Evolutionary Dynamics
Supreet Saini
Chemical Engineering
Indian Institute of Technology Bombay
Week 03
Lecture 12

Hi, welcome back, everybody. Let us continue our discussion where we left off last time. So the scenario that we are discussing is that we have a flask. We allow for E. coli growth in it. We start with one individual, which is a really hard thing to do—to seed a culture with only one individual—but it doesn't change anything.

And after allowing sufficient time for growth, the number of individuals increases to 10^11. And the number of cell divisions here is 10^11 minus 1 because, as we have seen previously, this is simply equal to NT minus N naught. And this number is roughly equal to 10^11. It is only 1 less than 10^11, roughly equal to 10^11. The mutation rate for E. coli, μ, is 10^-3 per cell per generation, and hence the number of mutants generated

in this growth process is simply equal to 10^11 (approximately, because this is an approximation) multiplied by 10^-3, which is just equal to 10^8. Now, that is a very large number. It is much larger than the total number of possible SNPs. which we saw is just equal to 1.5×10^7. So, in one sense, what we should realize is that in just one growth culture of 100 mL, an E. coli population generated enough diversity.

Then it generated more diversity than it could possibly do. And for natural selection to act, we need variation. And that variation has been supplied via this mutation rate in such a large amount in only one culture that there is sufficient variety to select for individuals which are better adapted in this environment, worse adapted, and so on and so forth. So the key lesson from this exercise for us is that the variation that is generated in a microbial population is via mutation, and that, as we will see, is extremely important. So we started with one genotype, but this genotype after some time for growth gives us individuals which are different from each other.

These are all individuals which are carrying one mutation or the other. So this individual is not, and so on and so forth. And now this process of mutations happening has generated variation—genetic variation, at least. And this variation is genetic because their genotypes are different. This genetic variation will lead to different physical manifestations, different phenotypes, and natural selection acts on this.

This phenotypic variation—so the idea here is that this variation on which natural selection acts is provided via mutations. All right, a couple more points before we leave this discussion for now: how many generations happened How many generations happened? As the population transitioned from N naught equal to one to NT equal to 10 to the power 11 in this case. And we know the formula that NT is simply equal to 2 to the power k times N naught.

$$N_t = N_o\, 2^k$$

Plugging in the numbers that we have, we get 10 to the power 11 equals 2 to the power k into 1 which is 2 to the power k equals 10 to the power 11 which just tells you that k is simply equal to log base 2 of 10 to the power 11 and we can work out what this number is. So, this many This many generations happened. The number of cell divisions, of course, was obviously equal to approximately 10 to the power 11, but this is the number of generations happened in this process.

One more feature associated with this growth process is that when we seeded our culture with this one individual, this one individual is going to divide first. So that happens after generation one. That's just, this is generation one. In generation 1, because we started with only one individual, only one cell division happened. And of course, what is important to realize is that this cell division, we know that statistically speaking, an error happens.

An error happens once every 1000 divisions. An error happens with that frequency. But that once in 1,000 is a stochastic process. It's a random process. Sometimes it might be the first division.

Sometimes it might be the last of the 1,000 divisions. Sometimes it may not happen in the first 1,000, but it may happen twice in the second 1,000, and so on and so forth. So when we make a statement like this, this is a statistical statement which is held true over a very large number of divisions. So it's possible that when this individual's DNA was copied, and I'm representing that by straight lines, it's possible that one mutation happened right there in the first division that happened in the flask itself. Let's say there was an A here which changed to G. So that is the change that happened.

But what we want to acknowledge here is that in the next division these two individuals will replicate giving us four individuals after two generations. So in generation two only two cell divisions. Similarly, in generation three, the number of divisions will be, each one

of these four will now divide. So the number of divisions will be four. And we can generalize this, that after K generations, in the Kth generation,

the number of cell divisions is simply going to be equal to 2 to the power k minus 1. What this means, so in the first generation, the number of cell divisions was 2 to the power 0. In the second generation, it was 2 to the power 1. In the third generation, it was 2 to the power 2. In the kth generation, it is 2 to the power k minus 1.

What that means is that this implies that number of divisions is increasing as generation number is increasing. But we know that number of mutations is just equal to number of divisions by 1000. or number of divisions times 10 power minus 3. The 10 power minus 3 is in the denominator here, so that's 10 power 3. So the more the number of divisions taking place in a generation, the more the number of mutations.

And because the growth in the flask is exponentially increasing like this, there are greater number of divisions taking place later on in the growth cycle as compared to the early one. As a result of this, there is a greater chance for mutations to come up late in the flask as compared to an early one. And this has implications because let us imagine that a mutation came up in the last division. So this is after some number of generations, there is only room for one more generation of growth and after that all nutrients are exhausted. Now, the maximum number of divisions are happening in this particular last generation.

And now, as this individual divides, maybe this individual picks up a mutation. This is the mutation, whereas these two are without the mutation. Now, because all nutrients have been exhausted in the flask, this individual has no opportunity to further divide and for more individuals for it to carry and doesn't have the chance to divide so that more individuals can carry this particular mutation that happened while this was being replicated. Compare this with the mutation that we discussed here that this happened very very early on. So every individual that is born out of this individual is going to carry this mutation A to G not because new mutations are happening because these individuals are simply inheriting the DNA of their parent.

which carried a G and not an A. So this G here in this individual is not a result of mutation, but is simply inheriting its parent's DNA. As a result of that, at the end of the growth process, every individual in this half of this branching process resulting in this tree-like structure, Every individual in this half of this sketch is going to carry that G mutation. So while the mutation happened only once. In this case, in the first division itself, when the process of growth started, that mutation happened only once.

Half the individuals in the population are carrying that mutation when the growth process halted, which means 10 to the power 11 divided by two individuals are carrying that mutation A to G. Because this mutation happened early in the growth process. Compare this with the scenario where mutation happened at the last cell division which was this mutation. This mutation only one individual carries because after this there was no more room for further cell division so that more individuals could carry that mutation. So the timing of mutations is extremely important.

But we know that these mistakes are happening randomly, stochastically. So there is a big element of chance involved in processes such as this. Early mutations are fortuitous. They can spread. They have a chance.

Late mutations—it's harder for them. But that has to be studied in the context of this observation: most mutations actually happen very late in the growth phase. And hence, most mutations do not get the chance to spread in the population. So timing is important, but timing depends on the random chance of when a mutation first appears in the process. In a few videos from now, we will formally discuss how the role of chance dictates evolutionary trajectories.

One more thing we want to discuss here—and this is more like a puzzle—is to imagine an E. coli. It has its DNA, and this DNA is being copied. Now, we want to zoom in and understand how DNA replication takes place. If we zoom into DNA and arrange it as a perfect circle—if we draw the DNA as a perfect circle—then replication is done by a machine called DNA polymerase. But DNA replication starts at a particular location, called the origin of replication.

DNA polymerase binds to this site, moves in one direction, and while moving, it copies the DNA, which acts as a template. And The replication process halts at the diagonally opposite end, called the terminus, when DNA polymerase has made its way through, covering half of the DNA and making an additional copy. The same process occurs at the other end, where another DNA polymerase molecule moves in the opposite direction, copying the DNA, and both polymerases meet at the diagonally opposite end.

The equalized genome we have seen is about 5 into 10 to the power of 6 nucleotides. In the example that we discussed in the previous slide, where growth is facilitated in a medium called LB medium, the temperature is 37 degrees Celsius, and there is shaking. So, oxygen is available for the cells. Under these conditions, the division time of E. coli is roughly 20 minutes. We know this from experiments.

We know this happens. We've made measurements of this and know that replication can take place in 20 minutes. This means that for replication to happen every single time, DNA polymerase must copy the entire organism's DNA because that is something that is faithfully transmitted from one generation to another. The speed at which these DNA polymerases move—so, DNA polymerase speed—so far, we have talked about the accuracy of DNA polymerase and saw that one error happens every 1000 divisions, so it is an extremely accurate machine. The speed at which it copies DNA is about 10 to the power of 3 nucleotides per second.

So, that's also very, very fast. It's putting together a thousand nucleotides every second and makes one error in every thousand divisions. But if we do some simple math and ask how long it takes for this DNA polymerase to copy one half of the DNA—and obviously, the same is happening on the other half— How long does it take for this machine to copy one half of the DNA? So, in one second, it is polymerizing 10 to the power of 3 nucleotides.

The total number of nucleotides that it has to polymerize is half the genome because the other half is simultaneously being done by the other machine. So, this is 5 into 10 to the power of 6 nucleotides. Divided by 2. So, that is 2.5 into 10 to the power of 6 nucleotides. And so, the time taken for DNA replication is

can be calculated as the total number of nucleotides that need to be copied, which is 2.5 into 10 to the power of 6 nucleotides, divided by the speed at which they are being copied, which is 10 to the power of 3 nucleotides per second. This cancels. Newt seconds comes on the top. This 10 to the power of 3 grows. This 6 goes to 3.

So this comes out to 2.5 into 10 to the power of 3 seconds, which is 2500 seconds. Now we just do a simple division by 60 to find out how many minutes this is. So, the time for DNA replication in minutes is simply 2500 divided by 60. And this is 41.8 minutes. And this is a really strange observation that if this is true, if this is the amount of time that DNA polymerase takes to copy DNA, then how can division take place in half the time?

Remember that DNA—all the DNA—is passed on to the progeny. It is not as if the progeny only gets half of it because, I mean, no organism can survive losing half its DNA. So, this is a problem. If it takes 40 minutes to copy DNA alone, how can replication occur in 20 minutes? I leave this as an exercise for you to ponder and consider the possibilities of how the cell can manage both these truths.

I will summarize this in a sentence, but I encourage all interested to look up how this is possible. How can E. coli divide in 20 minutes? If it takes 40 minutes to copy its DNA. That's the question I'd like you all to ponder and research. It's available online.

But it's a really cool mechanism—what E. coli does to ensure that while it needs this, it can also achieve this. One answer could be: if this is the DNA being copied, why not use four DNA polymerases? If these four start from here, each one takes about 20 minutes. You're sharing the load, but that's not what E. coli does. The answer is something entirely different.

Now, while we have studied one type of mutation, There are several other types of mutations too. We have only studied one type of error that DNA polymerase makes. The number of possible error types is actually wider than what we have discussed. So, types of errors or mutation rates.

The first one that we have been looking at is called a SNP, single nucleotide polymorphism. This is just a change from one base to another. And its implications can be quite variable. For example, let's imagine that we have a gene sequence. So, this starts with ATG, ends with TAA, and let us imagine that a change takes place, this mutation takes place somewhere here.

This SNP could lead to several possibilities. One possibility could be that an amino acid change takes place because now the triplet has changed, and hence, looking at the standard genetic code table where we saw the association between triplets and which amino acid is being assembled into the growing polypeptide chain, you could have an amino acid change, and that may or may not affect protein functionality. The other thing that could happen is no amino acid change. And that's also possible because we know that for several amino acids, more than one triplet corresponds to the same amino acid. For example, serine has six codons that encode the same amino acid.

So if the mutation happened such that the prior and the post-mutation codon both correspond to serine, then no nucleotide or amino acid change takes place despite there being a nucleotide change that occurred because of the mutation. That's also possible. The third possibility is that we get a truncated protein. Because this change led to TAA right here. So maybe the original codon here was AAA, but this A changed to a T, and hence it became a TAA, which meant the process of translation halted right here, and we didn't get the full protein that we were looking at prior to the mutation.

So these are three possibilities. This is called a non-synonymous change. This is called a synonymous change. The last one is just... We can have another type of SNP where, if this is the gene, then the mutation could happen not in the gene but actually in the region outside.

And what this could do is that for this gene to be expressed, RNA polymerase has to come and bind here. And maybe this mutation makes RNA polymerase coming and binding here extremely favorable. In which case, prior to the mutation, So this is prior to mutation. The cell was making X amounts of this protein after transcription and translation.

But after the mutation, RNA polymerase comes and binds, making a lot more mRNA. If it makes a lot more mRNA—maybe twice the amount—it's because this mutation has made RNA polymerase binding extremely favorable. So if it is making twice the amount of mRNA, then it is very possible that the amount of protein would also double. Hence, post-mutation, the cell makes twice the amount of protein. So just this one mutation did not change the amino acid sequence, nor did it alter the nucleotide sequence of the gene it was controlling.

But the mutation actually occurred outside the gene, and such mutations are called regulatory mutations. Because what has changed now is not the gene's sequence, but rather how much of it is being made—the aspects of protein production that are being controlled. Hence, in this particular case, the protein resulting from this gene is being made in twice the amount. So that's one example of a SNP. You could have other types of mutations, one of which is called an insertion.

In an insertion, let us imagine that the original DNA was like this, and there was a gene here. But DNA polymerase makes an error, and after the error, the resulting DNA looks like this. It just inserted a piece of DNA that wasn't there before this mutation. This inserted piece could be from another part of the genome, a gene acquired from the environment—not its own—or a piece of DNA traded with another species in the environment. But the point is, if this was a gene and insertion occurred like this, the gene here has been broken into pieces and is most likely non-functional.

So whatever job in this case this gene was doing is likely not going to happen in the progeny. Alternatively, insertions can also happen where you have a gene and an insertion took place, not inside the gene but next to it. And this insertion was carrying another gene which made a protein that was catalyzing, performing a certain function. The catalysis job in the cell, in this context, means this individual, after this insertion, has acquired a

functionality which the parent did not because the red gene wasn't there. Maybe this trait happened between species such that it acquired this mutation, and the insertion happened, giving it properties it earlier did not. The third type of mutation is called deletion, which is the opposite of insertion.

And in this case, let's imagine that you have a piece of DNA, and on this piece of DNA, you have one gene, another gene, and yet another gene. But during the process of replication, an error is made, and this error leads to the deletion of parts that existed. So the progeny's DNA looks like this and then this. The part in between got deleted. And as a result of this, the blue gene is lost, and whatever functionality it was conferring upon the individual is also lost.

Collectively, insertions and deletions are called indels. The last errors that we want to talk about are called duplications. One thing I should mention about insertions is that they could be really small. While the DNA was being copied, this entire red patch was just a single nucleotide A, which the DNA polymerase inserted by mistake. But what that's going to do is, the insertion of this one nucleotide makes the triplets go off frame.

Uh, they were being read in a triplet, in a triplet after transcription. The mRNA on the transcript was being read in a triplet, but now because of this, they are going to be read in a completely different frame because everything is shifted by one. So the amino acid sequence is going to be no longer functional. In duplication, you simply make another copy of a gene. So, if this was the parent DNA, in duplication the error was such that now the progeny has two copies of this DNA. Two copies of this part of DNA. And in this case, it has two copies of this red gene.

And the most natural outcome here is that if this was making X amounts of the protein, this one is going to make two X amounts of this protein. So it simply doubles, and in some cases this can be beneficial, but this has other interesting implications that we'll discuss towards the end of the course in some really interesting evolution experiments. So these are the types of mutations that we can have. We can have SNPs, we can have indels, we can have duplications, and we can have horizontal gene transfer where genes are acquired from another species or the environment and then inserted into your own genome.

This collectively represents the spectrum of mutations that can take place in an organism. The main idea of this lecture again being that the variation for natural selection The variation for natural selection to act on comes from mutations and the diversity generated.

Mutations and the diversity so generated. We will see how this process is fundamentally different when we look at eukaryotic populations in the next video.

Thank you.