Evolutionary Dynamics
Supreet Saini
Chemical Engineering
Indian Institute of Technology Bombay
Week 01
Lecture 07

Hi everybody, welcome to this lecture of the course. And up until now, what we have seen are the three fundamental principles of natural selection. And we saw two examples of how small evolutionary changes via the acquisition of just a handful of mutations can lead individuals to exhibit remarkably different phenotypes. We saw this through the example of moths, where populations shifted significantly. And then shifted back to light-colored.

And we also saw this through the example of the mice, which were camouflaged against the light-colored sand versus the dark-colored lava ash background surface that they were living upon. So, what we are going to do, of course, is discuss Darwin, Wallace, and 19th-century science, but in the 21st century, the study of evolution is increasingly molecular. So, we have to understand that the focus now of the field is understanding the molecular basis of how these changes take place, what mutations are, and what happens inside a cell that facilitates changes in the phenotypes that we have been discussing, which leads to the creation of diversity and then the principle of natural selection applying to the variation that has been so generated. So, with that focus, we are going to switch gears to the molecular side of things now.

But before we do that, we want to spend some time discussing the principles of biology for those of you who are from physical sciences and engineering backgrounds. So, this video and perhaps the next one also is going to be discussing bringing everyone on the same page with regards to understanding the molecular underpinnings of how a cell works. And what we saw was that the evolution of the mice against the two-colored backgrounds was facilitated by just the melanin protein production. So, what we are going to do is just get an understanding of how proteins are made and what is the link between the DNA sequence of an organism and the proteins that it makes. You might have heard of DNA as being called the blueprint of life and proteins as the workhorses of a cell.

DNA is called the blueprint of life because it contains all the information regarding what proteins are going to be made inside a cell. Proteins are called the workhorses of a cell because they perform all cellular functions, or virtually all cellular functions, that occur inside the cell. So, our effort in this lecture and the next one is going to be understanding how DNA contains information about making proteins in a cell. And obviously, because the DNA sequence of every organism is unique, every organism contains different information regarding what proteins to make, when to make them, and how much of each protein to make. And this is what distinguishes not just one organism from another or one species from another.

This is also what distinguishes different parts of my body, for instance. For example, the skin in my body contains the same DNA as a cell in my liver, but they look very different. And that is because They are sensing different environments. The environment that this cell on my hand senses is very different from the environment in the vicinity of a liver cell.

And hence, they make different proteins in different amounts, although the information contained in them is the same. And that leads to different manifestations, not just between species and between individuals, but also within individuals such as us. So, we'll try to understand how this process works. Evolution as a subject is steeped in the principles of logic, mathematics, and statistics. So, we don't need to do a lot of biology.

But this is one process that is central to all biology, which we have to understand. So. Essentially, we are going to understand the following process: DNA contains information. This information is read by cellular machinery and converted into RNA. DNA, of course, stands for deoxyribonucleic acid, and RNA is ribonucleic acid.

And this RNA is read by another machine to make what are called amino acid chains. And this chain then folds and acquires a functional form, which we call a protein. And this is the process that we hope to understand now, and that allows us to get into the molecular aspects of how evolution proceeds. And this process is referred to as the central dogma of molecular biology. This was a term that was coined by Francis Crick.

So we will dive straight into this. And essentially, what we want to understand is what these three species are as chemical entities. And what are these two processes that facilitate information contained in DNA being taken to a level where we have an amino acid chain? So that's our goal: understanding these three species and understanding these two processes. We'll start with DNA.

So, DNA. The DNA structure was discovered in 1953 in a classic paper by Watson and Crick. Other people, like Rosalind Franklin, were also involved in the discovery of the double-helical structure of DNA. Essentially, DNA is just a polymer of nucleotides. And that's pretty much it.

Each nucleotide consists of three parts. It has a sugar group, a phosphate group, and a nitrogenous base. A polymer of nucleotides, and this collectively is referred to as one nucleotide. The sugar is the same group of sugar in every nucleotide. The phosphate is the same.

However, the nitrogenous base comes in four forms, which are called adenine, thymine, guanine, and cytosine. We will just refer to them as A, T, G, C, and not worry too much about their structures and technical names. Although you can look them up. That is just a detail. So, this is a polymer.

So, what does one monomer look like? One monomer of this polymer looks like the following. Here is the carbon chain. OH group. There's a methyl group, and then you have the phosphate, and here you have a nitrogenous base.

So, this entire part of a nucleotide is the same. The only differentiating chemical signature between each nucleotide is which of these four nucleotide bases is attached. And if you were to number the carbon atoms on this chemical species, this is number 5, carbon number 5, and this is number 3. And because this is a polymer, each of the monomers is linked together. And the next monomer in this series is going to be linked like this.

You will have the phosphate, methyl group. A nitrogenous base. And so on and so forth. So, DNA has a directionality associated with it. And we read the DNA from the 5' end to the 3' end.

And when we are reading DNA, because these circular bits are identical in every nucleotide, we do not really look at them. We just list the order in which these nitrogenous bases make an appearance. So, for instance, if this nucleotide number 1, the nitrogenous base was G. Here it was C, and the next one was T, and then it was A. So I would list, I would read this piece of DNA as 5' G C T A 3', and what this is given to understand is this chemical detail that we have just looked at. This is one strand of DNA. You would have heard that DNA is a double helical structure and So, there is another strand of DNA which runs counter-parallel to this, which runs from 5' to 3'.

And the nucleotides that are present on this other strand are, if it is an A on the first strand, the other strand will always have a T. Similarly, if it is a T there, there is an A. If there is a C on the other strand, then this one has a G. And vice versa, if there is a G, there is a C. A and T form two hydrogen bonds between themselves, between these double helical structures, double-stranded helical structures. And G and C form triple hydrogen bonds between the two nitrogenous bases on the two different strands of this double-stranded structure. So as you can see, if I gave you the

information that there is a DNA strand whose sequence is AAAGCT3', if I gave you this information that an organism contains this sequence, then you automatically know that what this corresponds to is this chemical detail and the order of nitrogenous bases is AAAGCT. You also know that because you are only given one strand of the double-stranded structure, you automatically know that because of this complementarity rule that A is always going to pair with T and G is always going to pair with C, you automatically know that the other strand will be 5' A G C T T T 3'.

So this is obvious as soon as you are given this information. As a result, when we are talking of DNA sequences, we are almost always talking, we almost always write only this much because writing it like this automatically means that this is the structure that I am talking about, and writing it like this automatically implies that the other strand is going to be of this particular sequence. So in DNA sequence, we only give 5' to 3' and the order in which nucleotides appear. That's the chemical structure of DNA. Now, remember that our goal is to understand these three chemical entities and these two processes.

So, that's the DNA structure that we are done with. Next, we'll understand what happens in this step. So, that process which facilitates DNA information being read from DNA and the synthesis of an RNA molecule, that process is called transcription. In transcription, Let us say we have a DNA sequence of 5 prime nucleotides.

This could be any sequence. I am just writing some sequence, so on and so forth, and 3 prime. So, there is a long DNA sequence. Often, through this course, as we get into the molecular aspects and we discuss the details of DNA more and more, we'll just write this as 5 prime DNA 3 prime. Some sequence between these two ends of this fragment of DNA that we are talking about.

By the way, when we are talking like this, human DNA is of the order of length 3 into 10 to the power of 9 monomers. But that's just a detail. Bacterial DNA is often of the order of 10, 10 to the power of 6. This is humans, and this is bacteria. But genome sizes range.

Genome sizes are measured in terms of the number of monomers that you have, the number of nucleotides that a species has. And they vary through various orders of magnitude depending on the species we are looking at. So we have this double-stranded DNA, 5 prime to 3 prime. And what happens during transcription is that a molecular machine called RNA polymerase is recruited, which comes and sits on this DNA. This one is called RNA polymerase.

And as the name suggests, it's a polymerase which is responsible for the synthesis of RNA. It comes here, and the process is called melting of DNA. It basically separates the two strands from each

other. So what this machine does is it separates the two strands and breaks the hydrogen bonds that are present there. And then it reads one of the strands and basically makes a copy of the strand, which is going from the five prime to three prime end.

So it scans this piece of DNA and makes an identical copy of the five prime to three prime strand that it is reading. And at the end of this process, when the machine reaches here, it's going to look like this. So we have the two DNA strands as they are because their job was just to be read. The machine does its job and reaches the end of the DNA fragment that we are interested in. And in the process of reading, it has synthesized a piece of RNA, which is five prime to three prime.

And it is an identical copy of this 5 prime to 3 prime end with two differences. One is that instead of the nucleotide T, which is used in DNA, this molecule uses U, which is uracil. This is in RNA. And the second difference is that DNA uses deoxyribose, whereas RNA uses ribose. This makes RNA much more reactive compared to DNA, but that is just a chemical detail that we need to remember at the back of our minds.

It will become relevant in some cases, but that is just the process detail that we have. So, imagine if the DNA sequence here was GGG TACCG and so on and so forth. Then the resulting sequence on the mRNA, on the RNA molecule that is synthesized, will be GGG UACCG. So, the only difference is going to be that instead of a T, you have a U. And at the end of this process, we have what is called an mRNA. An mRNA molecule which is 5 prime to 3 prime, which is essentially just a copy of this particular strand of DNA, except for the fact that there is a U instead of T, and in the carbon base, it is not deoxy but it is ribose sugars.

This molecule is called messenger RNA. Or, in short, mRNA. And this is something that we'll refer to through the course because we'll see how messenger RNA can play an important role in facilitating evolutionary change. So, what we have looked at so far, if we go back to this one, is we are done with DNA. We've done with transcription, and this RNA molecule that we have looked at is called messenger RNA, and we know this.

In the next step now, this process we learned is called transcription, which is facilitated by a machine called RNA polymerase. Next, we are going to look at what happens to this mRNA that we have synthesized and how that information is read, resulting in an amino acid chain. So, that process of mRNA being read and an amino acid chain being synthesized is referred to as translation. And it's a key process that is responsible for the synthesis of proteins inside a cell. So, an amino acid is simply this.

You have a carbon atom. We know the valency of carbon is 4. You have a carboxylic group attached to it. You have an amino group attached to it. There is a hydrogen and there is an R group.

This is the central structure of an amino acid. And depending on the identity of R, depending on what this R is, you get different types of amino acids. Our bodies, the amino acids that we use, have 20 different types of amino acids. Hence, 20 different types of R are being used in our bodies when it comes to incorporating these amino acids into this chain that we are going to be talking about. How are these amino acid chains built from individual amino acids?

You have an N-prime end of a protein. This is where it is. And you have a C-prime end of an amino acid where this group is. And what happens is that when you have, let us say, this is R1, the first amino acid. And when these two have to be linked to each other,

R2, a dehydration reaction takes place here, which gives me NH2CHR1. And so on and so forth. And subsequent additions will keep on happening after this. And I will get a chain of this amino acid one. This is amino acid two, and so on and so forth.

So I will get a chain of amino acid one, amino acid two, and so on and so forth. And this is how this chain will form. What you should realize at this point is how similar this process is to the way we define the polymer of nucleotides, which results in the definition of DNA, where two nucleotides were brought together, and the chain led to us defining a nucleotide series that we call DNA. By the same token, a chain of amino acids brought together via this reaction leads to an amino acid chain, which then folds and is called a protein molecule. So, we will see how this process of translation works.

Remember, we are at this process where DNA was transcribed via this process of transcription, and we have an mRNA molecule which comprises A, U, G, C and is directional 5 prime to 3 prime. mRNA, unlike DNA, is single-stranded. So, what happens to this mRNA chain? We have this; let us imagine we have this mRNA chain that has resulted from the process of transcription. Now, what will happen is that another molecular machine will come and sit somewhere near the 5 prime end of this mRNA.

It has two subunits. The names of these subunits are sort of not so important for the purpose of our discussion. But these two subunits of this machine come together and sit somewhere near the 5 prime end of the mRNA. And this machine is called the ribosome. After identifying its binding site, the region of the mRNA which it recognizes and binds to near the 5 prime end is called the ribosome binding site.

site, simply abbreviated as RBS. It is also called the Shine-Dalgarno sequence, named after the people who discovered this. Once there, it starts to scan; it starts to scan the mRNA from the 5 prime end towards the 3 prime end. And when it is doing Roughly, and these are rules of thumb, biology is the study of exceptions; by rule of thumb, about 6 to 10 bases after the RBS, this will encounter a triplet which will be AUG.

Remember, there is no T in mRNA, so the sequence here will be AUG. AUG, when this ribosome is here and it is reading AUG, this is the signal for the ribosome to start the process of translation. Start the process of translation. And what that means is that AUG is a triplet. This is a triplet.

Every triplet on the mRNA encodes a particular amino acid. Remember, there are 20 different amino acids. And AUG is one such triplet that corresponds to a unique amino acid. And the amino acid that AUG corresponds to is called methionine. So, when this ribosome reads this AUG for the first time after starting from the ribosome binding site, that is a signal for the ribosome to begin the process of translation.

And the first step in that process is the methionine amino acid corresponding to AUG, which is the first amino acid assembled in the growing chain of this amino acid chain that is going to be built. So, that is methionine. Let us say after this triplet, it moves to the next triplet. So this ribosome is going to move from one triplet to another. So the next triplet, let us imagine, is UUU.

That means in the DNA sequence, it was TTT. Now UUU is another triplet, and this one corresponds to tryptophan, which is another amino acid. So what the ribosome does is it links this methionine. So now the ribosome is here. It recognizes that the mRNA sequence says UUU.

I'm sorry, this is not tryptophan. UUU is phenylalanine. This is phenylalanine. So the ribosome recognizes that UUU corresponds to phenylalanine as an amino acid. So it brings a phenylalanine molecule and links it with the methionine that corresponded to the previous triplet and so on.

And it links it. So this one, this circle, is phenylalanine. And it keeps scanning this mRNA sequence one triplet at a time. So, the first three it reads are AUG. So, let us call that number 1, 2, 3.

The second triplet it will read will be 4 to 6, then 7 to 9, and so on and so forth. This process continues until somewhere near the end of the three prime end of the mRNA will be this signal, which will be UAA. UAA is, of course, another triplet that corresponds to a stop signal. That is a signal for the ribosome

that is a signal for the ribosome to stop the process of translation, and the two units then disassemble from the mRNA. But what would be left is this amino acid chain, which was methionine, uracil,

some other, another methionine, because maybe the next one after UUU was again AUG, which means another methionine gets added, and so on and so forth. So we are left with This chain of amino acids, and this is a result of the process of translation that took place inside the cell. Just before we end this discussion, we should see that we are saying that the ribosome reads the mRNA in triplets.

It reads three nucleotides at a time, and these three correspond to a unique amino acid. Which means the question that arises is how many triplets are possible. And you should realize that since mRNA is comprised of AUGC, for the first position, I have four options because it could be any one of these four. For the second position, I again have four options because repeats are allowed. And for the third position, again, I have four options.

So in all, I have a total of 64 possible triplets that could be made. But when it comes to amino acids, We said that we only have 20 amino acids. Which means that every one of these triplets must correspond to one unique amino acid. But since there are more triplets than amino acids, that means you will have scenarios where more than one triplet corresponds to the same amino acid.

And this correspondence between a triplet and an amino acid is called the genetic code. So, for example, let us look at this table. This is the summary of what 64 triplets do, how the 64 triplets correspond to 20 amino acids. If you see here, the rows here indicate the first letter, so that is A. The columns indicate the second letter, so that is U, AUA. And these finer gradations here indicate the third letter.

So, this is AUG. AUG as a triplet corresponds to methionine and it is indicated in red here because this is also the start signal to start the process of protein synthesis. If you see these six codons, each triplet is called a codon. These six triplets or codons correspond to the same amino acid called leucine. You have two codons which are responsible for encoding phenylalanine.

UUU is something that we looked at in the last slide and so on and so forth. You have serine as another amino acid, which is encoded by six different codons. So, in all, we have 64 codons. All triplets correspond to 20 amino acids, except for the fact that three codons are responsible for not corresponding to any amino acid, but as a signal for the ribosome to stop the process of protein synthesis. And these are UAA, UAG, and UGA.

So, this is a one-to-one correspondence, and this is called the standard genetic code. Now, it's nearly universal. There are a few exceptions where a codon might correspond to something else in one species. But, by and large, this correspondence between codons and amino acids is nearly universal,

which means if you're talking about bacteria, viruses, horses, donkeys, plants, or us. AUU always corresponds to isoleucine, and so on and so forth.

We will continue this discussion in the next video. Thank you.