

## Evidence for evolution from experiments and genetics

A successful theory must be able to accommodate new data, and evolutionary theory has had to deal with a breadth of data that may well be unmatched in the history of science. Specifically, in the nearly one and a half centuries since Darwin presented his ideas, the mechanism for heredity has become clear and the information that is now available allows a host of detailed and profound tests of the basic ideas. In this class we will examine the genetic evidence and then talk about evolutionary experiments in the lab as well as observations of evolution in the field. We will start by giving an overview of how genetic information is stored in life on Earth.

### DNA

DNA (deoxyribonucleic acid) is an extremely long molecule shaped in a double helix (i.e., two spirals connected by rungs). It is the carrier of genetic information in all known non-viral organisms and some viruses. Each rung connects a “base pair”; the bases are molecules that are abbreviated A, C, G, and T, and the only possible pairings are A-T and C-G. If you pick one of the two strands, therefore, you can have a sequence like AATGTACT and the opposite strand has to be TTACATGA. Different species have different numbers of these pairs; humans, for example, have about three billion(!), which means that if you stretched out the molecule fully it would extend to about one meter(!!!). Given that every cell contains a copy of our DNA in its nucleus, you can therefore see that the molecule is coiled very tightly.

The sequence of base pairs codes for amino acids that are used in the synthesis of proteins. It has been determined that each triplet of base pairs codes for a given amino acid. However, since there are  $4 \times 4 \times 4 = 64$  possible triplets and only 20 amino acids are used in nature, there is some overlap. For example, GGT, GGC, GGA, and GGG all code for glycine, whereas TGT and TGC both code for cysteine. There are also triplets that code for the start or stop of a gene (genes can be thought of as units of inheritance). The redundancy is important to note for later, because we see that, e.g., changing GGT into GGC has *no effect whatsoever* on the organism (well, except that it might change the speed of production a bit). This means that, except for evolution and common descent, there is no reason at all to expect different organisms to have the same sequence of base pairs. We’ll get to that later.

For now, let’s step back a bit. This is a course on life in the universe, not just life on Earth. We can certainly argue that evolution as a process is likely to be universal to life. But is DNA going to arise on other planets? It is not nearly as clear. In fact, DNA in its current form is far too complex to have arisen in the very first life. It does turn out that some viruses use much shorter strands of another molecule called RNA (ribonucleic acid, which

also plays an important role in helping transcribe the instructions in DNA) as the carrier of genetic information. People have speculated that very early life could have been an “RNA world”. Note that in the very early days, without any robust DNA organisms around, even a much less effective system would have had a competitive advantage, and that’s all evolution needs to proceed.

Now let’s see how the details of proteins and DNA compare with expectations from evolution.

### **Case study: the cytochrome c protein**

We will focus on one particular protein, cytochrome c, which plays an essential role in the transport of oxygen. As a result, it is found in plants, animals, and many single-celled organisms. This protein typically consists of a sequence of 104 amino acids, with the exact number varying somewhat between species. Cytochrome c is an example of a “ubiquitous protein”, which means that it is basically found everywhere. To draw from Douglas Theobald’s webpage <http://www.talkorigins.org/faqs/comdesc/>, the importance of ubiquitous proteins in a test of evolutionary principles is:

- Ubiquitous proteins perform the same basic function in all organisms; the needed function does not depend on whether the organism is a plant or animal, for example.
- For a given function, there are an astronomically large number of different amino acid sequences that will do the same job. In the case of cytochrome c, we can see differences in the sequences between different organisms, but substitutions lead to no problems. As an example, yeast cytochrome c differs from the human version in about half of the protein, but human cytochrome c can be substituted into yeast and the yeast functions perfectly. It has been estimated that there are a minimum of more than  $10^{93}$  possible functional variants of cytochrome c, which is vastly more than all the particles in the observable universe, not to mention the number of organisms that have ever lived on Earth!
- As a result, for the purposes of functionality, there is no reason at all for a given sequence to exist in a given organism, and no reason that even closely related organisms *must* have the same sequence.
- However, heredity *demand*s that there be such relations. We said earlier that there could be more than  $10^{93}$  functional variants of cytochrome c. However, the total number of possible 104 amino acid sequences is  $20^{104} \approx 2 \times 10^{135}$ , a vastly larger number. Since the role of cytochrome c is so critical, this means that most variations will kill the organism. In an evolutionary model, one therefore expects that transcription of this sequence will be really, really robust, so that closely related organisms will have

amino acid sequences in cytochrome c that are close to each other. However, more distantly related organisms would be expected to have ever-decreasing similarity. The key question is then: is the tree of relations of life derived in this way consistent with what has been derived independently by the fossil record?

The answer is a resounding yes! Humans and chimpanzees have exactly the same cytochrome sequence. Their sequences differ by at most 10 amino acids from any other mammal. In contrast, a particular species of yeast differs in 51 amino acids. We also note that the sequence for bats is much more similar to that of other mammals than it is to birds (who might be thought to have closer to the same specific requirements), and similarly the sequence for dolphins is much closer to ours than to sharks. Similar correspondence exists for other proteins. This is a great success for the theory.

### **Similarity of DNA sequences**

There is another prediction that can be applied, given our understanding of how DNA codes for amino acids. Remember that we said there are 64 possible triplets, but only 20 amino acids? This means that for a given amino acid there are on average about 3 triplets that code for it. For cytochrome c, with 104 amino acids, this means that there are about  $3^{104} \approx 4 \times 10^{49}$  base pair sequences that give *exactly* the same protein, with zero functional change. To drive home that point, from the functional standpoint there is nothing at all to demand any given sequence out of the possible  $4 \times 10^{49}$ .

Heredity, however, implies similarity between the sequences, and in this case we can be specific. The background mutation rate per base pair per generation in mammals has been measured to be  $(1 - 5) \times 10^{-8}$ . An average primate generation is about 20 years, and the fossil record indicates that we diverged from chimpanzees less than 10 million years ago, so around 500,000 generations. We therefore expect a less than 3% difference in the sequence of base pairs in the gene that codes for cytochrome c. The actual difference is 4 base pairs out of the  $3 \times 10^4 = 312$ , or a 1.3% difference. Overall, in fact, human and chimpanzee DNA differ by only about 2%, which is expected if there is common descent but is not even remotely required by functionality.

### **Using evolutionary principles: the AIDS cocktail**

Our understanding of evolution is not merely academic. It has played an extremely important role in our treatment of disease. Bacteria and viruses have very short generational times, meaning that they can evolve rapidly in response to external pressures. One of those pressures involves the medicines and treatments that we apply to get rid of them. If we apply a treatment that does not get rid of all the invaders, and that they can adjust to without multiple simultaneous (and thus highly improbable) mutations, then we are just

helping them get stronger. Overuse of antibiotics has had this effect, leading to multiple drug-resistant strains of many maladies we thought we'd conquered (e.g., tuberculosis).

A happier story, at least at the present, has to do with treatment of patients with HIV (human immunodeficiency virus, the cause of AIDS). There is still no cure for AIDS, and millions of people die from it every year. There is, nonetheless, a ray of hope for those who can afford somewhat expensive therapy. If you look at a plot of AIDS deaths in the United States, you see that it rose rapidly to a peak of more than 40,000 in 1995. It then dropped quickly to 15,000-19,000 per year; still a horrifying number, but far less than before. What happened?

From studies of the virus it became clear that its spectacular replication rate in human hosts meant that it is poised to adapt quickly to medicines. The key, though, is that there is a limit to what evolution can do. If a simple flip of a gene is needed, sure, no problem. However, if multiple mutations are required *all at once*, and there is no selective advantage for any of the mutations individually, then it is highly improbable that the mutations will happen as needed. Without understanding of evolution, there would be no particular reason to think that a virus couldn't make the large leap.

With this background, scientists developed a three-drug regimen that is applied all at once. These hold back critical enzymes at the beginning and the end of the HIV replication, and so far HIV appears not to have breached the defenses. We are still a long way from a vaccine, and there is still no substitute for safety, but this is a clear case in which evolutionary knowledge has saved tens of thousands of lives per year.

### **Evolution in the lab: the Lenski experiments**

An elegant experiment in evolution has been carried out by Richard Lenski and colleagues at Michigan State University and elsewhere. For more details on this I strongly recommend that you go to his home page: <http://www.msu.edu/~lenski/>, and also to an article describing his experiments and others: <http://www.msu.edu/~lenski/sciencearticle.html>

The basic idea is simple. Bacteria reproduce rapidly, so they are ideal for evolutionary experiments. They can also be cloned easily, so in 1989 Lenski set up 12 flasks that contained cloned members of an original *E. coli* bacterium (this is a bacterium we have in our intestines). Into each flask he put a small amount of glucose (a type of sugar). Since then, daily, he has let the bacteria run through the sugar, then the next morning he would transfer some of the survivors in each flask to a new flask. By 2007 he had run through more than 40,000 generations of bacteria (i.e., several per day). This is the equivalent of nearly a million years of human evolution!

The point, of course, is to determine how mutation and other variational mechanisms

would change the properties of the bacteria. The answer: a lot! He has frozen samples every now and then, so that direct comparison is possible. The current bacteria average twice the size of their ancestors, and are 70% more efficient at metabolizing the glucose. Most of the improvement occurred early on, but the changes are not identical in each of the flasks. In fact, now that gene sequencing is fast and efficient, the scientists can follow the precise changes that occur.

Since the original experiments, others have tried different environments. Julian Adams at the University of Michigan also started with genetically identical *E. coli*, and grew them in a device called a chemostat that made sure that all the lines had identical conditions. However, commonly, he found that in a given flask more than one strain of *E. coli* would develop. He found that initially the *E. coli* evolved to metabolize glucose more efficiently. After a while, however, the amount of glucose was too much to metabolize, so the *E. coli* added another pathway that generated acetate as a waste product. Enough acetate was produced that a mutant emerged that used the acetate as fuel. Pretty impressive adaptation!

Even more recently, Lenski and collaborators reported that a variant had evolved around the 31,000th generation that was able to use citrate. This is not possible in wild-type *E. coli*, but investigations by the team showed that there had been at least three separate mutations required; the first two were non-adaptive (i.e., neutral drift), but the last one combined with the first two to allow a major biochemical change. Such “potentiating” mutations can be seen in the digital life simulation Avida (available at <http://avida.devosoft.org/>), in which one sees many cases where neutral or even somewhat harmful mutations allow subsequent development of new capabilities.

### **Evolution of fruit flies in the laboratory**

Being human, of course, we tend to focus on things closer to ourselves. Bacteria are pretty far away; has evolution ever been seen in animals? Again, of course, we need to look at things that have short generational times.

For this purpose, fruit flies are ideal subjects. They breed like mad and are easy to care for, so they have been the workhorses of genetics laboratories since the early 1900s. Tons of changes have been seen to arise spontaneously, such as eye color, but what we’re really interested in is whether a new species has emerged at any point.

To determine this, we need a definition of species. Biologists classify living things based on a taxonomic scale, in which the broadest category is the domain (bacteria, archaea, and eukarya, which constitute plants, animals, fungi, etc.; yes, all those really are more related than we are to bacterial!), followed by kingdom, phylum, class, order, family, genus, and species. The boundaries are often a little fuzzy, but for animals and plant species can be defined fairly specifically, if you’ll pardon the pun. Individuals of a species can produce

viable offspring (i.e., offspring that can produce offspring). Individuals of different species can't. Therefore, all horses are part of the same species, but horses and donkeys aren't because although they can reproduce the result (a mule) is sterile.

With this in mind, the laboratory of Dobzhansky saw speciation that occurred some time between 1958 and 1963. The strain in question started from a single inseminated female that was captured in Columbia, and as of 1958 it produced fertile hybrids when crossed with its original strain. As of 1963, however, crosses produced only sterile males. This satisfies the definition of the emergence of a new species.

### **Observed speciation in nature**

You may wonder, though, whether new species have ever been seen to emerge in nature, as opposed to a lab (although the principles are of course identical). Yes, using our previous definition, speciation has been seen in various plants (e.g., the evening primrose, fireweed, maize, and several others); by "seen" here I mean that people have actually observed different species produced over the course of years of observation.

Over longer times, of course, the changes can mount up, so that even larger organisms with longer generational times can speciate. A good example is the greenish warbler, a bird of central Asia that is the poster child of the "ring species" phenomenon. These birds live in trees in a wide portion of Asia that fully rings the Tibetan plateau, which does not have trees. Along the ring one can do tests of various aspects of the birds, such as their coloration, songs, and genetic makeup. There are continuous changes going from the north boundary of the region (Siberia), westward, southward, and back again north. However, in Siberia itself there are two varieties that appear to be different species! They don't mate, the males ignore each other's songs, and so on. It has been suggested that the original species started in the south, then as time went on some moved gradually west and north, whereas others moved gradually east and north.

There are also a number of cases of speciation that could be happening right now. On the Hawaiian island of Kauai, for example, the crickets have to deal with a rather vile fly parasite, which lays its eggs in the cricket and the young then eat the cricket. Not pleasant. Even worse is that the fly uses the sounds the male cricket makes to hone in on them. The cricket population responded, in fewer than 20 generations, by having more than 90% of the males mutate into a silent form. You have to feel sorry for the poor crickets, because this is how they signal to potential mates, but it is a great example of evolutionary adaptation!

### **Summary**

The evidence of fossils, molecular genetics, and laboratory and field observations is unanimous that there is not a single aspect of the development of life that is inconsistent with

evolutionary theory. More specifically, every aspect of life that we know about seems completely accessible via a series of minor changes, rather than large leaps. People experiment all the time and given the pace at which science advances we can expect more tests in the next decade than in all of previous human history. For now, though, evolution as a fact is as solidly established as any scientific principle, and the basic principles of evolution are established beyond any reasonable doubt (although many details are being explored, of course). However, for decades in the post-Darwinian period we did not have the molecular evidence discussed in this lecture. What data existed before the current epoch, and how convincing was it?