**Evolutionary Dynamics**

**Supreet Saini**

**Chemical Engineering**

**Indian Institute of Technology Bombay**

**Week 10**

**Lecture 46**

Hi everyone, welcome to the next video of the course. So, we will take about 10 minutes or so to wrap up our discussion on dN by dS, and starting from there to the end of the course, we are going to discuss experimental evolution in the lab and the lessons from it. So, continuing the dN by dS discussion, this quantity we saw is the ratio of the following, and all of this starts from the comparison of two different They could be two different species or two different groups of the same species. Let us call them A and B. These could be groups or different species.

And let us imagine that we are comparing sequences of a particular gene and we see that this sequence is different at a few positions. The nucleotides that are present in the two sequences are not identical at different these highlighted positions; everywhere else, they are identical. So, we saw that dN by dS is the ratio of the following. Number of non-synonymous differences between two sequences between the two sequences

divided by number of all possible non-synonymous changes that could have happened in the sequence number of all possible non-synonymous changes in the sequence And we saw this at the end of the previous video on how to calculate this. And this ratio has to be divided by number of synonymous differences between two sequences divided by number of all possible synonymous changes in the sequence, and this ratio gives us dN by dS. We just spent a few minutes trying to understand that when we actually implement this strategy on a particular gene in two species, for instance, what do the different values of dN by dS tell us? So typically, this dN by dS ratio, the number that we get, is classified into three different groups. In the first one, if this ratio is much greater than one, what that tells us is that this numerator is much greater than this denominator

And what that implies is that the number of non-synonymous changes that are taking place between these two sequences is far higher than the number of synonymous changes taking place between these two sequences when corrected for this total propensity of

synonymous and non-synonymous mutations in the sequence. So, non-synonymous mutations is thought to be adaptive because non-synonymous is leading to change of amino acid and hence protein functionality. And in this particular case since dn by ds is much greater than 1, what that means is that these two sequences are being selected for different types of functions of the same protein. So, that means non-synonymous changes

are far greater than synonymous changes when normalized for the sequence propensity which implies that the change is being driven by selection. For instance, this could be a species which is now living at 40 degree Celsius, this could be a species which is now living at 20 degree Celsius and originally these two came from a place where the temperature was 30. So, now these enzymes have to work at this elevated temperature and this colder temperature and hence enzyme might acquire mutations so that it has a necessary amount of efficiency working at these different temperatures to which it was already achieved historically. So, dn by ds much greater than 1 tells us that changes are being driven by selection. On the other hand, if dn by ds is much less than 1, what that tells us is that synonymous changes are far more frequent again when corrected for the propensity of the sequence than non-synonymous changes.

And remember, synonymous changes are being driven by drift alone. So drift, and we saw that the possibility of a mutation fixing by drift is 1 upon n, which in microbial populations is really small because n is typically a very large number. Despite that, synonymous changes are fixing much more frequently as compared to non-synonymous changes. So what is happening in this case is that for most of the non-synonymous changes that are occurring, you are making the protein that results from this DNA sequence worse off. And as a result, fitness decreases.

And as a result, it's weeded out from the population and natural selection just keeps rejecting it. So all non-synonymous changes are being rejected by the population by are rejected by natural selection and only synonymous changes because they are neutral are being accepted. So this in this case. Selection is operating. but selection is operating but rejecting the non synonymous changes.

Because non-synonymous changes the amino acid that is going to get incorporated here. And this new amino acid that is being incorporated because of this mutation makes the functionality worse off. Functionality becomes poorer. And as a result, natural selection rejects this variant. So, in cases like this, you will see very very few non-synonymous changes actually taking place.

And lastly, we have dn by ds approximately equal to 1. What that tells is that synonymous changes are approximately the same as non-synonymous changes. And since these are occurring only because of drift and non-synonymous changes are also occurring because of drift, what that implies is that there is no selection on this protein. For instance, maybe this gene corresponds to a protein which is useful for import of galactose. Import of this carbon sugar called galactose.

So, we are comparing the sequence of two galactose transporting genes. closely related species. However, if in the environment that they are present in, galactose is never present. And maybe let us say they grow on glucose only. So, the only carbon source that is present is glucose.

In which case, the sequence of the gene which is responsible for encoding the galactose transport is not under selection at all. And the Just because errors are made randomly, drift will ensure that changes will take place here and because these changes are synonymous and hence neutral. But because errors are being made randomly, suppose a mutation takes place here and this mutation is non-synonymous. But usually non-synonymous mutations are

not neutral they will have a selective coefficient associated with it either beneficial or deleterious but in this case this gene is not under selection so even if you make a non-synonymous change which compromises the function of the galactose transport it doesn't really matter because the cell is not using this transporter anyway the cell needs to use the glucose transporter and this gene is not under selection because in the environment that they are present in the galactose transporter is no longer needed. Hence, even a non-synonymous change gets treated as a neutral case. And that is only in the scenario when the gene that you are comparing is not under selection. So there is no selection on the protein whose gene sequence is being compared.

So you get these three variants: dN by dS greater than one, much greater than one, approximately equal to one, and much less than one, which have different implications and tell us different stories about what is happening with the gene that we are actually looking at. So that sort of wraps up our discussion of dN by dS. And now we sort of switch gears. So the first one-third of the course was sort of fundamentals of biology and evolution. The second one-third, roughly, was a discussion on these conceptual ideas in evolutionary biology: fixation probability, clonal interference, dN by dS, and ideas related to dN by dS.

And in the last third of the course, we are going to study lessons from evolution experiments in the lab. Lab—and as has been the case throughout this course—this will be using microbes, so Let us first see what the advantage is of using microbes and doing evolution experiments in the lab. First, we control the environment. When I am studying bacteria that colonize a garden on campus here,

I can study what species are present, but what I don't have control over is the selection that was placed on these bacteria. For instance, today it rained heavily here, so there is lots of water and moisture available for these nutrients, which will also bring in resources from neighboring areas and so on and so forth. But for the next seven days, it might be very dry, in which case the selection will be for survival in the absence of water. So I don't have control over the environment and hence the selection pressure out there in the ecological niche of the bacteria.

But in the lab, I can precisely control that these are going to be the exact conditions that I'm looking at. So using microbes, their generation time is really short. It is short; typically, we can easily do of the order of 10 generations in a microbial population every day. And that allows me to process a lot of generations in an evolution experiment. In an evolution experiment, the relevant time scale is not the quantum of time that has passed, but the relevant time scale is how many generations we processed.

So, in working with bacteria at this rate, we can do roughly three and a half thousand generations in a year's experiment. Then we can—cloning is easy—we can make manipulations. We can make manipulations in the DNA of the organism very easily and study its effect. We can make freezer stocks, and what this means is these are called frozen fossils. And what this means is that if I start an experiment today with this E. coli and I evolve it for a year—this is one year down the line, 3600 generations.

Then every 1000 generations, I was making freezer stocks and storing them at minus 80 degrees Celsius. Now the purpose of doing this is that at this temperature, they remain viable, so they are not dead. They remain viable and they are metabolically inert. So they're not doing anything. They're just sitting in a while in a deep freezer.

And suppose at the end of this experiment, I want to compare these 3600 generations. And I also want to look at this one after 1000 generations. I can simply take it out of the freezer stock, grow it in a culture media, and compare these two live in the same flask or in neighboring flasks, and so on and so forth. And obviously, now sequencing genomes is

very easy. And one more advantage that I will mention here of doing lab evolution experiments with microbes is that evolution is a stochastic process.

So, the trajectories of evolutionary experiments are noisy. They can go in this direction, that direction, and so on and so forth. So, every single evolution experiment will lead me to a novel outcome. Hence, when we do evolution experiments in the lab, we do not do just one experiment; we do several identical experiments and study whether these identical runs of the experiment lead to the same output or to different results in parallel runs. Using microbes—because they are small and large population sizes of these microbes occupy only a very small space in a lab—we can do hundreds of parallel evolution experiments without really worrying about space as a constraint.

So, logistically, they are just a lot easier to handle. So, when it comes to doing lab evolution experiments in the lab, we generally have three different ways of doing this. One is called serial subculture. This is by far the most commonly used technique for doing lab evolution experiments in the lab. The second one is called a chemostat, something that we have been looking at throughout the course.

And is not very frequently used in labs which study evolution. And last one is called mutation accumulation. Also just in short called as MA experiments. And these mutation accumulation experiments are done in a very limited number of studies when we are asking a very specific type of question. And what is the type of question is something that we will discuss when we come to discussion of MA experiments.

But these are broadly the three type of three ways of to do lab evolution experiment with microbes. So we'll discuss one at a time. We'll primarily, this is what we are most interested in and we'll spend some time discussing this because most of the studies that we discuss will come from a serial subculture experimental design. We'll quickly go through chemostat and then we'll discuss the premise of mutation accumulation experiments and what sort of lessons have we learned from them.

All right. So the first one, which is serial subculture. So, in this experiment, imagine that I want to evolve E. coli in an environment which has only glucose as the carbon source. Suppose I want to set up this kind of an experiment. Presumably, as E. coli evolves in an environment where glucose is the only carbon source, it is going to accumulate mutations which make it volatile.

which provide it with a greater ability to utilize that glucose faster, and those will be the kind of mutations that will happen in the genome of the organism. So, if we do this experiment via a serial subculture method, our strategy would be as follows: we will take a flask in which we will put nutrients. They will have all necessary nutrients needed for E. coli growth, and as a carbon source, they will have glucose and obviously no other carbon source. So, that's done. After that, we will introduce E. coli to this flask and let growth happen. And we will provide appropriate growth conditions for E. coli to grow. So, this will be at 37 degrees Celsius with shaking.

And we let that happen for some amount of time, let us say delta T time. Maybe delta T time is 24 hours. So, I start this experiment at 10 a.m. today, and I let this growth take place for 24 hours. But very soon, you will realize that in such a context, you run into a problem, and the problem is as follows.

When growth happens in this flask, the following will occur. So, let us say I have these two axes. This is time. On one axis, I am tracing the number of bacteria in the flask. And on the other axis, I am tracing the glucose amount still remaining in the flask.

So, let us say bacteria I represent with green and I started at t equal to 0, there was a small number of N naught amount of bacteria and exponential growth takes place and then saturates because all the glucose is gone. And let us say I represent glucose with a red curve and at t equal to 0, there was lots of glucose around. But then as growth started taking place, we ran out of glucose and all the glucose was gone. Maybe this goes to 0. Let me redraw this.

So glucose got, so glucose exhausted. So when I come in the following morning after 24 hours of growth and see that initially number of bacteria that are added was N naught. At the end of 24 hours, the number of bacteria that are present is Nt. So this growth facilitated increase in number of bacteria from N naught to Nt. And if I want to find out how many generations happened in this time, so in one generation the number would become 2 to the power n naught, in two generations the number would become 4 to the power n naught and so on and so forth.

After k generations, the number would become 2 to the power k into n naught, but this number is equal to n t. So, the number of generations k that happened in these 24 hours is n t upon n naught. So I can find out how many generations were processed in this one day of growth. But what do I do next? So this is why this technique is called serial subculture.

So at the end of this 24 hour period, I have the following setup that I have these individuals.

I have NT number of bacteria in the flask. And what we do at this point is let's imagine that this volume is V. I take a very small volume, let's say small V, which is typically let's say V by 1000. So, I take a small volume and I add this volume to a flask which is contains nutrients plus glucose. So the composition of nutrients of this flask is exactly the same as the composition of nutrients that were present in this particular flask.

So these two are identical. Basically, what I'm doing is after this growth process in flask number one, all the glucose is gone and NT has reached this point where it can't increase any further. So what I do? I go to a fresh flask and which has again this high amount of glucose. This is that flask.

And from this antibacteria, I randomly take N naught number of bacteria and I put it in this new flask. So this small volume goes in this new flask. And obviously, when I'm taking V by 1000 amount of volume, it also brings in some bacteria. This is the N naught number of bacteria that will seed the second flask in my experiment. Glucose is already present because this is a fresh flask and so on and so forth.

And I give it growth conditions of 37 degrees Celsius, shaking. and delta T of 24 hours and I let the same thing happen again. And in the second flask, I will keep on, I will process this many number of generations and so on and so forth, it will keep on progressing. So again, serial subculture is a simple design and it's sort of the most widely used and most popular design of doing lab evolution experiments with microbes. I have a flask with volume V, I introduce bacteria, I let growth take place for delta T amount of time.

This could be anything depending on what the experiment is. After this delta T amount of growth time, what is going to happen is that this flask is full of bacteria. no more growth can take place. At that point, I take a small volume, small v, which can be anything, but let us say is capital V upon 1000. And this volume goes in flask, so let us call it flask 1.

And this volume goes in a flask, which has volume v, has same nutrients, flask 1 and we introduce this and in this volume obviously there will be some bacteria and we again allow growth to take place for Delta T amount of time at the end of which this flask will be full of bacteria after which we take small v equal to V by 1000 and this goes in a fresh flask which has capital V and has the same resources. And this is flask 3, this is flask 2.

And as we just saw in the previous slide, that the number of generations K that are processed in every flask is ln of nT upon n naught.

So, every flask I am processing K, I am processing K generations. And what that means is that if I continue to do this for F number of flasks, each flask taking one day, then the total number of generations I have processed is simply equal to K multiplied by F. And very rapidly, this number can grow to a very large number of generations, enough for me to be able to see evolutionary change take place. So, we'll continue this discussion of serial subculture and the remaining two ways of doing evolution in the next video.

Thank you.