA molecules. These reactions, driven by nucleotide hydrolysis, end up producing a continuous DNA strand that runs from one end of the chromosome to the other, or in circular chromosomes, all the way around the circle. For a dynamic, and largely deterministic, look at the process check out this FLASH webpage (link)(→); it is clearer than more realistic simulations in part because it is "flat" and the proteins involved are not shown in order to reduce the complexity of the process.
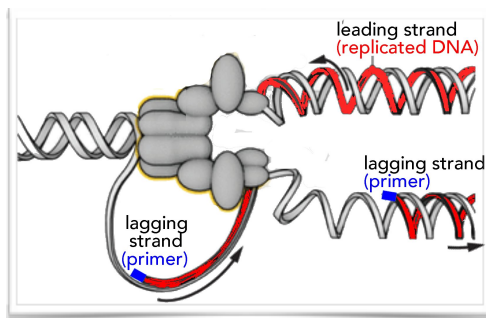
Evolutionary considerations: At this point you might well ask yourself, why (for heavens sake) is the process of DNA replication so complex. Why not use a DNA polymerase that does not need an RNA primer, or any primer for that matter? That should be possible, particularly given that RNA polymerase

does not need a primer. Why not have polymerases that can add nucleotides equally well to either end of a polymer? That such a mechanism is possible is suggested by the presence of enzymes in eukaryotic cells that can catalyze the addition of a nucleotide to the 5' end of an RNA molecule, the 5' capping reaction associated with mRNA synthesis that we will considered later on. But while apparently possible, such activities are not used in DNA replication in any known organism. The real answer to why DNA replication is as complex as it is is that we are not sure why. It could be its complexity is an evolutionary relic, based on a process established within the last common ancestor of all organisms and extremely difficult or impossible to change through evolutionary mechanisms, or simply not worth the effort, in terms of its effects on reproductive success. Alternatively, there could be strong selective advantages associated with the system that preclude such changes. What is clear is that this is how the system appears to function in all known organisms. For practical purposes, we will have to remember some of the key details involved, these include the direction of polymer synthesis (3' addition) and the need (in the case of DNA synthesis) for an RNA primer.

Replication machines

We have presented DNA replication (the same, apparently homologous process is used in all known organisms) in as conceptually simple terms as we can, but it is important to keep in mind that the actual machinery involved is complex. In part this complexity arises because the process is topologically constrained and needs to be highly accurate. In the bacterium *Escherichia coli* over 100 genes are involved in DNA replication and repair. To insure that replication is controlled and complete, replication begins at specific sequences along the DNA strand, known as origins of replication or origins for short. Origin DNA sequences are recognized by specific DNA binding proteins. The binding of these proteins initiates the assembly of an origin recognition complex, an ORC. Various proteins then bind to the DNA to locally denature (unwind and separate) and block the single strands from reannealing. This leads to the formation of what is known as a replication bubble. Multiprotein complexes, known as a replication fork, assemble on the two DNA strands. Using a single replication origin and two replication forks, moving in opposite directions, a rapidly growing *E. coli* cell can replicate its ~4,700,000 base pairs of DNA, which are present in the form of a single circular DNA molecule, in ~40 minutes. Each replication fork moves along the DNA adding ~1000 base pairs of DNA per second to the newly formed DNA polymer. While a discussion of the exact mechanisms involved is beyond our scope here, it is critical that DNA is complete before a cell attempts to divide - this implies that there are signaling systems within the cell that can be used to monitor the completion of DNA replication and to initiate the start of cell division. We will find such systems in a number of cellular processes. In many bacteria, the signaling system is based on the fact that the chromosome is circular, that DNA replication begins at a single site (the origin), and that replication forks collide with one another in a region of the chromosome known as the terminus.[279] We will consider eukaryotic processes (known generically as checkpoints) later on.

[279] Synchronization of Chromosome Dynamics and Cell Division in Bacteria

Questions to answer

128. Draw a diagram of the key steps in the replication of a circular DNA molecule. How might you adapt this system to replicate much longer linear molecules?

129. What key, non-deducible features of DNA replication do you need to remember (memorize) and why.

## Accuracy and error in DNA synthesis

DNA synthesis (replication) is a highly accurate process; the DNA-dependent DNA polymerase makes about one error for every 10,000 bases it adds. But that level of error would almost certainly be highly deleterious, and in fact most of these errors are quickly recognized as mistakes. To understand how, remember that correct AT and GC base pairs have the same molecular dimensions, that means that incorrect AG, CT, AC, and GT base pairs are either too long or too short. By responding to base pair length, molecular machines recognize a mistake in base pairing as a structural defect in the DNA molecule. When a mismatched base pair is formed and recognized, the DNA polymerase stops forward synthesis, reverses its direction, and removes the region of the DNA containing the mismatched base pair using a "DNA exonuclease" activity. It then resynthesizes the region, (hopefully) correctly. This process is known as proof-reading; the proof-reading activity of the DNA polymerase complex reduces the total DNA synthesis error rate to ~1 error per 1,000,000,000 (10^9) base pairs synthesized.

At this point let us consider nomenclature, which can seem arcane and impossible to understand, but in fact obeys reasonably straightforward rules. An exonuclease is an enzyme that can bind to the free end of a nucleic acid polymer and remove nucleotides through a hydrolysis reaction of the phosphodiester bond ($\rightarrow$). A 5' exonuclease cuts off a nucleotide located at the 5' end of the molecule, a 3' exonuclease, cuts off a nucleotide located at the molecule's 3' end. An intact circular nucleic acid molecule is immune to the effects of an exonuclease. To break the bond between two nucleotides in the interior of a nucleic acid molecule (or in a circular molecule, which has no ends), one needs an endonuclease activity.
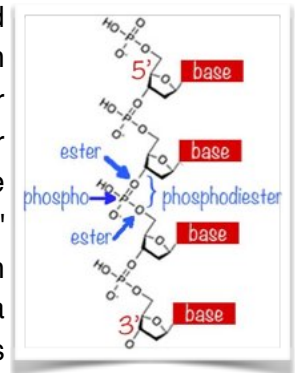
As you think about the processes involved, you come to realize that once DNA synthesis begins, it is important that it continues without interruption. But the interactions between nucleic acid chains are based on weak H-bonding interactions, and the enzymes involved in the DNA replication process can be expected to dissociate from the DNA because of the effects of thermal motion, imagine the whole system jiggling and vibrating – held together by relatively weak interactions. We can characterize how well a DNA polymerase molecule remains productively associated with a DNA molecule in terms of the number of nucleotides it adds to a new molecule before it falls off; this is known as its "processivity". So if you think of the DNA replication complex as a molecular machine, you can design ways to insure that the replication complex has high processivity, basically by keeping it associated with the DNA. One set of such machines is the polymerase sliding clamp - in this system, the DNA polymerase complex is held onto the DNA by a doughnut shaped protein, known as a sliding clamp, that encircles the DNA double helix and is strongly bound to the DNA polymerase (video link). So the question is, how does a protein come to encircle a DNA molecule? The answer is that the clamp protein is added to DNA by another protein molecular
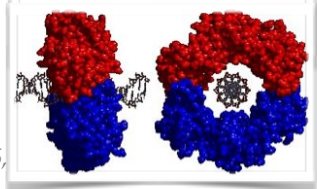
machine known as the clamp loader (→).[280] Once closed around the DNA the clamp can move freely along the length of the DNA molecule, but it cannot leave the DNA. The clamp's sliding movement along DNA is diffusive – that is, it is driven by collisions with other molecules, with the average strength of such collisions related to the temperature of the system. Its movement is given a direction because the clamp is attached to the DNA polymerase complex which is adding monomers to the 3' end of the growing nucleic acid polymer. This moves the replication complex (inhibited from diffusing away from the DNA by the clamp) along the DNA in the direction of synthesis. Processivity is increased since, in order to leave the DNA the polymerase has to disengage from the clamp or the clamp as to be removed by the clamp loader acting in reverse, that is, acting as an unloader.

Further replication complexities in eukaryotes: telomeres

The DNA molecules found in bacteria and archaea are circular; there have no free ends.[281] Eukaryotic cells can contain more than 1000 times the DNA found in a typical bacterial cell. Instead of circles, they contain multiple linear molecules that form the structural basis of their chromosomes (we will consider the details later on). The free ends of the chromosomes are known as telomeres. The linearity of eukaryotic chromosomes creates problems replicating the ends of the DNA molecules. Left alone, more and more of the lagging strand end of the chromosome would go unreplicated, the end of the chromosome would begin to disappear with each DNA replication cycle. To address this "design limitation" in the DNA-dependent, DNA polymerase system eukaryotes use another RNA-protein complex, known as telomerase.[282]

Telomeres have a repeated sequence; in the case of human (and all other vertebrates) chromosomes end in repeated copies of the sequence TTAGGG-3' (→). The RNA part of the telomerase enzyme is the product of the TERC gene (OMIM:602322); it combines with the protein product of the TERT gene (OMIM: 187270).[283] The TERC RNA contains a sequence complementary to the telomere DNA sequence and serves as the template for the synthesis of GGTTAG from the 3' end of the telomere's lagging strand - this process can occur multiple times, after which the primase and DNA-dependent, DNA polymerase can fill in the telomere end. A further discussion of the role of telomeres and telomerase is beyond this course.[284]

---
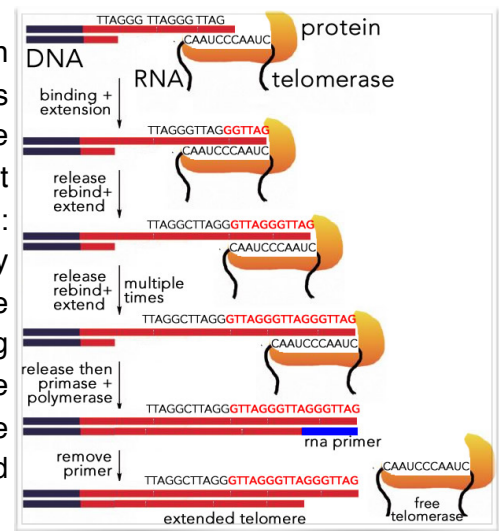
[280] see Clamp loader ATPases and the evolution of DNA replication machinery & DNA Clamp & Clamp Loader video

[281] The mitochondria and chloroplasts of eukaryotic cells also contain circular DNA molecules, another homology with their ancestral bacterial parents. ,

[282] http://en.wikipedia.org/wiki/Telomerase

[283] You can explore the known genetic diseases by using the web based On-line Mendelian Inheritance in Man (OMIM) database: http://www.ncbi.nlm.nih.gov/omim/

[284] more on telomerase: http://blogs.scientificamerican.com/guest-blog/aging-too-much-telomerase-can-be-as-bad-as-too-little/

The circular nature of prokaryotic chromosomes creates its own issues, issues based on molecular topology. After replication, the two double-stranded DNA circles are linked together. Long linear DNA molecules can also become knotted together within eukaryotic cells. In addition, the replication of DNA unwinds the DNA, and this unwinding leads to what is known as the supercoiling of the DNA molecule. Left unresolved, supercoiling and knotting will inhibit the separation of replicated strands and DNA synthesis (perhaps you can explain why).[285] These topological issues are resolved by enzymes known as topoisomerases, because they can interconvert topologically distinct versions of the same molecule. There are two generic types of topoisomerases that act on DNA (→). Type I topoisomerases bind to the DNA, catalyze the breaking of a single bond in one sugar-phosphate-sugar backbone, and allow the release of overwinding through rotation around the bonds in the intact chain. When the tension is released, and the molecule has returned to its "relaxed" form, the enzyme catalyzes the reformation of the broken bond. Both bond breaking and reformation are coupled to ATP hydrolysis. Type II topoisomerases (→) are involved in "unknotting" DNA molecules. These enzymes bind to the DNA, catalyze the hydrolysis of both backbone chains, but hold on to the now free ends. This allows another strand to "pass through" the broken strand. The enzyme also catalyzes the reverse reaction, reforming the bonds originally broken.
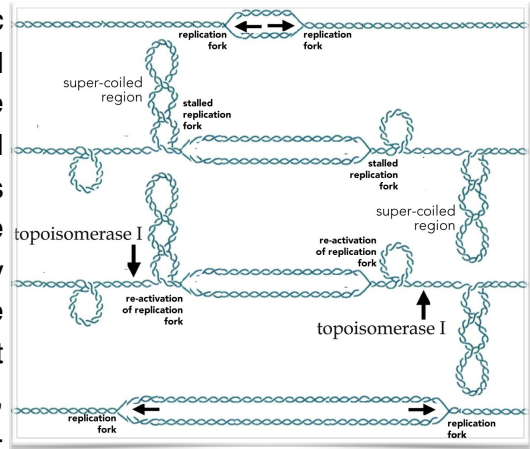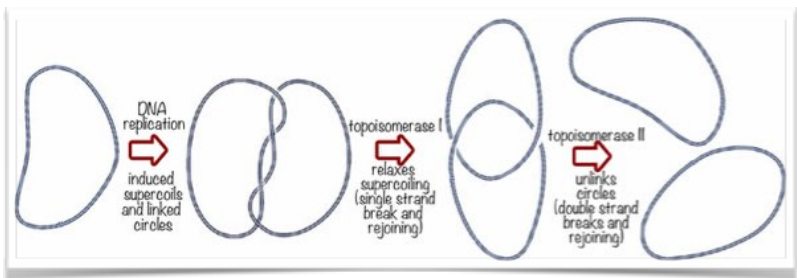
In addition to having typically much more DNA, the eukaryotic DNA replication enzyme complex is much slower, about 1/20th as fast as the prokaryotic system. While a bacterial cell can replicate its circular ~3 x 10^6 base pair chromosome in ~1500 seconds using a single origin of replication, the replication of the billions of base pairs of a typical eukaryote's DNAs involves the use of multiple (many) origins of replication, scattered along the length of each chromosome. So what happens when replication forks collide with one another? In the case of a circular DNA molecule, with its single origin of replication, the replication forks resolve in a specific region known as the terminator. At this point type II topoisomerase allows the two circular DNA molecules to disengage from one another and move to opposite ends of the cell. The cell division machinery forms between the two DNA molecules. The system in eukaryotes, with their multiple linear chromosomes, is much more complex, although topoisomerases are still involved in separating replicated chromosomes, and involves a more complex molecular machines that we will return to later, specifically in the complex of sexual reproduction (meiosis).

---

[285] see this video on DNA supercoiling and topoisomerases: http://youtu.be/EYGrEIVyHnU

Questions to answer

130. During DNA/RNA synthesis what is the average ratio of productive to unproductive interactions between an incoming nucleotide and the polymerase?
131. What are topological isomers?
132. Why do you need to denature (melt) the DNA double-helix to copy it?
133. How would DNA replication change if H-bonds were as strong as covalent bonds?
134. List all of the unrealistic components in the DNA replication video: http://bcove.me/x3ukmq4x

Questions to ponder:

- How would evolution be impacted if DNA were totally stable and DNA replication was error-free?
- Draw a diagram to explain how the DNA polymerase might recognize a mismatched base pair.
- What would be the impact of mutations that altered the proof-reading function of the DNA polymerase complex?
- How might mutations in the genes encoding the clamp/clamp-loader system influence DNA replication?

Mutations, deletions, duplications & repair

While DNA is used as the universal genetic material of organisms, it is worth remembering that DNA is a thermodynamically unstable molecule. Eventually it will breakdown into more stable and dramatically simpler components; and as it decomposes the information stored within its sequence will be lost. For example, at a temperature of ~13ºC, half of the phosphodiester bonds in a DNA sample will break after ~520 years.[286] But there is more. For example, cytosine groups within the DNA molecule can react with water, which (you might remember) is present at a concentration of ~54 M inside a cell. This leads to a deamination reaction that transforms cytosine into uracil (→). If left unrepaired the original CG base pair will be replaced by an AU base pair in one strand during DNA synthesis. But, uracil is not normally found in DNA and its presence will be recognized by an enzyme that severs the bond between the uracil moiety and the deoxyribose group.[287] The absence of a base, due either to its spontaneous loss or its enzymatic removal, acts as a signal for another enzyme system, the Base Excision Repair complex that removes a section of the DNA strand with the missing base.[288] A DNA-dependent DNA polymerase can then bind to the open DNA and use the existing strand as a primer and the undamaged strand as a template to fill in the gap. Finally, another enzyme (a DNA ligase) joins the newly synthesized segment to the pre-existing strand. In the human genome there are over 130 genes devoted to repairing damaged DNA.[289]
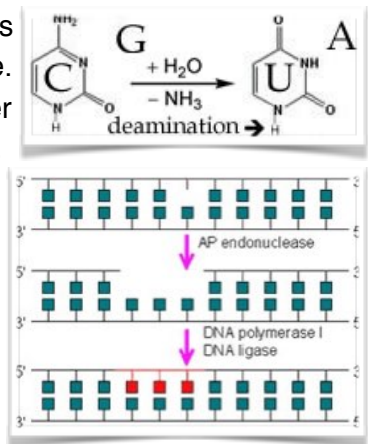
Other hydrolysis reactions include depurination: the loss of an cytosine or thymine group and depyrimidination: the loss of an adenine or guanine group, lead to the removal of a base from the DNA. The rates of these reactions increases at acidic pH, which is probably one reason that the cytoplasm is

---

[286] Here is the paper from which statement is derived: http://www.nature.com/news/dna-has-a-521-year-half-life-1.11555

[287] UNG: uracil-DNA-N-glycosidase http://omim.org/entry/191525

[288] absent purine/absent pyrimidine endonuclease http://omim.org/entry/300773

[289] Human DNA Repair Genes – video with lots of misspelled words here: http://youtu.be/g4khROaOO6c

not acidic. How frequent are such events? A human body contains ~10^{14} cells. Each cell contains about ~10^9 base pairs of DNA. Each cell, whether it is dividing or not, undergoes ~10,000 base loss events per day or ~10^{18} events per day per person. That's a lot! The basic instability of DNA and the lack of repair after an organism dies means that DNA from dinosaurs, the last of which went extinct ~65,000,000 years ago, has disappeared from the earth, making it impossible to clone (or resurrect) a true dinosaur.[290] In addition DNA can be damaged by environmental factors, such as radiation, ingested chemicals, and reactive compounds made by the cell itself. Many of the most potent mutagens known are natural products, often produced by organisms to defend themselves against being eaten or infected by parasites, predators, or pathogens.[291]

A step back before going forward: what, exactly, is a gene anyway?

Now that we have introduced you to DNA and have casually referred to genes multiple times in various contexts, it is probably well past time that we serious consider exactly what we mean by a gene.[292] Each organism (cell) carries is genomic DNA, which it replicates when it divides to produce an offspring. The DNA molecules (the genomes) of those organisms that survive and produce offspring become more frequent within a population than the genomes of those organisms that fail to reproduce to the same extent (or at all). As DNA is replicated and maintained within a cell, mutations arise. These mutations can influence the reproductive success of an organism. Over time this process (natural selection) leads to changes in the genomes of a population. When populations split into two (or more), their DNA molecules start changing independently of one another.

From a theoretical perspective there are two types of changes that can occur within a DNA molecule, those that influence the probability of reproductive success and those that do not. Those that influenced reproductive success can have either a positive or negative impact; over time they can become more frequent within the population, they are said to be under positive (beneficial) selection or they can become less frequent within the population, in which case they are said to be under negative (detrimental) selection. Again, beneficial and detrimental apply not to the well being of the individual who carries these changes (mutations) but rather on its reproductive success. In asexual organisms, without complicating processes like horizontal gene transfer (which we will consider later on), the changes (mutations) that have no effect on reproductive success are known as neutral mutations. They can be seen as a kind of molecular clock. If we count the number of neutral changes in the genome sequences of two isolated populations (or organisms) we can use that information to estimate how long ago they shared a common ancestor. Of course this is not a particularly good clock in that there are only three possible changes a mutation that alters a single position in a genomic DNA molecule can make, a mutation that leads to what is known as a single nucleotide polymorphism or SNP (pronounced "snip"). For example if the original base is an A, it can change to a C, G, or T - if it changed to an A, we would not be able to tell. Of course, that changed base could itself change; for example, if a A changed

---

[290] DNA has a 521-year half-life: http://www.nature.com/news/dna-has-a-521-year-half-life-1.11555

[291] Dietary carcinogens, environmental pollution, and cancer: some misconception

[292] Part of the issue here goes to the continuity of life and its long history. We are always considering living systems that contain a range of molecules and reactive systems derived from its immediate ancestor - there simply is no easy "starting off point".

to a C, the C could change to an A, T, or G. BUT, if it changes to an A, we could not tell whether it had changed at all. Over long periods of time, the ability to date the divergence between organisms using neutral mutations begins to lose resolution - a situation known as "long branch attraction".

Ah, but how do we know that a genomic change is neutral or subject to positive or negative selection? To begin to answer these questions, we need to know what mutations can do to a gene, and what changing a gene can do to reproductive success. The answers to these question are complex, but the path to such answers begins with recognizing what is stored in genomic DNA - namely information. Mutation, selection, and other evolutionary processes can add and remove information from the genome. Depending upon the circumstances, a mutation can have a positive or negative effects on reproductive success.

We can recognize changes (mutations) that give rise to a measurable change in phenotype as influencing what we will call genes. There are many genes in an organism, originally identified by the phenotypes they produced. In a completely over-simplified view (we will get more real later on) we find that a mutation in a particular region along a DNA molecule produces a similar or related phenotype. In some cases it was clear that a mutation alter the presence or activity of a particular enzyme, which led George Beadle (1903-1989) to put forward the one gene one protein (enzyme) model.[293] After awhile it became clear that many proteins are composed of the products of multiple genes, an example would be telomerase. Some genes encode RNAs that are used directly (e.g. the TERC gene) and some encode RNAs that are used to direct the synthesis of a polypeptide, such as TERT, while others encode RNAs that regulate the expression of genes. Understanding these interactions and their impact on the behavior of biological systems will be considered in detail in the second half of the course.

As we will see, and as you might probably already know, genes can be divided roughly into two domains: these are the regulatory and the transcribed domains of the DNA. Mutations (changes in DNA sequence) in the regulatory domains generally influence where, when, and how many RNAs are synthesized from the transcribed domain. You will note that we have not mentioned where these regions are with respect to one another. As we will consider in greater detail later on, genes can overlap with one another and defining all of the regulatory regions of a gene can be challenging, particularly since different regulatory regions may be used in the different cell types present within a multicellular organism. A gene's regulatory regions can span many thousands of kilobases of DNA and be located upstream, downstream, or within the gene's coding region. In addition, because DNA is double stranded, one gene can be located on one strand and another, completely different gene can be located on the anti-parallel strand. We will return to the mechanisms of gene regulation later on, but as you probably have discerned, gene regulation is complex and often the subject of its own course.

Transcribed domains can also be complex, particularly in eukaryotic genes: a single gene can produce multiple, functionally distinct gene products through the process known as RNA splicing.[294] How differences in gene sequence influence the activity and role(s) of a gene is not simple. A critical point to keep in mind is that a gene has meaning only in the context of a cell or an organism. Change the organism and the same, or rather, more accurately put, homologous genes (that is genes that share a common ancestor, a point we will return to) can have different roles.

_____

[293] One gene one protein & One gene one enzyme

[294] Expansion of the eukaryotic proteome by alternative splicing

Alleles, their origins and their impact on evolution

Once we understand that a gene corresponds to a specific sequence of DNA, we understand that different versions of a gene, known as alleles, correspond to genes with different sequences. Two alleles of the same gene can differ from one another by as little as a difference at one, out of thousands of nucleotide position or at many positions. In some cases, the differences between alleles can include deletions and duplications in the sequence. An allele can produce a gene product with completely normal function or no remaining functional activity at all, referred to as a null or amorphic allele. It can have less function than the "wild type" allele (hypomorphic), more function than the wild type (hypermorphic), or a new function (neomorphic). Given that many gene products function as part of multimeric complexes that are the products of multiple genes and that many organisms (like us) are diploid, there is one more possibility, the product of one allele can antagonize the activity of the other - this is known as an antimorphic allele. These different types of alleles were defined genetically by Herbert Muller, who won the Nobel prize for showing that X-rays could induce mutations, that is, new alleles.[295] The functional characterization of an allele is typically carried out with respect to how its presence influences a specific trait(s). Again, remember that most traits are influenced by multiple genes, and a single gene can influence multiple traits and processes.

The most common version of an allele is often referred to as the wild type allele ($\rightarrow$), but that is really just because it is the most common. There are often multiple alleles of a particular gene in the population and they all may be equally normal, that is have similar effects on reproductive success and in terms of the phenotypes they produce. If there is no significant selective advantage between them, their relative frequencies within a population will drift. At the same time, the phenotype associated with a particular allele can be influenced by which alleles are present at other genetic loci, known as the genetic background. Since most traits are the results of hundreds or thousands of genes functioning together, and different combinations of alleles can produce different effects, the universe of variation is large. This can make identifying the genetic basis of a disease difficult, particularly when variation at any one locus may make only a minor contribution to the disease phenotype. On top of that, environmental and developmental differences can outweigh genetic influence on phenotype. Genetic background effects can lead to a particular allele producing a disease in one person and not another.[296]

Mutations are the ultimate source of genetic variation – without them evolution would not occur. Mutations can lead to a number of effects, in particular, they can create new activities. At the same time these changes may reduce the original (and necessary) activity of an important gene. Left unresolved such molecular level conflicts would greatly limit the flexibility of evolutionary mechanisms. For example, it is common to think of a gene (or rather the particular gene product it encodes) as having one and only one function or activity, but in fact, when examined closely many catalytic gene products (typically proteins) can catalyze "off-target" reactions or carry out, even if rather inefficiently, other activities - they interact with other molecules within the cell and the organism. Assume for the moment

---

[295] Muller's morphs: https://en.wikipedia.org/wiki/Muller's_morphs

[296] Genetic background effects: https://www.sciencedaily.com/releases/2015/07/150716135104.htm

that a gene encodes a gene product with an essential function as well as potentially useful (from a reproductive success perspective) activities. Mutations that enhance these "ancillary functions" will survive (that is be passed on to subsequent generations) only to the extent that they do not (overly) negatively influence the gene's primary and essential function. The evolution of ancillary functions may be severely constrained or blocked altogether.

This problem can be circumvented based on the fact that the genome is not static (→). There are molecular level processes through which regions of DNA (and the genes that they contain) can be deleted, duplicated,
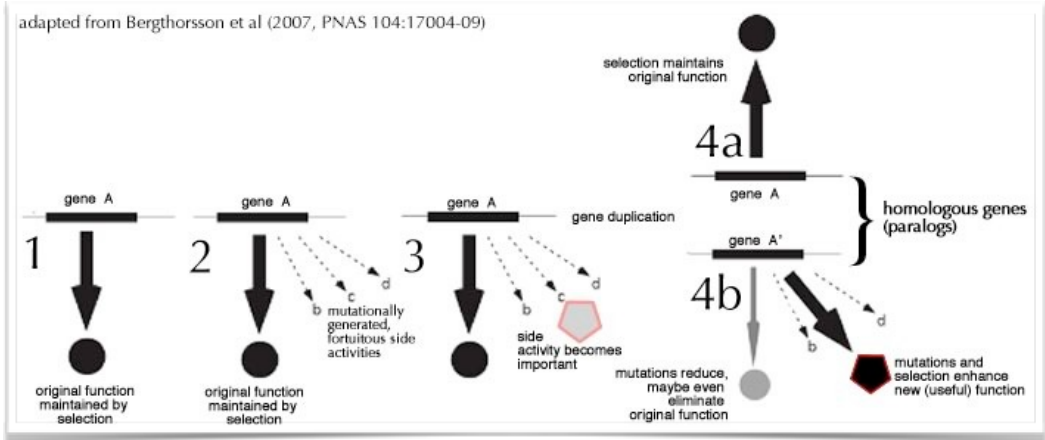

and moved from place to place within the genome. Such genomic rearrangements, which are mutations, may occur continuously during embryonic development. The end result is that while most of the cells in your body have very similar genomes (perhaps consisting of single base pair changes that arose during DNA replication), some have genomes with different arrangements of DNA. These differences can include deletions, duplications, and translocations, moving a region of DNA from one place to another in the genome. Not all cells in your body will have exactly the same genome.[297]

In the case above illustrated in the figure, imagine that the essential but multifuntional gene is duplicated. Now one copy can continue to carry out its essential function, while the second is free to change. While many mutations will negatively effect the duplicated gene, some might increase and refine its favorable ancillary function. A new trait can emerge freed from the need to continue to perform its original (and essential) function. We see evidence of this type of process throughout the biological world. When a gene is duplicated, the two copies are known as paralogs. Such paralogs often evolve independently.

DNA repeat diseases and genetic anticipation

While they are essential for evolution, defects in DNA synthesis and genomic rearrangements more often lead to genetic (that is inherited) diseases than to any benefit to an individual. While we will return to many mutational mechanisms and their their effects as we continue, here we briefly consider diseases associated with DNA replication, specifically the class of genetic diseases known as the trinucleotide repeat disorders (→). There are a number of such "triplet repeat" diseases, including several forms of mental retardation, Huntington's

---
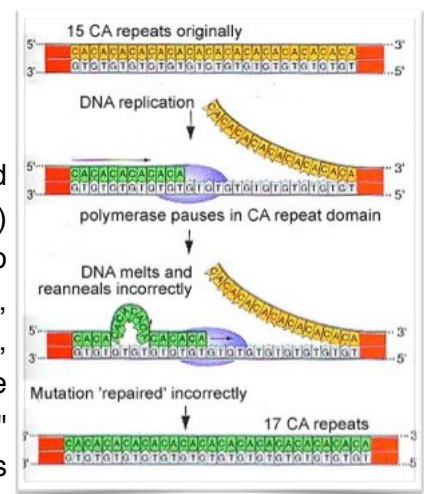
297 Copy Number Variation in Human Health, Disease, and Evolution and LINE-1 retrotransposons: mediators of somatic variation in neuronal genomes?

disease, inherited ataxias, and muscular dystrophy. These diseases are caused by slippage of DNA polymerase and the subsequent duplication of sequences. When these "slippable" repeats occur in a region of DNA encoding a protein, they can lead to regions of repeated amino acids. For example, expansion of a domain of CAGs in the gene encoding the polypeptide Huntingtin (OMIM:613004) causes the neurological disorder Huntingdon's chorea.[298]

A mechanistically related pathogenic syndrome is known as Fragile X (OMIM:300624); the underlying DNA replication defect is the leading form of autism of known cause (most forms of autism have no known cause). About 6% of autistic individuals have fragile X. Fragile X can also lead to anxiety disorders, attention deficit hyperactivity disorder, psychosis, and obsessive-compulsive disorder. Because the mutation involves the FMR-1 gene (OMIM:309550), which is located on the X chromosome, the disease is sex-linked and effects mainly males, who are XY, compared to XX females. In the unaffected population, the FMR-1 gene contains between 6 to 50 copies of a CGG repeat. Individuals with between 6 to 50 repeats are phenotypically normal. Those with 50 to 200 repeats carry what is known as a pre-mutation; these individuals rarely display symptoms but can transmit the disease to their children. Those with more than 200 repeats typically display symptoms and often have what appears to be a broken X chromosome – from which the disease derives its name. The pathogenic sequence in Fragile X is downstream of the FMR1 gene's coding region. When this region expands, it inhibits the expression of the FMR1 gene.[299] There are a number of processes that can mediate the pathogenic effects DNA repeat diseases, some of which we will consider when we discuss the inheritance of these conditions.

Other DNA Defects: Defects in DNA repair can lead to severe diseases and often a susceptibility to cancer. A OMIM search for DNA repair returns 654 entries! For example, defects in mismatch repair lead to a susceptibility to colon cancer, while defects in translation-coupled DNA repair are associated with Cockayne syndrome. People with Cockayne's syndrome (OMIM:216400 & 133540) are sensitive to light, short and appear to age prematurely.[300]

Our introduction to genes has necessarily been quite foundational and we will extend it in the second half of the course. There are lots of variations and associated complexities that occur within the biological world. The key ideas are that genes represent biologically meaningful DNA sequences. To be meaningful, the sequence must play a role within the organism, typically by encoding a gene product (which we will consider next) and/or the information needed to insure its correct expression" that is, where and when the information in the gene is used. A practical problem is that most studies of genes are carried out using organisms grown in the lab or in otherwise artificial or unnatural conditions. It might be possible for an organism to exist with an amorphic mutation in a gene in the lab, whereas organisms that carry that allele may well be at a significant reproductive disadvantage in the real world. Moreover, a particular set of alleles, a particular genotype, might have a reproductive advantage in one environment (one ecological/behavioral niche) but not another. Measuring these effects can be difficult.

---

[298] You will probably want to learn how to use the On-line Mendelian Inheritance in Man (OMIM) to explore various disease and their genetic components. OMIM is a part of PubMed: http://www.ncbi.nlm.nih.gov/pubmed

[299] Molecular mechanisms of fragile X syndrome: a twenty-year perspective.

[300] Cockayne syndrome: http://omim.org/entry/278760

All of which should serve as a warning to skeptically consider pronouncements that a gene, or more accurately a specific allele of a gene, is responsible for a certain trait, particularly if the trait is complex, ill-defined, and likely to be significantly influenced by genomic context (the rest of the genotype) and environmental factors. Intelligence is one such complex trait. A dramatic example of the difficulty in defining a gene product's functions is illustrated by the studies of Hutchinson et al; they produced a minimal bacterial genome containing 473 genes.[301] Of these 473 genes, the function(s) of 149 (~32% of the total genome) was unknown, a rather surprising result.

Questions to answer

135. During DNA/RNA synthesis what is the average ratio of productive to unproductive interactions between nucleotides and the polymerase?
136. How does a mutation generate a new allele? How is a mutation different from an allele?
137. What would be a reasonable way to determine that you had defined an entire gene?
138. Is it possible to build a system (through evolutionary mechanisms) in which mutations do not occur?

Questions to ponder:

- How could removing information from the genome enhanced reproductive success?
- Outline a strategy to approach defining the function of a "gene with unknown function"?

---

[301] Design and synthesis of a minimal bacterial genome. https://www.ncbi.nlm.nih.gov/pubmed/27013737

Chapter 8: Peptide bonds, polypeptides, proteins, and molecular machines

In which we consider the nature of proteins, how they are synthesized and assembled, how they get to where they need to go within the cell and within the organism, how they function, how their activities are regulated, and how mutations can influence their expression, stability, activity, and evolution.
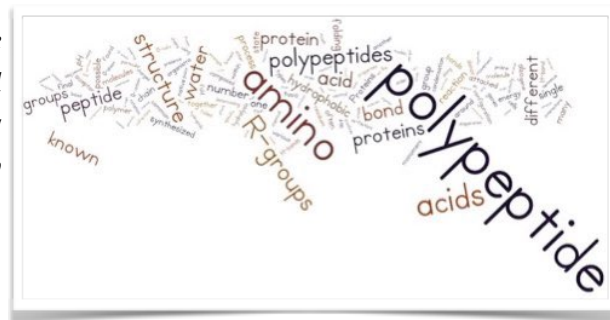
We mentioned proteins many times, since there are few biological processes that do not rely on them. Proteins act as structural elements, signals, regulators, and catalysts in a wide arrange of molecular machines. Up to this point, however, we have not said much about what they are, how they are made, and how they come to do what they do. The first scientific characterization of what are now known as proteins was published by the Dutch chemist, Gerardus Johannes Mulder (1802–1880).[302] After an analysis of a number of different substances, he proposed that all proteins contain a common chemical core, with the molecular formula $C_{400}H_{620}N_{100}O_{120}P_1S_1$, and that the differences between different proteins were primarily in the numbers of phosphate (P) and sulfur (S) atoms they contained. The name "protein", from the Greek word πρώτα ("prota"), meaning "primary", was suggested by the Swede, Jons Jakob Berzelius (1779–1848) based on the presumed importance of these compounds in biological systems.[303] As you can see, Mulder's molecular formula was not very informative, it tells us little or nothing about protein structure, but suggests that all proteins are fundamentally similar, which is confusing since they carry out so many different roles. Subsequent studies revealed that proteins could be dissolved in either water or dilute salt solutions but aggregated and became insoluble when the solution was heated; as we will see this aggregation reaction reflects a change in the structure of the protein. Mulder was able to break down proteins into amino acids through an acid hydrolysis reaction. Amino acids get their name from the fact that they contain both an amino (–NH$_2$) and a carboxylic acid (–COOH) group. While there many thousands of possible amino acids, only twenty (or rather twenty two - see below) different amino acids could be identified in hydrolyzed samples of proteins. Since their original characterization as a general class of compounds, we now understand that while proteins share a common basic polymer structure, they are remarkably diverse. Proteins are involved in roles from the mechanical strengthening of skin, the building of shells and claws to the regulation of genes, to the transport of oxygen, to the capture of energy, to the release of light, to the catalysis and regulation of essentially all of the chemical reactions that occur within cells and organisms.

While all proteins have a similar bulk composition, this obscures rather than illuminates their dramatic structural and functional differences. With the introduction of various chemical methods, it was discovered that different proteins were composed of distinct and specific sets of subunits, and that each

---

[302] From 'protein' to the beginnings of clinical proteomics: http://www.ncbi.nlm.nih.gov/pubmed/21136729

[303] While historically true, the original claim that proteins get their name from "the ancient Greek sea-god Proteus who, like your typical sea-god, could change shape. The name acknowledges the many different properties and functions of proteins." seems more poetically satisfying to us.

subunit is an unbranched polymer with a specific amino acid sequence. Because the amino acids in these polymers are linked by what are known as peptide bonds, the polymers are known generically as polypeptides. At this point, it is important to reiterate that proteins are functional objects, and specific proteins are composed of specific sets of distinct polypeptides; moreover, each distinct polypeptide is encoded by a distinct gene. In addition to polypeptides many proteins also contain other molecular components, known as co-factors or prosthetic groups (we will call them co-factors for simplicity's sake.) These co-factors can range from metal ions to various small molecules. A protein is a fully assembled and functional entity.

As you might remember from your chemistry courses carbon atoms (C) typically form four bonds. We can think of an amino acid as a (highly) modified form of methane (CH_4), with the C referred to as the alpha carbon (C_α). Instead of four hydrogens, in a biological amino acid there is an H, an amino group (-NH_2), an carboxylic acid group (-COOH), and a final, variable (R) group attached to the central C_α atom. The four groups attached to the α-carbon are arranged at the vertices of a tetrahedron (→). If all four groups attached to the α-carbon are different from one another, as they are in all biological amino acids except glycine, the resulting amino acid can exist in two possible forms, known as enantiomeric stereoisomers. Enantiomers are mirror images of one another and are referred to as the L- and D- forms. Only L-type amino acids are found in proteins, even though there is no obvious chemical reason that proteins could not have also been made using both types of amino acids or using only D-amino acids.[304] It appears that the universal use of L-type amino acids in the polypeptides found in biological systems is another example of the evolutionary relatedness of organisms, it appears to be a homologous trait, presumably established in the last universal common ancestor (LUCA). Similarly, even though there are hundreds of different amino acids known, only 22 (these include the 20 common amino acids and two others, selenocysteine and pyrrolysine) are found in proteins and presumably were present in LUCA.
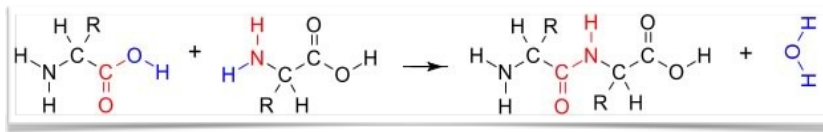
Amino acids differ from one another by their R-groups, which are often referred to as "side-chains". Some of these R-groups are large, some are small, some are hydrophobic, some are hydrophilic, some of the hydrophilic R-groups contain weak acidic or basic groups. The extent to which these weak acidic or basic groups are positively or negatively charged changes in response to environmental pH. Changes in charge will (as we will see) influence the structure of the polypeptide/protein in which they find themselves. The different R-groups provide proteins with a broad range of chemical properties, which are further extended by the presence of co-factors.[305]

As we noted for nucleic acids, a polymer is a chain of subunits. In the case of a polypeptide, amino acid monomers are linked together by peptide bonds. Under the conditions that exist inside the cell, this is a thermodynamically unfavorable dehydration reaction, and so polypeptide synthesis must be

---

[304] It is not that D-amino acids do not occur in nature, or in organisms, they do. They are found in biomolecules, such as the antibiotic gramicidin, which is composed of alternating L-and D-type amino acids - however gramicidin is synthesized by a different process than that used to synthesize proteins.

[305] Bioengineers are working to go Beyond the Canonical 20 Amino Acids: Expanding the Genetic Lexicon & to incorporation of non-canonical amino acids into proteins in yeast; something made possible due to the redundancy of the genetic code.

coupled to a thermodynamically favorable reaction. A molecule formed from two amino acids, joined together by a peptide bond, is known as a dipeptide. As in the case of each amino acid, the dipeptide has an N-terminal (amino) end and a C-terminal (carboxylic acid) end. To generate a polypeptide, new amino acids are added sequentially (and only) to the C-terminal end of the polymer – a reaction analogous to the synthesis of a polynucleotide, with addition of monomers to one end of the growing polymer. A peptide bond forms between the amino group of the added amino acid and the carboxylic acid group of the polymer; the formation of a peptide bond is associated with the release of a water molecule (↓). As you might suspect, this is a thermodynamically unfavorable reaction, so it is coupled to a favorable reaction, a nucleotide triphosphate hydrolysis reaction. When complete, the polypeptide synthesis reaction generates a new C-terminal carboxylic acid group. It is important to note that while some amino acids have a carboxylic acid group as part of their R-groups, new amino acids are not added there. Because of this fact, polypeptides are synthesized as unbranched, linear polymers. The process of amino acid addition can continue, theoretically without limit. Biological polypeptides range from the very short (5-10) to very long (many hundreds to thousands) of amino acids in length.[306] For example, the protein Titin (found in muscle cells) can be more than 30,000 amino acids in length.[307] Because there is no theoretical constraint on which amino acids occurs at a particular position within a polypeptide, there is a enormous number of possible polypeptides that can exist. In the case of a 100 amino acid long polypeptide, there are 20^{100} possible different polypeptides that could, in theory, be formed.

Questions to answer:
139. How does a polypeptide chain resemble and how does it differ from a nucleic acid molecule?
140. What are the "natural" limits to the structure of an R-group in a polypeptide?

Question to ponder:
- Why do we think that the use of a common set of amino acids is a homologous trait?

Specifying a polypeptide's sequence

Perhaps at this point you are asking yourself, if there are so many different possible polypeptides, and there is no inherent bias favoring the addition of one amino acid over another, what determines the sequence of amino acids within a polypeptide, presumably it is not random. Here we connect the specification of polypeptide sequence to the information stored in DNA. We begin with a description of the process in bacteria and then extend it to archaea and eukaryotes. We introduce them in this order because, while basically similar (homologous), the system is somewhat simpler in bacteria, although you might find it complex enough for your taste. Even so, we will leave most of the complexities for subsequent courses. One thing that we will do that is not common is that we will consider the network dynamics of these systems. We will even ask you to do a little analytics, with the goal of enabling you to

---

[306] Short polypeptides, or rather the genes that encode them, can be difficult to recognize since short "open reading frames" are difficult to identify unambiguous: see Peptidomic discovery of short open reading frame–encoded peptides in human cells

[307] OMIM entry for TITIN: http://omim.org/entry/188840

make plausible predictions about the behavior of these systems, particularly in response to various perturbations, mutations and such. Another important point to keep in mind, one we have made previously, is that the system is continuous. The machinery required for protein synthesis is inherited by the cell, and new copies of it are synthesized as the cell grows; each new polypeptide is synthesized in an environment full of pre-existing proteins and ongoing metabolic processes.

A bacterial cell synthesizes thousands of different polypeptides. The sequence of these polypeptides, the exact amino acids from the N-terminal start to the C-terminal end of the polypeptide, is encoded within the organism's DNA. The bacterial genome is a double-stranded circular DNA molecule that is millions of base pairs in length. Each polypeptide is encoded by a specific region of this DNA molecule. So, our questions are how are specific regions in the DNA recognized and how is the information present in nucleic acid-sequence **translated** into polypeptide sequence.

To address the first question let us think back to the structure of DNA. It was immediately obvious to people studying the question that the one-dimensional sequence of a polypeptide could be encoded in the one-dimensional sequence of the polynucleotide chains in a DNA molecule.[308] The real question was how to translate the language of nucleic acids, which consists of sequences of four different nucleotides, into the language of polypeptides, which consists of sequences of the 20 (or 22) different amino acids. As pointed out by the physicist George Gamow (1904-1968)[309] the minimum set of nucleotides needed to encode all 20-22 amino acids is three; a sequence of one nucleotide (4^1) could encode at most four different animo acids, a two nucleotide sequence could encode (4^2) or 16 different amino acids (not enough), while a three nucleotide sequence (4^3) could encode 64 different amino acids (more than enough).[310] Although the actual coding scheme that Gamow proposed was wrong, his thinking about the coding capacity of DNA influenced those who set out to experimentally determine the actual rules of the "genetic code".

The genetic code is not the information itself, but the algorithm by which nucleotide sequences are "read" to determine polypeptide sequences. A polypeptide is encoded by the sequence of nucleotides. This nucleotide sequence is read in groups of three nucleotides, known as a codon. The codons are read in a non-overlapping manner, with no spaces (that is, non-coding nucleotides) between them. Since there are 64 possible codons but only 20 (or 22) different amino acids used in organisms, the code is redundant, that is, certain amino acids are encoded for by more than one codon. In addition there are three codons, UAA, UAG and UGA, that do not encode any amino acid but are used to mark the end of a polypeptide, they encode "stops" or periods (→).

The region of the nucleic acid that encodes a polypeptide
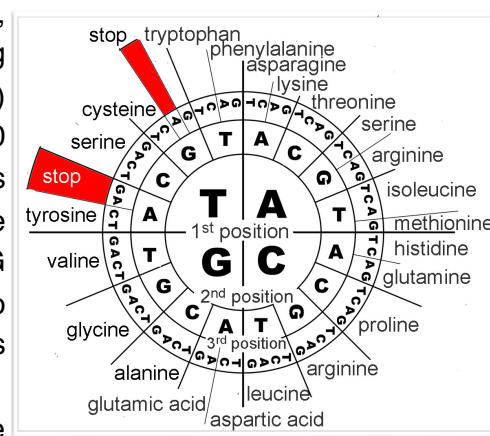
[308] Nature of the genetic code finally revealed!: http://www.nature.com/nrmicro/journal/v9/n12/full/nrmicro2707.html

[309] when he was a professor at UC Boulder

[310] The Big Bang and the genetic code: Gamow, a prankster and physicist, thought of them first

begins with what is known as the "start" codon and continues until one of the three stop codons is reached.[311] A sequence defined by in-frame start and stop codons, with some number of codons between them, is known as an open reading frame, an ORF. At this point it is important to note that while the information encoding a polypeptide is present in the DNA, the DNA copy of this information is not used directly to specify the polypeptide sequence. Rather, the process is indirect, it involves an intermediate. The information in the DNA is first copied (transcribed) into an RNA molecule, known as a messenger RNA or mRNA; it is the mRNA molecule that directs polypeptide synthesis. The process of copying information within DNA into an RNA molecule is known as transcription because both DNA and RNA use the same language, nucleotide sequences. In English, as opposed to molecular biology, transcription is the process of making a written copy of what someone says - the language of both is the same. In contrast polypeptides are written in a different language, amino acid sequences. For this reason the process of RNA-directed polypeptide synthesis is known as translation, which involves changing between languages, from nucleic acid-ese to polypeptide-ese.

The origin of the genetic code

There are a number of hypotheses as to how the genetic code originated. One is the frozen accident model in which the code used in modern cells is the result of an accident, a bottleneck event associated with the appearance of LUCA. Early in the evolution of life on Earth, there may have been multiple types of proto-organisms, each using a different genetic code. The common genetic code found in all existing organisms reflects the fact that only one of these proto-organisms gave rise to all modern organisms. Alternatively, the code could reflect specific interactions between RNAs and amino acids that played a role in the initial establishment of the code. It is not clear which model reflects what actually happened, it is likely to be theoretically unknowable, at least until unrelated forms of life are discovered on Earth or elsewhere. What is clear, however, is that the code is not absolutely fixed, there are examples in which certain codons are "repurposed" in various organisms. In fact there are efforts to re-engineer codons to produce proteins made using a range of more that 100 "unnatural" amino acids (uAAs).[312] What these variations in the genetic code illustrate is that evolutionary mechanisms can change the genetic code.[313] Since the genetic code does not appear to be predetermined, the general conservation of the genetic code among organisms is seen as strong evidence that all organisms, even the ones with minor variations in their genetic codes, are derived from a single common ancestor. It appears that the genetic code is a homologous trait between organisms.

Protein synthesis: transcription (DNA to RNA)

Having introduced you to DNA, mRNA, and the genetic code, however briefly, we now return to the process by which a polypeptide is specified by a DNA sequence. Our first task is to understand how it is that we might be able to find the specific region within a DNA molecule that encodes a specific

---

[311] There are situations in which non-start codons occur: see repeat-associated non-ATG translation (RAN translation)
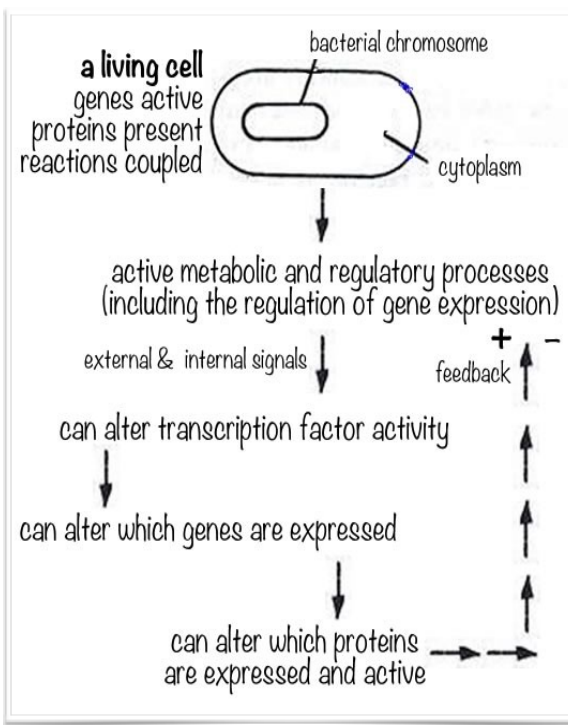
[312] Designing logical codon reassignment – Expanding the chemistry in biology

[313] The genetic code is nearly optimal for allowing additional information within protein-coding sequences & Stops making sense: translational trade-offs and stop codon reassignment:

polypeptide; we are looking for a relatively short region of DNA within millions (in prokaryotes) or billions (in eukaryotes) of base pairs of DNA. So while the double-stranded nature of DNA makes the information stored in it redundant, a fact that makes DNA replication straightforward, the specific nucleotide sequence that will be decoded using the genetic code is present in only one of the two strands. From the point of view of polypeptide sequence the other strand is effectively nonsense. One complexity associated with the double-stranded and anti-parallel nature of DNA is that information containing sequence can, in theory, run along either strand, although in opposite directions. This means that a gene's regulatory sequence must specify where, when and how often RNA synthesis starts and which of the two anti-parallel DNA strands is used to specify the "expressed" RNA's sequence.

If we think about this problem - we recognize one way to recognize a gene involves nucleotide sequences, together with something that can "read" (or recognize) a specific nucleotide sequence. Let us consider a specific form of the problem, say we want to uniquely specify one gene (one sequence) within the 3,000,000 base pairs of an *E. coli*'s cell's genomic DNA. For simplicity let us assume that the A:T ratio equals the G:C ratio. Clearly a 1 base pair sequence will not work, since we might expect that half of the base pairs will be recognized, either by directly binding to T or indirectly by binding to an A. To be unique the sequence we want must occur once in 3,000,000 base pairs $(1/3,000,000 = 3.33\ldots \times 10^{-7} = 0.000000333)$. If we use a two base sequence, it will occur $1/4 \times 1/4 = 1/16 = 0.0625$, a four base sequence 0.0039, an eight base sequence 0.00001523, but a 16 base sequence has a probability of occurring purely by chance of $\sim 2.32 \times 10^{-10}$, which is less than once per genome.[314]

Once a gene's regulatory region is identified (by the binding of a specific type of protein - see below), it can be "expressed". In fact, it is common to say that a gene is expressed only when RNAs are synthesized (transcribed) from it. If a gene is not expressed, that means that no RNAs corresponding to its sequence are being synthesized within the cell. In a sense, it is as if it is not there (at least in a particular cell type type or environmental condition). RNA synthesis is mediated by a DNA-dependent, RNA polymerase (which is encoded by genes)($\rightarrow$). Where, and in which orientation, the polymerase binds to the gene's DNA is determined by the gene's regulatory sequence(s), which is inherited from the organism's parent(s) and the protein(s), known as transcription factors, bound to it. Transcription factor proteins are themselves encoded by genes. Polymerase can bind to the DNA-transcription factor complex, the first step in the synthesis of a new RNA. Of course, since there are many genes in the genome, the stability of the DNA-Transcription Factor-Polymerase complex, as well as a number of other factors, will impact the number of RNAs from a particular gene that are

a living cell
genes active
proteins present
reactions coupled

bacterial chromosome

cytoplasm

active metabolic and regulatory processes
(including the regulation of gene expression)

external & internal signals

feedback

can alter transcription factor activity

can alter which genes are expressed

can alter which proteins
are expressed and active

---

[314] As we will return to, the CRISPR CAS9 system for mutagenesis uses a 22-base "guide RNA" to direct an endonuclease; this, in theory at least, would be expected to guarantee one target per genome.

synthesized per unit time. In addition to mRNAs, a number of other types of RNAs are synthesized, these include structural, catalytic, and regulatory RNAs. We will postpone discussion of further complexities to later on (and to subsequent classes).

At this point, it is useful to explicitly recognize some common aspects of biological systems. They are highly regulated, adaptive and homeostatic - that is, they can adjust their behavior to changes in their environment (both internal and external) to maintain the living state. These types of behaviors are based on various forms of feedback regulation. In the case of the bacterial gene expression system, there are genes that encode specific transcription factors. Which of these genes are expressed determines which transcription factor proteins are present and, in turn, which genes are actively expressed. Of course, the gene encoding a specific transcription factor is itself regulated. Transcription factors can act positively or negatively, which means that they can lead to the activation of transcription (by recruiting and activating the RNA polymerase) or blocking its recruitment and/or its activation. In addition the activity of a particular transcription factor can be regulated (a topic we will return to later on in this chapter).

All organisms are complex. A "simple" bacterium contains thousands of genes and different sets of genes are used in different environments and situations, and in different combinations to produce specific behaviors. In some cases, these behaviors may be mutually antagonistic. For example, a bacterium facing a rapidly drying out environment might turn on specific genes involved in rapid growth and division in order to prepare itself (through the expression of other genes that turn on) to survive in a more hostile environment. Our goal is not to have you generate perfectly accurate predictions about the behavior of an organism in a particular situation, but rather to be able to make plausible predictions about how gene expression will change in response to various perturbations. This requires us to consider, although at a rather elementary level, a few of the regulatory processes that are active in cells.

For a transcription factor to regulate a specific gene, either positively or negatively, it must be able to bind to specific sequences within the DNA. Whether or not a gene is expressed, whether it is "on" or "off", depends upon which transcription factors are expressed, are active, and can interact productively with the DNA-dependent, RNA polymerase (commonly referred to as RNA polymerase). You might speculate that groups of genes that are expressed together, under common cellular and environmental conditions, may be regulated by the same or related transcription factor proteins, and have similar regulatory sequences, a situation that makes it possible to regulate groups of genes in a coordinated manner. Inactivation of a transcription factor can involve a number of mechanisms, including its destruction, modification, or interactions with other proteins, so that it no longer interacts productively with either its target DNA sequence or the RNA polymerase. Similarly the activity of a transcription factor can be regulated (as we will see). Once a transcription factor is active, it can diffuse through out the cell and (in prokaryotic cells that do not have a barrier to control interactions with DNA) can bind to its target DNA sequences. Now an RNA polymerase can bind to the DNA-transcription factor complex, an interaction that leads to the activation of the RNA polymerase and the initiation of RNA synthesis, using one DNA strand to direct RNA synthesis. Once RNA polymerase has been activated, it will move away from the transcription factor-DNA complex. The DNA bound transcription factor can then bind another polymerase or the transcription factor can release from the DNA (in response to molecular level collisions), and can diffuse away, interact with other regulatory factors, or rebind to other sites in the
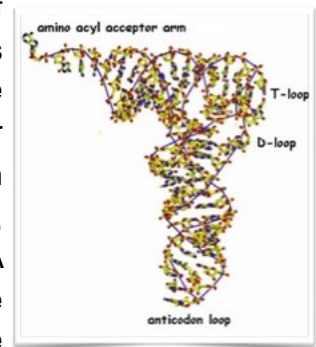
DNA. Clearly the number of copies of a particular transcription factor protein and its interaction partners and DNA binding sites will impact the behavior of the system.

RNA synthesis is a thermodynamically unfavorable reaction, so for it to occur it must be coupled to a thermodynamically favorable reaction, in particular nucleotide triphosphate hydrolysis. The RNA polymerase moves along the DNA (or the DNA moves through the RNA polymerase, your choice), to generate an RNA molecule (the transcript). Other signals within the DNA, and recognized by proteins associated with the transcription machinery, lead to the termination of transcription and the release of the RNA polymerase. Once released, the RNA polymerase returns to its inactive state. It can act on another gene if the RNA polymerase interacts with transcription factors bound to the gene's promoter. Since multiple types transcription factor proteins are present within the cell and RNA polymerase can interact with all of them, which genes are expressed within a cell will depend upon the relative concentrations and activities of specific transcription factors and their regulatory proteins, together with the binding affinities of particular transcription factors for specific DNA sequences (compared to their general low-affinity binding to DNA in general).

Protein synthesis: translation (RNA to polypeptide)

Translation involves a complex cellular organelle, the ribosome, which together with a number of accessory factors reads the code in a mRNA molecule and produces the appropriate polypeptide.[315] The ribosome is the site of polypeptide synthesis. It holds the various components, the mRNA, tRNAs, and accessory factors, in appropriate juxtaposition to one another to catalyze polypeptide synthesis. But perhaps we are getting ahead of ourselves. For one, what exactly is a tRNA?

The process of transcription is used to generate a number of other types of RNAs beside mRNAs; these play structural, catalytic, and regulatory roles within the cell. Of these non-mRNAs, two are particularly important in the context of polypeptide synthesis. The first are molecules known as transfer RNAs (tRNAs). These small single stranded RNA molecules (→) fold back on themselves to generate a compact L-shaped structure. In the bacterium *E. coli*, there are 87 genes that encode tRNAs (there are over 400 such tRNA encoding genes in humans). For each amino acid and each codon there are one or more tRNAs. The only exception being the stop codons, for which there are no tRNAs. A tRNA specific for the amino acid phenylalanine would be written tRNA[Phe]. Two parts of the tRNA molecule are particularly important and functionally linked: the part that recognizes the codon on the mRNA, in the mRNA-ribosome complex, and the amino acid acceptor stem, which is where an amino acid is covalently attached to the tRNA. Each specific type of tRNA can recognize a particular codon in an mRNA through base pairing interactions with what is known as its anti-codon. The rest of the tRNA molecule mediates interactions with protein catalysts (enzymes) known as amino acyl tRNA synthetases. There is a distinct amino acyl tRNA synthetase for each amino acid: there is a phenylalanine-tRNA synthetase and a proline-tRNA synthetase, etc. An amino acyl tRNA synthetase binds the appropriate tRNA and the appropriate amino acid and, through a reaction coupled to a

_____

[315] Can't stop yourself? go here for a more detailed description of translation.

thermodynamically favorable nucleotide triphosphate hydrolysis reaction, catalyzes the formation of a covalent bond between the amino acid acceptor stem of the tRNA and the amino acid, to form what is known as a charged or amino acyl-tRNA. The loop containing the anti-codon is located at the other end of the tRNA molecule. As we will see, in the course of polypeptide synthesis, the amino acid group attached to the tRNA's acceptor stem will be transferred from the tRNA to the growing polypeptide.

Ribosomes

Ribosomes are composed of roughly equal amounts by mass of ribosomal RNAs (rRNAs) and ribosomal polypeptides. An active ribosome is composed of a small and a large ribosomal subunit. In the bacterium *E. coli*, the small subunit is composed of 21 different polypeptides and a 1542 nucleotide long rRNA molecule, while the large subunit is composed of 33 different polypeptides and two rRNAs, one 121 nucleotides long and the other 2904 nucleotides long.[316] It goes without saying (so why are we saying it?) that each ribosomal polypeptide and RNA is itself a gene product. The complete ribosome has a molecular weight of ~3×10^6 daltons (please note, there is no reason to remember any of these numbers except to appreciate that the ribosome is encoded by over 50 distinct genes, it is a complex molecular machine). One of the rRNAs is an evolutionarily conserved catalyst, known as a ribozyme (in contrast to protein based catalysts, which are known as enzymes). This rRNA lies at the heart of the ribosome and catalyzes the transfer of an amino acid bound to a tRNA to the carboxylic acid end of the growing polypeptide chain (also attached to a tRNA). RNA based catalysis is a conserved feature of polypeptide synthesis and appears to represent an evolutionarily homologous trait.

The growing polypeptide chain is bound to a tRNA, known as the peptidyl tRNA. When a new aa-tRNA enters the ribosome's active site (site A), the growing polypeptide is added to it, so that it becomes the peptidyl tRNA (with a newly added amino acid, the amino acid originally associated with an incoming aa-tRNA)(↓). This attached polypeptide group is now one amino acid longer. A
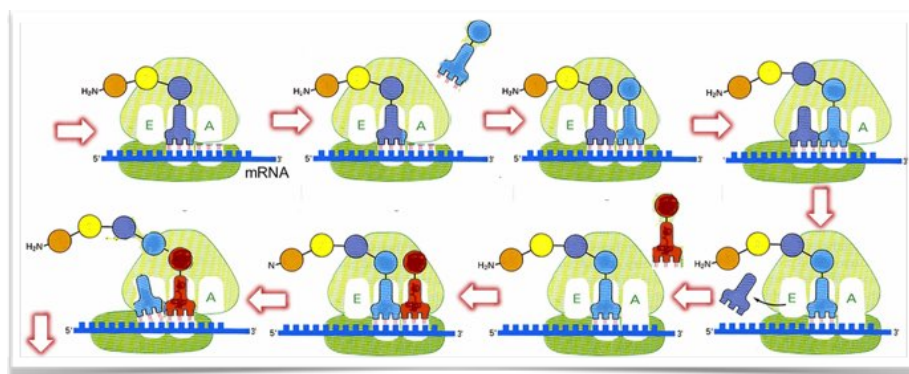
virtuallaboratory FLASH applet illustrating this process can be found here (link).

The cytoplasm of cells is packed with ribosomes. In a rapidly growing bacterial cell, ~25% of the total cell mass is ribosomes. Although structurally similar, there are characteristic differences between the ribosomes of bacteria, archaea, and eukaryotes. This is important from a practical perspective. For example, a number of antibiotics selectively inhibit polypeptide synthesis by bacterial, but not eukaryotic ribosomes. Both chloroplasts and mitochondria have ribosomes of the bacterial type; another piece of evidence that chloroplasts and mitochondria are descended from bacterial

---

[316] In the human, the small ribosomal subunit is composed of 33 polypeptides and a 1870 nucleotide rRNA, while the large ribosomal subunit contains 47 polypeptides, and three rRNAs of 121, 156, and 5034 nucleotides in length.

endosymbionts and a reason that translational blocking anti-bacterial antibiotics are mostly benign, since most of the ribosomes inside a eukaryotic cell are not seriously effected by them.

The translation (polypeptide synthesis) cycle

In bacteria and archaea, there is no barrier between the cell's DNA and its cytoplasm, which contains the ribosomal subunits and all of the other components involved in polypeptide synthesis. Newly synthesized RNAs emerge from the RNA polymerase directly into the cytoplasm, where they can begin to interact with ribosomes. In fact, because the DNA is located in the cytoplasm in bacteria, the process of protein synthesis (translation) can begin before mRNA synthesis (transcription) is complete.

We will walk through the process of protein synthesis, but at each step we will leave out the various accessory factors involved in regulating the process and coupling it to the thermodynamically favorable reactions that make it possible. These can be important if you want to re-engineer or manipulate the translation system, but (we think) are unnecessary details that obscure a basic understanding. Here we will remind you of two recurring themes. The first is the need to recognize that all of the components needed to synthesize a new polypeptide (except the mRNA) are already present in the cell; another example of biological continuity - mRNA translation can occur only because the cell, and all the needed components of the processes involved, already exist. The second is that all of the interactions we will be describing are based on stochastic, thermally driven movements (and collisions). For example, when considering the addition of an amino acid to a tRNA (the formation of an amino acyl-tRNA or aa-tRNA), random motions have to bring the correct amino acid and the correct tRNA to their binding sites on the appropriate amino acyl tRNA synthetase,. Once the aa-tRNA is formed, only the correct amino acid charged tRNA will bind productively to the ribosome-mRNA-nascent polypeptide complex. Generally, many unproductive collisions occur before a productive (correct) one, since there are more than 20 different amino acid/tRNA molecules bouncing around in the cytoplasm. The stochastic aspects of the peptide synthesis process are rarely illustrated.
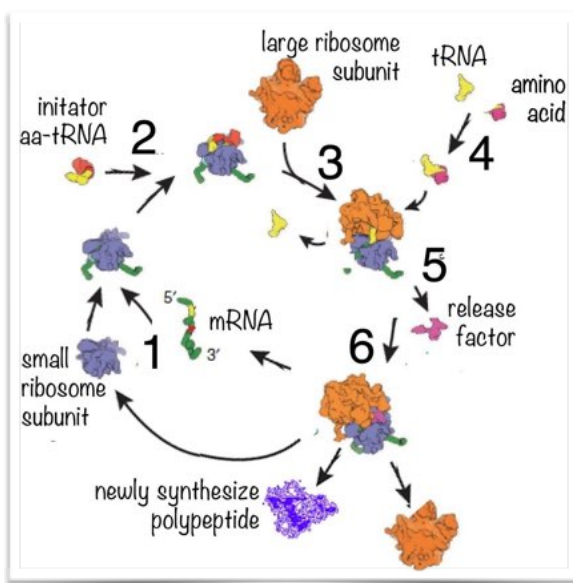
The first step in polypeptide synthesis is the synthesis of the specific mRNA that encodes the polypeptide (←). (1) The mRNA contains a sequence[317] that mediates its binding to the small ribosomal subunit. This sequence is located near the 5' end of the mRNA. (2) the mRNA-small ribosome subunit complex now interacts with and binds to a complex containing an initiator (start) amino acid:tRNA. In both bacteria and eukaryotes the start codon is generally an AUG codon and inserts the amino acid methionine, although other, non-AUG start codons are possible.[318] This interaction defines the beginning of the polypeptide as well as the coding region's reading frame. (3) The met-tRNA:mRNA:small ribosome subunit complex can now form

---

[317] Known as the Shine-Delgarno sequence for its discovers

[318] Hidden coding potential of eukaryotic genomes: nonAUG started ORFs: http://www.ncbi.nlm.nih.gov/pubmed/22804099

a functional complex with a large ribosomal subunit to form the functional mRNA:ribosome complex. (4) Catalyzed by amino acid tRNA synthetases, charged amino acyl tRNAs will be present and can interact with the mRNA:ribosome complex to generate a polypeptide. Based on the mRNA sequence and the reading frame defined by the start codon, amino acids will be added sequentially. With each new amino acid added, the ribosome moves along the mRNA (or the mRNA moves through the ribosome). An important point, that we will return to when we consider the folding of polypeptides into their final three-dimensional shapes, is that the newly synthesized polypeptide is threaded through a molecular tunnel within the ribosome. Only after the N-terminal end of the polypeptide begins to emerge from this tunnel can it begin to fold. (5) The process of polypeptide polymerization continues until the ribosome reaches a stop codon, that is a UGA, UAA or UAG.[319]   Since there are no tRNAs for these codons, the ribosome pauses, waiting for a charged tRNA that will never arrive. Instead, a polypeptide known as release factor, which has a shape something like a tRNA (→), binds to the polypeptide:mRNA:ribosome complex instead. (6) This leads to the release of the polypeptide, the disassembly of the ribosome into small and large subunits, and the release of the mRNA.[320]
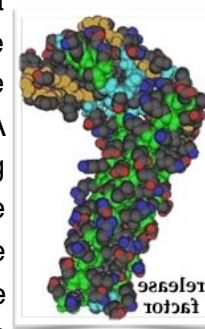
When associated with the ribosome, the mRNA is protected against interaction with proteins (ribonucleases) that could catalyze its degradation into nucleotides. Upon its release from the ribosome, an mRNA may interact with a new small ribosome subunit, and begin the process of polypeptide synthesis again or it may interact with a ribonuclease and be degraded. Where it is important to limit the synthesis of particular polypeptides, the relative probabilities of these two events, new translation versus RNA degradation, will be skewed in favor of degradation. Typically RNA stability is regulated by the binding of specific proteins to nucleotide sequences within the mRNA. The relationship between mRNA synthesis and degradation will determine the half-life of a population of mRNA molecules within the cell, the steady state concentration of the mRNA in the cell, and indirectly, the level of the encoded polypeptide present.

Questions to answer:
141. Why so many tRNA genes? How, in the most basic terms, do different tRNAs differ from one another?
142. How might the concentration of various tRNAs and the frequency of various codons influence the rate of polypeptide synthesis?
143. What is the minimal number of different tRNA-amino acid synthetases in a cell?
144. Would you expect a ribosome to make mistakes in amino acid incorporation or polypeptide termination? How are such mistakes similar to and different from mutations?

Question to ponder:
- How might a ribosome shift its reading frame while translating an mRNA?

---

[319] In addition to the common 19 amino and 1 imino (proline) acids, the code can be used to insert two other amino acids selenocysteine and pyrrolysine. In the case of selenocysteine, the amino acid is encoded by a stop codon, UGA, that is in a particular context within the mRNA. Pyrrolysine is also encoded by a stop codon. In this case, a gene that encodes a special tRNA that recognizes the normal stop codon UAG is expressed. see Selenocysteine

[320] Interested in learning more, check out eukaryotic translation termination factor 1

Effects of point mutations on polypeptides and proteins

Mutations in a gene's regulatory region can alter the gene's expression by regulating the frequency of transcription. Mutations in a gene's coding region generally do not influence transcription rate (unless of course regulatory regions are located within the coding region) but they can influence the sequence of the encoded polypeptide. We can define three types of mutations that involve changing a single base pair, known as a single nucleotide polymorphism or SNP: synonymous, mis-sense, and non-sense mutations. Because of the semi-redundant nature of the genetic code, it is possible that a single nucleotide change in a coding region can have no effect on the amino acid encoded – this is referred to as a synonymous mutation. That said, different codons for the same amino acid can be recognized by different tRNAs, which are the products of different genes, and may be present at different concentrations in the cell. The efficiency of translation is influenced by the rate of aa-tRNA binding. Different organisms can differ in codons they use to encode a particular amino acid, a fact that leads to what is known as codon bias. Codon bias can influence the efficiency of mRNA translation; when genetically engineering the synthesis of a mRNA from one organism in another, translational efficiency can be significantly increased by altering the gene that encodes the mRNA so that it uses the codon bias of the host cell, rather than the codon bias of the donor.

Another possibility is that the change of a single nucleotide in the coding region will change the amino acid encoded; this is known as a mis-sense mutation. The effect of a mis-sense mutation will depend upon where in the polypeptide it occurs. We can compare homologous polypeptides found in various organisms; regions that are similar in terms of amino acid sequence are referred to as conserved regions, compared to regions that are more variable, known happily as variable regions.[321] A mis-sense mutation that replaces an amino acid in a conserved region of a polypeptide is likely to have a more drastic effect on the polypeptide's function than a similar change in a variable region. Similarly, a mutation that replaces a large hydrophobic amino acid with a acidic or basic, that is, highly hydrophilic amino acid, is more likely to perturb polypeptide structure and function than replacing a large hydrophobic amino acid with a smaller one. The final type of single nucleotide mutation that we will consider here leads to the replacement of codon that specifies a amino acid with a stop codon; it is known as a non-sense mutation. The result of a non-sense mutation is a truncated polypeptide. As a first guess, the effect of a non-sense mutation will be more severe the closer it is to the beginning of the coding region, compared to its effect near the end of the coding region – although other factors that we will consider in a short while can influence such a mutation's effect.

A final type of mutation involves deletion or addition of a nucleotide. Such insertions or deletion can disrupt or alter the binding of proteins to a gene's regulatory region, influencing gene expression.   If they occur within the coding region, they can alter the reading frame.   In particular, the insertion or deletion does not involve three or a multiple of three nucleotides, the reading frame of the mRNA down-stream of the insertion/deletion will be changed, – the sequence of the polypeptide will be changed completely.   In contrast, if the insertion/deletion involves three or a multiple of three nucleotides, there will be insertion or deletion of amino acids from the final polypeptide, but the normal sequence downstream of that altered region will stay the same.

---

[321] In practice, the situation can be more complex, as a polypeptide can assume a three-dimensional shape - and that shape itself can be conserved, but we will not consider that possibility further - that is something for a later course.
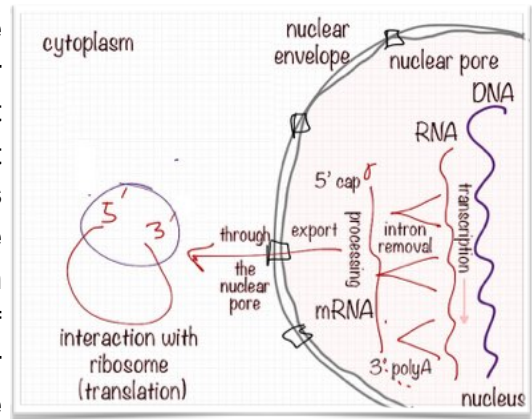
Questions to answer:
145. How would you explain the terms "up-stream" and "down-stream" in terms of gene structure.
146. What effects on polypeptide synthesis arise from neglecting codon bias?
147. Why don't release factors cause the premature termination of translation at non-stop codons?
148. What might happen if a ribosome starts translating an mRNA at the "wrong" place?

Question to ponder:
- When analyzing the effects of a particular non-sense or mis-sense mutation (allele), what factors would you consider first?
- How would you go about reengineering an organism to incorporate non-biological amino acid in its proteins

mRNA processing and nuclear export in eukaryotes

We will just briefly reiterate a few points on how gene expression in particular, and polypeptide synthesis differ between prokaryotes and eukaryotes. The first and most obvious difference is the presence of a nucleus, a distinct domain within the eukaryotic cell that separates the cell's genetic material, its DNA, from the cytoplasm, where the ribosomes are located (→). Aside from those within mitochondria and chloroplasts, the DNA molecules of eukaryotic cells are located within the nucleus. The barrier between nuclear interior and cytoplasm is known as the nuclear envelope: no such barrier between DNA and ribosomes exists in prokaryotes. In prokaryotes, both bacteria and archaea, the DNA is in direct contact with the cytoplasm. In eukaryotes, a newly synthesized mRNA molecule undergoes splicing (see below) and is modified (processed) at both its 5' and 3' ends. Only after RNA processing has occurred will the "mature" mRNA be exported out of the nucleus, through a nuclear pore, into the cytoplasm, where it can interact with ribosomes. Prokaryotic mRNAs are generally not processed.

The nuclear envelope complex (typically considered in greater detail in cell biology courses) consists of two lipid bilayer membranes punctuated by nuclear pores, which are macromolecular complexes (protein machines) of ~125,000,000 daltons (→). While molecules of molecular weight less than ~40,000 daltons can generally pass through the nuclear pore, larger molecules must be actively transported through a process coupled to a thermodynamically favorable reaction, in this case the hydrolysis of guanosine triphosphate (GTP). The movement of larger molecules into and out of the nucleus through nuclear pores is regulated by what are known as nuclear localization and nuclear export sequences, located within polypeptides. These are recognized by proteins (receptors) associated with the pore complex. A protein with an active nuclear localization sequence (NLS) will be found in the nucleus while a protein with an active nuclear exclusion sequence (NES) will be found in the cytoplasm. By controlling NLS and NES activity a protein can come to accumulate, in a regulated manner, in either the nucleus or the cytoplasm. As we will see later on, the nuclear envelope breaks down during cell division (mitosis) in many but not all eukaryotes. Tears in the nuclear envelope have also been found to occur when migrating cells try to
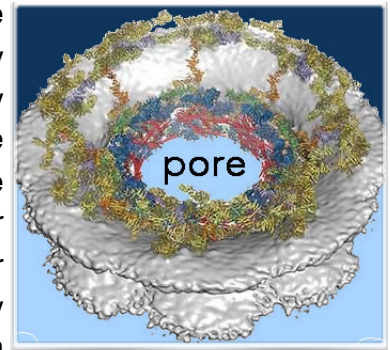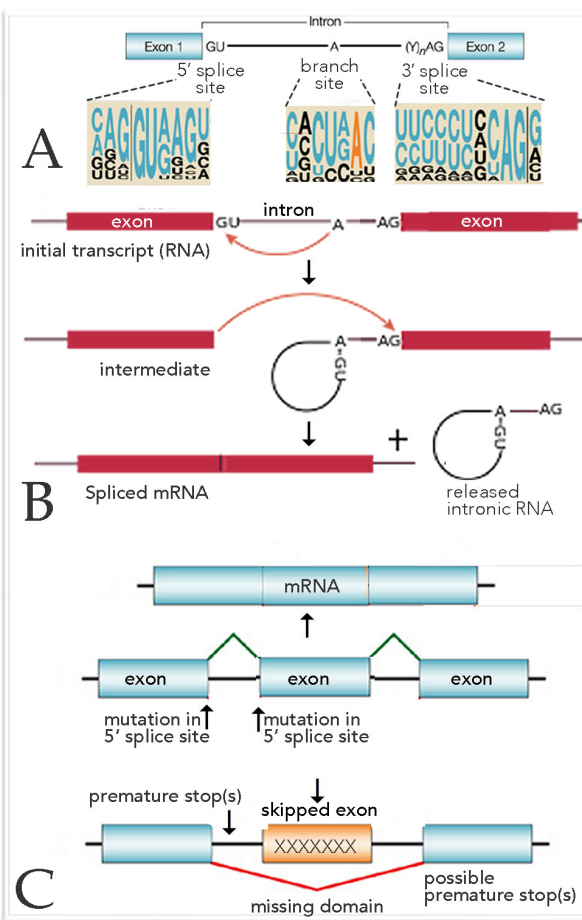
squeeze through small openings.[322] Once the integrity of the nuclear envelop is re-established, proteins with NLS and NES sequences move back to their appropriate location within the cell through active, that is energy driven, coupled reaction-based processes (discussed in more detail in Chapter 10).

Mutations influencing splicing

While there is much more detail we can consider, details best reserved for a subsequent course in molecular biology, it is worth noting a final class of point mutations, namely those that influence the splicing of an newly synthesized RNA molecule. Eukaryotic genes are generally broken up into coding regions, known as exons, and the non-coding regions between exons – these are known as intervening regions or introns. When a polypeptide-encoding gene is expressed, the RNA made, the initial transcript, contains both introns and exons. But ribosomes cannot distinguish between exon and intron sequences (probably one reason that prokaryotes do not have introns). In eukaryotes, introns need to be removed before the mature mRNA is exported across the nuclear envelope and into the cytoplasm (were the ribosomes are). So the obvious question is, how exactly are introns recognized and removed, what mechanism is used? As you might already have guessed, there must be information that identifies introns to be removed, and because information is encoded by nucleic acid sequences, that information must be in the form of specific sequences in the initial RNA molecule: there are sequences that indicate the end of an exon (and the start of an intron), known as the 5' splice site, as well as the start of an exon and the end of an intron, known as the 3' splice site, as well as a sequence within the intron, known as the branch site (A→). We can visualize this information through what are known as a "sequence logo" plot.[323] Such a plot indicates the information associated with a sequence; where there is no preference, that is, where any of the four nucleotides is acceptable, the information present at that site is 0. Where either of two nucleotides are acceptable, the information is 1, and where only one particular nucleotide

is acceptable, the information content is 2. 5' and 3' splice sites are identified by specific sequences associated with the ends of the exon and the intron, and the branch points (A→). Factors in the cell recognize these sequences, and using endonuclease activity, cut out the intron and join the two ends of the exons together (B↑), releasing the intervening intron sequence in a looped form. A point mutation that disrupts the normal splice recognizing sequences (C↑) can lead to an inhibition of splicing, so that

---

[322] Tearing the nuclear envelope: http://www.sciencemag.org/news/2016/03/cells-can-do-twist-sometimes-their-nuclei-burst

[323] Sequence logos: a new way to display consensus sequences: http://www.ncbi.nlm.nih.gov/pubmed/2172928

the introns remains in the final mRNA. Since introns do not encode polypeptides, there is no selection against the presence of stop codons in their sequence. A ribosome reading along a non-spliced RNA will first add a series of inappropriate amino acids to the growing polypeptide, and is likely to encounter a stop codon, leading to premature termination. Alternatively, for example, if a 3' splice site is disabled, a "down-stream" exon may be used for splicing; the result is that an exon normally include is lost from the spliced mRNA, the polypeptide sequence it encodes will be missing from the synthesized polypeptide, and it is possible that the down-stream reading frame will be wrong, leading to the synthesis of irrelevant amino acid sequences and stop codons. The result is that mutations that disrupt splicing can have dramatic hypomorphic, anti-morphic, and possible neo-morphic effects, and such mutations (alleles) have been associated with a number of human diseases.[324]

The complexity of eukaryotic genomes is greatly increased by the fact that most genes contain multiple exons and introns; different sets of exons can be spliced together in different cells and within a single cell to produce mRNA molecules that encodes variants of the same polypeptide. Exploring the implications of this added complexity is best left for later (molecular biology) courses.

Non-sense mediated RNA decay

As we will return to in detail, diploid organisms have two copies of each chromosome, and so two copies of each gene. A number of possible outcomes can be expected If one of these genes carries a allele containing a non-sense mutation. First, the encoded polypeptide will be truncated. This truncated polypeptide may be completely non-functional, and perhaps even unstable (that is, have a short half-life) – both of which can generate an amorphic or null allele – if this allele produces a phenotype, its presence leads to what is known as haploinsufficiency. It is also possible that the truncated polypeptide is not only non-functional, but that it may interact with and inhibit or alter the activity of the product of the other wild type (functional) allele – it can be anti-morphic. Such an anti-morphic allele can lead to more severe phenotypes than an amorphic allele, since it negatively effects the behavior of the wild type allele.

Organisms have developed something of a defense against non-sense mutations, particularly those that occur well up-stream of the normal stop codon. The normal stop codon typically occurs within in a particular sequence context, in part to insure that translation stops at the correct place. You should be able to generate plausible models for what happens if a stop codon is ignored. In fact, errors in translation are not so uncommon.[325] Because stop codons normally occur in a context that can be recognized, it is also possible to recognize that a stop codon is an inappropriate context, which is what a mis-sense mutation can generate. In particular, stop codon located within an exon located near the 3' end of the gene. During mRNA processing, introns are recognized and removed by the splicing system (see ↑); all introns have to be removed before the processed transcript (now an mRNA) is transported through the nuclear pore complex and into the cytoplasm. The removal of an intron is associated with

---

[324] The pathobiology of splicing: https://www.ncbi.nlm.nih.gov/pubmed/19918805

[325] In fact "amino-acid misincorporations during translation are estimated to occur once in every 1,000 to 10,000 codons translated‿At this error rate, 15% of average-length protein molecules will contain at least one misincorporated amino acid. Polypeptide errors can induce protein misfolding, aggregation, and cell death." from Drummond and Wilke: The evolutionary consequences of erroneous protein synthesis.
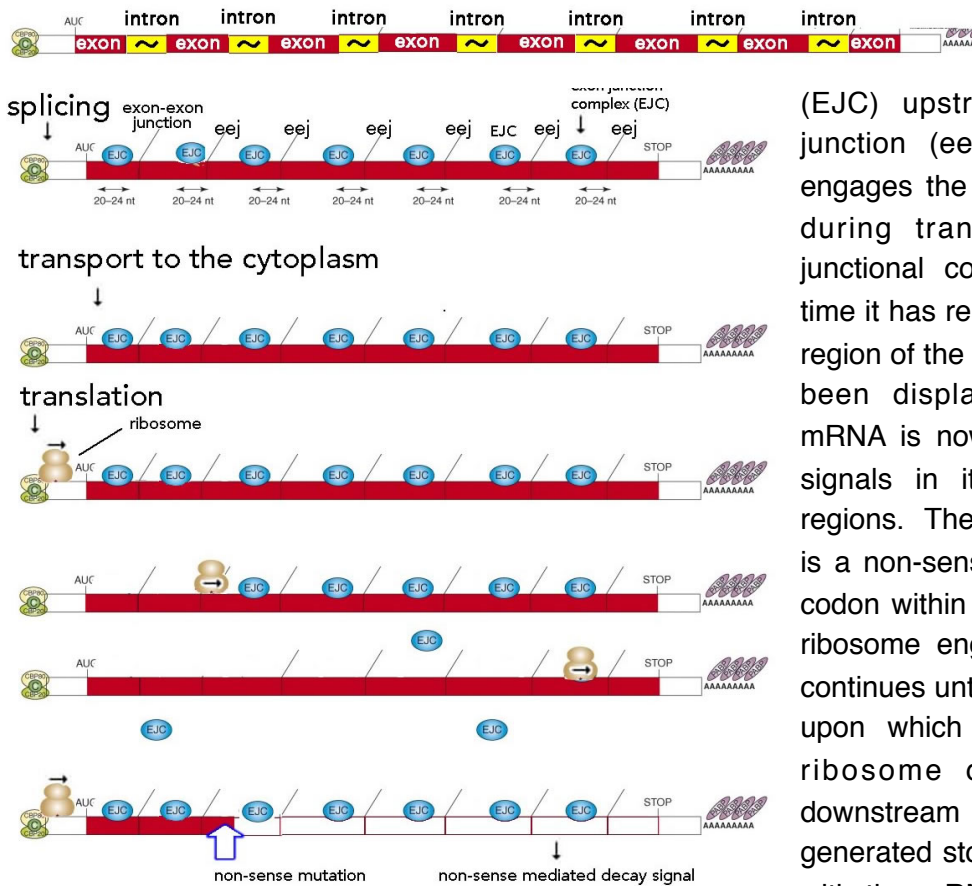
association of an exon-exon junction complex (EJC) upstream of each exon-exon junction (eej)(←). When a ribosome engages the mRNA and moves down it during translation, it removes the junctional complexes, so what by the time it has reached the end of the coding region of the mRNA, all of the EJCs have been displaced (←). The EFC-free mRNA is now stable, and regulated by signals in its 5' and 3' untranslated regions. The situation is different if there is a non-sense mutation generated stop codon within an upstream exon (←); the ribosome engages with the mRNA and continues until it reaches this stop codon, upon which release factor binds and ribosome disengages. All EJC downstream of the non-sense mutation generated stop codon remain associated with the mRNA. The failure to remove the EFCs marks the mRNA as aberrant and triggers a regulatory response known as "non-sense mediated decay" (NMD).[326] NMD leads to the degradation of an mRNA containing an out-of-context non-sense codon and dramatically reduces the synthesis of potentially anti-morphic polypeptides.

Alarm generation

The translation system is a major consumer of energy within the cell.[327] When a cell is starving, it does not have the energy to generate amino acid charged tRNAs (→). The result is that uncharged tRNAs accumulate. Since uncharged tRNAs fit into the amino-acyl-tRNA binding sites on the ribosome, their presence increases the probability of unproductive tRNA interactions with the mRNA-ribosome complex. When this occurs the stalled ribosome generates a signal (illustrated here: link) that can lead to adaptive changes in the cell that enable it to survive for long periods in a "dormant" state.[328]

---
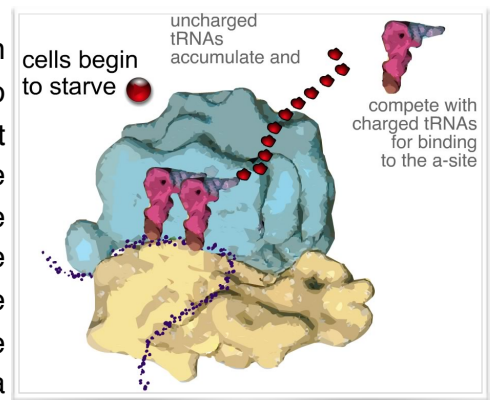
326 Mechanism and regulation of the nonsense-mediated decay pathway

327 Quantifying absolute protein synthesis rates reveals principles underlying allocation of cellular resources

328 Characterization of the Starvation-Survival Response of Staphylococcus aureus:

Another response that can occur is a more social one. Some cells in the population can "sacrifice" themselves for their generally closely related neighbors (remember kin selection and inclusive fitness.) By shutting down translation (and transcription), a cell containing an addiction module can undergo what is known as programmed cell death. The mechanism is based on the fact that proteins, like nucleic acids, differ in the rates that they are degraded within the cell. Just as ribonucleases can degrade mRNAs, proteases degrade proteins and polypeptides. How stable a protein/polypeptide is depends upon its structure, which we will be turning to soon. As discussed previously, interrupting protein synthesis leads to the rapid disappearance (turn-over) of the anti-toxin while the toxin persists, leading to cell death, which in turn leads to the release of the cell's nutrients, nutrients that can be used by its neighbors, in part to maintain active gene expression and protein synthesis.

Questions to answer:
149. A gene has many introns - provide a model for how it might encode functionally distinct polypeptides.
150. How can a mutation in splice site sequence influence gene expression and protein function?
151. How does NMD protect against potentially deleterious mutations (alleles)?
152. Why would a cell want to stop (rather than continue) polypeptide synthesis when it is starving?

Turning polypeptides into proteins

Protein structure is commonly presented in a hierarchical manner (video link). While this is an over-simplification, it is a good place to start. When we think about how a polypeptide folds, we have to think about the environment it will inhabit, how it interacts with itself and with other polypeptides. In a protein composed of multiple polypeptides, we need to consider how the polypeptide comes to interact with those other polypeptides, often termed protein subunits. As we think about polypeptide structure it is common to see the terms primary, secondary, tertiary, and quaternary structure. The primary structure of a polypeptide is the sequence of amino acids along the polypeptide chain, written from its N- or amino terminus to its C- or carboxyl terminus. The secondary structure of a polypeptide consists of local folding motifs: the α-helix, the β-sheet, and connecting domains. The tertiary structure of a polypeptide is the overall three dimensional shape a polypeptide takes in space, as well as how its R-chains are oriented. Quaternary structure refers to how the various polypeptides and co-factors combine and are arranged to form a functional protein. In a protein that consists of a single polypeptide and no co-factors, tertiary and quaternary structures are the same. As a final complexity, a particular polypeptide can be part of a number of different proteins – the universe of proteins that a polypeptide is a part of could be considered another level of structure. Some of these interactions are relatively stable, others more ephemeral and regulative. This is one way in which a gene can play a role in a number of different processes and be involved in the generation of a number of different phenotypes.

Polypeptide synthesis (translation), like most all processes that occur within the cell, is a stochastic process, meaning that it is based on random collisions between molecules. In the specific case of translation, the association of the mRNA with ribosomal components occurs stochastically. Given that a human cell contains ~24,000 genes that can generate mRNAs and ~2,000,000,000 ribosomes, most RNAs find a ribosome. Similarly, the addition of a new amino acid to the end of a growing polypeptide depends on the collision of the appropriate amino acid-charged tRNA with the RNA-ribosome complex. Since there are many different amino-acid charged tRNAs in the cytoplasm, the ribosomal complex

must productively bind only the amino-acyl-tRNA that the mRNA specifies, that is the tRNA with the right anticodon. This enables its attached amino acid to interact productively, leading to the addition of the amino acid to C-terminus of the growing polypeptide chain. You rarely see this fact illustrated in most presentations of polypeptide synthesis. From 12 to 21 amino acids are added to the end of a growing polypeptide chain per second in bacterial cells, and about half that rate in mammalian cells.[329]

Now you might wonder whether there are errors in polypeptide synthesis as there are in nucleic acid synthesis. In fact there are, as we have already noted above. For example, if a base is skipped by the ribosomal system, the reading frame will be thrown off. Typically, this leads to a completely different sequence of amino acids added to the end of the polypeptide, down-stream of the skip, and very often leads to a stop codon, which terminates translation, leading to the release of a polypeptide that cannot fold correctly and is (generally) rapidly degraded.[330] Similarly, if the wrong amino acid is inserted at a particular position and it disrupts normal folding, the can polypeptide disrupt normal cellular function or it may be degraded. There are molecular machines that can recognize mis-folded proteins and then mark such mis-folded protein for degradation. What limits the effects of mistakes made during translation is that most proteins (unlike DNA molecules) have finite and relatively short half-lives; that is, the time an average polypeptide exists before it is degraded by various enzymes. Normally (but not always) this limits the damage that a mis-translated polypeptide can do to the cell and organism.

Factors influencing polypeptide folding and structure

Polypeptides are synthesized, and they fold, in a vectorial, that is, directional manner. Synthesis occurs in an N- to C- terminal direction and the newly synthesized polypeptide exits the ribosome
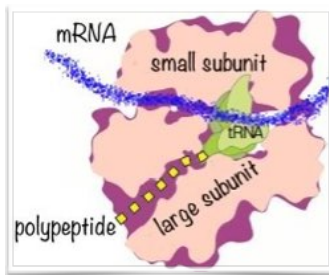
through a ~10 nm long and ~1.5 nm diameter tunnel (←)(video link). This tunnel is narrow enough to block the folding of the newly synthesized polypeptide chain. As the polypeptide emerges from the tunnel it begins to fold. At the same time it encounters the crowded cytoplasmic environment; the newly synthesized polypeptide needs to avoid low affinity, non-specific, and non-physiologically significant interactions with other cellular components.[331] If the polypeptide is part of a multi-subunit protein, it must also "find" its correct partner polypeptides, which again is a stochastic process. If the polypeptide does not fold correctly, it will not function correctly and may even damage the cell or the organism. A number of degenerative neurological disorders are due, at least in part, to the accumulation of mis-folded polypeptides (see below).

We can think of the folding process as a "drunken" walk across an energy landscape, with movements driven by intermolecular interactions and collisions with other molecules. The successful goal of this process is to find the lowest point in the landscape, the energy minimum of the system. This is generally assumed to be the native or functional state of the polypeptide. That said, this native state

---

[329] see http://bionumbers.hms.harvard.edu/default.aspx

[330] Quality control by the ribosome following peptide bond formation: http://www.ncbi.nlm.nih.gov/pubmed/19092806

[331] Remember, all molecules interact with each other via van der Waals interactions.

is not necessarily static, since the folded polypeptide (and the final protein) will be subject to thermal fluctuations; it is possible that it will move between various states with similar, but not identical stabilities.[332] The challenge to calculating the final folded state of a polypeptide is that it is a extremely complex problem, computationally. Generally two approaches are taken to characterizing the structure of a functional protein. In the first the structure of the protein is determined directly by X-ray crystallography or Nuclear Magnetic Resonance (NMR) spectroscopy (which, as you will notice, we are not going to explain here, but which you may encounter in chemistry classes). In the second, if the structure of a homologous (evolutionarily-related) protein is known, it can be used as a framework to model the structure of a previously unsolved protein. There are a number of on-line tools to generate such structural models.

A number of constraints influence the folding of a polypeptide. The first is the peptide bond itself. All polypeptides consist of a string of peptide bonds. It is therefore not surprising that there are common patterns in polypeptide folding. The first of these common patterns to be recognized, the α-helix (→), was discovered by Linus Pauling (1901-1994) and Robert Corey (1897-1971) in 1951. This was followed shortly thereafter by their description of the β-sheet. The forces that drive the formation of the α-helix and the β-sheet will be familiar, they are the same forces that underlie water structure, namely H-bonding interactions.
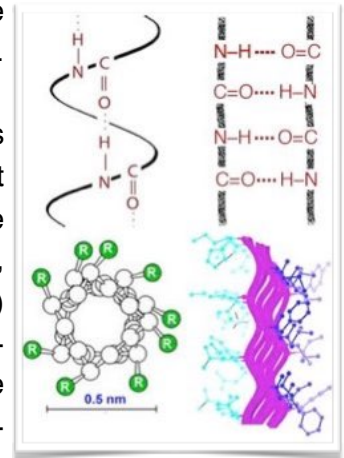
In an α-helix and a β-sheet, all of the possible H-bonds involving the peptide bond's donor and acceptor groups (–N–H : O=C– , with ":" indicating a H-bond) are formed within the polypeptide. In an α-helix these H-bond interactions run parallel to the polypeptide chain. In the β-sheet, these H-bonding interactions occur between polypeptide chains. The interacting strands within a β-sheet can run parallel or anti-parallel to one another, and can occur within a single polypeptide chain (folded back on itself in various ways) or between different polypeptide chains. In an α-helix, the R-groups point outward from the helix axis. In β-sheets the R-groups point in an alternating manner either above or below the plane of the sheet. While all amino acids can take part in either α-helix or β-sheet structures, the imino acid proline cannot - the N-group coming off the α-carbon has no H, so its presence in a polypeptide chain leads to a break in the pattern of intrachain H-bonds. It is worth noting that some polypeptides can adopt functionally different structures: for example in one form (PrPC) the prion protein contains a high level of α-helix (~42%) and essentially no β-sheet (~3%), while an alternative form (PrPSc), associated with the disease scrapie contains high levels of β-sheet (~43%) and ~30% α-helix (see below).[333] The result is two very different 3-dimensional protein structures, even though the primary sequences of the two are identical.

Peptide bond rotation and proline: Although drawn as a single bond, the peptide bond behaves more like a double bond, or rather like a bond and a half. In the case of a single bond, there is free rotation around the bond axis in response to molecular collisions. In contrast, rotation around a peptide bond requires more energy to move from the trans to the cis configuration and back again (↓), that is, it is more difficult to rotate around the peptide bond because it involves the partial breakage of the bond. In

---

[332] folding video: from YOUTUBE - Stoneybrook: https://youtu.be/YANAso8Jxrk

[333] http://www.ncbi.nlm.nih.gov/pmc/articles/PMC47901/ and prion disease: https://en.wikipedia.org/wiki/Prion

addition, in the cis configuration the R groups of adjacent amino acids are on the same side of the polypeptide chain. If these R groups are both large they can bump into each other. If they get too close they will repel each other. The result is that usually the polypeptide chain will be in the trans arrangement. In both α-helix and β-sheet configurations, the peptide bonds are in the trans configuration because the cis configuration disrupts their regular organization.
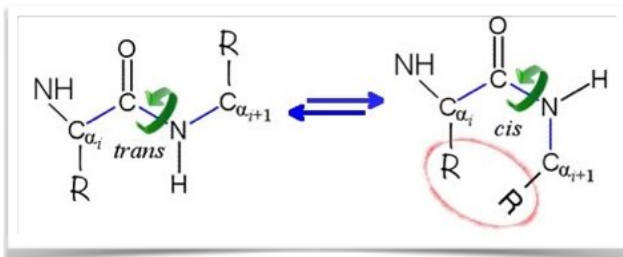
Peptide bonds involving a proline residue have a different problem. The amino group is "locked" into a particular shape by the ring and therefore inherently destabilizes both α-helix and β-sheet structures (see above). In addition, peptide bonds involving prolines are found in the cis configuration ~100 times as often as those between other amino acids. This cis configuration leads to a bend or kink in the polypeptide chain (→). The energy involved in the rotation around peptide bond involving a proline is much higher than that of a standard peptide bond; so high, in fact, that there are protein catalysts, peptidyl proline isomerases, such as PIN1 (OMIM:601052) that facilitate the cis-trans rotation.
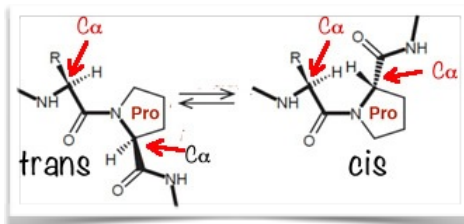
Hydrophobic R-groups: Many polypeptides and proteins exist primarily in an aqueous (water-based) environment. Yet, a number of their amino acid R-groups are hydrophobic. That means that their interactions with water will decrease the entropy of the system by leading to the organization of water molecules around the hydrophobic group, a thermodynamically unfavorable situation. This is very much like the process that drives the assembly of lipids into micelles and bilayers. A typical polypeptide, with large hydrophobic R groups along its length will, in aqueous solution, tend to collapse onto itself so as to minimize (although not always completely eliminate) the interactions of its hydrophobic residues with water. In practice this means that the first step in the folding of a newly synthesized polypeptide is, after it leaves the ribosomal tunnel, the collapse of the polypeptide chain onto itself so that the majority of its hydrophobic R groups are located internally, out of contact with water. In contrast, where there are no (or few) hydrophobic R groups in the polypeptide, the polypeptide will tend to adopt an extended configuration. On the other hand, if a protein comes to be embedded within a membrane (we will consider how this occurs later on), then the hydrophobic R-groups will tend to be located on the surface of the folded polypeptide that interacts with the hydrophobic interior of the lipid bilayer. Hopefully this makes sense to you, thermodynamically.

Acidic and basic R-groups: Some amino acid R-groups contain carboxylic acid or amino groups and so act as weak acids and bases. Depending on the pH of their environment these groups may be uncharged, positively charged, or negatively charged. Whether a group is charged or uncharged can have a dramatic effect on the structure, and therefore the activity, of a protein. By regulating pH, an organism can modulate the activity of specific proteins. There are, in fact, compartments within eukaryotic cells that are maintained at low pH in part to regulate protein structure and activity. In particular, it is common for the internal regions of vesicles associated with endocytosis to become

acidic (through the ATP-dependent pumping of H+ across their membranes), which in turn activates a number of enzymes (located within the vesicle) involved in the hydrolysis of proteins and nucleic acids.

Subunits and prosthetic groups: Many proteins contain non-amino acid-based components, known generically as co-factors. A protein minus its cofactors is known as an apoprotein. Together with its cofactors, it is known as a holoprotein. Generally, without its cofactors, a protein is inactive and often unstable. Cofactors can range in complexity from a single metal ion to quite complex molecules, such as vitamin B12. The retinal group of bacteriorhodopsin and the heme group (with its central iron ion) are co-factors. In general, co-factors are synthesized by various anabolic pathways, and so they represent the activities of a number of genes. So a functional protein can be the direct product of a single gene, many genes, or (indirectly) entire metabolic pathways.

Chaperones

The path to the native, that is, stable, functional state is not necessarily a smooth or predetermined one. The folding polypeptide can get "stuck" in a local energy minimum; there may not be enough energy, derived from thermal collisions, for it to get out again. If a polypeptide gets stuck, structurally, there are active mechanisms to unfold it and let the process leading to the native state proceed again (→). The process of unfolding misfolded polypeptides is carried out by proteins known as chaperones; we will call them folding/re-folding chaperones to distinguish them from other types of chaperones. Chaperones are protein-based molecular machines that are encoded by other genes. The unfolding of a misfolded protein by a chaperone requires energy, and so is coupled to a thermodynamically favorable reaction (such as ATP hydrolysis).
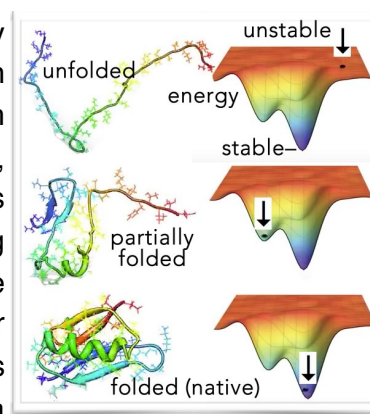
An important point to recognize is that chaperones do not determine the native state of a polypeptide–that is a function of the polypeptide's primary amino acid sequence. Consider, for example, the effect of a mis-sense mutation. Such a mutation can change the pattern of folding of a polypeptide; it may get caught more frequently in a mis-folded form. A folding/refolding chaperone can recognize such a mis-folded polypeptide, unfold it, either totally or partially, and release it to refold again, enabling the polypeptide to reach a functional structure.

There are many types of protein chaperones; some interact with specific polypeptides as they are synthesized and attempt to keep them from getting into trouble, that is, folding in an unproductive way. Others can recognize inappropriately folded polypeptides and, through coupling to ATP hydrolysis, catalyze the unfolding of the polypeptide, allowing the polypeptide a second (or third or …) chance to fold correctly. In the "simple" eukaryote, the yeast *Saccharomyces cerevisiae*, at least 63 distinct molecular chaperones have been recognized.[334]

Now you may well find yourself asking yourself, if most proteins are composed of multiple polypeptides, but polypeptides are synthesized individually, how do polypeptides come to be correctly

---

334 An atlas of chaperone–protein interactions in Saccharomyces cerevisiae: implications to protein folding pathways

assembled into functional proteins in a cytoplasm crowded with other proteins and molecules? Protein assembly often involves specific "assembly" chaperones, that bind to a newly synthesized polypeptide and either stabilize their folding, or hold them until they interact with other polypeptides to form the final, functional protein.[335] When proteins are synthesized in vitro, the absence of appropriate chaperones can make it difficult to assemble multisubunit proteins into functional proteins.

Another class of chaperones are known as "heat shock proteins." The genes that encode these proteins are expressed in response to increased temperature, assuming that the increase does not kill the cell immediately. At these higher temperatures collisions with surrounding molecules can lead a protein to unfold and misfold, it can become "denatured". Given what you know about polypeptide/ protein structure and gene expression, you should be able to develop a plausible model for how the expression of heat shock genes is regulated in response to temperature. Once expressed, heat shock proteins recognize denatured polypeptides, couple ATP hydrolysis reactions to unfold them, and then release the unfolded protein, giving them another chance to refold correctly.

Heat shock proteins help an organism adapt.[336] In classic experiments, when bacteria were grown at temperatures sufficient to turn on the expression of the genes that encode heat shock proteins, the bacteria had a higher survival rate when re-exposed to elevated temperatures compared to bacteria that had been grown continuously at lower temperature. Heat shock response-mediated survival at higher temperatures is an example of the ability of an organism to adapt to its environment - it is a physiological response. The presence of the heat shock system itself, however, is a selectable trait, encouraged by temperature variation in the environment. It is the result of evolutionary factors.

By now you might be asking yourself, how do chaperones recognize unfolded or abnormally folded proteins? In the case of a water soluble protein, most of the hydrophobic R-groups will be found within the interior of the correctly folded protein; in contrast, an unfolded protein will tend to have hydrophobic amino acid side chains exposed on its surface. The presence of these surface hydrophobic residues will lead to an tendency to aggregate; interacting hydrophobic regions will minimize hydrophobic-water interactions. Chaperones for water-soluble proteins recognize and interact with surface hydrophobic regions. For assembly chaperones, we can expect that specific sequences or structures in the target protein are recognized, which presumably is one reason that there are so many chaperone-like proteins, and specific chaperones for specific polypeptides and proteins.

Questions to answer
153. Why does it matter that rotation around a peptide bond is constrained?
154. How can changing the pH of a solution alter a protein's structure and activity?
155. Make a model of the structure of a polypeptide if all of its R-groups were hydrophilic or hydrophobic?
156. How might the presence of a folding/refolding-chaperone mitigate the effects of a mis-sense mutation?
157. How do assembly-chaperones facilitate the assembly of multi-polypeptide proteins?
158. Under what conditions might you expect heat shock proteins to be unnecessary for an organism?

Questions to ponder
- How does entropy drive protein folding and assembly?
- How might surface hydrophobic R-groups facilitate protein-protein interactions.

---

[335] Assembly chaperones: a perspective: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3638391/

[336] The heat shock response: life on the verge of death

Regulating protein activity, concentrations and stability (half-life)

Proteins act through their interactions with other molecules. Catalytic proteins (enzymes) interact with substrate molecules; these interactions lower the activation energy of the reaction's rate limiting step, leading to an increase in the overall reaction rate. At the same time, cells and organisms are not static. They must regulate which proteins they produce, the final concentrations of those proteins within the cell or organism, how active those proteins are, and where those proteins are located. It is primarily by altering proteins, which in turn influences gene expression, that cells and organisms adapt to changes in their environment.

A protein's activity can be regulated in a number of ways. The first and most obvious is to control the total number of protein molecules present within the system. Let us assume that once synthesized a protein is fully active. With this simplifying assumption, the total concentration of a protein, and the total protein activity in a system $[P_{sys}]$ is proportional to the rate of that protein's synthesis (dSynthesis/dt) minus the rate of that protein's degradation (dDegradation/dt), with dt indicating synthesis or degradation per unit time. The combination of these two processes, synthesis and degradation, determines the protein's half-life. Since both a protein's synthesis and degradation can be regulated, its half-life can be regulated.

The degradation of proteins is mediated by a special class of enzymes known as proteases. Proteases cleave peptide bonds via hydrolysis (adding water) reactions. Proteases that cleave a polypeptide chain internally are known as endoproteases - they generate two polypeptides. Those that hydrolyze polypeptides from one end or the other, generally release one or two amino acids at a time, are known as exoproteases. Proteases can also act more specifically, recognizing and removing specific parts of a protein in order to activate or inactivate it, or to control where it is found in a cell. For example, nuclear proteins become localized to the nucleus (typically) because they contain a NLS or they can be excluded because they contain an NES (see above). For these sequences to work they have to be able to interact with the transport machinery associated with the nuclear pores; but the protein may be folded so that they are hidden. Changes in a protein's structure can reveal or hide such NLS or NES sequences, thereby altering the protein's distribution within the cell and therefore its activity. As an example, a transcription factor located in the cytoplasm is, in terms of its effects on gene expression, inactive; it can become active if it enters the nucleus. Similarly, many proteins are originally synthesized in a longer and inactive "pro-form". When the pro-peptide is removed, cut away by an endoprotease, the processed protein becomes active. Proteolytic processing is itself often regulated.

The amount of a protein within a cell or organism is a function of the number of mRNAs encoding the protein, the rate that these mRNAs are recognized and translated, and the rate at which functional protein is formed, which in turn depends upon folding rates and their efficiency. It is generally the case that once translation begins, it continues at a more or less constant rate. In the bacterium E. coli, the rate of translation at 37ºC is ~15 amino acids per second.[337] The translation of a polypeptide of 1500 amino acids therefore takes about 100 seconds. After translation, folding and, in multisubunit proteins, assembly, the protein will function, assuming that it is active, until it is degraded.

---

[337] We are going to totally ignore the fact that different tRNAs are present at difference concentrations, which gives rise to what is known as codon bias. The presence of codons recognized by rare tRNAs slows down translation. to learn more look at Codon Bias as a Means to Fine-Tune Gene Expression: https://www.ncbi.nlm.nih.gov/pubmed/26186290

Many proteins within the cell are necessary all of the time. Such proteins are termed constitutive or house-keeping proteins. Protein degradation is particularly important for controlling the levels of "regulated" proteins, whose presence or concentration within the cell may lead to unwanted effects in certain situations. The regulated degradation of a protein typically begins when the protein is specifically marked for degradation (→). This is an active and highly regulated process, involving ATP hydrolysis and a multi-subunit complex known as the proteosome. The proteosome degrades the polypeptide into small peptides and amino acids that can be recycled. As a mechanism for regulating protein activity, however, degradation has a serious drawback, it is irreversible.
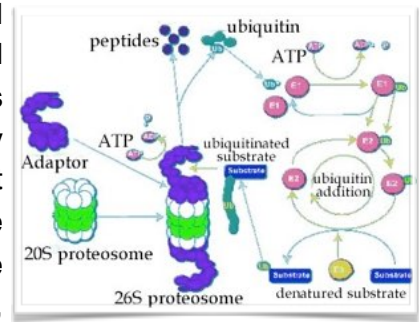
Allosteric and post-translational regulation

A reversible form of regulation is known as allosteric regulation, where a regulatory molecule binds reversibly to the protein altering the protein's structure, its activity, its location within the cell, and/or its stability (its half-life). When an allosteric effector binds to a protein, it is not covalently attached to the protein – its interactions are reversible, influenced by thermal factors. Allosteric regulators can act either positively or negatively. The nature of such factors is broad, they can be a small molecule or another protein. What is important is that the allosteric binding site is distinct from the enzyme's catalytic site. In fact allosteric means "other site". Because allosteric regulators do not bind to the same site on the protein as the substrate, changing substrate concentration generally does not alter their effects.

Of course there are other types of regulation as well. A molecule may bind to and block the active site of an enzyme. If this binding is reversible, then increasing the amount of substrate can over-come the inhibition. An inhibitor of this type is known as a competitive inhibitor. In some cases, the inhibitor chemically reacts with the enzyme, forming a covalent bond. This type of inhibitor is essentially irreversible, so that increasing substrate concentration does not overcome inhibition. These are therefore known as non-competitive inhibitors. Allosteric effectors are also non-competitive, since they do not compete with substrate for binding to the active site. That said, binding of substrate could, in theory, change the affinity of the protein for its allosteric effectors, just as binding of the allosteric effector changes the binding affinity of the protein for the substrate.

Proteins may be modified, through various covalent-modifications, after their synthesis, folding, and assembly - this process is known as post-translational modification. A number of post-translational modifications have been found to occur within cells. In general where a protein can be modified that modification can be reversed. The exception, of course, is when the modification involves protein degradation or proteolytic processing. There are many different types of post-translational modification, and we will consider them only generically. In general they involve the formation of a covalent bond linking a specific chemical group to specific amino acid side chains on the protein - these groups can range from a phosphate group (phosphorylation), an acetate group (acetylation), the attachment of lipid/hydrophobic groups (lipid modification), or carbohydrates (glycosylation). Such post-translational modifications are generally reversible, one enzyme adds the modifying group and another removes it. For example, proteins are phosphorylated by enzymes known as protein kinases, while protein
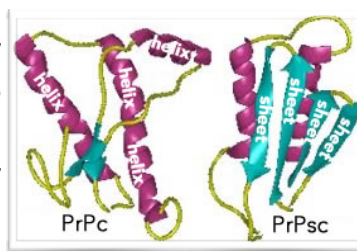
phosphotases remove such phosphate groups. Post-translational modifications act in much the same way as do allosteric effectors, they modify the structure and, in turn, the activity of the polypeptide to which they are attached. They can also modify a protein's interactions with other proteins, the protein's localization within the cell, or its stability.

Diseases of folding and misfolding

If a functional protein is in its native (or natural) state, a dysfunctional mis-folded protein is said to be denatured. It does not take much of a perturbation to unfold or denature many proteins. In fact, under normal conditions, proteins often become partially denatured spontaneously, normally these are either refolded, often with the help of chaperones or degraded through the action of proteosomes and proteases. A number of diseases, however, arise from irreversible protein mis-folding.

Kuru was among the first of these protein mis-folding diseases to be identified. Beginning in the 1950s, D. Carleton Gadjusek (1923–2008)[338] studied a neurological disorder common among the Fore people of New Guinea. The symptoms of kuru, which means "trembling with fear", are similar to those of scrapie, a disease of sheep, and variant Creutzfeld-Jakob disease (vCJD) in humans. Among the Fore people, Kuru was linked to the ritual eating of the dead. Since this practice has ended, the disease has disappeared. The cause of kuru, scrapie, and vCJD appears to be the presence of an abnormal form of a normal protein, known as a prion (mentioned above). We can think of prions as a type of anti-chaperone. The idea of proteins as infectious agents was championed by Stan Prusiner (b. 1942), who was awarded the Nobel Prize in Medicine in 1997.[339]

As we have noted previously, the protein (PrPc) responsible for Kuru and Scrapie is encoded by the PRP gene (OMIM:176640). It normally exists in a largely α-helical form. There is a second, abnormal form of the protein, PrPsc (the "sc" indicates scrapie); whose structure contains high levels of β-sheet (→). The two polypeptides have the same primary sequence. PrPsc acts to catalyze the transformation of PrPc into PrPsc. Once initiated, this reaction leads to a chain reaction and the accumulation

of PrPsc. As it accumulates PrPsc assembles into rod-shaped aggregates that appear to damage cells. When this process occurs within the cells of the central nervous system it leads to neuronal cell death and dysfunction, and severe neurological defects. There is no natural defense, since the protein responsible is a normal protein.

When the Fore ate the brains of their beloved ancestors, they inadvertently introduced PrPsc protein into their bodies. Genetic studies indicate that early humans evolved resistance to prion diseases, suggesting that cannibalism might have been an important selective factor during human evolution. Since cannibalism is not very common today, how does anyone get such diseases in the modern world? There are rare cases of iatrogenic transmission, that is, where the disease is caused by faulty medical practice, for example through the use of contaminated surgical instruments or when diseased tissue is used for transplantation.

---

[338] Carleton Gajdusek: http://www.theguardian.com/science/2009/feb/25/carleton-gajdusek-obituary

[339] Stanley Prusiner: 'A Nobel prize doesn't wipe the skepticism away' & http://youtu.be/yzDQ8WgFB_U

But where did people get the disease originally? Since the disease is caused by the formation of PrPsc, any event that leads to PrPsc formation could cause the disease. Normally, the formation of PrPsc from PrPc is very rare. We all have PrPc but very few of us spontaneously develop Kuru-like symptoms. There are, however, mutations in the gene that encodes PrPc that greatly increase the frequency of the PrPc → PrPsc conversion reaction. Such mutations may be inherited (genetic) or may occur during the life of an organism (sporadic). Fatal familial insomnia (FFI)(OMIM:600072) is due to the inheritance of a mutation in the PRP gene. This mutation replaces the aspartic acid normally found at position 178 of the PrPc protein with an asparagine. When combined with a second mutation in the PRP gene at position 129, the FFI mutation leads to Creutzfeld-Jacob disease (CJD).[340] If one were to eat the brain of a person with FFI or CJD one might well develop a prion disease.

So why do PrPsc aggregates accumulate? To cut a peptide bond, a protease (an enzyme that cuts peptide bonds) must position the target peptide bond within its catalytic active site. If the target protein's peptide bonds do not fit into the active site, they cannot be cut. Because of their structure, PrPsc aggregates are highly resistant to proteolysis. They gradually accumulate over many years, a fact that may explain the late onset of PrP-based diseases.

Questions to answer
159. A protein binds an allosteric regulator - what might happen to the protein?
160. How is the post-translational modification of a protein analogous to allosteric regulation? how is it different?
161. How is the proteolytic processing of a polypeptide like and unlike an allosteric effector or a post-translational modification.
162. Why do post-translational modifications (and their reversals) require energy?
163. How might a mutation that alters a signal sequence influence the translation, assembly, localization, and function of a polypeptide (protein)? What the effects of mutation on NLS or NES signals?

Questions to ponder
- Why is a negative allosteric regulator not considered a "competitive" inhibitor?
- How would a cell recover from the effects of an irreversible, non-competitive inhibitor?
- Why might a specific protein have a short half-life?

---

[340] OMIM entry for Creutzfeld-Jacob diease: http://omim.org/entry/123400

Chapter 9: Organizing and expressing genes in regulatory networks

In which we consider how DNA molecules, and the genes that they contain, are organized in a cell, how genes are recognized, and how their expression is controlled and organized into regulatory networks.

An important part of our approach to the study of biology is to think concretely about the molecules we are considering. Nowhere is this more important than with DNA. DNA molecules are very long and cells, even the largest cells, are (generally) small. For example, a typical bacterium is roughly cylindrical and ~2 µm in length and ~1 µm in circumference. Based on the structure of DNA, each base pair is ~0.34 nm in length. A region of DNA that is 1000 (10^3) base pairs long is therefore ~0.34 μm in length. A bacterium, like *E. coli*, has ~3 x 10^6 base pairs of DNA – that's a DNA molecule almost a millimeter in length or about 500 times the length of the cell in which it finds itself. That implies that at the very least the DNA has to be folded back on itself many times (←). A human cell has ~6000 times more DNA, resulting in a total length of greater than 2 meters of DNA per cell; these DNA molecules have to fit into a nucleus that is typically ~10 μm in diameter. In both cases, the DNA has to be folded and packaged in ways that allow it to fit within the cell and yet still be accessible to the various proteins involved in the regulation of gene expression and DNA replication. To accomplish this, the DNA molecule is associated with specific proteins; the resulting DNA:protein complex is known as chromatin.

doi:10.1038/nature21429

Top left: a diagram of a bacterial cell showing its DNA molecule. **Bottom left**: disrupting the cell membrane allows the DNA molecule within the cell to unfold. **Right**: the (color coded) chromosomes within the nucleus of a human cell in a compact (top) and an "exploded" view (bottom).

The study of how DNA is regulated is the general topic of epigenetics (on top of genetics), while genetics refers to the genetic information itself, the sequence of DNA molecules. A mutation will effect the sequence of DNA, it may or may not effect a gene, what a gene encodes and/or how the gene is regulated. If you consider a particular gene, based on our previous discussions, you will realize that to be expressed, transcription factor proteins must be able to find (by diffusion) and bind, through various intermolecular interactions, to specific regions of the gene's regulatory region(s), defined by their nucleotide sequences. But the way the DNA is organized

into chromatin, particularly in eukaryotic cells, can dramatically influence the ability of transcription factors to interact with and bind to their regulatory sequences. For example, if a gene's regulatory regions are inaccessible to protein binding because of the structure of the chromatin, the gene will be "off" (unexpressed) even if the transcription factors that would normally turn it on are present and active. As with essentially all biological systems, the interactions between DNA and various proteins can be regulated. At this point it is also worth remembering that there are typically only one to two copies of any particular gene within a cell, so we also have to consider the stochastic aspects of these molecular recognition processes.

Different types of cells can often have their DNA organized differently through the differential expression and activities of genes encoding proteins and non-coding RNAs involved in opening up (making accessible) or closing down (making inaccessible) regions of DNA. Accessible, transcriptionally active regions of DNA are known as euchromatin while DNA packaged so that the DNA is inaccessible to the binding of regulatory proteins is known as heterochromatin (→). A particularly dramatic example of this process occurs in female mammals. The X chromosome contains ~1100 polypeptide-encoding genes that play important roles in both males and females.[341] But the level of gene expression is influenced by the number of copies of a particular gene present within a cell. Only so many RNA polymerase complexes can move along a DNA molecule at a time, and each assembles a single RNA molecule as it moves; each ribosome assembles a single polypeptide as it moves along an mRNA molecule.

While various mechanisms can compensate for differences in gene copy number, this is not always the case. For example, there are genes in which the mutational inactivation of one of the two copies leads to a distinct dominant phenotype, a situation known as haploinsufficiency. This raises issues for genes located on the X chromosome, since XX organisms (females) have two copies of these genes, while XY organisms (males) have only one.[342] While one could imagine a mechanism that increased expression of genes on the male's single X chromosome, the actual mechanism used is to inhibit the expression of genes on one of the female's two X chromosomes. In each XX cell, one of the two X chromosomes is packed into a heterochromatic state, known as a Barr body, more or less permanently. The "decision" of which X chromosome is to be packed away ("inactivated") is made in the early embryo and appears to be stochastic - that means that it is equally likely that in any particular cell, either the X chromosome inherited from the mother or the X chromosome inherited from the father may be inactivated, that is, made heterochromatic. Importantly, once made this choice is inherited, the offspring of a cell will maintain the active/inactivated states of the X chromosomes of its parental cell – the inactivation event

---
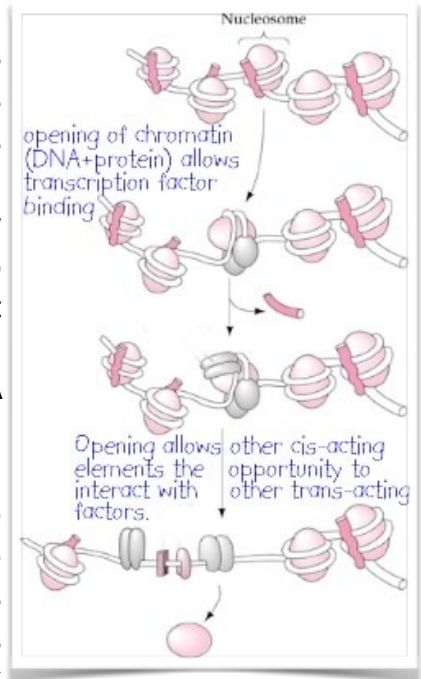
[341] Human Genome Project: Chromosome X: http://www.sanger.ac.uk/about/history/hgp/chrx.html

[342] The Y chromosome is not that serious an issue, since its ~50 genes are primarily involved in producing the male phenotype.

is inherited vertically.[343] The result is that XX females are epigenetic mosaics, they are made of clones of cells in which either one or the other of their X chromosomes have been inactivated. Many epigenetic events can persist through DNA replication and cell division, so these states can be inherited through the soma. There is even the possibility of evolutionary selection, for example, if the expression of one X chromosome leads to a reproductive advantage (more efficient cell division or survival) than is associated with the expression of the the other X chromosome; this can lead to evolutionary outcomes in which one clone out reproduces the other – a particular tissue may end up preferentially expressing genes on the maternal or the paternal X chromosome. A question remains whether epigenetic states can be transmitted through meiosis and into the next generation.[344] Most epigenetic information appears to be reset during the process of embryonic development.

Locating information within DNA

For genes to be useful there needs to be mechanisms by which specific genes can be recognized and expressed (transcribed) at specific times, at specific levels, and (in multicellular organisms) in specific types of cells.[345] Recognizing genes involves a two-component system consisting of regulatory nucleotide sequences that provide a molecular address; this molecular address (a type of bar code) identifies a specific region of a DNA molecule as well as which strand of the DNA should be transcribed, that is, used to direct RNA synthesis. The second component of this recognition system are proteins that recognize (and bind to) specific DNA sequences. The regulatory region of a gene can be simple and relatively short or long and complex. In some human genes, the regulatory region is spread over thousands of base-pairs of DNA, located "up-stream" and/or "down-stream" or within the coding region.[346] DNA (chromatin) within a chromosome can fold back on itself, allowing widely separated regions to interact.

The proteins that bind to regulatory sequences are known as transcription factors.[347] Many different transcription factors and transcription factor binding sites can be involved in the regulation of a gene's expression. In early genetic studies, two general types of mutations were identified that could influence the expression of a gene. "cis" mutations are located within the gene's regulatory region, often near the gene's coding (transcribed) region. In contrast "trans" mutations mapped to other, more distant sites, within the genome – often sites located on different chromosomes. Such mutations turned out to alter genes that encode transcription factors and other molecular components involved in gene expression, often proteins that bind specifically to sequences within the target gene's regulatory region. A particular transcription factor can influence the expression of many hundreds of genes. Transcription factors can act either positively to recruit and activate DNA-dependent, RNA polymerase or negatively, to block

---

[343]X Chromosome: X Inactivation: http://www.nature.com/scitable/topicpage/x-chromosome-x-inactivation-323

[344] Identification of genes preventing transgenerational transmission of stress-induced epigenetic states

[345] As an aside, are many transcribed DNA sequences that do not appear to encode a polypeptide or regulatory RNAs. It is not clear whether this transcription is an error, due to molecular level noise or whether such RNAs play a physiological role..
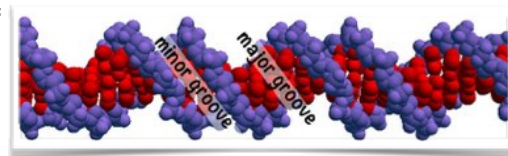
[346] Regulatory regions located far from the gene's transcribed region are known as enhancer elements.

[347] In prokaryotes transcription factors are often referred to as sigma (σ) factors.

polymerase binding and activation. Post-translational modifications and the binding of allosteric factors can alter the activity of transcription factors, while interactions with other proteins can alter binding specificity and down-stream effects on gene expression.

Genes that efficiently recruit and activate RNA polymerase will make many copies of the transcribed RNA and are said to be highly expressed. Generally, high levels of mRNA will lead to high levels of the encoded polypeptide. A mutation in a gene encoding a transcription factor can influence the expression of many genes, while mutations in a gene's regulatory sequence will directly effect only its own expression, unless of course the gene encodes a transcription factor or its activity influences the regulatory circuitry of the cell. Genes are organized in interacting systems, with associated feed-back mechanisms involved in homeostatic, adaptive, and developmental processes. An experimental point is often to determine whether a particular gene is a direct or an indirect target of a mutation or an environmental factor.

Transcription regulatory proteins recognize specific DNA sequences by interacting with the edges of base pairs accessible through the major and/or minor grooves of the DNA helix (→). There are a number of different types of transcription factors, with structurally distinct DNA binding domains; transcription factor proteins can be grouped in various (presumably evolutionarily related) families.[348] The binding affinity of a particular transcription factor to a particular regulatory sequence will be influenced by the DNA sequence as well as the binding of other proteins in the molecular neighborhood. We can compare affinities of different proteins for different binding sites by using an assay in which short DNA molecules containing a particular nucleotide sequence are mixed in a 1:1 molar ratio, that is, equal numbers of protein and DNA molecules:

$$DNA_{sequence} + protein \leftrightarrows DNA:protein.$$

After the binding reaction has reached equilibrium we can measure the percentage of the DNA bound to the protein. If the protein binds with high affinity the value is close to 100% and close to 0% if it binds with low affinity. In this way we can empirically determine the relative binding specificities (binding affinities for particular sequences) of various proteins, assuming that we can generate DNA molecules of specific length and sequence, which we can, and that we can purify proteins that remain properly folded in a native rather than in a denatured or inactive configuration, which may or may not be simple.[349] What we discover is that transcription factors (very much like the factors that mediate RNA splicing, see above) do not recognize a single, unique nucleotide sequence, but rather have a range of affinities for related sequences. This binding preference is a characteristic of each transcription factor protein; it involves both the length of the DNA sequence recognized and the pattern of nucleotides within that sequence. A simple approach to this problem considers the binding information present at each nucleotide position as independent of all others in the binding sequence, which is certainly not accurate but close enough for most situations. As noted before, the data is presented as a "sequence

---

[348] Determining the specificity of protein-DNA interactions: http://www.ncbi.nlm.nih.gov/pubmed/20877328

[349] Of course we are assuming that physiologically significant aspect of protein binding involves only the DNA, rather than DNA in the context of chromatin, and ignores the effects of other proteins, but it is a good initial assumption.

logo".[350] In such a plot, we indicate the amount of binding information at each position along the length of the binding site (→). Where there is no preference any of the four nucleotides is acceptable. The fewer the number of nucleotides that are acceptable the more information is present. Different transcription factor proteins produce different preference plots. As you might predict, mutations in a transcription factor binding site can have dramatically different effects. At sites that contain no specific information, a mutation will have no effect, whereas in sites of high information, any change from the preferred nucleotide(s) will have a severe effect on binding affinity, and can lead to a dramatic change in gene expression.
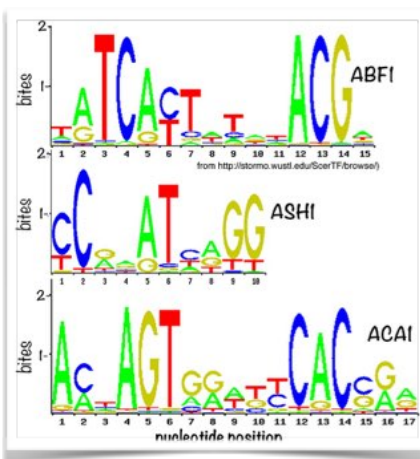
This is not to say that proteins cannot be extremely specific in their binding to nucleic acid sequences. For example, there are classes of proteins, known as restriction endonucleases and site specific DNA modification enzymes (methylases and acetylases) that bind to unique nucleotide sequences. For example the restriction endonuclease EcoR1 binds to (and cleaves) the nucleotide sequence GAATTC; change any one of these bases and there is no significant binding and no cleavage of the sequence. The recently described CRISPR CAS9 system for genetic manipulation is also highly specific, using a 22 nucleotide RNA to target an endonuclease to a specific site in the genome.[351] So the fact that transcription factors' binding specificities are more flexible suggests that there is a reason for such flexibility, although exactly what that reason is remains conjectural.

An important point to take away from this discussion is that most transcription factor proteins also bind to generic DNA sequences, but with low affinity. Such non-sequence specific binding is transient and rapidly broken by thermal motion. That said, since there are huge numbers of such non-sequence specific binding sites within a cell's DNA, much of the time transcription factors are found transiently associated with DNA. To be effective in recruiting a functional RNA polymerase complex to specific sites along a DNA molecule, the binding of a protein to a specific DNA sequence must be relatively long lasting. A common approach to achieving this outcome is for the transcription factor to be multivalent, that is, so that it can bind to multiple (typically two) sequence elements at the same time. This has the effect that if the transcription factor dissociates from one binding site, it remains tethered to the other; since the molecule is held, by this binding, close to the DNA it is more likely to rebind to its original site. In contrast, a protein with a single binding site is more likely to diffuse away before rebinding can occur. A related behavior involving the low affinity binding of proteins to DNA is that it leads to one-dimensional diffusion along the length of the bound DNA molecule.[352] Collisions are more likely to move the protein along the DNA molecule, rather than away from the DNA molecule. This enables a transcription factor protein to bind weakly to DNA and then move back and forth along the DNA molecule until it interacts with, and binds to, a high affinity site or until it dissociates completely. This

---

[350] Sequence logos: a new way to display consensus sequences: http://www.ncbi.nlm.nih.gov/pubmed/2172928

[351] The CRISPR-CAS9 system involves targeting a double-stranded DNA exonuclease to a specific site in a DNA sequence; it uses a RNA molecule to achieve very high levels of specificity. see CRISPR/Cas9 and Targeted Genome Editing

[352] As illustrated in the PhET applet:http://phet.colorado.edu/en/simulation/gene-expression-basics

type of "facilitated target search" behavior can greatly reduce the time it takes for a protein to find a high affinity binding site among millions of low affinity sites present in the genome.[353]

As the conditions in which an organism lives get more complex, the more dynamic gene expression needs to be. This is particularly the case in multicellular eukaryotes, where different cell types need to express different genes, or different versions (splice variants) of genes. One approach is to have different gene regulatory regions, that bind different sets of transcription factors. Such regulatory factors not only bind to DNA, they interact with one another. We can imagine that the binding affinity of a particular transcription factor will be influenced by the presence of another transcription factor already bound to an adjacent or overlapping site on the DNA. Similarly the structure of a protein can change when it is bound to DNA, and such a change can lead to interactions with DNA:protein complexes located at more distant sites, known as enhancers. Such regulatory elements, can be part of multiple regulatory systems.

For example, consider the following situation. Two genes share a common enhancer, depending upon which interaction occurs, gene A or gene B but not both could be active (→). The end result is that combinations of transcription factors are involved in turning on and off gene expression. In some cases, the same protein can act either positively or negatively, depending upon molecular context, that is, the specific gene regulatory sequences accessible, the other
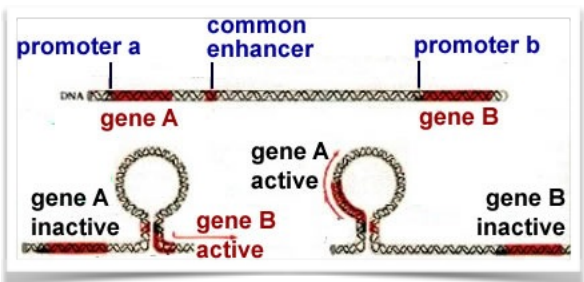
transcription factors expressed, and their various post-translational modifications. Here it is worth noting (again) that the organization of regulatory and coding sequences in DNA imposes directionality on the system. A transcription factor bound to DNA in one orientation or at one position may block the binding of other proteins (or RNA polymerase), while bound to another site it may stabilize protein (RNA polymerase) binding. Similarly, DNA binding proteins can interact with other proteins to control chromatin configurations that can facilitate or block accessibility to regulatory sequences. While it is common to see a particular transcription factor protein labelled as either a transcriptional activator or repressor, in reality the activity of a protein often reflects the specific gene under consideration, and its interactions with various accessory factors, all of which can influence gene expression outcomes.

The exact position on the DNA where RNA polymerase starts transcribing an RNA molecule is known as the transcription start site. Different regulatory sequences can lead to different transcription start sites. Similarly, in genes with introns, where transcription starts can determine which exons are included in the final transcript (mRNA molecule). Similarly, other factors can determine which exons are included and excluded in the final RNA, as well as where the encoded mRNA ends translation (↓). Where the RNA polymerase falls off the DNA, and so stops transcribing RNA, is known as the transcription termination site.
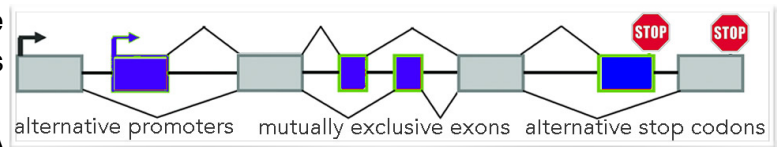
Once transcription initiates, the RNA polymerase moves away; as it clears the transcription start site, there is room for another polymerase

_____

[353] Physics of protein-DNA interactions: mechanisms of facilitated target search

complex to associate with the DNA, through interactions with transcription factors. Assuming that the regulatory region and its associated factors remains intact, the time to load a new polymerase will be much faster than the time it takes to build up a new regulatory complex *de novo*. This is one reason that transcription is often found to occur in bursts, a number of RNAs are synthesized from a particular gene in a short time period, followed by a period of transcriptional silence associated with the disassembly of the transcription start complex. A similar bursting behavior is observed in polypeptide synthesis (translation). The onset of translation begins with the small ribosomal subunit interacting with the 5' end of the mRNA; the assembly of this initial complex involves a number of components, and takes time to assemble, but once formed persists for awhile. While this complex exists multiple ribosomes can interact with the mRNA, each synthesizing a polypeptide, leading to bursts (multiple rounds) of translation. Once the translation initiation complex dissociates, it takes time, more time than just colliding with another small ribosomal subunit, for a new complex to form. The combination of transcriptional and translational bursting leads to noisy protein synthesis. Since cellular behavior can be influenced by changes in gene expression, these processes can lead to phenotypic differences between genetically identical cells. More details of the various intricacies of gene regulatory systems is best postponed to more advanced courses.

Questions to answer:
164. Make a model for how a transcription factor determines which DNA strand will be transcribed.
165. Make a model for how one could increase the specificity of a transcription factor protein for the regulation of a specific gene.
166. Describe the possible effects of mutations that alter the DNA-binding specificity of a transcription factor or a DNA sequence normally recognized by that transcription factor.
167. Consider a particular gene, what factors are likely to influence the length of its regulatory region?
168. How might you tell which X chromosome was inactivated in a particular cell of a female person?

Questions to ponder:
What factors might drive the evolution of overlapping genes?
How can over-lapping genes, or genes on different DNA strands influence each others' expression?

Interaction networks and model systems

As we come to analyze the regulation of gene expression, we recognize that they represent interaction networks. Interaction networks are a universal feature of biological systems, from the molecular and cellular to the social, ecological and evolutionary. These are generally organized in a hierarchical and bidirectional manner. So what exactly does that mean? Most obviously, at the macroscopic level, the behavior of ecosystems depends upon the interactions of organisms with one another. As we move down the size scale the behavior of individual organisms is based on the interactions between cells and tissues formed during the process of embryonic development and maturation. Since many of these interactions have a stochastic nature, chance plays a role. At the same time there are regulatory interactions and feed-back loops that can act to suppress some stochastic effects and serve to make biological behaviors more predictable. All of these interactions, and the processes that underlie particular biological systems, are the result of evolutionary mechanisms and historical situations, including past adaptations and non-adaptive events in ancestral populations.

Scientific studies of systems are driven by a number of factors, including ego, status, obsession, and financial reward. More idealistically, there is a curiosity to understand how it is that biological systems, in and of themselves, came to be, to behave and how they work. A related driver is the desire to understand in order to fix or avoid a disease, to be able to manipulate the world for the betterment of humanity, or more prosaically to make money or build a reputation. But there are a number of reasons that some questions cannot be answered directly; it might not be possible (or ethical) to carry out the necessary experiments. But here the evolutionary relationships between organisms come to our aid; we can choose organisms that are easier to study, develop faster, or are "simpler" in a way. By studying various "model" organisms, we can come to identify what can be relevant mechanisms. At the same time, it is important to explicitly recognize that the various "types" of organism that have been useful experimentally are each adapted to a specific environmental niche, generally evolving independently of others for millions to hundreds of millions of years. Even the most closely related of organisms, such as the great apes, a group that includes humans, display functionally significant differences. Once isolated, and maintained in the laboratory, we put organisms in an unnatural situation, which subjects them to different selection pressures. At the same time, isolated organisms are often maintained under conditions that reduce genetic variation - they become inbred. Such inbreeding can be desirable (for science), since it reduces variability and makes experiments more interpretable, while at the same time making them less realistic or relevant to "real" organisms.

Not withstanding the complexity of biological systems, we can approach them at various levels through a systems perspective, using specific models to study specific processes and behaviors. At each level, there are objects that interact with one another in various ways to produce specific behaviors – many of these systems are conserved, related to one another evolutionarily. To analyze a system at the molecular, cellular, tissue, organismic, social, or ecological level we need to define, understand, and appreciate the nature of the objects involved, how they interact with one another, and what behaviors and outcomes emerge from such interactions, in particular how such interactions influence the components of the system – does the system move to a new state or does it return, after a perturbation, to its original state. There are many ways to illustrate this way of thinking but we think that it is important to get concrete by looking at a (relatively) simple system and to consider how it behaves at the molecular, cellular, and social levels. Our next model system will be the bacterium *E. coli* and some of its behaviors, in particular how it behaves in isolation and in social groups and how it metabolizes the milk sugar lactose.[354] Together these illustrate a number of common regulatory principles that apply more or less universally to biological systems at all levels of organization.

E. coli as a model system

Every surface of your body harbors a flourishing microbial ecosystem. This is particularly true of the gastrointestinal system, which runs from your mouth and esophagus (with a detour to the nose), through the stomach, into the small and large intestine and the colon. (↓)[355] Each of these regions supports is own unique microbial community, known as a microbiome. These environments differ in terms of a number properties, including differences in pH and O_2 levels. Near the mouth and esophagus

---

[354] The Lac Operon: A Short History of a Genetic Paradigm

[355] The gut microbiome: scourge, sentinel or spectator?

O$_2$ levels are high and microbes can use aerobic (O$_2$ dependent) respiration to maximize the extraction of energy from food. Moving through the system O$_2$ levels decrease until anaerobic (without O$_2$) mechanisms are necessary. At different positions along the length of the gastrointestinal track microbes with different ecological preferences and adaptabilities are found.[356]

One challenge associated with characterizing the complexity of the microbiome present at various locations is that often the organisms present are dependent upon one another for growth and survival; when isolated from one another (and their normal environment) they do not grow. The standard way to count bacteria is to grow them in the lab using plates of growth media. Samples are diluted so that single bacteria land in isolation from one another on the plate surface. When they grow and divide, they form macroscopic (visible) colonies; we count the number of "colony forming units" (CFUs) per original sample volume; this number provides a measure of the number of individual viable bacteria present, or rather the number of bacteria capable of growing and dividing. If an organism cannot form a colony under the assay conditions, it will appear to be absent from the population. But as we have just mentioned some bacteria are totally dependent on others and therefore do not grow in isolation. To avoid this issue, newer molecular methods use DNA sequence analyses to identify which organisms are present without having to grow them.[357] The result of this type of analysis has revealed the true complexity of the microbial ecosystems living on and within us.[358]
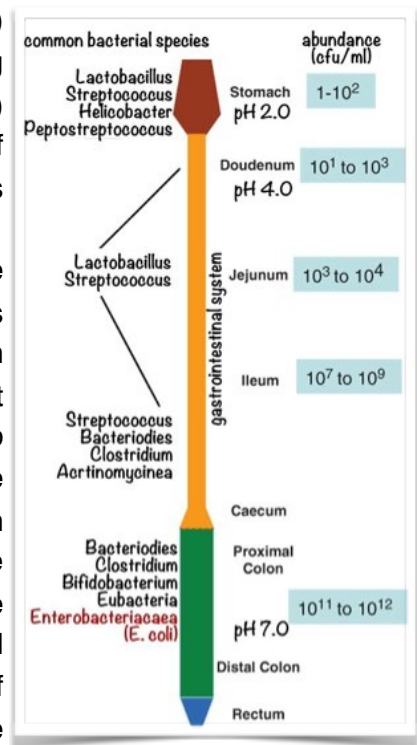
Much early work in molecular biology was carried out using a relatively minor member of this microbial community, *Escherichia coli*.[359] *E. coli* is a member of the Enterobacteriaceae family of bacteria and is found in the colons of birds and mammals.[360] *E. coli* is what is known as a facultative aerobe, it can survive in both anaerobic and an aerobic environments. This flexibility, as well as *E. coli*'s generally non-fastidious nutrient requirements make it easy to grow in the laboratory. Moreover, the commonly used laboratory strain of *E. coli,* known as K12, does not cause disease in humans. That said, there are other strains of *E. coli*, such as *E. coli* O157:H7 that are seriously pathogenic (disease-causing). *E. coli* O157:H7 contains 1,387 genes that are not found in the *E. coli* K12 strain and it is estimated that the two strains diverged from a common ancestor ~4 million years ago. The details of what makes *E. coli* O157:H7 pathogenic is a fascinating topic, but beyond our scope here.[361]

---

[356] The Gut Microbiome: Connecting Spatial Organization to Function and Gut biogeography of the bacterial microbiota

[357] Application of sequence-based methods in human microbial ecology: http://www.ncbi.nlm.nih.gov/pubmed/16461883

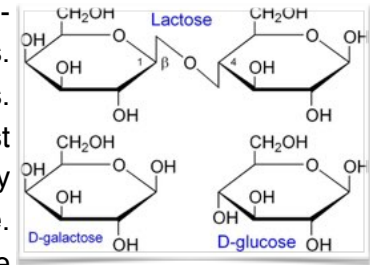[358] The human microbiome: our second genome: http://www.ncbi.nlm.nih.gov/pubmed/22703178

[359] virtual lab on E. coli: http://virtuallaboratory.colorado.edu/BioFun-Support/labs/EColi%20introduced/Coli.html

[360] Evolutionary ecology of E.coli

[361] Enterohemorrhagic E. coli (EHEC) pathogenesis: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3417627/

Adaptive behavior and gene networks (the lac response)

Lactose is a disaccharide (a sugar) composed of D-galactose and D-glucose (→). It is synthesized, biologically, exclusively by female mammals. Mammals use lactose in milk as a source of calories (energy) for infants. One reason, it is thought, is that lactose is not easily digested by most microbes. The lactose synthesis system is derived from an evolutionary modification of an ancestral gene that encodes the enzyme lysozyme. Through a gene duplication event and mutations, a gene encoding the protein α-lactoalbumin was generated. α-lactoalbumin is expressed in mammary glands, where it forms a macromolecular complex with a ubiquitously expressed protein, galactosyltransferase, to form the protein lactose synthase.[362]

E. coli is capable of metabolizing lactose, but only when there are no better (easier) sugars to eat. If glucose or other compounds are present in the environment, the genes required to metabolize lactose are turned off, they are not expressed. Two genes are required for *E. coli* to metabolize lactose. The first encodes lactose permease (see below). Lactose, being large and highly hydrophilic cannot pass through the *E. coli* cell membrane. Lactose permease is a membrane protein that allows lactose to enter the cell, moving down its concentration gradient. The second gene involved in lactose utilization encodes the enzyme β-galactosidase, which catalyzes the reaction that splits lactose into D-galactose and D-glucose, both of which can be metabolized by proteins expressed constitutively (that is, all of the time) within the cell. So how exactly does this system work? How are the lactose utilization genes turned off in the absence of lactose and how are they turned on when lactose is present and energy is needed. The answers illustrate general principles of the interaction networks controlling gene expression.

In *E. coli*, like many bacteria, multiple genes are organized into what are known as operons. In an operon, a single regulatory region controls the expression of multiple genes. It is common that multiple genes involved in a single metabolic pathway are located in the same operon, the same stretch of DNA. A powerful approach to the study of genes is to look for mutations that abolish a specific process, and so produce a discernible phenotype. As we said, wild type (that is, normal) *E. coli* can grow on lactose as their sole energy sources. So to understand lactose utilization, we can look for mutant *E. coli* that cannot grow on lactose.[363] To make the screen for such mutations more relevant, we first check to make sure that the mutant can grow on glucose. Why? Because we are not really interested (in this case) in mutations in genes that disrupt standard metabolism, such as the ability to use glucose. We seek to understand the genes involved in a specific process, lactose metabolism. Such an analysis revealed a number of distinct classes of mutations: some led to an inability to respond to the presence of lactose in the medium, others led to the de-repression, that is the constant expression of the two genes involved in the ability to metabolize lactose, lactose permease and β-galactosidase. In these mutant strains both genes were expressed whether or not lactose is present.

---

362 Molecular divergence of lysozymes and alpha-lactalbumin: http://www.ncbi.nlm.nih.gov/pubmed/9307874

363 The basic experimental approach involves a technique known as replica plating

By mapping, using the Hfr horizontal gene transfer system (see below), where these mutations are in the genome of *E. coli*, and a number of other experiments, the following model was generated (↓). The genes encoding lactose permease (lacY) and β-galactosidase (lacZ) are part of an operon, known as the lac operon. This operon is regulated by two distinct factors. The first is the product of a constitutively active gene, lacI, which encodes a polypeptide, the lac repressor, that assembles into a tetrameric protein that acts as a transcriptional repressor. In a typical cell there are ~10 lac repressor proteins present, and generally one or two copies of the lac operon. The lac repressor protein binds to sites in the promoter of the lac operon. When bound to these sites the repressor protein blocks the expression (transcription) of the lac operon. The

repressor's binding sites within the lac operon promoter appear to be its only functionally significant binding sites in the entire *E. coli* genome. The second regulatory element in the system is known as the activator site. It can bind the catabolyte activator protein (CAP), which is encoded by a gene located outside of the lac operon. CAP is a homodimer, that is, it is composed of two identical polypeptides. The DNA binding activity of CAP is regulated by the binding of an allosteric co-factor, cyclic adenosine monophosphate (cAMP). cAMP accumulates in the cell when nutrients, specifically free energy delivering nutrients (like glucose) are low. An increase in cAMP concentration [cAMP] acts as a signal that the cell needs energy. In the absence of cAMP, CAP does not bind to or activate expression of the lac operon, but in its presence (that is, when energy is needed), the CAP-cAMP protein is active, binds to a site in the lac operon promoter, and recruits and activates RNA polymerase, leading to the synthesis of lactose permease and β-galactosidase RNAs and proteins. However, even if energy levels are low, and [cAMP] is high, the lac operon will be inactive (not expressed) if lactose is absent because binding of the lac repressor protein to sites (labeled 0_1, 0_2, and 0_3) in the lac operon's regulatory region blocks polymerase recruitment.
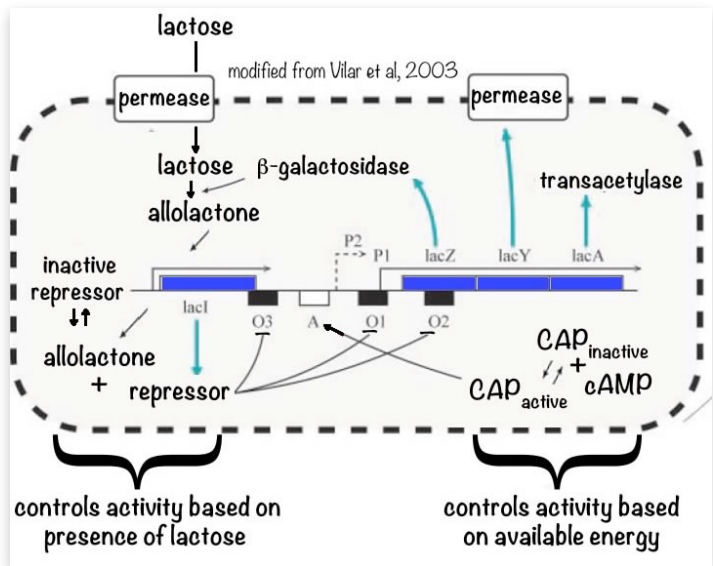
So what happens when lactose appears in the cell's environment? Well, obviously nothing, since the cells are expressing the lac repressor, so lactose permease is not present. Lactose cannot enter the cell without it. But that prediction assumes that, at the molecular level, the system works perfectly and deterministically. But this is not the case; the system is stochastic, that is, it is subject to the effects of random processes - it is noisy and probabilistic. Given the small number of lac repressor molecules per cell (~10), there is a small but significant (non-zero) chance that, at random, the lac operon will be free of bound repressor. If this occurs under conditions in which CAP is active, β-galactosidase and lactose permease will be expressed independently of the presence of lactose. If, however, lactose is present, we see the effect of a positive feedback loop (↓).[364] Those cells that have, by chance, expressed both the lacY (lactose permease) and lacZ (β-galactosidase) genes, a small percentage of the total cell

---

[364] Modeling network dynamics: the lac operon, a case study: http://www.ncbi.nlm.nih.gov/pubmed/12743100

population, will respond. Lactose will enter these cells, since the permease is present. Since β-galactosidase is also present, this lactose will be converted to allolactone, in a reaction catalyzed by β-galactosidase. Allolactone binds to, and inhibits the activity of the lac repressor protein. In the presence of allolactone the repressor no longer inhibits lac operon expression and there is a further increase (~1000 fold) in the rate of expression of the lacZ and lacY genes. In addition to generating allolactone from lactose, β-galactosidase catalyzes the hydrolysis of lactose into D-galactosidase and D-glucose, which are used to drive cellular metabolism. Through this process, the cell
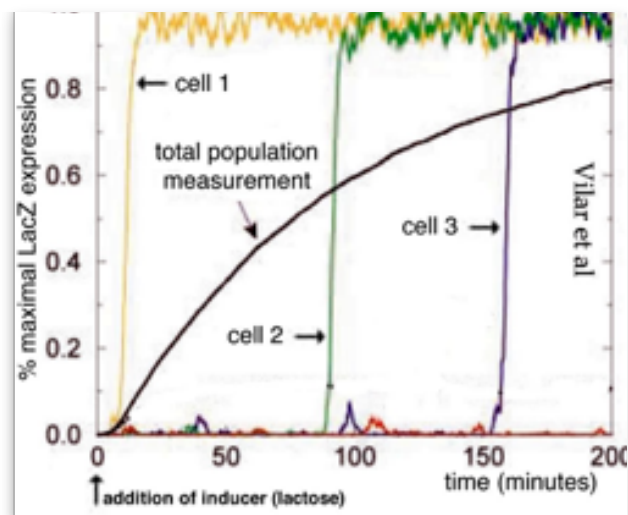
goes from essentially no expression to the full expression of the genes in the lac operon. Full expression allows the cells to metabolize lactose. At the same time, those cells that did not (by chance) express lactose operon will not be able to metabolize lactose, even though lactose is present outside the cells. So even though all of the *E. coli* cells present in a culture may be genetically identical, they can express different phenotypes due to the stochastic nature of gene expression.[365] You can play with these processes using the PhET gene expression applet (link here). In the case of the lac system, over time the noisy nature of gene expression leads to more and more cells activating their copy of the lac operon. Once "on", the operon will be expressed as long as lactose is present, since allolactone, derived from lactose, binds to and inactivates the lac repressor protein.

What happens if (and when) lactose disappears from the environment, what determines how long it takes for the cells to return to the state in which they no longer express the lac operon? The answer is determined by the effects of cell division and regulatory processes. In the absence of lactose, the [allolactone] falls and the lac repressor protein returns to its active (repressive) state, inhibiting lac operon expression. No new lactose permease and β-galactosidase proteins will be synthesized and their concentrations will fall based on the rate of their degradation (proteolysis). At the same time, and again because their synthesis has stopped, with each cell division the concentration of the lactose permease and β-galactosidase decreases by ~50%. With time the proteins are diluted, so the cells return to their initial state, that is, with the lac operon off and no copies of either lactose permease or β-galactosidase present.

Final thoughts on (molecular) noise, for now

When we think about the stochastic behaviors of cells, we can identify a few reasonably obvious sources of molecular and cellular level noise. First, there are generally only one or two copies of a particular gene within a cell. The probability that those genes are accessible and able to recruit transcription factors, associated proteins, and RNA polymerase molecules is determined by the frequency of productive collisions between regulatory sequences and relevant transcription factors together with their dissociation rates. Cells are small, and the numbers of different transcription factors

[365] An example of such behavior here: http://www.elowitz.caltech.edu/publications/Noise.pdf

can vary quite dramatically. Some transcription factors are present in high numbers (~250,000 per cell) while others (like the lac repressor) may be present in less than 10 copies per cell. The probability that particular molecules interact will be controlled by relative concentrations, diffusion, binding, and kinetic energies. This will influence the probability that a particular gene regulated by a particular transcription factor is active or not. Once on, transcriptional and translational bursting will produce gene products that can alter the state of the cell such that secondary, down-stream changes occur in gene expression and other cellular processes. These changes may (like the lac operon system) be reversible once the stimulus (lactose) is removed or they may be more or less irreversible, as occurs during cellular differentiation and embryonic development.[366]

Types of regulatory interactions

A comprehensive analysis of the interactions between 106 transcription factors and (many more) regulatory sequences in the baker's yeast *Saccharomyces cerevisiae* has revealed the presence of a number of common regulatory motifs.[367] These include (→):

- Auto-regulatory loops: A transcription factor binds to sequences that regulate its own transcription. Such interactions can be positive (amplifying) or negative (squelching).
- Feed forward interactions: A transcription factor regulates the expression of a second transcription factor; the two transcription factors then cooperate to regulate the expression of a third gene.
- Regulatory chains: A transcription factor binds to the regulatory sequences in another gene and induces expression of a second transcription factor, which in turn binds to regulatory sequences in a third gene, etc. The chain ends with the production of some non-transcription factor products.
- Single and multiple input modules: A transcription factor binds to sequences in a number of genes, regulating their coordinated expression. In most cases, sets of target genes are regulated by sets of transcription factors that bind in concert.
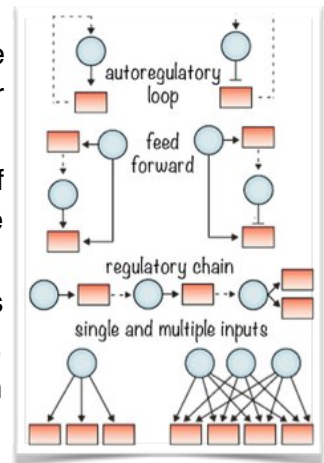
In each case the activity of a protein involved in an interaction network can, like the lac repressor, be regulated through interactions with other proteins, allosteric factors, and post-translational modifications. It is through such interactions that signals from inside and outside the cell can control patterns of gene expression leading to maintenance of the homeostatic state or various adaptations.

**Questions to answer & ponder:**
169. How would you design a regulatory network to produce a steady level of product?
170. How would you design a regulatory network that oscillates like a clock?
171. Draw out the predicted behavior of the various regulatory interactions as a function of time.
– How would you design a gene regulatory system (switch between states) that is irreversible?
– Can you imagine other regulatory schemes, in addition to the one's listed?

---

366 A single molecule view of gene expression: http://www.ncbi.nlm.nih.gov/pubmed/19819144

367 Transcriptional regulatory networks in *Saccharomyces cerevisiae*: http://www.ncbi.nlm.nih.gov/pubmed/12399584

Chapter 10: Cellular topology and intercellular signaling (Draft for FALL 2018).
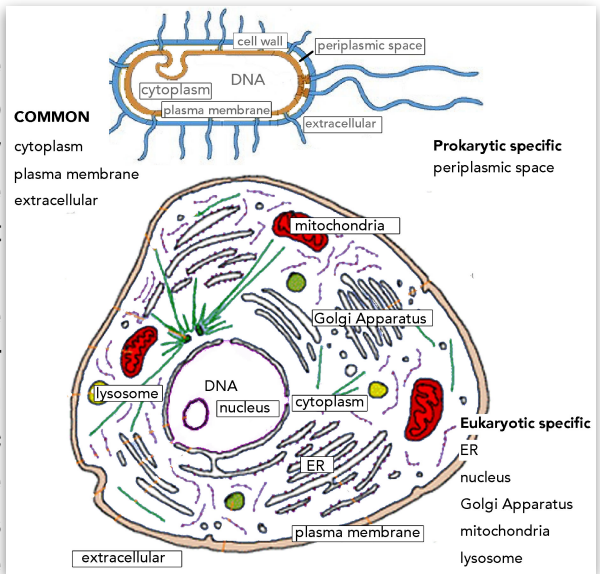
In which we consider the signals and receptors that control how proteins come to be where they are needed within cells and organisms, and how cells interact with one another through various signaling systems.
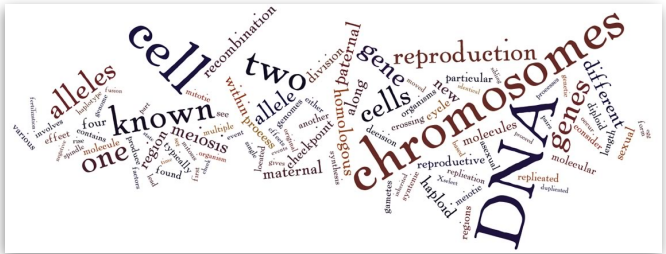
Prokaryotes, eukaryotes & cellular topology

As noted earlier (chapter 5), the cell is a bounded non-equilibrium system. The plasma membrane forms the unambiguous boundary between the rest of the universe and the cell. In prokaryotes, the cell is typically surrounded by a cell wall, that protects the cell from osmotic effects (among other things). Between the cell membrane and the cell wall is the periplasmic space. As we have discussed, the cell's metabolic activities occur within the space defined by its cell membrane, the cytoplasm. A polypeptide synthesized within the cytoplasm has a number of places it might end up, and like other biological processes these choices are controlled by signals and receptors. In a prokaryote (upper image →), a newly synthesized polypeptide can remain in the cytoplasm (where it can interact with the organism's genetic material, its DNA, since it is also located directly in the cytoplasm). Alternatively, a newly synthesized polypeptide can end up embedded within the plasma membrane (an integral membrane protein) or it pass through the membrane and be secreted. Secreted proteins (in a prokaryote) can remain within the periplasmic space, can become part of the cell wall, or can pass through the cell wall and into the external environment. The situation is more complex in eukaryotic cells (→), which contain a distinctive double membrane structure, the nucleus. The cell's genetic material, its DNA, organized into chromosomes, is located within the nucleus. The synthesis of RNA molecules, occurs within the nucleus. The membranes of the nucleus are elaborated within the cytoplasm into a network of membranes, known as the endoplasmic reticulum, the ER. There are a number of other intracellular membrane, the Golgi apparatus and small vesicles, involved in moving molecules to and from the plasma membrane. Finally, there are the mitochondria and (in plants) chloroplasts, double-membrane structures, with their own genomic DNAs, derived from apparent endosymbiotic events early in the history of eukaryotes (see chapter 6). The details of these membrane systems is a topic generally addressed in a course on cell biology, but from our point of view here, these different structures involve specific proteins that must, after their synthesis in the cytoplasm, be targeted to these various "organelles."

Targeting proteins to where they need to be: membrane proteins

So the question is, what determines where a polypeptide ends up? As you might suspect, there are signals and receptors involved, signals that are part of the polypeptide's primary (amino acid)

sequence and receptors that are encoded for by a number of other genes. The receptors are already present in the living cell, they are part of what is inherited from a cells progenitor, part of the continuity of life, that is, the cell theory. We will begin our description of polypeptide (protein) targeting with prokaryotes, because they are simpler, and we will consider how a newly synthesized polypeptide comes to end up in the cytoplasm, the plasma membrane, or outside of the plasma membrane.

In prokaryotes, the genomic DNA is located in the cytoplasm; there is therefore no barrier between a newly synthesized RNA molecule (produced by transcription) and the ribosomes and the other components involved in the translation of the RNA to generate a polypeptide. The newly synthesized mRNA molecule can interact with the ribosome, and direct polypeptide synthesis. For a water-soluble polypeptide, as opposed to a polypeptide that resides in, or passes through the membrane, no further "signals" are necessary. The ribosomal complex moves along the mRNA, the polypeptide is synthesized, passing through the ribosomal channel, and emerging into the cytoplasm. When the ribosome reaches a stop codon, release factor binds, leading to the disassembly of the ribosomal-mRNA-polypeptide complex. The ribosomal components, as well as the mRNA can then initiate a new mRNA -ribosome complex, to produce another polypeptide. The released (newly synthesized) polypeptide may fold on its own, or associate with other polypeptides to form a functional protein. Some of these folding steps may involve interactions with chaperones, but they all occur within the cytoplasm.

So what is going on with a polypeptide destined for the membrane. Clearly it has a different structure than a water-soluble protein. The first step is to recognize such a polypeptide. The general mechanism (and the only one we will consider) involves what is known as a signal sequence. A signal sequence is a sequence composed primarily of hydrophobic amino acids, typically 8 to 12 amino acids in length, and generally located near the polypeptide's N-terminus, the first part of the polypeptide to be synthesized. The presence of such a signal sequence marks the polypeptide as a membrane protein. As a polypeptide's signal sequence emerges from the ribosomal tunnel, its signal sequence is recognized through the binding of a cytoplasmic receptor, the signal recognition particle (SRP). SRP is composed of both polypeptides and a structural RNA. The binding of SRP a signal sequence causes translation to halt (although the mRNA-ribosome-nascent polypeptide-SRP complex remains intact. The mRNA-ribosome-nascent polypeptide-SRP complex diffuses within the cell until it engages an SRP-receptor located on the cytoplasmic surface of the plasma membrane; the SRP receptor is associated with a transmembrane polypeptide translocator (FIGURE ↓). When the mRNA-ribosome-nascent polypeptide-SRP+SRP Receptor complex forms; SRP can now disassociates from the ribosome-nascent polypeptide complex, and when it does, translation resumes and the nascent polypeptide interacts with the translocon and either folds to become embedded within the membrane, or passes through the membrane, and is released (secreted) on the other side. Typically, if the polypeptide is secreted, the signal sequence is removed by proteolytic processing.

INSERT signal sequence FIGURE HERE

Now let us consider the situation in eukaryotic cells. Although more complex, topologically, the same basic process applies. The difference is that the SRP receptor is not located in the plasma membrane, rather it located in the ER membrane. A protein with a signal sequence will be delivered to the ER membrane, or released into the lumen of the ER. From there, other signals will determine

whether the protein stays in the ER, moves to the Golgi apparatus, where is post-translationally modified, and may then move to the plasma membrane, or to some other membrane compartment within the cell. A protein in the lumen of the ER is effectively outside of the cytoplasm, and can be retained within a membrane compartment (such as the ER) or secreted from the cell. At this point, we will not concern ourselves with further details, except to say that whenever a protein is targeted to a specific cellular compartment, we can assume that the protein contains signals that are recognized by receptors that lead to its localization.

Nuclear targeting and nuclear exclusion in eukaryotes

All proteins are synthesized in the cytoplasm, but what happens if the protein functions in the nucleus, say as part of the DNA replication, DNA repair, RNA transcription, or RNA processing machinery? And what about a cytoplasmic protein that might interfere with such processes, if it were to find its way into the nucleus? Again we find the same pattern, there must be signals, typically sequences, within the polypeptide that indicate the protein should be located to or excluded from the nucleus. Such signals exist, and are referred to as nuclear localization or nuclear exclusion sequences. Such sequences interact with receptors, that is, molecular machines associated with the nuclear pore complex, that move the molecule from the cytoplasm into the nucleus in the case of a nuclear localization sequence (NLS) or from the nucleus to the cytoplasm, in the case of a nuclear exclusion sequence (NES).

An important point is that a protein can contain both nuclear localization and exclusion sequences; which are active can be determined by allosteric effector binding or post-translational modifications. An nuclear localization sequence may be hidden because the protein that contains it is interacting with another protein that makes it unavailable for interactions with its receptors. In this way, where a protein is within a cell, nucleus, cytoplasm, or both, can be regulated. As you can appreciate, regulating whether a protein, such as a transcription factor, is within the nucleus will also influence its function. Nuclear localization of a positively acting transcription factor can lead to the activation of a gene, while excluding a negatively acting transcription factor from the nucleus can de-repress the expression of a gene. Similarly, moving a transcription factor, whether positively or negatively acting, from the cytoplasm to the nucleus, will influence the expression of the genes the transcription factor regulates.The situation is different from that found in membrane targeting (the signal sequence-SRP system), which is considered irreversible - once a protein is inserted into a membrane or excreted from the cell, it cannot (generally) go back to being in the cytoplasm.

Questions to answer:
172. What are the components needed to insert of polypeptide/protein into or through a membrane? How might mutations in these proteins influence a polypeptide's localization within a cell?
173. Predict what would happen if a signal sequence were mutated. Which (in general) would have a greater effect, mutation to a hydrophobic or a hydrophilic residue?
174. How might you activate a NLS or NES sequence within a protein? How might such a sequence be rendered inactive?

Questions to ponder:
- Why might it be harmful if a membrane protein were synthesized in the cytoplasm (rather than directly inserted into a membrane)?

Intercellular signaling: signals, receptors & responses

The ability of cells to place proteins on their surface, and to secrete proteins into the extracellular space, opens up the possibility of various forms of signaling between cells, allowing cells to influence each other in various ways.[368] Here we will only consider the basics of such processes, more details will be found in later courses. Clearly such a system begins with the synthesis of the signaling molecule, which starts with turning on expression of the gene encoding the signaling molecule, its synthesis, processing, and secretion (or localization) to the cell surface. Similarly, for a cell to respond to a signal from another cell (or from itself), a cell has to express a receptor for the signal, and this receptor (typically a protein) needs to be on the receiving cell's surface. When the signal binds to the receptor it acts as an allosteric effector, changing the behavior of the receptor. Different signal-receptor combinations produce different types of changes in receptor activity, which typically initiates a cascade of events, that lead to changes in gene expression or changes in some other aspect(s) of cell behavior.

One aspect of signaling systems is the distance over which they act. Juxtacrine signaling occurs when the signaling and receiving cells need to touch one another molecularly, through membrane proteins on the two cells. Forming the interaction leads to changes in molecular conformations, that initiate a signaling cascade. It is possible that such an interaction influences both cells involved in the interaction. Such cell-cell interactions can involve cell adhering to one another, a critical step in forming multicellular organisms and appropriately organized tissues. Controlling when and where cells adhere to one another, and the signals such interactions send to the interacting cells plays a key role in embryonic development and cellular differentiation.

INSERT Signaling FIGURE HERE

When signaling molecules are released from a cell into the extracellular space they are free to diffuse. They can, in fact, interact with receptors present on the surface of cells within the immediate neighborhood of the cell secreting the signal. Those cells that have receptors on their surface for the signal can respond. In autocrine signaling, the cell that released the signal also has receptors for the signal; in a sense it is talking to itself.[369] If the signal interacts with receptors on neighboring cells, it is referred to as paracrine signaling. There is a third form of signaling, in which the signal is released from one type of cell (or cells in one region) and is transported throughout the body of the (multicellular) organism, typically through the blood stream, which is referred to as endocrine signaling. The details of these types of signaling systems are addressed in courses that deal with physiological processes, included in these are the functioning of the nervous system, the immune system, and various hormonal systems.

Molecules that provoke a signaling response are typically called agonists. Different agonists interact with agonist-specific receptors, typically composed of one or more integral membrane proteins, and interact to produce distinct "down-stream" molecular cascades that generally exploit post-translational modification or allosteric effects to activate or inactivate various enzymes and transcription

---

[368] Antebi et al. 2017. An operational view of intercellular signaling pathways

[369] as an example, see Glucagon regulates its own synthesis by autocrine signaling

factors. In general for each component of a signaling system, there are molecules (generally proteins) that act antagonistically; they inhibit the signaling process. For example, there are molecules, known as antagonists, that can bind signaling agonists or their receptors and so block signaling. Moreover, any one particular cell may express a number of different signaling pathway components; cells of different types will express different combinations of signaling systems, so they will be responsive to different incoming signals.

In cases where signaling leads to changes in gene expression, these changes can modify the behavior of the cell, or lead to it becoming a different type of cell. For any particular signaling input to a cell, there will be direct and indirect effects. For example, activation of a signaling system may lead to the activation (or repression) of a specific set of transcription factors. These can directly regulate the expression of a set of target genes. Some of these genes may themselves encodes transcription factors, or polypeptides that regulate transcription factor activity. The expression of these genes will, in turn, regulate other genes – these are considered indirect targets of the signaling system. Since which genes will be turned on or off will be influenced by the total set of transcription factors (regulated by various cellular processes) that are expressed and active in a cell, the response of different types of cells to the same signal can be different, and characteristic of the cell type. For example, a muscle cell might respond differently from a kidney cell to the same signal. Similarly, once a cell has been signaled to, the changes in the patterns of gene expression can lead to subsequent changes in cell morphology and behavior, including more changes in patterns of gene expression, it can differentiate, that is become different from what it was originally. The process of embryonic development consists of a series of signals and cellular responses that lead to the specialization of cells, the development of tissues, and organ systems. Normally, this process of signal-driven differentiation is irreversible. It proceeds in one and only one direction. The processes result in what is known as terminal differentiation. Only recently have strategies been developed that can reverse these effects.

Cellular reprogramming: embryonic and induced pluripotent stem cells

An important question that was asked by early developmental biologists was, is cellular differentiation due to the loss of genetic information? Is the genetic complement of a neuron different from a skin cell or a muscle cell? This question was first approached by Briggs and King in the 1950s through nuclear transfer experiments in frogs. These experiments were extended by Gurdon and McKinnell in the early 1960s; they were able to generate adult frogs via nuclear transfer using embryonic cells.[370] The process was inefficient however - only a small percentage of the nuclei taken from differentiated cells supported normal embryonic development. Nevertheless, these experiments suggested that it was the regulation rather than the loss of genetic information that was important in embryonic differentiation. That said, it is increasingly clear that, particularly in cells that no longer divide, such as neurons and muscle cells, that there is an accumulation of mutations, which appear to influence their behavior.[371]

---

[370] The egg and the nucleus: a battle for supremacy: http://www.nobelprize.org/mediaplayer/?id=1864

[371] see: Individual neurons may carry over 1,000 mutations

In 1996 Wilmut et al used somatic cell nuclear transplantation to clone the first mammal, the sheep Dolly. Since then many different species of mammal have been cloned, and there is serious debate about the cloning of humans. In 2004, cloned mice were derived from the nuclei of olfactory neurons using a method similar to that used by Gurdon. These neurons came from a genetically engineered mouse that expressed the fluorescent protein GFP in most cell types. After the nuclei of a mature (haploid) oocyte was removed, a neuronal nucleus was introduced. Blastula derived from these cells were then used to generate totipotent embryonic stem cells from cells of the inner cell mass. It was the nuclei from these cells that were then transplanted into enucleated eggs. The resulting embryos were able to develop into fully grown and fluorescent mice, proving that neuronal nuclei retained all of the information required to generate a complete adult animal.

The process of cloning from somatic cells is inefficient – many attempts had to be performed, each using an egg, to generate an embryo that is apparently normal (most embryos produced this way were abnormal). At the same time, there are strong ethical concerns about the entire process of reproductive cloning. For example the types of cells used, embryonic stem cells, are derived from the inner cell mass of mouse or human embryos. Embryonic stem cells can be cultured *in vitro* and under certain conditions can be induced to differentiate into various cell types. Since the generation of totipotent human embryonic stem cells involves the destruction of a human embryo, it raises a number of ethical issues and concerns, particularly given the persistent inequalities in modern society.[372]

In a breakthrough series of studies, Takahashi and Yamanaka (2006) determined that introducing a set of four transcription factors (Oct3/4, Sox2, c-Myc, and Klf4) into terminally differentiated cells led some of the transfected cells reverse their differentiation, and return to a more pluripotent state, that is a state that can subsequently differentiate into any of a number of other cell types.[373] This process of dedifferentiation has been found to be robust, and the dedifferentiated cells produced are known as "induced pluripotent stem cells" or iPSCs. These "induced pluripotent stem cells" behave much like embryonic stem cells. The hope is that iPSCs derived from a patient can be used to generate tissues or even organs that could be transplanted back into the patient, and so reverse and repair disease-associated damage.

Questions to answer:

175. What cellular factors determine how (or whether) a cell responds to a particular signaling molecule?
176. What is necessary for cells to become different from one another - for example how do muscle cells and skin cells come to be different from one another?
177. Based on your understanding of the control of gene expression, outline the steps required to reprogram a nucleus so that it might be able to support embryonic development.

Questions to ponder:

⁻ Why, if differentiation is normally uni-directional and irreversible, is it possible to artificially reprogram somatic cells to an "earlier" state, that is, induced pluripotent stem cells? Why doesn't this happen all the time in your body?
⁻ What are the main ethical objections to human cloning? What if the clone were designed to lack a brain, and destined to be used for "spare parts"? Does that change anything, or make things worse?

---

[372] J. Gray. 2017. A History of the Future: how writers envisioned tomorrow's world

[373] Takahashi & Yamanaka. 2006. Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors.

In which we consider the behavior of genes during the course of asexual and sexual reproduction, how they interact with one another, both along chromosomes and more generally, and how they can be manipulated and studied.

Chapter 11: A short review of concepts with which you should already be familiar

In which we reflect back on what we have learned and review the meanings and implications of various terms and processes associated with molecular mechanisms, genes, gene products, and cellular reproduction, together with a few key terms that we will be using over and over again this semester.

Words, terms, and processes we (really) need to understand: As you will have noted, biology is full of words. Some of these words are familiar and others are strange or have biological meanings that differ from their common usage. Up through this point, you have probably run into a number of these words. Science is generally referred to as a discipline, and one critical aspect of a discipline is discipline, meaning that there are strict rules involved, there are ways to act, behave, and speak. The discipline of science in general, and biology in particular, is that what words mean is unambiguously defined, and not subject to personal preference – in a sense we are not allowed to be creative in the meaning of scientific words; this is probably one reason people like to invent new words to describe new phenomena and new ideas. Therefore, we will begin by considering a number of words and what they mean scientifically, as opposed to colloquially. To understand the meaning of a word, you have to be able to use it correctly, apply it when appropriate, and understand its implications. Don't be afraid to ask a question if you are unsure whether a particular word is appropriate to a particular situation.

So just to review (and if you feel the need, look at chapters 7 & 8), consider what we mean by a **gene**: within a cell, a gene is a stretch of DNA that can be **expressed.** A gene includes the DNA sequences involved in determining where and when the gene is expressed. So what does it mean to be "expressed"? To say that a gene is expressed, we mean that an **RNA molecule**, complementary to the sequence of the DNA, is generated through the process of **transcription**. Transcription is mediated through the action of DNA-dependent, RNA polymerases; where such enzymes bind to the DNA and act, that is, where RNA synthesis starts, is determined by where specific sets of transcription factors (proteins) bind to specific sequences within the DNA; these sites of transcription factor binding are part of the gene's regulatory sequences. The binding of transcription factors acts to recruit and activate the RNA polymerase. A gene's regulatory sequences can be located near to the gene's transcribed region (generally referred to as the gene's promoter) or at more distant sites, known as enhancer elements.

The sum of all the DNA molecules within a cell constitutes the cell's genome. Generally, each cell in an organism contains a full copy of the genome. The cell's genome is characteristic, and serves to define the type of organism, the species, the cell is, or is part of. Organisms of the same species have extremely similar genomes, more similar than do organisms of different species – genome differences define a species. A cell can have either one complete genomic copy, in which case it is known as **haploid**, or it can contain two, in which case it is known as **diploid**. In certain cases, a cell can have more than two copies of its genome, in which case it is termed triploid (3 copies), tetraploid (4 copies), or polyploid (> four copies).

As we will go into greater detail, a haploid or a diploid cell can divide asexually to produce two haploid or diploid cells, respectively. So what does sexual and asexual mean, exactly? **Asexual reproduction** involves only a single cell, a single individual. The genome of the cell is duplicated through the process of **DNA replication** (which is mediated primarily by DNA-dependent, DNA polymerases and other factors) and then the cell splits into two. Similarly, at the organismic level, there is no essential need for cooperation between different organisms, or different cells, for reproduction to occur. Of course, for such a process to continue there has to be growth of the cell between cell division events. Generally the cell doubles in volume and mass between one division and the next. The growth of the cell involves the import of energy and other materials into the cell, and their metabolic transformation into various cellular parts, proteins, nucleic acid polymers, lipids, etc. There is a continuity, one cell becomes two; this is the simplest version of the cell theory of life.

The process of **sexual reproduction** is more complex. Two different cells, generally but not always from two different organisms, have to cooperate with and fuse with one another. Such cooperation requires them to recognize one another as appropriate fusion partners. Sexual reproduction involves a diploid cell that first generates a number of haploid cells, known as **gametes**, through a process known as meiosis. Typically, gametes from two different organisms come into proximity through the process of mating and fuse with one another; their initially distinct plasma membranes become one, thereby forming a new diploid cell (organism). Some people might say that this is when life begins, but they would be confused – life began ~3.4 billion years ago. Both gametes are alive, as is the **zygote**, the cell formed by their fusion. That said, the fusion of gametes generates a genetically distinct (and so new) organism.

The two modes of reproduction have different characteristics. In asexual reproduction, the versions of genes, known as **alleles,** within a cell evolve together, as a group - there is no simple way to remove deleterious alleles from future progeny, although the processes of horizontal gene transfer, that is, transformation, conjugation, or transduction, can modify genomes (something we will discuss). In contrast sexual reproduction (including the process of meiotic recombination, which we will consider in detail) enables alleles to move more or less independently of one another. Sexual reproduction is also associated a number of features, particularly in multicellular organisms. Sexual dimorphism means that the two gametes, and the organisms that produce them, can be different in morphology and behavior. Such differences can lead to sexual selection, a distinctive process associated with the evolution of a range of traits and with a range of evolutionary implications.[374]

Questions to answer and ponder:
178. How are transcription and translation similar, how are they different?
179. Within a gene, what signals and signal binding proteins are involved in gene expression? make a diagram.
180. How would having two copies of a gene (in a diploid cell) alter the behavior the cell?

Where do genes, alleles, and mutations come from?

When we think about genes, there are two issues to consider. The first is where do genes come from? The most obvious (and perhaps unsatisfying) answer is that our genes come from our ancestors,

---

[374] here is an interesting book on the topic: The Mating Mind by Geoffrey Miller.

our parents (at least for the sexual among us), through the process of DNA replication. Unfortunately, this leaves the ultimate origin of genes shrouded in mystery. As discussed earlier all life on Earth appears to be descended from a **last universal common ancestor** (LUCA), and this organism already had lots of genes. These genes arose even earlier, through processes involving various molecular systems that were active before the appearance of LUCA. New genes have been observed to appear *de novo* out of DNA sequence, in various organisms, in particular the fruit fly *Drosophila*.[375] Perhaps even more surprising, many of these de novo genes appear to have become essential rather quickly. [376] A number of putative *de novo* genes have been identified in humans.[377]

Once DNA (nucleic acid) molecules and genes existed, new versions of genes (alleles) can appear through processes of **mutation** and recombination, which lead to alterations in DNA sequence. Moreover, an existing gene can appear through the process of gene duplication, leading to the production of paralogs, and disappear through gene deletion. A number of studies, beginning with the classic Luria-Delbruck experiment (which we will discuss in detail), indicate that these processes (mutation, recombination, deletion, and duplication) occur largely randomly, based on the molecular nature of DNA, various molecular mechanisms active in the cell, and environmental effects (chemicals and radiation). They appear randomly, and not to meet the adaptive needs of the organism. Once a mutation arises it can, however, effect the organism's **phenotype**, that is the traits displayed by an organism. These phenotypic effects can include effects on reproductive success. The most severe of such effects is lethality, generally arising because the mutation inactivates an essential gene, a gene whose activity (gene product) is necessary for the organism's survival, that is the maintenance of life. Evolutionary processes act to "select" against mutant alleles that reduce reproductive success (**negative selection**) and increase the frequency of mutant alleles that improve it (**positive selection**). Generally, environmental factors and preexisting adaptations and behaviors determine the selective pressure on a new allele. There are also processes, such as **genetic drift** and various **founder** and **bottleneck** effects, that can influence which alleles are found within a population. These principles apply both to the cells within a multicellular organism (somatic selection) as well as organisms within a population.

Alleles

The specific version of a gene, defined by the gene's DNA sequence, is known as an allele. In a diploid organism, the two copies of the gene can have different sequences, they can be different alleles. If the two alleles in a diploid organism are the same, the organism is said to be **homozygous** for that gene, if they are different it is said to be **heterozygous** for that gene. An organism can be homozygous for some genes and heterozygous for others. Different alleles can be expressed differently, due to differences in their regulatory sequences, and they can encode different gene products due to differences in their transcribed and coding (in the case that the gene encodes a

---

[375] see: Schlotter. 2015. Genes from scratch – the evolutionary fate of de novo genes and Fact or fiction: updates on how protein-coding genes might emerge de novo from previously non-coding DNA.

[376] see New genes in Drosophila quickly become essential and The Goddard and Saturn Genes Are Essential for Drosophila Male Fertility and May Have Arisen De Novo.

[377] De novo origin of human protein-coding genes

polypeptide) regions. Within a population of organisms of the same species, there can be multiple versions of a particular gene, multiple alleles. Later we will use the EXAC browser to visualize the alleles of various human genes. Closely related species often share many genes, organized along chromosomes in similar patterns, a situation known as synteny, something that can be visualized using the Genomicus web tool. Different species are likely to have different alleles, a result of the divergent evolutionary histories; they can differ in terms of synteny, with genes deleted, duplicated, or moved to different chromosomal positions within the genome (genomic rearrangements.) Some of the differences between alleles have little or no impact on the function of a gene or the gene product that it encodes, these allelic variants can all be considered normal or **wild type**. In contrast, other alleles are associated with specific traits, or versions of a trait – in some cases these are traits associated with disease, disease susceptibility, developmental defects, or cellular and organismic lethality; in other cases, they are associated with evolutionary novelties, the traits that distinguish one species from another. A mutation in a wild type allele is much more likely to lead to a defect than an improvement in the gene product's function, or a useful new trait, but such beneficial mutations do occur; they appear (together with other environmental and selective factors) to drive evolutionary processes.

Phenotypes

The traits of an organism, including how it develops and responds to its environment, are determined by its genome, that is all of the genes it contains. The various regulatory interactions that occur between genes, gene products, and the cells' metabolic processes are known as its epigenome. The epigenome includes non-DNA sequence components, including how the DNA is packaged within the cell, which can influence which genes are, or can be, expressed in a particular cell type, or in response to particular signals. As we will explore in detail, all of the observable or measurable aspects of an organism constitute its phenotype. Phenotypes can range from blood type, allergic reactions, susceptibility or resistance to disease, height, skin color, eye color, the speed of reflexes, essentially anything and everything about an organism that you can observe and measure. In some, relatively rare cases there is a 1 to 1 correspondence between which allele of a gene an organism carries and the specific trait(s) it displays. This type of allele:trait association was used by Gregor Mendel to establish his rules of inheritance (see below). In the case of haploid organisms, genotype (the alleles present in an organism) maps to the organism's traits in a relatively simple manner – an example that we will consider in greater detail is involved in antibiotic resistance. A bacteria that contains a functional copy of a gene that confers resistance to an antibiotic is resistant to that antibiotic. A mutation that inactivates that gene's expression or the gene product it makes can leave the bacteria susceptible to the antibiotic. Of course it is a mistake to think that the gene and the product that it encodes are the only components needed for antibiotic resistance; no gene acts alone – for a gene to influence a phenotype (such as antibiotic resistance) the gene needs to be recognized and expressed (transcribed), the encoded protein synthesized (translated), and delivered to the right location (targeting). Even a simple gene (allele) → phenotype relationship is based on the functioning of a complex biological system, a system composed of thousands of genes and gene products. Most traits are based on many gene products, and often the impact of a particular allele of a particular gene is subtle, something that can be identified through complex molecular genetic studies, which we will consider anon.

The relationship between an allele and a phenotype is more complex in a diploid organism since there are two copies of most genes (with the possible exception of genes associated with sex determination, which we will return to). The two copies of the gene present can be the same, in which case the organism is referred to as homozygous for that genetic locus (gene), or they can be different, or heterozygous at that genetic locus. These terms always refer to a specific genetic locus. If an organism is homozygous for all genetic loci, it is generally the result of extensive in-breeding. Of course, these terms do not apply to prokaryotes which are generally haploid.

Now consider a trait that is associated with the presence of a particular allele. If the trait is visible when the locus is heterozygous for that allele, the allele is referred to as **dominant** to whatever the other (different) allele might be. On the other hand, if the trait is not apparent when the locus is heterozygous, but is visible when the locus is homozygous for the allele, it is referred to as **recessive**. Finally, if the trait displayed by an organism that is heterozygous for a particular locus is different from either of the homozygous versions, the alleles are referred to as **co-dominant** or **semi-dominant**. In such cases, the nature of the phenotype observed will depend on exactly which alleles are involved. We will return to, and consider all of these topics in greater detail, when we consider the interactions between combinations of alleles. Similarly, the extent of the appearance of a phenotype, known as its penetrance and its expressivity can be influenced by the other alleles within the genome, the organism's genetic background. Remember however, the terms recessive and dominant refer to alleles that are associated with visible traits. Most alleles are neither strictly recessive nor dominant, and contribute in complex ways to a number of traits. Because it is easier to make sense of things we will generally start, at least initially, with strictly dominant and recessive alleles, and then get more complex.

Questions to answer and ponder:

181. Based on your understanding of DNA, draw out (schematically) the relationship between a specific allele and the phenotypic traits it is associated with.
182. Why might the mutation of gene not be associated with any one specific phenotypic trait?

Muller's Morphs

Another way to look at alleles is from a functional perspective. This was the approach taken by Herman J. Muller in the 1920s and 30s. He exploited work done in the fruit fly *Drosophila*, in which a number of gene and regional chromosomal duplications and deletions had been constructed by geneticists, something made possible by unique aspects of chromosome organization in the salivary glands of the fly (→). These cells are polyploid, and each chromosome contains more than 1000 DNA strands.[378] Based on the analysis of various


modified from Zhimulev et al., 2012
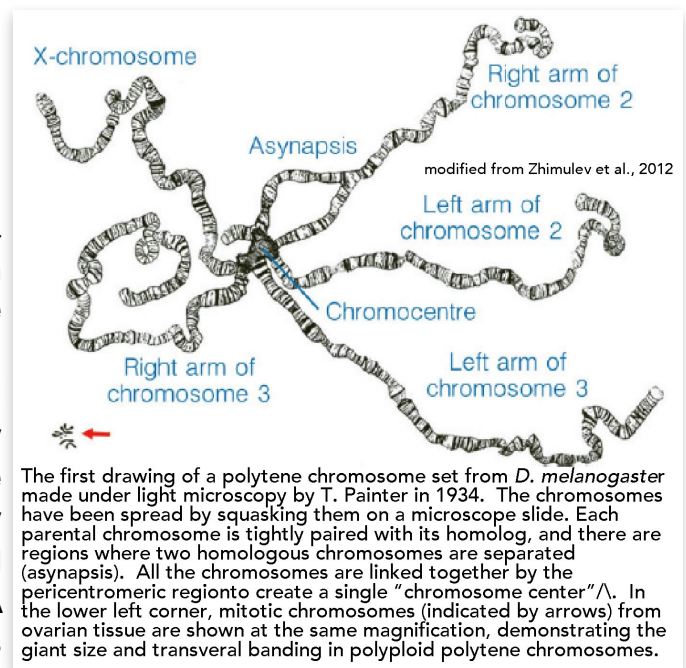
The first drawing of a polytene chromosome set from *D. melanogaster* made under light microscopy by T. Painter in 1934. The chromosomes have been spread by squashing them on a microscope slide. Each parental chromosome is tightly paired with its homolog, and there are regions where two homologous chromosomes are separated (asynapsis). All the chromosomes are linked together by the pericentromeric region to create a single "chromosome center"/\. In the lower left corner, mitotic chromosomes (indicated by arrows) from ovarian tissue are shown at the same magnification, demonstrating the giant size and transveral banding in polyploid polytene chromosomes.

---

[378] Banding patterns in Drosophila melanogaster polytene chromosomes correlate with DNA-binding protein occupancy.

mutations he was able to place mutations into distinct functional (with respect to a particular phenotype) classifications: that is amorphic, hypomorphic, hypermorphic, antimorphic, and neomorphic. These classes are compared to the wild type ("normal") version of the allele. At this point we will not consider mutations that have no functional effects on the gene. Compared to the level of functional gene product produced by a wild type allele, an **amorphic** allele has no function - it might not be expressed, or if expressed, the gene product may not carry out the trait-specific functions of a wild type gene product. Importantly, an amorphic allele does not interfere in any way with the expression or functioning of the wild type gene product encoded by the other allele in a diploid cell. Amorphic alleles are also known as **null** or loss of function (LoF) alleles (see below). In a similar manner, a **hypomorphic** allele has less functional activity, whatever that might be, compared to a wild type allele, whereas a **hypermorphic** allele has more, but the same, functional activity as the wild type allele. Again, for both hypo- and hypermorphic alleles, the mutant gene product does not interact with the wild type gene product. In contrast, an **antimorphic** allele is not only non-functional with respect to a trait-specific function, but it interacts with and inhibits the activity of the wild type gene product.

The final class of mutation (allele) is known as **neomorphic**; it changes the activity of the gene product, producing a new (neo-) function. There are a number of ways a new function can happen, for example the mutation can change the specificity of an enzyme, something that can happen in the course of cancer development.[379] To illustrate one such neomorphic mutation, consider the myogenic transcription factor MyoD, a protein that regulates the differentiation of skeletal muscle cells. There are mutations (alleles) associated with an aggressive form of embryonal rhabdomyosarcoma, a cancer of skeletal muscle. One mutant allele changes the DNA sequence so that the leucine found at position 122 of the wild type MyoD protein is replaced by an arginine.[380] Such a mutation is known as a **missense** mutation. So what is the effect of this change in the MyoD protein? To understand, you need to remember that MyoD is a transcription factor, a protein that recognizes specific sequences in DNA and leads to a change in gene expression. The wild type MyoD protein recognizes a consensus sequence (top panel →); in contrast the mutant allele encodes a protein whose DNA sequence specificity is altered (bottom panel →); it now binds better to a sequence that is also recognized by the transcription factor Myc. Myc regulates genes associated with active cell division. The result is that a gene product that normally inhibits cell division and encourages cell differentiation into non-dividing muscle cells (MyoD), acquires a new function, the ability to bind to different binding sites and induce cell division – a key feature of cancer cells. The mutation is neomorphic because the mutated MyoD protein (known as MyoDAla$_{122}$→Arg) has a new function, and probably much weaker if any binding to its original target sequence.[381]
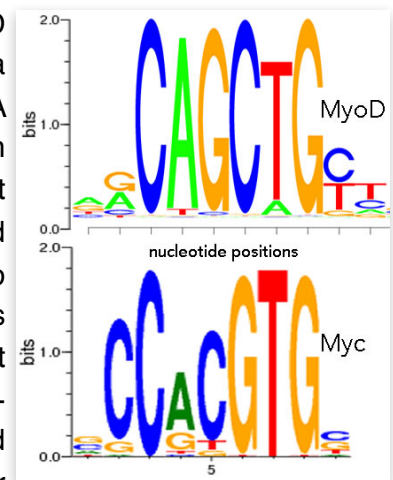
It is worth noting explicitly, that the relationship between the type of mutation (in Muller's terminology) and recessivity or dominance is not simple. An amorphic allele could be dominant, a

---

[379] Neomorphic mutations create therapeutic challenges in cancer

[380] from http://crunch.unibas.ch/ENCODE_REPORTS/Myers_HudsonAlpha/BG_5_8/report_BCLAF/JASPAR.Myf.wm.html and Deep Sequencing of MYC DNA-Binding Sites in Burkitt Lymphoma

[381] we will return to this topic toward the end of book: see Neomorphic mutations create therapeutic challenges in cancer

behavior known as halpoinsufficiency, arising because one copy of the gene does not produce the necessary amount of the gene product, or it can be recessive, if one functional copy of the gene is sufficient.

Before we move on, let us consider (again) the effects of mutations in a coding region of a gene. We have already mentioned missense mutations, mutations that lead to the replacement of one amino acid by another, different amino acid. There are mutations that do not change the amino acid sequence of the encoded polypeptide, but do change the DNA sequence – these are known as **synonymous** mutations, and as will see produce what is known as single nucleotide polymorphisms (SNPs), a feature in the DNA that can be detected by various molecular methods and is often used in the analysis of genomic similarities and differences (including human ancestry). There are two other types of generic names for alleles. A **non-sense mutation** is one that leads to a stop codon replacing a sequence encoding an amino acid in a polypeptide. Non-sense mutations lead to the premature truncation of the encoded polypeptide; their effects on gene function depend upon where they occur within the gene. In eukaryotic genes, which often have many exons and introns, there can be mutations that disrupt the sequences involved in recognizing and removing introns following transcription. These are generally referred to as splice-site mutations, since the process of RNA processing to generate an mRNA involves splicing out (removing) of the introns before the RNA is transported from the nucleus to the cytoplasm. Depending upon their effects on the final polypeptide, both non-sense mutations and mutations that alter an intron-exon junction can result in what is known as a **loss of function** (LoF) mutation, or one of Muller's morphs (although which type of morph is formed is dependent upon the mutation). Similarly, they can produce recessive or dominant alleles. Finally, it is worth remembering that essentially all traits are dependent upon a number of gene products, and so are polygenic, whereas a particular gene product may have a functional role in a number of processes; its mutational alteration can influence some or all of these processes, in which case it is considered pleiotrophic.[382] Don't get confused, all biological processes are complex, it is just that some alleles in some genes generate easily recognizable (distinctive) phenotypes.

Questions to answer

183. Draw out the relationship between gene - RNA - polypeptide, and describe the effects of missense, non-sense, and intron-exon junction mutations on gene expression.
184. How does the position within a gene of any of the mutations mentioned above influence their effects on the function of the gene's product?
185. Why is the MyoD mutation (mentioned above) neomorphic? What would you call it, if the mutated MyoD protein blocked the binding of wild type MyoD to its target DNA sequences?
186. Describe how a DNA change (missense, non-sense, junction mutation) produce Muller's morphs.
187. Describe how a neomorphic mutation alters the behavior of transcription factor and an enzyme.

Questions to ponder

- A *Drosophila* polytene chromosome can have over 1000 DNA molecules (strands). How, do you imagine, does the banding pattern observed in Drosophila polytene chromosomes relate to the genes on the chromosome?
- How does the polyploid nature of these chromosomes make visualizing chromosomal duplications and deletions possible? What are its limits, do you think?

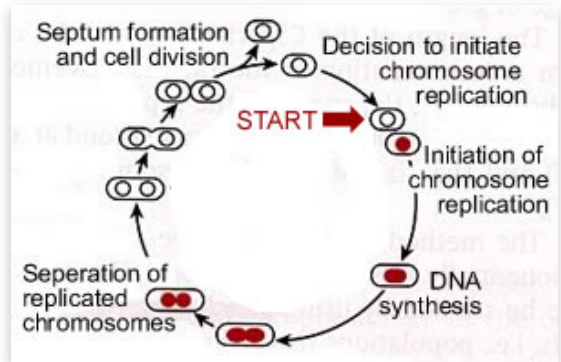[382] Pleiotropy: One Gene Can Affect Multiple Traits

Chapter 12: Reproduction in prokaryotes and horizontal gene transfer

*In which we consider how prokaryotic cells replicate asexually,
and how they can (under specific conditions) pass genetic
information to one another and acquire up such information
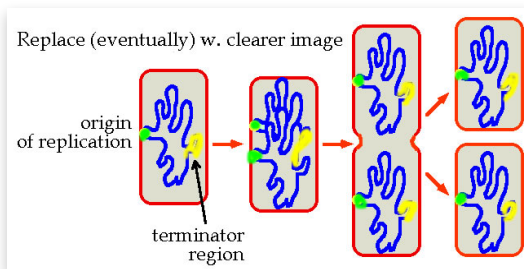from their environment.*

Asexual reproduction in bacteria and archaea

The simplest type of biological (cellular) reproduction is probably the asexual process found in with prokaryotes. In both bacteria and archaea, the genome typically consists of a single large circular DNA molecule, known as the bacterial chromosome. In some cases, the cell also contains smaller circular DNA molecules, known as plasmids. For the moment we will ignore plasmids and focus on the chromosome. The chromosome contains two importance sequence elements, the origin of replication (ORI) and the terminator (TER). When conditions are appropriate, a cell will pass through a decision point, a molecular switch, known as start (→). This switch activates the proteins that bind to the ORI region of the chromosome, and initiates the assembly of the DNA replication complex. A replication bubble forms, and the replication forks begin to move around the DNA molecule, making a copy. As the ORI sequence is replicated, the two ORI sites remain associated with the plasma membrane. The replication forks move around the DNA molecule, and collide in the TER region (←). As the DNA replication forks collide, they generate a signal that indicates that DNA replication is complete. During this period the cell itself is also growing, adding mass and volume. The division of one cell into two is mediated by the formation of a septum, an extension of the plasma membrane and the cell wall. Septum growth initiates between the two membrane-bound ORI sequences, which insures that each daughter cell receives one complete chromosome, one total genome.

If we consider the chromosome itself, it is worth noting that the order of genes around the circular molecule is conserved between organisms of the same species. The genes along the chromosome constitute a **syntenic** linkage group, the same genes in the same order along a chromosome (discussed further below). In the standard asexual mode of replication, all of the alleles are inherited together, the result is that a mutation in any particular gene (generating a new allele) acts in concert with the other alleles (in other genes) present. Over time, each organism produces a clone, and various clones interact with the environment and each other independently. These clones can display different levels of reproductive success, some clones can take over the population, while others can become extinct. In the case of studies on the evolution of bacterial antibiotic resistance (see below, page 259), each clone has to develop antibiotic resistance independently of every other clone; a similar

situation was observed in long term bacterial evolution studies.[383] There is no cross talk between lineages in such situations. Of course, if DNA is passed from clone to clone, as occurs within Griffith's transformation experiments (see above, page 157), things can get more complex. The movement of genes between lineages is known as horizontal gene transfer. Here we will review three versions of horizontal gene transfer found in prokaryotes.

Conjugation: what counts as sex in prokaryotes

The process of conjugation in bacteria allows DNA to move from one cell to another, with the moving (donor) DNA replacing the host DNA through the process of homologous recombination, a mechanism that we will consider, only briefly and at seriously over-simplified molecular detail (→). Homologous recombination is used in many systems, and is based on the recognition of a DNA sequence by a similar sequence.

over-simplified homologous recombination

mutation removed

Conjugation is a major pathway for horizontal gene transfer in bacteria.[384] In contrast to transformation (see below), conjugation "forces" DNA into what may be a reluctant recipient cell. In the process of conjugation, we can distinguish between two types of bacterial cells (of the same species). One contains a plasmid known as the sex factor (F), the other does not, it is referred to as a F⁻ cell. The F plasmid can exist independently of the host chromosome or it can be integrated into it; cells in which the F-plasmid is integrated into the host chromosome are known as a Hfr (high frequency recombination) cells (↓). The F plasmid contains the
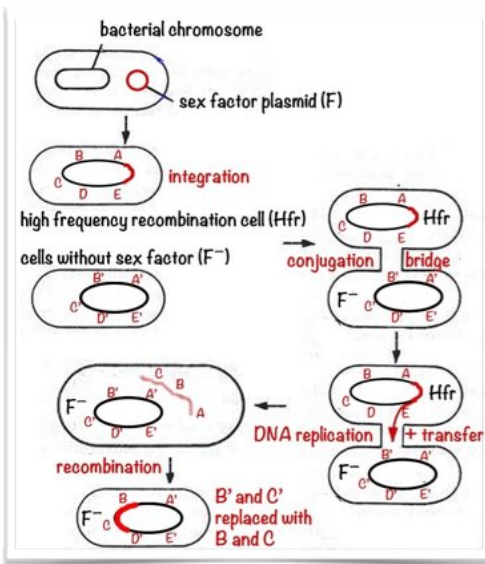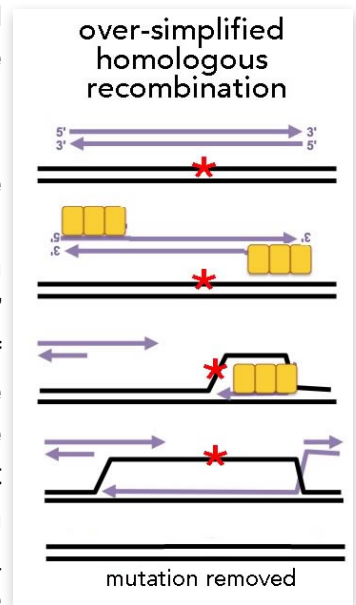
genes needed to transfer a copy of its DNA into a cell that lacks an F-plasmid. In this manner, an F-plasmid can colonize a population. In Hfr cells, the chromosome integrated F-plasmid can transfer host and plasmid genes into a F⁻ cell. To help make things a little simpler, we will refer to the Hfr cell as the DNA donor and F⁻ cells as DNA recipients.

To initiate conjugation, the Hfr/F⁺ cell makes a physical (conjugation) bridge to the F⁻ cell (←). A break in the donor DNA initiates a process by which single stranded DNA is synthesized and moved into the recipient F⁻ cell. The amount of DNA transported is determined largely by how long the bridge between the cells remains intact. It takes ~100 minutes to transfer the entire donor chromosome from an Hfr to an F⁻ cell. Once inside the F⁻ cell, the donor DNA is integrated into the recipient's chromosome, replacing the recipient's versions of the genes transferred, through homologous recombination - which we will return to later. Using Hfr strains carrying different alleles of various genes, and by controlling the duration of conjugation by breaking the conjugation bridge by placing the cells in

---

[383] see A cinematic approach to drug resistance and E. coli Long-term Experimental Evolution Project

[384] review of prokaryotic conjugation (prokaryotes)

a kitchen blender, the experimenters were able to determined the order of genes along the chromosome. The result was the discovery that related organisms had the same genes arranged in the same order, their genes where in syntenic groups (see above).[385] The typical drawing of the circular bacterial chromosome is like a clock going from 0 to 100 (→), with the genes placed in their respective positions, based on the time it takes to transfer them (in minutes).

If the entire F-plasmid sequence is transferred, the original F⁻ cell becomes an Hfr cell. If the Hfr cell loses the F-plasmid sequence, it reverts to a F⁻ state. The end result of the conjugation process is similar to that obtained in sexual reproduction in eukaryotes, namely the original F⁻ cell now has a genome derived in part from itself and from the "donor" Hfr cell. Because the outcome of an Hfr/F⁻ cell interaction can lead to a cell with a different set of alleles than either of the "parental" cells, this process is often referred to as bacterial (prokaryotic) sex.
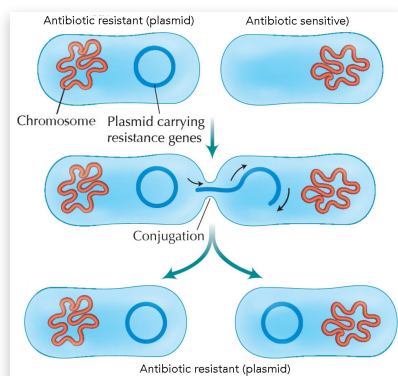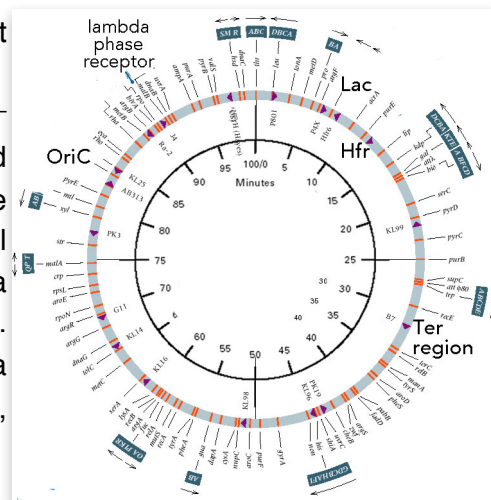
Versions of this process are involved in the transfer of plasmids from cell to cell within a community (←).[386] A plasmid contains its own "origin of replication"; some (low copy number) plasmids exist in one to two copies per cell, while others (high copy number plasmids) are present in multiple copies, which is which is determined in large part by their origin of replication sequences. Some plasmids can exist in as many as 700 copies per cell.[387] Plasmids can encode genes responsible for antibiotic resistance and the rapid dispersion of antibiotic resistance phenotype is a cause of increasing concern.[388] Many plasmids, also known as mobile genetic elements, are more selfish, that is, their presence in a cell might not directly benefit that cell. Once a plasmid has a Hfr-like element, it can move through (and parasitize) a population. We (that is you) might even be able to generate a plausible mechanism by which viruses could have evolved from such "selfish" plasmids.

Questions to answer & ponder:
188. What factors act to insure that each (prokaryotic) cell generated contains a complete genome?
189. How would mutating the origin or terminator regions of a prokaryotic chromosome influence the cell's reproduction?
190. Describe what you would expect to happen, and why, if a prokaryotic cell received an incomplete genome.
191. Describe (diagram) what happens to the DNA molecule that is introduced to a cell via conjugation?
- How might the regulation of plasmid ORI regions be different in low and high copy number plasmids?

[385] Synteny: http://en.wikipedia.org/wiki/Synteny

[386] Plasmids Spread Very Fast in Heterogeneous Bacterial Communities: https://www.ncbi.nlm.nih.gov/pubmed/12524329

[387] Plasmids 101: Origin of Replication

[388] Addgene: Mechanisms of Antibiotic Resistance

Other naturally occurring horizontal gene transfer mechanisms

Many horizontal transfer mechanisms are regulated by social and/or ecological interactions between organisms.[389] It is important to note that the mechanisms involved can be complex, one could easily imagine an entire course focused on this topic alone. We introduce only the broad features of these systems. Also, we want to be clear about the various mechanisms of DNA uptake. First recognize that when an organism dies its DNA can be eaten as a source of energy, as well as carbon, nitrogen, and phosphorus. When eaten, any information in the DNA, the result of mutation and selection, is lost.[390] Alternatively, the nucleotide sequence of a DNA molecule can be integrated into another organism's genome, resulting in the acquisition of whatever information developed (evolved) within that lineage. This is information that might be useful, harmful, or irrelevant to the organism that acquires it – imagine how inserting a piece of DNA into a genome could be harmful. The study of these natural DNA import systems has identified specific molecular machines that mediate DNA transfer. Some organisms use a system that preferentially imports DNA molecules that are derived from organisms of the same or closely related types as themselves. You can probably even imagine how they do this – they must have receptor systems that can recognize species-specific "DNA uptake sequences." The various mechanisms of horizontal gene transfer, unsuspected until relatively recently, have had profound influences on evolutionary processes, particularly among microbial communities, where they are more common. It turns out that a population of organisms does not have to "invent" all of its own genes, it can adopt genes generated by evolutionary mechanisms in other organisms in other environments for other purposes. So the question is, what advantages might such information uptake systems convey, and (on the darker side), what dangers do they make possible?

Transformation

There are well established methods used in genetic engineering to enhance the ability of bacteria to take up plasmids from their environment.[391] We, however, focus on natural transformation, the process associated with the transfer of DNA molecules from the environment into a cell. Transformation is an active process that involves a number of components, encoded by genes that can be expressed or not depending upon environmental conditions. Consider a type of bacteria that can import DNA from its environment. If the density of bacteria is low, there will be little DNA to import, and it may not be worth the expense to express the genes and synthesize the proteins involved in the DNA uptake and integration machinery. In fact, bacteria can sense the density of organisms in their environment using quorum sensing (see above, pp. 91-95). Bacteria use quorum sensing to generate the DNA uptake system, apparently by activating specific transcription factors. When present in a crowded environment, the quorum sensing system turns on the expression of the DNA uptake system.

---

[389] DNA uptake during bacterial transformation: http://www.ncbi.nlm.nih.gov/pubmed/15083159

[390] This is of course why genes are rarely if ever transferred from food to the organism doing the eating.

[391] Making Calcium Competent (bacterial) Cells: http://mcb.berkeley.edu/labs/krantz/protocols/calcium_comp_cells.pdf

Here we outline the process in one type of bacteria but functionally similar mechanisms are used in other bacterial and archaeal species. Double-stranded DNA binds to the cell's surface through a variety of DNA receptors. In some cases these receptors bind specific DNA sequences, in others they bind DNA generically, that is any DNA sequence. As shown, Gram negative bacteria have two lipid membranes, an outer one and an inner (plasma) membrane, with a periplasmic space in between (→). In an ATP-hydrolysis coupled reaction, DNA bound to the exterior surface of the bacterium is moved, through a protein pore, through the outer membrane and into the periplasmic space, where it is passed to the DNA channel protein. Here one strand is degraded by a nuclease while the other moves intact through the channel into the cytoplasm of the cell in a 5' to 3' direction. Once inside the cell, the DNA associates with specific single-stranded DNA binding proteins and, by homologous recombination, it is inserted into the host genome.[392] While the molecular details of this and functionally similar processes are best addressed elsewhere, what is key is that transformation enables a cell to decide whether or not to take up foreign DNA and whether to add such DNA sequences to its own genome.
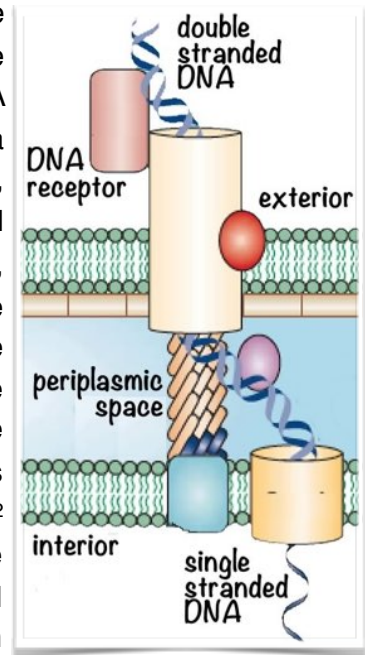
Viruses moving genes: transduction

The final form of horizontal gene transfer that we will consider involves viruses. The structure and behavior of viruses is a complex topic, the details of which are largely beyond us here, but it is not unreasonable to consider viruses as nucleic acid transport machines. Viruses are completely dependent for their replication on the infected host cell, they have no active metabolic processes and so are not alive in any meaningful sense, although they can certainly be infectious, that is they can spread through a population. Viruses cannot be killed, because they are not alive, but they can be inactivated by various treatments.

The simplest viruses contain a nucleic acid genome and a protein-based transport and delivery system. We briefly consider a typical bacterial virus, known as a bacteriophage or bacteria eater. The bacterial virus we consider here, the T4 bacteriophage, looks complex and it is, other viruses are simpler (→). The T4 phage (short for bacteriophage) has a ~169,000 base pair double-stranded DNA genome that encodes 289 polypeptides, almost as many as a minimal cell (see above).[393] The assembled virus has an icosahedral protein head that contains a DNA molecule attached to a tail assembly that recognizes and binds to target cells. Once a suitable host is
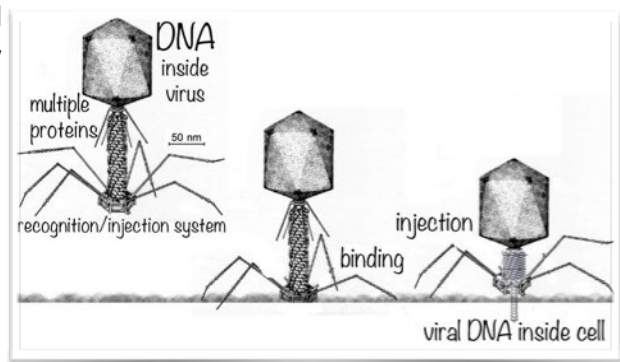
[392] Bacterial transformation: distribution, shared mechanisms and divergent control & Natural competence and the evolution of DNA uptake specificity

[393] http://en.wikipedia.org/wiki/Bacteriophage_T4

found, based on tail binding to cell surface molecules, the tail domain attaches and contracts, like a syringe, punching a hole through the cell's external wall and plasma membrane. The DNA emerges from the bacteriophage and enters the cytoplasm, infecting the cell. Genes within the phage genome are expressed, leading to the replication of the phage DNA molecule and the fragmentation of the host cell's genome.[394] The phage DNA encodes the proteins that are used to assembled new phage heads. DNA is packed into these heads by a protein-based DNA pump, a pump driven by coupling to an ATP hydrolysis reaction complex (→).[395] In the course of packaging virus DNA, the system will, occasionally, make a mistake and package a fragment of the host cell's DNA. When such a phage particle infects another cell, it injects that cell with a DNA fragment derived from the previous host. Of course, this mis-packaged DNA may not contain all of the genes the virus needs to make a new virus or to kill the host. The transferred DNA can be inserted into the newly infected host cell genome, with the end result being similar to that discussed previously for transformation and conjugation. DNA from one organism is delivered to another, horizontally rather than vertically.
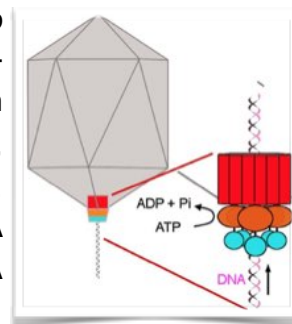
Because the horizontal movement of DNA is so common in the microbial world, a number of defense mechanisms have evolved.[396] These include the restriction/DNA modification systems used widely for genetic engineering, and the CRISPR-CAS9 system, which enables cells to recognize and destroy foreign (viral) DNA. These systems, evolved as part of prokaryotic immune systems, form the basis of modern genetic engineering methods. They illustrate how studying apparently arcane aspects of the biological world can have dramatic impacts on modern technological, medical, and economic systems.

Questions to answer:
192. What is an asexual clone? How would you recognize it.
193. What is the effect of a amorphic allele / mutation on the behavior of a prokaryotic clone.
194. What are some possible (evolutionary) advantages to the ability to take up and integrate, as opposed to simply eat foreign DNA?
195. Why might the "source" of foreign DNA matter?
196. Present a plausible model that would identify host from foreign DNA.
197. What factors are necessary for homologous recombination?
198. Propose the steps that would be involved in the evolution of a "selfish" plasmid into a virus.

Questions to ponder:
- How might a prokaryotic organism protect itself from invading viruses?
- How might the importation of DNA through transformation be harmful to the host?

[394] An infected bacterial cell can protect is neighbors, often its clonal relatives, if it can kill itself before the virus can replicate. This is an example of a simple altruistic behavior.

[395] The Structure of the Phage T4 DNA Packaging Motor Suggests a Mechanism Dependent on Electrostatic Forces

[396] see The phage-host arms-race: Shaping the evolution of microbes

## Chapter 13: Asexual and sexual reproduction in eukaryotes

In which we consider the processes of asexual and sexual reproduction in eukaryotes. We note the molecular processes, mitosis & cytokinesis, involved in somatic cell reproduction and how they are modified in meiosis and gamete formation within the germ line. We consider the implications of chromosome pairing, recombination & independent segregation as well as dimorphism of gametes leading to maternal and paternal effects, including mitochondria inheritance and sex determination.
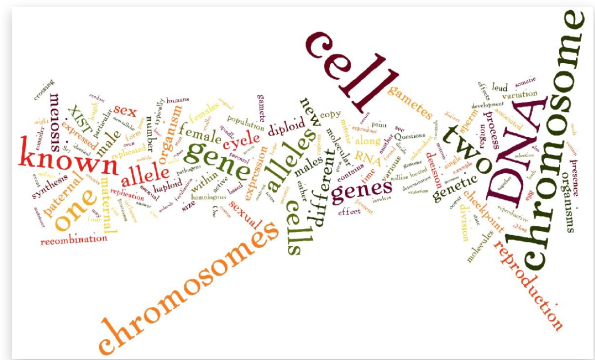
Asexual reproduction in a eukaryote: making a (somatic) clone

Asexual reproduction in a eukaryote is similar to that in a prokaryote, the cell grows and at some point there is a molecular decision to divide. At that point the genome of the cell is replicated and the cell is then divided into two, each receiving one complete copy of the genome. In addition, all eukaryotes have cytoplasmic organelles (mitochondria, and in algae and plants, chloroplasts) with their own, albeit reduced in size, genomes. In the course of asexual, what is termed somatic, reproduction, each of the sibling cells also receive a number of mitochondria (and in plants, chloroplasts).[397] In the eukaryotes that we will concern ourselves with, most of the cells of the organism are diploid (rest assured, we will let you know when they are not).

Somatic (asexual) reproduction involves what is known as a cell cycle. We think of the cell cycle as beginning with the process of cell division (D in the figure below)($\downarrow$). The process of dividing one cell into two, known as cytokinesis, results in two sibling cells, each with identical genomes. Cytokinesis involves cytoskeletal and cytomuscular systems that are generally discussed in detail in a later cell biology course (not here!) Generally, cell division is symmetrical, so that the two sibling cells are half the volume of the parental cell and very similar. Division is followed by a period of cell growth (known as G_1)($\rightarrow$), during which energy and materials are imported from the external environment, or previously stored within the parental cell, are converted into lipids, nucleic acids, proteins, and other molecules leading to an increase in cell volume, the growth of the cell. As the cell grows, there is a decision to be made, will it continue to grow (and perhaps divide) or will it stop growing and enter a steady state where it maintains itself (building and disassembling molecules, repairing DNA, etc) – a state known as G_o ($\uparrow$). The majority of cells in any particular tissue are in the G_o state; in G_o there is no new DNA synthesis, so the possibility of mutation is lower than when DNA is being

---
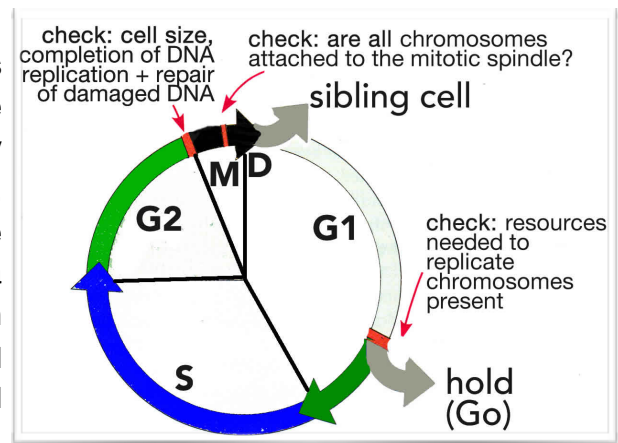
[397] Plants and algae, which we will not be discussing in any detail, contain a second type of intracellular, DNA-containing organelle, known as chloroplasts. Their inheritance is similar to that of mitochondria.

replicated. If, however, various external and internal signals act on and within the cell, the cell can reverse the G_o decision and resume growth and eventually divide (note that it is difficult to talk about these systems without personalizing them, even though they are certainly not conscious).

Once the decision to proceed has been made, that is, the molecular switch has been flipped, the cell will encounter what is known as a checkpoint, discussed in more detail below.[398] A checkpoint is a molecular feedback system by which the cell essentially calculates various aspects of its internal state and makes a decision to wait or proceed with DNA synthesis. The decision to start DNA synthesis is based in part on whether the cell has, or will have, sufficient resources to replicate its DNA molecules, which requires (in a human cell) ~12 billion nucleotide addition reactions (both strands of a total of ~6 billion base pairs). The DNA synthesis decision point is known as "start"; once that decision is made the cell will continue to grow and proceed into the part of the cell cycle during which DNA synthesis occurs, known as S (↑). At the end of this phase of the cell cycle, DNA synthesis will be complete and the cell will continue to grow; the cell has entered into what is known as the G_2 phase of the cell cycle. The start decision is particularly critical for the cell (and the organism), since failure to complete DNA replication will likely lead to changes in gene number and increased mutagenesis and inaccurate repair of single stranded DNA molecules.[399] There are mutant alleles, originally described in yeast and known as "wee" mutations, in which the molecular switch controlling entry into S is damaged; these mutations lead to a disconnect between growth and division and result in smaller and smaller cells and eventually cell death.[400]

During the asexual reproduction cycle the ploidy (the number of copies of a each chromosome) is conserved. A haploid cell gives rise to a haploid cell, while a diploid cell gives rise to a diploid cell. The one detail that is altered during S-phase of the cell cycle is that there are now two copies of each chromosome. While a cell is diploid during G_1, it is effectively tetraploid during G_2. This can have physiological effects because two copies of a gene can, in theory and generally in practice, support the synthesis of more RNA molecules per unit time than one copy of a gene. Based on this logic, we would expect to see changes in the rates of gene expression in G2 compared to G1 cells.

In contrast to circular prokaryotic genomes, which typically have a single origin of replication (the site where DNA synthesis begins), the much larger size of eukaryotic genomes and the presence of multiple linear chromosomes requires multiple sites per chromosome at which DNA synthesis starts. These replication origins are regulated during S phase such that each is activated once and only once, so that each region of the genomic DNA is replicated once and only once. Before cell division (cytokinesis), a checkpoint monitors the presence of unreplicated DNA and delays the cell cycle until that DNA has been replicated. The process of DNA replication can lead to mutations, so this checkpoint also monitors the completion of various DNA repair processes. The presence of such a DNA repair checkpoint explains the observation that damaging DNA, for example by radiation, or inhibiting DNA synthesis enzymes using drugs, leads to delays in the cell cycle. Pathogens, such as the bacteria

---

[398] The quorum sensing systems we discussed previously is a version of a checkpoint system.
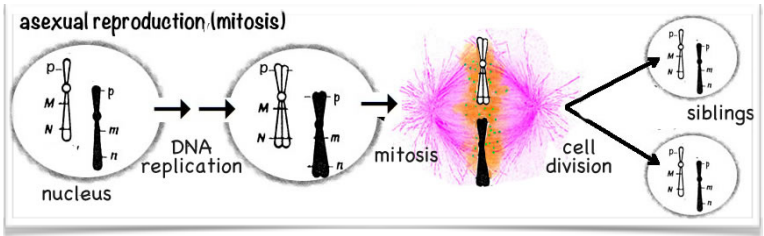
[399] DNA replication is complex process, with details considered in subsequent molecular biology courses: this review addresses some of these complexities Forks on the Run: Can the Stalling of DNA Replication Promote Epigenetic Changes?

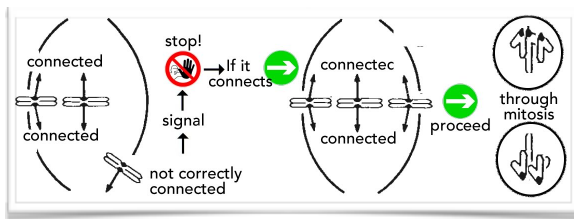[400] Paul Nurse and Pierre Thuriaux on wee Mutants and Cell Cycle Control: https://www.ncbi.nlm.nih.gov/pubmed/27927897

Listeria, exploit this DNA damage checkpoint to enhance their own replication.[401]

Molecular choices and checkpoints

Once the DNA replication/repair checkpoint has been passed, the cell can divide. The first step of this process is known as mitosis (→); it involves a molecular machine based on protein polymers (αβ-tubulin-based microtubules) that are organized into a "mitotic spindle". There is a molecular checkpoint switch that monitors the assembly of the mitotic spindle, and a second checkpoint that monitors that each replicated chromosome has connected correctly to the spindle (←). Each of the replicated chromosomes interacts independently with the mitotic spindle (this is different from their behavior during meiosis, as we will see below). Together these two checkpoints serve to insure that each of the two future sibling cells gets one and only one copy of each and every chromosome present in the parental cell.[402] The presence of the second chromosome attachment mitotic checkpoint was recognized in experiments in which chromosomes were manipulated so that they could not connect correctly to the mitotic spindle, such a manipulation caused a delay or halt in mitosis.[403]

Once the chromosomes are segregated to the opposite ends of the parental cell, the parental cell divides, using another protein polymer-based (actin/myosin-based microfilaments) molecular machine, known as the contractile ring, to produce two sibling cells. It is worth noting that while these two cells are genotypically identical, as they inherit the same set of alleles as were present in the parental cell, they may behave differently due to differences in their environment and differences in internal components - factors that we will return to (rather briefly) when we consider developmental processes.

The cell cycle decision check points are composed of multicomponent interaction networks. While we consider check point mechanisms only briefly here, they play a number of important roles in development and disease. The typical check point is built around a protein kinase, an enzyme that can phosphorylate various targets – such phosphorylation (a post-translational modification) can lead to changes in protein-protein interactions and activities. Checkpoints involve a particular class of kinase, known as cyclin-dependent kinases (CDKs)(→). The activity of these CDKs is regulated positively by the binding of a small regulatory protein, known as a cyclin, as

---
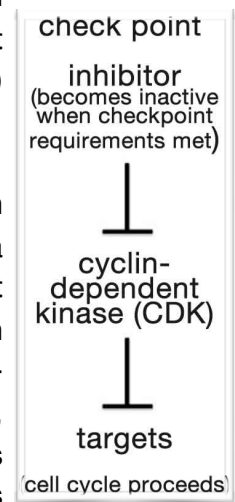
401 Listeria monocytogenes induces host DNA damage and delays the host cell cycle to promote infection

402 Kinetochores, microtubules, and spindle assembly checkpoint signaling

403 Mitotic forces control a cell-cycle checkpoint

well as other interacting proteins and a number of post-translational modifications. Cyclin's themselves are the target of various forms of regulation, including proteolytic degradation, triggered by their post-translational modification. Typically the activity of the cyclin-CDK complex is inhibited by various factors (proteins). When the conditions involved in the checkpoint are met, this inhibitor is itself inactivated, allowing the cyclin-CDK complex to become active; the active kinase then phosphorylates and regulates the activity (and stability) of its targets, allowing the cell to pass through the check point and proceed along the cell cycle. One effect of activating the CDK is the rapid degradation (removal) of the cyclin, this makes the switch irreversible until such time as cyclin levels increase again, during the next cell cycle.

Questions to answer:
199. How do chromosomes interact with one another during mitosis/cytokinesis?
200. How do checkpoints work and what makes them irreversible?
201. What does it mean that a checkpoint acts to "make a decision based on evidence"?
202. Make a graph of CDK activity and the concentration of the cyclin regulating it, as a function of the cell cycle.
203. What can go wrong if a checkpoint is ignored (start with a cell cycle diagram)?
204. How can a mutation in a checkpoint influence cell behavior during the somatic (mitotic) cell cycle?
205. How does gene expression change over the course of the somatic cell cycle?

Questions to ponder:
- Why is the decision to start a new cell cycle critical?
- When is the decision to start a new cycle made?

Sex-determination and its chromosomal basis

In eukaryotes, the generation of a new organism, distinct from previous organisms, involves the process of sexual reproduction. Different types of organisms determine an individual's sex using different mechanisms, and in some cases, a single individual, known as a hermaphrodite, can display traits of both sexes at either the same time or sequentially.[404] There are basically two general mechanisms that determine the sex of an organism: genetic and environmental. In environmental sex determination various external signals influence the sex of the organism. For example in a number of reptiles (and other organisms), the sex of the adult is determined by temperature during key developmental periods, with different temperatures associated with male and female outcomes.[405] Recently, climate change has been implicated in sea turtle sex ratios.[406] In other organisms, all organisms originally develop into one sex or the other and, as they mature and

---
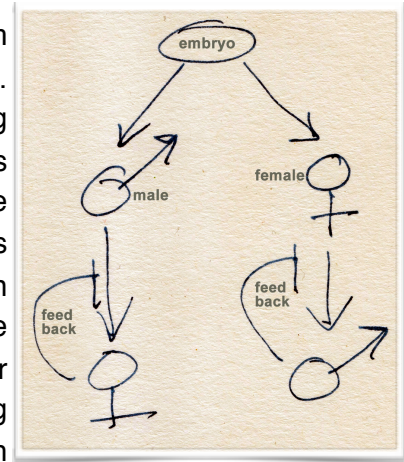
[404] We will not go into any great detail about hermaphroditic models of reproduction, but this is an interesting paper related to the subject: Sexual selection: lessons from hermaphrodite mating systems: https://academic.oup.com/icb/article/46/4/349/634174/Sexual-selection-lessons-from-hermaphrodite-mating

[405] Environmental sex determination mechanisms in reptiles

[406] Climate change is turning 99 percent of these baby sea turtles female

(often grow larger), transform into the other sex.[407] In some cases the presence of a mature animal of one sex can inhibit the sex change in smaller individuals (↑). As an example, the largest clownfish in a group is typically female; if that female is removed, one of the smaller males will develop into a female (think about the impact on Nemo). In other species, the situation is reversed, the largest animal is a male, and if this male is removed, one of the smaller females develops into a male.[408]

In humans, and most mammals, birds, and reptiles the phenotypic sex of an individual is determined chromosomally, that is, by which sex chromosomes their cells contain. The other, non-sex determining chromosomes are known as autosomes.[409] In humans the sex (23rd) chromosome comes in two forms, known as X and Y (→).[410] An XX individual typically develops as a female, while an XY individual typically develops as a male. Most of the X and Y chromosomes are non-syntenic, as you might have suspected given that the Y chromosome has only ~50 genes, while the X-chromosome has between 800 and 900 genes. The X and Y chromosomes are syntenic in what are known as their psuedo-autosomal regions. As we will see below, the organization of these chromosomes will have effects on how they behave during the course of meiosis (sexual reproduction).
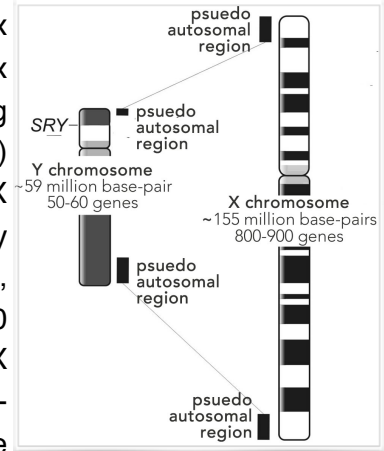
One key difference between X and Y chromosomes in therian mammals (marsupials and placental mammals, which includes humans), is the presence of the SRY gene on the Y (there is no copy of SRY on the X chromosome). The SRY gene is not found in monotremes (egg-laying mammals) and other vertebrates.[411] The SRY gene appears to have originated in the therian mammal lineage ~150 million years ago, derived by duplication of a Sox-type DNA binding protein/transcription factor, which contains a high-mobility group or HMG box, DNA binding domain. The presence of a Y chromosome, and so an active Sry gene, leads to male sexual development, whereas loss of function mutations in Sry lead to female development, even if the Y chromosome is present (→).[412]

---
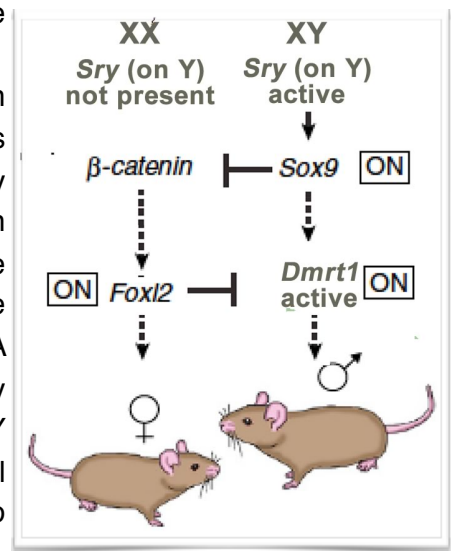
[407] Phylogenetic Perspectives on the Evolution of Functional Hermaphroditism

[408] Functional hermaphroditism in teleosts

[409] In other species (e.g. birds, some reptiles, and some insects) the system is based on Z and W sex chromosomes. In contrast to the XY system, males are ZZ while females are ZW.

[410] X chromosome regulation: diverse patterns in development, tissues and disease: https://www.ncbi.nlm.nih.gov/pubmed/24733023 and Y-chromosome: https://ghr.nlm.nih.gov/chromosome/Y

[411] Vertebrates use a number of mechanisms to determine sex, these include "Environmental sex determination is widely employed in fish, where a range of stimuli from social cues to temperature establishes sex. Temperature sex determination is also extensively utilized in reptiles. see Sex determination in mammals--before and after the evolution of SRY

[412] see Molecular Mechanisms of Male Sex Determination: The Enigma of SRY for more details.

Sry encodes a transcription factor that initiates a down-stream cascade, activating some genes and inhibiting others, with the end result being the generation of the various developmental difference associated with male and female anatomy and behavior.[413] In females other genes are expressed (actively transcribed) and they act to inhibit the male differentiation system, just as Sry and its "downstream" targets act to inhibit female differentiation. In molecular studies, it is possible to show the importance of Sry, since the Sry gene can be transferred to one of the other chromosomes, and its presence still leads to male determination. The details of these processes are complex, so we refer further details to more advanced classes.[414]

At this point we should mention that there are other sex determination strategies that you might come across in your subsequent studies, but which we will ignore here.[415] For example, in some organisms (plants and algae), the haploid (gametic) stage can persist and live independently,[416] but generally the haploid stage of a eukaryote, and particularly animal's life cycle is short.

So what are the benefits of sexual reproduction, a process that requires social collaboration.[417] As will noted in our discussion of meiosis (see below), the simple answer is the generation of genetic variation. So why is this variation important. One major reason arises from the presence of rapidly reproducing pathogens. Viruses, bacterial and microbial (eukaryotic) organisms typically reproduce over a period of minutes to hours to days, whereas larger organisms reproduce (generate new organisms) over a period of months, years, and decades. Susceptibility to infection by pathogens is itself a phenotype, one with a genetic component. The genetic variability within a population can serve as insurance against pathogens; even the most lethal pathogens known, viruses like smallpox and bacteria such as those that cause plague, do not kill all of the organisms they infect. And those organisms that survive infection are often immune to subsequent infections, a phenomena that is the basis of vaccination and various other processes, including the CRISPR CAS9 system of prokaryotes.

The level of genetic variation within a population is important as insurance against infectious disease. Similarly, but on somewhat longer time scales, the level of genetic variation within a population enables a population adapt to a changing environment. The larger the population size, the more likely there is some genotypic combination present that will make adaptation to a changing environment possible. The reduction in genetic variation is one of the reasons that reductions in population size have been linked to an increase in the probability of extinction.[418]

In addition to the generation of variation, the process of sexual reproduction offers mechanisms by which to isolate populations reproductively, that is, to create two species from one. Generally males and females have to cooperate to reproduce. They have to be producing functional gametes at the same time, these gametes have to be able to meet each other, recognize each other, and fuse together, the

---

[413] In a recent study, the primary sex determination event in humans has been found to be associated with changes in ~6500 genes: see 6,500 Genes That Are Expressed Differently in Men and Women

[414] Sex determination: a primer

[415] The evolutionary dynamics of haplodiploidy

[416] see wikipedia – gametophyte: https://en.wikipedia.org/wiki/Gametophyte

[417] Origins of Eukaryotic Sexual Reproduction: http://cshperspectives.cshlp.org/content/6/3/a016154.ful

[418] Timing and causes of mid-Holocene mammoth extinction

diploid cell that forms has to develop normally, which then has to be able to form functional gametes, which involves the pairing of homologous chromosomes, and so on and so forth. Incompatibilities in any of these processes can lead to a reproductive barrier between individuals within populations - that is, speciation.

Questions to answer:
206. If you were design a temperature sensitive form of sex determination, how would you go about it?
207. What might happen if you removed the regions of the Y chromosome that are homologous to the X?

Question to ponder:
- Any thoughts on why different vertebrates would have adopted such different modes of sex-determination, and their evolutionary benefits and drawbacks?

Meiosis, fertilization, and embryogenesis

In contrast to asexual reproduction, which produces clones that are identical (with the exception of newly arising mutations) to their progenitor, the result of sexual reproduction is a genetically distinct organism, different from either parent. There have been a number of explanations for why sexual reproduction is so common, essentially all visible (macroscopic) organisms reproduce (or can reproduce) sexually.[419] One view considers the fact that most parasites and pathogens are small, and reproduce quickly. Populations of such organisms exploit the generation of variations, through mutation, to evolve quickly. In contrast, larger macroscopic organisms typically reproduce much more slowly. So how can they keep up with their parasites and pathogens? Sexual reproduction offers, as we will see, a mechanism to generate huge amounts of genetic variation within a population; this view of the selective advantage of sex is often referred to as the Red Queen Hypothesis, since organisms have to "run"

"It takes all the running you can do, to keep in the same place."
says the Red Queen to Alice

constantly, in terms of generating genetic variation, to keep up with their parasites and pathogens.[420] In addition, there is the possibility to eliminate lethal alleles from a lineage, as opposed to having to have the lineage itself go extinct, and the fact that sexual reproduction can speed the appearance of beneficial combinations of alleles (→), combinations that would take significantly longer to appear if they had to occur independently in a particular lineage.

One aspect of the haploid state associated with sexual reproduction, is that it can reveal the presence, and lead to the elimination, of highly deleterious recessive alleles. Haploid cells that contain, and are dependent upon the expression of such alleles will be eliminated, removing the allele from the population, which can have a strong evolutionary effect.[421]

_____
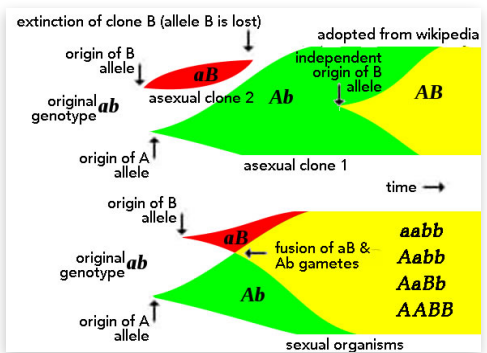
[419] C. Zimmer. 2009. On the Origin of Sexual Reproduction

[420] see Sexual reproduction as an adaptation to resist parasites

[421] see Evolution of haploid selection in predominantly diploid organisms: http://www.pnas.org/content/112/52/15952.full and Haploid selection in animals: http://www.sciencedirect.com/science/article/pii/S0169534704002381

Steps in meiosis

Sexual reproduction begins with two diploid cells, generally found in two distinct individuals. These two individuals are of different "mating types", which in macroscopic organisms are referred to as the two sexes, male and female. In species with mating types, the gametes produced appear identical, and both share an equal investment in reproduction. Where gametes are different in size (an example of sexual dimorphism), the two sexes can have discordant investments in reproduction, one can spend more energy generating gametes than the other. The sex with the greater investment in gamete production is known as female, one with less of an investment is referred to as male. This difference can become even more pronounced in terms of parental investment, a fact that underlies sexual selection, one of the key aspects of modern (Darwinian) evolutionary theory.[422]

The basic process of sexual reproduction can be summarized as follows: a diploid cell generates, through the process of meiosis, one or more haploid gametes. In females the process of meiosis typically generates a single gamete, known as an egg, and three mini-cells, known as polar bodies. In males, meiosis produces four gametes, known as sperm. Each gamete will contain one and only one copy of each autosomal chromosome present in the original diploid cell (→). Historically, chromosomes were numbered based on their apparent size in histologically stained specimens. In humans, the largest of these chromosomes, chromosome 1, contains ~250 million base pairs of DNA and over 2000 polypeptide-encoding genes, while the smallest, chromosome 22 contains ~52 million based pairs of DNA and around 500 polypeptide encoding genes.[423] Homologous chromosomes are also

schematic of human karyotype (from genomicus)
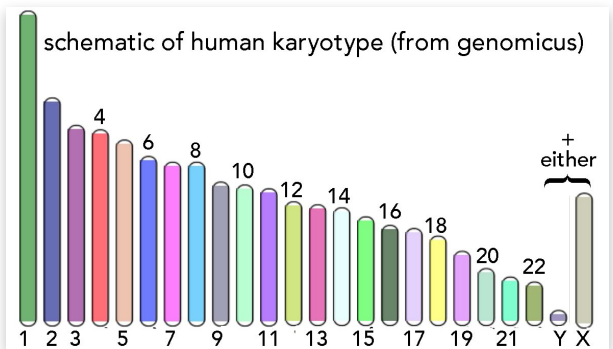
defined by the order of genes found along their length. Human chromosome #5 contains different genes than are found on chromosome #6. Moreover, the maternal version of each chromosome can contain different alleles of the genes present compared to those that are found in the paternal version. In (therian) mammals males have both an X and a Y chromosome; meiosis generates four gametes that contain one copy of each of the autosomes and either an X or a Y chromosome. Females have two X chromosomes, so all gametes they produce contain an X chromosome. A male gamete (a sperm) fuses with a female gamete (an egg) to form a new diploid cell, a new organism - if the male gamete contains a Y chromosome, the new (diploid) organism is chromosomally male, if the male gamete contains an X chromosome, the new organism is chromosomally female.[424] The fusion event, known as fertilization, is the most discontinuous event in the process of (sexually reproducing) life. Even so,

---

[422] A peacock's tail: how Darwin arrived at his theory of sexual selection and Mate choice and sexual selection: What have we learned since Darwin?

[423] We are only discussing polypeptide-encoding genes because it remains unclear whether (and which) other transcribed regions are genes, or physiologically significant.

[424] While we not deal in detail with this topic, aspects of gender are complex traits: see Beyond XX and XY: The Extraordinary Complexity of Sex Determination

fertilization does not represent a true discontinuity – both sperm and egg are alive, as is the fertilized egg.[425] In a critical sense life (in the post-LUCA world) never begins – it continues and is transformed.

The fused cell that results from fertilization is known as a zygote; through somatic (asexual) cell division (mitosis and cytokinesis) it will develop into a adult, composed of diploid cells. The cells of the adult that produce gametes are known as germ cells, and together are known as the organism's germ line; whereas the rest of the adult is composed of somatic cells, cells that divide (if they divide) by mitosis. Meiosis is restricted to germ line cells.

Recombination & independent segregation

We will begin our description of meiosis with a germ line cell, a cell that contains two copies of each autosome, and either two X chromosomes in a female and an X and Y chromosome in a male. The chromosome derived from the progenitor female gamete is known as the maternal copy of the chromosome, the one derived from the progenitor male gamete is known as the paternal copy of the chromosome. In order to generate gametes, a diploid germ cell enters meiosis (see video link). Meiosis (↓) consists of a single round of DNA replication followed by two cell division cycles.
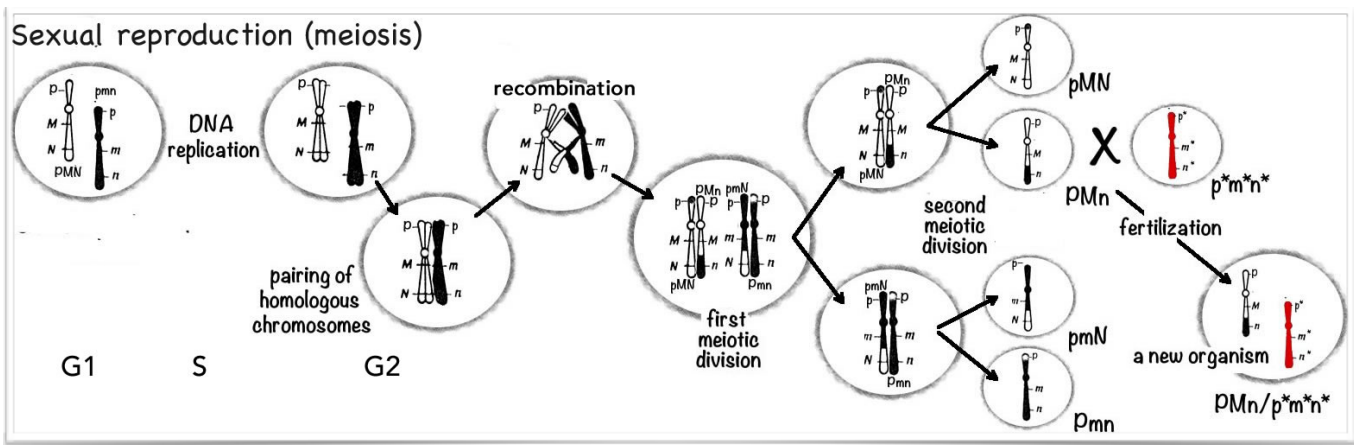
As a cell enters meiosis the first step is the replication of its DNA - the cell goes from a diploid G1 state into S, just as in mitosis. Each of its individual chromosomes (46 in humans, 2 copies each of the 23 homologous chromosomes) is duplicated (→). The two resulting replicated (double-stranded) DNA molecules remain attached to one another through a structure known as the centromere. Here is were meiosis diverges from mitosis. In asexual (mitotic) cell division each replicated chromosome remains independent of its homolog and interacts independently with the mitotic spindle (through its centromeric region). In meiosis, during G2 the (now) duplicated homologs (the maternal and paternal

---
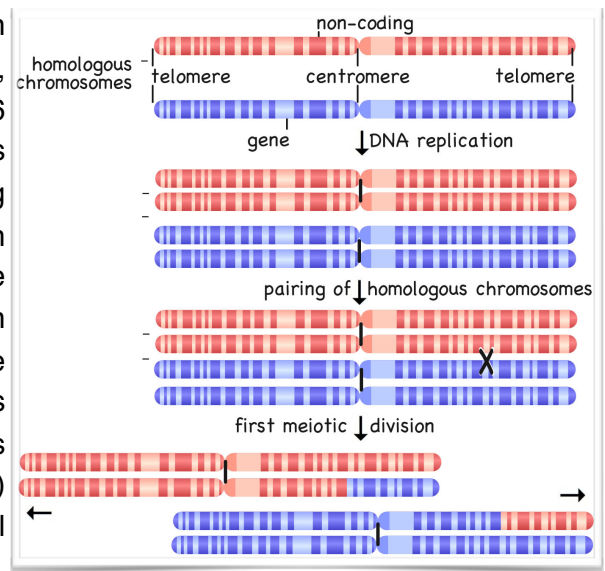
[425] In fact, there are examples of cell fusion within organisms - as an example, during the development of skeletal muscle, muscle precursor cells fused to generate large multi-nuclear cells, known as myotubes.

chromosomes) align with one another to form a structure containing four (double stranded) DNA molecules; these four DNA molecules are known historically as a "tetrad". Homologous chromosome pairing is based on the association of syntenic regions of the chromosome. The DNA sequences along the homologous chromosomes, while not identical, are extremely similar, with the same genes located in the same order on each (if they are not - due to chromosomal rearrangements, things can get messy). After chromosome pairing, and at essentially random positions along the length of the

chromosome, there occurs what are known as cross-over or recombination events (←). An enzyme (a DNA endonuclease) produces double-strand breaks in two of the four (double-stranded) DNA molecules (at the sites marked by "X" above or by "cross over" to the left).[426] The DNA molecules are then rejoined, either back to themselves (maternal to maternal, paternal to paternal) or to the other DNA molecule (maternal to paternal or paternal to maternal), leading to a visible "crossing-over" event (crossing over maternal to maternal or paternal to paternal is generally invisible). Typically, multiple "cross-over" events occur along the length of each set of paired (replicated) homologous chromosomes. Whenever maternal-paternal crossing over occurs the recombinant chromosome contains a different set of alleles than either the original paternal or maternal chromosomes. You can convince yourself by following any one DNA molecule from beginning to end.

In addition to shuffling alleles crossing over can create new alleles. Consider the situation in which two alleles of a particular gene are different from one another (→). Let us assume that each allele contains a distinct sequence difference (as marked). If, during meiosis, a crossing over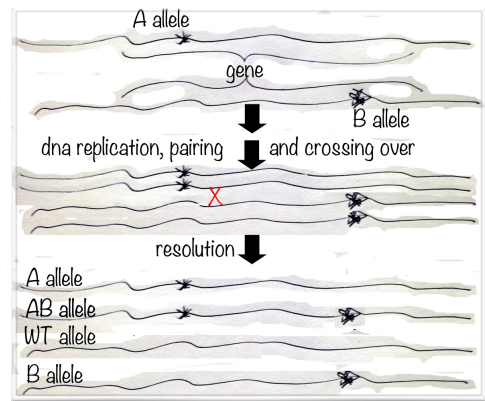 event takes place between these sites, it results in one allele that contains both molecular sequences (AB), and another allele with neither (indicated as WT). A new allele (AB) has been created, without a new mutation!

In the case of the X and Y chromosomes, the chromosomes

pair with one another through their common psuedo-autosomal regions (see above), which are syntenic. Outside of these regions there is no significant synteny between the X and Y chromosomes, leading to the suppression of crossing over over much of the X and Y chromosomes' length in males. In contrast, crossing over can occur normally (that is, just like for autosomes) between the two X chromosomes in a female.

In addition to the shuffling of alleles along chromosomes, and (occasionally) generating new alleles, meiosis leads to a second source of variation (←). At the first meiotic division, the duplicated (and recombined) chromosomes remain attached at their centromeres, so that each of the two resulting daughter cells receives either the duplicated maternal or paternal chromosome centromere region. However, what set of chromosomes (defined by

_____
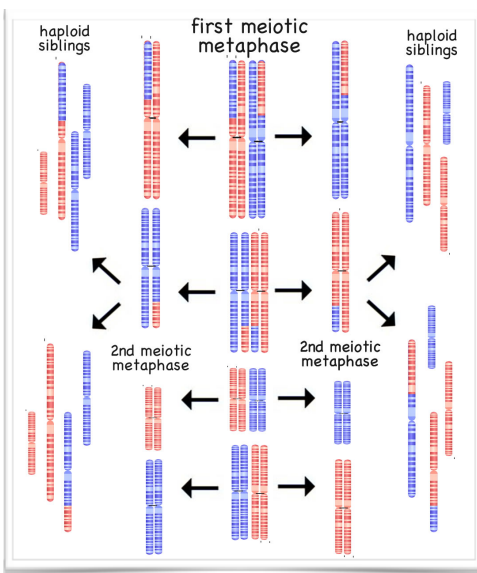
[426] adapted from The Centenary of Janssens's Chiasmatype Theory Koszul et al., 2012. Genetics **191:** 309-317.

their centromeres, maternal or paternal) they inherit is determined by chance. For an organism with 23 different chromosomes (such as humans), the first meiotic division produces 2^{23} possible different daughter cells. The process is known as the independent assortment of homologous chromosomes during the first meiotic division, or independent assortment for short.

There is no DNA replication between the first (M1) and the second (M2) meiotic divisions. During the second meiotic division the replicated chromosomes, held together at their centromeres, attach to the spindle. Because of recombination, the two chromosomes are not necessarily identical, which further increases (to rather astronomical levels) the number of different chromosome sets a particular haploid cell can inherit. When they separate the two resulting sibling cells each receives one and only one copy of each chromosome (a double-stranded DNA molecule); again, which molecule they inherit is stochastic. The four haploid cells that are generated by meiosis are known as gametes (or at least are potential gametes). In males, all four haploid cells differentiate to form sperm cells, whereas in females, typically one of the four haploid cells differentiates to form an oocyte, which becomes an egg, which can be fertilized by fusion with a sperm cell, and the other three form what are known as polar bodies. The result of meiosis is to generate gametes in which the alleles present in the maternal and paternal chromosomes have been shuffled in various ways, so that the resultant offspring has a genome related to, but distinct from that of either of its parents.[427] Fertilization (the fusion of gametes) combines two such genomes, one maternal and one paternal, to form a new organism, with a novel combination of alleles. Most phenotypes are influenced, to a greater or lesser degree, by the set of alleles within a genotype, and new combinations of alleles will generate new phenotypes and phenotypic variation that can impact reproductive success, and so lead to evolutionary effects.

Questions to answer:
208. Consider the odds of an organism obtaining the 3 new mutations necessary for the appearance of a new trait. If you were to predict, which would be faster (in terms of the number of generations required) in achieve this goal, a sexual or an asexual organism. Generate a drawing that illustrates your thinking.
209. You are working with an organism with 5 autosomes and 1 sex chromosome. Considering only the effects of independent assortment during meiosis, how many different types of gametes could be generated? A drawing of the process could help.
210. Indicate (in a drawing and associated explanation) how a deleterious mutation within a gene could be generated by or eliminated from a gene.
211. How would genetic diversity be altered if meiotic recombination occurred during meiosis II, rather than during meiosis I?

Questions to ponder
- Under what conditions might you expect the evolution of sexual reproduction to be selected against?
- Why are parents and their siblings not necessarily good donors for organ transplantation?

Linkage & haplotypes

An important feature of meiotic recombination is that it can "disconnect" the alleles of genes located near one another along a chromosome. Consider the situation when a mutation occurs that creates a new allele in gene X; let us call it X^{select}. Now let us assume that this allele is subject to strong positive or negative selection. That means that the presence of the X^{select} allele in an organism has a strong effect on reproductive success. Because it is either strongly selected for (positive selection, positive

---

[427] This even applies to hermaphrodites, in which one organism acts as both mother and father!

effect on reproductive success) or against (negative selection, negative effect on reproductive success) the frequency of the allele will tend to increase or decrease in subsequent generations. The change in the frequency of the X^{select} allele also influences the frequency of alleles near the X gene. If X^{select} is subject to strong positive selection, so will alleles in genes neighboring (linked to) gene X; similarly, if X^{select} has a negative selective effect, the frequency of the alleles in genes neighboring (linked to) gene X will also decrease over time. The closer the genes are to each other along the chromosome, the longer (over more generations) this "linkage" effect will last. Why? because the probability of recombination between two sites along a chromosome (two genetic loci or positions) is a positive function of distance. As the distance between two genetic loci increases, the probability that the original alleles at these positions will be separated by recombination increases. When the probability of a recombination event between two genes reaches 50% or greater, the genes behave as if they are on different chromosomes – they become "unlinked." Linkage distances are calculated in terms of centimorgans, named after the geneticist Thomas Hunt Morgan (1866-1945). A centimorgan corresponds to a 1% chance of a crossing over event between two specific sites along a chromosome. In humans, a centimorgan corresponds to ~1 million base pairs of DNA. Two genetic loci that are 50 centimorgans (or more apart) are separated by ~50 million or more base pairs. In the context of meiosis, two genetic loci on the same chromosome, but separated by >50 centimorgans, have the same probability of being inherited together as if they were on two different chromosomes. We will return to this again, when we consider the interpretation of genetic crosses.
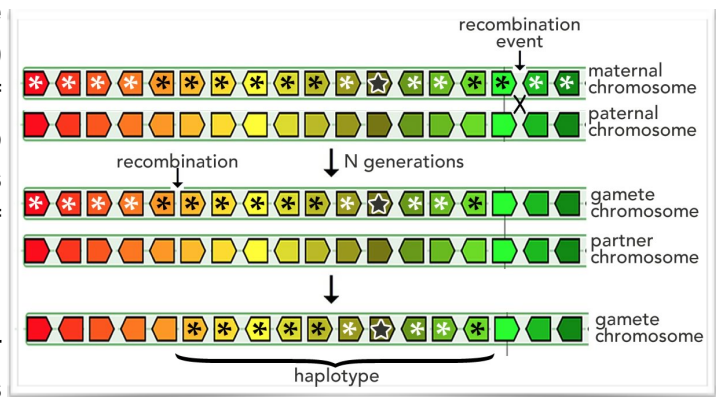
Consider a particular allele of a particular gene, marked by the star (★) here (→); let us assume that this allele is associated with a visible trait. We will mark the alleles found in neighboring genes on this chromosome with asterixes (*). For the sake of clarity assume that different alleles (un-marked) are found on the homologous chromosome. During meiosis, recombination events will occur randomly across these chromosomes. Over time independent recombination events occur that will increasingly reduce the size of the region of the original chromosome (containing the ★ allele). This original region is known as a haplotype; it is a group of alleles that are inherited together from a single parent. From a formal point of view, it is not clear which variation within the haplotype region is responsible for the trait observed. In the era of genetic (pre-molecular days), multiple rounds of crosses (breeding cycles) were required to locate i) on which chromosome the allele (gene) responsible for a particular trait was located, and ii) where, more or less exactly, the allele (gene) was located along the length of the chromosome. With more and more generations, the size of haplotype regions becomes smaller.

Now consider how the alleles within a particular region can be maintained together. Let us assume that the original allelic variant has effects on the expression of neighboring genes (→); how could this occur? Two obvious ideas suggest themselves: the allele can influence the packaging of the chromosome region, so that the genes' accessible to other regulatory

factors is modified or the allele can itself effect or be in an gene regulatory element (an enhancer) that plays an important role in the regulation of multiple genes in this molecular neighborhood. Both options could lead to selective effects based on the maintenance of the integrity of the chromosomal region (a haplotype) - that is, recombination events within the region can occur, but because they have a negative effect on reproductive outcomes they would be selected against.

Questions to answer:
212. Graph, as a function of distance, the likelihood that recombination will disconnect a selected (whether positively of negatively) allele from alleles in surrounding genes.
213. Why might a crossing over event inhibit nearby crossing over events?
214. How can you use the size of a conserved genomic region to estimate time of isolation of a population?
215. What are the benefits of recombination in terms of environmental adaptation?

Questions to ponder:
- How does the size of haplotype regions reflect the reproductive history of a population?
- How does the presence of a deleterious allele influence the selective pressures on an organism? How might it open up (over generational) time, new evolutionary possibilities?

Maternal and paternal effects

One of the implications arising from the dimorphism in gamete size and functional roles is that some alleles preferentially influence oocyte/egg and sperm functions. For example, in a number of organisms, particularly those that develop rapidly and external to the maternal parent after fertilization, most of the gene products and nutrients needed to support early development of the new organism are supplied by the much larger egg (→), accumulated during the course of oogenesis. Defects in the oocyte, due for example to recessive alleles in a homozygous mother, may lead to defects in the behavior of the fertilized egg and embryo that cannot be rescued by a sperm cell carrying a wild type (dominant) allele. Similarly, since mitochondria are supplied to the zygote by the egg and not the sperm, defects in the mitochondrial genome cannot be rescued by sperm, even if the sperm is generated by a male with normal mitochondria and a normal mitochondrial genome. Similarly, sperm supply components of the mitotic apparatus to the zygote, fertilization by an aberrant sperm can lead to an early defect in the embryo.
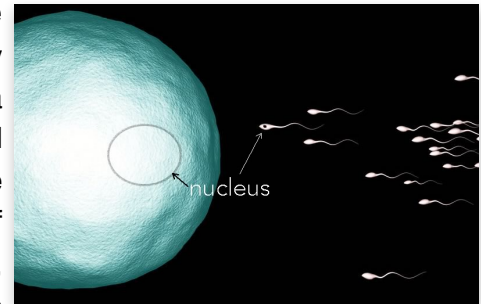
Imprinting:

Another molecular factor that we will consider anon is the fact that for a number of genes, the DNA of sperm and egg are modified differently, in a process known as imprinting.[428] Imprinting involves sequence specific modifications of the DNA; they are epigenetic in that they do not alter the gene's sequence but rather influence when and where a gene is expressed. Because patterns of imprinting are different in males and females, the maternal and paternal alleles present in a new diploid organism

---

[428] Genomic Imprinting: http://learn.genetics.utah.edu/content/epigenetics/imprinting/

may be expressed differently, that is in some cells only the maternal allele of a gene will be expressed, whereas in other cells only the paternal allele will be expressed. As we will see as we come to consider the genetic behaviors of genes (alleles), imprinting complicates things.

X-inactivation and sex-linked traits

One aspect of the XY chromosome-based system of sex determination is that the two sexes have different genotypes, at least with respect to these chromosomes. As mentioned above, the Y chromosome is ~59 million base pairs in length and encodes ~50 genes, while the X chromosome is ~155 million base pairs in length and encodes ~1000 genes. This creates a genetic imbalance between the two sexes in terms of gene copy numbers. A single gene can direct the synthesis of only so many RNA molecules per unit time, based on the rate of RNA polymerase binding, activation, and RNA synthesis along a DNA molecule. Without some "balancing" mechanism, we would predict that female cells would have about twice as many RNAs for genes on the X as do similar cells in a male (and most cells in males and females are the same or very similar). This is the reason for halpo-insufficiency; an amorphic (null) allele can be associated with a dominant phenotype if a single functional copy of the gene is insufficient to produce enough gene product. There would seem to be a need for some form of "dosage compensation"; either genes on the X in males have to be expressed more efficiently or genes on the X in females have to be expressed less efficiently.

The strategy used in humans and other placental mammals is a process known as X-inactivation. At random points during development, one or the other of a female's X chromosomes becomes associated with specific RNAs and proteins, and is packed into a compact structure that can no longer support gene expression (RNA transcription). Once the choice of which X chromosome to inactivate is made, it is stable and inherited through mitotic cell division, generating clones of cells with the same X chromosome active and the other inactive. A failure of X-inactivation leads to developmental arrest and embryonic death. While gene expression from the inactivated X is largely inhibited, the replication of the inactivated chromosome continues with each cell cycle. We can see the effect of this choice in female calico cats (→), in which the different coat colors reflect domains in which one or the other X chromosomes is actively expressed, while the other X chromosome is inactivated (a gene for coat color is located on the X).

The X-chromosomal inactivation system consists of two genes, XIST and TSIX. XIST encodes a functional ~19.3 kilobase long non-coding RNA, that is, not an mRNA, and is expressed only in cells with two X chromosomes – so it is not expressed in males.[429] Which of the two X-chromosomes expresses XIST is determined stochastically. When expressed, the XIST RNA associates with regions adjacent to the XIST gene and eventually comes to be localized along the entire length of the X-chromosome on which the active XIST gene is located. The XIST RNA comes to associate with a number of protein complexes involved in inhibiting gene expression and producing the compact state of the inactivated X, also known as a Barr body.

On the other (active) X-chromosome, located on the other strand of the XIST gene is a over-lapping gene known as TSIX. As you should be able to explain, the promoter of TSIX is distinct from

---

[429] X-inactivation-specific transcript ()

that of XIST. The TSIX gene encodes a ~40 kilobase long non-coding RNA that is partially complementary to the XIST RNA. The TSIX RNA acts to inhibit XIST activity, and so blocks the action of XIST on the active X chromosome, blocking its inactivation. Together the XIST/TSIX system insures that one and only one of the two X chromosomes is active.

X-linked diseases and mono-allelic gene expression

While calico spots occur only in female cats, there are a number of genetic susceptibilities that are seen in males; these arise because males have only a single X chromosome. The result is that, in contrast to the rest of the genome, genes on the X are effectively haploid in males. The result is that the phenotypes associated with recessive alleles are visible in males. In contrast, in females that are formally heterozygotic for that gene, some cells express one allele while others express the other. This situation (in females) leads to what is known as random monoallelic expression. Recent studies have revealed that random monoallelic expression occurs throughout the genome, even in autosomal genes. In a typical diploid cell, one gene may be active while the other copy of the gene, on the homologous chromosome may be inactive, due to various stochastic events.[430] In some cases of stable monoallelic expression there is what is known as somatic selection, which we will return to (→). Given that there are two alleles, when they are different which is expressed may differentially influence cell growth and division, or even cell survival, so that over time, cells expressing one allele may come to dominate (in numbers) those that express the other. The extent to which random monoallelic expression influences human development and disease is just now being recognized and examined carefully. We will return to this subject in the context of cancer evolution and brain development.
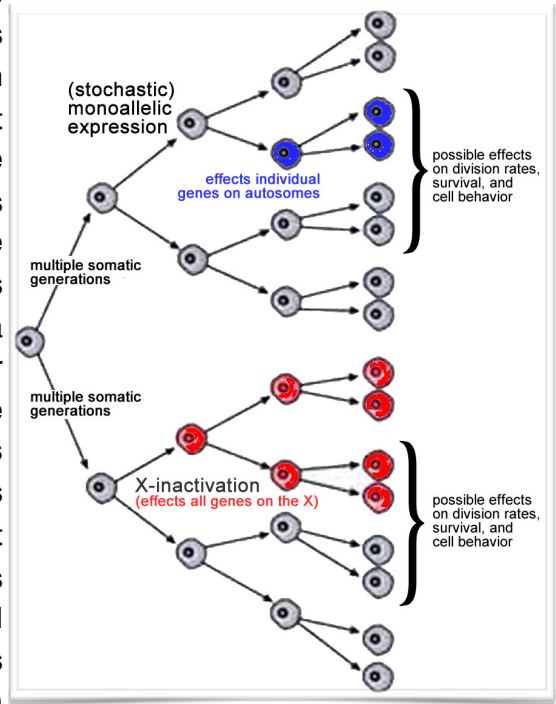
Questions to answer:
216. What does it mean to be mosaic for an allele?
217. Why do males and females differ in the traits they display?
218. Why do males and females differ in the display of phenotypes associated with genes on the X chromosome?
219. Can you provide a plausible mechanism to explain why (autosomal) random monoallelic expression occurs?
220. How can monoallelic expression impact an organism?

Question to ponder:
- Under what conditions might mono-allelic (autosomal) gene expression be beneficial?

---

430 Monoallelic Gene Expression in Mammals

Chapter 14: Generating mutations and becoming alleles

In which we consider how mutations appear and become alleles within a population. Distinguishing between randomness and purpose.

We are far enough along to recognize that beginning with a particular genome, any change in that genome, whether due to errors in its replication or environmental (chemical reaction or radiation induced) damage that goes unrepaired, that is, that is not returned to the original state, results in a mutation. If the mutated cell/organism survives and gives rise to offspring, and if the mutation lies within a gene, it becomes an allele - a variant within a population. If it lies outside of a gene, it becomes a polymorphism. With the advent of genomic sequencing, and related technologies, it is possible to estimate the rates of mutation in a particular organism or a particular cell type.[431] Here we distinguish between mutations in the germ line (leading to eggs and sperm) and in somatic cells, the cells of the body. As a first approximation, mutations occur randomly within genomes, but in fact there are what are known as mutational hotspots - for example, CpG dinucleotides are mutated more frequently (~10X) than other dinucleotides. In addition to single nucleotide changes, there are also mutations that involve small insertions and deletions, known collectively as indels; these are defined to be less than 20 base pairs (bps) in length, and are distinguished from larger changes, known as structural variants (> 20 bps).[432] It has been estimated that each generation sees the addition of ~3 indels and ~0.16 structural variants in the germ line of each person. In addition, there are copy number variations (CNVs), these can lead to changes in the number of copies of a particular gene or sequence.[433] Many mutations are associated with the process of cell division. Mutations occur more frequently in the soma because there are more cell divisions involved, trillions in the human. Similarly, there are fewer cell divisions involved in the generation of oocytes in females than the generation of sperm in males, and the number of mutations, particularly in the male germ line, increases with age. As you can probably predict, germ line mutations can be passed from generation to generation, while somatic mutations are lost with the death of the host. The current estimate is that the chance of a de novo germ line mutation in humans is ~1×10^{-8} per base pair per generation (remember the human genome contains ~6×10^9 bps). Somatic mutations appear to be the prime driver of cancer, and we will discuss both germ line and somatic mutations, and their effects, in awhile.
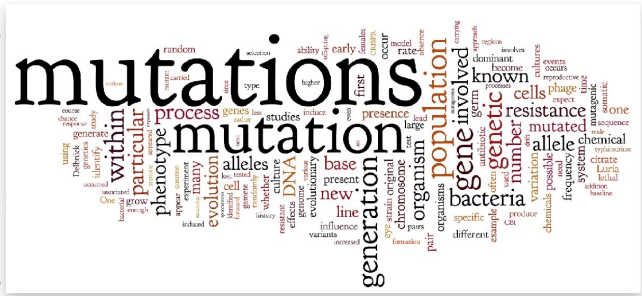
Mutations into alleles

For a mutation to become an allele the first criterion is that it does not have a dominant lethal phenotype; why dominant? In a diploid organism a new mutation will involve only one of the two genes

---

[431] see: The origins determinants, and consequences of human mutations

[432] Indels

[433] Copy Number Variation

present; for it to have a phenotype, it needs to be dominant over the other allele present. Of course this is not the case in prokaryotes, which are typically haploid. If the mutation is not dominant lethal, and if it occurs in the germ line, it can be passed to a gamete and from there into the next generation. Again, this assumes that the allele does not produce a lethal phenotype in the gamete or the early zygote, since where and when a gene is expressed, has a lot to do with the phenotypes it is associated with.

A non-lethal, non-dominant mutation has a chance to become an allele within a population, but first it has to avoid elimination. Remember that when it first appears in the germ line of a sexually reproducing organism (we will ignore somatic mutations for the moment, since they are "trapped" within a particular organism), there is only one copy of the mutated allele, it is possible that the gamete carrying it fails to fuse with another gamete, if so, this new mutant allele would be lost. This is a version of genetic drift. Similarly, the mutant allele may make it into the next generation even if it is deleterious (although not too deleterious), again just by chance.

If a mutant allele survives these early events, it comes to be referred to as an allele when it is found to occur in >1% of the population. Mutations that occur outside of a gene become what are known as polymorphisms, and are part of the differences between organisms, although unless they lie within an enhancer, are unlikely to influence gene expression or visible phenotype. The difference between allele and polymorphism lies in the ability to recognize what is, and what is not, a gene, something that can be tricky. The genetic variation within a population reflects its past history (the combination of selective pressures and non-adaptive events, such as founder effects, bottlenecks, and genetic drift) and serves as the basis for subsequent evolutionary change.

Luria & Delbrück: Discovering the origin of mutations

Keeping in mind that Darwin and Wallace had no clear understanding of exactly where genetic variation came from, an important question that arose early in the history of evolutionary theory was whether mutations (a prime source of phenotypic and genetic variation) associated with the evolution of new and complex traits – such as the eye – were the result of random (stochastic) events or were they somehow directly and purposefully generated in response to the needs of the organism. In the absence of a clear understanding of how genetic information arose in a population or how it was passed from generation to generation, there was really no way to distinguish between Darwinian (random variation + selection) and Lemarckian (adaptation based on need) evolutionary mechanisms, although Lemarckian mechanisms seemed kinder.[434]

To understand how this question was resolved, we will consider a classic experiment, known as the Luria-Delbrück experiment after the researchers, Salvador Luria (1912-1991) and Max Delbrück (1906-1981).[435] The study was published in 1943, before DNA was recognized as the genetic material and well before anyone understood how genetic information was stored.[436] In their study, Luria and Delbrück examined bacterial resistance to viral infection. The bacteria they used could be infected and killed by a specific type of bacteriophage (phage for short). Some of the mutations that arise

---

[434] This led to what was known as the "Eclipse of Darwinism"; biology emerged from this "darkness" with the development of an understanding of genes and genetic mechanisms to produce what became know as the "Modern Synthesis".

[435] Luria–Delbrück experiment

[436] Mutations of bacteria from virus sensitivity to virus resistance: http://www.genetics.org/content/genetics/28/6/491.full.pdf

spontaneously in the bacteria rendered them, and their off-spring, resistant (immune) to phage infection. The question Luria and Delbrück asked was, are phage resistance mutations appearing randomly all of the time in bacteria or is it the presence of the virus that induces their appearance in response to the bacteria's "need" to be immune. Is immunity learned or lucky?[437] If the mechanism is random, then we can expect that the number of mutational events will vary dramatically from one population (culture) to the next - the variation in the frequency of phage resistance (and the mutations that produce it) between independent populations will be large. On the other hand, if the generation of phage resistance mutations is an adaptive process, then we would expect that the frequency of phage resistance (mutations) will be more or less uniform from one population to the next – repeating experiments on different cultures should produce resistant bacteria at approximately the same rate in each (→). In a sense, as proposed by Darwin, evolution involves random mutations in individuals, whereas a Lemarckian mechanism involves induced responses by the population as a whole.[438]

Luria and Delbrück started a number of bacterial cultures to which they then added enough virus (at the time of the horizontal red line) to kill every sensitive bacterium. They then plated out the culture and counted the number of phage-resistant bacteria present, each of which can grow up into a macroscopic (asexual) clone, a colony.[439] The number of such cells in a culture will reflect when in the history of the culture the resistance mutation appeared; for example, if it appeared early in the history of the culture, as in the red-boxed culture (↑) in the spontaneous mutation model, it would be common, whereas if it appeared late, it would be rare. The two models (induced/Lemarckian versus spontaneous/Darwinian) make dramatically different predictions. In the induced/Lemarckian model, the variation of resistant bacteria between cultures is expected to be low, since resistance arises through a common "inductive", physiological process, even though we do not know how that process works. In contrast, in the spontaneous/Darwinian model we expect large variations, with many cultures having no resistant bacteria and some having many, depending upon whether and when the mutation occurred, a chance event. If the mutation occurs late, most bacteria will be killed (population 2); if the mutation occurs early (as in population 5, boxed in red) there will be many resistant bacteria present. In the lower panel, Luria and Delbrück calculated what they expected from their experiment if the spontaneous/Darwinian model occurred - their observed
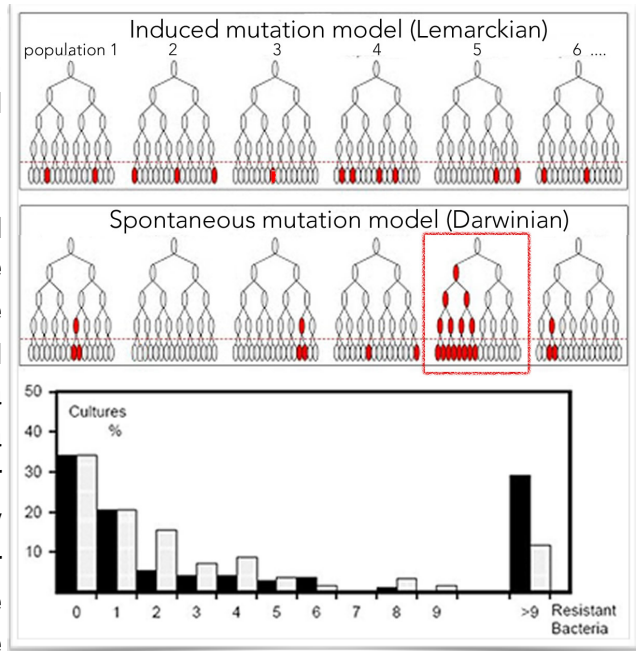
---

[437] As we will see later on, there are molecular mechanisms, such as the CRISPR CAS9 system that can learn and lead to acquired immunity.

[438] This is perhaps one reason that collectivist ideologies, such as the Soviet Union under Stalin, so disliked Darwinian evolution (and harshly prosecuted geneticists). see http://blogs.plos.org/scied/2017/04/10/science-politics-marches/

[439] The logic and details of their experiment are the subject of this virtual lab lab on the Luria-Delbruck experiment

results (black bars) matched this prediction, allowing them to conclude that, at least in this system, mutations were occurring independently of the presence of the virus.

To date there is no evidence that environmental factors can specifically induce the generation of beneficial or useful mutations. What can happen, however, is that the general (non-specific) mutation rate can increase in response to various stress conditions, arising from internal or environmental effects. Typically an increased mutation rate involves effects on the efficiency of DNA error repair systems, which leads to increased levels of genetic variation upon which selection can act.[440] The ability to control mutation rates occurs within the vertebrate immune system, through a process known as somatic hypermutation.[441] This process is involved in the maturation of the immune response and the generation of increasingly specific antibodies, a topic well beyond our scope here. That said, the mechanism is known; these cells activate a gene that encodes an "activation-induced deaminase" or AID (OMIM:605257). AID acts on cytosine residues to generate uracils, which when repaired generate an A:T base pair, replacing the original C:G base pair. The other genes in these cells appear to be at least partially protected by "selective targeting of AID and gene-specific, high-fidelity repair of AID-generated uracils".[442]

Generating mutations experimentally

When we think about a particular trait or a behavior, a specific phenotype, we often want to know how many different genes are involve in producing that phenotype. One approach to begin to answer that question is to determine how many genes can be mutated so as to disrupt the formation of that phenotype. Such a search for mutations that disrupt a particular phenotype is known as a "forward genetic screen", and has, historically, been used to identify the molecular components involved in the process. The first step in such a screen is to generate mutations. Waiting for naturally occurring mutations is too slow for the ambitious (and mortal) researcher, so steps are taken to induce large numbers of mutations. Among the first of these mutagenesis methods was irradiation using X-rays. In 1927, H.J. Muller, who we have met before, was the first to create a mutation using X-rays (→).[443] He examined the generation of mutations on the X chromosome of the fruit (or more accurately vinegar) fly *Drosophila melanogaster*, an organism chosen in part because of its small size (which

---
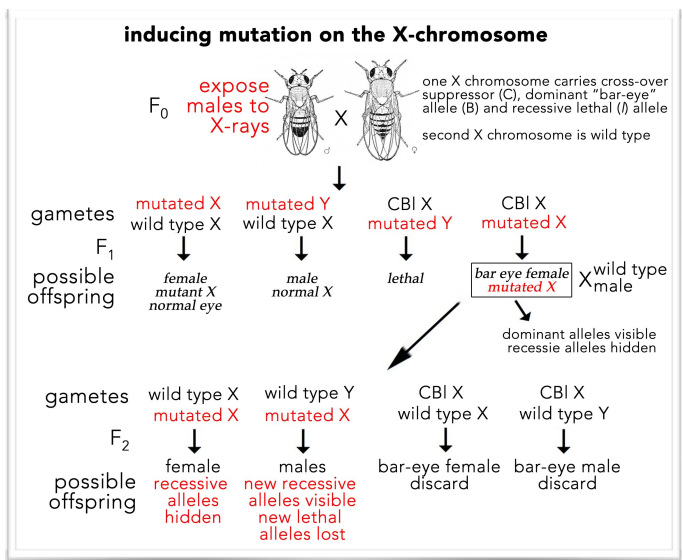
[440] A trade-off between oxidative stress resistance and DNA repair plays a role in the evolution of elevated mutation rates in bacteria

[441] Somatic hypermutation: wikipedia

[442] Two levels of protection for the B cell genome during somatic hypermutation

[443] Hermann J. Muller (1890-1967) demonstrates that X rays can induce mutations

allows lots of animals to be raised in a limited space), rapid life cycle, and the large number (~400) offspring that are produced by a single female after a mating. In previous studies, he had isolated a
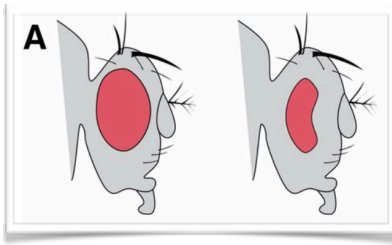
version of the X-chromosome, known as CBI, that carries a dominant allele that produces bar eyes (←), a recessive lethal mutation, and a large inversion (a flipped region) in the chromosome; the presence of this inversion generates embryonic lethal mutations if recombination occurs within the inverted region (think about how a chromosome with a large inversion will interact with a normal chromosome during meiosis.) Muller took wild type male flies and irradiated them, so that mutations were induced in their testes, producing sperm with mutations. He then mated females carrying the altered CBI X-chromosome with the irradiated males (you should be able to explain why he could not have used males carrying the CBI chromosome). Based on the markers present, he could identify females that carried the CBI X chromosome and a mutated X chromosome from an irradiated male (↑ previous page). When these first filial generation (F_1) females were mated with a wild type male, the offspring that carried a mutated X chromosome could be identified and analyzed. Males displayed phenotypes associated with recessive alleles (mutations) on the X, while dominant mutations were visible in females. Through this analysis, Muller identified hundreds of new mutations and, more importantly, showed that the genetic material could be damaged, or rather altered, by radiation.

Since these studies, a number of other methods have been found to induce mutations, all act by damaging the DNA in one way or the other. For example, animals can be fed potent mutagenic chemicals, such as ethyl methane sulfonate (EMS)(→). EMS reacts, through an esterification reaction, with guanosine residues in DNA, modifying them by the addition of an ethyl group. The modified G base (G*) pairs with T rather than C; when the modified DNA is replicated, one copy is wild type while the other generates an aberrant AG* base pair, which is then repaired to produce a mutation, replacing the original CG base pair with an TA base pair.
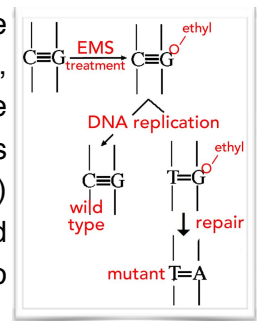
To identify chemicals that can induce mutations, Bruce Ames (b.1928) and colleagues developed a test using the bacterium *Salmonella typhimurium*.[444] They began by using a strain of *S. typhimurium* that carries a mutation that rendered it unable to grow in the absence of the amino acid histidine; they termed this strain his−. The his− strain can be reverted to a his+ strain by mutation. To test whether a chemical is mutagenic in *S. typhimurium*, his− cells were grown up in the presence of histidine (to allow for growth) together with the chemical to be tested. Typically, a number of different concentrations of the chemical are tested. After some time the cultures are plated out onto agar plates in the absence of histidine. The result is that only those bacteria that have acquired a mutation that converts them from a his− to a his+ phenotype can grow into macroscopic colonies (→). There is, of course, a low rate of spontaneous mutation, that is mutation in the absence of test chemical; this enables us to estimate the baseline mutation frequency for the *S. typhimurium* strain used. If the chemical to be tested is mutagenic, then the frequency of mutations
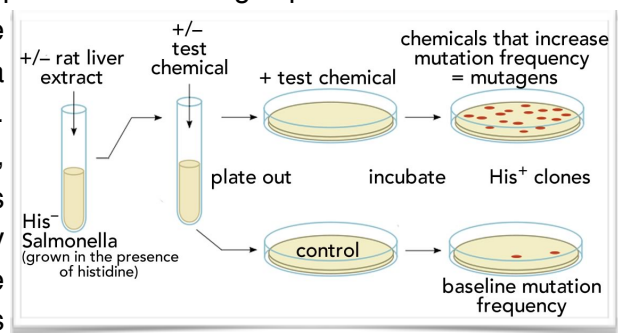
444 Ames test (wikipedia)

should increase above this baseline rate; we also expect that the mutation rate will increase as a function of the concentration of the chemical tested. Hopefully you appreciate (but we will remind you) that while we are assaying for the appearance of his⁻ to his⁺ mutations, mutations are occurring randomly throughout the genome of the organism - most fail to produce a discernible phenotype.

An important variation of this assay, needed to adopt it to organisms such as humans, was the recognition that many chemicals that you might be exposed to are metabolized in the liver. Such reactions generate related chemicals, which may well be significantly more (or less) mutagenic than the original compound. To mimic such metabolic effects, it is possible to add liver extracts to the original culture. Because cancer arises due to somatic mutations, it is clear that we would like to minimize our exposure to mutagenic chemicals. But often a particular chemical is significantly mutagenic only at high concentrations, much higher than you would ever be exposed to. So while many chemicals can induce mutagenesis many fewer are carcinogenic, in part because most mutations are repaired and exposure levels are low enough to have little effect on the baseline mutation frequency.[445]

Questions to answer:
221. How would increasing the mutation rate influence the outcome of the Luria-Delbrück experiment.
222. What are the advantages (for a geneticist) for choosing an organism with hundreds of offspring per mating event?
223. What is the advantage of studying traits that alter non-essential structures?
224. Why is it not possible to identify every gene involved in the formation of a complex trait by a simple mutagenesis approach?
225. What is responsible for the baseline mutation frequency (in the Ames test)?
226. A compound produces mutations in the Ames test; what factors would influence your decision about whether to worry about exposure to that compound?

Questions to ponder:
- Given the frequency at which phage resistance arises, can you provide a plausible reason for why resistance to bacteriophage is not already a universal trait in prokaryotes?
- How would it change your perspective if mutations occurred because organisms need them, rather than randomly?
- How does the apparent fact that evolution depends upon random mutations to generate new genes and new "types" of organisms, new species, influence your view of the meaning of existence?

Longer term mutation / evolution studies

We can see the spontaneous mutation model applies throughout the biological world, where ever we look mutations appear to arise randomly. If they can persist within the population (see above), they become alleles. It is worth reiterating that because of non-adaptive processes such as genetic drift, new neutral or beneficial mutations are often lost because initially they are extremely rare within the population, while mildly deleterious mutations can become fixed.

To study such evolutionary processes in a laboratory setting is not easy, but the now classic example of such a study has been carried out by Richard Lenski (p. 1956) and his associates; they have been growing 12 originally identical populations of the bacteria *Escherichia coli* for more than 25 years and > 60,000 generations.[446] One, of many, characteristics of *E. coli* that distinguish it from other

---

[445] "All substances are poisons; there is none which is not a poison. The right dose differentiates a poison…" Paracelsus [link]

[446] *E. coli* long-term evolution experiment: wikipedia and the Lenski lab's *E. coli* Long-term Experimental Evolution Project site

bacteria is that it cannot metabolize citrate in the presence of O_2. In the course of their studies, Blount et al observed the appearance of variants of *E. coli* that could metabolize citrate in the presence of O_2 in one of their cultures; a beneficial evolutionary adaptation, since it provided a previously un-utilized energy and carbon source.[447] By tracking backward, the investigators identified a "pre-disposing" mutation that occurred in this lineage around generation 20,000; the presence of this mutation made it more likely that subsequent mutations would enable cells to grow on citrate, the Cit+ phenotype. Molecular analyses indicated that the initial Cit+ phenotype, which appeared around generation ~31,500, was weak. Molecular analyses established that it involved a ~3000 bp genomic duplication that led to increased expression of the citT gene, which encodes a protein involved in the import of citrate into the cell. Subsequent studies identified mutations in other genes in the Cit+ strain that further improved the mutant cells' ability to metabolize citrate.[448] One of these mutations led to increased expression of DctA, a gene that encodes a membrane transport protein that increases the cell's ability to import various nutrients normally released into the media, given the cell a reproductive advantage when grown on citrate. An interesting aspect of these studies was the backlash from some creationists, who reject the possibility of the evolution new traits via mutation and selection.[449]

A second more recent study on bacterial evolution, this time looking at the evolution of resistance to an antibiotic, used a giant agar plate (a "megaplate") with a gradient of antibiotic on it (→). Bacterial cells were placed in the regions free of antibiotic, and over time their ability to grow into regions of higher and higher antibiotic concentrations was visualized directly (video link). It is possible to watch the emergence of new variants at the boundary regions, as new mutations arise.[450]
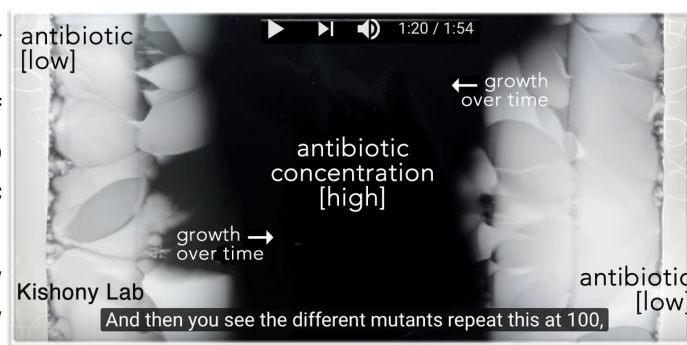
An important point to recall about the bacterial evolution studies is that these organisms are reproducing asexually, as clones. That means that they have no issue with interbreeding with other organisms in the population, but it also means that (in the absence of horizontal gene transfer) all necessary mutations need to occur in a single clonal population. As we discussed in the evolution section, if such mutations lead to a reproductive advantage they can, barring accidental death, take over the population – a process known as a reproductive sweep. This can lead to the loss of alleles present in other clones within the population; if they are useful, they will need to appear again, through mutation and selection (or transferred horizontally). In sexually reproducing (diploid) organisms, the various stochastic events involved, including gamete formation and fertilization, can lead to the loss of

---

[447] see Historical contingency and the evolution of a key innovation in an experimental population of Escherichia coli.

[448] see Genomic analysis of a key innovation in an experimental Escherichia coli population.

[449] The evolution of citrate metabolizing *E. coli*: the "Lenski affair"

[450] Baym et al., 2016 Spatiotemporal microbial evolution on antibiotic landscapes.

alleles (by genetic drift). Also the importance of reproductive barriers between subgroups within the original population, adapting to different ecological niches, can impact evolutionary processes.

Forward and reverse genetics

Originally, genetic analyses were carried out through what are known as forward genetics. Forward genetics involves the generation of mutations, essentially at random, and then identifying individuals carrying mutations that disrupt a particular process of interest. As an example, consider eye color in the fruit fly *Drosophila melanogaster* (↓). Eye shape and color are experimentally accessible because a

from Sturtevant & Beadle, 1938
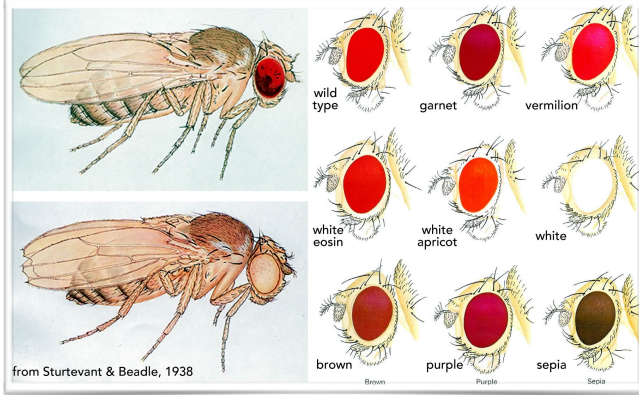
Drosophila embryo can develop into a fertile adult without an eye; this makes it possible to identify mutations (alleles) that alter the eye but allow other aspects of embryonic development to occur (more or less) normally. On the other hand, if the product of the mutated gene plays multiple roles in the developing organism, perhaps in processes distinct from those involved in eye formation, the embryo may die before eyes form, and no mutations in that gene will be recovered, even though the gene's product plays a key role in eye development or pigmentation. It is for this reason that forward genetic screens for mutations that influence a particular process are rarely if ever complete, that is, they do not identify every gene involved in a process.

While early geneticists worked with forward genetics, often known as classical genetics, there are reasons that this approach generally fails to generate a complete map of the genes involved in a particular process. An alternative approach is to determine whether a specific gene is involved in a particular process. While there are a number of ways to mutate a particular gene, the mechanisms involved are largely beyond the scope of this course. One exception is the recently developed CRISPR CAS9 system, which is one of a number of anti-viral infection systems found in bacteria and archaea.[451] The Cas9 enzyme is an endonuclease that creates double-stranded breaks in DNA. What makes the system distinctly different, and extremely powerful, is that the site at which the endonuclease cuts the DNA is determined by a ~23 base pair RNA sequence, a guide RNA (gRNA) – a sequence that is long enough to (often) occur once and only once within the genome of an organism, even an organism with a genome of more that a billion base pairs, such as humans. This gives an extremely high degree of specificity to the system. After the double-strand break is made, host cell DNA repair systems act to join the two ends of the DNA molecule back together again, but this joining is not accurate – base pairs can be lost or added, generating a mutated form of the original DNA sequence. More and more sophisticated genetic manipulations can be generated using variants of the CRISPR Cas9 system; for example, regions of a gene can be deleted by using pairs of gRNAs (→). If the gRNA sequence is present in both alleles of a gene, both alleles

Each Cas9 complex generates a double strand break in the DNA

repair by non-homologous end-joining (leading to deletion of the intervening DNA)

---
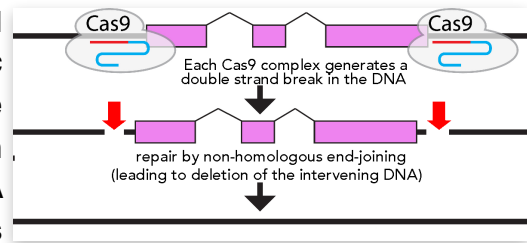
[451] over-view reference for the Crispr cas9 system: <u>wikipedia</u>.

can be mutated at the same time. Finally, if the CRISPR CAS9 system is activated (or introduced) early in the development of an organism all or most cells can be mutated, which can lead to multiple phenotypes. One way to control the effects of a mutation is to use various molecular strategies to express the CRISPR CAS9 system at specific times and in specific cells within a multicellular organism, something that will be expanded upon (we expect) in later classes.

Questions to answer:
227. How can a "predisposing mutation" influence the possible directions of subsequent evolution?
228. In the antibiotic resistance video (watch!), why is there often (but not always) a delay before the bacteria grow into a region of higher antibiotic resistance?
229. How would the presence of horizontal gene transfer impact the megaplate experiment?
230. How would an evolutionary sweep effect a human population?

Question to ponder:
- How would evolution be altered if the mutations (alleles) were induced rather than selected?

Chapter 15: Becoming Mendelian: analyzing alleles in terms of phenotypes & pathways

In which we consider the contributions of Gregor Mendel, namely the realization that distinct and stable genetic elements influence specific traits. These genetic elements behave in predictable ways during sexual reproduction. The behavior of chromosomes during meiosis leads to Mendel's rules of allele segregation and independent assortment. We come to recognize that the traits Mendel studied reflect a unique subclass of genetic elements. We consider how exceptions to Mendel's rules, including linkage, synthetic phenotypes, and epistatic behaviors, arise.

As we think about the origins of genetics, it is worth considering some of the biases imposed by the way that Gregor Mendel did his work, these reflect the realities of science – understanding does not appear fully formed, like religious revelation, rather it is build up by insights, some productive and others a distraction. Subsequent observations and experiments lead to the recognition of the implications and limitations of original ideas (tentative hypotheses and working models), and drive their modification, so as to explain a larger number of observations more and more mechanistically. To make genetic behaviors intelligible, Mendel purposefully grew plants (peas) that displayed highly reproducible and distinguishable traits. These were traits (phenotypes) that were not dramatically influenced by environmental or genetic background effects, but were due to alleles of a single genetic loci, single genes. Moreover, these genes were unlinked to one another – they were located on different chromosomes, and so were segregated independently of one another during meiosis. Finally, he used traits (as we will see) that were well behaved in terms of their interactions with one another (little or no interaction), and an experimental organism that produced high numbers of progeny that could be obtain from a well controlled matings. Moreover, the traits displayed clear dominant-recessive behaviors with respect to one another. It is not that Mendel knew anything about chromosomes and molecular mechanisms, it is just that his choices made the data he obtained intelligible.[452]

At the same time, it is worth recognizing explicitly that most alleles do not behave in these ways. What phenotypes a particular allele is associated with are often partial, difficult to recognize, and influenced by the genetic background of the organism. In this context it is important to remember that many laboratory studies (including Mendel's) are carried out in in-bred backgrounds, that is, all of the organisms in the study share a common genetic background (overall genotype). Such genotypic homogeneity is an artifact of the way such experiments are conducted; natural populations display much more background genotypic variation - there are many more different alleles present. Such background variation influences the phenotypes associated with a particular allele, whether hetero- or homozygous. Consider a dominant allele; even though the allele is present within the genome of the organism, the associated trait may vary - this variation is characterized in the terms expressivity and penetrance. So, what does that mean exactly? Variable expressivity refers to the observation that even

---

in the presence of the associated (dominant) allele, the phenotypic trait can vary. As an example, consider a pea; is each pea really wrinkled to exactly the same extent, or do they vary – are some a little more or less wrinkly, although not smooth? Such behavior indicates variable expressivity. Similarly, it is possible that out of 100 individuals (or peas) that carry a particular dominant allele, all of them display the trait associated with that allele, but only in an in-bred background. If in a wild, out-bred background, only 50% of the organisms carrying the dominant allele express the trait, it is said to be incompletely penetrant, presumably due to various other factors, including different sets of suppressor and enhancer alleles, located in other genetic loci, other genes.[453]

Hidden alleles within a population

Except for those traits that are fully (100%) penetrant and expressive, for all types of alleles their effects on phenotype can vary dramatically depending on genetic background, that is the rest of the alleles present in the organism – they can disappear and reappear within a family lineage (a family tree or pedigree). It was for that reason that Mendel restricted his studies to only fully penetrant dominant and recessive alleles; otherwise, the results of his studies would not have revealed the simple rules of inheritance that he discovered.

HOW MENDEL MADE HYBRIDS:

From: The Cartoon Guide to Genetics by Larry Gonick & Mark Wheelis

FIRST HE SNIPPED OFF THE ANTHERS WHILE STILL IMMATURE TO PREVENT "SELFING."

THEN HE DUSTED THE STIGMA WITH POLLEN TAKEN FROM THE DESIRED 'FATHER.'

FINALLY, HE TIED BAGS OVER THE FLOWERS TO KEEP OUT ANY STRAY POLLEN.

IN THIS WAY MENDEL WAS ABLE TO CONTROL THE PARENTAGE OF EACH GENERATION.

PSST! I THINK THE MONK IS PLAYING GOD!!

In his studies, Mendel carefully identified dominant and recessive alleles of a number of unlinked genetic loci with completely penetrant and expressive phenotypes, working primarily with the garden pea *Pisum sativum*.[454] These plants can reproduce by either self-fertilization, that is sperm (pollen) and eggs are derived from the same plant, or by out-breeding, which involves sperm (pollen) and eggs from different plants. He developed or identified pure-breeding lines, that is plants that produce offspring with phenotypes similar (identical) to their own when bred with themselves. One reason to choose the pea was that it is possible to control who breeds with whom (↑); the investigator can block self-fertilization. In such studies, the parents are known as the F_0 generation and their offspring are known as the F_1

---

[453] here is a particularly relevant recent study: Genetic background limits generalizability of genotype-phenotype relationships

[454] It might be worth considering the distinction between a study and an experiment. In an experiment, the system is subject to some perturbation, and we examine how the system responds. Typically the experiment begins with a hypothesis, a guess on how a particular perturbation, which we might think we understand, will influence the system. A study is more about observing and collecting data about a system. From such observations, we can make hypotheses about how the system under study might act under different conditions (another observational study) or how a perturbation (an experimental study) might alter the system's behavior. Our prediction of the outcome is known as the null hypothesis - and we generally examine the data collected to determine whether the null hypothesis is supported or not.

generation. The requirement that the lines breed true means, in practice, that they are homozygous for the alleles controlling the trait under consideration. For example, when the F_0 parents are homozygous for either recessive or dominant alleles of a genetic locus, all of the gametes they produce carry the same version of the allele.

Now consider what happens in what is known as a monohybrid cross. In such a cross we are looking at a single trait, presumably controlled by which alleles are present at a particular genetic locus (gene). We assume that we have in bred lines that display either of two different versions of the trait, for example either smooth or wrinkled peas. What do the peas of their offspring look like? Turns out they are round. These F_1 plants must have one allele from each homozygous parent, and these alleles most be different (understand why?); this leads us to assume that the allele associated with the smooth peas' phenotype is dominant to the (recessive) allele associated with the wrinkled pea phenotype. We can then show that all of the F_1 organisms are genetically similar for this genetic locus. How? We can cross each plant to itself and to any of its siblings. Assuming that the number of offspring is large enough, we will find that 75% of the F_2 plants have smooth seeds, while 25% have round seeds, a 3:1 ratio.

An obvious question is are all of the round and wrinkled seeded F_2 plants similar? The answer for the wrinkled seed plants is yes, if we self-cross these plants, or cross them to another wrinkled seed plant, only wrinkled seed plants are produced in the F_3 generation. In contrast, we discover two types of round seed F_2 plants. To reveal these classes, we cross the plants to one another, or to a true breeding wrinkled seed plant. We discover that self-crosses lead to two different outcomes. One third of the smooth seed plants generate only smooth seed plants upon self-crossing, while two-third act like F_1 plants, producing both smooth and wrinkled seed plants in a 3 to 1 ratio (75% to 25%). If we cross the round seed F_2 plants to a true breeding wrinkled seed plants, we again find two different outcomes. Again, one third of the F_2 smooth seed plants generate only smooth seed plants upon self-crossing, while two-thirds of the F_2 smooth seed plants produce smooth and wrinkled seed plants in a 1:1 ratio.

We have approached these studies already knowing that peas are diploid, that genes exist in multiple forms (alleles), and that chromosomes segregate independently during meiosis. Mendel deduced all of these "laws" from his studies (rather amazingkt). Moreover, he found that the same rules held for a number of distinct traits, including plant height, green/yellow seeds, purple/white flowers, flower position, seed pod shape, and seed pod color. From this type of data, Mendel deduced that: each these phenotypic traits were controlled by 1) a single genetic element that exists in two different forms, alleles, where one is dominant to the other, recessive allele; 2) that each plant contained two copies of these genetic elements (which we call genes), and 3) that upon breeding, the parent passes one and only one of its alleles to its offspring, but which allele is passed is random ($\rightarrow$).
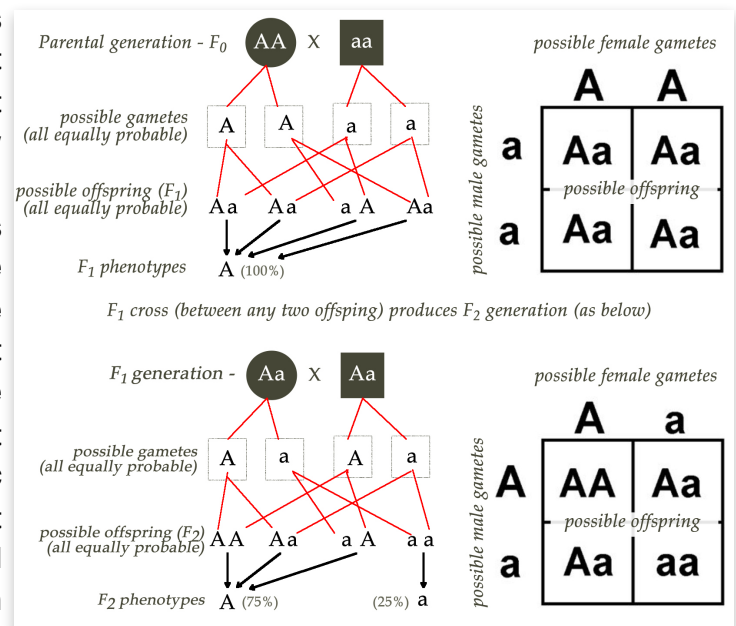
Questions to answer:

231. Why was it critical for Mendel's studies to be able to control crosses between individual plants?
232. What led Mendel to be able to discover recessive alleles?
233. Describe, in terms of meiotic behaviors, how the results of a monohybrid cross are produced.
234. Explain why, when small numbers of offspring are generated, the ratio of phenotypes in a F2 cross can differ from the expected 3:1 ratio.

Questions to answer:

- How are backcross to homozygous recessive individuals informative? Are similar backcrosses to homozygous dominant individuals useful?
- How does not determine, in practice, that a homozygous recessive individual is homozygous recessive?

Chi square analysis, hypothesis testing, and dealing with numbers that are less than infinity

One limitation of Mendel's work is associated with the limited number of plants he could examine. The various ratios he predicted are expected to be true only when the number of individuals examined becomes large, at smaller numbers of individuals, there can be serious divergences between what is observed and what is (according to the hypothesis or model being tested) predicted. This is a situation similar to one we considered previously in the case of other stochastic processes, since which gametes contain which alleles and which fuse with one another are both stochastic processes.[455] Consider the general question, how many rolls of a die would you need to perform to convince yourself that the die is fair? While the stochastic nature of meiosis and fertilization does not effect the F_1 generation of a cross between homozygous dominant and recessive plants, in which all offspring are predicted to be the same (heterozygous), it certainly influences the 3:1 ratio of phenotypically dominant to recessive plants predicted to occur in the F_2 generation. How do we evaluate whether our observations are consistent with our model, or contradict it, and so force us to abandon or substantially revised it?

The answer is a statistical test known as a χ^2 analysis.[456] Such an analysis uses the equation ($\downarrow$) together with two other concepts: degrees of freedom and null hypothesis.[457] If we are testing a model that makes a mathematically precise prediction as to the frequency of the various

$$\chi^2 = \sum_{}^{n} \frac{(\text{observed} - \text{expected})^2}{\text{expected}}$$

phenotypic classes observed, our null hypothesis is that our model is correct, that there will be no significant difference between the observed data and the predicted data. We will then try to determine whether our data are consistent with the null hypothesis. Remember, we cannot prove anything, we can only conclude that the data we observe is consistent or inconsistent with our prediction, our null hypothesis.

To define the degrees of freedom, we need to know how many independent variables there are. In our two phenotype system (wrinkled and round), we assumed that all individuals have either one or the other phenotype, if we know the number of individuals involved and the number of either phenotype, we

---

[455] It is similar to the question of which unstable isotope atom will decay next.

[456] Here is an alternative presentation from GENETICS AND GENE PROBLEMS

[457] chi square tutorial: http://www.radford.edu/rsheehy/Gen_flash/Tutorials/Chi-Square_tutorial/x2-tut.htm

automatically know the number of the other. In the case of two phenotypic classes, the degree of freedom is 1 (if there are four classes, the degree of freedom is 3, and so on). What is the degree of freedom for a six-sided die? By convention, which is currently under some discussion[458], we take an observation to be consistent with the null hypothesis if it can be expected to occur by chance at less that 1 time out of 20 (0.05); otherwise we have a good case to reject the "null" hypothesis.

For any particular experiment, we make observations to test our null hypothesis, are our predictions supported or rejected? Just for fun, let us consider here (and as a classroom assignment) Mendel's monohybrid crosses (→). The prediction of his model is that round : wrinkled seeds in the F_2 will occur in a ratio of 3:1. In his study, he reported that he examined 7324 plants. Given his model, he would have predicted that 5492 of these plants would have round seeds, while 1849 plants would have

Results of all of Mendel's monohybrid crosses

	Parental phenotype	F_1	F_2
1.	Round×wrinkled seeds	All round	5474 round; 1850 wrinkled
2.	Yellow×green seeds	All yellow	6022 yellow; 2001 green
3.	Purple×white petals	All purple	705 purple; 224 white
4.	Inflated×pinched pods	All inflated	882 inflated; 299 pinched
5.	Green×yellow pods	All green	428 green; 152 yellow
6.	Axial×terminal flowers	All axial	651 axial; 207 terminal
7.	Long×short stems	All long	787 long; 277 short

Griffiths et al., 2000

wrinkled seeds. We can now do our χ^2 calculation. We have 5474 (observed) − 5492 (expected)^2 = (-18)^2 = 324/5492 (expected) equals 0.059 and 1850 (observed)−1849 (expected)^2 = 1^2 = 1/1849 (expected) = 0.00054. The sum ($\sum$) of these two numbers = 0.0595. To determine whether these observations are consistent with our null hypothesis, we need to consult a χ^2 probability table (↓). The higher the χ^2 value the more likely the difference between observed and expected data is

Selected percentile values of the X^2 distribution

df*	.99	.95	.50	.10	.05	.01
1	.000157	.00393	.455	2.706	3.841	6.635

due to chance, rather than because our assumption, our null hypothesis is correct. Our value of 0.059 lies well below the 0.05 probability value of 3.841, suggesting the observed numbers are consistent with, but by no means proving, that our model is "true." In fact, there have been suggestions that Mendel's observed numbers are too good, to close to what would be predicted.[459] Be that as it may, Mendel's conclusions for the behavior of the types of traits he chose to study have been repeatedly verified - we can trust his general conclusions given his assumptions.

Dihybrid crosses and linkage

Now we can move to more complex questions. As an example, let us consider two distinct traits (smooth/wrinkled and yellow/green seeds), we can ask, do the alleles involved behave independently of one another or do they interaction in some way? We begin, based on a monohybrid analysis, knowing which traits are determined by recessive and dominant alleles. We can begin with a null hypothesis, namely that the two traits behave, in meiosis, independently, that is they do not interact with one another. Assume that we begin with two lines that breed true for these traits (↓). As before, each

---

[458] Statistical errors and Colquhoun. 2014. An investigation of the false discovery rate and the misinterpretation of p-values

[459] see On Fisher's Criticism of Mendel's Results With the Garden Pea

parental F_0 organism can produce only one type of gamete, and all F_1 organisms will have the same AaBb genotype (which is independent of which parent was AA and which was BB). We can then predict the outcome of a cross between F_1 individuals (F_1 x F_1). Assuming that the two genetic loci behave independently, then each F_1 individual can produce four different types of gametes, and these gametes can fuse (randomly) with gametes from the other F_1 individual. We can visualize this behavior, and the outcome of the cross, using what
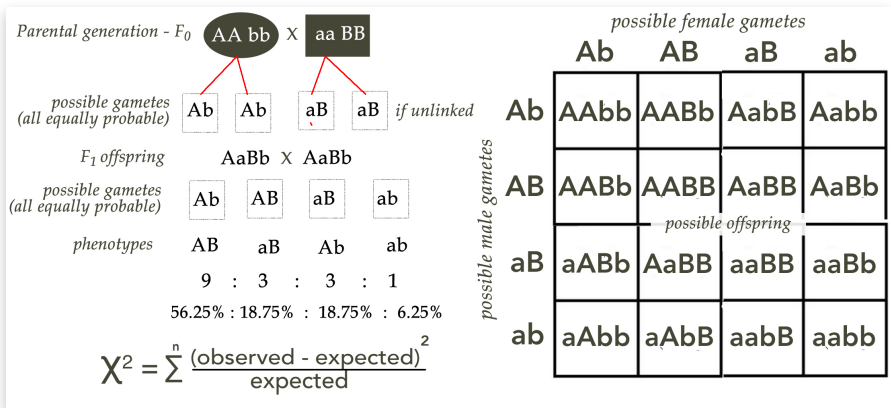
is known as a Punnett square, which enables us to determine the various possible phenotypically distinct outcomes and their relative frequencies.[460] There are 16 possible combinations of these alleles in the F_2 generation, of these 9 display a dominant:dominant phenotype: AABB (1), AABb (2), AaBb (4), AaBB (2); two display a dominant:recessive phenotype: AAbb (1), Aabb (2) or a recessive:dominant phenotype: aaBB (1), aaBb (2); and one (aabb) displays a recessive:recessive phenotype. This produces F_2 progeny in the ratio of 9:3:3:1. Test crosses to recessive:recessive organisms can be used to identify the genotypes (allele composition) of these various classes of organisms. We can, again, use a χ^2 analysis to determine whether the outcome of a particular dihybrid (two trait) cross is consistent with the hypothesis that the alleles involved do not interact with one another, that they are unlinked.

But what happens if we find that the cross produces the same phenotypic combinations BUT that numbers observed for the various phenotypic classes of the F_2 offspring do not match our expected values - what can we conclude? The simplest conclusion, and one not made by Mendel (because he excluded such traits), was that the genetic loci involve in these traits are somehow linked together, and only occasionally separated during the process of meiotic recombination.[461] Let us consider one such example, we generate a dihybrid F_2 generation from AB phenotype F_1 offspring (the result of a AB X ab cross), and observed the following outcome :

observed

AB : aB : Ab : aa

981 : 72 : 86 : 964

552 : 394 : 394 : 131

expected

We carry out a χ^2 analysis and obtain a value of 6219. A quick look at the probability table (↓) confirms our suspicion, namely that our null hypothesis, that the genes are unlinked, is rejected. We are forced to assume that the genes are linked, and we can now generate an estimate of how closely linked they are on the chromosome.

Chi Square Values and Probability

Degrees of Freedom	P = 0.99	0.95	0.80	0.50	0.20	0.05	0.01
3	0.115	0.352	1.005	2.366	4.642	7.815	11.345

[460] Who was this Punnett fellow? see Reginald Punnett

[461] Why did he missing this type of genetic behavior, because i) he did not have linked traits in his analysis or ii) because he excluded traits that behaved in this way from his analysis - I have not checked with was the actual situation.

We know from our cross that the parents (F_0) were AB and ab, and that there chromosomes were AB and ab respectively. If the A and B genes are located on the same chromosome, we can assume that, in the absence of recombination, only [AB] and [ab] gametes would be generated and that all F_1 organisms were [AB][ab], with the brackets indicating that the alleles are linked on the same chromosome. Again, in the absence of recombination, we can assume that F_1 organisms can produce only [AB] and [ab] gametes. To produce aB or Ab gametes, we must assume that a recombination event occurred between the A and B loci. To calculate the frequency at which such recombination (cross-over) events occurred, we add the number of aB and Ab organisms and divide by the total number of organisms, in our case this results in 72 + 86 / 2103 = 0.0751. This indicates a recombination frequency of ~7.5%. Recombination frequencies are typically referred to as map units or centimorgans, named in honor of the early geneticist Thomas Hunt Morgan.[462] A 7.5% recombination frequency equals 20 map units or 50 centimorgans.

unlinked genetically & physically

linked genetically & physically

unlinked genetically, linked physically
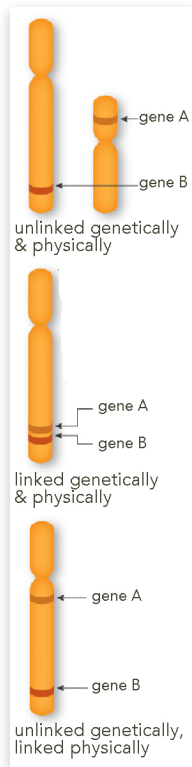
We should note that when the linkage distance exceeds 50 centimorgans (cM), the two genetic loci behave as if they are unlinked, that is, located on different chromosomes, even if they are actually located on the same chromosome (←). It is, of course, possible to walk along a chromosome using pairs, or sets of three loci. In this way, we find that a typical chromosome is more than 50 cM in length (→). Because recombination (crossing-over) can be influenced by the physical state of the chromosome, for example crossing over is often inhibited within the centromeric region of the chromosome, centimorgans do not directly or consistently convert into DNA lengths in base pairs. That said, on average (in humans) a 1 centimorgan distance between genetic loci corresponds to ~1 million basepairs of DNA, 1 megabase (appreviated Mb). From an evolutionary standpoint it is worth remembering that linkage can influence the inheritance of alleles (see above); the closer two genetic loci (and their alleles) are do one another the longer (the more generations) it will take recombination to separate them, so that they are inherited independently.

MUTANT	centimorgan	WILD TYPE
Short aristae	0	Long aristae
Black body	48.5	Gray body
Cinnabar eyes	57.5	Red eyes
Vestigial wings	65.5	Normal wings
Brown eyes	104.5	Red eyes

Using conventional genetic methods, we can extend our analysis of linkage from two to three or more genes, in order to identify the order of genes along a chromosome. If two different genes are linked to the same gene (for example, the m gene is linked to the w and the y genes) (→), they can be in various orientations with respect to one another. Genetic crosses using organisms that are originally homozygous for all three alleles, assuming that at least two forms of the alleles at each locus can be identified and that these homozygous organisms are viable, can be used to map genes with respect to one another. In this example (↑), you should be able to predict what you would expect from a cross if the w gene were located upstream or downstream of the m gene. In an era (like today) of full genomic sequence data, it

---
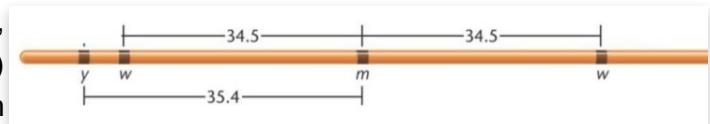
462 Thomas Hunt Morgan

is generally easier to use web based tools such as Genomicus [link], since finding the organisms needed to carry out a multigenerational cross, particularly in humans, can be challenging.

Questions to answer:

235. What does it mean if the null hypothesis is not supported?
236. A dihybrid cross produces offspring that do not fall into the expected 9:3:3:1 distribution, what kinds of conclusions can we make?
237. In a dihybrid cross, the individuals that are homozygous for both recessive alleles are absent, what might you conclude and why?
238. What might you, at least tentatively, conclude if expected individuals (from a dihybrid cross) that were heterozygous for both dominant alleles, failed to appear?
239. Alleles in two different genes appear linked to an allele in a third gene, but they do not appear to be linked to each other. What can you conclude and why?

Question to ponder:

- Do genes on opposite sides of the centromeric region of a chromosome appear closer or further away (genetically) than they are molecularly?

Using web-based bioinformatic tools: Genomicus

In Genomicus (to be illustrated in class), the user inputs a gene name (rest assured, we will talk more about gene names soon), and the system displays the gene in its genomic context (within a chromosome) as well as the genomic positions "of all its orthologous and paralogous copies in all the other sequenced metazoan genomes" together with "predicted ancestral genome structure".[463] In the example below (↓), we inputed the gene name *Lct* (OMIM: 603202), a gene than encodes the enzyme
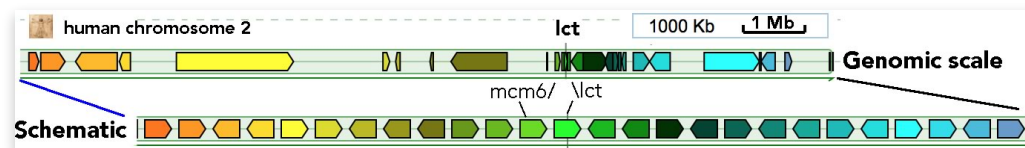
lactase, the enzyme that enables mammals to digest lactose, and so survive on their mother's milk, one of the defining traits of mammals. In most mammals, the *Lct* gene is expressed in infants and then turned off as they mature into adults. In populations of humans known to raise domesticated animals from which milk can be harvested, and so provide a significant source of energy and nutrients, we find the trait "adult lactose tolerance". Adult lactose tolerance is associated with a failure to turn off expression of the *Lct* gene in adults.[464] Molecular studies indicate that expression of the *Lct* gene in adults is negatively regulated by an enhancer element ~14 kbs upstream of the *Lct* gene, located within an intron of the *Mcm6* gene. Mutations within this enhancer element are found in populations in which adult lactase tolerance is common, apparently due to positive selection.[465] Genomicus enables us to analyze the region around the *Lct* gene. Two views are possible, in the genomic scale view, the genes are displayed based on their actual size in base pairs), relative locations, and direction of transcription, indicated by a pointed box (↑). Different genes get different colors and the direction of the box indicates the direction of RNA synthesis; here are two genes that are

---

[463] Genomicus update 2015: a genome-wide perspective to multispecies comparative genomics.

[464] Lactose digestion and the evolutionary genetics of lactose persistence

[465] World-wide distributions of lactase persistence alleles and the complex effects of recombination and selection.

transcribed in opposing directions ▭▭, hopefully you can explain how such a thing is possible. While each pointed box indicates the region of the gene, it does not show the positions of introns and exons. Intergenic regions (the regions between genes) are indicated, with their relative lengths accurately displayed. In the schematic view, each gene is again indicated by a pointed box, but all genes, no matter their actual length, are indicated by the same size box. It can be easier to recognize genes in the schematic view. On the web, holding your cursor on a gene (in either view) will display the gene name and more information about it (we will work with Genomicus in class). Note that the *Mcm6* gene is located adjacent to the *Lct* gene. We could, if we wanted to, walk along the chromosome (the *Lct* gene is located on human chromosome 2), by inputing genes at each end of the region displayed. Genomicus also presents syntenic regions in other organisms, and provides predictions of the genomic organization of evolutionary ancestors.

To use Genomicus to study evolutionary change, let us consider a gene we have already introduced, the *Gulo1* gene. Recall (*page 71*) that, in contrast to most vertebrates the *Haplorhini* or dry nose primates are dependent on the presence of vitamin C (ascorbic acid) in their diets. A plausible scenario for this situation is that a functional L-gulonolactone oxidase (*Gulo1*) gene was lost due to mutation in the last common ancestor of the *Haplorhini*. The remains of the *Gulo1* gene found in humans and other *Haplorhini* genomes is mutated and non-functional, leading to our requirement for dietary vitamin C. If we use the human genome as a reference, Genomicus fails to find the non-functional *Gulo1* gene. In contrast, if we enter *Gulo1* using the mouse or a *Strepsirrhini* (wet nose primate) genome, Genomicus finds the gene (↓). Each horizontal line in the diagram represents a segment of a chromosome from a particular species selected, together with phylogenic (evolutionary) relationships based on synteny between species. We find a *Gulo1* gene in the mouse together with orthologs in a wide range of eukaryotes, including single-celled eukaryotes such as baker's yeast, which appears to have diverged from other eukaryotes about ~1,500,000,000 years ago. Moreover, we find that
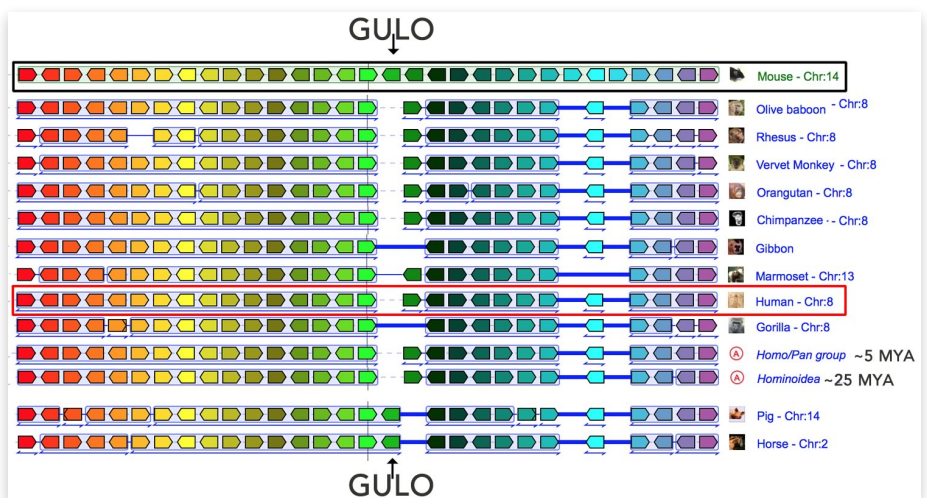
the genes surrounding the *Gulo1* locus in mammals are also (largely) the same; mammals are estimated to have shared a common ancestor ~184 Mya. The syntenic region around the *Gulo1* gene, and the presence of a *Gulo1* gene in yeast and other distantly related organisms, suggests that the ability to synthesize vitamin C is a trait that was present in the ancestor of all eukaryotes.

Humans are eukaryotes, but an examination of the resulting map reveals the absence of humans (*Homo sapiens*) and other Haplorhini primates – Whoa!!! what gives? The explanation, it turns out, is rather simple (see link). Because of mutation there is no functional *Gulo1* gene in any *Haplorhini* primate. But the *Haplorhini* are related to the rest of the mammals, aren't they? We can test this assumption (and circumvent the absence of a functional *Gulo1* gene) by exploiting synteny – when we

search for genes in the neighboring region, we find that this region, with the exception of *Gulo1*, is present and conserved in the *Haplorhini* (↑). The *Gulo1* syntenic region (without *Gulo1*) lies on human chromosome 8 (highlighted by the red box) and similar syntenic regions are found in the homologous chromosomes of other *Haplorhini* primates. Our Genomicus analysis enables us to make a number of readily testable predictions. A newly discovered *Haplorhini* primate would be predicted to share the same syntenic region and to be missing a functional *Gulo1* gene, whereas a newly discovered *Strepsirrhini* primate, or any mammal that does not require dietary ascorbic acid, should have a functional *Gulo1* gene within this syntenic region. We might also predict that adding a functional *Gulo1* gene, for example from a mouse, would make a human cell (or a human) vitamin C independent (perhaps something a future genetic engineer with do).[466] Such an analysis also indicates that gene can move around within the genome.

Questions to answer:
240. If you were to add a mouse Gulo1 gene to a a human genome, where would you put it and why?
241. If a gene is missing from a syntenic region, what might have happened to it?

Question to ponder:
- Given what you know about meiosis, how would the deletion of a gene influence the genotypes of the gametes; what about a translocation, in which part of one chromosome was moved to another chromosome?
- What would happen if the homology domains on the Y chromosome were deleted?

Genetic complementation

When we make mutations in various traditional ways (such as X-rays or exposure to mutagenic chemicals – see above), the organisms carrying these mutations can be selected for further study based on their phenotypes, typically chosen because they effect a particular process. The first aspect of such a study is to carry out various "back-crosses" in order to remove unwanted mutations; remember, mutation is random, and generally carried out so as to produce hundreds of mutations within the genome so as to insure that genes of interest are mutated. Organisms that carry mutations that influence a specific process need to have the mutations in other genes removed (through sexual reproduction) before they can be studied. The strategies involved in "cleaning up" a mutation vary between different genetic systems, and we will not considered them in detail here.[467]

A priori we do not know whether mutations (alleles) producing similar or related phenotypes generated following mutagenesis, are in the same or different genes. One way to answer this question is through genetic complementation tests. Let us assume that two (newly defined) mutant alleles influence molecular processes leading to the discernible traits. We can use dihybrid crosses to carry out a preliminary examination of the various types of interactions between these alleles. These are outlined in this table (next page ↓). As an example, consider two independently derived alleles that produce the same apparent phenotype. Let us assume that we can generate organisms that are

---

[466] [Functional rescue of vitamin C synthesis deficiency in human cells using adenoviral-based expression of murine l-gulono-γ-lactone oxidase](#)

[467] If you are interested, you can check out: [The art and design of genetic screens: Drosophila melanogaster](#) which has some interesting properties, such as the lack of meiotic recombination in males!

homozygous for these alleles (which implies that they are not homozygous lethal). If we cross these, let us call them a1/a1 and b1/b1 organisms, we expect that all of the F_1 generation will be genetically the same, at least at these loci. If the F_1 organisms exhibit a wild type phenotype, we can tentatively conclude that these alleles are located in different genetic loci (genes), and have an a1/+ b1/+ genotype. If they display a mutant phenotype, we could tentatively conclude that these are alleles of the same gene, with a a1/b1 genotype. We could seek to confirm these conclusions by asking whether the alleles are linked, although this can be difficult (or impossible) if a1/a1 and b1/b1 have similar phenotypes. We could avoid this problem if we had enough phenotypically distinct genetic markers; that would enable us to determine whether the two genes are linked to the same or different genes. If

Allelic interactions

Independent	allele a in a gene is associated with a particular phenotype allele b in a different gene is associated with different phenotype	a/b organism displays both phenotypes.
synthetic	allele a in a gene is associated with a particular phenotype allele b in a different gene is associated with different phenotype	a/b organism displays a new phenotype (such as lethality)
complementary	allele a in a gene is associated with a particular phenotype allele b in the same or a different gene is associated with the same or a different phenotype	a/b organism displays wild type phenotype
enhancement	allele a in a gene is associated with a particular phenotype allele b is in a different gene	phenotype of a/b organism is more severe than a/+
suppression	allele a in a gene is associated with a particular phenotype allele b is in a different gene	phenotype of a/b organism is less severe than a/+
epistasis	allele a in a gene is associated with a particular phenotype allele b in a different gene is associated with different phenotype	a/b organism expresses only one of the two phenotypes.

they were found to be linked to the same markers (allelic versions of other genes), we would conclude that they are alleles of the same gene, if linked to different genetic markers, then it is likely that these are alleles of different genes.

Another formal possibility is that these two alleles are in the same gene, but display what is known as intragenic complementation, that is, while the a1 and a2 alleles are both recessive, leading to a mutant phenotype as homozygotes (that is, as either a1/a1 or a2/a2) while the a1/a2 heterozygote is wild type. This type of intragenic complementation is relatively rare, since generally both allelic versions of the gene product are inactive (amorphic/null, or hypomorphic), but there are cases, particularly involving proteins composed of multiple copies of the same gene product, in which the combination of allelic polypeptides retains sufficient activity to produce a wild type phenotype. We will consider the various other types of genetic interactions, all of which can combine to various extents, through interactions with other allelic variants present throughout the genome, to modify the phenotype displayed by an allele (genetic background effects.)[468] This is one reason that research often examines an allele's phenotype(s) in a number of genetic backgrounds - crossing mutant animals with wild (by which we mean really wild) type animals. Genetic backgrounds can have substantial effects on phenotypes.[469] Given that different species (such as mice and humans) have dramatically different

---

[468] Genetic Background Limits Generalizability of Genotype-Phenotype Relationships (a paper cited above)

[469] Analysis of 589,306 genomes identifies individuals resilient to severe Mendelian childhood diseases

genetic backgrounds, it is not surprising that the same mutation (for example, a null mutation) defined in one organism can produce a different phenotype in another.[470]

Interacting traits: synthetic lethality and co-dominance

Physical linkage of genetic loci is only one way that genes interact, another involves interactions between gene products and the biological processes they mediate. Perhaps the most dramatic is known as synthetic lethality.[471] In such a situation, often but not necessarily, carried out with dominant alleles of two distinct genes, both heterozygotes, on their own, are viable, while the double heterozygote is dead, the combination is lethal (→). Similarly, it can be the case for recessive alleles, that individually homozygous organisms are viable, while double homozygous individuals die, or display a different phenotype. We can detect the presence of synthetic lethality through various crosses in which individuals with specific combinations of alleles (such as the dominant A and B alleles) fail to appear in the progeny of a cross (←). Again, as log as we can identify expected progeny phenotypes, and so count their presence in a population, such deviations from expected outcomes can be detected using a χ^2 analysis as we did previously to identify linkage.

The appearance of synthetic lethality suggests that the two gene products are involved in a common, essential process. Less extreme interaction outcomes are associated with other types of synthetic interactions between alleles of different genetic loci; these are recognized because the phenotype produced by the presence of both alleles is different from the phenotype of either allele on its own. This is different from the behavior of Mendel's genetic factors whose phenotypes are (because of Mendel's choices) independent of one another.

Synthetic phenotypes can arise in a number of different ways. As an example, a process may depend upon multiple gene products interacting to form a functional complex, necessary to produce a trait. Two, often paralogous, genes may produce functionally similar gene products. If one is mutated so as to produce little or no gene functional product (amorphic or hypomorphic alleles), the product of the second gene may be sufficient, but if both are mutant, not enough of the functional complex is present, resulting in a new version of the trait or lethality. In some cases, alleles of both genes may be recessive, but when present together, they may appear dominant. Such a situation can be generated

---

[470] Null mutations in human and mouse orthologs frequently result in different phenotypes

[471] Synthetic lethality and cancer

using various molecular methods, generating what is known as a "sensitized background" that reveals the roles of gene products in specific tissues.

Questions to answer:
242. Generate a concept map of all the plausible ways that the products of two distinct genetic loci could interact to produce a synthetic phenotype (we will probably do this in class)
243. If a gene is missing from a syntenic region, what might have happened to it?
244. How might the level of expression of one gene influence the phenotype associated with another?

Question to ponder:
- Why did Mendel exclude interacting alleles from his analysis. How did he do this?

Interacting traits: epistasis

Once mutations (alleles) that alter a particular phenotype, such as eye shape or color, limb formation, or a specific behavior have been identified, they can be used to study the underlying cellular and molecular processes involved. Our first task is to determine whether the mutations are in the same gene or different genes. Different genes are recognized by the fact that they are (generally) unlinked or genetically separable.[472] In the context of any study in which mutations are generated, it is necessary to remember there are number of possible effects on the gene product, as well as the phenotype, that can arise from a mutation – it is important to characterize the nature of the mutation, an amorphic mutation will behave differently from an anti-morphic or neomorphic mutation.

Most gene products function within networks in which particular gene products interact with each other and regulatory molecules to produce specific phenotypes. Within such a network, we can consider the types of effects that a particular mutation will have on the phenotype. As an example, let us return to the lac operon. We can generate a schematic of the interactions between genes, gene products, and regulatory molecules - in this case lactose, allolactose, and cyclic AMP (↓). Based on such a scheme, we could, if we were so motivated, generate a mathematical (graphical) model to serve as the basis for making predictions about the effects of mutations in the various genes involved in the process. If those predictions are confirmed experimentally, we have increased faith that our understanding of the system is complete; if the predictions are not confirmed, it is possible (likely) that we have missed important

---
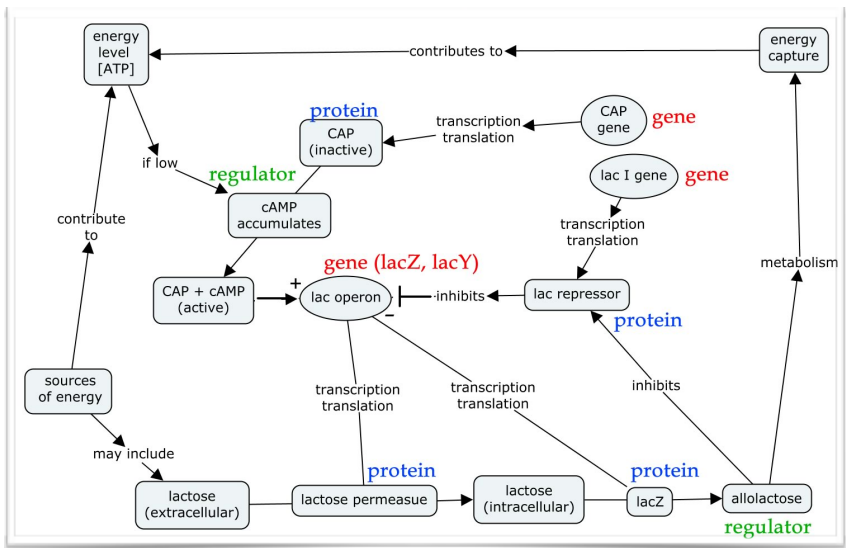
472 One point to keep in mind is that normally the process of generating mutants generates lots mutants throughout the genome, which can complicate the analysis. To remove these "background" mutations, mutated organisms that display the trait under study are crossed to wild-type animals, this is known as a backcross. Those organisms that display the trait in subsequent generations selected for further study

components of the system. We might have missed a gene/gene product that influences the behavior of the system. At the same time, while DNA-dependent, RNA polymerase is a necessary component of the system, required to expressed the genes involved, it is not explicitly included in our model because mutations that alter polymerase function would be expected to disrupt many (essentially all) systems within a cell or an organism, and produce complicating phenotypes. These are known as pleiotropic phenotypes.[473] Similarly, if any of the components of the system we include are involved in other processes, the model may be influenced by effects on those systems and processes.

In a number of systems, there are often parts of the network that are linear, or perhaps best termed sequential, with one gene product acting on another, "down-stream" aspect of the system. An example is the testosterone/estradiol system; both testosterone and estradiol are derived from cholesterol and both play key roles in the generation of male and female sexual characteristics in mammals. If we begin with cholesterol (ignoring the pathway of reactions involved in cholesterol synthesis), we find a number of gene products, identified by their On-line Mendelian Inheritance in Man (OMIM) designations, that catalyze the various steps in this pathway (↓), reactions that occur in both the cytoplasmic and
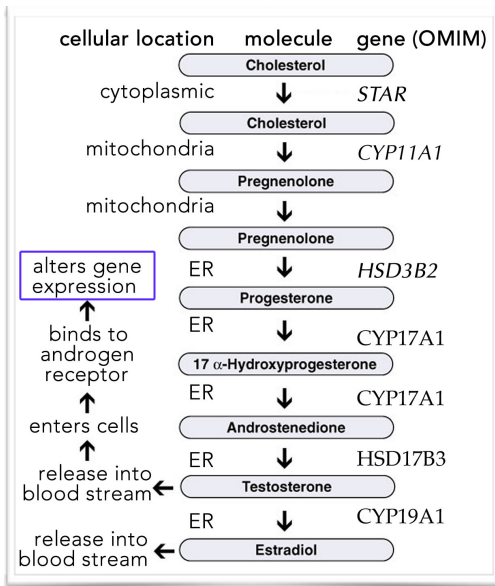
mitochondrial compartments of the cell. Entry of cytoplasmic cholesterol into mitochondria is facilitated by the STAR gene product; within mitochondria, an enzyme (a gene product) catalyzes the chemical reaction that transforms cholesterol into pregnenolone, which then leaves the mitochondria and accumulates in the endoplasmic reticulum (ER). A series of reactions then leads to the formation of testosterone, the "male" hormone, which can then be transformed into estradiol, a "female" hormone, which is also involved in male reproductive function.[474] Both testosterone and estradiol are released into the blood stream, allowing them to interact with cytoplasmic receptor proteins (androgen/estrogen receptors) in various cell types. Testosterone and estradiol act as allosteric effectors of these transcription factor proteins, activating them to enter the nucleus and regulate the expression of specific target genes.

In the context of such a pathway analysis, we find that the effects of mutations/alleles of genes can be ordered. For example, assume that there is a mutation in the CYP17A1 gene which leads to a non-functional (amorphic or null) version of the encoded protein. In an individual homozygous for this CYP17A1 mutation, we would expect to see the accumulation of progesterone in the ER. Now consider a second null mutation in the CYP11A1 gene, an individual that is homozygous for this mutation would be expected to accumulate cholesterol in mitochondria. So, you should now be able to predict the phenotype, in molecular terms, of an organism homozygous for null alleles in both CYP17A1 and CYP11A1 genes, as well as predicting the phenotype resulting from a genetic cross between CYP17A1 and CYP11A1 homozygous individuals, assuming of course that both are viable and fertile. The result of such a genetic analysis allows us to establish what is known as the epistatic relationship between

---

[473] Pleiotropy: One Gene Can Affect Multiple Traits

[474] see The role of estradiol in male reproductive function

genes (or more accurately gene products) in a particular process.[475]

A complicating aspect of most actual interaction pathways is that there are various forms of feed-back and feed-forward interactions that can influence the behavior of a pathway when its normal functioning is inhibited or perturbed. As an example, the accumulation of one compound might influence the expression of other genes, or the activity of other enzymes. In some cases, this can result in a by-pass of the block, so that phenotypic effects are minimized. Consider the cholesterol to testosterone/estradiol pathway - both testosterone and estradiol influence gene expression by serving as allosteric effectors of transcription factors; just as their presence can activate or inhibit the expression of genes, their absence can activate or inhibit the expression of a range of genes. At this point, what is important is to consider what the phenotypes of various genetic crosses might tell you about underlying molecular and cellular systems, while recognizing the limitations of such predictions.

Temperature sensitive alleles

A final type of mutation (allele) is known as a temperature-sensitive mutation / allele. In the case that a gene encodes a polypeptide, changing the amino acid sequence of that polypeptide can influence how the polypeptide chain folds, as well as its stability as a function of temperature. In some cases, the polypeptide (or protein) can be more sensitive to its surroundings. A mutant protein may no longer behave normally when the temperature is reduced (cold-sensitive) or increased (heat-sensitive). This underscores the fact that each organism typically has an optimal growth temperature; as part of its evolutionary adaptation, its polypeptides/proteins are optimally functional at that temperature, and are relatively less functional at higher temperatures, where they may denature, or lower temperatures, where they may adopt non-functional configurations.

Questions to answer:
245. Under what conditions might self-fertilization be selected for or selected against?
246. What factors limit the usefulness of genetic crosses to establish epigenetic relationships?
247. How are genetic pathway maps useful, and what are there limitations?
248. Why is a forward genetic screen unlikely to identify all components of a particular process?
249. Consider a dominant allele in which the associated phenotype is lost on a particular genetic background. How might you reveal the presence of such an allele through a genetic analysis?

Measuring evolution's impact on allele frequencies: Hardy-Weinberg

If we consider a population, each gene is represented by some set of alleles; these occur at various frequencies in different genes. To determine whether evolution is occurring within a population, we use what is known as the Hardy-Weinberg (H-W) equation, based on the work of G.H. Hardy (1877-1947) and Wilhelm Weinberg (1862-1937) – published independently in 1908. Their analysis was based on the assumption that evolutionary processes were <u>not</u> occurring within a population, they assumed that: 1) the population was infinite, so that processes such as genetic drift do not occur; 2) the population is isolated, so that no individuals leave or enter; 3) that no new mutations occur; 4) that mating between individuals is random (no sexual selection); and 5) there are no differential reproductive effects, that is,

---

[475] Epistasis — the essential role of gene interactions in the structure and evolution of genetic systems

natural selection is not occurring.[476] Under these conditions, the allele frequencies found in the initial population do not change over time. If, on the other hand, allele frequencies are found to change, selection (or some other process) must be occurring.

Before Hardy-Weinberg analysis there was a belief that dominant alleles were somehow "stronger" than recessive alleles, that "dominant alleles must, over time, inevitably swamp recessive alleles out of existence. This incorrect assumption was called "genophagy", literally "gene eating""[477], but this is not the case unless the alleles influence reproductive success, that is, unless positive or negative selection is occurring.

So let us consider the situation in which there are two alleles (A and a) of a particular gene; if the frequency of A in the population is p, the frequency of a equals q. It is clear (hopefully) that $p + q = 1$. We can then calculate the frequency of homozygotes and heterozygotes by expanding the term $(p+q)^2$; simple mathematical considerations indicate that within this population, the probability of an AA homozygote is p^2, the probability of an aa homozygote is q^2, and the probability of an Aa heterozygote is $2pq$, such that

$$p^2 + 2pq + q^2 = 1.$$

How is this possible? remember, both p and q are less than 1. Our null hypothesis is that these alleles are NOT subject to natural selection, which means that they have no effect on reproductive success within the population. Now we can look at the frequency of recessive homozygotes in a population and calculate the χ^2 value and use it to estimate whether the population is at equilibrium, that is, no evolutionary changes are occurring, or whether there is active selection for or against certain alleles. For example, it might be that homozygous recessive individuals are either not viable, they die, or they are not fertile. Alternatively, the heterozygote might have a reproductive advantage compared to the recessive homozygote; such a heterozygote reproductive advantage can maintain significant levels of an allele that is deleterious as a homozygote within a population. If allele frequencies change over time, one of the assumptions of the model must be wrong - the most obvious is that genotype-based differential reproduction effects (natural selection) are active.

Tracking the fate of mutations (becoming an allele or disappearing from the population)

At this point, you might well ask yourself, given the effectiveness of natural selection, why do alleles that produce severe diseases occur or persist at all? There are a number of possible scenarios that the previous discussion should help you consider. One is that new mutations are continuously arising, either in the germ line of the organism's parents or early in the development of the organism itself, and that these alleles (mutations) disappear from the population with the death of the organism. The prevalence of the disease will then reflect the rate at which such pathogenic mutations arise *de novo* together with the rate at which the individuals carrying them are eliminated (before they have off-spring). The second, more complex reason involves the fact that in diploid organisms there are two copies of each gene and that carrying a single functional copy of a disease-associated allele might have no discernible effect on the organism's reproductive success – that is, the allele is recessive. If we remember that whether an allele is recessive or dominant depends upon the phenotypic trait being

---

[476] Hardy-Weinberg Equilibrium: http://www.tiem.utk.edu/~gross/bioed/bealsmodules/hardy-weinberg.html

[477] genophagy

considered. As noted above, it is possible that the heterozygotic state conveys a reproductive advantage, that is, the allele has both a dominant (positive) and a recessive (negative) phenotype. In this case, the heterozygote will be subject to positive selection (leading to an increase in allele frequency), while the homozygote will be subject to negative selection (leading to a decrease in allele frequency); this can be sufficient to maintain the allele in the population at a significant levels. Similarly the effects of a dominant allele associated with a pathological condition can be be ameliorated, or even beneficial in the presence of various genetic modifiers (enhancers or suppressors)(we will go into more detail below). Eventually the population will reach a point where negative and positive effects balance. This is better considered a "steady state" than an equilibrium, since selection is active, but positive and negative effects the final balance. Of course this steady state is sensitive to changes in the environment that influence phenotype and their effects on reproductive success. If we were more mathematical, one could model the system based on such effects.

The pace of selective effects depends upon population size and the strength of selective (both positive and negative) effects. As selection acts, and the population's allele frequencies change, the degree to which a particular trait influences reproductive success can also change. The effects of selection are not static, but evolve over time. For example, a trait that is beneficial when rare may be less beneficial when common, and competition between individuals that express the trait increases. New mutations that appear in the same or different genes can influence the trait and selective effects, leading to changes in the population over time. The example of the evolution of the ability to utilize citrate (described above) appeared in a population pre-disposed to such a change.

Questions to answer:
250. Consider conditions in which the deletion of a gene might lead to a selective advantage.
251. How might you determine whether the appearance of an allele in a population is due to a new mutation, as opposed to some other mechanism (or is there no other way?)
252. How can combinations of alleles in different genes lead to new traits?

Questions to ponder:
- Do genomes always become more complex over (evolutionary) time? Why might they become simpler?
- Are there broader implications arising from the maintenance of deleterious alleles within a population?

Chapter 16: Germ line alleles and human pathologies

We consider inherited alleles, their effects on embryonic development and their roles in human disorders and diseases. We consider their patterns of inheritance and techniques used to identify when and where a gene is expressed. We consider how to identify the genes involved in more complex phenotypic traits, traits influenced by allelic variation at tens to hundreds of genes.

Up to this point, we have been considering genes, mutations, and alleles from the perspective of distinct traits. We now focus on how mutations, and the alleles they can become, influence human health, leading to genetic dispositions and diseases. We will begin by a short review of embryonic development, although critical details are clearly beyond us here. We focus first on germ line alleles, present in maternal or paternal parents that influence gametes and early developmental

"In human genetics, we try to avoid referring to patients as "mutants," even when it is fully justified scientifically; the word carries unfortunate cultural connotations." D. Botstein. Decoding the language of genetics. 2016. CSH press.

processes. In the next chapter we will consider how these inherited alleles can interact with mutations that occur during the process of development (cell division and differentiation). This is a hugely complex topic, so our intent is to identify core concepts rather than specific molecular and cellular processes.

Developing multicellular organisms: from egg to embryo and more

Complex multicellular organisms undergo a process known as embryonic development.[478] Development begins with the fusion of a haploid sperm and egg, produced through meiosis, to form a diploid zygote that then divides (by mitosis) to produce a multicellular embryo that develops into an adult. Egg and sperm differ in their composition. For example in vertebrates, zygotic mitochondria are supplied in the egg only. Similarly, genomic imprinting (p. 250) means that in the zygote, only the egg-derived or the sperm-derived alleles are expressed. Following the formation of the zygote, cell division leads to a multicellular organism.[479]

While the fertilized egg is totipotent - that is, it can generate all of the cells found in the adult, the cells formed during development become more and more restricted with respect to the types of progeny that they can produce – they become committed to one or another specific fate. In part this fate restriction is due to the fact that as cells divide, different cells come to have different neighbors, and so they experience different environments and different combinations of signals from other cells, leading to the expression of different genes and different cellular behaviors. Differences in gene expression can

---

[478] Multicellularity: The Evolution of Differentiation

[479] This process of multicellularity is described in the supplemental chapter.

be used to identify cells, as in this figure in which the cells of an 8 day after fertilization mouse embryo are characterized (in the mouse birth occurs ~21 days post zygote formation, that is fertilization)(→). As development proceeds, cells in various regions of the embryo behave differently from one another, they differentiate into various types of cells, such as neurons, muscle cells, epithelial (surface) cells, etc.[480] The process of differentiation is associated with, and driven by various internal asymmetries, signals from neighboring cells, and differences in which genes are expressed in the various cells.
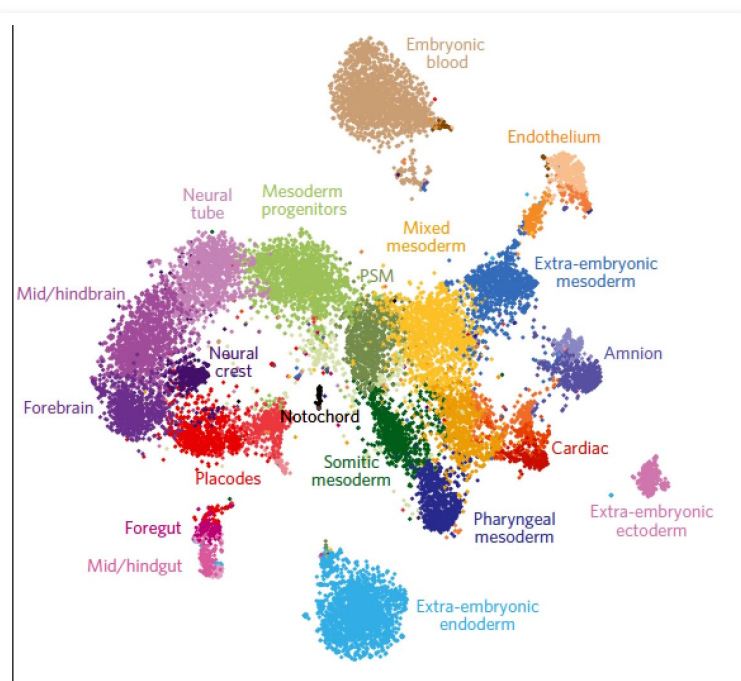
Gene expression analysis identified 20 general cell types (in 33 different clusters, indicated by the different colors) in the 8 day mouse embryo. Clusters was identifed based on the expression of marker genes. Several cell types are composed of two or more clusters. adapted from Ibarra-Soria et al., 2018

The process of development is complex, and somewhat different in different organisms, leading to the different morphologies of different species and different individuals within a single species. Development is influenced by the gene regulatory systems active, their regulation, together with stochastic effects and environmental influences. As an example, excessive exposure of the human embryo to ethyl alcohol leads to a developmental defect known as fetal alcohol syndrome (FAS), associated with a range of effects and defects, including irreversible brain damage and a number of growth problems, including some minor malformations of facial structures (↓).[481] The extent of the effects of fetal alcohol exposure are also influenced by the genotypes of the mother and the developing embryo, in particular by genes involved in the metabolism of ethanol.[482] In a similar way, the recent outbreak of Zika virus has been associated with developmental defects, specifically microcephaly, a drastic disruption of brain growth, associated multiple functional (cognitive) defects.[483] Again, the severity of these defects is likely to vary based on the timing of infection in terms of embryonic development, together with the genotype of the mother, particularly in terms of her immune response to the virus, and the genotype of embryo, in terms of its susceptibility to perturbation.

---
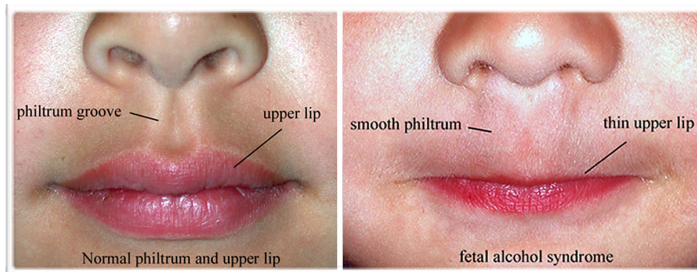
[480] Defining murine organogenesis at single-cell resolution reveals a role for the leukotriene pathway in regulating blood progenitor formation

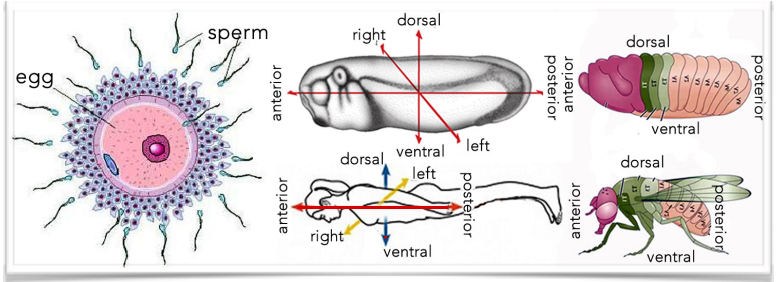[481] The effects of alcohol on fetal development.

[482] Genetic and epigenetic insights into fetal alcohol spectrum disorders

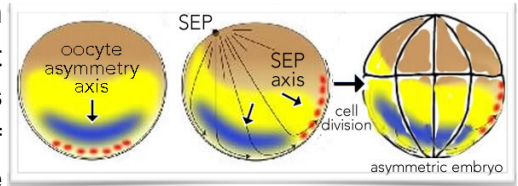[483] Zika Virus in the Americas — Yet Another Arbovirus Threat

Establishing embryonic axes

During the process of embryonic development, the various axes of the organism, typically anterior–posterior, dorsal–ventral, and left–right in vertebrates form (↓). While the details of embryonic development beyond our scope here, there are two basic, and interacting, processes – asymmetries and inductive interactions – that drive the formation of embryonic axes, the changing shape of the embryo, from fertilized egg to adult, and the associated cellular differentiation and tissue formation. The mechanisms involved reflect evolutionary pressures, for example, when an egg is fertilized externally (outside of the mother), and without subsequent parental protection, the new organisms are highly vulnerable to predators and their development is normally quite rapid. The eggs are often large and contain all of the nutrients required for development to proceed up to the point where the new organism can feed on its own. To facilitate rapid development, the egg is pre-organized, that is, it is highly asymmetric, with specific factors that can influence gene expression, either directly or indirectly, positioned in various regions of the egg (↓). As an example, in the frog *Xenopus laevis*, the sperm entry point (SEP) establishes a new asymmetry axis that can lead to the reorganization of the cytoplasm before the first embryonic cell division. The oocyte's asymmetrical structure is transformed into an asymmetrical egg through the process of meiosis and the cellular events associated with it. The asymmetric cytoplasmic determinants include various RNAs and proteins, RNAs. The various asymmetries in the fertilized egg are stabilized by the rapid cycles of DNA replication and mitotic cell division, with growth based on the utilization of maternally supplied nutrients. As distinct cells are formed, they become different from one another as they inherit different determinants; the presence of these determinants leads to changes in gene expression and cell differentiation (see above).

On the other hand, in a number of organisms, and specifically mammals, embryonic development occurs within the mother, so there is no compelling need to stockpile nutrients within the egg and the rate of development is (generally) dramatically slower than that seen in externally developing embryos. In such developmental systems, it is often not the asymmetries associated with the oocyte and fertilized egg that are critical, but rather the asymmetries that arise during embryonic development. As the zygote divides, a major factor that drives differentiation is whether a cell comes to lie on the surface of the embryo or within the interior (↓). In mammals, the cells on the exterior form the trophectoderm, which goes on to form
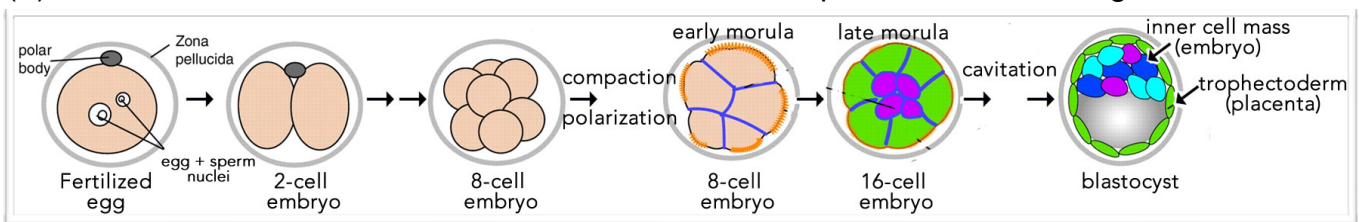
extraembryonic tissues, in particular the membranous tissues that surround the embryo and

become part of the placenta, the interface between the embryo and the mother. Cells within the interior form the inner cell mass that produces the embryo proper. Changes in gene expression will lead to changes in the ability to produce and respond to inductive signals, which will in turn influence cell behavior and gene expression. Through this process, the cells of the inner cell mass come to form the various tissues and organs of the organism. It is easy to tell a muscle cell from a neuron from a bone cell from a skin cell by the set of genes they express, the proteins they contain, their shapes (morphology), their internal organization, and their behaviors.

Genetic analysis of developmental processes: maternal and zygotic effect mutations

Embryonic development, like any other process or trait, can be studied and underlying mechanisms identified through the generation and analysis of mutations in the genes that influence the processes involved. From a genetic perspective, there are two general types of mutations (alleles) - there are those that effect the formation of gametes, particularly the egg, and those that effect the process of embryonic development directly. Mutations (alleles) that influence oocyte formation, and then embryonic development are known as "maternal effect mutations". They can be recognized based on their behavior in crosses. Take for example a recessive effect allele "a" - it may be a typical zygotic effect allele or a "maternal effect" allele. Let us consider how they can be distinguished. Let us assume that the mother is homozygous for a. For a typical zygotic effect allele, in cross to a wild type (paternal) homozygote (↓), they behave as expected, with all offspring displaying a wild type phenotype. A cross between two of the resulting heterozygotes (↓) produces all wild-type offspring. But let us consider a maternal effect allele. When we cross the homozygous mother to a male of any genotype we find that all offspring are mutant (↓). We recognize maternal effect mutations by their non-Mendelian behavior (circled genotypes). Similarly, but not shown here, a dominant maternal effect allele will, when crossed to a male of any genotype, produce all mutant offspring.
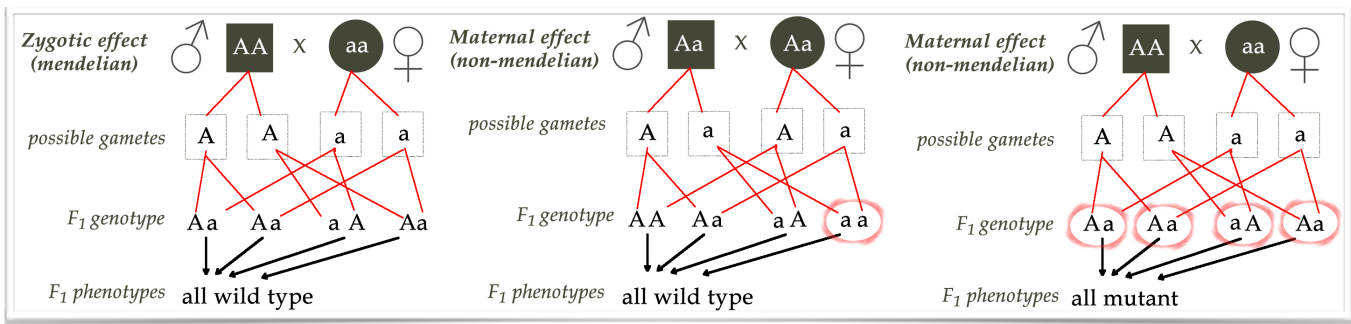
Mitochondrial inheritance

Eukaryotic cells (such as our cells) have one or more type of intracellular organelle, either mitochondria (all eukaryotes) or mitochondria and chloroplasts (in algae and plants). These organelles have their own genomes, circular DNA molecules known as mtDNAs. A number of genes are encoded by the mtDNA: 37 in human. Also, the mtDNA can, like any DNA molecule, accumulate mutations when it is replicated or in response to free radicals generated during the course of aerobic respiration (something that we will not consider further).

One aspect of typical sexual reproduction is that only the mitochondria of the oocyte are inherited by the fertilized egg; the mitochondria present in the sperm cell, either do not enter the egg or if they do, they and their DNA is destroyed – degraded in various ways, by activated endonucleases and other processes. Mutations in mitochondrial DNA can lead to dysfunctional mitochondria, which can have a number of phenotypes.[484] The genetics of these mutations are often non-Mendelian and include maternal effects.

One complexity in the study of mitochondrial DNA mutations is that each mitochondrion contains a DNA molecule, and the cell contains many mitochondria (hundreds to a few thousand); different cell types within the same organisms can contain different numbers of mitochondria and differ in their dependence on mitochondrial function. The result is that we are looking a population of mitochondria, with a number of different mitochondrial genotypes. Moreover, the numbers of mitochondria can change, raising the possibility of population bottlenecks and associated changes in genotype, which raises the possibility of somatic selection. In any one cell or tissue, mitochondrial dependent phenotypes will reflect, and be influenced by the multiple mitochondrial DNA genotypes present – that is, the percentage of mutant (dysfunctional) to wild type (functional) genotypes. A detailed consideration of mitochondrial influences on disease phenotypes in humans and other organisms is beyond us here, but the interested can find a database of mitochondrial DNA mutations at the MitoMap web site.[485]

Questions to answer:
253. Schematically consider mechanisms that would lead to differential gene expression in the various regions of an embryos - how is that even possible?
254. Describe how imprinting can impact Mendelian allele behavior(s)?
255. Most of the genes involved in mitochondrial function are nuclear; how might that influence the phenotypes of mutations in mitochondrial DNA?
256. If you were to predict which tissues would be more severely effected by mutations in mitochondrial DNA, what would you base your predications on?

Questions to ponder:
- What has to happen to change the events or timing of early developmental events?
- Explain the evolutionary pressures egg and sperm behavior and the speed of early development.

Traits and the number of genes involved

Mutations that become alleles (enter the germ line) can be seen as lying along a continuum. At one end of this continuum are alleles that behave as do the alleles that Mendel used; these are alleles of a gene that control what we might term discrete features of a particular trait, such as pea color or in the case of humans blood type, or a number of genetic diseases that you either have or you do not have (↓ left side). As the number of genes (and the alleles) that influence a particular trait, the distribution of versions of the trait, say for example, height, approaches a smooth curve, a curve often termed a bell curve (right side ↓). Such a distribution is characterized by a mean, a median (which is the same as the

modified from
Tobin/Dusheck, Asking About Life, 2/e
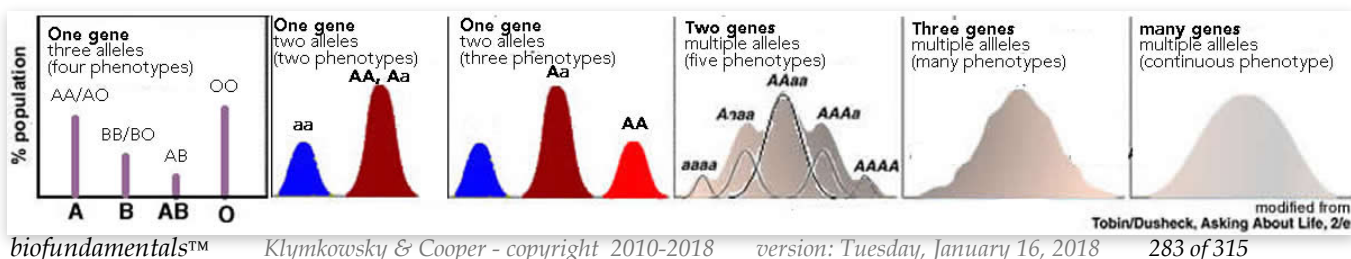
mean when the curve is symmetrical), and a standard deviation, which reflects the width of the distribution. The alleles in the various genes involved in a trait can display dominant, recessive, or what we synergistic (interactive) behaviors.

An important feature of germ line alleles is that all cells of the resulting organism (with the exception of the gametes produced by that organism and any new somatic mutation - last chapter) will have the same genotype (↓). That said, the phenotypes associated with a particular allele can vary between different regions of the organism, different tissues, organs, and organ systems. Genes that encode common, often termed house-keeping functions, generally have global effects, while those expressed in only one or a few cell types may have effects in only these cells. The fact that many genes have been duplicated during evolution, to form paralogous genes, which often have similar (although rarely identical) functions can also influence the phenotypes associated with various alleles. A gene may be expressed in a particular cell type, but the behavior of the gene product may be more or less critical in those cells because of the presence of functionally complementary gene products (both due to expression of a paralogous gene, or genes in various compensatory or parallel molecular processes and pathways. We will see this effect in our discussion of somatic mutations; a germ line mutation can be inherited but not have a discernible phenotypic effect. However, if a subsequent somatic mutation occurs that disables the functioning copy of the gene, or compromises the function of a complementary gene, a phenotype can arise. Such events are involved in some heritable cancer susceptibilities (see below).
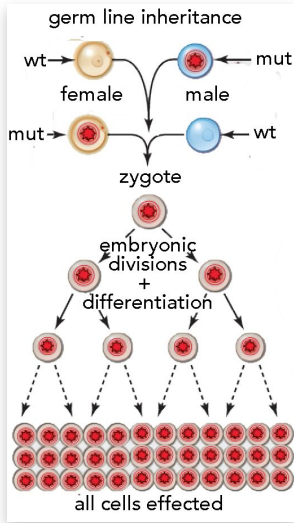
Where is a gene expressed? RT-PCR based systems

The following discussion might encourage you to ask, exactly how do we determine where and when specific genes are expressed within an organism? There are a number of applicable mechanisms that fall into two basic types - there are those that detect transcribed gene products (RNAs) and those that detect the polypeptide encoded by an RNA. We consider them briefly here.

A transformative technology, made feasible by the discovery of heat stable DNA-dependent, DNA polymerases, isolated from archaea that live in very high temperature environments (thermophiles and hyperthermophiles), polymerase chain reaction (PCR) has been a powerful technique in manipulating genes, and visualizing gene expression. In the context of gene expression analysis, we can use PCR to quantify the amount of a particular transcribed (expressed) RNA within a particular tissue, cell type, or together with single cell isolation technology, a single cell. This process involves first making a DNA copy of the transcribed RNA, so that we avoid the genomic DNA copies of the genes present in every cell. We isolate RNA from a tissue and then use the virally-derived reverse transcriptase (RT) enzyme; RT uses a DNA primer and makes a DNA copy complementary to the RNA strand, a cDNA (→). The two strands are then separated (typically by increasing the temperature of the system), and then a second DNA primer is added together with a thermostable DNA-dependent, DNA polymerase in order to generate a copy of the cDNA. Now we begin the amplification stage of the reaction. The two strands are separated by increasing system temperature. The two DNA primers are present in excess,

so that when the temperature is reduced, they bind back to the DNA strands, and initiate a new round of DNA-dependent, DNA synthesis. With each cycle the number of DNA strands doubles, a exponential growth in DNA molecule numbers with each cycle. Because the primer sequences, which are synthesized in vitro, are complementary to, and specific for, a particular gene sequences, one can amplify one and only one of the RNAs (gene products) present in the tissue under analysis. If the gene is not expressed, no amplified DNA will be synthesized. By using various tricks (beyond us here, but relatively simple to employ with the right equipment) the process can be made quantitative, so that it is possible to compare the numbers of a particular type of RNA molecule (the products of a particular gene) present in the original sample, a measure of the levels of gene expression, at least at the RNA level. With different sets of primers, it is possible to quantify the various splice forms of a gene expressed.

    More recently, it has become possible to isolate and sequence the RNAs (or rather cDNAs derived and amplified from mRNAs) in even a single cell and to then sequence those DNA molecules to characterize the genes expressed in that cell.[486] Because mRNA is used, only exon sequences are included - and the result is known as an exome sequence. This is a method that can be particularly useful in characterizing the genes expressed in a particular cell type, or in a cancer.[487]
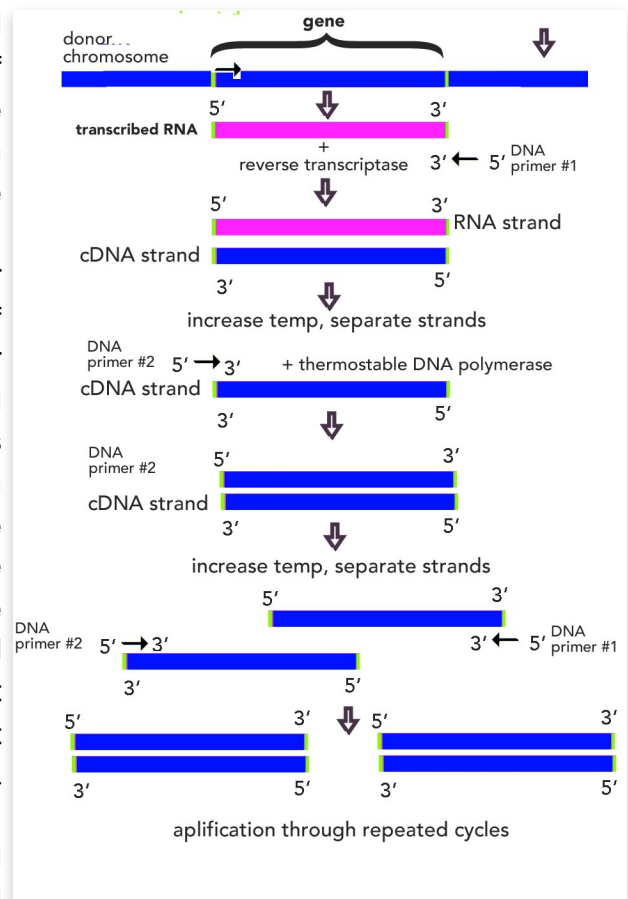
In situ hybridization: A limitation of the RT-PCR approach is that it generally works on tissue samples, which contain multiple different types of cells. To achieve spatial resolution, we need to use other methods. Perhaps the most common is known as in situ hybridization. When a gene is expressed, an RNA molecule complementary to one strand of the gene is synthesized, and these sense RNAs accumulate in the cells that express the gene (there is little evidence for significant transport of RNA from cell to cell, across the plasma membrane.)[488] To identify cells that express a gene, we generate modified "anti-sense" RNA molecules. Typically, we first isolate and subclone a DNA molecule that encodes the sense (mRNA) and antisense RNA of a gene's expressed (exonic) region – this can be based on a cDNA generated from an mRNA or a genomic exon. Using specific primers, recognized by different bacteriophage-derived DNA-dependent, RNA polymerases, we can generate either sense or anti-sense RNA molecules. In these reactions modified (with either fluroscein or digoxygenin) forms of

---

[486] A practical guide to single-cell RNA-sequencing for biomedical research and clinical applications

[487] see Defining murine organogenesis at single-cell resolution reveals a role for the leukotriene pathway in regulating blood progenitor formation

[488] although things may actually be somewhat more complex: see Brain Cells Share Information With Virus-Like Capsules

the RNA nucleotide UTP are used; this modified nucleotide can be used by the polymerase and is incorporated into the newly synthesized RNA (→).

The over all process is relatively simple. The tissue is chemically stabilized, and then incubated with either sense or anti-sense probe. Because of the complementary nature of nucleic acids, the anti-sense probe RNA will bind to RNA transcripts, generated during the gene's expression. In contrast, the sense probe is the same sequence as the RNA transcript, and so does not bind - it is used as a control, since (generally) such a sense RNA probe is not complementary to any mRNA present. By controlling the hybridization temperature, we can remove low affinity, non-specific interactions, leaving only the high affinity sense (transcript)-anti-sense complexes. The probe will be retained in regions that
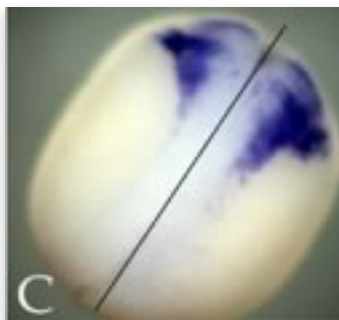
express the gene, and washed away from regions where the gene is not expressed (the level of binding to genomic sequence is too low to be visible). Antibodies, conjugated with various enzymes (typically alkaline phosphotase or horse radish peroxidase) can then be used to recognize the modified probe RNA:mRNA complex, and color-generating reactions, catalyzed by the enzyme, allow the distribution of probe to be visualized. The example here (←) is a neurula stage *Xenopus* (clawed frog) embryo in which a gene (Snail2/Slug) that is expressed in the neural crest has been visualized.[489] In situ hybridization can provide single cell resolution, distinguishing cells that do, from those that do not, express a particular gene. The specificity of the technique is influenced by the length of the probe and the hybridization temperatures used.
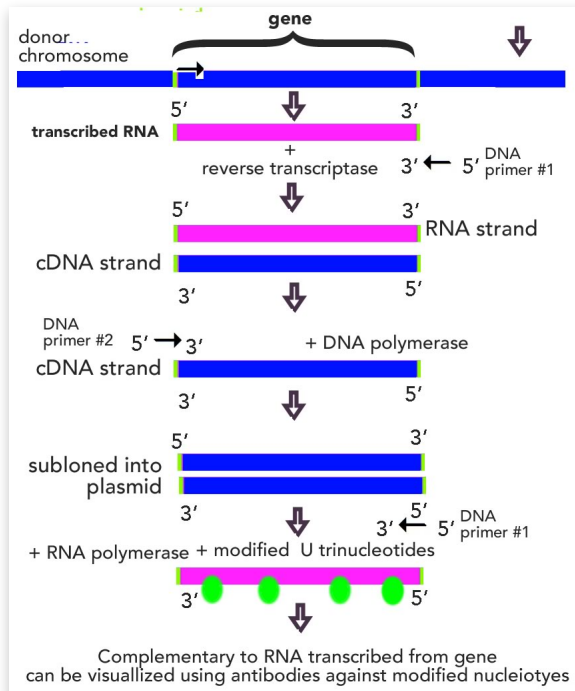
Immunocytochemistry: One limitation of RT-PCR and in situ hybridization methods is that they monitor RNA levels. In cases where the ultimate gene product is a polypeptide, it can be the case that RNA levels are not strictly correlated with level of the accumulated polypeptide. One approach to avoid this disconnect is to use antibodies, proteins generated by the vertebrate immune system that can bind specifically to specific molecular targets. We will ignore how antibodies are generated, but basically the act very much like in situ probes, binding to specific molecular (protein) targets. A full characterization of the proteins present in a cell or tissue relies on physicochemical approaches, such as mass spectrometry, to define the proteome (a subject beyond us here).[490]

---

[489] from: An NF-κB and Slug Regulatory Loop Active in Early Vertebrate Mesoderm

[490] Here is an example of proteomic analysis: Region and cell-type resolved quantitative proteomic map of the human heart

Questions to answer:

257. How can observed variation in a trait be used to develop a model for the number of genes involved in determining the trait. How might you test your model?
258. A gene can be spliced various ways - design primer sets to distinguish the splice variants of a gene.
259. Explain why a sense strand RNA probe serves as a useful control for in situ hybridization studies; what does it control for, and why does it work?

Questions to ponder:

- Why might the number of polypeptides in a cell differ from the number of RNAs that encode it?

Back to Mendelian determinants

Returning to the effects of various alleles, we will begin with discrete traits in humans that behave in a strict Mendelian manner. Perhaps the best known is blood type (see above) which is determined by three different alleles at a single locus. These alleles behave in a dominant A to O, and B to O, and a co-dominant (A with B) manner. The distribution of these alleles in different human populations appears to be due, at least in part to selective advantages associated with specific alleles in specific environments. For example, "*Mourant suggested that the major differences in the geographical distribution of ABO blood groups may be the consequence of epidemics that occurred in the past. The concept of evolutionary selection based on pathogen-driven blood group changes is currently supported by studies on the genetic characterization of the ABO blood group in Neanderthals and ancient Egyptian mummies. These studies suggest a potential selective advantage of the O allele influencing the susceptibility to several different pathogens responsible for diseases such as severe malaria, H. pylori infections and severe forms of cholera*".[491]

There are three common alleles that control blood type in humans, A, B, and O. Because blood type can be determined unambiguously, the mode of interaction of these alleles is well defined, it is possible to trace their inheritance across multiple generations. If we know an individuals blood type, we have an initial (although incomplete) model of their genotype. As we examine the phenotypes of their progeny, we can further constrain their genotypes. In such studies, we assume that we know with certainty who mated with whom, something that may or may not be a realistic assumption. For example, the presence of an AB individual in the second generation (→), indicates that the male parent had to have an AB or BO genotype, other genotypes could not have been produced by the parental (A X B) cross. Similarly in the lineage giving rise to the O individual, we can conclude that its male parent had to be BO, while its female parent had to be OO. The more of the individual phenotypes we know in a pedigree, the more we can constrain the genotypes of members of the lineage.
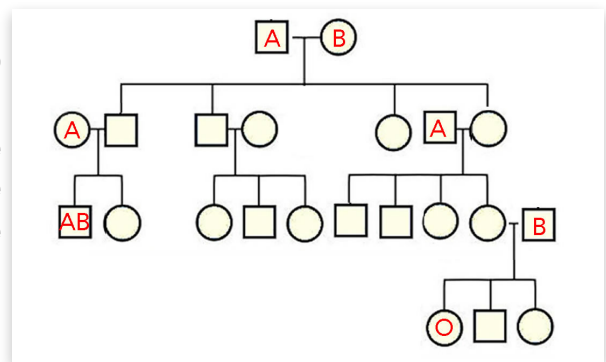
It is also worth noting that in the modern world, we use molecular markers to identify the alleles present in specific individuals. One issue with such pedigree analysis is that it can lead to potentially embarrassing or disruptive conclusions, for example revealing that a father cannot be the genetic father of a child (generally, but not always, who the mother of a child is is more unambiguous).

---

[491] Beyond immunohaematology: the role of the ABO blood group in human diseases

Disease-associated alleles

There are a number of identified genetic disorders with clear Mendelian inheritance (see Specific Genetic Disorders). What does this mean? Basically that the alleles associated with the disease act in a simple dominant or recessive manner. In the case of dominant disease-associated alleles, to be inherited means that they are not lethal as homozygotes, and result in fertile individuals, otherwise they could not pass the allele on to the next generation. Recessive alleles can be lethal in the homozygous state (as might be dominant alleles), but heterozygotes much survive and be able to reproduce. One point to keep in might is that the terms recessive or dominant are always in reference to specific phenotypic traits. An allele can be recessive with respect to one phenotype and dominant with respect to another. The classic example of such behavior are mutations associated with the hemoglobin B (HBB) gene of humans (→ and below). Alleles of this gene are associated with a dominant trait, resistance to malarial infection, as well as a homozygous (often lethal) trait, sickle cell anemia. While the recessive trait is subject to strong negative

Gene-Phenotype Relationships				
Location	Phenotype	Phenotype MIM number	Inheritance	Phenotype mapping key
11p15.4	Delta-beta thalassemia	141749	AD	3
	Erythremias, beta-			3
	Heinz body anemias, beta-	140700	AD	3
	Hereditary persistence of fetal hemoglobin	141749	AD	3
	Methemoglobinemias, beta-			3
	Sickle cell anemia	603903	AR	3
	Thalassemia-beta, dominant inclusion-body	603902		3
	Thalassemias, beta-	613985		3
	{Malaria, resistance to}	611162		3

selection, the dominant trait is subject to strong positive selection in environments where malaria is endemic. The same allele is responsible for both traits.

Concordance between monozygotic twins and genetic influence on a trait

An interesting phenomenon that can be used to characterize the genetic contribution to a trait involves twins. There are two generic types of twins. Fraternal twins involves two eggs, and two sperm, with the two result that two distinct embryos develop within the mother simultaneously, and are generally born at in rapid succession. Fraternal twins are no more or less closely related than are two siblings born years apart. Fraternal twins are also termed dizygotic twins, since they involve two distinct zygotes. In animals that typically have multiple offspring (litters), all of the individuals born arise from distinct zygotes. In contrast, identical twins are known as monozygotic twins. Identical twins occur because a single sperm fertilizes a single egg, and generates a single zygote, which then begins development. During development, for one reason or another, the embryo fragments into two distinct embryos, which then develop independently of one another. So, with the exception of (somatic) mutations that may have occurred during embryonic development, the two are genetically (genotypically) identical. This genetic identity enables us to measure the genetic concordance of a trait.[492] For example, if a trait is totally dependent upon a shared allele, together with a shared genetic (genomic) background, the concordance is 100%. In other cases, while genotype plays a role, it is not completely determinative. As an example, in the auto-immune muscle weakness disease myathenia

---

[492] Does Higher Concordance in Monozygotic Twins Than in Dizygotic Twins Suggest a Genetic Component?

gravis, the genetic concordance is ~35%, a level of genetic concordance that implies other factors play an important role in the appearance and progression of the disease.[493]

As we are talking about twins, it is also worth noting another type of outcome, which is known as a chimera.[494] In a chimeric embryo, two embryos fuse into one - such that a single organism develops, but it has two distinct "sibling" genotypes.[495] When this dizygotic fusion is complete, a single normal, albeit mosaic, embryo and mature organism is generated. When fusion is incomplete, or occurs at a later developmental stage, incompletely fused embryos are formed - what are known as conjoined twins.

Using web-based bioinformatic tools: Exac Browser

When studying a disease that appears to have a genetic component, it is common to identify the particular allele(s) involved. In the case of recessive alleles, such studies often involve pedigree analysis of more or less in bred families. Once a disease-associated allele is identified, it can be important to determine whether that allele is found in individuals who do not display the disease trait. Particularly for dominant alleles, the presence of an allele without the disease phenotype indicates the influence of genetic background effects that influence the disease allele's penetrance and expressivity. Over the last decade, there has been in increasing number of human genome or exon sequences; the exome is all of the DNA sequences, the exons, that make it into mature RNA, and even more specifically into mRNA. Most genomic DNA is not transcribed into RNA, which makes generating exomic sequences easier and less expensive - less DNA to sequence.

The accumulating library of exomic sequence data now includes more than 60,000 people from around the globe. This data library can be searched using the ExAC Browser.[496] To search the ExAC database, the user (you, for example), inputs a gene's official name, as listed in OMIM or GenBank. ExAC then displays sequence data from 60,706 (as of July 2017) unrelated individuals; this allows for the identification of alleles and mutations present in a range of human populations. Let us try using the gene associated with sickle cell anemia, the HBB gene (hemoglobin, beta, OMIM: 141900). Mutations (disease-associated alleles) in HBB have been associated with a number of human diseases (see above ↑). The allele associated with the sickle cell phenotype involves a missense mutation from GLU to VAL, now known as GLU7VAL (↓). We discover that within the ExAC database of "normal", that is

Variant	Chrom	Position	Consequence	Filter	Annotation	Flags	Allele Count	Allele Number	Number of Homozygotes	Allele Frequency
11:5248232 T / A (rs77121243)	11	5248232	p.Glu7Val	PASS	missense		532	121340	1	0.004384
11:5248233 C / T	11	5248233	p.Glu7Lys	PASS	missense		149	121340	0	0.001228

disease free individuals, this allele occurs with a frequency of ~0.0044 (with a single homozygous individual identified). The heterozygotic individuals would not be expected to display any overt

[493] Immunopathogenesis in myathenia gravis and neuromyelitic optica.

[494] It is even possible to generate chimeric embryos between different species: Humanized mice and porcinized people.

[495] Such human chimeras have been identified: see 3 Human Chimeras That Already Exist and One Person, Two Sets of DNA: The Strange Case of the Human Chimera

[496] Genomics, Big Data, and Medicine Seminar Series – Daniel MacArthur

phenotype under most conditions, while the homozygous individual would be expected to have sickle cell disease. The vast majority of the people with the HBB Glu7Val allele are of African descent, as is the one homozygous individual (→). At the time of this writing (January 2018) there was only one other homozygous individual within the library (Glu122Gln). 71 out of 85 of the people carrying this allele are of African descent, as is the homozygous individual.

Population Frequencies

Population	Allele Count	Allele Number	Number of Homozygotes	Allele Frequency
African	505	10404	1	0.04854
Latino	12	11548	0	0.001039
South Asian	9	16512	0	0.0005451
European (Non-Finnish)	6	66734	0	8.991e-05
East Asian	0	8620	0	0
European (Finnish)	0	6614	0	0
Other	0	908	0	0
Total	**532**	**121340**	**1**	**0.004384**

Data from ExAC enables us to make informed guesses as to the impact of various genetic differences on the activity of a gene product.[497] If, for example, a dominant allele has been linked to a disease and yet that allele is detected in the ExAC database, we might suggest either that that allele is not the cause of the disease, or that the effects of the allele are influenced by variation (alleles) in other genes, leading to reduced penetrance and/or expressivity. If an allele is present in a heterozygous condition, but not a homozygous one, we can tentatively assume that negative selection is acting on the allele. If, on the other hand, alleles are present at different frequencies in different populations, that may be evidence for the action of positive selection dependent on environmental factors. In addition, the frequency of alleles in different populations often reflects the effects of founder effects, bottlenecks, and drift. Take for example three other HBB alleles, p.Gly70Ser, p.Glu122Gln, and p.Gln40Ter (Ter=stop)(↓). We see that the Gly70Ser and Glu40Ter alleles are present primarily in non-Finnish Europeans, while the

Population Frequencies

Population	Allele Count	Allele Number	Number of Homozygotes	Allele Frequency
Other	1	908	0	0.001101
European (Non-Finnish)	48	66736	0	0.0007193
Latino	2	11556	0	0.0001731
African	0	10404	0	0
East Asian	0	8624	0	0
European (Finnish)	0	6614	0	0
South Asian	0	16512	0	0
Total	**51**	**121354**	**0**	**0.0004203**

Population Frequencies

Population	Allele Count	Allele Number	Number of Homozygotes	Allele Frequency
South Asian	71	16512	1	0.0043
Other	2	908	0	0.002203
Latino	3	11570	0	0.0002593
European (Non-Finnish)	9	66740	0	0.0001349
African	0	10406	0	0
East Asian	0	8636	0	0
European (Finnish)	0	6612	0	0
Total	**85**	**121384**	**1**	**0.0007003**

Population Frequencies

Population	Allele Count	Allele Number	Number of Homozygotes	Allele Frequency
Other	1	908	0	0.001101
European (Non-Finnish)	48	66736	0	0.0007193
Latino	2	11556	0	0.0001731
African	0	10404	0	0
East Asian	0	8624	0	0
European (Finnish)	0	6614	0	0
South Asian	0	16512	0	0
Total	**51**	**121354**	**0**	**0.0004203**

Glu122Gln allele is found in South Asians. It is not clear exactly what the effects of such missense mutations will be on functions of the polypeptide – it could change folding, change interactions with other polypeptides and molecules, add or remove sites of post-translational modification, or change catalytic activity, if the polypeptide has such an activity, but clearly the non-sense mutation will produce a short 39 amino acid polypeptide, compared to the 147 amino acid long full length polypeptide. It is unlikely that such a severely truncated protein is functional, but if it accumulates it could interfere with the function or molecular interactions of the full length polypeptide.

Using web-based bioinformatic tools: BLAST

There are other web based tools to identify evolutionarily conserved regions in related gene products. Perhaps the most useful is BLAST. It enables you to take either a nucleotide or a polypeptide

---

497 The ExAC browser: displaying reference data information from over 60 000 exomes.

sequence and search for similar sequences throughout all sequenced genes that have been deposited in a central repository (GenBank). The program returns similar sequences in other organisms. The presence of such sequences can be best explained through either evolutionary relationships (inherited from a common ancestor) or horizontal gene transfer. The BLAST tool is also useful for identifying those parts of nucleic acid or polypeptide sequences that are conserved, that is, that vary the least from organism to organism – we might well expect such regions to be particularly sensitive to mutational change. The absence of allelic (missense/non-sense) variants (in ExAC) in such regions would argue for the action of positive selection.

Questions to answer:
260. Outline your strategy to determine whether someone is not telling you the truth about parentage, given a family tree and a simple dominant recessive trait.
261. You find a frequent allele in a population, but no individuals homozygous for that allele - how might you make sense of that observation?
262. Why aren't missense mutations necessarily loss of function mutations?
263. Looking at two populations, you find a particular allele to be much more common in one than the other - what processes and historic events could explain such an observation?

Questions to ponder:
- Provide a model for why an individual homozygous for the Glu7Val allele not have sickle cell disease?

Genetic anticipation

There is a type of inherited allele that differs in interesting ways from conventional alleles, these are alleles that change from generation to generation, a behavior that has been termed genetic anticipation (see pp. 175-176). Such alleles are associated with what are known as "trinucleotide repeat" (TNR) expansion diseases, although some involve sequences longer than repeating triplets, and are known as microsatellite expansion mutations. Such repeated microsatellite sequences (3 to 6 repeating units) account for ~30% of human genome sequence. Nucleotide repeat expansion diseases include several forms of mental retardation, Huntington's disease, inherited ataxias, and muscular dystrophies.[498] Within the genes involved, there are regions of repeating nucleotides. Because of the slippage of the DNA polymerase during the process of DNA replication, the number of such repeats can grow bigger or smaller. The result? the allele delivered to an offspring can be more deleterious than the allele present in the parent - over generations, the symptoms of such an allele grow more and more severe. The length of the repeat correlates with the age of disease onset, but the age of onset is variable between individuals with the same repeat length, suggesting the impact of various genetic modifiers. In addition to standard inheritance, many of these genes play roles in the function of nervous tissue, and it is possible that somatic (as opposed to germ line mutations) can influence the allele associated phenotype. As an example, there is evidence that genetic anticipation is important in the context of schizophrenia and bipolar disorder, which together occur in ~1% of the population and have an estimated ~80% heritability risk, which means that on average, about 80% of the differences between individual organisms is due to genetic factors.

---

[498] A Brief History of Triplet Repeat Diseases

Mechanisms: Given the number of sites in which nucleotide repeats are found, and where their expansion can lead to disease (→) implies a number of possible mechanisms behind the pathogenic state. First, all of the pathology-associated nucleotide expansion regions appear to occur within the transcribed region of the gene, and that includes the 5' and 3' untranslated regions, as well as within introns and exons. For example, if
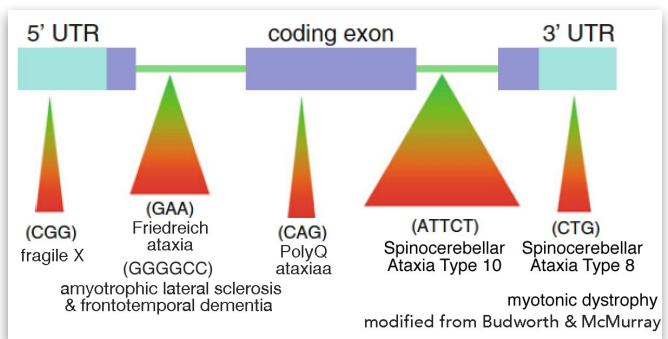
such a domain occurs in a coding region they can lead to increased stretches of repeating amino acids in a polypeptide. Alternatively, they may reflect toxic interactions between the transcribed RNA and other cellular components. To illustrate the potential complexity (a full exploration is clearly beyond our scope here), consider recent work on the role of a nucleotide expansion domain in the gene C9ORF72 (OMIM: 614620), which encodes a polypeptide implicated in vesicle trafficking within the cell. Expansion domain effects within a region of the C9ORF72 gene have been linked to both amyotrophic lateral sclerosis (ALS) and frontotemporal dementia (FTD). Studies indicate that the expanded nucleotide region is targeted for inappropriate transcription; RNAs are synthesized bidirectionally from both DNA strands (sense and anti-sense) and "that RAN (repeat-associated non-ATG translation) translation [499] occurs from both sense and antisense expansion transcripts, resulting in the expression of six RAN proteins (antisense: Pro-Arg, Pro-Ala, Gly-Pro; and sense: Gly-Ala, Gly-Arg, Gly-Pro). These proteins accumulate in cytoplasmic aggregates in affected brain regions".[500] Interestingly, another gene product, encoded for by the Supt4H1 gene (OMIM: 603555) appears to play a role in the inappropriate transcription of the C9ORF72 gene; reducing the levels of the Supt4H1gene product ameliorates the phenotypic effects of nucleotide expansion in C9ORF72 (complex and weird, huh?).[501] Rest assured, the exact mechanisms of these types of alleles and associated phenotypes are complex, but based on the effects of altered transcription on the functional roles of specific cell types.

Genome-wide Association Studies (GWAS)

The majority of phenotypic traits are not associated with simple Mendelian inheritance, rather a number of different genetic loci (genes) and the combination of alleles present determines the genetic aspect of the trait. In addition, there are non-genetic, that is environmental factors involved. How much nutrition an organism gets when developing, the effects of pathogens and such, combine to influence the final phenotype. A classic example of a trait influenced by both genetics and environment is height, because it is what is known as a quantitative trait – we characterize it by a simple number (although in fact, posture can influence our measurement).[502] The estimates for the heritability of height are not all that accurate (and differs between populations), ranging from between ~60 to ~80% of the variation

---

[499] Non-ATG–initiated translation directed by microsatellite expansions

[500] RAN proteins and RNA foci from antisense transcripts in C9ORF72 ALS and frontotemporal dementia.

[501] Spt4 selectively regulates the expression of C9orf72 sense and antisense mutant transcripts

[502] How much of human height is genetic and how much is due to nutrition?

attributed to genetic differences and ~20 to ~40% environmental (nutritional) factors. In addition, height (in humans) is a sexually dimorphic trait - on average males are taller than females.

So how, if many genes are involved, do we identify those genes involved in a particular trait?[503] We begin with a trait that can be accurately measured. In this regard, height is better than friendliness (for example). Then we need a method to identify the various differences found between different organisms (people in this case). Typically between 500,000 to 1 million single nucleotide polymorphisms (SNPs) are used. A useful SNP occurs at high frequency (>10 to 30%) in the population - it does not need to be located within a particular gene, but with a high enough density of SNPs, it will be near specific genes, and inherited with the gene (allele). Of course meiotic recombination can influence who is linked to whom.

The different SNPs present are identified based on nucleotide complementarity (not unlike the basic process behind in situ hybridization (see above). Samples of the person's genome is taken (often from white blood cells, which have nuclei and DNA, in humans - in contrast to enucleated red blood cells). Since alleles and SNPs differ in their nucleotide sequences, two perfectly complementary (single-stranded) DNA molecules bind more strongly to one another than two mis-matched molecules. We can use this difference in binding stability to identify which SNP or allele is present at a particular position. Finally, we ask how the presence of particular SNPs/alleles relates to the level of the trait, for example the height of the person or the levels of LDL and HDL (low and high density lipoproteins) in their blood. Of course you see some of the issues right away. People are different heights at different times of their lives, and different levels of LDL and HDL depending on their diet, and when they last ate. So the trait we are trying to study has to be accurately and reproducibly measureable.

We than ask which markers (SNPs or alleles) are found in correlation with the trait phenotype (height, LDL/HDL levels, etc.). With a large enough population of people (genotypes and phenotypes) we can identify those markers (alleles and SNPs) that are in or near specific genes and are associated with the phenotype in question. However correlation does not imply (or better put prove) causation. It may be that the allele/SNP is simply linked on the chromosome to the actually functionally significant allele. This is one reason that it is important that there has been time (generations) to separate, by meiotic recombination, one allele from another. To prove that a particular allele plays a functionally significant role in producing or modifying a trait, further experimental studies are necessary.[504]

Questions to answer:
264. In the case of genetic anticipation, what is the impact if the repeat domain gets shorter?
265. How might the synthesis of small polypeptides influence normal cell behavior?
266. How would a repeat domain influence a coding region?
267. What is critical before one can even consider beginning a GWAS study?

Questions to ponder:
- You discover a gene linked to a particular trait through a GWAS study, how might you go about establishing a significant physiological role for the gene in influencing that trait?

---

[503] Chapter 11: Genome-Wide Association Studies

[504] The interplay of common, rare variation in autism

Chapter 17: Genome dynamics and pathogenic somatic mutations

In which we consider how genes move around within the genome, through the action of transposable elements of various types, and how that can influence phenotype. We consider how mutations arising in somatic cells can interact with inherited alleles or on their own to impact the development of the nervous and other systems, including the origination and progression of cancer.

Up to this point, we have been considering mutations that have become alleles and that are inherited from one's parents. We have considered the shuffling of alleles through meiosis and the formation of a new diploid organism from haploid gametes. Now we introduce the reality of mutations that occur during the development of the organism. First let us reiterate, an inherited allele is present in all cells of the developing and adult organism. With the exception of processes such as X-inactivation and monoallelic expression (p. 252), it can be expected to have its effects on all tissues in which it is expressed. In contrast, when a mutation occurs within a somatic cell, it is passed on as part of a clone, through asexual reproduction. When during the development of the organism the mutation occurs will determine what percentage of the cells in the organism carry the mutation. Of course, if the mutation leads to a lethal phenotype, the cell that carries it will die, so no cells in the organism will carry the mutation. More often, such mutations are not lethal, but many influence the rate and outcomes of cell divisions, a number have been found to influence the development of the organism, and its various tissues and organs.

Normally, when and where a cell divides is under strict regulatory control, involving both internal regulatory networks, as well as signals from other cells. Another class of somatic mutations underlie the appearance of cancer cells. Cancer itself is a complex process, often involving a number of steps, a number of somatic mutations within a particular clone (cellular lineage); a complete study of cancer is beyond us here, but certain common features of carcinogenesis and progression are worth considering - the most important is to recognize that (as we noted previously) a multicellular organism is a social system. Cells are expected to cooperate in defined ways to keep the society functioning smoothly. Somatic mutations serve to disrupt that coordination. In particular, somatic mutations can lead to cells ignoring various signals meant to control their growth and behavior. As each somatic cell is clonally related to its ancestors and progeny, these clonal populations can (in the absence of appropriate regulation) compete with each other in destructive ways, at least destructive to the goals, survival and reproduction, of the organism as a whole.

Rates and effects of somatic mutation

The rates at which mutations occur within a particular cell type is based on the number of rounds of DNA replication, the error rate associated with that process, and the efficiency of DNA error repair. In mammals, most germ line errors are associated with males, because there are more cell divisions giving rise to sperm than are involved in generating eggs in the female germ line. DNA error rates differ (apparently) between species (you might speculate on why). In the mouse the current estimate is of ~5 x 10^{-9} per base pair per generation. The number is estimated to be higher in humans, closer to 1.2 x 10^{-8} per base pair per generation. A number of studies have been carried out to determine the mutation rate in somatic cells in the two species (mouse and human); it appears that these rates are higher in the soma than in the germ line.[505]

To think clearly about the effects of a particular somatic mutation, we have to think in some context. How does the new mutation interact with the pre-existing genome. For example, if you inherit a amorphic or hypomorphic allele of a particular gene, that allele may well act in a recessive manner. But what happens if, in a somatic cell, the other (wild type) allele is mutated to an amorphic state, and assume that that cell is capable of, and called upon within the context of the organism, to divide. If the new mutation does not produce a lethal phenotype, the cell will divide to form a somatic clone, and each daughter cell will carry the mutation. If this occurs in the cells that give rise to the brain, it can lead to dysfunction in specific regions, and neurological symptoms. As an example, autism (along a spectrum of severity) is common, occurring in ~1% of the population. Both germ line alleles and somatic mutations have been implicated.[506] Autism appears to share genetic risk factors with schizophrenia and bipolar disease.[507] The occurrence of a somatic mutation, in the context of a "sensitized genetic background", that is in the presence of an allele that influences neural development, can lead to defects in neuronal development. Even as few as 10% of cells that carry the somatic mutation can lead to a neuronal pathology.[508]

In the context of non-neuronal cells, the effects of somatic mutations can lead to the loss of growth control, and subsequent over-proliferation - the formation of a tumor, both benign (non-malignant) and malignant. While the steps in the formation of a cancer can be complex, and reflect a number of regulatory pathways - what is clear is that once a somatic mutation has occurred, it can establish a clone that continues to divide. The mutation turns the well behaved somatic cell into a social cheater (see chapter 4). Subsequent mutations can then accumulate that enable the cancer clone to get adequate nutrients and avoid host responses. The evolution of the cancer clone is, however, futile - at best from the clone's perspective, it will continue to divide and grow, but in the end such growth is incompatible with the survival of the host, both the clone and the host will die of the disease, the cancer.

---

[505] see Differences between germline and somatic mutation rates in humans and mice

[506] The interplay of common, rare variation in autism

[507] Genetic overlap between autism, schizophrenia and bipolar disorder

[508] see Somatic Mutation, Genomic Variation, and Neurological Disease

There are a number of ways that genes can be mutated to lead to cancer, and a number of ways such somatic mutations can interact with inherited alleles, the details complex and beyond our scope here.[509]

Nevertheless, we will consider one type of allele/somatic mutation combination, that is involved in retinoblastoma, a cancer of the retina. There is a trait that we will call "susceptibility to retinoblastoma"; it is conferred by the presence of a dominant, loss of function (amorphic) allele in the RB1 gene, let us call this allele Rb⁻. In those that inherit the Rb⁻ allele have ~90% chance of developing retinoblastoma early in childhood, ~10% will not – they will be "silent" carriers.[510] Inheriting a single copy of the Rb⁻ allele is not, by itself sufficient to lead retinal cells to become cancerous, a second, somatic mutation needs to occur in order to inactivate the wild type copy of the RB1 gene. But having the Rb⁻ allele dramatically increases the probability that when such a mutation occurs, cancer will result. We see this effect because people (children) homozygous for the Rb⁻ allele typically develop multiple tumors in each of the two eyes; these tumors appear early in childhood. The presence of the Rb⁻ allele leads to a hereditary disposition toward developing retinoblastoma (→).

But people who do not inherit this allele can also get retinoblastoma; the difference is that they have to accumulate two separate mutations, which is a much rarer (improbable) event – you might consider how much rarer. Such rare events do occur, but they tend to occur later in development, so it would be very unlikely that to occur cells whose decedents come to be present in both eyes. When sporadic forms of retinoblastoma appear, they are almost always restricted to one tumor in one eye, and appear in older individuals. A similar pattern of inheritance is associated with breast cancer susceptibility gene 1 (BRCA1).[511]
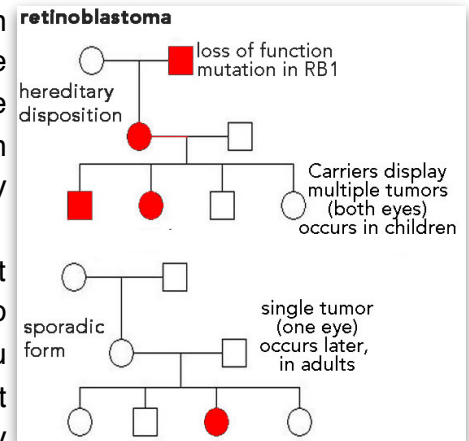
Non-disjunction: a disease of aberrant chromosome segregation meiosis

There is one more genetic disorder that we will consider, but only briefly, namely non-disjunction. Non-disjunction refers to the situation where there is a failure of normal chromosome segregation. In the case of somatic (mitotic) cell division, one daughter cell may receive two copies of a chromosome, while the other daughter receives none. This can lead to lethality or differential reproduction (somatic evolution) within the two resulting clones.

In the germ line, non-disjunction can lead to a gamete containing extra copies of one or more chromosomes, a situation known as chromosomal aneuploidy. Given that each chromosome, even the smallest ones, contain hundreds of genes, the presence (or absence) of the correct number of chromosomes leads to changes in patterns of gene expression. Generally, when a chromosomal aneuploidy occurs, the resulting embryo fails to complete normal development; recent studies indicate

---

[509] Neomorphic mutations create therapeutic challenges in cancer

[510] Genetics of Retinoblastoma.

[511] BRCA1 and BRCA2: Cancer Risk and Genetic Testing

that chromosomal abnormalities are surprisingly common in humans.[512] For example, when a human embryo carries three copies of one of the smaller human chromosomes, chromosome 21 (the basis for Down Syndrome), it is estimated that ~80% of such embryos perish *in utero* or in the neonatal period.[513] In cases where the early embryo is mosaic for chromosomal abnormalities, somatic evolution in which euploid blastomeres (embryonic cells) replace aneuploid cells appears to lead to normal embryos (and people!!!!).

Questions to answer:
268. A somatic mutation occurs early in development, what factors will influence the % of cells in the organism over time that carry the mutation?
269. How would you characterize dominant from recessive disease susceptibility alleles?
270. How does exposure to mutagens lead to increased risk of cancer development?
271. What types of molecular defects would lead to chromosomal aneuploidy?
272. How might having three (or one) copy of a chromosome influence normal cell behavior (and gene expression)?
273. In the context of the Rb⁻ allele, how might loss of the chromosome or chromosomal region in which Rb resides influence cellular phenotypes?

Questions to ponder:
- Why doesn't inheriting a cancer associated Rb⁻ or BRCA1 allele not lead to increase risk of cancer in other (all) tissues?
- Under what conditions would a somatic mutation become an inheritable allele?
- How would a mutation in a checkpoint gene influence a somatic cell's clonal evolution?

Genome dynamics

Aside from the insertion of "external" DNA through horizontal gene transfer, something that is rare in eukaryotes, and abnormal meiotic recombination events (see below), we might assume that the genome itself, is static. It is, however, becoming increasingly clear that genomes are more dynamic than previously thought. For example, consider the number of new mutations (single nucleotide polymorphisms, and small insertions and deletions and such) that arise in each generation. The frequency of such events can be estimated based on the number of times a DNA molecule has been replicated in the course of developing from a fertilized egg to the formation of its own gametes – about 400 replication events in a human male, fewer in a female – together with the error rate of DNA replication ($\sim 1 \times 10^{-10}$ per nucleotide per division.) Since each diploid cell contains $\sim 6 \times 10^9$ nucleotides, we can expect ~1 new mutation for every two rounds of DNA replication. It has been estimated that, compared with the chromosomes our parents supplied us, we each have between 60 to 100 new mutations in the chromosomes found in our germ line. Given that less than ~5% of our DNA encodes gene products, that is polypeptides and functional RNAs, only of few of these new mutations are likely to influence a gene coding region or its expression.[514] Even when they occur in a gene's coding region,

---

[512] Chaos in the embryo

[513] Morris et al. 1999.: Fetal loss in Down syndrome pregnancies. Prenat Diagn. 19: 142-145.

[514] It is, admittedly more more difficult to estimate the percentage of the genome involved in the regulation of gene expression, since these regions are harder to recognize that coding regions (can you guess why?) Also, there is an increase recognition that there is regulated transcription of regions that do not encode polypeptides (or at least longer polypeptides). Such long non-coding RNAs (lncRNAs) have been found to regulatory roles. see: Long noncoding RNAs: past, present, and future

the redundancy of codons means that many SNPs lead to what are termed synonymous mutations, which (generally) do not lead to functionally significant alterations in the gene products. That said, even apparently "neutral" mutations can lead to changes in genotype that can have effects on phenotype, and so evolutionary impacts. For example, they might influence the regulatory region of a gene.[515] As we have already discussed, in small populations genetic drift can influence whether new alleles (with non-lethal effects) are retained in the population.

In addition to the point mutations that arise from mistakes in DNA replication, a whole other type of genomic variation has been uncovered in the course of genome sequencing studies, these include the movements of transposable elements, discussed below. These are known as "structural variants." They include flipping of the orientation of a DNA region (an inversion) and sequence insertions or deletions, known as copy number variations.[516] It has been estimated that each person contains about 2000 "structural variants".[517] Large chromosomal inversions or the movements of regions of DNA molecules between chromosomes can have effects on chromosome pairing during meiosis, and can lead to hybrid sterility and inviability. You can work out for yourself what might happen if recombination events occur in such regions. The mechanisms that lead to these genomic changes can be complex, and largely beyond our scope here.[518]

An important point with all types of new genetic variants is that if they occur in the soma, that is in cells that do not give rise to the haploid cells (gametes) involved in reproduction, they will be lost when the host organism dies. Moreover, if a mutation disrupts an essential function, the affected cell will die, to be replaced by surrounding normal cells, a version of somatic selection (see above). Finally, as we have discussed before, multicellular organisms are social systems. Mutations, such as those that give rise to cancer, can be seen as cheating the evolutionary (cooperative) bargain that multicellular organisms are based on. It is often the case that organisms have both internal (cellular) and social (organismic) policing systems. Mutant cells often actively kill themselves (through apoptosis) or in organisms with an immune system, they can be actively identified and killed.

Gene duplications and deletions

While meiotic alignment generally occurs accurately, there are times were mis-alignment happens. For example, what happens if there are repeated sequences within a chromosomal region. If the homologous chromosomes misalign (→), crossing over can lead to haploid cells that emerge from meiosis with either gene duplications or

---
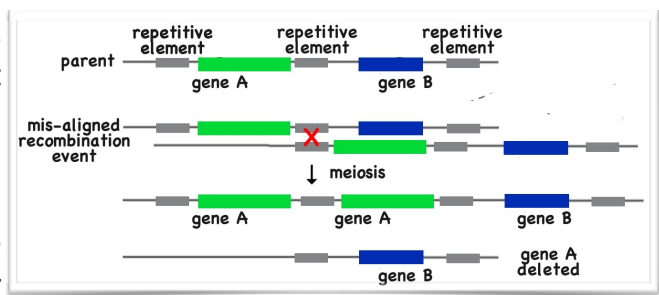
[515] this appears to have occurred with the human genome: see Exploring the genesis and functions of Human Accelerated Regions sheds light on their role in human evolution

[516] Copy number variation in humans:

[517] Child Development and Structural Variation in the Human Genome

[518] Mechanisms of Gene Duplication and Amplification

deletions. Such duplication events can have a kind of liberating effect on subsequent evolutionary pathways.[519] Most obviously, having two copies of a previously single copy gene means that it is possible for the cell/organism to make twice as many transcripts per unit time. This extra activity can be useful. For example, imagine that the original gene product was involved in inactivating a toxin; one copy of the gene might not make enough polypeptide/protein to allow the cell/organism to grow or survive, whereas two copies might. When one analyzes bacterial (or cancer) cells that can grow in the presence of a toxic compound, it is not uncommon to find that a gene that encodes a polypeptide/ protein involved in the degradation or export of the toxin from the cell has been duplicated one or more times.[520]

Another adaptive mechanism depends upon the fact (noted above) that while a particular gene product may have a clear "primary" activity, it may also have weaker, often much weaker, secondary activities. It may catalyze various off-reactions, these are sometimes referred to as off-target or promiscuous activities.[521] Assuming that a gene product's primary function is essential for survival or reproductive success, changes that negatively influence survival or reproductive success will be strongly selected against, even if they improve valuable secondary activities. In this context, the duplication of the gene allows the original activity to be preserved, while the duplicated gene can evolve freely, often in ways that improve its various, and useful, off-target activities or alter when and where the gene is expressed.

Orthologs and paralogs

When a gene with similar sequence properties is found in distinct organisms, our general assumption is that an ancestor of that gene was present in the organisms' common ancestor and that the two genes are homologs, or orthologs, of one another. Because of gene duplication events, a gene in an organism (and eventually a population) can be duplicated (→). Even more dramatically, entire genomes, particularly in plants, appear to have been duplicated multiple times during the course of their evolution.[522] In any gene duplication event, the two duplicated genes can have a number of fates, they can act as a "back-up" for one another, they can be re-purposed, or one can be lost. Repeated gene duplication events can generate families of

---
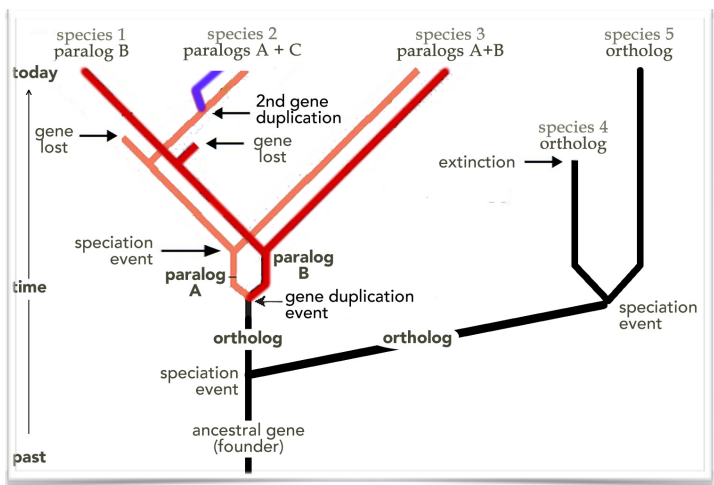
[519] Ohno's dilemma: evolution of new genes under continuous selection: and Copy-number changes in evolution: rates, fitness effects and adaptive significance

[520] Dihydrofolate reductase amplification and sensitization to methotrexate of methotrexate-resistant colon cancer cells:

[521] Enzyme promiscuity: a mechanistic and evolutionary perspective & Network Context and Selection in the Evolution to Enzyme Specificity

[522] Genome and gene duplications and gene expression divergence: a view from plants

evolutionarily-related genes that are recognized by the presence of similar nucleotide and amino acid sequences and structural motifs in the encoded polypeptides. In the analysis of gene families, we make a distinction between paralogs and orthologs.

Orthologs are homologous genes found in different organisms; they are presumed to be derived from a gene present in the last common ancestor of those organisms. Paralogous genes are derived from a gene duplication event; they are present together in a particular organism. If one paralog of a pair is subsequently lost, it can be difficult to distinguish the remaining gene from the original ortholog. A particular paralog in one organism can be orthologous to a gene in another organism, or it could have arisen independently in an ancestor, through a gene duplication event.

When both paralogs are present in a species, detailed gene/polypeptide sequences comparisons can often be used to distinguish the evolutionary family tree of a gene. That said, the further in the past that a gene duplication event occurred, the more mutational noise can obscure the relationship between the duplicated genes. For example, when looking at a DNA sequence there are only four possible bases at each position. A mutation can change a base from an A to a G, and a subsequent mutation can change the G back to A. With time, this becomes more and more frequent, making it difficult to accurately calculate the number of mutational events that separate two genes, since it could be 0, 1, 2 or a greater number. We can only generate estimates of probable relationships. Since many multigene families appear to have their origins in organisms that lived hundreds of millions of years ago, the older the common ancestor, the more obscure the relationship can be. The exceptions involve genes that are very highly conserved, which basically means that their sequences are constrained by the sequence of their gene product and natural selection. In this case most mutations produce a lethal or highly disadvantageous phenotype, meaning that the cell or organism with that mutation dies or fails to reproduce. These genes evolve (change sequence) very slowly. In contrast, gene/gene products with less rigid constraints, and this includes many genes/gene products, evolve more rapidly, which can make determining the relationships between genes found in distantly related organisms more tentative and speculative. Also, while functional similarities are often seen as evidence for evolutionary homology, it is worth considering the possibility, particularly in highly divergent genes and gene products, of convergent evolution. As with wings, the number of ways to carry out a particular molecular level function may be limited.

Transposons: moving DNA within a genome (and weird genetics)

As we are thinking about DNA molecules moving into the genome through horizontal (lateral) gene transfer, and between genomes through conjugation, we can consider another widely important molecular system known as transposons. A transposon is a piece of DNA that can move (jump) from place to place in the genome.[523] The geneticist (and Nobel prize winner) Barbara McClintock (1902–1992)(→) first identified transposons, although she did not know the molecular basis of the effect, while studying maize (*Zea mays*).[524] In particular, she studied
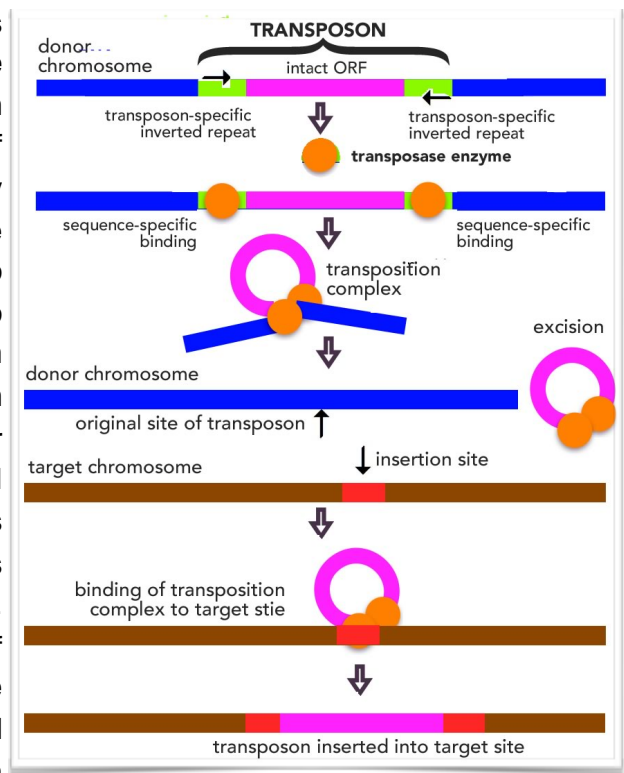
---

[523] Transposons: The Jumping Genes: http://www.nature.com/scitable/topicpage/transposons-the-jumping-genes-518

[524] Barbara McClintock: http://www.nobelprize.org/nobel_prizes/medicine/laureates/1983/mcclintock-bio.html

the phenomena of variegation in the pigmentation of kernels in maize. The variegation phenotype (↑) is due to what are known as unstable alleles; these are pairs of alleles in which one allele is associated with one phenotype (e.g. dark pigment) and the other allele is associated with another phenotype (e.g. lighter pigmentation or a different color). During development an allele can change from one state to another. Since tissues are built from (asexual) clones of somatic cells, the earlier in development an allele change occurs, the larger the region associated with the phenotype in the organism, due to the presence of the "alternative" allele.[525]

Transposons can have a number of different effects on the expression of the genes in which they are found.[526] For example, some transposons are found in the coding region of a gene, and are then spliced out of the RNA, resulting in the synthesis of a normally functioning gene product.[527] In other cases, the movement of a transposon can inactivate the gene into which it inserts. Transposons are classified into two general types - those that move a DNA sequence from one place in the genome to another with no increase in total transposon copy number – these are known, for historical reasons, as type II transposons (→). Type II transposons come in two types, known as autonomous and non-autonomous (dependent). Autonomous transposons encode a protein known as transposase. The transposon is characterized by the presence of repeat nucleotide sequences at each end. The transposase protein recognizes these sequences and catalyzes the removal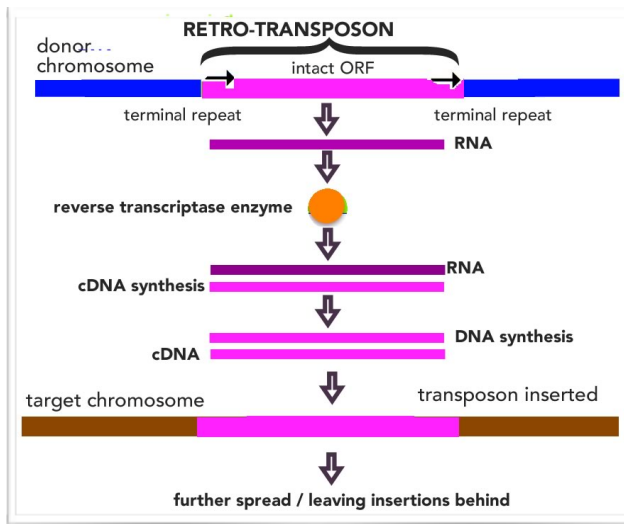 of the intervening sequence from the original site on the DNA and its subsequent insertion into another site, which can be located anywhere in the genome, for example, on another chromosome. In non-autonomous (dependent) type II transposons, mutations have led to the loss of a functional transposase gene within the transposon. By itself, such a dependent transposon cannot move; if there is an autonomous transposon within the cell, however, then the transposase it encodes can catalyze the excision and insertion of a dependent (non-autonomous) transposon. Why? because when the transposase protein is synthesized (in the cytoplasm) it can move around the cell (and within the nucleus) and interact with multiple transposons (DNA regions).

The second type of transposon, known as a type I transposon, is also a DNA sequence, but it uses a different mechanism to move. Again type I transposons come in autonomous and non-autonomous

---

525 In you can't stop yourself, check out: Controlling elements in maize – https://www.ncbi.nlm.nih.gov/books/NBK21808/. We will not go into the genetics of corn, that is something to look forward to in an advanced class in plant genetics.

526 Transposable Elements, Epigenetics, and Genome Evolution: http://science.sciencemag.org/content/338/6108/758

527 The Maize Transposable Element Ds Is Spliced from RNA: https://www.ncbi.nlm.nih.gov/pubmed/3039661

RETRO-TRANSPOSON

donor chromosome

intact ORF

terminal repeat — terminal repeat

RNA

reverse transcriptase enzyme

cDNA synthesis — RNA

cDNA — DNA synthesis

target chromosome — transposon inserted

further spread / leaving insertions behind

(dependent) forms (←). The autonomous form encodes a protein known as reverse transcriptase. When expressed, the type I transposon leads to the generation of an mRNA that encodes the reverse transcriptase (or RNA-directed, DNA polymerase) protein. The reverse transcriptase can recognize and make a complementary DNA (cDNA) copy of the transposon encoded RNA. The cDNA can, in turn, be used as the template to generate a double-stranded DNA molecule that can then be inserted, more or less randomly, into the genome. In contrast to a type II transposon, the original transposon's DNA sequence remains in place, and a new transposable element is created and inserted into the genome. If the transposon sequence is inserted into a gene, it can create a null or amorphic mutation in that gene by disrupting the gene's regulatory or coding sequences. It can also act as a regulatory element, leading to changes in when and where the gene is expressed. In contrast to an autonomous type II transposon, an autonomous type I transposon encodes a functional reverse transcriptase protein, copies itself, and leads to an increase in the number of copies of the transposon in the genome. In dependent (non-autonomous) type I transposons, mutations in the transposon sequence render the reverse transcriptase non-functional; it can only make copies of itself if an autonomous type I transposon is present and actively expressed within the genome.

Because transposons do not normally encode essential functions, random mutations can inhibit the various molecular components involved in their recognition, excision, replication, and insertion within a genome. They can be inactivated (killed) by random mutation. If you remember back to our discussion of DNA, human and many other types of genomes contain multiple copies of specific sequences - these are clearly derived from once active transposons, but most are now "dead" – they are the remains of molecular parasites. It is estimated that the human genome contains ~1,000,000 copies of the Alu type transposon (~11% of the total genome); they are dependent, type I transposons that rely on the presence of autonomous transposons to move.[528] About ~50% or more of the human genome consists of various dead transposons. It is probably not too surprising then that there is movement within genomes during the course of an organism's life time, since some transposons are still active.[529] Moreover, since transposon movement is generally stochastic, as populations separate from one another, the patterns of transposons within the genome diverge from the ancestral population.[530] In addition, various stresses within an organism can enhance transposon movement, which may play a role in the generation of genetic variation - a primary driver of evolutionary diversity and adaptation.[531]

---

[528] Wikipedia: Alu element

[529] Active transposition in genomes

[530] The impact of retrotransposons on human genome evolution: https://www.ncbi.nlm.nih.gov/pubmed/19763152

[531] Stress and transposable elements: co-evolution or useful parasites? https://www.ncbi.nlm.nih.gov/pubmed/11012710

Questions to answer:
181. How could the movement of a transposon influence gene expression?
182. What are the selective pressures on the maintenance or destruction of active transposons?
183. How could the movement of a transposable element NOT produce a mutation?

Questions to ponder:
Does the presence of molecular parasites represent an evolutionary design feature or an unintended consequence of molecular machines involved in"normal" DNA dynamics and mutational repair?

Forensics and ancestors (to be completed - stay tuned)

A good paper to read: Crime Scene Genetics:Transforming Forensic Science through Molecular Technologies
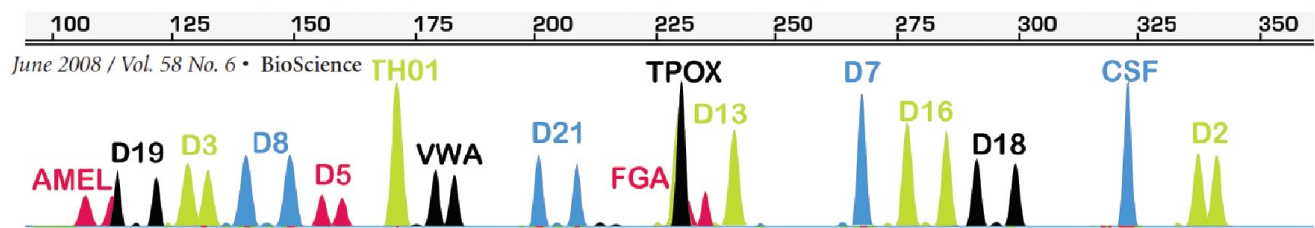
Profile of 15 human STRs + the amelogenin gene (AMEL). The lenght of each DNA fragment in base pairs along the top reveals the number of copies of a particular STR the frament contains. At loci with two differently sized peaks, the individual has two different STR alleles; a single pears indicates. the same STR copy number in both alleles. Image from John butler, National Insstitute of Standards and Technology.

Forensics: How does DNA fingerprinting work?
   - Naked Science Scrapbook. [532]

forensic studies: https://youtu.be/ZxWXCT9wVoI

Supplemental materials

Multidimensional scaling of the identity-by-state sharing dissimilarity between pairs of reference individuals.

---
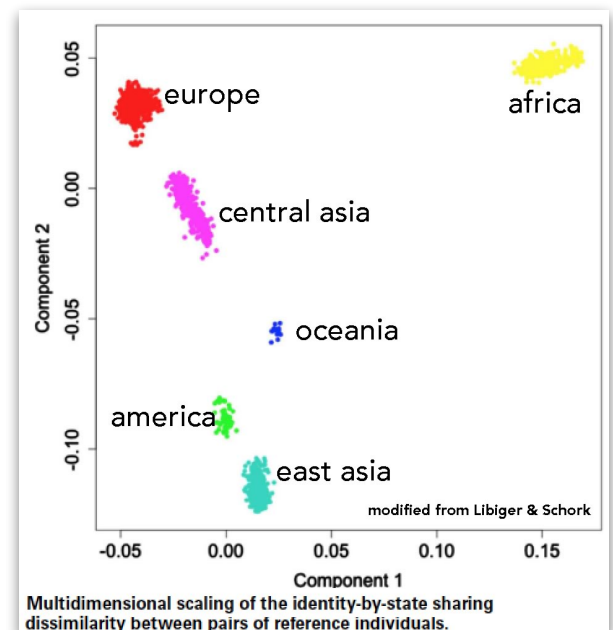
[532] A Method for Inferring an Individual's Genetic Ancestry and Degree of Admixture Associated with Six Major Continental Populations.

Extra-chapter 1. How genes control social and developmental systems

In which we consider the dynamics of social systems, from bacterial quorum sensing to the development of a complex metazoan embryo (sorry no plants!) We examine some of the signaling and gene regulatory systems involved, and the general principles of how they how work and how they were initially identified, genetically.
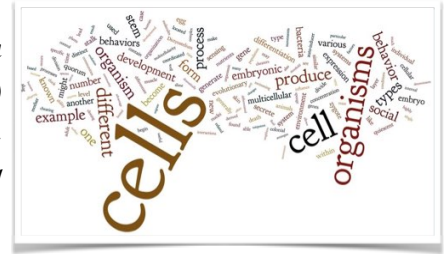
Thinking about organisms, we have to consider the combinations of molecular and cellular subsystems that combine to produce specific structures, responses, and behaviors. So what exactly do we mean by a system, and how do genes (shaped by evolutionary mechanisms) influence such systems? We have already introduced the idea of molecular systems (genes, proteins, membranes, allosteric regulators), in the context of the lac operon, together with the recognition that cells, and the organisms they form, are historically continuous bounded non-equilibrium systems. At each level of a system's organization we can identify the objects that interact, how they interact, and the outcomes (changes in cell structure and behavior leading to changes in the organism) that can result from such interactions. At the molecular level it is common to focus on the interactions between the proteins and DNA sequences that control gene expression, while at the cellular level we generally focus on the release and reception of signaling molecules, together with the responses that they induce. As we have seen, systems often (always?) include checkpoints associated with a range of cellular processes, such as the cell cycle. Other interactions are structural, such as the control of DNA packing within chromosomes, most dramatically illustrated by X-chromosome inactivation in mammalian females as a way to regulate gene expression. These molecular level interactions play an important role in determining how cells behave, how they respond to external stimuli, and how they maintain their internal organization, growth, differentiation, and reproduction. Interactions between cells also influence the behaviors of the interacting cells, as well as the overall behavior(s) of biological communities and multicellular organisms. Interactions between organisms, ranging from mutual dependencies to host-pathogen and predator-prey interactions, underlie social and ecological systems. Interaction systems are often complex, in part so that responses are controlled and regulate-able. Interactions between cells influence both lower (molecular level) and higher (organismic and social) systems. Moreover systems change over time and will respond to environmental perturbations in various, often unexpected ways. For example, systems (organisms) can age and, through gamete formation and fertilization, be reset (reborn). Systems thinking provides an analytical context to consider biological systems at all levels, from the gene to the ecosystem.

On the social behavior of microbes

The various types of organisms within a particular community are often critically dependent upon one another, and generally influence each other's behavior through both competition and cooperation. Some organisms will secrete nutrients that are needed by others for their survival. Our own need for vitamins, small organic molecules obtained from our diet, reflects this interdependence. Some organisms secrete toxins to control the growth or to influence the behaviors of other organisms, including ourselves - this is one reason that considering the microbiome has become so important in

understand human physiology.[533] There are complex molecular level conversations going on between the various organisms within an ecosystem (as well as the cells that make up a multicellular organism.) Organisms are not independent of one another, their behaviors are altered by their environment and they in turn, alter their environment.

An example of how even the simplest of organisms can cooperate is the phenomena known as quorum sensing (considered previously). A bacterium of a particular species can secrete factors that are useful, for example, in the breakdown of food into soluble nutrient molecules that it can ingest (transport through the plasma membrane) or in the generation of toxins that suppress the growth of other types of bacteria. But when growing in sparse situations (that is, low concentrations, with few organisms per unit area or volume), such a strategy may not efficient. Consider what happens if organisms at low density in a aqueous environment commit to the expense of producing complex secreted molecules, such as enzymes or antibiotics, a process that involves gene expression (RNA synthesis), and polypeptide/protein synthesis and targeting. Once secreted, such molecules are likely to diffuse away and so become effectively useless to the organism that produced them. Similarly, if anchored to the cell's surface, the products are also likely to diffuse away before being imported into the organism.

Such secreted molecules are "public goods"; once secreted they, or the product molecules they generate, are freely available to be used by other organisms.[534] For the cost of generating public goods can lead to reproductive benefits to the organisms or its relatives (inclusive fitness), the release of such public goods should be limited to when the environment contains high numbers of related organisms. In the case of quorum sensing, when the numbers of related organisms reaches a high local concentrations (organisms/unit volume), the process becomes more efficient – the concentration of the secreted molecules increases to reach useful levels and the nutrients released can be effectively captured. By cooperating with their neighbors to produce a mutually beneficial behaviors or structures (such as biofilms), individuals may benefit. The classic iconic example of a such a "communal" mutualistic system are the lichens, composed of fungi and algae.[535]

How might such types of cooperation work? In bacteria a common strategy is for individuals to produce and secrete small (relatively inexpensive, energetically) molecules known as auto-inducers, together with a cellular receptor that binds the auto-inducer, an interaction that initiates a downstream signaling cascade that control cellular and community behavior.[536] The auto-inducer-receptor system enables organisms of the same type to recognize each other's presence. The system works because different types of organisms produce different auto-inducers, and the level of auto-inducer produced by a single bacterium is insufficient to activate its own receptors or those of rare neighbors. The system is tuned – only when the density of individuals reaches a threshold level does the concentration of auto-inducer become high enough to activate the receptor system. Activation of the auto-inducer-receptor

---

[533] Book Club Reference: I contain multitudes by Ed Yong.

[534] Once again, the importance of basic economic, cost-benefit, analyses becomes apparent

[535] Evolution of complex symbiotic relationships in a morphologically derived family of lichen-forming fungi

[536] Bacterial Quorum-Sensing Network Architectures

generates a signal that in turn influences the bacterium's behavior, including gene expression.[537] There are a number of different induced responses, depending on the organism involved. These include the activation of light-emitting systems, the activation of the complex DNA import machinery, the secretion of digestive enzymes, and various cellular differentiation decisions, as to whether to form a biofilm, to actively divide, die, or to withdraw into a quiescent state.

For example in situations where nutrients become scarce a quorum sensing controlled behavior can lead some of the cells in the population to die, a process known as programmed cell death, releasing their nutrients for their neighbors to use. Which cell's die and which survive is a stochastic process. Programmed cell death can be seen as a form of altruism, since it helps the neighbors, who are likely to be relatives of the sacrificing cell, to survive and prosper at a severe cost to an individual.[538] In some cases, hostile conditions can lead to a decrease in the rate of the repair of replication associated mutations – this leads to higher (general) levels of mutations, some of these mutations may be useful survival – clearly a desperate, last ditch strategy. Another behavior is the appearance of "persisters", these are cells (organisms) that grow slowly or not at all, while the rest of the population continues to grow.[539] Persisters are formed stochastically. If the environment turns seriously hostile, persisters have a higher probability of survival than do the actively growing cells. When conditions improve, persisters reverse their behavior and establish an actively growing population. The ability to generate persisters allows the organism to survive in a range of highly fluctuating environments. For example, during antibiotic treatment, certain individuals can retreat from an actively dividing state; in such a metabolic quiet (quiescent) state, the cells are largely unharmed by the presence of the antibiotic, but when the antibiotic is gone, they can re-emerge into an actively dividing state. The ability to form persisters can greatly complicate the treatment of microbial diseases.[540]

On the other hand if the conditions never get hostile the growing cells will have a strong evolutionary advantage, and the ability to form persisters may be selected against. This is a example of group selection, based on environmental stability. One type of fluctuation could be the presence of an antibiotic; a population of bacteria that can form persisters might survive a treatment of antibiotic that would kill a population that could not. A similar persister behavior has been found to occur within populations of cancer cells.[541] We have already seen, in the context of the lac operon, an initially uniform population of organisms can produce distinct phenotypes through stochastic processes; similar random events play an important role in the determination of cell fates in many social situations.

An interesting example of a quorum sensing behavior involves the symbiotic relationship between the Hawaiian nobtail squid Euprymna scolopes and the marine bacterium Vibrio fischeri. V. fischeri is bioluminescent, which means that it can emit light using the luciferase system; but synthesizing the molecules and running the reactions involved in costly – an individual bacteria will not emit enough light

---

[537] Bacterial quorum-sensing network architectures: http://www.ncbi.nlm.nih.gov/pubmed/19686078

[538] Programmed cell death in bacteria and implications for antibiotic therapy: http://www.ncbi.nlm.nih.gov/pubmed/23684151

[539] "Persisters": Survival at the Cellular Level: http://www.ncbi.nlm.nih.gov/pubmed/21829345

[540] Bacterial Persister Cell Formation and Dormancy

[541] Evolution of cooperation among tumor cells: http://www.ncbi.nlm.nih.gov/pubmed/16938860

to be detected, only a high-density population emits light at useful levels. In this system, the squid has a specific organ within which V. fischeri bacteria accumulate, while other types of bacteria are excluded. The light emitted by the bacteria in these organs is used as a form of camouflage that allows the squid to avoid predators at night, it "emits luminescence from its ventral surface to match downwelling moonlight and starlight, thereby casting an obscured silhouette"; during the day the squid bury themselves in the sand to avoid predators.[542] V. fischeri can swim using flagellar (rotary) motors driven by the electrochemical gradient across the bacteria's plasma membrane.[543] A new born squid is born sterile, it does not have bacteria – it needs to recruit V. fischeri, but not other types of bacteria, from the environment (↓). Once this symbiosis has been established, both bacterial growth and squid immune
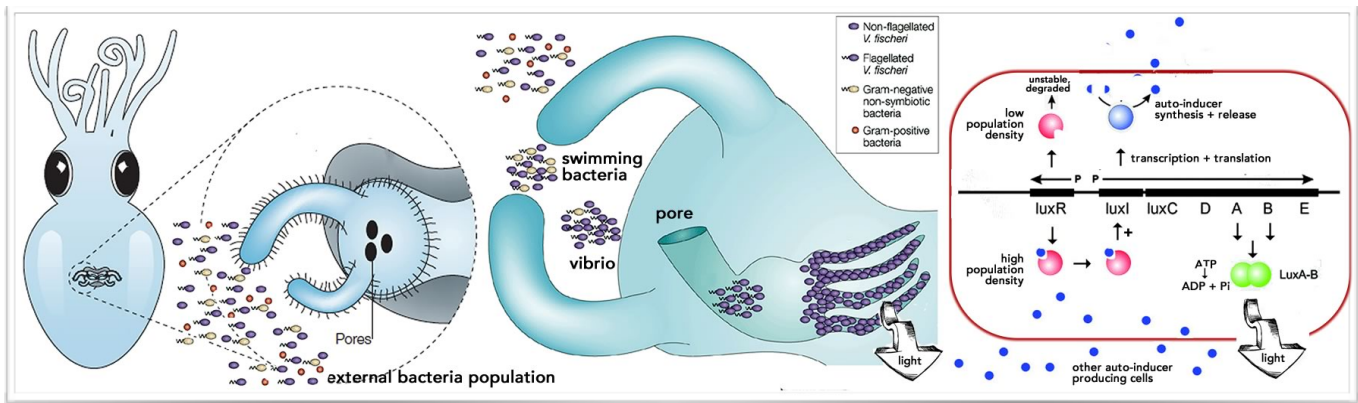
responses need to be regulated to maintain the relationship. This is analogous to the process by which bacteria colonies the surfaces (internal and external) of a newborn baby.

Within the light organ, V. fischeri activate a quorum sensing system that, when bacterial concentrates gets sufficiently high, activates the gene (operon) that encodes the proteins involved. The system is control by the expression of the regulator (LuxR) which generates an unstable transcription factor protein. The stability of this protein is regulated by an allosteric effector, the same molecule, generated by the protein encoded by the LuxI gene, that serves as the quorum sensing autoinducer. The system acts as a positive feedback loop. When autoinducer binds to the LuxR protein, the protein is stabilized and it acts to turn on the expression of Lux operon, leading to increased levels of LuxI (more autoinducer) and the expression of the genes LuxA and LuxB, which together encode the luciferase, the enzyme that catalyzes the light emitting reaction (→). A number of mutations in V. fischeri have been identified that disrupt this process; within the light organ there is a selection against strains that cannot produce light.
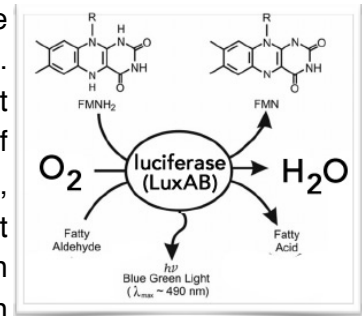
Social cheating and social defenses

---

[542] Establishing the squid-vibrio symbiosis

[543] We will not go into the details of bacterial swimming (known as chemotaxis). Basically, in the absence of a gradient of attractant (such as released by the squid), the motor reverses periodically causing the cell to tumble and change direction. When moving up a gradient (toward higher concentration) of attractant, or down a gradient of repulsant, tumbling is suppressed; the end result is directed movement.

Social cooperation between cells can provide benefits, but also opens up the system to selfish behaviors, known as social cheating.[544] Such selfish behaviors can benefit, at least temporarily, one cell within a multicellular organism or an individual organism within a social system. In the context of quorum sensing, suppose an individual does not make the auto-inducer, but continues to make its receptor. The cheater gains the benefits of communicating with other bacteria, but minimizes its contribution to the process (it does not use energy to synthesize the autoinducer). It might well gain an advantage in that the energy used to make the auto-inducer could instead be used for growth and reproduction.

Another obvious example of social cheating is cancer, in which the normally strict control of cell division is lost, and the cancer cell reproduces in a aberrant manner, leading to primary tumor formation. Subsequent mutations lead to metastasis (cell migration) and secondary tumor formation. The cancer cell, instead of behaving as a disciplined part of the whole organisms, goes out on its own - eventually leading to the death of the organism and its own extinction.

The occurrence of social cheaters raises the evolutionary question, what can be done to protect a social system against the emergence and effects of social cheaters? One is evolutionary, associated with group selection; if enough members of a particular population become cheaters, the quorum sensing/cooperative system will fail because not enough members of the community secrete the auto-inducer. Assuming that the social behavior is critical for the survival of the population, a group with too many cheaters may become uncompetitive and die out (become extinct).

There are other more pro-active strategies that can be used to suppress cheaters. It may be that the production of the auto-inducer (for example through a mutation in the LuxI gene in the V. fischeri example above) is a by-product of an essential reaction. In this case, loss of the ability to produce the auto-inducer could lead to death; the organisms become addicted to the auto-inducer. Many bacterial species synthesize toxins to which they themselves are immune, but which kill cells of related species. It can be that toxin immunity is coupled to auto-inducer expression. In more complex organisms there can be the development of the ability to recognize cheaters by their behavior in certain situations. Social systems also interact. It is possible that one strain of bacteria can make an inducers that inhibits the quorum sensing system of another.[545]

Questions to answer:
172. What are the characteristics of a "cheater" within a social system?
173. Consider possible benefits of inhibiting another organisms quorum sensing system - construct a plausible favoring or selecting against such interactions.
174. How does social cooperation influence the ecological niches an organism can inhabit?

Questions to ponder:
Are there condition in which social interactions would be selected against? Is it possible for a non-social organism to cheat?

---

[544] Safeguards for cell cooperation in mouse embryogenesis shown by genome-wide cheater screen

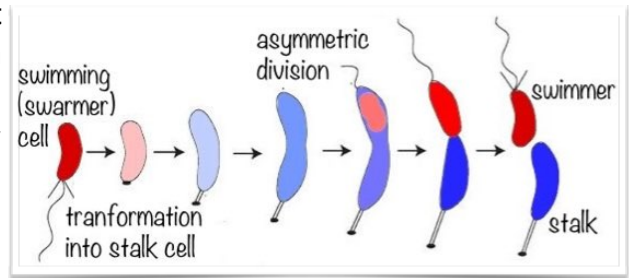[545] Social conflict drives the evolutionary divergence of quorum sensing

Cellular differentiation, introduced

So far, we have been considering the cells that interact through quorum sensing, and other mechanisms, as if they are all the same, but in any group of cells, even cell's with exactly the same genotypes, there will be differences due to the stochastic nature of many molecular processes. We described this explicitly during our discussion of the lac operon in E. coli, but such noisy processes are present and active in all cells/organisms. They lead to phenotypic changes between genetically similar (identical) systems. Often these changes are reversible, and represent adaptive metabolic responses to environmental factors, such as the presence of lactose. In other cases, the phenotypic change is more extreme and in higher organisms, normally irreversible, although in some cases it can be reversed by artificial manipulation.

We will examine two organisms that display different types what we call differentiation behaviors; we leave more molecular details for later courses (that you may or may not take). The first is the bacterium Caulobacter crescentus; these cells adhere to solid substrates through a stalk-adhesion system.[546] Only such adherent cells can replicate their DNA in preparation for cell division. When they divide, they divide asymmetrically (→). The original stalk cell remains a stalk cell, but the daughter develops differently, instead of a stalk it forms a flagellum, a screw-type propeller-like structure that drives cell swimming. When cell division is complete, the "swimmer" cell swims away until it begins the process of substrate attachment; it then retracts is flagella and forms a stalk. The process generates two cell types, the stalked cells that can replicate and the swimmer cells that can migrate within the environment and colonize new surfaces. C. crecentus has established two different cell types, two distinct phenotypes, that enable it to exploit and explore its environment.[547]

A second example of cell differentiation event is illustrated by the process of spore formation. Spore formation is an adaptation that allows an organism to survive hostile environmental conditions, while at the same time optimizing growth when conditions are "comfortable" (→). Under comfortable condition, the cells divide in a standard manner, and division can be rapid (depending upon the presence of nutrients). Under hostile conditions such "vegetative" cells would not only not be able to divide, but would like die. To avoid this outcome, the cells' respond by differentiating. After DNA replication, one copy of the DNA is packed away in a specialized compartment, the developing spore. Genes expressed from the non-spore DNA encode the structural components used to form the spore membrane/cell wall. The spore is released when the mother cell dies and its plasma membrane breaks open (lyses). The released spore is highly resistant to dehydration, radiation, and chemical insult. The
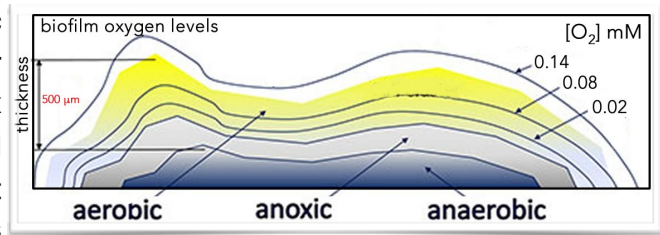
---
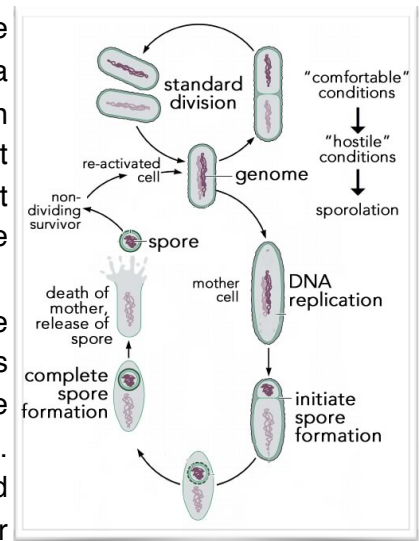
[546] a quick guide to Caulobacter

[547] Want to go deeper, try this: Cell-cycle progression and the generation of asymmetry in Caulobacter crescentus.

process is analogous to that displayed by the cryptobiotic state of the tardigrad (remember those guys?) As in the case with C. crescentus, the evolutionary logic is clear - compared to a non-spore forming relative, a spore forming population can survive harsh environmental conditions and re-emerge when conditions improve.

Microbial social complexity

In the real world, microbes (bacteria, archaea, and unicellular eukaryotes) often do not live alone, but rather in communities held together by extracellular "slime", various types of proteins that they secrete into the extracellular space. Biofilms, such as the plaque that forms on your teeth, are microbial communities and consist of a number of different types of often co-dependent organisms. The structure of a biofilm can be quite complex and often changes over time in terms of numbers and types of organisms present, as well as the levels of nutrients available, and signaling molecules (auto-induces) and toxins present. As an example, the levels of O_2 at the base of a biofilm can be very low, forcing organisms to either become quiescent or anaerobic (→); similarly levels of nutrients and other molecules can be expected to display complex concentration differences between surface and deeper regions of the film. We can also expect that the system will change over time; which microbes are most common, their distribution within the biofilm, and the types of phenotypes they display will depend upon their surroundings as well as the history of the biofilm. Different organisms may have a competitive advantage as the biofilm grows older. Biofilms can also also be subject to significant founder effects, genetic bottlenecks, and the effects of genetic drift. For example, a biofilm exposed to an antibiotic (or expresso, vodka, tobacco or marijuana smoke) might be distinctly different, in terms of the types of organisms present, compared to one that has not been so exposed. All together these factors serve to influence the biofilm and its evolution over time. In much the same way, the appearance of mutations can change the dynamics between organisms within the biofilm and, for biofilms associated with multicellular organisms, like us, the multicellular organism's response to the biofilm. At the same time to appearance of mutations within biofilm organism may influence its behavior and effects on the host. Similarly, the density of organisms in a biofilm community can facilitate horizontal gene transfer between the organisms within it.[548] The current interest in the microbiome, the various biofilm associated organisms with multicellular organisms, particularly humans, reflects an awareness of how such interactions can influence both biofilm and host. The importance of biofilm behavior, while beyond the scope of this course, is clearly an important and growing area of research.[549]

Questions to answer:

196. Why might an organism grow well in a biofilm but not in isolation (or in a pure culture)?

---

[548] The role of bacterial biofilms in chronic infections a

[549] Exploiting social evolution in biofilms & Experimental evolution in biofilm populations.

197. Describe conditions that would favor and those that might select against the ability for a spore or the ability to physiologically adapt to changing environmental conditions.

Making metazoans

The next obvious level of organization, above that of quorum sensing and biofilms, is what we will call a discrete colonial organism. In such organisms individual cells are attached to one another, generally through the extracellular materials they secrete, not unlike the organisms within a biofilm (→). They gain advantages associated with their larger size, for example, they may be able to swim faster or be too big to easily swallow but these advantages are constrained by the fact that the individual cells retain their individuality. In a pure colonial organism, each cell within the colony retains its ability to reproduce independently, either sexually or asexually. The result is that, in an evolutionary sense, they are both cooperating with one another, and competing with one another for nutrients, etc.
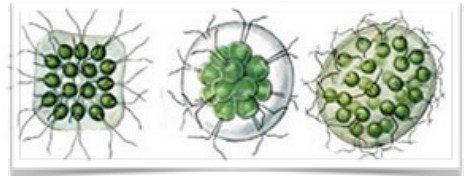
Such colonial organism can retain the ability to return to a unicellular lifestyle, whereas true multicellular organisms are, with the exception of their single-celled gametic (sperm and egg) life-stages, always multicellular. Previously we introduced the terms soma for the cells of the body that reproduce asexually and are responsible for the growth and repair of the organism, and the germ line, the cells that are responsible for producing the next generation of organisms. In a purely colonial organism, all cells are both germ cells and somatic cells. There is no central system for coordinating behavior.

So we might ask, what is the next step in the evolution of a truly integrated multicellular organism, one with a soma and a germ line? A critical step in this evolutionary process is the accumulation of adaptations that make the commitment to multicellularity irreversible, what has been termed adaptive adaptive or evolutionary "ratcheting".[550] A ratchet allows for movement in only one direction (→). In the context of evolving multicellularity, mutation-based adaptations that make multicellularity more successful (reproductively) also reduce the ability of cells to succeed (reproduce) as isolated cells, that is, outside of the multicellular organism.
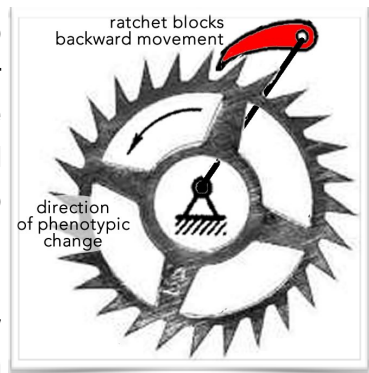
The evolution of multicellularity appears to be reasonably straightforward process since, based on fossil and genetic evidence, such as the presence of orthologous genes linked to the multicellular phenotype, multicellularity appears to have arisen independently at least 25 times during evolution.[551] Animals, plants, and fungi all appear to have taken independent strategies leading to multicellularity. It has even been possible to drive the first steps in the evolution of multicellularity in the laboratory by changing selective constraints.[552] A classic

---

[550] Stabilizing multicellularity through ratcheting

[551] The evolutionary-developmental origins of multicellularity

[552] The evolution of multicellularity: a minor major transition?

example is the study by Borass et al[553]; they examined the effect of the introducing a predator, the flagellated protist Ochromonas vallescia, on the behavior of the (normally) unicellular green alga Chlorella vulgaris. Within less that 100 generations (cell division cycles), the normally unicellular algae had adopted a multicellular (8-celled) form that were resistant to the predator. That said, this is only the first step in the path to true multicellularity, as there is no evidence for a germ line-soma division.

A temporary metazoans: Dictyostelium

We can consider some aspects of the evolution of multicellularity by studying modern organisms. These are not necessarily similar to any real ancestor; after all, these are modern organisms, well adapted to their current environment and the result of their own evolutionary histories. Nevertheless they illustrate examples of how multicellularity has arisen. The origins of multicellularity lies in the increasing importance of social interactions, originally between isolated individuals, and eventually in an increasing integrated context. A well studied example of such interactions is provided by a group of organisms, known as the slime molds, in particular Dictyostelium discoideum.[554]

Cellular slime molds live in soil and eat bacteria - they are unicellular predators.[555] Most of the time they are small, amoeba-like, haploid cells (→). Upon starvation they can undergo a dramatic aggregation process. Aggregation is triggered by the release from individual cells of pulses of cyclic adenosine monophosphate (cAMP); a process analogous to quorum sensing in bacteria (see
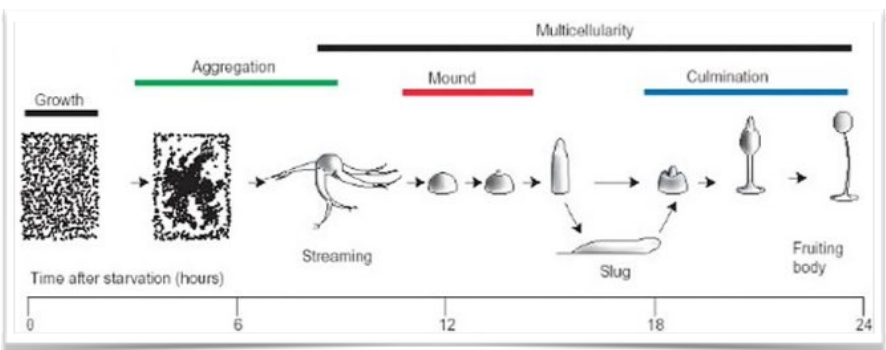
above). The result is that individual cells begin to migrate up the cAMP concentration gradient, where they interact with and adhere to one another. Groups of cells produce and secrete more cAMP leading to the formation of cellular aggregates known as slugs; a slug typically contain between 10,000 to 100,000 discrete cells. These slugs migrate in a coordinated manner. Eventually a slug will stop its migration and begin a process of known as differentiation, a process that involves changes in gene expression and cellular behavior - not unlikely the differentiation of cells in Caulobacter. Some of the cells of the slug differentiate to form stalk cells. The coordinated elongation of these stalk cells lifts the rest of the slug "body" into the air. The non-stalk cells differentiate to form spores, cells like the quiescent persisters we mentioned above. When released into the air, spores are widely dispersed and, if they land in an appropriate environment, can go on to form single celled amoebae.

By now you may be able to generate a plausible scenario to explain exactly how the self-sacrificing behavior of stalk cells is possible. The answer lies in inclusive fitness. The purpose of the slug and stalk

---

[553] Phagotrophy by a flagellate selects for colonial prey: A possible origin of multicellularity

[554] The Evolution of Aggregative Multicellularity and Cell-Cell Communication in the Dictyostelia.

[555] Cellular slime molds

are to enable Dictyostelium cells to escape a hostile environment and (hopefully) colonize new, more hospitable environments. In fact, in a number of cases the spores carry with them bacteria that inoculate their new environments, helping to insure that a new environment will be hospitable. The slime mold could be considered migrating bacterial farmers.[556] Since individual amoeboid Dictyostelium cells cannot migrate far within soil, most of the cells in any particular region, that is, the cells that combine to form a slug, are likely to be closely related to one another - they are part of a clone. Within the slug, the cells that come to lie in the anterior region, that is, the forward part of the migrating slug (about 20% of the cells present), differentiate into stalk cells, while the rest differentiate to become spores. The sacrifice of the stalk cells is more than made up for by the increased chance that the spore cells will survive and produce lots of offspring. Of course there is a danger that some cells will diverge (through mutation) and cheat the system. That is, they will avoid becoming stalk cells. Such cheating has been observed in wild type Dictyostelium and cheating is a challenged faced by all multicellular systems.[557] There are a number of strategies that are used to suppress cheaters, generally they are similar to those exploited in the context of quorum sensing.[558]

Within a population of Dictyostelium, there is a possibility that social cheaters will arise; these are cells that can respond to the aggregation signal, join with others to form a slug and a fruiting body, but end up preferentially forming spores rather than stalk cells – they rely on the sacrifices of non-cheaters to form the stalk. In the wild, cheater's will have a temporary advantage, since their chance of forming a spore, migrating, and forming a new population are higher than non-cheaters. But what happens to a new cheater-based population. When it comes time to aggregate and form a fruiting body, these cells may not be able to form a functional or useful stalk, and the process will be frustrated – without the stalk spores cannot escape from a hostile environment. Similarly, the presence of cheaters can lead to fruiting bodies with smaller stalks, impacting the fate of spores formed by non-cheaters. [559]

Questions to answer:

198. How might you determine whether a single celled, free-living Choanoflagellate evolved independently or was derived by the "simplification" of a more complex organism?

199. If you were to build an evolutionary ratchet, what properties would it have?

200. Generate a plausible model to explain why the cell's in a slime mold slug behave in a coordinated manner migrated in a single direction?

201. You are at a social event; how might you explain to others why Dictyostelium cooperate, sacrifice themselves, and how that behavior evolved and is maintained against the effects of cheaters.

Developing the germ-line soma divide

The first true multicellular organisms appear to have arisen ~1 billion years ago, apparently multiple times independently in different lineages; remember, the last common ancestor of all life probably lived

---

[556] Small molecules mediate bacterial farming by social amoebae. http://www.ncbi.nlm.nih.gov/pubmed/23975931

[557] Altruism and social cheating in the social amoeba D. discoideum

[558] Kin Recognition Protects Cooperators against Cheaters: http://www.ncbi.nlm.nih.gov/pubmed/23910661

[559] Facultative cheater mutants reveal the genetic complexity of cooperation in social amoebae.

around 3.5 to 3 billion years ago.[560] In true multicellular, as opposed to colonial organisms, different cells become highly specialized and once specialized generally can no longer divide. These "somatic" cells are relieved of the need to produce a new organism; that task is taken up by specialized cells (germ cells).

To get a better idea of the evolutionary history of multicellularity it is helpful to look in detail at the organization, both cellular and genomic, of current organisms as well as those unicellular organisms most closely related to a particular metazoan lineage.[561] We can then speculate on the various steps between the unicellular and multicellular forms. In the case of the animals, it appears that their (our) unicellular sister group are the choanoflagellates.[562] Choanoflagellates have cells that are characterized by a single flagellum surrounded by a distinctive collar structure.[563] Choanoflagellates exist in both unicellular and simple colonial forms.

Sponges (porifera) are among the simplest of the multicellular animals, known collectively as metazoans (→). Fossils of extinct sponges (such as the Archaeocyathids) have been found in Cambrian rock over 500 million years old. Earlier sponge-like organisms have been found in even older Pre-cambrian rock. Sponges contain only a few different types of cells: these include the cells that form the outer layer of the organism (pinococytes) and those that form the pores in the organism's outer layer (porocytes). The skeletal system of the sponge, the spicules, are produced by sclerocytes. A distinct cell type (archaeocytes) function in digestion, gamete production, tissue repair and regeneration. Because of their ability to form gametes, these cells correspond to the sponge germ line. Sponges also include cells, known as choanocytes, that move fluid through the body. It is the striking anatomical resemblance of these cells to the unicellular choanaflagellates that led to the hypothesis that choanoflagellates and animals are sister groups. This relationship has been confirmed through genomic analyses.[564]
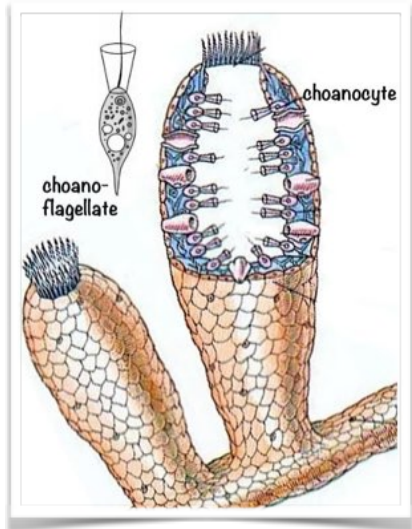
The next level of metazoan complexity is represented by hydra and related organisms, the hydrozoa, which include jellyfish. Some of these organisms alternate between a sessile (anchored) and benthic, or floating, lifestyles.[565] The hydrozoa contain more distinct cell types than the porifera. The most dramatic difference is their ability to produce coordinated movements associated with swimming and predation. While sponges behave as passive sieves, the hydrozoa have a single distinct mouth, an

---

[560] Multicellularity arose several times in the evolution of eukaryotes

[561] A video: The Origin Of Multi-Cellular Life

[562] http://www.nytimes.com/2010/12/14/science/14creatures.html?_r=0

[563] Introduction to the Choanoflagellata

[564] The genome of the choanoflagellate Monosiga brevicollis and the origin of metazoans

[565] The life cycle of jellyfish

internal stomach-like cavity, and motile arms specialized to capture prey. Their mouth also serves as their anus, through which waste is released.

Hydrozoan movements are coordinated by a network of cells, known as a nerve net, that acts to regulate contractile muscle cells (→). Together the nerve net and muscles cells generate coordinated movements, even though there is no central brain (which in its simplest form is just a dense mass of nerve cells). A hydra can display movements complicated enough to capture and engulf small fish. Stinging cells, nematocysts, are located in the "arms". Triggered by touch, they explode outward, embedding themselves in prey and delivering a paralyzing poison.[566] Hydrozoans are complex enough to be true predators.

Questions to answer:
202. What are the advantages of a closed gut versus a sieve? Which conditions might favor which approach?
203. Why is the presence of highly specialized cells considered evidence for common ancestry?
204. Dictyostelium does not have a nervous system, why do hydra (and us) need one?

Questions to ponder:
Does coordinated movement require a brain? Does having a brain equal self-awareness?

---

[566] How do jellyfish sting