LECTURE 10 of Biological Physics (PHY3040)
Evolution

Richard Sear

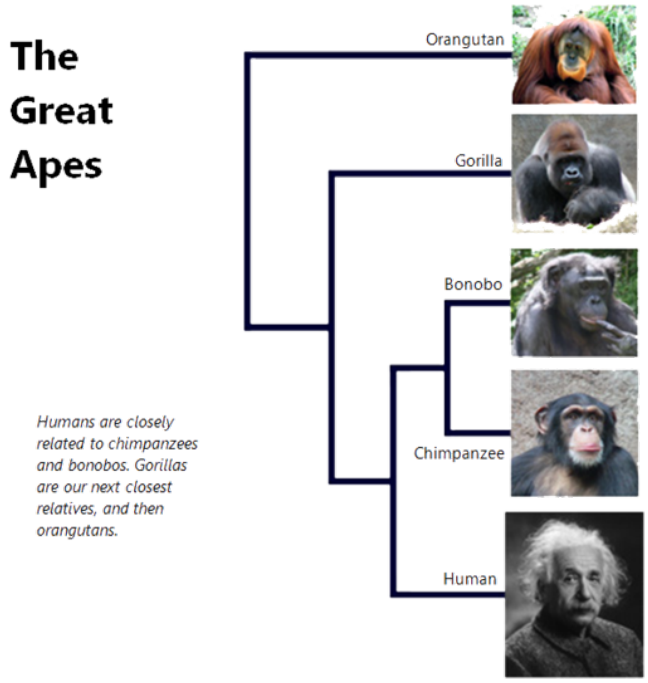

FIG. 1. Our close evolutionary family tree. We are most close related to chimpanzees and bonobo chimpanzees, the last ancestor common to chimpanzees and us lived around 6 to 8 million years ago. For both chimpanzees and us, our next-nearest relative is the gorilla. Image from Wikimedia.

In this lecture we will look at evolution. The defining feature of all living organisms, including ourselves, is that we are the product of evolution.

The evidence is that life arose on Earth approximately 4 billion years ago, and that everything alive today is descended from these very early organisms. We are related to all other living organisms on Earth. Our closest living relatives are shown in Figure 1. The fossil record suggests that our ancestors and those of chimpanzees diverged around 6 to 8 million years ago, i.e., the last common ancestor of chimpanzees and us was around 6 to 8 million years ago.

ALL LIFE ON EARTH IS RELATED, BUT WE ARE MOST CLOSELY RELATED TO CHIMPANZEES AND MUCH MORE DISTANTLY RELATED TO BACTERIA. WE CAN QUANTIFY HOW CLOSELY RELATED WE ARE BY USING DIFFERENCES IN OUR DNA.

Evolution relies on mutations in DNA. In the nuclei of our cells, we have around 3 billion base pairs of DNA, and we also have around 17,000 base pairs of DNA separately in our mitochondria. Our mitochondria are small structures in our cells that play an important role in the cell's energy metabolism. Below is a small piece, 180 monomers in length, of human mitonchondrial DNA (mtDNA), compared to the equivalent parts of the mitochondrial DNA of a Neanderthal, a chimpanzee and a gorilla:

human	GTTTATGTAGCTTACCTCCTCAAAGCAATACACTGAAAATGTTTAGACGGGCTCACATCA
Neanderthal	GTTTATGTAGCTTACCTCCTCAAAGCAATACACTGAAAATGTTTAGACGGGCTCACATCA
chimp	GTTTATGTAGCTTACCCCCTCAAAGCAATACACTGAAAATGTTTCGACGGGTTTACATCA
gorilla	GTTTATGTAGCTTACCTCCCCAAAGCAATACACTGAAAATGTTTCGACGGGCTCACATCA

All 180 monomers (often called bases or base pairs) of human and the Neanderthal's DNA are identical, but there are a few differences between ours and chimpanzee and gorilla DNA. These are shown in pale grey (red). Note that there are four different DNA monomers (bases): A, T, C and G.

Table I has a matrix with the number of base pairs at which human, Neanderthal, chimpanzee and gorilla mitochondrial DNA differs. We see that our mtDNA only differs from that of Neanderthals at 168 of the 17,000 possible positions. We are very closely related to Neanderthals. Our DNA

differs at more than a thousand places to that of our more distant cousins, chimpanzees and gorillas. Our mitochondrial DNA is a bit more similar to that of chimpanzees than that of gorillas — we are slightly more closely related to chimpanzees than we are to gorillas. The fact that we have approximately ten times fewer mtDNA differences to Neanderthals suggests that our last commmon ancestor with Neanderthals was roughly 600,000 to 800,000 years ago.

	human	Nean.	chimp	gorilla
human	0	168	1305	1605
Neanderthal	168	0	1290	1597
chimpanzee	1305	1290	0	1557
gorilla	1605	1597	1557	0

TABLE I. Table of the number of the 17,000 mitochondrial bases of DNA that are different, between us, Neanderthals, chimpanzees and gorillas.

Rate of evolution in humans & chimpanzees

Our mtDNA differs in about 1,300 places from chimpanzees, and our last common ancestor was around 6 million years ago, or of order 10^7 years ago. This gives an average mutation rate

$$\begin{matrix}\text{average mutation}\\ \text{rate of}\\ \text{human mtDNA}\end{matrix} \sim \frac{1300}{10^7} \sim 10^{-4}/\text{year}$$

Per monomer of the 17,000 mtDNA monomers, this is a rate $\sim 10^{-8}$/monomer/year.

In humans the rate of reproduction is roughly 25 years per generation. At 25 years/generation, 6 million years is approximately 200,000 generations, i.e., if you go back 200,000 generations our ancestor was the same as that of modern chimpanzees. So 1300 mutations in 200,000 generations gives

$$\begin{matrix}\text{average mutation}\\ \text{rate of}\\ \text{human mtDNA}\end{matrix} \sim \frac{1300}{200,000} \sim 10^{-2}/\text{generation}$$

Note that we are not descended from chimpanzees, or chimpanzees from us, both species have a common ancestor roughly 6 million years or 200,000 generations ago.

Evolution of viruses: *much* faster than the evolution of apes

As we have seen, the genes of apes such as chimpanzees and ourselves are changing at around 10^{-8}/monomer/year. So to see any significant change takes a long time, even a 1% change takes ~ 1 million years.

But apes, especially ourselves, are some of the most slowly reproducing organisms on the planet. We take decades per generation, but some bacteria can reproduce in 20 minutes and so can evolve in as little as weeks. Earlier we saw that apes can evolve a little in hundreds of generations, which is thousands of years. For

$$\begin{array}{c}\text{bacteria under}\\ \text{ideal growth}\\ \text{conditions a}\\ \text{1000 generations}\end{array} \sim 1000\times 20 = 20,000 \text{ minutes}$$

and with 60×24 minutes in one day

$$\begin{array}{c}\text{bacteria under ideal}\\ \text{growth conditions a}\\ \text{1000 generations}\end{array} \sim 14 \text{ days}$$

Many viruses can also, in ideal conditions, reproduce very rapidly. So, bacteria and viruses evolve much more rapidly than we do.

For example, a study by Gojobori *et al.* [T. Gojobori, E.N. Moriyama and M. Kimura, *Proc. Nat. Acad. Sci.* **87**, 10015 (1990)] estimated the rate of change of the genes of both the influenza and HIV (Human Immunodeficiency Virus) at around 10^{-2}/monomer/year, i.e., about 1% per year. They are evolving a million times faster than us apes.

This has consequences. For example, the flu vaccination has to be changed every year, because every year the flu virus has evolved and so is different. Also, HIV evolves so rapidly it can evolve during its infection of a single individual.

1. *Evolution of SARS-CoV-2*

SARS-CoV-2 is the virus that causes COVID-19. An electron microscopy image of this virus is in Figure 2. It evolved to infect humans in late 2019, and is still evolving, the Alpha variant (also sometimes called the Kent or British variant) evolved in the UK in 2020, and then in late 2020 the Delta variant evolved in India and now dominates worldwide, i.e., the Delta variant has been selected by natural selection at the expense of Alpha etc variants. This gives a time to evolve a new variant and for it to spread worldwide that is a few months.

These new variants have evolved small differences in their spike proteins, that enable them to spread faster, and so outcompete less infectious earlier variants. We are watching the evolution of the spike protein very carefully as vaccines use this protein, and so if a new variant arises with a heavily mutated spike protein, vaccines may then be much less effective against it, than they are for current variants.

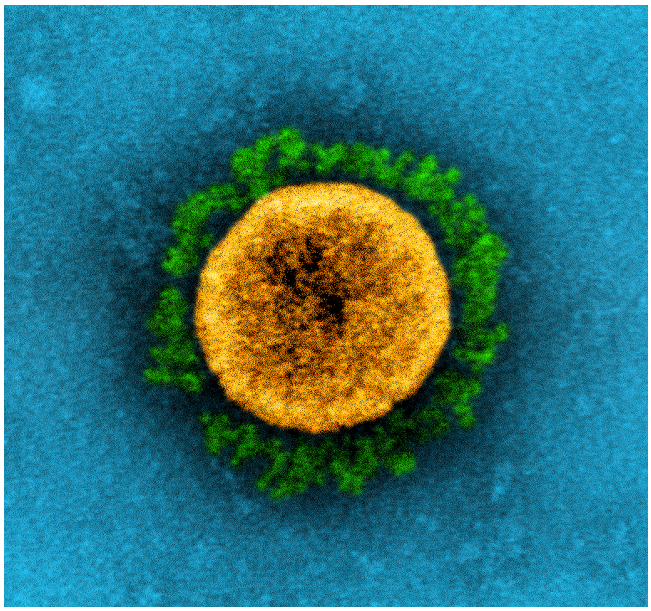

FIG. 2. False-colour transmission electron micrograph of an Alpha variant coronavirus — the variant believed to have evolved in the UK. The virus is about 100 nm across. The variant's increased transmissibility is believed to be due to changes in the structure of the spike proteins, shown here in green. Image from Wikimedia. Image obtained at the NIAID Integrated Research Facility (IRF) in Fort Detrick, Maryland, USA.

GENES AND ALLELES

We need to understand the difference between a gene and an allele. Consider the example of the genetic disease sickle cell anaemia. This is a disease of the protein hemoglobin in our red blood cells, so it is caused by a mutation in the β-globin gene.

Now we all have a copy of the β-globin gene, both people with and without sickle-cell anaemia. The difference between someone with sickle-cell anaemia and someone who does not, is that they have different versions of this gene. Versions of a gene are called alleles. So people with sickle-cell anaemia have a different β-globin allele that produces a different protein, that can damage their red blood cells, leading to the disease.

Population	% who have mutation at position 13,910
European American	77.2%
Swedish & Finnish	81.5%
French	43.1%
African American	14.0%
Yoruba (Nigeria)	0%
Makrani (Pakistan)	34.0%
Han (China)	0%
Papuan (New Guinea)	0%

TABLE II. Percentages of people sampled who have one specific mutation (at position 13,910) in their lactase gene. This mutation increases production of the protein enzyme lactase in adults. This helps the adults digest and derive energy from this sugar which is found in milk. The people with this mutation may have a common ancester around 5,000 to 10,000 years in the past. Note that human populations are not precisely defined, for example, Barack Obama could fall under both the European-American and African-American populations. Thus these categories are a little bit arbitrary. The higher % in African Americans versus African populations such as Yoruba is due to African Americans often having significant amounts of European ancestry. Data from Bersaglieri *et al.* American Journal Human Genetics **74**, 1111 (2004).

EVOLUTION IS NATURAL SELECTION FOR ADVANTAGEOUS ALLELES. EXAMPLE CASE STUDY: POSITIVE SELECTION FOR DRINKING MILK IN ADULTHOOD

There is evidence [Bersaglieri *et al.* American Journal Human Genetics **74**, 1111 (2004); Tishkoff *et al.* Nature Genetics **39**, 31 (2006)] in a few genes of evolution in the relatively recent past, over the last 10,000 years. One of these is the gene for lactase, the enzyme needed to digest the molecule lactose. We are a species of mammals, and as such human babies, like other baby mammals, drink milk. Milk contains the sugar called lactose. Lactose is abundant in milk but rare outside it. So baby mammals of all mammal species produce an enzyme called lactase, which helps them digest the lactose in their mother's milk. However, milk is not present in the diet of almost all adult mammals, and so adult mammals do not produce the protein lactase — producing this protein would be a waste of resources without lactose in the diet for it to digest.

So, mammals have evolved to produce the lactase protein as babies but not as adults. Then roughly 10,000 years ago, some of our ancestors invented farming, and the organisms they farmed included some species of mammals. Our species now farms goats, sheep, cows, etc, all of which are mammals. This advent of mammal farming meant that some populations of adult humans, unlike adults of essentially all other species of mammals, had access to milk.

Some of these populations of humans then underwent evolution in their lactase gene, so that instead of switching off production of lactase when they were still young, it continued to produce the enzyme lactase for their whole life. In Table II there are the fractions of seven different human populations, that have one specific mutation in the lactase gene. This specific mutation arose in a ancestor common to many modern day Europeans, and some in the Indian subcontinent. Other mutations independently arose in the same gene that have similar effects occurred in populations in eastern Africa [Tishkoff *et al.* Nature Genetics **39**, 31 (2006).].

It appears that in both cases, the populations

domesticated mammals and then there was selection for individuals with a gene that allowed them to obtain nutrition from the animals' milk. These mutations then spread through populations in Europe, India and Africe, but not, for example, China.

A MUTATION THAT GIVES EVEN A SMALL ADVANTAGE SPREADS RAPIDLY THROUGH A POPULATION, THE NUMBER GROWS EXPONENTIALLY

Fitness

In evolution, whether a mutation is good or bad is measured by the mutation's effect on the organism's fitness to survive and reproduce, or just fitness for short. Fitness is measured by the average number of offspring. If organisms carrying this mutation have more children, on average, than those without the mutation, then in successive generations the mutation spreads through the population, i.e., organisms with the mutation become more common. By definition, this mutation increases fitness. Whereas if organisms without the mutation have fewer children, the mutation will tend to disappear. That would be a mutation that decreases fitness.

Spread of a high fitness allele through a population, over successive generations

A mutation of a gene produces a different allele of a gene. So, for example, before the mutation we have an allele of the lactase gene that only results in the lactase protein being produced in babies. Then a mutation in the lactase gene produces a new allele, that results in lactase being produced in adulthood. This second allele will spread through the population if it leads to a higher fitness (more offspring on average) than the original allele.

Note that, we have two copies of most genes. These two copies makes the maths a bit complicated, so I will just study a very simple model in which an organism has only one copy of a gene. This is true in bacteria for example.

Fitness is measured by the number of offspring a carrier of this allele has:

$$\text{fitness} \propto \text{Number of offspring} = n$$

These offspring are per generation, and are the number of offspring that survive long enough to reproduce. So after G generations, the number of descendants is

$$\text{Number of descendants of single parent after } G \text{ generations} = n^G$$

So, as with the allele of the lactase gene, say an allele appears due to mutation that increases the number of offspring by a fraction δ, from n to $n(1+\delta)$. Then for this allele, the number of offspring is then

$$\text{Number of offspring of higher fitness allele} = n(1+\delta)$$

and after G generations, the number of descendants is

$$\text{Number of descendants with allele after } G \text{ generations} = [n(1+\delta)]^G$$

then

$$\text{Number of descendants with allele after } G \text{ generations} \simeq n^G \exp(\delta G)$$

where we used the math approximation

$$(1+\epsilon)^m \simeq \exp(\epsilon m)$$

which holds when $\epsilon \ll 1$. A increase in fitness exponentially increases the number of descendants.