THE MODERN SCHOLAR

GREAT PROFESSORS TEACHING YOU!

THE BASICS OF GENETICS

COURSE GUIDE

Professor Betsey Dexter Dyer
WHEATON COLLEGE

The Basics of Genetics

Professor Betsey Dexter Dyer
Wheaton College

The Basics of Genetics
Professor Betsey Dexter Dyer

Executive Editor
Donna F. Carnahan

RECORDING
Producer - David Markowitz
Director - Matthew Cavnar

COURSE GUIDE
Editor - James Gallagher
Design - Robin White

Cover image: © Scott Bowlin/shutterstock.com
#UT132 ISBN: 978-1-4361-6994-3

Course Syllabus

The Basics of Genetics

Photo courtesy of Betsey Dexter Dyer

About Your Professor

Betsey Dexter Dyer

Betsey Dexter Dyer is a biology professor at Wheaton College in Norton, Massachusetts, where her courses include bacteriology, genetics, parasitology, and invertebrate evolution. She earned her Ph.D. in biology at Boston University in 1984. Dyer's research interests include DNA sequence analysis, cell evolution, symbiosis, and field microbiology. Dyer considers herself to be a curious naturalist and a generalist, with lots more to learn. She has written three books: *Perl for Exploring DNA* (with coauthor Mark LeBlanc, Oxford University Press, 2007), *A Field Guide to Bacteria* (Cornell University Press, 2003), and *Tracing the History of Eukaryotic Cells* (with coauthor Robert Obar, Columbia University Press, 1994).

Dyer grew up on a family farm in Rehoboth, Massachusetts, which was a great influence on her development as a biologist and naturalist. She had many opportunities to be puzzled about why the offspring of animals (such as a new litter of kittens in the barn) might look so different from their parents. She lives in Walpole, Massachusetts, with husband Robert Obar, a protein chemist, two children Alice and Sam, and a Brittany, Genevieve. She loves reading, writing, cooking, and dancing.

Introduction

Genetics is the study of the hereditary information of organisms, how it is used, and how it is transferred through generations. Genetics has deep, practical roots in animal and plant husbandry and has its scientific beginnings in natural questions (such as the following) asked over and over by our human ancestors:

"Why do the offspring of those particular parents look and act the way they do?"

Recently, genetics has become quite sophisticated and seemingly far afield of such basic questions. Now that scientists seem to know so much about DNA (the hereditary information) and are able to manipulate DNA in so many astounding ways, the nature of the questions seems to have changed. We are more and more called upon to tackle such enormous philosophical conundrums as the following:

"Should I be carrying an identification card containing all of the sequences of my genes?"

"Is it okay to clone a beloved, deceased pet? A human? An organ of a human with the plan to transplant it? And what IS cloning anyway?"

Indeed, to be a well-educated, conversant human is to be able to pull together some sort of coherent response concerning the latest genetic developments. However, I would argue that there is still a great deal of mystery in the fundamental question:

"Why do the offspring of those particular parents look and act the way they do?"

The goal of this course is to create a foundation in the practical, everyday genetics that fascinated our ancestors and which is still fascinating. The plan is to not skip quickly over it but to linger on examples that will leave you feeling more empowered to think about and talk about what genes do. Many examples will be drawn from easily visualized traits of domestic animals and plants. As a curious naturalist myself, that sort of information alone gives me a great deal of satisfaction—having that insight about my everyday world. Indeed, it was exactly the sort of information that was missing from my life as a young naturalist growing up on a farm. I had access to the results of all sorts of matings: new kittens, new calves, new chicks, new guinea pigs, but almost no understanding of any mechanisms of genetics and no particular confidence in learning it from the few complicated-looking diagrams I could find on the topic.

How about the big, important philosophical problems for which we are asked more and more often to have at the ready some sort of well-informed response? I think those follow naturally (as they have in the history of genetics) right after the foundations are set out. First I will tell you all about kittens and puppies and other familiar animals (including zebras, birds, and goldfish) and all about vegetables and flowers, such that the grocery or florist shop could be your laboratory and seed catalogues could be a sort of textbook. I do not intend to pass too quickly over any of those basics. Throughout those lectures I will interweave terminology, concepts, and topics from current problems in genetics. For example, in the lectures about Gregor Mendel, I will include some information on cloning.

In this course we also will make our way toward an understanding of the amazing properties of DNA sequences themselves and how those properties lend themselves to such things as "genetic engineering" or "genetic modification," viruses, genetic diseases and conditions (cancers, birth defects), genetic testing, and therapies. "All that?" you might be wondering. Well, the foundations for all that. Those topics may appear to be springing up as complete entities unto themselves, each with its own thicket of questions, but all have deep roots in practical genetics and all are themes of a grander overall scheme of DNA structure and function.

Lecture 1:
The Long History of Practical Genetics

The Suggested Reading for this lecture is Juliet Clutton-Brock's *A Natural History of Domesticated Mammals*.

This course focuses on the basics of genetics:

- The rules and patterns by which genes work
- The many fascinating exceptions to those rules and patterns
- Genes: what they are, how they function
- DNA (deoxyribonucleic acid): what it is, how it functions

A goal for this course is to give you a firmer foundation in those basics of genetics so that you can be that much more conversant and thoughtful about some major current topics in genetics. Another goal is to make genetics accessible to you, no matter what your background in science; indeed, you will be encouraged throughout to explore and investigate at home or in gardens or pet shops or other venues, where genetic variations are readily observed.

How accessible is genetics? It is so accessible and so practical that our human ancestors for thousands of years have been cleverly manipulating and evaluating genetic crosses without any particular scientific training in the modern (seventeenth-century) sense of science. Genetics is a foundation of agriculture from its deepest roots ten thousand years ago and without the practice of genetics, agriculture would not have evolved to become such an efficient and abundant supplier of animals and plants for our use. The transition from gathering and hunting food to cultivating and breeding plants and animals must have been accompanied by a close scrutiny and manipulation of the reproduction of those plants and animals. Arranging matings, monitoring the results (the offspring), and adjusting future matings were the obvious activities of the most successful early farmers. Thus wild species of animals and plants were brought into domestication, usually becoming more enhanced in ways useful to humans. The evidence of these ancient endeavors in genetics is all around us.

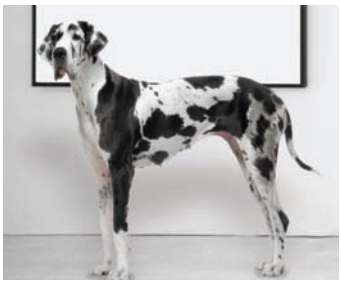
Wild apples (crab apples) are tiny, hard, and sour. Domesticated apples (the products of agricultural practices) are enormous, tender, and sweet. Wild wheat has fewer and smaller grains than domesticated wheat. The original, wild wheat, a version of which still grows wild, is hypothesized to have had just one grain per spiklet and therefore just a few grains per spike, thus the name "einkorn," meaning one seed. Furthermore, wild wheat is more difficult to harvest because the grains are more fragile and more likely to drop off before

A male *mouflon*, the species that is thought to be one of the ancestors for all modern domestic sheep breeds.

harvesting. Domestic wheat has been bred for efficient, abundant processing, whether by ancient methods such as thrashing and winnowing or by their mechanical counterparts. Domestic egg-laying, milk-producing, wool-producing, and meat-producing animals of all sorts are doing so on a far more extravagant scale than any of their wild counterparts.

Canis lupus, North American Grey Wolf

However, domesticated dogs provide some of the very best evidence of the skill of our ancestors worldwide in manipulating genetics. Humans paid close attention to the behaviors of wolves and chose the tamest and most cooperative of wolf puppies to produce what would become the many lineages of dogs. Dogs have been bred for a great variety of work (guarding, hunting) but also for aesthetics perhaps more so than any other domestic animal. Get a large picture book of dogs (or go to a dog show) and begin by just looking at place names (from all over the world) associated with particular breeds, indicative of widespread and enormous success in getting dogs to look and act pretty much any way that the human breeders could imagine. Thus we have from an ancestral wolf-like wild canine, the most extravagant ranges of sizes (Chihuahua to Great Dane), colors (brindle, piebald, golden, chocolate), and shapes (bull dogs, Pekinese, corgis, greyhounds), not to mention diverse dispositions and other behavioral tendencies such as pointing and fetching. If you'd like to see more evidence of extravagant genetics, get picture books from the library of breeds of fancy goldfish, fancy pigeons, fancy mice, poultry, sheep, or almost any domestic animal that has been bred for thousands of years by imaginative humans. The word "fancy" usually implies that the breeding has gone far beyond mere practicalities and into some of the more individual tastes in aesthetics.



A harlequin-coated Great Dane

A long-haired Chihuahua

Next take a field trip to the grocery store and be amazed at the varieties of plants all in the genus *Brassica* (also called "mustards" or "crucifers") and which abound in the produce section. They all derive from small, spindly, seemingly inedible wild mustards and include the following:

- Broccoli
- Broccoli rabe
- Brussels sprouts
- Cabbage
- Canola (or rape)
- Cauliflower
- Chinese cabbage
- Collard
- Kale
- Mustard (as in mustard seed)
- Swiss chard
- Turnip


Broccoli


Brussels Sprouts


Cabbage


Broccoli rabe


Brassica rapa, wild mustard


Cauliflower


Turnips


Chinese cabbage


Kale

Note that many of these are closely associated with regional cuisines and are indicative of preferences for one or another plant part. For example, flowers (as in cauliflower) were selected, or unopened flower buds (as in broccoli), or oil (as in canola), or roots (as in turnips). You probably have some version of wild mustard growing in your area for comparison. Get a field guide to wild flowers and look up wild mustard or wild radish and find some.

So how much of the underlying genetics did our amazing ancestors understand? They understood almost none of it! They were persistent and observant and made some good guesses. Most of their mistakes no longer exist. Practical genetics is all about making choices: allowing this organism to breed, but not that organism; allowing this organism to live to reproductive age, but not that organism; saving the seeds of this organism, but not that one and so on through many generations.

Our ancestors also must also have pondered many surprise outcomes ranging from the mild (such as the unexpected colors in a litter of new kittens) to the horrifying or disturbing (such as birth defects of all kinds that used to be classified as “monstrous”). Furthermore, our ancestors became clever at animal and plant husbandry, monitoring reproductive cycles, carefully taking seasons into account, and experimenting with means of propagation such as the grafting of trees and vines. The practices by which the domestication and specific enhancements of animals and plants occurred were mostly about common sense. No one had any knowledge of “genetics” *per se*, that is, the study of how genes (the hereditary information of organisms) are transferred and how they function.

Now and then superstitions must have arisen due to the unpredictable nature of some heritable traits. Humans being humans, there were attempts (in a pre-science world) to rationalize and find causes, and thus superstitions and religious beliefs were often evoked as explanations. For example, there was a notion that a baby might have a short life if the pregnant mother cut her hair. (Many babies of our ancestors had short lives and many pregnant women cut their hair. Such correlations can be difficult to sort out.) However, superstition must have played a relatively small part overall. The rudiments of scientific logic seem to have prevailed. The evidence is in the many successful results with domesticated plants and animals that could only have come by repeated and thoughtful experimentation and not by wild guesses or magical spells. The most effective (most logical) breeders of animals and plants most likely were also the most successful. For important lineages such as racehorses and special dogs, humans developed ways to keep records and thereby to establish "pedigrees."

By the way, there still are plenty of surprises in genetics, even among knowledgeable breeders. For example, striving to produce a faster and faster lineage of race horses might inadvertently produce some with overly long, brittle, breakable bones or some other skeletal or muscular defect.

Humans are also prodigious manipulators of their own reproduction and thus the world over there are distinctive regional differences in human populations based on hundreds of generations of some lineages being able to reproduce more often and abundantly than others. Medical doctors are particularly interested in regional differences among humans because along with the distinctive positive traits of a location and lineage such as skin color, height, and shape of face there may be surprise negative traits such as a tendency to a particular eye problem or blood disease.

Our human ancestors considered this important enough that pedigrees of humans or genealogies are one of the most ancient forms of record keeping, whether by oral or written methods, and are the foundation of history. Our ancestors developed and used elaborate vocabularies to indicate relatedness as in "second cousin," "great-great-aunt," and so on. There was even an aspect of manipulation of genetics through arranged marriages not only for religious, economic, or other cultural reasons, but also from the perception that certain traits such as royalty or strength or wisdom might be passed down through generations. It was difficult then as it still is now to sort out accidental effects of good fortune versus the effects of hard work or good teachers versus the effects of heritable traits passed down from a parent's genes. The analysis still goes under the rubric of "nurture versus nature," although a simple dichotomy like that does not do justice to the problem.

After thousands of years of good work in genetics, a few humans began to understand some of the science behind it in the last part of the nineteenth century. Then it was a major project of the twentieth century to establish the fundamental principles and mechanisms with a major turning point being the discovery of the structure of DNA in 1953. Genetics is still a work in progress, with new and amazing things being discovered every year.

FOR GREATER UNDERSTANDING

Questions and Activities

1. How does domesticated wheat differ from wild wheat?
2. What is the evidence of the enormous success of our human ancestors in practical genetics?
3. Develop a repertoire of vivid examples from among your own favorite domesticated animals and plants.

Suggested Reading

Clutton-Brock, Juliet. *A Natural History of Domesticated Mammals*. 2nd ed. Cambridge: Cambridge University Press, 1999.

Other Books of Interest

Brown, Jack, and Peter Caligari. *An Introduction to Plant Breeding*. Oxford: Wiley-Blackwell, 2008.

Lecture 2:
What Did Gregor Mendel Do?

The Suggested Reading for this lecture is James Schwartz's *In Pursuit of the Gene: From Darwin to DNA*.

Before Mendel

Although the history of genetics typically begins with the investigations of Gregor Mendel in the 1850s, the actual practical beginnings of genetics already were deeply rooted worldwide. Humans had been asking questions about genetics for many thousands of years, probably ever since they were able to put into words puzzles like "Why does this child from these parents seem to have some traits or mixture of traits in common with the parents and yet some traits that seem different from either parent?" By the time Mendel was doing his experiments on peas in the 1850s, many humans worldwide were experts many times over in the practices of genetics. They were simply missing the mechanisms. They knew the practicalities of "how" but they didn't know "why."

Public Domain

Gregor Johann Mendel
(1822–1884)

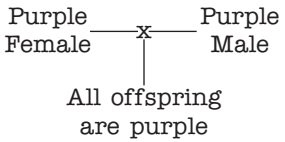
So what did Mendel do that was such a turning point in genetics? He decided to look at one simple trait at a time for an organism that could be made to reproduce quickly and abundantly, the pea, *Pisum sativum*. It might seem obvious now, but it was not obvious at the time. Practical genetics in the 1850s had advanced to the point of being an essential part of any comprehensive breeding program by which any animal or plant might be improved. Racehorses are a good example from that time period. Their reproduction was carefully controlled and followed closely and analyzed with detailed pedigree charts. Horse breeders wanted as much as possible from their expensive, potentially lucrative, and relatively slow-reproducing animals. They wanted speed, stamina, strength, intelligence, and determination. They wanted winners. If that wasn't enough, they wanted beauty: lovely coat colors, flowing tails and manes, nicely shaped heads and bodies. That's why none of it was ever successfully sorted out, so that the hereditary mechanisms could be understood. Genetics seemed to be something like mixing various colors and amounts of paints from a paint box and then trying to understand how the shades and gradations of resulting colors might have come about. Indeed, paint mixing was a popular metaphor for what was going on. The traits of the stallion and mare were somehow "mixing" in the colt, but exactly how was not understood. Furthermore there must have been occasional inexplicable surprises as there still are today in animal and plant breeding (especially when

breeders are overeager to possess all valuable traits at once). These surprises include amazingly fast and winning horses from a couple of improbably slow parents and (at the other end of good fortune) unpredictable, catastrophic fractures of delicate leg bones in horses of noble racing lineages.

In contrast to the elaborate, high-stakes schemes of horse breeders, Mendel's endeavors were quite modest and that seems to have been key to his success. Not only did Mendel look at only one trait of his peas at a time, but he chose trivial traits of relatively minor interest to pea growers. He did not look at flavor or aroma or shelf life in storage or any other trait that might be valued in a food plant. Instead, he chose traits that were easily visualized, easily counted and recorded. These included flower color (purple or white), seed color (yellow or green), plant height (tall or short), and seed shape (wrinkled or smooth). These particular varieties of peas (different flower colors, seed colors, and so forth) already existed and were available from seed companies and were the results of many generations of pea breeding, just as different colors of horses or dogs were the results of generations of breeding. Mendel set out to understand the basic mechanisms by which plant and animal varieties occurred and interacted using those simple pea traits, one at a time.

True-Breeding Peas

Mendel's simplest mating (or cross) between two pea plants demonstrated "true breeding." This is also one of the crosses of great interest to any breeder of animals or plants who has found a set of desirable traits in an organism and would like those traits to be passed down as completely as possible to all subsequent generations. Mendel simply took each of his varieties and crossed like traits with like: purple flower with purple flower, yellow seeded with yellow seeded, and found perhaps to no great surprise that purple-flowered parents yielded purple offspring and yellow-seeded parents yielded yellow offspring. Indeed, that is what "true breeding" means. Let's represent one of those crosses like this:



So true breeding is not much more than the phenomenon of similar parents yielding offspring similar to themselves. This tendency to "breed true" is one that a seed company might advertise as one of the reliable attributes of a particular seed. It is also the foundation of the idea of "pedigree," "thoroughbred," "purebred," and "pureblood" lineages of animals. Humans even refer proudly to themselves with adjectives like pureblood, dating from a time when the blood was thought to confer genetic traits. According to the *Oxford English Dictionary*, the concept of true breeding dates from at least as far back as the time of William Shakespeare. It has come to mean that identical, or as nearly identical as can be managed, organisms mate with each other and produce identical or nearly identical offspring. It can also imply true breeding for a particular set of traits but not all of the traits. For example, a breeder of miniature

poodles might take great pride in producing a consistent lineage of desirable miniature traits, generation after generation of carefully chosen matings. However, the dogs also come in a large variety of coat colors, which are not necessarily true breeding. Surprises can occur too, such as shapes and sizes that are not quite of dog-show quality, temperaments that may be problematic and even genetic disabilities such as hip dysplasia. In this and in subsequent lectures, I will explain more about how such surprises might occur in seemingly stable lineages of supposedly "true breeding" genetic traits.

"Inbreeding" (the breeding of like organisms) is a way to get a true (or nearly true) breeding lineage, but inbreeding can have negative connotations. Indeed, before our human ancestors understood exactly what might go wrong, they had some deep misgivings, taboos, and prohibitions (as well as natural inclinations) against incest or mating with their closest relatives. However, despite taboos against inbreeding, our human ancestors nonetheless were probably quite inbred. This would simply be a consequence of being in a small population, with little mobility in or out, isolated from other populations for many generations. Geneticists studying human variation have concluded that for most of the one million years of evolution of our species, we've existed mostly in small, inbred populations. For example, church records for tiny villages in England reveal inbred genealogies, unavoidable because people rarely traveled more than several miles from their birthplaces to find a spouse. There are plenty of cousin and second-cousin marriages in all of our ancestries. For the most part things went well in our ancestral, small, partly inbred villages. However, rare genetic diseases sometimes appear in higher frequencies in particular lineages of humans because of that tendency to inbreed.

Meanwhile, in some human lineages, especially ones that considered themselves royal, a certain amount of inbreeding (presumably to maintain the purebred lineage of royalty) was considered positive and was deliberately arranged—and still is to some extent—to this day among some royals. As a result, some "royal diseases" can arise, such as the prevalence of hemophilia (a blood-clotting disorder) in the inbred royal families of Europe.

Similar to our human ancestors, many wild populations of animals and plants are inbred to some degree, a natural result of little migration with the advantages of mostly good traits being perpetuated in an environment that remains relatively stable. On the other hand, for domestic animals and plants, there was probably considerable experimentation with close inbreeding, for the purpose of establishing thoroughbred or purebred lineages. Lineages of domestic animals and plants are often inbred because of the conservative breeding practices of their human facilitators. Some breeders in an attempt to perpetuate desirable traits may have bred siblings. Sometimes the consequences were fine. In other cases rare genetic disorders appeared as they still do today. Hip dysplasia, epilepsy, and cataracts in dogs are just a few examples. Some of the negative connotations of inbreeding come about because some deleterious genes function only when there are double copies of them. For example, animals typically have two copies of each gene. If a gene "q" is deleterious in double copies but I have only one copy, I am merely a carrier for "q," but will not suffer any consequences of "q" myself. However, if two

organisms (perhaps relatives or perhaps just coincidently each bearing a copy of "q") get together, then there will be possibilities for offspring that are "qq" with negative consequences. The *Oxford English Dictionary* provides some insight in our ancestors' understanding of undesirable aspects of inbreeding:

> **In•breed•ing** |'in,brēd,ing|
> Breeding from animals of the same parentage or closely related; breeding in-and-in. ***c*1842** E. J. LANCE *Cottage Farmer*, An in-breeding soon breaks up the inattentive keeper of domestic animals. **1881** J. P. SHELDON *Dairy Farming* 4/2 *In-breeding*, that is, the breeding for a time amongst near relations generally results mischievously on the systems and on the fertility of the stock. **1882** *Standard* 23 Aug. 2/1 Over-preserving and 'in-breeding' are deteriorating the herds.

Laboratory white rats bred for generations are a good example of a true-breeding lineage developed via inbreeding for the purpose of having as many genetic traits identical as possible. In that lineage, even very close relatives (twin-like) are mating and presumably the rat breeders are getting away with it because so many deleterious effects of inbreeding were culled early on in the lineage. Furthermore, white rats are maintained in controlled laboratory conditions with plenty of food and little competition. In the wild, some of their gene combinations might be detrimental. However, in spite of the extraordinary uniformity of white rats, they still have differences in their genomes because of important complications concerning chromosomes (lecture 10), gene regulation (lecture 12), and the surprisingly unstable structure and arrangement of genetic material that will have to be explained in lecture 13 on transposons (jumping genes) and viruses.

So how about Mendel's peas—the ones from the original seed packets he purchased? Were they true breeding? Were they inbred? Yes, we could use those words for them and that is what Mendel confirmed in his first sets of crosses. We don't know whether they were perfectly true breeding for all of their thousands of genes, but certainly they were true breeding for purple flower, white flower, yellow seed, and so forth, as well as a host of other traits considered desirable by the original seed grower. Furthermore, being plants, they were probably much more true breeding and inbred than any mammal with which they might be compared.

Cloning

The idea of true breeding is also relevant to understanding identical twins and clones, which genetically are the same phenomenon, albeit with different mechanisms. Identical twins or clones have the same (or very nearly the same) complement of genes (hereditary material). Essentially, those true-breeding purple-flowered plants of Mendel were "twins" of each other, although we don't usually use the word "twin" for plants. Some true-breeding plants are even clones of each other, although now we typically reserve the word "clone" for particular results of procedures in modern genetics labs.

In order to properly sort out the word "clone," I again went to the *Oxford English Dictionary* to get a better idea about the evolving use of "clone," including the development of emotionally charged connotations. It turns out

that a discussion of “cloning” presents an excellent opportunity to present some crucial (although sometimes overlooked) differences between animals and plants that have profound effects on their genetics. Actually, our human ancestors were acutely aware of these differences because they involve so many practical aspects of plant and animal breeding. It matters very much that Mendel chose plants and not animals for his experiments. He would not have gotten the same clear results otherwise. The problem comes in when we use a term like “cloning” in general as though it could be universally applied to any organism with identical results. That’s not the case. In a sense, even when trying to understand modern genetics in a theoretical way, it is difficult to get away from the everyday practicalities of handling the organisms themselves and understanding their differences. Let’s sort it out.

It turns out that plants have many natural mechanisms of propagation by which they can bypass the complications of having parents all together. These mechanisms include some that gardeners will recognize, such as the following:

1. Taking cuttings and rooting them in moist sand or water.
2. Taking cuttings and grafting them onto other plants.
3. Dividing roots, bulbs, and tubers.
4. Establishing new plants from runners.

It is actually extremely easy to get a set of identical plants and thereby (if one chooses to mate them) to develop a true-breeding (even clonal) lineage. Plants in the wild make great use of their asexual (non-sexual, clonal) mechanisms for reproduction, often to the exclusion of the relatively transient opportunities for sexual reproduction via flowers and seeds. For example, a stand of aspen trees may be identical, all connected underground having sprouted up from a common root system. Another way to think about it is that plants easily can avoid the hassle of having two parents and do so regularly.

In contrast, mammals (for example, humans and dogs) have no asexual reproduction by which to avoid the necessity of having two parents. Mammals are obligately sexual and from the point-of-view of a plant, strangely limited in their reproductive repertoires. There are no naturally occurring mechanisms to get a perfectly true breeding lineage of mammals. Mating identical twins is not an option, as identical twins are (by definition) of the same sex. Very recently some techniques in the laboratory have been developed to attempt to bypass the step of having two parents in mammals and thereby to accomplish an asexual, clonal means of propagation, that which plants do on their own routinely. The genetic material from a single individual (one parent) is implanted into an egg and, if all goes well, the offspring that develops from the egg is an identical twin of its sole parent.

An additional characteristic of plants allows them to “self.” Their sexual reproduction (which occurs less often than their asexual, clonal methods) involves pollen (the equivalent of sperm) fertilizing an ovum (the equivalent of an egg). The pollen can come from almost anywhere via the wind or a pollinating animal such as a honeybee. It can be any pollen: pollen from a completely different species (and no fertilization), pollen from a completely different species (yet a surprise fertilization), pollen from a distant plant of the

same species, pollen from a nearby, identical plant, or pollen from the very same plant, which is called "selfing." Many plants have both male and female parts on the same flower or at least on the same plant and many have no mechanisms against selfing (although some do). One term for having both male and female organs is hermaphrodism, but more typically plants in that condition are called "monecious."

In contrast, mammals have little or no such flexibility as to how and where their sperm and eggs are joined. Fertilization always requires two partners, is internal, and generally only is successful if those partners are of the same species. Indeed, "species" for mammals and some other animals (with similar limitations in reproduction) is often defined by the ability to mate and produce offspring who in turn can mate, and so on.

There is another complication and intriguing difference between mammals and plants: plants seem to have a much greater tolerance for the results of true breeding. Their indeterminate growth (for example, with widely varying numbers and shapes of branches) is in contrast to the more determined body plan of mammals, in which (for example) extra limbs would be considered a birth defect.

All of this means that a breeding pair of pedigree black Labrador retrievers is never "true breeding" in the strictest sense, that one could use for plants. A looser definition of "true breeding" describes what most breeders attempt to do with just a small subset of genetic traits that define a particular breed, such as size, shape, color, and behavior.

But there is more, especially with regard to plants: Plant breeders must have noticed early on that plants were capable of a great variety of propagation methods (including asexual mechanisms) and were tolerant of a much wider range of "inbreeding" and "true breeding" practices. For some of these cases the word "clone" was coined. Currently, this word has been taken over by science fiction writers and modern geneticists alike, as well as becoming a metaphor for a host of phenomena involving imitation. However, "clone" (as either a noun or a verb), has deep roots in the *Oxford English Dictionary*. It was first applied to plants in the early twentieth century.

Clone | klōn |
—noun
1. *Botany*
- **a.** A group of cultivated plants the individuals of which are transplanted parts of one original stock, the propagation having been carried out by the use of grafts, cuttings, bulbs, etc.
- **b.** In wider use in *Biol.* Any group of cells or organisms produced asexually from a single sexually produced ancestor.

2. a. *fig.* A person or animal that develops from one somatic cell of its parent and is genetically identical to that parent. Also (*colloq.*), a person who imitates another, esp. slavishly.
- **b.** A thing produced in imitation of, or closely resembling, another; *spec.* a microcomputer designed to simulate the functions of another (usu. more expensive) model.

—verb

a. To propagate or cause to reproduce so as to form a clone. (And as of 2004 in the *OED*) a new verb definition: To make copies of (a DNA molecule, base sequence, gene, etc.).

b. (*loosely*) To propagate or reproduce (an identical individual) from a given original; to replicate (an existing individual). Chiefly *fig.*

1903 H. J. WEBBER in *Science* 16 Oct. 502/2

"Clon[e]s . . . are groups of plants that are propagated by the use of any form of vegetative parts."

[Professor Dyer: Such as most commercially grown fruit trees.]

1929 *Bibliographia Genetica* V. 234

"In Bacillus coli communis *. . . a biotype was also found having lower motility than the remainder of the clone from which it came."*

[Professor Dyer: This is the first use as applied to bacteria (*Bacillus coli*, which is now called *Escherischia coli*.) At the time, bacteria were often classified together with plants.]

1958 *New Scientist* 20 Feb. 13/1

"Various techniques have been devised for producing these 'clone cultures' from single cells."

[Professor Dyer: The British biologist Honor Fell developed methods of taking single cells from mammalian tissues and growing them on Petri dishes. She is one of the founding researchers of what is now thought of as "stem cell research." Note that plant "stem cell research" can be done by almost any clever gardener or botanist. Plants lend themselves well to such manipulations because asexual propagation is normal for most plants. The logistics for animals are more challenging.]

1959 *Nature* 22 Aug. 648 (*heading*)

"A New Technique for Isolating and Cloning Cells of Higher Plants."

[Professor Dyer: In addition to typical gardeners' techniques for asexual propagation of plants (a legacy of eons), there was still much to be done with getting the techniques refined for laboratory use.]

1959 *Genetics* XLIV. 1259

"A number of variants were obtained from two recently cloned lines of strain HeLa S3."

[Professor Dyer: In this groundbreaking work cancer cells isolated from a woman named Henrietta Lacks who died of cancer in 1951 were established on Petri dishes. This became the first (of many) immortal lineages of cloned human cells.]

1968 *Observer* (Colour Suppl.) 10 Mar. 9/1

"One of the most extraordinary of the possibilities now being explored . . . is referred to as 'cloning people' the creation of genetically identical individuals from body cells."

[Professor Dyer: This may be the first written mentions of this possibility. It must have caught the imaginations of popular science writers, social commentators, and science fiction writers, given the flurry of subsequent reference, many from the popular press in the *OED*.

Note that the Ultimate Science Fiction Web Guide (http://www.magicdragon.com/UltimateSF/clone.html) has analyzed this topic. The idea of cloning (such as making multiple copies of humans) is a popular theme but predates the actual use of the word "clone" in that context. They cite A. E. Van Vogt's 1945 serial and later novel *The World of Null-A*.]

1970 *Nature* 19 Apr. 210/2
"The Jockey Club was . . . understandably cool when asked to comment on the possibility of a dozen cloned Arkles thundering neck and neck round the course at Epsom."

1970 A. TOFFLER *Future Shock* ix. 197
"Those most likely to replicate themselves will be those who are most narcissistic, and . . . the clones they produce will also be narcissists."

1974 *Proc. National Acad. Sci. U.S.A.* **71** 1747/2
"The procedure reported here offers a general approach utilizing bacterial plasmids for the cloning of DNA molecules from various sources." AND Proc. National Acad. Sci. U.S.A. *71 3459/1 ColE1 has been shown to serve as an effective molecular vehicle for cloning and amplifying specific regions of unrelated DNA."*

[Professor Dyer: This is groundbreaking work by which "cloning" went from being bounded and constrained by the organism to being about *any* genetic information. Previously, it was known that plants and bacteria were easy to clone and animals much more difficult and limited. In this work, animal genes were inserted into bacteria and behold, suddenly animal genes were easy to clone. This launched genetic engineering.]

1979 *Whig-Standard* (Kingston, Ontario) 13 Nov. 23/1
"The 32-year-old is not one of a myriad of Elvis clones who came out of the woodwork when the King died two years ago."

1982 *Sci. Amer.* May 112/1
"Individual organisms that arise asexually from the somatic, or body, cells of the parent rather than from the specialized sexual cells are called clones."

[Professor Dyer: A new, more general scientific definition is established. Note, it no longer matters which organism.]

1993 D. SHAY & J. DUNCAN *Making of Jurassic Park* p. ix
"The research seems to echo Jurassic Park*, the novel about scientists who bring dinosaurs back to life by cloning their DNA."*

2000 *U.S. News & World Rep.* 20 Mar. 52/2
"With the biotech revolution of the 1960s and 1970s, scientists cloned the relevant gene and inserted that DNA into nonhuman cells."


Public Domain

Dolly the sheep, 1998

What about famous cloned animals such as Dolly the sheep and other various well-publicized attempts to clone whole mammals from single cells? These will be discussed in subsequent lectures. (See the discussion of a cloned calico cat in lecture 10, as well as the complexities of environment and gene regulation in lectures 6 and 12.) The technical challenges are many and fascinating. (Again for plants, this is relatively easy and has been for decades.)

So how about Mendel's peas. Were they clones of each other? That is, were all the purple-flower peas in his original "true breeding" packets of seeds clones of each other? If Mendel (who by the way did not know the word "clone") had chosen a different plant to work with—one that had a history of being propagated asexually, such as most domesticated fruit trees—then yes, he could have been working with clones! However, peas generate quickly by sexual reproduction (a major reason Mendel chose them) and are not commonly propagated by asexual manipulations by breeders. Therefore, although Mendel would have been fascinated to learn all about cloning (if he could time-travel to the present), his own plants were "just" true breeders.

So to reiterate, Mendel's first crosses confirmed the well-established true-breeding nature of his plants. What he did next was outbreeding (outcrossing, crossbreeding) in as many ways as he could imagine, keeping careful notes of the outcomes.

FOR GREATER UNDERSTANDING

Questions

1. What is true breeding and what are some examples?
2. Why does it matter sometimes whether a genetic subject is a plant or an animal?
3. What are some of the early uses of the word “clone”?

Suggested Reading

Schwartz, James. *In Pursuit of the Gene: From Darwin to DNA*. Cambridge, MA: Harvard University Press, 2008.

Other Books of Interest

Mawer, Simon. *Gregor Mendel: Planting the Seeds of Genetics*. New York: Harry N. Abrams, Inc., 2006.

Websites to Visit

1. The Compact Oxford English Dictionary is available free online — http://www.askoxford.com
2. The full version of the *Oxford English Dictionary* is available by subscription online — http://www.oed.com

Lecture 3:
What Did Mendel Do Next?

The Suggested Reading for this lecture is James Schwartz's *In Pursuit of the Gene: From Darwin to DNA*.

Crossbreeding or Outcrossing

Simply put, Mendel crossbred or outcrossed his plants and monitored the results carefully.

Let's continue perusing the *Oxford English Dictionary* for selected definitions to see that Mendel's next set of experiments with his pea varieties were also straight from the annals of plant and animal breeding practices. In this case, I will concatenate the *OED* historical records for various forms of two relevant, interchangeable concepts: outcross and crossbred to produce one chronological record of use. All together it reads like an outline for the history of practical genetics. Here and there I've added annotations. Try looking up the words *hybrid* and *hybridization* yourself to get a chronology of their rich history.

Notice the mixture of references to humans and to their domestic animals. Notice too the many familiar words that come up in the definitions and descriptions, such as "mongrel," "blood," "race," and "hybrid." Feel free to look up those words and more in the *OED* to get a more complete picture.

Crossbreed |'krôs,brēd|
To breed across the lines which separate varieties or races; to breed (animals or plants) from individuals of different species or races. Hence **cross-breeding** *vbl. n.* A breed of animals (or plants) produced by crossing; a mongrel or hybrid breed.

Outcross |ˌout'krôs|
An introduction of unrelated breeding stock to an established, usually somewhat inbred or homogeneous, line; a mating between unrelated individuals; the result or offspring of such a mating.

Hybrid |'hī,brid|
The offspring of two animals or plants of different species, or (less strictly) varieties; a half-breed, cross-breed, or mongrel.

And here are a few selected uses of crossbreed and outcross to show you their antiquity, with commentary.

1675 WYCHERLEY *Country Wife* II. i,
"They are come to think cross breeding for themselves best, as well as for their dogs and horses."

1774 WILKES *Corr.* (1805) IV. 185
"The family of monsieur Louvet . . . emigrated to England; and made a cross-breed with those who [etc.]."

1882 J. MACDONALD & J. SINCLAIR *Hist. Polled Aberdeen or Angus Cattle* 259 In President 3rd 246,

". . . a very judicious outcross was taken. This animal had in his veins an almost equal proportion of Keillor and Ardovie blood."

1844 DISRAELI *Coningsby* III. v,

"It seems to me a barren thing, this Conservatism, an unhappy cross-breed; the mule of politics that engenders nothing."

Here Disraeli refers to a mule that is an offspring (unable itself to reproduce) of an extreme crossbreed between a horse and a donkey.

1918 *Genetics* **3** 475

"When double-throwing Matthiola is used as egg parent in an outcross to ordinary singles, half the offspring receive a factor for doubleness."

Wow, there's a bit of very early technical jargon from a genetics research paper. As scientific jargon often does, this suffers from having been coined before a full understanding of the phenomena had been settled. *Matthiola* is a plant. Double and single here refer to shapes of chromosomes as viewed under the microscope. Chromosomes were suspected to contain the hereditary material (and they do) but in 1918 nobody knew how that worked.

1928 D.F. JONES *Selective Fertilization* vi. 117

"The chromosome situation in these flies is not known, and it is not clear what the reduced fertility in outcrosses is due to."

At this point in the history of genetics, the fruit fly *Drosophila* was well on its way to being a favorite model organism for geneticists. There will be more about fruit flies in subsequent lectures.

Mendel's Methods

Here is some notation we will be using to sort this out as Gregor Mendel did with his outcrosses. I will be using a deck of cards analogy:

Most of the organisms (animals and plants) to be described in this course have two of each of their genes. Bacteria presented only briefly in these lectures are an exception. For animals and plants, one of each gene pair came from its mother, one from its father. Therefore, the genes of a true-breeding pea plant will be shown here as pairs. Simple genetics problems often symbolize genes with a single letter code, either upper or lower case. (That means for the twenty-six-letter alphabet we can describe fifty-four genes with a single letter code. That isn't nearly enough, but it's a start. Humans and most other animals have about twenty-five thousand genes and several to many versions of each of those genes.)

Here are some of the genes of a true-breeding organism such as a pea plant or a black Labrador retriever:

GG ff PP TT rr ww QQ NN vv ll xx ZZ…

The dots at the end are important. They imply that all the rest of the twenty-five thousand genes continue along in the same manner. Note that no matter which symbol we are using (whether uppercase QQ or lower case ff) the gene pairs are identical to each other because the organism is true breeding.

If we perform a true breeding cross (of identical parents) it looks like this:

Parents:

GGffPPTTrrwwQQNNvvllxxZZ… X GGffPPTTrrwwQQNNvvllxxZZ…

Offspring:

GGffPPTTrrwwQQNNvvllxxZZ…

Note the genetically identical parents and identical offspring.

Now here is a technicality—a sort of dichotomy between the ideal and the practical. Think about those black Labrador retrievers once more. If we were trying to understand just a few of their genes (and not the entire set of twenty-five thousand) and if those genes just happened to be some of the genes represented as letters (GGffPP…) in the cross above (with each gene pair identical), then *for those genes*, the dogs *would* be true breeding. Most of genetics is practiced on a practical, functional level with just a few genes at a time, not the entire idealized set. In the case of the black Labradors, the dog breeder might think of the pedigree dogs as being true breeding just for the specific sets of genes for black coat color and a big chunky body and head and other pedigree Lab traits. If we were to look at the entire set of dog genes, we would find that many were not true breeding. It is impossible to mate identical twin dogs, as they would have to be of the same gender.

For the majority of problems that I will be presenting, we will consider one gene pair or at most two gene pairs at a time, as Mendel did. However, now and then I will remind you that there are in fact twenty-five thousand genes. It is our choice to simplify down to one or two genes. In doing so we are not declaring that genes are simple; however, by simplifying, we are getting a glimpse at some basic rules for how genes function. That we deliberately are choosing to simplify (to one or two genes at a time) may seem too obvious to even say, but I think it is the source of a common misconception about genetics. If you read about a new scientific development that sounds like this: "They just discovered *the gene* for autism" or "They are looking for *the gene* for homosexuality," then you are encountering the misconception that genes function as simple, isolated individual entities and can be understood completely or sufficiently that way. In subsequent lectures we will return to thinking about many genes at once with intricate, convoluted interactions.

Meanwhile, here is a simplified true-breeding cross for a dog:

Let's say that all individuals bearing the genes BB (one from each of their parents) all have black coat color and let's say the bb animals all have brown coat colors.

BB x BB (two black dogs)
BB (black offspring)

Here is another.

bb x bb (two brown dogs)
bb (brown offspring)

So the two true-breeding crosses show that black dogs beget black dogs, brown dogs beget brown dogs, and so on forever and ever until we decide to outcross black with brown. Let's do that:

BB x bb
black x brown
Bb
All offspring are black

Notice that we need just one copy of that "B" gene to get a black dog. That's what Mendel noticed too, except that he was working with pea colors and other pea traits. Sometimes a gene functions just fine with one copy, as with "B" for black, and sometimes two copies are needed, as in "bb" for brown.

Here is what did *not* happen: The two colors did not mix like paint, which was a common misconception about heredity previous to Mendel's analyses. We did not get a blackish-brown color. We got black.

Card dealing is a useful analogy here. Each dog was holding two cards in its hand (or paw.) Each dog contributed one card to the future offspring. For this out cross, there was not a lot of choice as to which card each dog contributed.

To continue the card-playing analogy, clearly these cards are not quite equal. In fact, one seems to trump the other. If we have "B," it doesn't seem to matter what that second card is. We get black.

Indeed, there are three combinations of gene pairs we can observe from this round of play:

BB black
Bb black
bb brown

Rather than use the word "trump," geneticists use the terms *dominant* for "B" and *recessive* for "b." A dominant gene is one that needs only one copy in order to cause a particular trait (blackness). A recessive gene needs two copies to show its trait (brownness).

Surely somebody had noticed this before, that inherited traits sorted like playing cards and did not mix like paint. After all, we have thousands of years of serious dog and horse breeding before Mendel's time. I think it must have been noticed over and over, but apparently nobody had thought to apply a set of symbols by which nearly every possible combination could be figured out and the amazing regularity revealed. (And recall too that dog and horse breeders were and continue to be generally uninterested in examining exactly one trivial trait at a time). To confound things a bit, not all traits sort themselves out as crisply and obviously as the ones in this example. Mendel may also have been fortunate in his choice of organism and choice of traits. In fact, much later Mendel continued his investigations with different plants and different traits and was considerably less successful.

Mendel must have been eager to try out every possible combination of crosses that he could using just this pair of genes. The symbols applied here either to peas or dogs; it doesn't matter.

Here are all the combinations that are not just true breeding:

BB x Bb
Offspring could be either BB or Bb

bb x Bb
Offspring could be either Bb or bb

Bb x Bb
Offspring could be either BB or Bb or bb

Recall we are dealing out these genes as though they were cards that would make up a new "hand." The offspring can be one or another of all the possible combinations that could be made from the parents' hands.

Just as in card playing, it is fun to know the odds of getting those combinations. If I am dealing out cards blindfolded and my fellow card player is doing the same, what are the odds (or probabilities or chances) of us each dealing a "B" card such that we get a "BB" pairing? And what are the odds of our cards producing a mismatched pair "Bb." It is convenient to set up the problem on a Punnett square (named for Reginald Punnett, who devised the method in the early twentieth century).

BB x Bb yields all black puppies

	B	b
B	BB black	Bb black

Bb x bb yields a 1:1 ratio of black to brown puppies.

That means there is a one-half chance of getting a black or a brown puppy.

	B	b
b	Bb black	bb brown

Bb X Bb yields a 3:1 ratio of black to brown puppies.

That means there is one-quarter chance brown and a three-quarter chance of black.

	B	b
B	BB black	Bb black
b	Bb black	bb brown

So does that mean that in a litter of four puppies with parents Bb and Bb that I will get exactly one brown and three black? No, it does not mean that. Those are ideal ratios that one might see if the dogs were to have dozens to hundreds of offspring. Each puppy represents a sort of independent hand of cards. Notice this phenomenon in families. There is a 50 percent chance of getting either a boy or a girl and yet plenty of families do not show a 1:1 ratio of boys and girls. It is entirely possible to have a lucky run of six boys or eight girls in a row. This is an important concept for genetic counselors explaining to parents the probability of some deleterious gene being passed to offspring. Take Tay Sachs disease as an example.

TT	normal recycling of wastes from neurons
Tt	carrying but not showing the trait of failing to recycle wastes
tt	Tay Sachs disease: a progressive, incurable deterioration of muscles and nerves due to a failure to recycle wastes in neurons, resulting in death by the age of three

Let's say the parents (Tt and Tt) have already three normal children (either TT or Tt) and are discussing plans for a fourth child with their genetic counselor.

Are they "due" for a child with Tay Sachs (tt) after that run of good fortune with their first three? No, the probability is exactly the same as with their first child. Likewise, parents who have already lost three children in a row to Tay Sachs are not due to have a TT or Tt child. One of the reasons this can be a challenge to explain is because the popular misconceptions about probabilities are pervasive. For example, many people do consider themselves "due" for a win, after a string of losses. How about you? Answer these two questions honestly, to find out whether you are vulnerable to magical thinking about numbers:

1. Which of these lottery numbers is more likely to be a winner: 2222 or 7569?
2. If 7569 won a million dollars yesterday, is it therefore unlikely to win today?

Answers: Both 2222 and 7569 are equally likely (or unlikely) to win. And 7569 is as likely (or unlikely) to win today as it was yesterday. If you got those wrong, you are in good company with billions of people. Mistakes of this sort about numbers seem to be deeply ingrained.

Mendelian Ratios in Nature

Can you see 1:1 or 3:1 ratios of particular traits in natural populations of organisms? For reasons that will be described in lectures 6 and 7 about environment, it is not typical to be able to do so. Rather, if you can spot ratios (by doing lots of counting and evaluating of a trait) they are likely to be highly skewed away from Mendelian ideals. But they will be interesting nonetheless. Here are some suggestions for investigations.

1. Look through a guidebook to wild flowers in your area, watching for ones that are described as varying in color, typically in the range of blues-pinks-whites (rather than yellows and oranges). Then set out to find a large stand of your

plant and see if you notice any variations. If you do not, you may have found a true-breeding population for that trait or maybe there is some environmental reason for only one color. If you do, have fun counting and recording and seeing whether any interesting ratios emerge.

2. Watch for reports of unusual albino or melanistic mutants of wild animals, such as white deer or black squirrels. Sometimes there will be just one sighting. However, there are interesting cases of populations with exceptionally high ratios of color mutations, for example, the population of black squirrels in some areas of northeastern United States. If you are in the right area, try counting them at your bird feeder.
3. Go to the grocery store during harvest season and purchase colorful, decorative corn. These are not from natural populations and therefore are likely to display Mendelian ratios or some variation of a more complex ratio. The more decorative the corn, the more likely it is to have genetic ratios that are more complex than 1:1 and 3:1. Expect them especially in ears with three or four or even more different colors of kernels. Sometimes, though, you can analyze an ear of corn such that you ignore subtle variations in color but just declare everything to be either pigmented (all of the reds, purples, blues) or not pigmented (all of the yellows and whites). That is, do as Mendel did, ignore everything except for one trait at a time.

Here in brief are a few somewhat advanced additional topics, typically part of a genetics course. You may want to consider these further, once you are quite comfortable with more basic principles:

Dominant and recessive relationships between genes can be complicated. Here are some possibilities:

The variations of genes are called “alleles.” Each gene has many potential alleles, way too many to account for with a system like “B,” “b,” which is why we are resorting to superscripts for some of these. Note that in any given organism, we have only two alleles, one from the mother, one from the father. It is when we consider a population of organisms that we need to bring in more complex notation to name all the different possibilities.

Bb	Simple dominance. The “B” trait is expressed as in “black.”
Bb	Dosage dependent dominance. We have just one copy of B, so the animal has a somewhat lighter coat color than it would if it were BB.
B^1B^2	Co-dominance, which is more likely to appear in a structure. It is almost as though B^1 and B^2 were two different types of bricks being used equally to build a wall. Structures attached to the outside of red blood cells are often co-dominant. An example is the AB blood group, where the two co-dominant alleles are I^AI^B.
B^1B^2	B^1 is dominant to B^2.
B^2B^3	B^2 is no longer recessive because the relationships are relative. In this pairing, B^2 is dominant to B^3.

I have not yet provided a definition for "gene." That will come later. It turns out that throughout his career as our official first geneticist, Mendel did not know what a gene was. Indeed, throughout the first part of the twentieth century, geneticists worked on genes as concepts, but did not know what they were. We will follow that same trajectory and in the first few lectures treat "genes" as though they were some sort of "hereditary information" with many as yet unknown properties and functions. Meanwhile, here are two more useful terms:

Genotypes: A genotype is the string of symbols representing all of the genes under consideration for an organism, such as:

Bb

or:

GGffPPTTrrwwQQNNvvllxxZZ…

Phenotypes: Phenotypes are the description (as long and detailed as they need to be) of the traits associated with a particular genotype:

"black"

or:

"long tailed, floppy eared, short snout, brindle coat"

It is possible to analyze two genes at a time, three genes at a time and more, except that each increment becomes much more complicated with potential combinations. Such analyses (in search of Mendelian ratios) work best when genes are acting as independently and separately as individual cards in a deck. However, there are some circumstances in which the genes are linked together (positioned right next to each other) and therefore tend to be sorted out as though they were like playing cards glued together. That is, they do not sort independently. Mendel didn't know about that and he would have been confused if he had encountered it. For this next example, we'll consider two genes that are independent and provide you with some more ratios to watch for.

Let's try various crosses involving two traits relevant to the flavor and texture of peas.

AA or Aa sweet

BB or Bb early

aa bland

bb late

AaBb (sweet early pea) x AABb (sweet early pea)

Notice the result (next page) is a 6:2, which reduces to a 3:1 ratio of sweet-early to sweet-late peas.

	AB	Ab	aB	ab
AB	AABB sweet-early	AABb sweet-early	AaBB sweet-early	AaBb sweet-early
Ab	AABb sweet-early	AAbb sweet-late	AaBb sweet-early	Aabb sweet-late

Try AaBb x AaBb (both are sweet early peas). The result is a 9:3:3:1 ratio of sweet-early (the nine not labeled below): bland-early; sweet-late; bland-late. It is one of the more complex ratios from a two-gene cross.

	AB	Ab	aB	ab
AB	AABB	AABb	AaBB	AaBb
Ab	AABb	AAbb sweet-late	AaBb	Aabb sweet-late
aB	AaBB	AaBb	aaBB bland-early	aaBb bland-early
ab	AaBb	Aabb sweet-late	aaBb bland-early	aabb bland-late

FOR GREATER UNDERSTANDING

Questions

1. What are examples of outcrossing, especially ones you can observe yourself?
2. What is an allele and what makes an allele either dominant or recessive?

Suggested Reading

Schwartz, James. *In Pursuit of the Gene: From Darwin to DNA*. Cambridge, MA: Harvard University Press, 2008.

Other Books of Interest

Mawer, Simon. *Gregor Mendel: Planting the Seeds of Genetics*. New York: Harry N. Abrams, Inc., 2006.

Lecture 4:
A Cookie Factory in a Black Box Designed by Rube Goldberg

The Suggested Reading for this lecture is James Schwartz's *In Pursuit of the Gene: From Darwin to DNA*.

This lecture introduces an extensive metaphor that I believe will help you to envision some of the more complex genetics processes that go beyond simple Mendelian ones. In subsequent lectures I will say something like, "recall that cookie factory from lecture 4" to transport you briefly back to being able to visualize it, and then I will layer on some bit of genetics. So please, even though the relevance may not seem perfectly apparent now, read through and prepare to use this as a sort of memorization system. It is similar to a "method loci," which is a classical mnemonic system sometimes based on visualizing a temple or palace in order to keep track of a large number of items and relationships.

One of the main reasons that Mendel and other early geneticists found so many confusing exceptions to their ratios (such as the 3:1 ratio) is because most genes can not be examined in isolation from their relationships or interactions with other genes. Those interactions make up the functions (the chemical reactions) of cells and therefore of organisms. Let's say there is an automated cookie factory with assembly lines arranged in a logical order like the following:

1. A dough-mixing machine for combining flour, sugar, butter, and so on.
2. The first fork in the path: some cookies receive chocolate chips, some do not.
3. A shaping machine by which cookies are formed from the dough.
4. A baking machine.
5. A cooling machine.
6. A second fork in the path: sorting machine (to remove broken and burned cookies).
7. A third fork in the path: some cookies receive frosting, some do not.
8. A counting machine to set out the right number of cookies for each package.
9. A packaging machine.
10. A labeling machine.

Each of the ten steps require at least one machine. Each of the machines requires its own little set of instructions for assembly and function.

And now, let's put it all into a black box so that you cannot see what is going on inside. You can only look at the output: cookies of one sort or another. You wait patiently and observantly for some little mistake to be made inside the factory and then you diligently record exactly what happened to the cookies.

Here is how your analysis goes if you are trying to decipher this as a geneticist would.

Some mistakes (mutations to the instructions for the machines) seem trivial and have little effect on the important aspect of the cookies (their edibility and deliciousness). For example:

- The labeling machine spells cookie wrong: “bookie.”
- There are three cookies in a package instead of two.

Other mistakes seem to cause a complete absence of an expected cookie (but we do get a “default” cookie). For example:

- There are no chocolate chip cookies at all, just those without chips.
- There are no frosted cookies at all, only unfrosted.

Still other mistakes have dramatic effects. For example:

- The cookies are raw and therefore cannot be sorted by the sorting machine, cannot be frosted, cannot be counted, cannot be packaged. Only horribly deformed and inedible cookie dough pours out of the factory.
- No cookies emerge at all. Could there be something wrong with the dough mixer? Perhaps no ingredients were measured in.

Things to notice:

- The hierarchy of seriousness of the mutations has something to do with the order of events. Certain early-acting machines can have a cascading effect on the subsequent machines. Some early-acting machines can even prevent the entire pathway from functioning.
- At the places where paths fork, we can have mutations that eliminate one possibility but not the other (frosted versus not-frosted).
- Events at the end of this pathway seem to be less important to the overall product.

This is what geneticists think about—metaphorically—when they are using the results of mutations to try to understand and reconstruct the order of events of a hidden cell pathway. In addition to watching for single mistakes (mutations) they also watch for combinations of mutations (double mutations and triple mutations) and use those to decipher the pathway.

Notice that some double or triple mutations might have interesting and not especially disastrous results. How about cookies that are miscounted *and* mislabeled?

However, any double or triple mutation that involves a broken dough mixer may not be observable at all. If one combines a broken dough mixer with a broken cookie counter, the effects of the broken counter disappear.

Over the years, geneticists have come up with a multitude of jargon by which to analyze pathways hidden in (metaphoric) black boxes by examining the end-products. I think having a good picture of an assembly-line analysis as I've just described is useful and will help you with subsequent lectures. Memorizing a lot of jargon that goes with such an analysis is less useful. However, I will introduce the jargon and etymologies here for another purpose: to demystify and deconstruct some of the struggles and challenges that geneticists have had and continue to have in their analyses.

Epistasis

When a mutation at the beginning of a pathway completely obliterates your view of any subsequent parts of the pathway, we call that epistasis. If we observe such a mutation (like uncooked dough), we declare it to be important and early in the pathway we are deciphering. The etymology of the word "stands upon" implies the early confusion as to what was going on with epistatic mutations. I still have my undergraduate genetics text book, written in 1969. Epistatic genes are described there as "covering up" or "hiding" (or standing upon) other genes, but there is not a word in that chapter about understanding genes as parts of pathways. It was still too early in 1969 for that sort of analysis to emerge from the black box.

Modifier

The frosting machine and the chocolate chip machine are "modifiers." Notice that they most likely are parts of their own pathways that we have not described. (That is, there must be a frosting-making machine somewhere.) But from the point-of-view of the main pathway, the frosting machine is just adding a little extra to a cookie that otherwise will turn out just fine. So we use this word "modifier" for those genes that (in our judgement) cause seemingly optional, and only slightly different variations.

Pleiotropy

A single mistake may have far-ranging and diverse consequences. How many different cookie types are we getting from the factory? I count at least six, each pertaining to a particular branching section of the pathway:

Unfrosted, no chocolate chips
Unfrosted, chocolate chips
Frosted, no chocolate chips
Frosted, chocolate chips
Broken without chocolate chips
Broken with chocolate chips

A geneticist would analyze this by looking for how many different results a particular pathway is producing. If there is more than one result we call that pleiotropy and we assume that branching pathways are the cause of it. And of course we can use epistasis to figure out whether some of those branching nodes are more important (near the beginning of the pathway) than others.

An example would be the chocolate chip machine node. If that machine is broken, no chips will appear in any subsequent cookie regardless of the cookies' other attributes; for example, frosting or wholeness.

Penetrance

And now we are entering the realm of "weasel words" that are used when geneticists are not exactly sure what is going on in the black box, but are pretty sure it's complicated and needs to be described somehow. Let's say I expect a particular machine to behave a certain way and every now and then it doesn't. Let's say that 10 percent of the time, the cookie-wrapping machine fails to wrap some cookies. I'd love to know why. It's probably something fascinating that will give me a more complete understanding of the cookie factory. However, right now I am forced to declare the cookie-wrapping machine to be "90 percent penetrant," which means it wraps cookies about 90 percent of the time. Interestingly, this word gets used most often for important genetic traits, the ones that are most difficult to sort out. A geneticist might say that a particular birth defect such as polydactyly (extra digits on fingers or toes) is incompletely penetrant. That is, a certain percentage of people who have the mutated gene that causes extra digits will have normal numbers of digits.

Expressivity

This is another vague word, used when we do not have a full understanding of a pathway. Let's say the cookie-shaping machine sometimes made extra-large cookies and sometimes tiny cookies, in addition to those of normal size. If we could figure it out, we would have a better understanding of the cookie-shaping function. Meanwhile, we are obliged to say that it has "variable expressivity." It turns out that polydactyly also may be used as an example of variable expressivity. Sometimes it affects one hand, sometimes both, sometimes just one foot, and sometimes both.

Predisposition

This is a useful word often in conjunction with penetrance and expressivity. For example, breast cancer is complex. If I think I have identified a mutation in one of many genes in the complex pathway by which breast tissue develops, I might say I have found a putative breast cancer allele. People who have the allele are most likely not guaranteed to get breast cancer. However, I could say they are predisposed to it. How predisposed? That's where I would bring in an analysis of penetrance and expressivity if I had enough data to do so.

Cascade

Biologists sometimes use this word evocative of a multilevel tumbling waterfall to describe complex multibranching pathways.

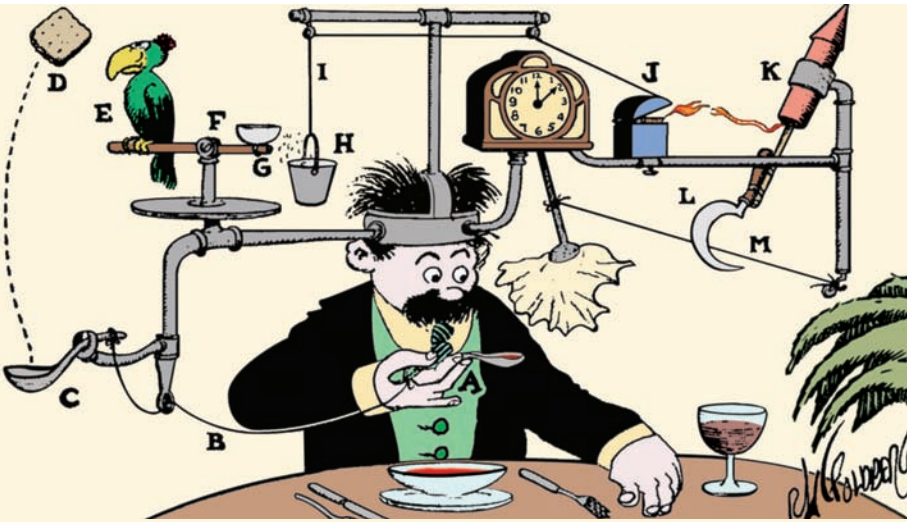
Alleles

There are many possible different ways to modify (including break) a machine. Those are all of its different variants or alleles.

Environment

This is so important that it will be addressed in lectures 6 and 7. It is important to never forget that genes operate within the contexts of environments.

The factory may seem complex enough at this point, but there is one more layer of detail to add. Let's say this factory was designed by Rube Goldberg (or Heath Robinson, or Dr. Seuss) using baling wire, duct tape, clothespins, and other odds and ends from their inventor's workshop. In spite of the seemingly linear (or cascading) logic described above, the factory is actually quite convoluted, loaded with redundant processes, baroque functions, and seemingly nonfunctional (at least for cookie-baking) cowbells and horns. And this is actually a truer picture of any pathway in any cell. The picture is wildly nonlinear, convoluted, and seemingly pieced together from unrelated items as though from the corner of an eccentric inventor's basement. This is what geneticists must deal with when they set about to decipher genes. As usual (in the tradition of Mendel) the most important first step is to simplify, to look at a few parts at a time and to assume none of it has any sort of logically designed, carefully, linearly assembled structure.



Professor Goldberg walked in his sleep, strolled through a cactus field in his bare feet, and screamed out an idea for a self-operating napkin:

As you raise spoon of soup (A) to your mouth it pulls string (B), thereby jerking ladle (C) which throws cracker (D) past parrot (E). Parrot jumps after cracker and perch (F) tilts, upsetting seeds (G) into pail (H). Extra weight in pail pulls cord (I), which opens and lights automatic cigar lighter (J), setting off sky-rocket (K) which causes sickle (L) to cut string (M) and allow pendulum with attached napkin to swing back and forth thereby wiping off your chin. After the meal, substitute a harmonica for the napkin and you'll be able to entertain the guests with a little music.

Rube Goldberg (roob gold'berg) n. a comically involved, complicated invention, laboriously contrived to perform a simple operation—*Webster's New World Dictionary*

FOR GREATER UNDERSTANDING

Questions

1. Can you explain the cookie factory metaphor to someone and use it to explain genes and mutations in the context of pathways?
2. What is epistasis?
3. What are modifiers?

Suggested Reading

Schwartz, James. *In Pursuit of the Gene: From Darwin to DNA*. Cambridge, MA: Harvard University Press, 2008.

Other Books of Interest

Goldberg, Rube. *Rube Goldberg: Inventions!* New York: Simon & Schuster, 2000.

Robinson, Heath. *Contraptions*. New York: The Overlook Press, 2007.

Website of Interest

Rube Goldberg, Inc. provides a website dedicated to the late engineer-turned-cartoonist — http://www.rubegoldberg.com

Lecture 5:
How Mammals Get Their Colors

The Suggested Reading for this lecture is Juliet Clutton-Brock's *A Natural History of Domesticated Mammals*.

All mammals are very closely related (being relatively recently evolved) and all use the same complex pathway to make their melanin pigments by which their hair (or coats) and skins are colored. The pathway that I will describe here was worked out mostly for mouse coat colors and dog coat colors, but pertains to any mammalian pigmentation, including humans. A major difference is that human pigmentation is mostly quite limited to a range of dark brown pigments to pale pigments. Humans evolved such that they are not using the full potential of the melanin pathway; for example, humans tend not to come in stripes and spots and multicolors (or if they do, it is declared to be an exceptional genetic condition such as Waardenburg syndrome, in which patches of white appear against a background of darker hair). Note too that many wild animals (whether they have stripes or spots or just plain grayish-brown coats) also are quite limited in their normal range of pigments.

In contrast dog breeders and fancy mouse breeders have delighted in achieving as wide a range of colors and combinations as possible. Therefore, to see the full capabilities of the melanin pigment pathway we turn to fancy mice and dogs.

Recall the cookie factory. I am going to lead you step by step through a set of activities and functions just as I did with cookie making. Only here it will be the relatively unfamiliar melanin pigment-making pathway. The goal is not to memorize the pathway, but rather it is to

- appreciate its intricacy
- see the possibilities for how it was analyzed (as we did in the hypothetical analysis of the cookie factory)
- be intrigued by the connectedness of the melanin pathway to other pathways
- get a little glimpse of the potential effects of environment
- never look at dog (or cat or other mammalian) coat colors the same way again

The problem of exactly where to begin and where to end and which other auxiliary pathways and branching pathways to include is difficult. That is intrinsic to any biological pathway. I will make a choice to begin arbitrarily right outside of a specialized cell called a "melanocyte" that is destined (or not) to become filled with the pigment melanin. This is a simplified pathway, but it features all nine of the major pigment genes that are fairly well understood. In case you'd like to read more (for example, at a genetics for dog breeders website), I am using the typical letter symbols (K, A, E, C, B, and D, S, T, and M) used for genes of the melanin pathway.

1. The "K" gene product binds (if it can) to the outside of the melanocyte (the specialized cell that is destined [or not] to become full of pigment). If K *does* bind, we will be on our way to dark pigmentation—producing the dark pigment "eumelanin."
2. But here is our first branch. There is another gene product "A." Some forms (some alleles) of A prevent K from binding. If that happens, we are on our way to making various degrees of pale pigmentation—producing the pigment "phaeomelanin."
3. But actually *none* of this happens without the gene product "E," which forms the place on the melanocyte where K would like to bind.

You might be noticing something right about now. There is a sort of nursery rhyme or even Dr. Seuss-like quality to the description. It is very difficult to put it all into a neat straight line. Indeed, very little in the cell is linear, no matter how hard we might try to depict it as linear. In this sense, the cookie factory metaphor is much simpler.

4. Okay, so let's say we made it through those first three nonlinear steps and the following happened: K and E both were functioning and therefore K did bind to E and, by the way, the form of "A" that was available did not prevent this from happening.

Were there any other options? Sure, lots. E might have been nonfunctional, not allowing E to bind. Or K might have been nonfunctional, unable to bind E. Or A might have prevented K from binding E. We have scarcely begun and we already have lots of possibilities.

5. Just inside of the cell (the one on which K just successfully bound) is a gene product called "C" waiting for the signal to go ahead with its function. Successful binding by K is that signal. C triggers a multibranching pathway. We'll just look at two of the branches to start with. One version of C is very efficient at converting the chemical tyrosine to dopaquinone. If it does, we are well on our way to making the dark pigment eumelanin with whatever B genes are available: If "B," we'll get black; if "b," we'll get brown.
6. Furthermore, we can modify that eumelanin (made by "B" or "b") a bit. Recall modifiers do not dramatically change a product but just cause little conversions and tweaks for slightly different products. The "D" genes do that modification, producing either a richer or more diluted version of black or brown.

There was another completely different option, though. There are many versions of C, some of which send us down the other fork in the path to making the paler pigment phaeomelanin.

So where are we so far?

In order to get a completely black Labrador retriever dog, I need a minimum of the following:

- K bindng to E
- E allowing K to bind
- the lower-case version (recessive) allele "a," which does not interfere with K and E

- C, which allows reactions to move strongly ahead to making eumelanin
- B (the upper case, dominant version) makes black eumelanin pigment (if b, it would be brown)
- D helps make full, rich, undiluted black eumelanin

Now notice this. Almost *any* variation to the pathway will result in a lack of black eumelanin and the possibility of defaulting to the paler pigmentation pathway. There is one major branch of the pathway to get "black" and several different modifications to get all of the non-black (paler) pigmentations. Take, for example, the following potential modifications, taken individually:

1. "A" (upper case, dominant A) interferes with binding. This is one way to get a yellowish-colored dog. (There is actually a range of varieties of A and, depending upon which are present, there will be a range of yellowish colorations or markings against a darker background), or
2. "e" (lower case e) does not allow K to bind. We get a cream colored dog, or
3. "c" (almost any version of "C" besides the maximum-strength upper case C version) gives us a range of pale coat colors, often seen in Siamese cats.

Consider what I explained in the first two lectures of this course about true breeding and outbreeding:

A purebred black Labrador retriever that is true breeding for its rich black coat color (and most likely true breeding for a host of other black Lab traits) will have the following genes in double copies:

KK EE aa CC BB DD

However, if the dog is not exactly truebreeding for black coat color but rather carries a gene for brown (or chocolate, as Labrador breeders call it), it will have the following genes:

KK EE aa CC Bb DD

Let's isolate the only pair of genes with any possibility of giving us variation in color and ignore the rest:

Bb

And next let's mate two dogs with the Bb combination:

Bb x Bb

	B	b
B	BB black	Bb black
b	Bb black	bb brown

And that is how we get a 3:1 ratio.

How about a cream- or yellow-colored retriever? There are lots of ways to do that by deviating just a bit from the black pathway. In fact, for some breeds of yellowish or cream-colored dogs it is not yet known exactly which gene is causing the deviation from black to a pale color.

Here are a few of the possibilities varying just one set of genes at a time:

- KK EE aa CC BB DD: black
- KK ee aa CC BB DD: pale color
- KK EE aa cc BB DD: pale color
- KK EE AA CC BB DD: pale (often yellow) color

But we are not done yet. I have not explained patches of white, including spots and stripes. For that we need to consider the packaging gene "M" and transport genes "S" and "T" for moving pigments. Again, it is difficult to put these into a simple linear pathway. In a sense, M, S, and T are more important than K, E, A, C, B, and D in that we will not be able to distribute pigment all over the animal if we cannot package and transport the pigment uniformly. The results of nonuniform packaging and transport are fascinating. It is how we get striping and spotting and in some cases completely white coat colors. The crucial activities take place very early in the development of the embryonic mammal, when it has just a few thousand cells. Along the future backbone of the embryonic animal, melanocytes either produce eumelanin or phaeomelanin (according to the genes present). Here is what happens next.

1. The recessive "m" gene product packages the pigment into specialized compartments in the cell. Alternatively, it does not package the pigment quite right. If we have one copy of upper case "M," we will get some pale cells (with incomplete packaging), even if the pigment pathway is producing black. In addition, we may get odd pigmentations such as one blue eye and one brown and in some cases a tendency to deafness in one or both ears. If we have two copies of M, it is lethal. Clearly, something pretty dramatic is going on with "M." This is a good example of another important, complex pathway interacting profoundly with the pathway we are examining. Why is packaging of pigments so important and with such far-reaching effects, including deafness and lethality? (We'll get back to that.)
2. Meanwhile "S" and "T" get to work and cause those melanocytes to migrate all over the body of the developing embryo, to one degree or another. Double copies "SS" efficiently transport pigment all over the body. Solid black Labs (or any completely solidly colored animal) are SS. There are lots of alternative, recessive versions of the transport gene "S." It does not suffice to call them all "s." Instead, we need to resort to superscripts, which is typical in genetic nomenclature. Here is a series of some of four variants (alleles) of "s" and their consequences (phenotypes).

S	solid-colored animal
s^{i}	Irish markings: white on feet, legs, and chest
s^{p}	piebald spots
s^{w}	extreme white (range of completely white to one or two spots of black, including just one bullseye-like patch around the eye)

A Gypsy Vanner Horse

The Gypsy Vanner horse is the preferred wagon and riding horse of the Romany Gypsies. While the brightly colored horses with profuse feather, mane, and tail are often called Gypsy Cobs, and Irish Tinkers, the Gypsy Vanner is a designation bestowed upon the "best of the best" of a half century breeding program. The program has the intention of producing the most spectacular and well-bred Gypsy Horse to complement the culture and colorful lifestyle of the Romany Gypsies.

The T genes (also a series of alleles) effect a more subtle pigment migration, with versions of "t" giving some mammals little spots on their noses and paws, called "ticking."

In contrast to the extravagant pigmentations of domestic animals, human evolution has been rather conservative. Most humans are pigmented along a short gradient of pale to dark. Exceptions are notable. The National Center for Biotechnology Information (NCBI) hosts the Online Mendelian Inheritance in Man (OMIM) site. Once there, use the term "melanin" as the search word (see "Website of Interest" at the end of this lecture).

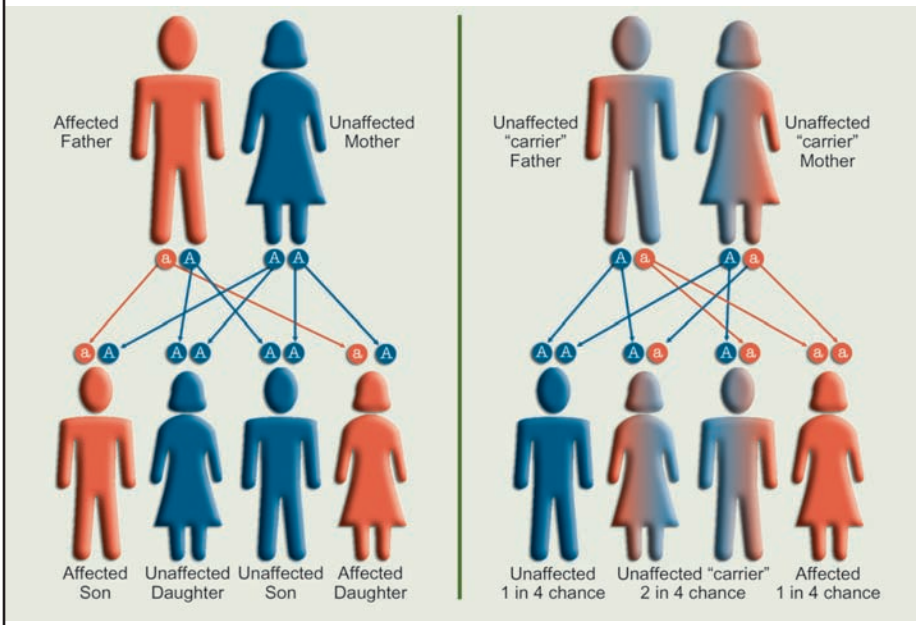
Many of the hits will be for various laboratory mammals, but some will be for humans. Note that OMIM tends to present some of the most extreme examples of a particular gene and therefore some of the descriptions may suggest that pigment anomalies in humans always have very negative consequences. That is not the case at all. However, for those descriptions at OMIM, notice that nervous system disorders often accompany the extreme pigment disorder. This is a good example of pleiotropy in a very complex pathway. On the way to making eumelanin and melanin, there are a few other branches that lead to production of some neurotransmitters and hormones. That is one of the reasons that the "M" packaging gene described above is lethal (or extremely detrimental) in double doses and somewhat detrimental in single doses. The gene "M" does not simply affect color, but also development of the nervous system. In dogs a double dose of the milder alleles of M "mm" results in a coat color called "merle," which may include heterochromia iridis (different colored eyes) and some hearing loss too.

Search for "Waardenburg syndrome" at the OMIM website. That is the human pigment condition in which there may be a blaze of white hair (against darker hair) and there may be "heterochromia iridis," that is, one dark and one light eye. Sometimes considerably more severe traits may accompany the pigmentation, including deafness. Although the details are still being worked out, it appears that Waardenburg syndrome is caused by a version of the "M" alleles. That is, Waardenburg seems to be a version of merle.

Albinism (the condition of having little or no melanin pigmentation) may be searched at OMIM. Note that in humans, it is a genetic condition of some concern, while in many domestic animals (such as laboratory rats) and some arctic wild animals (such as polar bears), a lack of pigment is normal. How does a lack of pigment occur? There are many ways in the long, complicated pathway to derail the production of dark eumelanin. Therefore, there are lots of genes for albinism. That means that two albino humans getting together to have children may not have albinos. It depends on which genes each are carrying.

Let's say that having either aa or bb could result in a deviation from melanin production. One albino individual is AAbb. The other albino is aaBB. They have children: all of them are AaBb. All of them have normal melanin production. The same could happen in a mating of two pale-colored dogs or cats. It is results like that, that reveal the complexity of the pigment pathway.

One major factor that has not been included yet in all of this discussion of genes is the environment. The pigment pathways (and any genetic pathways) occur within a greater context. That will be introduced in the next two lectures.



Left: Waardenburg syndrome is usually inherited in an autosomal (non-sex determining) *dominant* pattern. Right: Types II and IV Waardenburg syndrome may sometimes have an autosomal *recessive* pattern of inheritance.

FOR GREATER UNDERSTANDING

Questions and Activities

1. How does the evolution of human pigmentation compare to that of domestic animals?
2. What is Waardenburg syndrome?
3. Make a little pocket guide (that is, some notes on an index card) for yourself of dog-coat color genotypes and use it to diagnose the possible genotypes of dogs that you meet.

Suggested Reading

Clutton-Brock, Juliet. *A Natural History of Domesticated Mammals*. 2nd ed. Cambridge: Cambridge University Press, 1999.

Other Books of Interest

Coile, D. Caroline. *Encyclopedia of Dog Breeds*. 2nd ed. Hauppauge, NY: Barron's, 1999.

Helgren, J. Anne. *Encyclopedia of Cat Breeds*. 2nd ed. Hauppauge, NY: Barron's, 1997.

Website of Interest

The National Center for Biotechnology Information (NCBI) provides the Online Mendelian Inheritance in Man page; use "melanin" as a search word — http://www.ncbi.nlm.nih.gov/sites/entrez?db=OMIM

Lecture 6:
The Importance of Environment: In the Short Term

The Suggested Reading for this lecture is Michael Rutter's *Genes and Behavior: Nature-Nurture Interplay Explained*.

One of my favorite genome projects is the canine project, because it is accompanied by a wealth of information from hundreds of years of dog breeding. Even an amateur breeder of dogs with no particular knowledge of genetics cannot help but notice the natural tendencies of particular breeds for certain ranges of behavior (both positive and negative) and predispositions to diseases. If you read a detailed description of a particular breed you will find information concerning behavior under the rubric of trainability and temperament. These observations are now being paired with gene sequences and, little by little, the genetics of behavior is being deciphered. Whatever is learned from dogs will be applicable in general to any mammal, including humans.

Dog breeders are usually well aware of the effects of environment on dog behavior. That is, a dog raised in a calm, disciplined, nurturing household will have a different set of behaviors from its identical twin brought up in a chaotic, undisciplined, unpredictable environment. Behavior is not simply genetic; the environmental setting is important. However, dog breeders also know that they are working within a range of natural (genetic) behavioral tendencies. For example, you do not have to teach a puppy from a lineage of pointers to point. It does so on its own. Within that range of natural (genetic) pointing behavior, however, there is plenty of room for a trainer to bring out nuances, such as holding the point until commanded to do something else, pointing to the right game, and pointing at the right distance from the game.

What do geneticists mean by environment? Sometimes they are referring to the quality of air, water, and food, presence of pollutants, and other such external parameters, commonly associated with environment. However, more often, "environment" means the intimate setting in which genes are expressed. That setting includes the body, the various organs and tissues,

© Linn Currie/shutterstock.com

Small Münsterländer on Point

The Small Münsterländer is a hunting-pointing-retrieving dog breed that reached its current form in the area around Münster, Germany. Small Münsterländers bear a resemblance to both spaniels and setters, but are more versatile.

and the cells. It includes other genes and their particular products. The first few cell divisions of a new mammal take place within the influential environment of the uterus. Those early dividing cells have internal environments influenced by whatever conditions were in the newly fertilized egg. Organs, tissues, and cells communicate and coordinate with each other via hormones, neurotransmitters, and other signaling factors. Often the receipt of a signal results in expressing one or another gene. Interactions with our own species and with other species are also part of the environment and influence neuronal and hormonal communications. Whatever genes are present represent only part of the picture. It is all about the complex environmental context in which the genes are used or expressed.

Visualize the cookie factory. Not only is there an environment external to the factory itself, but there is also an intricate internal environment of relationships among the machines. Changing one small parameter, such as the temperature of the oven, can have a cascading effect on other functions.

"Nature versus nurture" is a favorite dichotomy for arguments about which is more important: the genes ("nature") or the environment ("nurture"). Well before any understanding of genetics, our human ancestors recognized that some aspects of traits seemed to be inherited while others seemed to be malleable products of training and teaching and other external factors. It is quite an interesting discussion to try to sort it out for any trait. The problem is that nature versus nurture is a false dichotomy. It is not an either/or situation but rather an intricate play of many environmental conditions (mostly unobserved by us) with hundreds of genes and their interactions (mostly undeciphered by us). If you find yourself in a nature versus nurture discussion, try introducing the useful (more nuanced) concept of predisposition, which allows much more flexibility and a more realistic picture of how a genetic trait is manifested. For example:

- My family is predisposed to diabetes (but we manage to control it with diet).
- That breed of dog is often predisposed to dominant behavior (but we've trained our dog not to act on her dominant impulses).
- A particular breast cancer gene predisposes about 5 percent of the women who carry it to getting breast cancer. What about the other 95 percent? There are gene and cell interactions in those women that we do not fully understand yet. The breast cancer gene seems to be expressed differently in them.
- The separated identical twins seemed to be predisposed to having anxieties—but one was raised in a calm, stable environment, the other in a hectic, unpredictable environment, and so you can see some differences in their behaviors.

Note that this is not "predetermination" or "determinism," which is a politically charged concept that sends us back to the false dichotomy of nurture versus nature. Predetermination is used simplistically to describe traits that are determined solely by our genetics, a situation that never turns out to be true once the particular trait is examined closely. Nonetheless, you may hear of predetermination in arguments in which one or another party does not fully

understand the complexities of genetics. A famous example is when Edward O. Wilson's *Sociobiology: The New Synthesis* was first published in 1975. The book caused great consternation among those who misunderstood and thought that the message was that all human behavior is predetermined by genes. For several years after the publication, research in the genetics of human behavior was discouraged, out of concern that proving genetic predetermination was the goal of such research. Indeed, the arguments were so heated that they are sometimes referred to as the "sociobiology wars." For example, Richard Lewontin and Stephen Jay Gould (both influential Harvard professors) and others equated a study of the genetics of human behavior to be "eugenics" in the most negative sense, as it was practiced in Nazi Germany. The most extreme arguments, many of which had political roots, have been mostly dispelled and the genetics of mammalian behavior, including that of humans, is now a thriving research area.

One caveat about the genetics of behavior is that it is easy to fall into the trap of thinking of single gene traits for complex behaviors, sometimes dubbed by the popular press with eye-catching names. About that, Lewontin, Gould, and others were right to be concerned. There is no "math gene" or "psychopath gene" or "athleticism gene" or "warrior gene." How might such misconceptions arise? Here is an example: A gene in mammals codes for the enzyme monoamine oxidase "MAO," which breaks down neurotransmitters after they have been used to pass along some neuronal communication. Laboratory rodents that have had their MAO gene switched off have excesses of some neurotransmitters and are more anxious than control rats. Some researchers have examined MAO genes in human families, trying to decipher any correlations with behavior. It appears that certain versions of MAO genes are connected to risk-taking behaviors. Is this difficult to sort out? It sure is. Any behavior is the end result of complex pathways with dozens to hundreds of genes involved. Environment in all of its many aspects is important and difficult to control out of a laboratory setting. Against the wishes of most serious scientists working on MAO and other genes, some popular writers have called it the "warrior gene." A good rule of thumb is that if you encounter a simplistic name for a gene that purports to be the primary explanation for a complicated trait, it probably isn't. But it makes a good headline.

There is a famous, ongoing study in Russia that was begun in the 1950s with silver foxes. It continues to be an excellent example of the intricacies of behavioral genetics. The goal of the experiment was to take an undomesticated canine and select the tamest animals every generation for as many generations as are necessary to get a lineage of domesticated animals. One surprise was how quickly domestication of the foxes occurred, suggesting that it was not all that difficult for our human ancestors to establish their own lineages of domesticated wolves, which became dogs. Another surprise was that other traits that were not being directly selected by the researchers nonetheless accompanied the complex behavioral trait of tameness. For example, researchers were looking for young foxes that would approach an offered human hand (neither cowering away nor growling) and they were looking for adults that were calm and even sociable around humans, rather than hypersensitive and anxious. These tamer foxes were allowed to breed the next generation, and so on. As each generation became more domesticated,

coat colors and shapes changed too, although the researchers were not selecting for those traits. The foxes became piebald (that is, two-colored such as black and white) with curly tails, floppy ears, and somewhat flattened (puppy-like) faces. Their rates of maturation and reproductive cycles changed such that they retained puppy-like characteristics much longer, even into sexual maturity. This is a phenomenon called neoteny, the retention of juvenile characteristics. It seems to be a trait of most well-integrated domestic breeds of dogs. (Exceptions might be some dog lineages bred deliberately to be aggressive.) So what does it mean that docile behavior and certain coat colors and ear and tail configurations as well as developmental changes all occur together. This is a good example of the way genes actually work, never one single gene at a time, but rather in complex pathways. Actually, it makes sense that the docile, sociable puppy-like behavior was accompanied by neoteny and cute puppy coloring, tails, and ears.

Two domesticated silver fox puppies from the Russian fifty-year breeding program.

Dog breeders have long noticed that certain coat colors of dogs can be correlated with behaviors. Some of the explanation for that will allow us to revisit the dog coat color pathway elaborated in a previous lecture:

Recall the C series of allele products, the ones just inside of the melanocyte (pigment cell) at which the pathway branches to make either dark eumelanin or lighter phaeomelanin or perhaps stops right there to make no melanin at all. It turns out that another branch from the C allele junction leads to the synthesis of certain hormones and neurotransmitters. Therefore, a failure to make enough or any melanin results not only in a pale animal but also may be accompanied by deafness, some eye disorders, and even some differences in temperament.

Furthermore, the C-alleles are a wonderful, accessible example of the role of environment in the activity of a gene product. The C enzyme (tyrosinase) works best in slightly cooler-than-core body temperatures. Siamese and Burmese cats have versions of C that produce pale coat color, but dark extremities (ears, feet, nose, tail) where temperature is a little cooler. This effect can be enhanced and can even fluctuate depending on what temperature environment the cat is experiencing. It is the same phenomenon in rodents and dogs with “Himalayan” pigmentation. Some forms of albinism in humans are due to the C-series alleles and, in some rare cases, an albino may darken a little, sometimes in the extremities. Note that in general, young mammals (including humans) are paler than adults and seem to develop more melanin as time passes. For example, light blond babies often develop darker hair through maturity, possibly due to variations in the C-series alleles.

Another example involves the S series of allele products that are involved with the distribution of pigment from the backbone area to the rest of a

mammal during development. A nonfunctional version of S results in a failure to distribute melanin and therefore is one of the many ways to get a white animal. Various patterns of stripes and spots are due to somewhat active forms of S and the most dominant form of S yields a fully colored animal (assuming all other parts of the pathway function). It turns out that neurons use the same S-allele series to grow out from the backbone area at around the same time in development as pigment is being distributed. Zebra striping is a good way to envision that pigment migration. Note that many mammals are darker along the backbone and paler beneath because of the way pigments migrate. Therefore, once again, it may be no surprise that interesting pigmentation can go together with particular neuronal conditions, given the interconnectedness of the pathways.

A demonstration of the importance of environment in plant genetics is one that can be set up in your kitchen. Boil some red cabbage and save the pigmented water. The color is from anthocyanin, which is found in various flowers, fruits, and leaves (including bright red autumn leaves). The pathway for the production of anthocyanin is complex. The pigment itself is susceptible to changing color depending upon the pH of the plant cells or sap. Pour some of the cabbage pigment into each of five glasses. Make a solution of baking soda in water (a few teaspoons in a cup of water) and have on hand some white vinegar. These are your basic (high pH) and acidic (low pH) solutions respectively. Add a couple of drops of acid to one glass and a couple of drops of base to another. Notice the color change. Try getting a slightly different color in two of your other glasses by adding more drops. Leave one glass of pigment untouched, so that you can compare.

So if you are trying to explain the importance of environment to someone, use the visual examples of the dark extremities of Siamese cats (effect of temperature) and the beautiful range of purples, reds, and pinks of anthocyanins in plants (effect of pH). However, also remind your listener that these are specially chosen examples to be the most dramatically visible. All genes and gene products are affected by environment with varying results, sometimes difficult to decipher. The superficial pigment genes are just metaphorical for a much greater phenomenon. Also, if you have a chance to talk to a serious dog breeder or trainer, you may find them to be a wealth of information about behavioral predispositions, the limits of trainability, and even the correlations between some coat colors and temperaments.

© Chepko Danil Vitalevich/Shutterstock.com

A Siamese kitten exhibiting traits associated with C-alleles (pale coat, but dark extremities).

FOR GREATER UNDERSTANDING

Questions

1. Can you explain the siamese cat example in the context of the influence of environment on genes?
2. What insights do we have on the domestication of animals from the silver fox experiment?

Suggested Reading

Rutter, Michael. *Genes and Behavior: Nature-Nurture Interplay Explained*. Oxford: Wiley-Blackwell, 2006.

Other Books of Interest

Grandin, Temple, and Catherine Johnson. *Animals in Translation: Using the Mysteries of Autism to Decode Animal Behavior*. New York: Harvest Books, 2006.

Wilson, Edward O. *Sociobiology: The New Synthesis*. 25th anniversary ed. Cambridge, MA: Belknap Press/Harvard University Press, 2000.

Article of Interest

Trut, Lyudmila. “Early Canid Domestication: The Farm-Fox Experiment.” *American Scientist*, vol. 87, pp. 160–169, 1999.

Lecture 7:
Environment Over the Long Term:
The Genetics of Populations

The Suggested Readings for this lecture are the examples of evolution in Kenneth R. Miller and Joseph S. Levine's *Biology* (or any other introductory biology textbook).

In lecture 3, when the idea of looking for ratios of genetic traits was introduced, you may have begun looking for ratios in natural populations, something that Gregor Mendel might have done too. I suggested some places to look, listed again below. You will see variability if you look carefully; however, you are not likely to find Mendelian-type ratios. Those are the results of careful breeding experiments. Something else is going on in natural populations and figuring that out is the topic of population genetics.

1. Look through a guidebook to wild flowers in your area, watching for ones that are described as varying in color, typically in the range of blues-pinks-whites (rather than yellows and oranges). Then set out to find a large stand of your plant and see if you notice any variations. If you do not, you may have found a true-breeding population for that trait, or maybe there is some environmental reason for only one color.
2. Watch for reports of unusual albino or melanistic mutants of wild animals such as white deer or black squirrels. Sometimes there will be just one sighting. However, there are interesting cases of populations with exceptionally high ratios of color mutations; for example, the population of black squirrels in some areas of the northeastern United States. If you are in the right area, try counting them at your bird feeder.
3. Or take any organism of interest and observe enough of them and in enough detail that you begin to notice variations. Focus in on one or another readily identifiable traits. Examples of natural genetic variability that I enjoy are variegated (white or reddish and green) markings in some clover leafs and pink and white water lilies. I also like to look at photos of wolves and other mammals to see the variations in coat colors. In an example below, I suggest counting spots in ladybugs. But really, if you are going to make a project of this, choose an organism you like (and can work with easily) because you will have to scrutinize a great many of them, perhaps at close range.

So let's take my example of a wild flower that can be either pink or white. If I find a large field of them and count one hundred, would I expect to get any of the ratios Mendel observed, such as 3:1 pinks to whites? Probably I would not, because out of the controlled conditions of the laboratory or greenhouse, environment plays a major role. Some alleles are selected above others. Some combinations of genes allow an organism to leave more offspring than others. Skewed ratios of traits are one of the first things to be noticed by a population geneticist. Then begins the challenging task of figuring out why. In the case of that pink pigment, which is anthocyanin, it has been noticed that in drought conditions, pink is favored over white perhaps

because the pigment is conferring some drought protection. In situations in which dryness seems not to be a factor, there may be less of a skewed ratio toward pink.

Pigmentation patterns in populations of mammals can often be interpreted fairly easily in terms of camouflage coloration. White animals are rare in temperate and tropical zones. They stand out against the background of vegetation. In the arctic, white coat color is a common trait (polar bears, rabbits, and arctic foxes). The same pigment pathways are present in all mammals. Which variants of the pathway are present is a consequence of selection by environmental conditions. Environment in this context (or any genetics context) should be interpreted broadly. The environment of a rabbit includes its predators; more white rabbits survive predation in the arctic, while more brown rabbits survive predation in the temperate forest.

What if we have a particular trait in a range of variations and the particular variation does not seem to matter much to the survival of the organism. Indeed, that is the case with some traits, which is why population geneticists use some restraint when they try to come up with explanations for their observations. For example, sometimes when an organism is new to an environment and does not yet have any natural predators, camouflage coloring or warning colorations can become quite variable. That seems to be the case with *Harmonia axyridis*, an Asiatic ladybird beetle, introduced as a natural pest-control insect to the United States. The patterns of orange coloration of these beetles are probably a warning of non-palatability to their natural predators, but there are no predators in their new environment. Therefore, a wonderful array of different markings has been occurring with no particular discernable patterns of selection. In contrast, native ladybird beetles have many more constraints in their warning colorations (their numbers and arrangements of spots) because they do have natural predators. Note that some Asiatic ladybird beetles like to overwinter in warm places, such as in your house. So if you happen to have an autumn infestation of colorful orange and black beetles, before you sweep them out the door, you might examine their varied color patterns.

Population genetics is the genetics that Charles Darwin would have loved to have known about. Darwin was fascinated with variations in traits. He was a close observer of variability in artificially selected

Asiatic ladybugs display multiple patterns and shades of orange to almost red.

domesticated organisms, and this strongly influenced his theory of natural selection. He wrote *The Variation of Animals and Plants Under Domestication* in 1868 and did some experiments to try to sort out the way in which traits were being inherited. For example, Darwin bred snapdragons and according to the data he collected, the traits were present in a 3:1 ratio, but unrecognized by Darwin. Meanwhile, around the same time, but completely unknown to Darwin, Mendel was doing the same sorts of experiments with peas and did manage to recognize a 3:1 ratio and its significance. How close was Darwin to figuring out some simple patterns of inheritance? Possibly he never would have spotted and correctly interpreted any ratios in his plant experiments. However, he had one more chance. A copy of Mendel's publication is in Darwin's library. If Darwin had read it, he might have understood the significance. In those days, inexpensive, bound publications sometimes had their pages doubled over, uncut, and the reader was obliged to cut each one, to reveal each new page. Darwin's copy of Mendel is still uncut. He never read it. By the way, a copy of Darwin's *On the Origin of Species* is in Mendel's library and Mendel did read it.

Genetics was the major component missing from Darwin's otherwise thorough explanation of natural selection. Genetics was not added into the picture of evolution until the twentieth century. The two major components of evolution are:

1. Genetic variability, easily observed but less easily interpreted and analyzed.
2. Selection by a host of environmental parameters for particular traits and combinations of traits. That is, some combinations of genes allow an organism to leave more offspring than others.

And then there is the component of time. There must be many generations of reproduction of genetic variants of organisms accompanied by selection for particular variants, over and over again. Thus originated the extraordinary richness of species on Earth as well as all the variants within species.

Population genetics is the genetics that medical doctors are beginning to pay attention to. It begins with human migrations or more likely less purposeful wanderings, starting with a population of *Homo sapiens* in east Africa about two hundred thousand years ago. Their coloring was dark, and therefore protective, which makes lots of sense for hairless organisms dwelling under a tropical sun. There was probably little genetic variability between individuals. This was a small, interbreeding, therefore inbred population, as were most human populations through most of human history.

Around eighty thousand years ago (a mere four thousand generations if you let each generation be twenty years), migrations or wanderings out of Africa began. By about forty to fifty thousand years ago (just two thousand generations), Europe, Asia, and Australia had their own small inbred populations of *Homo sapiens*. Therefore we are all east Africans with just a few variations due to our diverse wanderings. Around fifteen thousand years ago (less than one thousand generations) during the ice age, some humans made it to the Americas and wandered all over. And finally, once-separated populations of *Homo sapiens* met after coming full circle around the globe when Europeans first reached the Americas by boat, less than one thousand years ago (less

than fifty generations). In the latter part of the twentieth century some *Homo sapiens* began to be extraordinarily mobile (at least those with a desire to do so and who could afford it). However, we have spent many more generations *not* being mobile, with pockets or islands (or valleys) of genetically inbred groups—made even more so by cultural differences. Inbreeding has resulted in recognizably different-looking populations of humans with different skin colors, face shapes, statures, and other recognizable characteristics. In general, populations along the equator have maintained protective skin pigmentation. That protection was lost in European populations and then apparently lost at a different place and time by temperate Asian populations. However skin pigmentation is a rather superficial trait compared to what your doctor would really like to know about you. Therefore, it is becoming passé to fill out a simple form with just four boxes from which to choose, such as:

- Caucasian
- African descent
- Asian/Pacific Island
- Native American

Far more important is a description of you or me such as:

> "Most of my ancestors are from Sicily, but my grandmother always said she was half Native American and I am about to marry someone whose ancestors have lived in Sichuan Province of China for many generations."

A young Californian of Asian-Caucasian descent.

Or:

> "I am approximately half Inuit and half Norwegian and I am about to marry a Ghanian who can trace extensive ancestry within the Fanti tribe."

Your doctor is interested because your ancestors spent many more generations being isolated in Sicily or Siberia or Ghana than they did meeting each other and intermarrying in the nineteenth and twentieth centuries. Along with distinctive facial characteristics and pigmentations and other visible traits are predisposition to particular genetic conditions such as diabetes and cancers. Knowing which small populations are part of your ancestry will help to individualize medical testing and treatment.

Let's take the Amish of Lancaster County, Pennsylvania, as an example of an isolated population by their own choice due to their religious beliefs. They are a favorite example of human population geneticists. Because of the inbreeding that naturally occurs in such a population, the Amish have an unusual prevalence of certain rare genetic conditions, including some

muscular, neurological, and metabolic genetic diseases. But you can say the same for any population, whether geographically remote or isolated by choice. They will have their own particular suite of genetic predispositions.

Oliver Sacks writes wonderfully about an example of exceptional color blindness in two island populations (one in Scandinavia, one in the Pacific) in *The Island of the Colorblind.* What is notable is that the physical remoteness of the island populations was only part of the story. For many genetic conditions, people make conscious choices to live together. The color blindness in Sacks's examples was accompanied by extreme sensitivity to light and therefore a life-style that involved being sheltered indoors during the day and being more active outdoors at night. So color-blind people were doing more socializing (and marrying) within their group than with others. This is sometimes the case with social groups of deaf people and little people who share major lifestyle choices.

The concept of "eugenics" (good genes) was developed in the 1880s when some of the science behind heredity and natural selection had been elucidated. Francis Galton, a cousin of Charles Darwin, first developed the idea that there might be social programs based in scientific fact by which humans might be improved. The basic idea was far from new. It is the foundation of thousands of years of selective breeding and "improvement" (by some human criteria) of domestic animals and plants. What was new was the possibility of supporting scientific facts. Galton's goal was to lessen human suffering by disease, to increase intelligence, strength, and other desireable qualities, and to conserve limited resources by directing them to the healthiest, "fittest" individuals. In just a few decades, eugenics became associated with (indeed a justification for) the genocides of Nazi Germany. However, extreme policies and political agendas based on eugenics were far from being unique to Nazi Germany. Other manifestations of eugenics (throughout the world, including the United States) included strict immigration quotas to eliminate less desirable people and sterility programs to prevent some "less fit" individuals, such as those of less-than-average intelligence, from reproducing.

However, before we dismiss eugenics as the root of some of the most misguided social policies of the twentieth century, let's look at the deeper roots. Presented here in a list format, approximately from the mildest to the most extreme, are eugenic-type practices, many of which were and still are entirely routine ways by which human lineages are "improved" by some criteria. Where do we draw the line? Perhaps it is somewhere part way down the list, depending on your own beliefs and choices. When is it genocide, for all intents and purposes? And what are the criteria for "good" traits and how are they measured? You may disagree with the order in which I present these and with some of the items on this list or some left off. All of that places many issues of eugenics well into the realm of the ethics of genetics, not just the science of genetics. Furthermore, some of the most extreme public policies of eugenics come from pseudoscience (intentional or not) rather than science.

- Choosing mates for various positive qualities (perhaps subjectively determined)
- Arranged marriages for various positive traits
- Social policies calling positive attention to strong marriages (an early Nazi tactic, but could be an innocently framed policy of encouragement)
- Premarital genetic testing
- Genetic counseling
- Sperm banks and egg banks (including from donors with desireable characteristics)
- *In vitro* fertilization
- Birth control for oneself, or imposed on others (Margaret Sanger was part of an early eugenics movement)
- Sterilization chosen for oneself
- Prenatal testing and then decisions about terminating a pregnancy
- Shunning or ousting from sects
- Limiting immigration (by nationality, by abilities)
- Marriage laws (preventing certain marriages)
- Gene therapies (if it works; there are no routinely used examples)
- Designer gene therapy (if it works; there are no routinely used examples)
- Cloning (if it works; there are no routinely used examples in humans)
- Institutionalizing people with undesirable traits and thus preventing their marriages
- Rationing healthcare by group (that is, not using it excessively in hopeless cases)
- Selective abortion—as done to others
- Sterility programs—as done to others
- Forced emigration
- Infanticide
- Euthanasia
- Genocide—*ad hoc* or systematic

By the way, some of the more intriguing outcomes of eugenics policies in the United States in the early twentieth century were public policies to show more sympathy to people who were deemed to be defective genetically in some way. That included reform of some practices with disabled people in institutions and adjustments in the education system to accommodate special needs. It also included reevaluations of the use of capital punishment, especially in cases where the defendant might have some genetic deficiency that led to the crime. Clarence Darrow successfully used a defense of genetic disability in the famous murder trial of Leopold and Loeb.

FOR GREATER UNDERSTANDING

Questions and Activities

1. What is indicated by the many different colorations of Asiatic ladybird beetles?
2. What are the two major components of evolution?
3. Try working up a few detailed examples to use in conversations concerning the natural variability of animals and plants. Focus on your own favorites and try to get some experience with them, whether observing from nature or photographs.
4. Do the same for eugenics by coming up with a range of examples from routine, innocuous practices to more elaborate schemes that require careful analysis. The extremes of deliberate and methodical genocides are obvious examples, but it might be more interesting to have on hand some more subtle ones.

Suggested Reading

Miller, Kenneth R., and Joseph S. Levine. *Biology*. Upper Saddle River, NJ: Prentice Hall, 2007.

Other Books of Interest

Peterson, Roger Tory, and Margaret McKenny. *A Field Guide to Wildflowers*. New York: Houghton Mifflin Harcourt, 1998.

Sacks, Oliver. *The Island of the Colorblind*. New York: Vintage, 1998.

Website of Interest

1. Idea.org provides simulations of various types of color blindness — http://www.idea.org/vision-demo.html
2. The University of Missouri, Kansas City School of Law, provides an in-depth look at the 1924 murder trial of Leopold and Loeb, in which Clarence Darrow referred at length to the genetic factors of the case — http://www.law.umkc.edu/faculty/projects/FTrials/leoploeb/leopold.htm

Lecture 8:
DNA (Deoxyribonucleic Acid)

The Suggested Reading for this lecture is James Schwartz's *In Pursuit of the Gene: From Darwin to DNA*. (Also see "Isolating DNA from Strawberries" activity on page 102.)

I introduce DNA about halfway through my college genetics course because I think it is valuable to get a good feel for genes as information without the complexities of discussing the molecule actually carrying that information. It is a little like the difference between software and hardware for a computer. The information of genes is software-like. The DNA molecule is the physical reality of how that information is encoded: hardware-like. For many decades the chronology of discoveries concerning DNA as a molecule ran parallel to and independent of the studies of genetics. At first there were no particular strong connections between the two.

Dr. Frederich Miescher, a Swiss working in Germany, isolated DNA in 1869 from the pus (white blood cells) of the bandages of surgical patients. Actually, following a set of modern instructions included at the end of this lecture, it is not all that difficult to isolate DNA. The end result is a bit slimy or mucousy if you get enough of it. Dr. Miescher described DNA as a "white wolley precipitate." Note that the doctor was looking at all sorts of molecules in pus, trying to get a better understanding of how to treat wounds.

By 1889, the four bases or building blocks that make up DNA had been discovered and were understood to make up DNA. Here they are in order and with the original materials from which they were first isolated.

Guanine	1850	from bird guano and later from the silvery sheen of fish skin
Adenine	1885	from glands such as ox pancreas and spleen
Cytosine	1894	from calf thymus gland
Thymine	1894	from calf thymus gland

These bases usually are referred to as A, C, G, and T. In addition, there is a fifth base uracil "U" found in RNA instead of "T."

Throughout the twentieth century, there was a growing awareness that there must be a type of information-bearing molecule in the cell and most likely in the nucleus. However, until the 1940s, DNA was not considered the most likely candidate because it seemed too simple with only four types of building blocks.

Erwin Chargaff's research in the 1940s provided some important clues about the nature of DNA. He cut up DNA molecules into subunits, counted the bases, and found that the number of As equaled the number of Ts and the number of Cs equaled the number of Gs.

The story of how the structure of DNA was discovered has been reiterated from many different points of view in several books, recommended at the end of this lecture. In short, three research groups found themselves in competition for making the discovery.

Linus Pauling at the California Institute of Technology was already an expert in the necessary technique, X-ray crystallography, and had already discovered the structure of protein, for which he won a Nobel Prize. Pauling came close to collecting the Nobel Prize for DNA too.

Meanwhile in England were Francis Crick and James Watson at Kings College, and Rosalind Franklin and Maurice Wilkins at the Cavendish labs. It would have been a powerful and versatile team if all four had worked together. Watson and Crick were imaginative, adventurous builders of models using X-ray data, while Franklin and Wilkins were experts in X-ray crystallography, that is, producing the data. However, the two teams had quite different personalities and worked separately with only occasional, albeit significant interactions. The day-by-day accounts of their work in 1952 and early 1953 are well worth reading to get a flavor of how politics, friendships, rivalries, and communication can affect the outcomes of scientific research.

That DNA was a helix of two strands (like a spiral staircase) was revealed by X-ray crystallography. That DNA fit together such that each step of the staircase was A paired with T or C paired with G was revealed by model building. T and C are both single-ring structures. A and G are both double-ring structures. By pairing according to Chargaff''s numbers, each step was three rings side by side—A:T and C:G.

Three back-to-back papers appeared in *Nature* on April 25, 1953, to announce the nearly simultaneous discoveries of the structure of DNA. First were Watson and Crick, then Wilkins and his colleagues Stokes and Wilson, and finally Franklin and her colleague Gosling. At this point Wilkins and Franklin were no longer working harmoniously together. In 1962 Watson, Crick, and Wilkins shared the Nobel Prize. Franklin had died of cancer perhaps from exposure to X-rays.

In 1953, it was already obvious that the structure of DNA was relevant to genetics. Watson and Crick wrote the following:

> *We wish to suggest a structure for the salt of deoxyribose nucleic acid. This structure has novel features which are of considerable biological interest. . . . It has not escaped our notice that the specific pairing we have postulated immediately suggests a copying mechanism.*

Apparently Crick wanted to include the word “beauty” in the paper, but it was struck by a reviewer perhaps as too emotional. But it *is* a beautiful molecule. Try to find a model to look at from all angles, perhaps even a large model at a science museum. The spiral staircase effect is stunning. Look closely at the chemical structure and notice something of an M.C. Escher quality to the staircase. The two strands are anti-parallel; that is, one strand is in reverse orientation to the other. Furthermore, DNA in real life almost never gets the space and conditions to stretch out as a smooth double helix but rather is stuffed in a kinked and tangled mass into the tiny confines of the cell nucleus.

How does DNA work? The structure allows it to replicate so that if we begin with a double helix (flattened out) like this:

```
ACGTTACCGTAC
||||||||||||
TGCAATGGCATG
```

It can be made to split apart down the middle (splitting each step) like this:

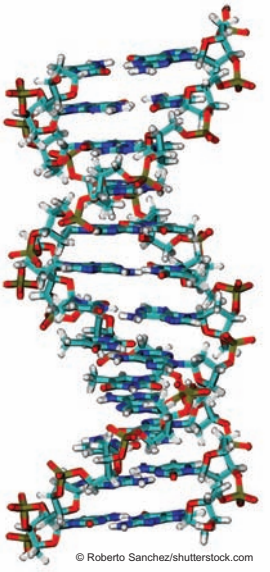
```
ACGTTACCGTAC

TGCAATGGCATG
```

And then each separate strand can have its complementary strand built. And now there are two DNA molecules, each bearing the same sequence (and therefore the same information).

```
ACGTTACCGTAC
||||||||||||
TGCAATGGCATG

ACGTTACCGTAC
||||||||||||
TGCAATGGCATG
```



What makes DNA interesting and in fact what makes possible its use as a biological information molecule is that sometimes the replication isn't perfect. Think of how tediously identical all DNA would be (and all biological information would be) if replication were perfect. It is the lack of perfection that makes the extraordinary diversity of life possible. We (or any organism) are the products of millions of little mistakes in replication or mutations. Although the word *mutation* sometimes has a negative context (the roots of which will be addressed further in lecture 11), it really is mostly positive in the biological sense. That is, it is positive if you consider the vast, versatile range of living beings and our own amazing complexity as positive attributes. Otherwise the molecule of "life" would be nothing but a short string of information precisely copying itself endlessly for generations in the same little warm pond of its origin. We certainly would not be here to talk about it.

How about the context of the cookie factory? Those blueprints for every machine are made of DNA. Building a new factory begins with the copying of the blueprints.

Sometimes there are errors. Whoever or whatever is doing the copying does not do so precisely. Therefore every new cookie factory, built with its own newly copied set of blueprints, is a little different from the others.

One of the original metaphors developed by DNA researchers for understanding the function of DNA was that of a medieval monk copying a manuscript in a scriptorium. In the next lecture, watch for the way some of the terminology used to talk about DNA reflects that metaphor.

FOR GREATER UNDERSTANDING

Questions and Activities

1. How did Dr. Frederich Miescher isolate DNA?
2. How is mutation "positive" in the biological sense?
3. Try isolating some DNA from strawberries and other materials using reagents made from common household chemicals (see "Isolating DNA" on page 102 for instructions).

Suggested Reading

Schwartz, James. *In Pursuit of the Gene: From Darwin to DNA*. Cambridge, MA: Harvard University Press, 2008.

Other Books of Interest

Crick, Francis. *What Mad Pursuit: A Personal View of Scientific Discovery*. New York: Basic Books, 1988.

Judson, Horace F. *The Eighth Day of Creation: Makers of the Revolution in Biology*. New York: Touchstone, 1979.

Maddox, Brenda. *Rosalind Franklin: The Dark Lady of DNA*. New York: Harper Perennial, 2003.

Watson, James. *The Double Helix*. New York: Signet Books, 1968.

Lecture 9:
Transcription and Translation:
The Metaphor of an Error-Prone Medieval Monk in a Scriptorium

The Suggested Reading for this lecture is James Schwartz's *In Pursuit of the Gene: From Darwin to DNA*.

Imagine a medieval monk laboring over a manuscript in a scriptorium. He is somewhat error-prone and therefore introduces little mistakes (mutations) here and there. Overall the manuscripts produced are mostly readable and mostly quite functional. That is, in the short run, many mutations may not matter much. For example, lots of DNA information is redundant so that a mistake in one section is balanced by the correct information being elsewhere. (And remember the nature of the cookie factory. It was designed by Rube Goldberg. It works well enough in spite of the baroque convolutions.)

Now and then, the monk introduces an enormous error, such as eliminating a few pages or badly misspelling the most meaningful word of a chapter. Here is where selection comes in. We do have a wastebasket for truly egregious manuscripts. That is, if DNA is accidently copied so badly as to cause an organism not to survive and therefore not to reproduce, then that DNA will not have opportunities to replicate again.

When the information of DNA is actually used as information (not just replicated with little errors) the first step is *transcription*. Think of the medieval monk, copying (transcribing) a short bit of a sequence onto a piece of scrap paper, for immediate use, down the hall in a different section of the monastery. It is as though the main manuscript (the main blueprint) is stored safely (and copied regularly in its entirety), but in order to actually build something we don't want the entire set. We just want a little section—

THE GENETIC CODE

	U	C	A	G	
U	UUU Phenylalanine UUC Phenylalanine	UCU Serine UCC Serine UCA Serine UCG Serine	UAU Tyrosine UAC Tyrosine	UGU Cysteine UGC Cysteine	**U**
	UUG Leucine UUA Leucine		UAA **Stop** UAG **Stop**	UGA **Stop** UGG Tryptophan	**C** **A** **G**
C	CUU Leucine CUC Leucine CUA Leucine CUG Leucine	CCU Proline CCC Proline CCA Proline CCG Proline	CAU Histidine CAC Histidine	CGU Arginine CGC Arginine CGA Arginine CGG Arginine	**U** **C** **A** **G**
			CAA Glutamine CAG Glutamine		
A	AUU Isoleucine AUC Isoleucine AUA Isoleucine	ACU Threonine ACC Threonine ACA Threonine ACG Threonine	AAU Asparagine AAC Asparagine	AGU Serine AGC Serine	**U** **C**
	AUG Methionine		AAA Lysine AAG Lysine	AGA Arginine AGG Arginine	**A** **G**
G	GUU Valine GUC Valine GUA Valine GUG Valine	GCU Alanine GCC Alanine GCA Alanine GCG Alanine	GAU Aspartic acid GAC Aspartic acid	GGU Glycine GGC Glycine GGA Glycine GGG Glycine	**U** **C** **A** **G**
			GAA Glutamic acid GAG Glutamic acid		

actually, we just want *a gene*. And here at last is an official definition (albeit one with lots of exceptions) for "gene." A gene is that sequence which can be transcribed (jotted down on a scrap of paper) and sent elsewhere (out of the scriptorium, out of the nucleus) for use. Scrap paper is a reasonable analogy here because the format in which the transcript is made is temporary. The message must be used within a reasonable period of time; otherwise it gets torn up. The actual DNA transcript is made of a single-stranded RNA molecule called messenger RNA or mRNA.

Here is the original DNA:

```
ACGTTACCGTAC
||||||||||||
TGCAATGGCATG
```

One strand serves as a template. Let's say it is this one:

ACGTTACCGTAC

A complementary mRNA is made using "U" instead of "T":

UGCAAUGGCAUG

And now UGCAAUGGCAUG, freshly transcribed (message-like) is being rushed down the hall so that it can be used as instructions to build a machine before it can be torn up.

How does the message get used? It gets *translated*; that is, the information is converted from one language to another. Therefore, down the hall is a set of bilingual translation machinery capable of reading two languages: DNA language and protein language. Organisms are made mostly of protein. Proteins form almost all of the distinctive structures and functions of the body. By analogy, proteins are the machines and all the infrastructure of the cookie factory.

The deciphering of the DNA code and its correspondence to protein code was a major accomplishment of the early 1960s. There are four bases (A, C, G, and T) in DNA; think of a four-letter alphabet. There are twenty amino acid subunits of proteins; think of a twenty-letter alphabet. Is it possible to "say" twenty different things with words comprising only four letters? Here are some of the possibilities:

If "words" in DNA language are of length one, then there are just four words possible: A, C, G, and T, which is much too parsimonious a language. How long do the words need to be in order to code for twenty amino acids?

Length of Word	Number of Possible Words (if alphabet size is four)
One Letter	Four Words
Two Letters	Sixteen Words
Three Letters	Sixty-four Words

And the answer is that the DNA code comprises three-letter words (or triplets), which is more than enough to code for twenty protein words. Indeed, with sixty-four possible triplets, there are lots of synonyms and even four "punctuation" words.

The laboratory breakthrough in the discovery of the triplet code was a little like the breakthrough of Helen Keller, who was blind and deaf. Her teacher Annie Sullivan simultaneously poured water into Helen's hands and signed the word for water and Helen made the connection. In the lab, an artificial sequence of messenger RNA was concocted of only uracils:

UUUUUUUUUU...

Then the message was sent through the bilingual translational machinery and the result was a string of the amino acid phenylalanine:

Phe Phe Phe Phe...

And thus the code was broken.

Each of the twenty amino acids had at least one DNA triplet (DNA codon) and most had many more than that. For example, here are the four synonyms of the amino acid serine: UCU, UCC, UCA, UCG. Note that there is a great advantage to having synonyms. It means that the system can tolerate some mistakes (mutations) quite easily without changing the meaning of the word.

In addition, there are four punctuation codons:

- AUG marks the beginning of a gene (that which can be transcribed); it also codes for the amino acid methionine.
- UGA, UAA, and UAG are synonyms for "stop," marking the end of a gene.

So a simplified gene (with spaces inserted between codons) looks like the following:

AUG UUU UUU UCA UCG UUU UAG

And it is translated to a tiny protein (typical proteins have hundreds of amino acids):

Met Phe Phe Ser Ser Phe

Returning to the metaphor of the monk in the scriptorium, let's say there are two major types of mutations he can cause in the manuscript:

1. He can *substitute* one letter for another or write the letter such that it is unintelligible. These are "point mutations."
2. He can *insert or delete* letters. Both have the same effect and unless you know what the original manuscript was supposed to have looked like, it can be difficult to tell which happened: an insertion or a deletion of a letter. That is why these are sometimes called "indels."

Many point mutations will have no major effect on the final product (a protein). That is in part because there are so many synonyms in the language. It is also because changing just one amino acid in a long string of hundreds making up a protein in many cases does not result in any major change in protein function.

However, indels can have dramatic consequences. They cause shifts in reading frame, "frame shift mutations," which might render the code unreadable. Consider this example using three-letter words in English:

THE CAT SAT

Remove the T and read it:

HEC ATS AT*

Or add a letter:

THE *B*CA TSA T**

Either way, the remainder of the sentence is spoiled. The same happens in DNA frame shifts, typically resulting in a nonfunctional protein.

So at this point, let's review what a gene is, and then move on to a bit of surprising complexity that will necessitate invoking a completely different metaphor than the monk in the scriptorium.

A gene is defined by its functionality. It is a sequence of DNA with a start (AUG), a sequence of triplets coding for amino acids, and a stop (for example, UAG). It is capable of being transcribed (mRNA, the temporary scrap of paper) and translated (from DNA to protein language).

Genes of complex organisms (like humans) are themselves complex in that they comprise coding sequences interspersed with non-coding sequences. An English example (using three-letter words) might look like this:

THE CAT SAT XZFYYH THE MEN BAT LFFFDCG GLOAAAB WET HAT

In order to make this into plain English (or into plain DNA language), first we transcribe and then we splice, removing all those intervening, unreadable sequences called "introns." The readable parts are called "exons."

THE CAT SAT THE MEN BAT WET HAT

And then we translate that spliced sequence.

It seems a bit convoluted, doesn't it? What are those introns doing in there, except to add a seemingly unnecessary step?

Time to switch metaphors: Let's say that exons and introns were a building toy and that there were all sorts of combinations that could be built of various exons.

Start with this set of building toys:

```
EXON#1 INTRON EXON#2 INTRON EXON#3
```

How many ways can I mix and match (splice) those exons?

There are seven possibilities, assuming they stay in order:

```
EXON#1
EXON#2
EXON#3
EXON#1 EXON#2
EXON#1 EXON#3
EXON#2 EXON#3
EXON#1 EXON#2 EXON#3
```

And each of those seven combinations will give us a different (albeit related) protein assembled in a sort of modular fashion. It means that one gene has the potential to make many different products. This helps to explain why it is that with only twenty-five thousand genes (typical of complex, multicellular organisms), there are hundreds of thousands of potential proteins that can be made. Our coding capacity has something in common with the possibilities from a large set of building toys in the hands of an imaginative child.

FOR GREATER UNDERSTANDING

Questions

1. Why is scrap paper a good analogy for the way a gene functions?
2. What is the difference between introns and exons?
3. Explain transcription and translation using the medieval monk in the scriptorium analogy.

Suggested Reading

Schwartz, James. *In Pursuit of the Gene: From Darwin to DNA*. Cambridge, MA: Harvard University Press, 2008.

Other Books of Interest

Crick, Francis. *What Mad Pursuit: A Personal View of Scientific Discovery*. New York: Basic Books, 1988.

Judson, Horace F. *The Eighth Day of Creation: Makers of the Revolution in Biology*. New York: Touchstone, 1979.

Maddox, Brenda. *Rosalind Franklin: The Dark Lady of DNA*. New York: Harper Perennial, 2003.

Watson, James. *The Double Helix*. New York: Signet Books, 1968.

Lecture 10:
Chromosomes, Mitosis, and Meiosis:
How DNA Is Packaged for Moving

The Suggested Reading for this lecture is James Schwartz's *In Pursuit of the Gene: From Darwin to DNA*.

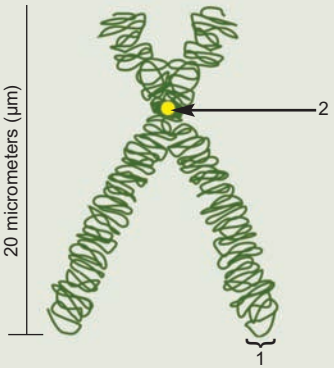
Thus far in the course, we have considered DNA as information (strings of letters spelling out the sequences for protein) and DNA as a molecule capable of being imperfectly replicated. There is an additional aspect of DNA having to do with the vast quantities of it packed into the tiny nuclei of each cell. If all of the DNA from just one human cell were set out end-to-end, stretched out to full length, forming one string, it would be 1.8 meters long. That's right: meters. And all of it is packed into a cell of perhaps 20 micrometers in diameter. Within that cell, the DNA is packed into an even tinier compartment, the nucleus, about 5 micrometers in diameter. To put that into perspective:

1.8 meters of DNA =
180 centimeters =
1,800 millimeters =
1,800,000 micrometers
packed into a 5 micrometer diameter (0.000196850394″) space

Is that a problem? It sure is. Enough of a problem that cells have entire sets of structures and mechanisms to deal with it. These include chromosomes, mitosis, and (for special occasions) meiosis. Can anything go wrong with chromosomes, mitosis, and meiosis? Yes, pretty much anything that can go right in a cell can go wrong too, and that's what can be so interesting. There are whole categories of DNA mutations on a grand scale having to do with the necessities of hauling around such vast quantities of it and the mistakes that can be made.

Chromosomes ("colored bodies") are enormous rope-like cell structures that have been observed with ordinary light microscopes since the nineteenth century. Thus the name, implying that chromosomes readily take up stain and become colorful, highly visible structures at certain times in the life of a cell under observation. In particular, it was noticed that chromosomes become enlarged and highly visible and active right before cell division. Here is what is happening. Notice that the first step cannot be observed under the microscope:

1. All of the DNA of the cell replicates. Now there are two copies of each DNA string, non-visible, under a microscope.



A chromosome that has replicated to make two copies. The copies called chromatids (1) are held together temporarily with a structure called a centromere (2).

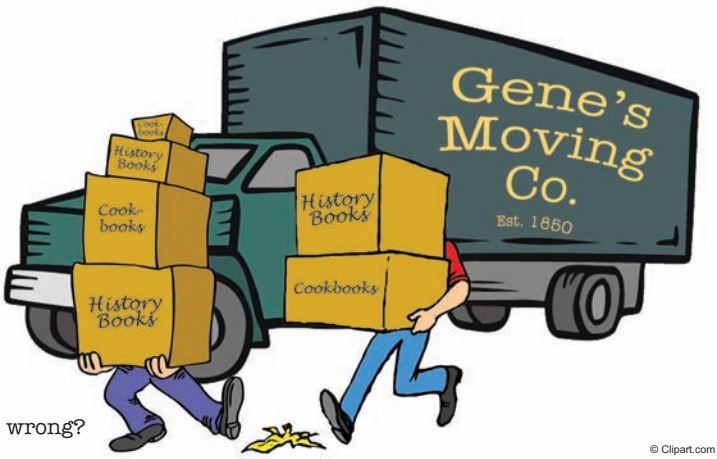
2. The DNA begins to coil up into thick rope-like structures stabilized with proteins, that is, chromosomes—visible under a microscope. The copies remain held together by a centromere. (Actually, DNA is in some chromosomal form all the time, but usually more loosely coiled and therefore not visible.)
3. The centromeres split. The chromosomes move and separate in an organized fashion such that each of the two copies of the cell's DNA (from step #1) become poised to separate into two cells (visible).
4. The cell divides; if all has gone well each new cell has a complete set of DNA (the DNA returns to invisibility under the microscope).

What is going on here is a little like moving day. Let's say you and your roommate own hundreds of books and are about to go your separate ways and would each like a copy of every book in your library. Step one: copy all those books. Maybe hire a medieval monk in a scriptorium to do it. Step two (the highly visible step): package up all those books into large moving boxes and shuffle them around such that when you go your separate ways, you'll each have a set of boxes. That is what chromosomes are all about. They are the moving boxes, packaging long stringy DNA such that it can be more efficiently sorted and separated. The process of shuffling and separating is "mitosis," and it contains all of the steps that you might expect:

1. Pack the boxes (coil up the chromosomes).
2. Pair up the boxes so that we can be sure that each roommate will get one of a pair.
3. Slide the boxes to opposite corners (my boxes versus your boxes).
4. Move boxes into two new apartments (cell division). Done.

What could go wrong? Plenty! Maybe I'll get two boxes of history books but no cookbooks and you'll get all the cookbooks and no history, which wasn't the intent, but with all that shuffling it could happen by accident. The consequences (for DNA) turn out to be fascinating. There will be more on that later.

Meanwhile, to help you visualize and to add in a few crucial nuances:



What could go wrong?

We humans have in most of our cells, most of the time, two sets of DNA. One set originally came from our mother, one set from our father. That DNA is bundled up (loosely or tightly) into 23 different chromosomes. Since we have two sets of each, we actually have 23 pairs of chromosomes or a total of 46 chromosomes. It is a common misunderstanding among people who have not studied genetics that DNA and genes and chromosomes are all something like synonyms. As you know, they aren't. Chromosomes, DNA, and genes exist on completely different scales, ranging from visible to conceptual.

Each chromosome consists of a long strand of DNA tightly (or loosely) wound up and coiled with the help of proteins. Each chromosome (each DNA strand) has thousands of gene sequences. That is why mistakes on moving day can have such dramatic consequences. I'm not losing just one cookbook. I'm losing hundreds of them.

When the genome project was not yet underway, there was plenty of speculation as to how genes might be organized on chromosomes. If you are thinking about books shelved in libraries, you might be thinking of something as orderly as the metaphor of boxed books on moving day, such as all the cookbooks in one section, the history books in another. That was certainly a hypothesis about gene organization too. One big surprise from completed genome projects of complex organisms was that the genes were arranged mostly in no particular discernable order. That is, the books are scattered all over the house. Using a particular set of books (genes) in some coordinated fashion apparently requires first getting them together somehow. There will be more about this in lecture 12, but for the moment consider this: The linear book-shelf metaphor begins to fall apart (as most metaphors eventually do) when you consider the logistics of 1.8 meters of DNA stuffed into that tiny nucleus. There is nothing linear about it. So rather than orderly rows of books, think of a dynamic ball of yarn with lots of potential for creative new combinations. Think of the yarn as capable of coiling and uncoiling such that in any given fraction of a second, this or that set of genes might be juxtaposed in the tangle. That may be how genes are being used in coordination, despite the seemingly (at least to our human perception) lack of library-order.

By the way, right after DNA replication (right after our 46 chromosomes are replicated), if you were to go in and count up all the chromosomes, you might expect 92 chromosomes. We just doubled 46: 46 x 2 = 92. And if you and I were writing the earliest genetics textbooks, that is what we would say: 92 chromosomes. However, for various reasons, mostly due to some early misunderstandings about what was important about mitosis, we never say "92" chromosomes. It is in the nature of scientific jargon that often early misunderstandings become nonetheless petrified permanently in the terminology. Genetics, being such an early topic of investigation, is rife with difficult terminology and concepts, not just because it is complex, but because of the amazing stability of vast amounts of jargon, coined well before a full understanding of the concepts.

Mitosis (that mostly orderly dividing of chromosomes) is happening all the time, all over your body, every single time you make new tissues (new skin, new hair, new intestinal linings, and so on). It is also the mechanism of growth. From infant (a single cell at fertilization) to adult (with ten trillion cells) there are many millions of mitoses. Mitosis is also the means by which single-celled

organisms reproduce. Mitosis is how cancer cells duplicate themselves out of control. And that is why rapid mitosis is sometimes the Achilles heel of cancer cells. Some chemotherapies set about to poison those cells that are dividing rapidly by poisoning certain aspects of mitosis. That is why side effects of chemotherapies can affect our other cells that divide frequently, such as hair, skin, and digestive lining.

There is another way to divide up chromosomes, used only in the special circumstance of making sperm or eggs. It is called meiosis and it goes like this:

Sexual reproduction (as opposed to simple reproduction by just doing mitosis) entails some logistical problems with numbers of chromosomes. Sexual reproduction is the union of two cells (sperm and egg) to make one cell (zygote)—followed by lots of cell divisions (lots of mitoses) to make an embryo. So I have 46 chromosomes and you have 46 chromosomes. We have sex—a fusion of sperm and egg, "fertilization." Our new zygote has 92 chromosomes. And that is a problem. Somehow, we have to begin this whole process, several steps back, or we'll be doubling the number of chromosomes every new generation. We need to start by reducing the number of chromosomes (in the sperms and eggs) in half. That requires a special cell division—meiosis—which is like mitosis, but instead we go from 46 chromosomes to 23 chromosomes, a sort of modified moving day.

Meiosis is very specialized, occurring only in the gonads. It requires two cell divisions. For this sequence I am going to use the typically forbidden number of "92" to make the point about what is happening:

1. Start with 46 (23 pairs) of chromosomes.
2. Replicate the DNA so now we have "92" (23 quadruples).
3. Do the first division. Now we are down to 46 (23 pairs).
4. Do the second division. Now we are down to 23 single copies of each chromosome.

So, sperm have 23 chromosomes, eggs have 23, and zygotes (newly fertilized eggs) have 46, or 23 pairs (one set from each parent).

Is that it? Well, no. With genetics there is always something more. The following are three (of many possible) additional nuances concerning chromosomes and meiosis.

Chromosomal Mutations

All sorts of mistakes (chromosome mutations) are possible in packing and moving chromosomes, and the consequences are interesting. The severity of those mutations depends on which organisms we are talking about. Plants have a great capacity for tolerating chromosomal mutations, while in humans and other mammals, most major chromosomal mutations are lethal or have dramatic negative consequences. There are two major types.

- *Changes in the number of chromosomes*. Recall we have 23 pairs of chromosomes typically referred to by number, as in Chromosome 1, Chromosome 2, and so on (or in the moving-day analogy, Box 1, Box 2). If we move those boxes wrong, I might get an extra copy of Chromosome (Box) 3 and you might get none. Meiosis that results means an extra chromosome or missing chromosome in a sperm or egg turns out

to be important enough in humans that the majority of cases are lethal and miscarriage occurs, or even immediate failure of the fertilized egg to implant. The major exception in humans is Chromosome 21, in which an individual receives three copies of that chromosome and the result is a form of mental retardation called Down's syndrome or Trisomy 21. Other exceptions in humans are for the sex chromosomes refered to not as any number, but rather as "X" and "Y." Females have XX and males have XY. Sometimes there are extra or missing copies of Xs and Xs.

X0—Turner syndrome. Women with short stature and some fertility problems; also in many cases lethal in utero.

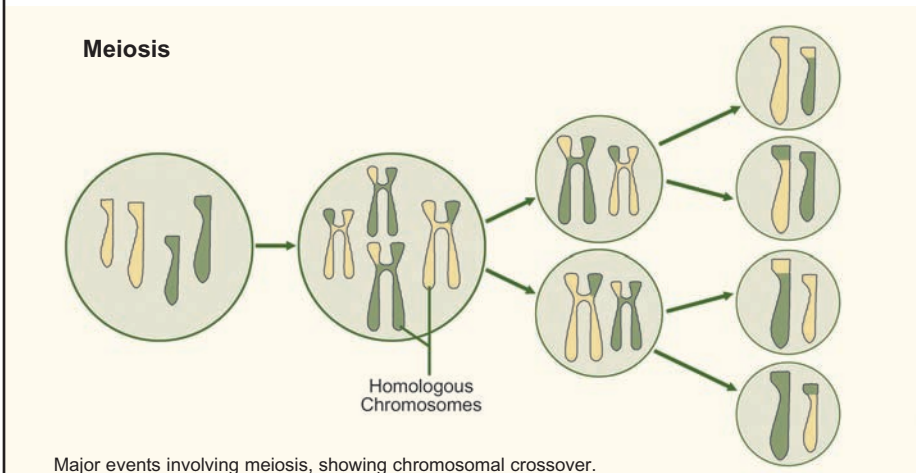
XXX or XXXX. These may be discovered as a result of investigating a fertility problem in women.

Xyy. Not often noticed because the characteristics are not typically followed up with chromosome studies: men with taller than average height, increased acne, and perhaps a tendency to some developmental delays.

XXy or XXXy—Klinefelter syndrome. Men with some ambiguous sexual characteristics and fertility problems.

- *The sex chromosomes and mosaics*. The sex chromosomes "X" and "y" present an interesting problem and solution in the way that the genes on those chromosomes are used. In particular, the X is a large chromosome with thousands of genes on it, including some for color vision and blood-clotting factors. Indeed, there is no particular association between the X and any sexual characteristics.

Males have just one X. That means that for color vision or blood clotting and a host of other characteristics, there is no "dominant" or "recessive" characteristics for genes on the X for males. Whatever version of an X chromosome gene they have manifests itself directly. That means color blindness and hemophilia are either present or not, but never "hidden" as recessives in



Major events involving meiosis, showing chromosomal crossover.

males. However, females have two copies of each X chromosome gene and therefore can be carriers of hidden recessive genes. That is why "sex-linked" genes are sometimes referred to as skipping a generation. Sometimes (but not always) they do skip because they may be hidden in the mother's generation and then appear suddenly in the son's generation.

Another consequence of two X chromosomes versus one is that the cells of females compensate to prevent an "overdose" of the genes on the X. That means that all over the body of females, in every cell (other than those leading to eggs), one or the other of the two Xs is permanently turned off. Which X is inactivated is random and it occurs very early *in utero*. That means that females are mosaics of characteristics of the X chromosome, especially for those genes in which there was a real difference in alleles—present, for example, as Aa or Bb.

It seems not to matter too much in hemophilia (a disorder of blood clotting) and color-blindness, in which the female has one copy of the normal gene and one of the mutated gene. All over her body she is either expressing normal blood clotting or abnormal blood clotting. However, those traits are not used all over the body but just in the bone marrow. The rest of the mosaic doesn't matter. Furthermore, in the bone marrow and in the retina there seems to be typically enough compensation from the normal cells to overcome the abnormal cells. This seems to be the case for most color-blindness too. The cells of the retina of the eye compensate so that females with one copy of a color-blind gene nonetheless see color.

That is why we switch to cats to get a really striking example of X inactivation and mosaics in action and highly visible. The gene for black or orange pigment is on the X in cats and are alleles of each other. Normal males can be either black or orange but never both. Females carrying both black and orange alleles are a wonderful mosaic (called tortoise shell) of black and orange splotches, randomly placed all over the body and of varying sizes. The size depends on how early the inactivation of one or another X occurred. If the cat was in very early development—just 4 to 8 cells—the splotches will be bigger than if it occurred later in the 16 to 32 cell stage. Add a gene for white (not on the X) to tortoise shell, and you get a three-colored cat called a calico.

It is because of X inactivation that identical twin female mammals, whether human or cat or any other, are not identical in respect to their X chromosome genes. Inactivation occurs randomly in each. This is especially important for traits in which mosaic might matter (although may not be visible to our eyes as tortoise

A tortoise shell calico kitten.

shell is). It is X-inactivation that makes the first cloned cat a rather unfortunate example. For some reason, researchers chose to clone a calico female, setting up a situation in which the cloned offspring could not possibly have the same coat color patterns. Sure enough, they look quite different, and keep in mind that coat color is just a visible manifestation (symbolic) of all the other thousands of genes on the X, also present in mosaic patterns.

- *Crossovers (recombinations)*. DNA, that amazing linear sequence of As, Cs, Gs, and Ts, arranged in gorgeous double helix, is surprisingly, disconcertingly unstable. If you have already had a chance to look at a three-dimensional model (such as at a museum), you probably have the impression of an amazingly stable construction not unlike a well-built set of stairs. Actually, DNA is capable of shuffling sequences from one strand to another (a process called crossover or recombination). Sometimes recombinations are between similar (homologous) strands, as in this English-language example:

  The cat sat on the hat.

  The dog ate on the mat.

  Recombine to make:

  The cat sat on the *mat*.

  Where "mat" crossed over.

At other times, DNA can recombine between completely different sequences, sometimes even between different organisms (part of a natural process that humans have called "genetic engineering"). For example:

The cat sat on the hat.
I like ice cream.

Recombining to:

The cat ice cream.

The last is a new combination that may or may not be nonsensical in a particular context. DNA is quite promiscuous about crossovers and does them without regard to making sensible combinations.

Although crossovers can happen any time, there is a special time right at the beginning of meiosis (the division to make sperm and eggs) during which homologous crossovers are facilitated and therefore occur in abundance. The consequences are significant. What it means (in terms of moving day with boxes of books) is that right before you start sliding those boxes to opposite corners of the room, there is a flurry of activity—book exchanges from box to box with no particular rhyme or reason except that it tends to occur with similar books (similar DNA sequences). So imagine at that crucial time, we are trading furiously for a few minutes—my Keats poetry book for your Yeats, my pancake cookbook for your waffle cookbook, etc. Within the context of this introductory course, there will not be much more elaboration on crossovers, which would lead to some advanced topics. However, I think the image of the book exchanges is useful even if just to keep in mind that DNA is full of surprises and complexities.

FOR GREATER UNDERSTANDING

Questions

1. How are chromosomes like moving boxes?
2. What is the difference between a gene and a chromosome?
3. What is the difference between DNA replication and mitosis?
4. When (or for what purpose) do cells do mitosis and when meiosis?

Suggested Reading

Schwartz, James. *In Pursuit of the Gene: From Darwin to DNA*. Cambridge, MA: Harvard University Press, 2008.

Other Books of Interest

Miller, Kenneth R., and Joseph S. Levine. *Biology*. Upper Saddle River, NJ: Prentice Hall, 2007. (Students are encouraged to read the sections on chromosomes. Any up-to-date introductory biology textbook will also serve for this purpose.)

Website of Interest

YouTube video clip of chromosomes moving through mitosis — http://www.youtube.com/watch?v=s1ylUTbXyWU

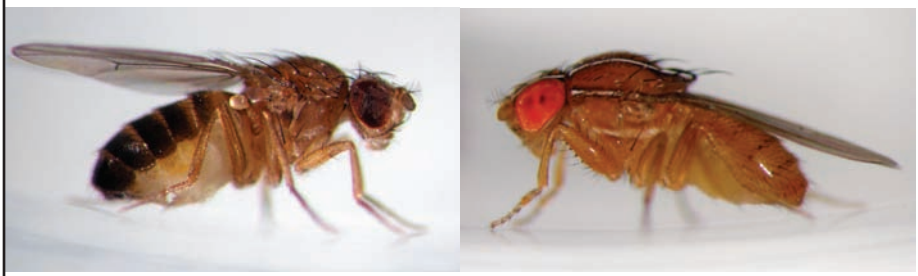
Lecture 11:
Mutations

The Suggested Reading for this lecture is James Schwartz's *In Pursuit of the Gene: From Darwin to DNA*.

The following is a quotation from the fruit fly geneticist Hermann Muler, taken from a lecture he gave in 1929. It is a *celebration* of mutations, which is a major take-home story of this course. The quotation appeared in *In Pursuit of the Gene* by James Schwartz:

> *All types of mutations, large and small, ugly and beautiful, burst upon the gaze. Flies with bulging eyes or with flat or dented eyes; flies with white, purple, yellow, or brown eyes, or no eyes at all; flies with curly hair, with ruffled hair, with parted hair, with fine and with coarse hair, and bald flies; flies with swollen antennae, or extra antennae, or legs in place of antennae; flies with broad wings, with narrow wings, with upturned wings, with downturned wings, with outstretched wings, with truncated wings, with split wings, with spotted wings, with bloated wings and with virtually no wings at all. Big flies and little ones, dark ones and light ones, active and long-lived and short-lived ones. Flies that preferred to stay on the ground, flies that did not care about the light, flies with a mixture of sex characteristics, flies that were especially sensitive to warm weather. The roots of life—the genes—had indeed been struck, and had yielded.*

Now it is true that in many circumstances, in particular contexts, any one of the above-named mutations could be considered negative. For example, flies "with virtually no wings at all" placed in competition with winged flies to get at some limited supply of food would probably be at a disadvantage. That is, *unless* the circumstances were such that flying was ridiculous. Let's say that the flies found themselves having to crawl into tiny crevices to get to the food



© Nicolas Gompel

Drosophila is a genus of small flies, belonging to the family *Drosophilidae*. The red-eyed fly (right) is *Drosophila melanogaster*, which has been heavily used in research in genetics and is a common model organism in developmental biology. The more darkly pigmented fly on the left is one of many that involve mutations to the pigment pathway for eye and body colors. The terms "fruit fly" and "*Drosophila*" are often used synonymously with *D. melanogaster* in modern biological literature. The entire genus, however, contains more than two thousand species and is very diverse in appearance, behavior, and breeding habitat.

and that the richest sources were in the narrowest passages, excluding the flies with bulky wings. Indeed, plenty of insects use walking as the preferred method to get to food and flying is more of an occasionally used evasive tactic for avoiding predators. In many documented cases, insects suddenly on a remote island (having been blown or washed there), and without their natural predators, gradually become wingless. That is, when wingless mutations appear accidentally, those insects end up doing a little better than their relatives, still lugging around large useless appendages.

It turns out that fruit flies—and probably nearly any insect that can be cultured in the lab—are outstanding subjects for the study of mutations. This may be because insect and other arthropod bodies have a somewhat modular construction, such that mutations to one part are isolated from other parts. Thus you can have a fly with misplaced antennae or extra wings or extra segments but otherwise normal parts. If you take good care of them in the lab (lots of food and mates and no competition) they do just fine and are easy subjects for all sorts of investigations. Fruit fly researchers have a long tradition of rejoicing in their mutants and even eagerly anticipating as yet undiscovered ones for the privilege of naming them. Some fly labs even save a list of good names, awaiting assignment to the right mutant. Here are a few of the examples. Longer lists easily may be found online.

Tin man: Embryos developing (albeit briefly) without hearts.

Methuselah: Long-lived flies.

Amontillado: Referring to the Edgar Allan Poe story; embryos unable to hatch, that is, walled in.

Van Gogh: Flies having swirly bristle patterns.

Road block, *Gridlock*, *Sunday driver*, and *Red Tape*: All referring to blocked neuron communication.

Yippee: From the margin of a lab notebook of an exuberant graduate student.

Here are some things I'd like to reinforce before we do a little analysis of how "mutant" came to be such a negative word in common parlance. All genes are mutants—no exception. Maybe somewhere, sometime, four billion years ago at the origin of life there was some little sequence of DNA that qualifies to be called the original sequence, the one from which all others derive. But after that it's nothing but mutations all the way. The entire mechanism of replication is set up so that it is impossible to *not* have mutations. It is a normal situation. Alleles of genes are mutations; variations of genes (same as alleles) are mutations. Mutations are the reason we have millions of species of organisms and why there is so much variation within each species.

Perhaps it is right that the small subset of negative, detrimental mutations get most of the research money and the attention of the popular press. After all, those are the mutations that cause cancers, birth defects, and other tragic genetic conditions that could be cured if we understood them better. However, those are the exceptions, and a grander more inclusive understanding of mutations may help to better inform us when investigating the few noteworthy negative ones.

Science fiction writers may be in part responsible for the popular misconceptions about mutations. According to the *Oxford English Dictionary*, "mutation" was used from the fourteenth to the eighteenth centuries to mean "change." In the nineteenth and twentieth centuries, scientists trying to make sense of genetics began to use mutation to talk about differences in traits of organisms. In 1938 (according to the *OED*), *mutation* was first used in science fiction to imply something freakishly abnormal. However, the stage already was well set for such a use in that fantasy fiction of the nineteenth and twentieth centuries abounds with images of the grotesque and bizarre, as in *Frankenstein*, *Dr. Jekyll and Mr. Hyde*, and *The Island of Dr. Moreau*. It was probably a fairly easy step to begin assigning the word *mutant* to monsters.

A misunderstanding of the word *mutation* has helped fuel misunderstandings about evolution. A common question is, if mutations are so bad, how could incremental accumulations of mutations lead to anything good? Look back to lecture 7 to be reminded of the importance of the environment in selecting mutants, not necessarily the very best ones (if such a thing exists), but rather the ones that are merely good enough for a particular place and time. Evolution is not about attaining perfection and mutations are "good" or "bad" only within the context of their use. The majority of truly negative mutations are selected out, leaving those that work well enough that the organisms containing them can pass along their DNA to the next generation. The proof is in the diversity of extant organisms, the products of millions of mutations and relentless selection.

Another misunderstanding of mutation is about the significance of their random occurrences. Again, the question is, How could random changes, whether good or bad or neutral, ever lead to anything like a functional change in an organism? The idea of selection is usually the missing component in this question. If you flip a coin, the appearance of heads or tails will be random. If you flip one hundred coins, the results will still be random. However, if you add in selection, such that for some reason in this particular environment, at this time it is advantageous to be heads and disadvantageous to be tails, then by and by you would have a pile of coins consisting only of heads.

The naming conventions for most mutations of any organism (including the whimsically named fruit fly mutations) follow certain conventions that sometimes inadvertently introduce misunderstandings. So in this part of the lecture, some of those conventions will be deconstructed or demystified, such that when you encounter mutant names you will be able to place them into greater context of your understanding of genetics.

Mutants are typically named backwards.

Take a look again at those fly mutant names. None of them sound very "normal." All imply some sort of aberration. Thus you might be led to believe that all mutations are about deviations from some normal trait. (And this is yet another source of the idea that mutations are bad). However, the names of mutants have more to do with the nature of genetic analysis than with the actual function of the mutant gene. Recall the cookie factory in the black box. If all is running smoothly (normally) and edible cookies are emerging from the back door, it is extremely difficult (with ordinary genetic analysis) to

figure out what each machine is doing. It is only when the machines malfunction one at a time or in various combinations that we can get a glimpse of what the machines were supposed to be doing. So our first look at the dough-mixing machine is really a look at its mistake—badly mixed dough—by which we infer that when it is operating correctly dough is mixed correctly. And now that we have had our first look at the dough machine, it is time to name it. Unfortunately, what we understand best about the machine so far is going to be reflected in the newly coined name: "Unmixed dough." Months or years later we may acquire a much more nuanced understanding of the dough-mixing machine, but too late, the original mutant name withstands.

Here is an example from biology. There are many mutant versions of genes named "cancer genes" of one sort or another and these names (being for genes of great interest) tend to appear in the popular press. You might wonder how it could be that we humans are carrying around such a load of cancer genes and may ask, where did they come from? Actually, none of them typically are cancer genes; they were just named that way because researchers caught their first glimpse of them going awry in the context of cancer. "Cancer genes" tend to be cell-division genes, either facilitating normal division or controlling the rate of normal division. As long as cell division is proceeding along correctly, which it does the vast majority of the time, it is as difficult to decipher as a well-functioning cookie factory in a black box.

Mutants often are named too simply for what they are.

This point may seem obvious, but nonetheless sometimes needs to be pointed out to dispel misunderstandings. Jargon is a sort of shorthand code by which people who understand the jargon can communicate efficiently. Let's take the example of "colon cancer gene," which might appear in the popular press. The name has the appearance of singularity as though it were solely responsible for a trait and therefore a sort of lynch pin or keystone by which an entire disease might be finally understood and cured. However, by now you are used to my evocation of the cookie factory. Cell division is complicated. There are many ways for it to go wrong. Any number of errors might result in one or another cancerous condition. This is not "the" colon cancer gene (or mutation), but "a" colon cancer mutation, one of many in an intricate pathway, by which *normal* cell division in the colon is regulated. And it is typical of the way in which scientific research is published, that the report of "a colon cancer gene" will precede by months and years any further understanding of the greater process of which it is a small part.

Mutants often are named with ridiculous acronyms.

Using long acronyms is the other extreme of naming that, unlike "cancer gene" or "Van Gogh" (the fruit fly with swirly bristles), almost completely excludes lay persons from any discussion. The custom began innocently enough with bacterial mutations, of which there were just a few for many decades. Therefore, simple three-letter acronyms worked just fine and we had not yet run out of letters in the English alphabet. For example, "pen" was the acronym for bacteria mutants with some particular response to penicillin. Adding superscripts made some distinctions: pen^r means penicillin resistant and pen^s means penicillin sensitive. So far, so good. Except that now there

are hundreds of thousands of bacterial genes (mutants) named and we have run out of letters of the alphabet and acronyms are part of an insider system of arcane knowledge. And the situation is by no means limited to bacteria. For many other model genetic organisms, the acronym system is in use, extravagantly so. It took me about a minute in a professional database of biological literature to come up with this mouse mutant: "Cftr-/-hCFTR-G542X." To deconstruct it a bit, "Cf" means cystic fibrosis and the "/" in the middle indicates that the name refers to two genes. Recall that in complex organisms there are two sets of each gene. So that notation is a little like B/b. Also relevant is that this mutant mouse is something of a human construct; that is, humans put together this particular set of mutations in order to get a closer look at a phenomenon relative to cystic fibrosis. What to do about this as a layperson? On one hand it is a wonderful thing that good research is being done on cystic fibrosis and however the experts want to name their mutants really is just fine. But if you should find yourself in a conversation with a genetics researcher who begins to sling around acronyms, you are perfectly within your right to tell the person to define their acronym and then perhaps for the duration of the conversation to use some simple English phrase to stand in for the acronym. Scientists often need a reminder of this sort of thing.

The hundreds of completed genomics projects are changing the way we name mutants (albeit very slowly).

We humans have only about twenty-five thousand genes and so do mice and fruit flies and tiny worms called *Caenorhabditis*. Many of the really fundamental genes, such as those that control cell division, are in common. That is why we can learn so much about the basics of human biology by exploring the same phenomena (often much more easily) in flies or worms or

mice. The more we get a look at the sequences of genes for these diverse (yet not so diverse) organisms, the more we know this to be true. However, for many decades researchers went separate ways in their naming systems for mutants. Therefore, a set of human mutations will have one set of names. That very same set of mutations in flies will have another set of names (amusing ones most likely), and the same set in mice will go by a completely different set of names. Now imagine wishing to do a comprehensive search to gather together *all* of the information available on a set of mutants. You will need to search each mutant under several names. If you want to set up a database to pull together the information, you will need many extra columns to accommodate all the redundant jargon. Slowly the situation is being rectified and perhaps some day there will be a few hundred thousand logically named genes that can be easily cross-referenced from one species to another. It would be nice if each gene name reflected something about the actual work that the particular "machine" was doing in the context of all of the rest of the machines in a pathway. One example is the "white eye" mutant of fruit flies. "Normal," wild fruit flies have brick-red eyes. White-eyed fly mutants were discovered almost one hundred years ago and the name is evocative of the trait. However, now we know that "white" is a gene that codes for an "ATP binding cassette transporter" that moves substances into cells, not just into fruit fly cells but also human, mouse, worm, and so on. And it is a very important transporter not just for making fruit fly eye colors, but for all sorts of behavioral traits in many species. Wouldn't it be nice if we could research ATP binding cassette transporters without having to learn dozens of different names for the same phenomenon? That is the goal of many researchers trying to organize the databases of information from gene projects, striving for a more unified, more searchable naming system that ultimately will facilitate research.

FOR GREATER UNDERSTANDING

Questions

1. How does the mechanism of replication make it impossible to *not* have mutations?
2. What are the main problems with the naming conventions for mutants?
3. Explain the backwards-naming tradition for many mutations. (Connect this back to mishaps in the cookie factory.)
4. Explain the problem with declaring a mutation or gene to be "the" gene and not "a" gene.
5. Why are mutations basically a positive phenomenon in spite of the commonly held belief that they are negative?

Suggested Reading

Schwartz, James. *In Pursuit of the Gene: From Darwin to DNA*. Cambridge, MA: Harvard University Press, 2008.

Other Books of Interest

Venter, Craig. *A Life Decoded*. New York: Viking, 2007.

Website of Interest

The University of Georgia website provides examples of the creative naming system used for *Drosophila* — http://jpetrie.myweb.uga.edu/genes.html

Lecture 12:
Regulating the Genes and the Dramatic Consequences of Regulatory Mutations

The Suggested Reading for this lecture is James Schwartz's *In Pursuit of the Gene: From Darwin to DNA*.

The topic of how genes are regulated is advanced, complex, and not even all that well understood by researchers. However, the consequences of gene regulation (and of gene regulation gone awry) can be highly visible and often readily interpretable even by laypersons. Therefore, the goal of this lecture is not to dwell on the jargon-laden, intricate mechanisms of gene regulation, the background for which might take an entire course. Instead, the goal is to empower you to talk about the basics of gene regulation using helpful metaphors and to recognize confidently the traits of gene regulation in organisms all around you. Gene regulation is one of my favorite topics to contemplate and the mutations of regulation are among the most interesting. Many are quite distinctive and (at least for those not involved with disease) are even surrealistically beautiful. As for the disease-causing regulation problems, they are among the most important, including some major, debilitating birth defects.

Recall the cookie factory. It would seem that we have covered nearly every aspect of it: the machines (functioning or not), the blueprints for the machines, the replication (with errors of those blueprints), the shuffling of boxes of blueprints on moving day so that they are divided equally, and the importance of environment to the functioning of the factory both short and long term.

Here is what we have not yet considered. These blueprints (those genes) are actually pretty simple in their function. Recall that a gene is that which can be transcribed and then translated to make a protein. To switch metaphors here (from cookie factory metaphor to a grammar metaphor), a gene is a simple imperative statement like this:

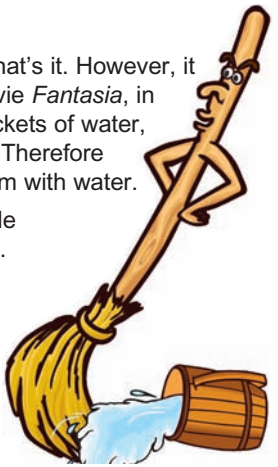
Make this.

Make what? Make whatever protein is encoded. And that's it. However, it isn't enough. You might remember the Walt Disney movie *Fantasia*, in which Mickey Mouse commands the broom to fetch buckets of water, but does not indicate *how much* water or *when to stop*. Therefore the broom (and then hundreds of brooms) flood the room with water.

That is what is missing when genes are used as simple commands with none of the extra nuances of grammar.

What we really need is the following:

- Make this.
  - If it is needed.
    - As many as are needed or until we have this number.
      - If it is in the right place.



© David Krupa/shutterstock.com

- If this is the right time.
- If there are other genes making their particular products nearby.
- And so on, and so on.

In order to use genes precisely and in a coordinated fashion with other genes and with the environment, there must be grammar of instruction for use. That is gene regulation. The grammar exists on either side of (upstream and downstream) nearly every gene. Recall that humans have about three billion bases of DNA, but only about twenty-five thousand genes. Those genes account for less than 10 percent of the entire genome. What is the other 90 percent of DNA doing? We don't know what most of it is doing, but some of it is providing the grammar for gene use. There are short sequences in some of that DNA that can be juxtaposed with other short sequences and with proteins such that they form "complete sentences" of instructions:

> Make hemoglobin, if you are in a red blood cell, and if more hemoglobin is needed and until we have enough.

The "make hemoglobin" part is the gene. All the rest resides up or downstream of the gene. How does the sentence get put together? Recall that DNA is linear only in our imaginations. It is really a tightly coiled, dynamic ball of twists and kinks. At any given moment, short sequences plus associated genes may find themselves all kinked up in the same tiny area, forming a sort of nuanced command.

The nature of gene regulatory grammar seems to be as fuzzy as that of any natural language, like English. For example, there are many more than one ways to say something, including ungrammatical and misspelled ways to get a point across. Don't envision it as being as precise as a computer program. Rather, it is like noisy party chatter, with all sorts of extraneous noise and overlapping, mostly unintelligible talk. There are lots of redundancies, lots of ways to say the same thing and more or less get a good enough result. That is, in part, a source of the variations we see from individual to individual. And it is why it can be difficult to tease apart and decipher the grammar, because if one part of a "sentence" is mutated, there may be other ways to "say" the same thing.

Exactly how all this looks on a molecular level is too advanced a topic for a single lecture. However, the consequences of regulation (whether it is going right or wrong) are something fairly easily visualized.

Gene regulation in a normal multicellular organism.

"Elbow/eye" is your mnemonic for this example. You have the very same set of genes in a cell of your elbow as you do in a cell of your eye. Yet those genes are being used in different ways to make either an elbow or an eye. It is as though each cell had an entire dictionary of words (or simple commands) as any other cell, but was writing a different essay, for instance, "The Essay of the Eye," using those words. The grammar and syntax are what make the essays different. Any functional multicellular organism with its body parts (both internal and external) in the right places and with its functional pathways operating in the right sequence is testament to the power and importance of gene regulation. That said, it is through mutations to gene regulatory sequences that we have a better view of how it was working.

Mutations to gene regulatory system in humans and other mammals.

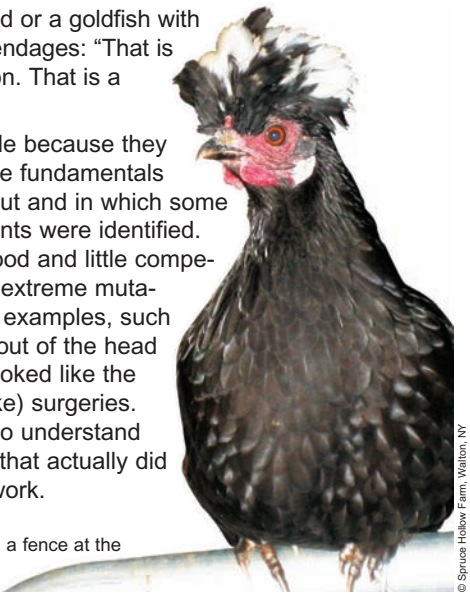
These are birth defects in their most extreme forms or they are just interesting variations in morphologies and placements of structures in the milder versions. For example, “charcot-marie tooth disease” is a birth defect of limb development including muscles and nerves. It is caused by failures to regulate large sets of genes crucial in embryonic limb formation.

However, most variation is within some defined range of “normal” and simply gives us interesting genetic differences such as in shapes of noses and ears and faces in general. Some of the most dramatic examples are found in pedigree dogs, in which human breeders have been enormously successful in getting dog faces throughout the range from truncated (flattened, even concave) to elongated. As always with most extreme dog morphologies, such mutations are accompanied by negative side effects such as problems with eating and breathing. Where does aesthetics cross the line into birth defect (in the most detrimental sense)? It should be kept in mind that many extreme breeds of dogs would never survive on their own in the wild in competition for scarce resources with other canines. Those breathing and eating problems would perhaps be better defined (under those circumstances) as lethal or nearly so.

Mutations to gene regulatory system in insects, birds, and fish.

Once you know what you are looking for, you will see these everywhere. Go to a pet store or a poultry show or leaf through any comprehensive picture book on a particular type of animal. Notice the wonderful diversity of shapes and positions of structures. Notice especially the extremes of the “fancy” domesticated animals. Maybe you can’t explain exactly what grammatical elements are at play. Don’t worry. You are in good company. Most researchers don’t have the full explanations yet either. But rest assured you can say with some confidence when pointing to a chicken with an enormous powder puff of feathers covering its head or a goldfish with its eyes perched on periscope-like appendages: “That is not caused by an ordinary gene mutation. That is a gene regulatory mutation.”

Fruit flies are their own special example because they are the organisms for which many of the fundamentals of gene regulation have been worked out and in which some of the first examples of regulatory mutants were identified. Also, flies in captivity (given plenty of food and little competition) can survive quite well with some extreme mutations. When some of the most extreme examples, such as antennapedia (having legs growing out of the head where the antennae should be), they looked like the results of little diabolic (Frankenstein-like) surgeries. Indeed, some early experiments to try to understand these mutations were surgical projects that actually did elucidate some of the mechanisms at work.



A white-crested black Poland chicken perches on a fence at the Spruce Hollow Farm in Walton, NY.

The regulation of fruit fly genes is very similar to that of human genes, so much so that a "grammar gene" for the development of eyes in humans can be placed into a fly embryo and it works. The proof is in a dramatic experiment in which researchers inserted that human gene everywhere in the body of an embryonic fly and then "turned on" the gene simultaneously in a dozen different places to see what would happen. What happened was eyes (albeit somewhat misshapen ones) popping out all over the fly. By the way, the wingless flies and the "tin man" (heartless) flies discussed in lecture 11 are also regulatory mutants, as are nearly any arthropod with multiplied or modified body parts. Such mutations have been essential in arthropod evolution, for example, resulting in a wide range of numbers of legs in millipedes, centipedes, crustaceans, spiders, and insects.

Mutations to gene regulatory system in plants.

Plants seem to have a great tolerance for regulatory mutations, so much so that we can easily take them for granted and not even notice how strangely varied plant parts can be. Think about all those *Brassica* in the produce section of the grocery store—cabbage, brussels sprouts, turnips—all derived from a skinny wild mustard. Regulatory mutations are the key to how such variation in morphology occurred.

There are many types of regulatory mutations in plants, but one of my favorites is "fasciation," coming from the word *fascia*, meaning holding together a bundle of sticks. (The word *fascism* also is derived from fascia.) Think of a single thin plant stem, say of a wild mustard, and multiply it many fold on one thick stem, looking a little like conjoined twins. That is how we get the thick stems of broccoli and cauliflower. Multiply the leaves of *Brassica* so that they are packed into a dense sphere and you have cabbage. Multiply small leaf-spheres more subtlely and in a more spaced-out manner along a stem and you have brussels sprouts. An entire grammar of multiplications are displayed in the produce bins of your grocery store.

Have you ever noticed that the fruits of some large heirloom tomatoes and peppers look like two or three fruits melded together? And when you cut them open, there appears to be more than one central core of seeds? That's fasciation too. Now and then, if you are examining a field of dandelions quite closely you may notice multiplied stems and flowers: fasciation. And fasciation is your one-word answer to "How do we get four-leafed clovers?"

Daffodils are one of my favorite plant regulatory mutants and often many different types are featured in the spring at botanical gardens (although any domesticated flower is likely to have just as dramatic variations). Look through a comprehensive picture book of your favorite flowers and be amazed at multiplications and juxtapositions of petals. In the case of daffodils, they can range all the way from pom-pom-like multiplications of dozens of petals to petite, subtle blossoms in bell-like clusters along a stem.

FOR GREATER UNDERSTANDING

Questions and Activities

1. How is the “elbow/eye” mnemonic helpful in understanding gene regulation?
2. What happened when a “grammar gene” for the development of eyes in humans was inserted in an embryonic fly?
3. Plan a garden of regulatory mutations that displays as many varieties as possible of one plant type, with a focus on size, shape, and numbers of structures.
4. Come up with your own repertoire of examples of regulatory mutations for either favorite plants or animals. You may not be able to explain exactly which mutations are involved but consider major changes in morphology to be due to some difference in the regulation of sets of genes.

Suggested Reading

Schwartz, James. *In Pursuit of the Gene: From Darwin to DNA*. Cambridge, MA: Harvard University Press, 2008.

Other Books of Interest

Andrews, Chris. *Guide to Fancy Goldfish*. Surrey, UK: Interpet Publishing, 2002.

Green-Armytage, Stephen. *Extraordinary Pigeons*. New York: Harry N. Abrams, Inc., 2003.

Website of Interest

1. Spruce Hollow Farm in Walton, NY, provides an excellent website that includes photographs of a variety of animals living at the farm and showing variations from breeding — http://www.sprucehollowfarm.org
2. Aunty Animal website (Small Animal Fostering and Rescue), provides images of a wide variety of small animals, including many from fancy breeds — http://www.auntyanimal.me.uk/animalpics.htm

Lecture 13:
Viruses, Jumping Genes, and Symbionts: Whose DNA Is It and What Are the New Philosophical Questions?

The Suggested Readings for this lecture are the "Microbe Cards" available from the American Society for Microbiology.

Professor Dyer: *Introductory biology textbooks will have a section or chapter on viruses. Note however that the emphasis will be on pathogens. This will be true also for textbooks devoted entirely to virology. The field of virology is slowly catching up with the abundance of presumably non-pathogenic viral DNA in our genomes and eventually may give at least equal treatment to non-pathogenic viruses.*

Brace yourself. We are near the end of the course and you are about to find out that most of the DNA we are hauling around is not ours. In fact, we may have to come up with new definitions for "we" and "us" and "our." An editor of the journal *Nature* (May 29, 2008) wrote:

> *There is a compelling new take on humankind's place in the world—a realization that "Who am I?" cannot be fully answered until it is fully understood who "we" are.*

The editor was referring to "the wild profusion of bacteria, fungi, and viruses that colonize the human body; unseen passengers number in the trillions. They contribute so much to human biology that it is difficult to say where the body ends and the microbes begin."

This lecture is an introduction to the complex, advanced topic of what it means to have so many genes, so much DNA that is not strictly our own.

Our genomes are mostly viral.

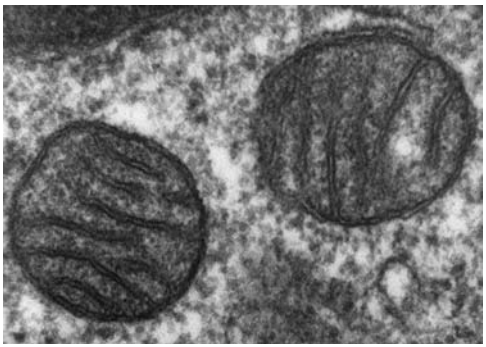
This was and continues to be one of the big surprises from the completion of the human genome project. We are not quite sure what it means. We just know that it is the case. We have three billion bases of DNA sequences (actually six billion if you consider that we have double copies of all of it). Less than 10 percent of six billion comprise genes (of which we have twenty-five thousand) that we consider to be "human genes," although most are very similar or identical to those of other animals. A certain unknown amount of the human genome is about gene regulation. However, most of the human genome, five-sixths of the six billion bases, has been identified as viral or viral-like, or perhaps derived from viruses. This includes actual identifiable viruses and pieces of viruses as well as a category of DNA called "transposons" or "jumping genes," many of which seem to be of viral origin. Viruses and transposons have (if they are fully functional) independent activities, sometimes involving replicating on their own or even jumping around from one section of the genome to another by recombinations. (See recombinations briefly discussed in lecture 10.) However, many or most seem to be not particularly active and simply wait passively for the entire genome to be replicated, which of course includes themselves. However, there is much we do

not understand about these viral entities. Consider this a frontier of genetics, If you are chatting with a geneticist at a party, ask for his or her opinion as to what all that viral DNA is about. Meanwhile, here is a tally thus far:

Total DNA per human cell = 6 billion bases
Viral or viral-like DNA per human cell = 5 billion bases
Our genes = 600 million
The rest of it—possibly regulatory DNA = 400 million

Our cells are full of bacterial symbionts with their own genomes.

About 2.5 billion years ago, complex cells such as what make up humans and other animals, fungi, and plants began to evolve. This greatly predates multicellularity. The first complex cells were single amoeba-like cells. The evolution of complex cells occurred primarily through symbioses between bacteria, which, up to that point (since the origin of life four billion years ago), were the only organisms. One major change in the transition from bacterial to complex cells was the nucleus, a protective interior compartment in which DNA may be stored and used. Another major change was the acquisition (by the ancestral complex cell) of bacterial symbionts capable of performing an especially efficient form of energy metabolism abundantly enough to support themselves and their host. These symbionts, called *mitochondria*, are roughly analogous to little furnaces burning fuel for energy and transferring some of that energy to various cell functions. Each of our active cells is full of mitochondria, without which there would be insufficient energy to maintain all of the functions we need. Every time our cells divide, our mitochondria each divide too and thus maintain their numbers. Mitochondria use oxygen as an essential part of their activity. A lack of oxygen shuts them down and soon after every single one of our essential, energy-using cells shuts down too. After six minutes or so without oxygen, it can be lethal. That's how important mitochondria are—indeed, so important they can be easily taken for granted. So let's say in each human cell are about one hundred mitochondria and in each mitochondrion are about ten copies each of a genome about twenty thousand bases in size. That means each of our cells has twenty million bases of ancient bacterial (now mitochondrial) DNA. And most of it is composed of genes. None appears to be viral. Compare that to the DNA representing our own genes from the calculation above:



Transmission electron microscope image of a thin section cut through an area of mammalian lung tissue. The high magnification image shows a mitochondria.

Mitochondrial DNA per cell, mostly mitochondrial genes:
20 million bases
DNA for our genes per cell: 600 million bases

Maternal inheritance of mitochondria and mitochondrial diseases.

Each of our cells carries about one hundred mitochondria, which provide most of our energy needs. They are so important that a lack of function is lethal. Decreased functions can be seriously debilitating. In healthy humans, mitochondria are easy to overlook because the essential assumption must be that they are there and working efficiently enough to keep us alive and active. However, there are plenty of subtle variations in mitochondria (due to variations in mitochondrial DNA) from person to person. At one extreme, there are debilitating mitochondrial diseases that affect those tissues that need the most energy, such as muscles and nerves. Some milder differences in mitochondria may be manifested in variations in our abilities to process energy. These may result in differences in athletic ability or in fat metabolism or even in rates of aging, involving maintenance and renewal of tissues. Some of these traits are of such general importance and part of such complex sets of cell functions that it can be difficult to sort out the mitochondrial contribution.

One of the first things noticed about some mitochondrial diseases (even before their mechanisms were understood) is that they are inherited maternally. This is due to the size difference between sperm and eggs. An egg arrives to fertilization loaded with mitochondria. A tiny sperm arrives with just one, enough to provide energy to propel itself. That single sperm mitochondrion is either lost or its effects are greatly diluted at fertilization. The number of egg mitochondria is overwhelming. They become the mitochondria of the new embryo. This means that if a woman has a mitochondrial disease but is healthy enough to have children, she will pass that disease to them. Meanwhile, a man with a mitochondrial disease will not. How about tendencies to obesity or athletic ability or other complex traits that sometimes are associated with mitochondria? Are those inherited maternally? Those important traits rely on so many different genes, both our own and those of the mitochondria, that inheritance does not get manifested in patterns that simple.

Meanwhile, the DNA of mitochondria, passed down from mothers to offspring (of both sexes) for many generations serves as a good genealogical record. By analyzing changes in mitochondrial DNA, it can be established that humans (as a nearly modern species) began with a small population in east Africa about one hundred thousand years ago. Sometimes the phrase “mitochondrial Eve” has been used as a metaphor. However, it was not just one woman, but rather a group of closely related women, from whom all of us (males and females) received our mitochondria.

Our body surfaces and cavities are loaded with hundreds of species of apparently beneficial (or at least benign) bacteria.

We humans have about 10 trillion cells. However, we are hauling around within our body cavities and on all of our surfaces about 100 trillion bacteria. Bacterial genomes are smaller than those of complex cells, but they also tend to be more densely packed with genes (albeit fewer of them), compared to ours. Let’s say bacteria have genomes of an average size of one million bases, and let’s give them just one copy, although lots of bacteria have multiple copies. So how much bacterial DNA are we carrying?

10 trillion (“our” cells) x 6 billion (our genome, shared mostly with viruses)

100 trillion (“their” cells) x 1 million (their genomes)
Total for “us” = 6×10^{22}, of which 6×10^{21} are “our genes”
Total for “them” = 1×10^{20}, most of which are “their genes”

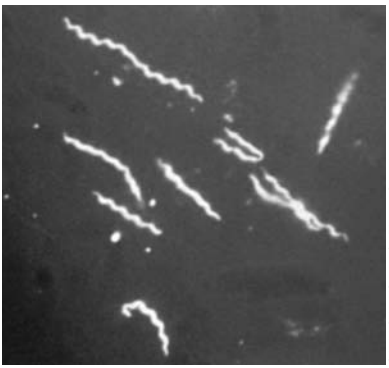
Okay, we (humans) win, by an order of magnitude. But that is far from providing any complacent satisfaction that we are independent genetic entities. Indeed, since bacteria outnumber us 10:1, we are considered by some bacteriocentric people as bacterial communities with human bodies attached.

The genetic significances of our great diversity of DNA (bacterial and viral) is mostly unknown, but here are some of the more tangible consequences:

1. Our immune systems are kept on alert so that when we encounter pathogens we are more ready to deal with them, which is better than if we lacked bacterial symbionts.
2. Our immune systems are somewhat less likely to produce inappropriate responses such as asthmas and allergies, because a constant presence of bacteria seems to keep the immune system more finely tuned to respond specifically to pathogens.
3. Our digestion and use of nutrients may be more efficient.
4. Our myriad symbiotic bacteria may be placeholders of sorts, preventing pathogens from getting an easy place to establish themselves.

A new understanding of bacterial pathogenicity.

The vast majority of bacteria on Earth (perhaps millions of species) are not pathogenic and have no particular associations with us. The few that are notable pathogens (about fifty species) have some particular characteristics. The most serious ones, such as *Yersinia pestis* (black plague), *Borrelia burgdorferi* (Lyme disease), and *Mycobacterium tuberculosis* (tuberculosis), have the ability to get deep inside our tissues, sometimes by being injected in. Lyme disease bacteria are injected by ticks; plague bacteria are injected by fleas. Often the serious pathogens have no particular free-living existence and have uncanny abilities to bypass our immune system defenses. Tuberculosis bacteria actually live inside our cells. Many serious pathogens are out of synchrony with the usual rhythms of our cell divisions. Rather than waiting (as mitochondria) do to time their divisions with our own, and thus maintain steady manageable numbers, pathogens overrun their hosts using up resources and releasing toxins. Many appear to be “new” in their relationships with us. These characteristics are so remarkable that they provide clues as to



This photomicrograph reveals the presence of spirochete, or “corkscrew-shaped” bacteria known as *Borrelia burgdorferi*, the pathogen responsible for causing Lyme disease. These bacteria are transmitted to humans by the bite of an infected deer tick, and caused more than twenty-three thousand cases of Lyme disease in the United States in 2002.

what makes a pathogen and how pathogenicity evolves. Being a pathogen (especially a lethal one) is actually a tough life (for the pathogen). Every time they kill a host, they are left (possibly) in a dead end situation, unable to get to a new host. The tendency in the evolution of bacteria/human relationships seems to be one of increasing synchrony and subtlety, away from extreme pathogenicity and closer to benign and even beneficial interactions. The vast majority of our personal bacteria are in just those sort of benign to beneficial relationships with us.

A new understanding of viruses.

The lessons of bacterial pathogenicity may be applied to viruses and indeed must be applied if we are to understand the enormous load of viruses we carry in our genomes. The history of virology has been a history of pathogenicity for practical reasons: we cannot easily see viruses, but rather we detect their presence by looking for their activities. The easiest viral activities to observe are those that disrupt some normal situation; that is, we see disease and nothing else of viruses. Therefore, there is a long tradition of equating viruses with diseases. There simply were no ways to "observe" viruses in benign or beneficial relationships. It now appears that the vast majority of viruses (like the vast majority of bacteria) are not associated in any way with humans. Of those that are associated with humans, the vast majority are benign or maybe even beneficial in ways that we do not yet understand. There are about twenty-five seriously pathogenic viruses. Like the bacterial pathogens, these unusual (but well-studied and publicized) viruses tend to get injected deep into tissues, evade immune system defenses, and perhaps most importantly, have new relationships with us and a lack of synchrony with our cell processes. Just as with bacteria, this is a tough existence for a lethal virus. The evolutionary tendencies seem to have been in the direction of greater synchrony and more subtle manifestations (or none at all). The evidence is that our DNA is mostly viral and yet we are not all suffering from constant, relentless viral pathogenicities.

By the way, you may need a handy answer to the old question (which actually is a false dichotomy): Are viruses alive or not? You cheerfully answer: "Neither." They are information encoded in DNA or sometimes RNA. Viruses are genetic entities just as genes are. And we don't ask about genes, "Are they alive or not?"

Plant viruses that you can easily observe.

Plants as usual are amazingly tolerant of some rather exotic (from a human point of view) genetic conditions. Some viruses and some transposons or jumping genes (which are virus-like) implant themselves in pigment genes and other plant genes. They pop in and out, resulting in splashy variegation of color in leaves and flowers. Look through any comprehensive flower catalogue and watch for the most extravagant colorings: splotches of one pigment against another, streaks of color, irregular spots, the more unstable and bizarre looking, the better. Whether or not it gets acknowledged in the print of the catalogue, the showiest variegations are probably due to viruses or transposons. One hint might be if the colors are described as not breeding true or unstable.

At the art museum, look for seventeenth-century Dutch paintings of tulips. The wildly variegated "Rembrandt type," some versions of which can still be purchased today, were a huge sensation in the seventeenth century. Speculation on exotic tulips (complete with fortunes being made and bankruptcies being filed) was called "tulipmania." The tulip variegation that was so desired by speculators was due to viral or transposon activities.

© Sebastian Barnard/shutterstock.com

A "Split" tulip—a modern variety of the seventeenth-century "Rembrandt" tulip.

In the grocery store, you may find decorative colorful ears of corn at harvest time. Look closely, perhaps at hundreds of colored kernels to find just one kernel with speckles, due to transposons. That is how transposons were first discovered by Barbara McClintock, as rare events in corn pigmentation. For many years, transposons were mistakenly thought to be just that: rare events occurring in plants and nothing much more. It makes all the more surprising the news of their ubiquity and abundance in all genomes, especially of complex organisms.

© Elena Elisseeva/shutterstock.com

FOR GREATER UNDERSTANDING

Questions

1. What effect does lack of oxygen have on mitochondria?
2. Why are mitochondrial diseases inherited maternally?
3. What is a virus?
4. Are viruses alive? (Or explain why that is not an appropriate question.)

Suggested Reading

The American Society for Microbiology (http://www.asm.org) distributes "Microbe Cards." These are flash cards of viral, bacterial, and other pathogens. They are a quick source of information by which one might try to further understand some of the common parameters of pathogenicity.

Other Books of Interest

Comfort, Nathaniel C. *The Tangled Field: Barbara McClintock's Search for the Patterns of Genetic Control*. Cambridge, MA: Harvard University Press, 2003.

Website of Interest

The National Center for Biotechnology Information (NCBI) provides the Online Mendelian Inheritance in Man page; for information about mitochondrial mutations, use "mitochondria" as a search word; note that some hits will involve mitochondrial conditions that are actually coded for by genes in the nucleus; to be sure that it is a maternally inherited mitochondrial disease, also use the search phrase "maternal inheritance" in quotes—http://www.ncbi.nlm.nih.gov/omim

Lecture 14:
Conclusions

The Suggested Reading for this lecture can be any current textbook on genetics, for example, Daniel Hartl and Elizabeth Jones's *Genetics: Analysis of Genes and Genomes*.

You have just listened to a course in basic genetics. Here are the major topics we've covered, phrased with confidence about your achievements. You have a repertoire of examples and metaphors by which you understand basic genetics principles better and you can even explain some aspects to others. You are also ready for further reading on topics of special interest to you.

- You can recognize some Mendelian ratios and "simple" genetic traits. You also can explain why there are no "simple" genetic traits and why that is just a convenient way to talk about some traits.
- You understand some phenotypes as spectrums of possibilities in multigene pathways, where many of the genes may be represented as allele series. You realize that most traits (perhaps all traits?) are parts of complex pathways.
- You try always to think of genes in context of "environment." You have a broad definition of environment that includes other genes, the cellular (or tissue or organismal) environment, and the environmental parameters outside of the organism (including other organisms).
- You have a growing appreciation for the importance of DNA mutations (DNA variability) of all sorts in evolution. Those mutations (of which there are many different types and degrees) are an essential part of the simplified events of natural selection.
- You have some interesting examples of chromosome activities and disorders and an idea of how some of those disorders occur.
- You can recognize or at least make good educated guesses about regulatory mutations. You are aware of the importance of gene regulation in the context of pathways. This is an advanced topic, but you have some accessible examples in your repertoire.
- You can recognize or make good guesses about some patterns of non-Mendelian inheritance, such as of mitochondria, and can explain how those occur.
- You can recognize the possible influences of viruses and/or transposons on phenotypes. (And you know that those phenotypes are mostly non-pathogenic.) And you are as puzzled (and eager to learn more) as any professional geneticist is about the enormous load of viruses and transposons we carry.
- You have a better idea about that most difficult concept in any science, distinguishing between what we know, what we *think* we know (but could be mistaken), what we don't know, and what we don't know yet that we don't know.
- And you might agree with me that if you could invite guests from the past to a dinner party, you'd like to invite Gregor Mendel and get into a wonderful discussion with him about how far we have come with genetics.

FOR GREATER UNDERSTANDING

Questions

1. What are the most prevalent misconceptions about genetics?
2. How has your own perception of genetics changed over the course of these lectures?

Suggested Reading

Hartl, Daniel, and Elizabeth Jones. *Genetics: Analysis of Genes and Genomes*. 7th ed. Sudbury, MA: Jones & Bartlett Publishers, 2008.

Other Books of Interest

Wells, Spencer. *The Journey of Man: A Genetic Odyssey*. New York: Random House, 2004.

GLOSSARY

Adenine: One of the building blocks of DNA or RNA, typically called "A."

Albino: Having little or no pigment (typically melanin pigment).

Allele: A variant form of a gene. Genes have many alleles. New alleles occur through mutation. Any given animal or plant usually has just two alleles at a time for each gene. If both alleles are the same, that is a homozygous condition. If they are different, that is heterozygous.

Asexual: Reproduction by division and without the extra steps of making gametes (sperm and eggs or pollen and ova) and getting them fertilized.

Bacteria: Microscopic single-celled organisms with simple cell structure that comprise the vast majority of organisms on Earth.

Birth defects: Deformities and malfunctions (ranging from major to minor) in organs, limbs, or systems that occur during development. Many genetic birth defects often are not well enough understood to have any particular gene associated with them. However, for those birth defects that have been well analyzed, the associated genes typically are found to involve the regulation of many other genes at some critical point in the development of an embryo.

***Caenorhabditis*:** A model organism for the understanding of cells, genes, and development. It is a tiny (one millimeter long) nematode worm with just about one thousand cells, easily viewed because its body is nearly transparent.

Cascade: A metaphor (of a waterfall) used by biologists to describe most pathways of chemical reactions that occur in organisms. Biological pathways are not linear, although textbooks sometimes depict them linearly. Rather, they branch in complex patterns, often resulting in many different products. (See *cookie factory*)

Chromosome: "Colored body," a structure comprising a strand of DNA tightly coiled with various proteins, often seen under the microscope during cell division when DNA must be sorted and moved into each of two cells.

Chromosome mutations: A change in the structure or number of chromosomes, often due to a mistake during cell division (mitosis or meiosis).

Clones: Genetically identical (or nearly identical) organisms; For asexually reproducing organisms (the vast majority of organisms) this is easy and routine. Just divide a cell and all of its genes into two cells. For obligately sexual organisms like mammals, in which cloning never occurs naturally, there have been some laboratory methods developed that approximate the cloning typical of asexual organisms.

Complement: In genetics, this often refers to the relationships or base pairings between the building blocks of DNA and RNA. For example, in DNA "C" complements (or binds, or base pairs with) "G," and "A" complements "T." In RNA, "U" is used instead of "T" to bind "A."

Cookie factory: A metaphor used in this course to describe pathways of chemical reactions controlled by genes. It may be more useful than the metaphor of a "cascade" (see *cascade*). In addition, it is best to think of the cookie factory as being convoluted and full of extra elaborations such as what Rube Goldberg might design.

Crossover: Also called recombination; a natural tendency of DNA strands to exchange pieces with each other. This is especially enhanced during some cell processes such as during first steps of meiosis. It is easy to get crossovers to occur in lab and this is a fundamental technique of "genetic engineering."

Cytosine: One of the building blocks of DNA or RNA, typically called "C."

DNA: An elongated molecule comprising building blocks (or bases) As, Gs, Cs, and Ts, forming two linear chains intertwined and bound together to form a double helix. This

is the primary information molecule of all organisms. The sequence of As, Cs, Gs, and Ts encodes the information.

Deletion: Used in two ways by geneticists: it can refer to a mutation to a strand of DNA in which a base (A or C, or G or T) is accidentally left out. It can also refer to a major change in a chromosome in which a piece is missing. That piece could represent thousands of bases.

Determinism (predetermination): Describes a misunderstanding about genetics that has occasionally caused great controversy. It is the idea that having a particular gene or set of genes essentially guarantees an organism to have a particular trait regardless of any of the usual complexities typical of genetic systems, such as the environment.

Dominant: A relationship between alleles (variants of genes) such that in any pair of alleles, if one of them is able to manifest its particular trait regardless of whether it is present in one copy or two, then it is dominant. Sometimes there is no clear-cut dominance of one allele over another. Both alleles of a pair get to manifest their traits. They are *co-dominant*. Sometimes it matters whether an allele is present in one or two copies. If two copies, the trait is manifested more strongly than in one. This is called either *partial dominance* or a *dosage effect*.

***Drosophila*:** A model organism—a small fly, easily cultured and displaying many wonderful, highly visible, and fascinating mutant traits. Much of what we understand about the workings of fly genetics informs genetics in general for all animals, including mammals.

Egg: A gamete of female organisms containing one set of chromosomes. The female has two sets of chromosomes and must undergo a specialized cell division, meiosis, in order to produce a gamete. In plants this is an ovum.

Environment: All of the parameters both biological and physical ranging from the immediate environment within a cell or within an organism or within a uterus to the greater "outdoor" environment. It also includes interactions of all kinds between organisms.

Epistasis: Used by geneticists to refer to activities of a gene (or allele of a gene) that produces a product used early in a sequence of chemical reactions. The activities of that early gene product can have profound effects on all subsequent chemical reactions, including stopping them from occurring. In the cookie factory, epistasis occurs early on the assembly line (such as with a malfunctioning dough mixer) and the result is that no subsequent machine on the assembly line can get its job done properly.

Eugenics: An appropriation of some aspects of genetics—including some simplifications and misunderstandings about genetics—to develop political and social policies. The goals typically are to make improvements in a population by allowing some to reproduce and others not. The means by which this is accomplished can range widely from voluntary birth control and abstinence to forced emigrations and genocides. Breeders of pedigree animals practice a version of eugenics, although it typically is not given that name.

Eumelanin: Dark melanin pigment that comprises the coat, skin, and eye colors of many animals as well as parts of some fungi.

Exons: Sections of a gene that contain sequences that eventually may be translated to protein. But first, all the intervening sequences (introns) must be clipped out and the exons connected up (spliced) into one continuous string.

Expressivity: Geneticists use this word when they realize that a particular gene (or allele) is not operating in any simple manner, easy to interpret. Probably much more work will need to be done to figure out where and how this allele is working in a pathway with other genes. Meanwhile, to describe the range of possible traits for

the allele, "variable expressivity" is used. An example is white patterns in dog coat color, produced by a complex system only partly understood. Some of the alleles involved yield various numbers and arrangements of white patches and spots for no easily interpretable reason.

Frame shift: When a base (A, C, G, or T) is either inserted into or deleted from a DNA sequence for a gene, it can change the reading frame and render the rest of the string unreadable. Genes are read in "words" or codons of three. In English (using just three-letter words), a frameshift of THE-CAT-SAT might result in an unreadable sequence: HEC-ATS-AT due to the deletion of the first "T."

Fungi: A large and diverse group of mostly microscopic organisms, including molds, yeasts, and mushrooms.

Gamete: A generic term for specialized reproductive cells with one set of chromosomes—sperm, egg, pollen, ova.

Gene: A sequence of DNA that has (or recently had, or at least theoretically has) the potential to be transcribed and translated into protein.

Gene therapy: An idea still mostly in an experimental phase that some genetic diseases might be cured by inserting (by genetic engineering) a normal functional copy of a gene to replace the function of one that is causing a disease. The resulting patient would in essence be "genetically modified."

Genetic code: The set of 64 triplet combinations of A, C, G, and T (such as ACT, GGG, GAT) and the corresponding amino acid (protein building block) or "stop" (to indicate end of a gene) for each triplet.

Genetic counselor: A medical professional trained to interpret genetic tests and to provide information and advice to patients with concerns about particular genetic conditions.

Genetic cross: Taking two sexual organisms and putting them together such that their gametes fuse and produce an offspring (or potential offspring). Asexual organisms (that is, most organisms) don't need to be crossed.

Genetic disease: A condition serious enough to be called a "disease" and not merely a trait that is a result of having a particular gene or set of genes operating differently than what is considered in the range of "normal."

Genetic engineering: Laboratory procedures taking advantage of the natural tendency of DNA to easily recombine (crossover) and to be easily inserted into other organisms. In the natural world, DNA is constantly recombining and being taken up by other organisms. In the lab, these functions are facilitated and directed for particular purposes, such as developing gene therapies.

Genetic testing: Examining either gene sequences themselves or the protein products of gene sequences to determine whether a gene is functioning within some range considered normal. This can also be done on a larger scale by examining visually entire sets of chromosomes (also called karyotyping) to see if there are any changes.

Genetically modified: Characterizes organisms that have been genetically engineered to contain foreign genes, usually for some purpose, such as to confer resistance to disease.

Genome: All the DNA contained within an organism.

Genotype: The genes (often represented by symbols) of an individual that contribute to a particular trait or phenotype.

Guanine: One of the building blocks of DNA or RNA, typically called "G."

Heirloom: A colloquial name for a strain of domesticated plant or animal implying that it is true breeding (or at least true breeding for a certain set of traits) and has been for many generations, having been selected and propagated by humans long ago.

Hybrid: An organism resulting from a cross of two organisms that are true breeding (or at least true breeding for certain traits). The hybrid offspring may show traits of one or another parent and, in some cases, traits that result from combinations and interaction between the two parental sets of genes. The hybrid offspring will not be true breeding, which is why seeds saved from hybrid plants will not yield reliably the characteristics of the parents.

Inbreeding: Matings between related organisms.

Insertion: A mutation of DNA by which one or more bases (As, Cs, Gs, and Ts) are added into a DNA strand.

Intron: An intervening sequence of DNA within a gene that is like a spacer between sections of DNA that code for proteins. The coding sequences are exons. The introns eventually are spliced out.

Jumping gene: See *transposon*.

Karyotype: A methodical examination of the chromosomes of an individual to determine the number, shapes, and patterns of the chromosomes.

Maternal inheritance: The passing along of genes (including viral genes and mitochondrial genes) that reside mostly in the cytoplasm of a cell. Egg cells of many sexually reproducing organisms often are larger than sperm. Therefore, eggs easily pass along their various cytoplasmic genes. Thus, those genes are maternally inherited; they come from the mother.

Meiosis: The process of moving chromosomes during a specialized cell division such that each newly divided cell potentially receives half the number of chromosomes compared to the starting cell. This is the division that occurs in ovaries and testes to produce eggs and sperms. In plants, the gametes are pollen and ova.

Melanin: A dark pigment, typically of animals and fungi. In animals, it is found in the coat (or hair), skin, and eyes.

Melanistic: Characterizes an organism typically found in the wild with a pale skin or coat or eye color, but in this case has a mutation that confers darker pigmentation (albinism is the opposite situation).

Mendelian: (After Gregor Mendel) a simplified genetic investigation in which the researcher sets up experiments and/or gets results that elucidate the activities of one or two or just a few genes.

Merle: A coat color of mammals that occurs when melanocytes (melanin-filled cells) fail to package melanin efficiently and therefore produce patches of pale skin, coat, and eye color that should have been darker with melanin.

Mitochondria: Membrane-bound structures within complex cells in which occur chemical reactions relevant to transferring and storing energy. Mitochondria originated as free-living bacteria and still retain vestiges of their bacterial genomes.

Mitosis: The process of moving chromosomes during cell division such that each newly divided cell potentially receives a complete set of chromosomes identical to that of the starting cell.

Modifier: Refers to a gene that produces a product that has some moderate, often subtle but enhancing or reducing effect on the product of some other gene.

Mosaic: Refers to genes that are expressed in a sort of patchwork pattern in and on a female mammal. In particular the genes are on her two X chromosomes. During

development one or another X chromosome is shut off in each cell of the female's body. Therefore, in some areas one allele of a gene may be expressed and in other areas a different allele.

Mutation: A change in sequence to a DNA molecule or a change on a larger scale to the structure of or number of a chromosome.

Nature versus nurture: A false dichotomy suggesting that the characteristics of an organism might be either due to the genes of an organism or to the environment of an organism, but not to any combination or interaction of the two. There is no need to choose one or another. Genes are expressed within the context of environments. Environments influence the expression of genes.

Neoteny: The retention of juvenile characteristics in the development of an organism. This is often the characteristic (albeit a complex one) sought by animal and plant breeders whether they realize it or not.

Nucleus: The membrane-bound compartment of complex cells in which most of the DNA for the organism is stored. However, other DNA such as for some viruses and in mitochondria is outside of the nucleus in the cytoplasm.

Outcross: A cross of two organisms that are true breeding (or at least true breeding for certain traits), resulting in a hybrid organism, displaying traits from a combination of its parents' genes.

Pathogenicity: A relationship between two organisms, a pathogen and its host or between a virus and its host, marked by a lack of coordination of reproductive cycles. This may result in the pathogen growing more quickly than the host and affecting the host by using up resources and producing wastes. Pathogenicity is often connected to a newness of the relationship and may also include certain functions that allow the pathogen to avoid detection and defensive responses by the host.

Pathway: As used by biologists, a set of sequential chemical reactions, albeit often highly redundant, convoluted, and baroque, by which particular functions are accomplished by a cell.

Penetrance: A word used, often accompanied by a percent (as in 25 percent), to refer to the chance of a particular organism with a genotype actually displaying the phenotype for those genes. Why aren't all organisms (100 percent) displaying a phenotype that corresponds to a particular genotype? Usually, the answer to that is unknown; there is likely to be a large, complex pathway with functions hidden (to us) that results in individuals responding differently to having the same genes.

Phaeomelanin: A lighter version of the mammalian (and fungal) pigment melanin.

Phenotype: The description of a particular trait associated with a genotype.

Pleiotropy: Characterizes a gene that functions early in a large, multibranched pathway such that its function may result in several different traits at the end of the pathway.

Population: In genetics, it is a group of organisms of the same species, typically in a geographic area such that they seem to be more or less related to each other. If sexual, it is expected that a population would be interbreeding.

Predisposition: A tendency for an individual with a particular gene or set of genes to manifest a particular phenotype as long as environmental conditions and other gene products (perhaps with unknown functions) are manifested as well.

Protein: A linear molecule comprising a string of amino acids (of which there are twenty commonly used ones); think of a string of beads with the beads being of twenty different colors. And then consider that proteins never remain in a linear format but rather become folded and twisted in particular ways to make distinctive shapes. Proteins do most of the work of facilitating chemical reactions to occur in organisms.

They also comprise most of the structures of organisms. Gene sequences code for the sequences of amino acids in proteins.

Punnett square: Reginald Punnett had the idea of arranging the genes involved in a genetic cross on a table (called a Punnett square) so that the various combinations could be readily seen.

Recessive: A relationship of a pair of alleles such that a particular allele is expressed (produces a phenotype or trait) only if it is present in two copies, that is, with no other competing allele.

Recombination (also *crossover*): The common tendency of DNA strands to exchange pieces with each other, especially when the sequences have some similarities (are homologous).

Regulation: In genetics, it is the organization and coordination of gene use, such that complex, multipart, time-sensitive traits result. For example, during the development of the eyes, hundreds of genes must be turned on and expressed in order to make such a complex structure.

Replication: Making a complementary copy of a DNA strand. Since DNA is double-stranded, each of the two strands gets a complement and thus one strand is replicated to two.

RNA: A single-stranded linear molecule comprising four bases A, C, G, and U, and with many diverse functions. These include messenger RNA, which delivers sequence information (on a metaphoric piece of scrap paper) from a gene to be translated to protein. RNAs also include ribosomal RNA and transfer RNA, both of which are part of the translational function.

Selection (artificial, natural): If natural selection, it occurs when a set of environmental conditions allows organisms with a particular set of traits to reproduce more efficiently and abundantly than organisms with a different set of traits. If artificial, the process is controlled by humans who allow organisms with one set of traits to reproduce and prevent reproduction in others. Thus, for example, dog breeds quickly evolve, especially if the human breeder is determined to make that happen. Darwin was greatly influenced by the practices of artificial selection when he develop the concept of natural selection.

Selfing: A mating easily done with many plants such that the male (pollen producing) parts of a plant fertilize the female (ova producing) parts of the same plant. Selfing requires that a plant have both male and female parts, which many do. It also requires that the plant not have mechanisms to prevent self-fertilizations.

Sexual: A type of reproduction requiring the production of gametes (such as sperm and eggs) with half the number of chromosomes of the parents. The gametes then fuse and produce offspring.

Species: A surprisingly difficult concept to define: if an organism reproduces sexually, then others of its species are those with which it presumably could mate and produce offspring that in turn were fertile. However, few species have been put to the actual test and many instances are known in which organisms assumed to be of different species mate. Furthermore, many organisms have both sexual and asexual reproduction or are exclusively asexual. In those cases, defining a species is more about observing enough similar characteristics (at least from a human point of view).

Spectrum: Used in genetics to refer to a range of traits (or phenotypes) resulting from a range of variations (alleles) for a gene or set of genes. In the cookie factory metaphor, we get a spectrum of cookies with various icings and decorations resulting from a spectrum of subtle differences in the frosting and decorating machines.

Sperm: A gamete of male organisms containing one set of chromosomes. The male has two sets of chromosomes and must undergo a specialized cell division, meiosis, in order to produce a gamete. In plants, the corresponding cell is pollen.

Splicing: An activity that occurs with a newly transcribed messenger RNA, which contains both exons (sequences that code for parts of proteins) and introns (sequences that interrupt the exons). Splicing occurs to cut out the introns and join together the exons into one continuous coding sequence.

Substitution: A mutation of change to a sequence of DNA such that one base is replaced by another.

Thymine: One of the building blocks of DNA , typically called "T."

Transcription: The process of making a messenger RNA complementary to a sequence of DNA that is a gene. Indeed, genes are defined in part by their being able to be transcribed.

Translation: The process following the transcription of a gene, forming a messenger RNA. The information in the message is converted to a string of amino acids, forming a protein.

Transposon: Also called a "jumping gene." A sequence of DNA capable of moving around in the chromosome and sometimes making extra copies of itself. It also includes sequences that appear to have once had the capability but no longer do.

True breeding: Organisms with a set of genes for a set of traits of interest in which those genes are all in the homozygous condition. That is, each gene is represented by a pair of identical alleles. This means that during reproduction, when sorting out of various alleles occurs, there isn't any sorting out to do. The true breeding parents and all of their offspring have the same allele pairs.

Twins: In sexual organisms, identical twins result from the splitting of a fertilized egg and presumably have identical or nearly identical sets of genes. Fraternal twins are from two different eggs, each with its own fertilization by different sperm. Asexually reproducing organisms produce identical twins all the time, by simple cell division or by various plant propagation mechanisms. However, the word "twin" is not typically used. Instead the word "clone" is more typical.

Uracil: One of the building blocks of RNA, typically called "U" and capable of pairing with "A."

X chromosome: A chromosome of mammals (and some other animals), comprising a large set of diverse genes, none of which are particularly associated with sex development. Females have two X chromosomes and males have one. The other chromosome in males is the Y chromosome.

X inactivation: Male and female mammals have different numbers of X chromosomes. A mechanism to compensate for that (to even up the numbers) is X inactivation. It occurs early in female development. In each cell one or another X is shut off or inactivated. Therefore, effectively, females have just one functional X per cell.

Y chromosome: A small chromosome of mammals (and some other animals) bearing just a few genes, some of which are essential to producing male primary and secondary sex characteristics. Males have one (or in rare cases two) Y chromosomes. Females have none.

Zygote: An egg newly fertilized with a sperm.

Isolating DNA from Strawberries[†]

1. Cut up one strawberry and place it in a zip-lock bag.
2. Add about 50 ml (1/4 cup) of chilled extraction buffer (ingredients below).

   Extraction Buffer (keep in refrigerator—makes enough for three extractions)

   1/8 cup (33 ml) of simple shampoo (without conditioner) OR dishwashing liquid

   2/3 tsp (5 grams) salt

   1-1/8 cup (300 ml) water

   The shampoo contains a detergent "sodium lauryl sulfate" that breaks apart cell membranes, releasing the DNA and other cell contents. The salt helps to make a solution that is similar in saltiness to the inside of the cell.

3. Seal the bag and squish the strawberry and extraction buffer, being careful not to squirt the mixture from the bag.
4. Pour the strawberry mixture through cheesecloth to remove the pulp.
5. Pour the red liquid into a tall, narrow vessel.
6. Slowly pour ice-cold ethanol ("grain alcohol") down the side of the vessel to form a layer on top of the red liquid. (Keep the ethanol in the freezer until you are ready to use it.) If grain alcohol is not available, try using the strongest proof vodka available.
7. Notice the filmy white DNA precipitating where the ethanol and red liquid are interfaced.
8. Use a glass rod—a toothpick or skewer will also work—to gently spin some DNA from the interface.
9. Note the thready, mucousy nature of the DNA.
10. When done, rinse all solutions down the sink and wash glassware.

[†]Strawberry procedure described on www.carnegieinstitution.org—Diane Sweeney, Pearson Education

Isolating DNA from Other Materials

Strawberries work well for isolating DNA because they can be macerated by hand in a plastic bag. Bananas and kiwi fruit work too. For tougher fruits and vegetables or fresh (or frozen) meat, you need to chop finely and then run through a blender. If you do choose to try animal tissue (meat), the best choice is organ meat, like liver or kidney or "sweetbread" (thymus), frozen, or as fresh as possible.

Isolate Your Own DNA[‡]

1. Start with 1 tsp of salt dissolved in a glass of water PLUS another glass in which you have placed 1 tsp of dishwashing liquid and 3 tsp water. ALSO have ice-cold alcohol greater than 100 proof. (This could be vodka, gin, whiskey, or rubbing alcohol.)
2. Wash your mouth vigorously for 30 seconds with the salty water and then spit into the glass with the diluted dishwashing liquid.
3. Stir this firmly for several minutes.
4. Then gently layer on the ice-cold ethanol down the side of the glass.
5. Look for spindly white threads forming in the alcohol.

[‡]This method is from Mike O'Hare's *How to Fossilize Your Hamster: And Other Amazing Experiments for the Armchair Scientist* (New York: Henry Holt & Company, 2008).

Suggested Readings:

Clutton-Brock, Juliet. *A Natural History of Domesticated Mammals*. 2nd ed. Cambridge: Cambridge University Press, 1999.

Hartl, Daniel, and Elizabeth Jones. *Genetics: Analysis of Genes and Genomes*. 7th ed. Sudbury, MA: Jones & Bartlett Publishers, 2008.

"Microbe Cards." The American Society of Microbiology. http://www.asm.org/

Miller, Kenneth R., and Joseph S. Levine. *Biology*. Upper Saddle River, NJ: Prentice Hall, 2007.

Rutter, Michael. *Genes and Behavior: Nature-Nurture Interplay Explained*. Oxford: Wiley-Blackwell, 2006.

Schwartz, James. *In Pursuit of the Gene: From Darwin to DNA*. Cambridge, MA: Harvard University Press, 2008.

Other Books of Interest:

Andrews, Chris. *Guide to Fancy Goldfish*. Surrey, UK: Interpet Publishing, 2002.

Brown, Jack, and Peter Caligari. *An Introduction to Plant Breeding*. Oxford: Wiley-Blackwell, 2008.

Coile, D. Caroline. *Encyclopedia of Dog Breeds*. 2nd ed. Hauppauge, NY: Barron's, 1999.

Comfort, Nathaniel C. *The Tangled Field: Barbara McClintock's Search for the Patterns of Genetic Control*. Cambridge, MA: Harvard University Press, 2003.

Crick, Francis. *What Mad Pursuit: A Personal View of Scientific Discovery*. New York: Basic Books, 1988.

Goldberg, Rube. *Rube Goldberg: Inventions!* New York: Simon & Schuster, 2000.

Grandin, Temple, and Catherine Johnson. *Animals in Translation: Using the Mysteries of Autism to Decode Animal Behavior*. New York: Harvest Books, 2006.

Green-Armytage, Stephen. *Extraordinary Pigeons*. New York: Harry N. Abrams, Inc., 2003.

Helgren, J. Anne. *Encyclopedia of Cat Breeds*. 2nd ed. Hauppauge, NY: Barron's, 1997.

Judson, Horace F. *The Eighth Day of Creation: Makers of the Revolution in Biology*. New York: Touchstone, 1979.

Maddox, Brenda. *Rosalind Franklin: The Dark Lady of DNA*. New York: Harper Perennial, 2003.

Mawer, Simon. *Gregor Mendel: Planting the Seeds of Genetics*. New York: Harry N. Abrams, Inc., 2006.

Other Books of Interest (continued):

Peterson, Roger Tory, and Margaret McKenny. *A Field Guide to Wildflowers*. New York: Houghton Mifflin Harcourt, 1998.

Robinson, Heath. *Contraptions*. New York: The Overlook Press, 2007.

Sacks, Oliver. *The Island of the Colorblind*. New York: Vintage, 1998.

Venter, Craig. *A Life Decoded*. New York: Viking, 2007.

Watson, James. *The Double Helix*. New York: Signet Books, 1968.

Wells, Spencer. *The Journey of Man: A Genetic Odyssey*. New York: Random House, 2004.

Wilson, Edward O. *Sociobiology: The New Synthesis*. 25th anniversary ed. Cambridge, MA: Belknap Press/Harvard University Press, 2000.

These books are available online through www.modernscholar.com or by calling Recorded Books at 1-800-636-3399.