Evolutionary Dynamics
Supreet Saini
Chemical Engineering
Indian Institute of Technology Bombay
Week 06
Lecture 27

Hi everyone, welcome back to the next video of the course. We'll continue our discussion of fitness landscapes and how experimentally people try to work out how evolution works in the context of these adaptive walks being studied on the analogy of fitness landscapes. And just as a recap of what we did in the previous one or two videos, we discussed this paper by Daniel Hartle and coworkers where we compared five mutations in a single gene, and the effect of these five mutations was that it increased the MIC by 1000-fold.

The increase of 1000-fold in MIC was accomplished via these five mutations, and we saw that transitioning from the strain which had a lower MIC to the one which had this thousand-fold MIC, there were 120 different paths or orders in which these mutations were acquired, which we could take to go from one strain to the other. And out of these 120 paths, if we use MIC as a proxy for fitness, only 18 were adaptive or permissible in a Darwinian sense, where natural selection would keep increasing the MIC or fitness after the acquisition of every subsequent mutation. But more important than that, and perhaps more surprisingly than that, 102 of these 120 paths were not accessible in a Darwinian sense, which meant that in between these 102 paths, there was a fitness peak that one encountered, which meant that you were trapped in a local maximum and wouldn't be able to reach the five mutations.

So while this makes protein evolution unpredictable because we wouldn't know when we would get trapped in a local maximum, and this local maximum is accomplished because of something called sign epistasis, which we discussed two videos ago. So epistasis has this characteristic of making paths inaccessible. Sign epistasis does that. But it also means that instead of worrying about any of the 120 paths that this mutational pathway could have taken, I only have to worry about the 18 which were permissible in a Darwinian sense. So that was the first paper that we looked at.

We will continue our discussion of the second paper, which was published in 2011 in Science. And the senior author of this is Tim Cooper. And just as a small recap of the experiment, the experiment was as follows: we have E. coli, which is being evolved in an environment that has low amounts of glucose. As the only carbon source available for the

cells, E. coli is being evolved in this environment for a long time. Obviously, mutations are going to accumulate in the bacteria that are in this flask, and these mutations are going to be the ones that make E. coli fitter in the environment they are dealing with.

So after the acquisition of five mutations, The fitness—if this was the ancestor E. coli, and let us say this was the genome of this ancestor E. coli—then after the acquisition of five mutations, let us say this is evolved E. coli, and this is its genome. This evolved genome has five mutations, which are spread in five different parts of the DNA and are associated with five different genes that are concerned with five different functionalities inside the cell. If the fitness of the starting genotype—let us say the fitness of the starting genotype is Let us call this 1.

Then after the acquisition of these five mutations in the genome, the fitness has increased to 1.35. So there is a 35% increase in the fitness of the organism after the acquisition of these five mutations. Let us quickly—I want to give a sense of what the nature of the genes that are targeted in a process like this is. So we will not get too much into the details of what the mutations were doing, but we want to list the names of the genes in which these mutations occurred. And the order in which they occurred in the experiment is the following.

The first mutation to have occurred was in a gene called RBS operon. That is the first mutation. The second mutation occurred in a gene called TOPA. The third one, this was the second mutation. The third mutation occurred in a gene called SPOTI.

Fourth one happened in the promoter region of SPOTI. operon which drives the expression of GLM-US and the fifth one occurred in a gene called PYKF. So these are the five regions of the genome that were targeted in this specific order as the strain acquired these five mutations and became 35% fitter. Four of these mutations are in the coding region, which means that if I have a gene, Let us say this is the region of the gene.

This is the gene. It starts with ATG, ends with a stop codon, which could be any one of the three, but let us say it is TAA. So, in four of these cases, 1, 2, 3 and 5, the mutation happened somewhere in the middle of the gene where a change in nucleotide happened. That's happening in these four cases. In the fifth case, something else happened where I have these two genes, GLMU, GLMS, and they are collectively written as GLMUS.

And the mutation that happened in this case occurred in the promoter region. This region—a mutation in this region—can often have one of two effects: it increases the rate at which

transcription happens by facilitating transcription factor binding or facilitating the binding of RNA polymerase. Conversely, it can also reduce the amount of transcription taking place for the genes under this particular promoter. So, this region of a gene is called the promoter region. So, these are the five mutations that occurred, leading to a change in fitness from 1 to 1.35—a 35% increase.

What we should also note is that a 35% increase is facilitated by the acquisition of five mutations, which means that roughly each beneficial mutation—each of these is a beneficial mutation—confers a roughly 7% advantage. So, this is also the first time we are seeing a typical benefit associated with a beneficial mutation. So far, all we've been saying is that a beneficial mutation occurs and fitness increases. But without any experimental data from a real evolutionary experiment, what is the quantum of benefit that a beneficial mutation confers?

And this is the first time we are seeing that data—that five to seven percent is a typical number associated with a strong beneficial mutation. OK, so as we did in the previous study, we are going to look at adaptive paths that are accessible or not accessible in this case as well. The same idea follows: this is the number of mutations. This is the ancestor, which carried zero mutations, or the ancestor that I started with. This is the evolved strain, which carried five mutations.

Evolved. As I mentioned, this experiment has been going on for more than 80,000 generations, and many more mutations have subsequently been acquired by these strains. This particular paper was published, keeping in mind only the first five mutations that got selected in these lines. And here is fitness. Which we mentioned was quantified by measuring growth rate as a proxy.

So, our ancestral strain conferred a fitness of 1—let us say this is 1—and the evolved strain conferred a fitness of 1.35. And now, as we saw in the last example, going from 0 mutations to 5 mutations, there are 120 paths. Let us call those mutations A, B, C, D, E. So, again, for those of you who did not understand that, How do we get these 120 mutations? When moving from 0 mutants to 1 mutant, I can pick any one of these 5 mutations.

Let's say I pick B. So that's my first mutation. Now, I have to pick the second mutation. So, I have 5 choices to pick the first mutation, out of which I picked B. For the second mutation, I now have 4 choices available, and let us say I pick D for the second mutation. So, I had 4 choices. For the third mutation, I have 3 choices available, and let us say I pick A.

That is 3. And then similarly for the fourth one, I only have two options available to me. This is my fourth mutation. And the fifth one, whatever is left, I have to pick that. So that is the only one.

In all, there are 120 paths. And again, if a path is monotonically increasing, 2 mutations, 3 mutations, 4 and 5, this type of a path is a monotonically increasing function, then this is called permissible in a Darwinian sense. permissible path. On the other hand, I could also have paths which look like this. In this case, after acquisition of the second mutation, the third mutation when it happened was deleterious because it led to decrease in fitness.

In this case, this is a path which has a local peak and hence it is non permissible in a Darwinian sense. To understand A little bit more about why these local peaks arise. Let's imagine that the order of mutations in the green path was A, B, C, D and E. Which means this genotype was A, this genotype was AB, this genotype was ABC, this one is ABCD. And finally, we get to the last genotype, which is carrying all five mutations.

It becomes ABCDE. On the other hand, in another path—the blue path—where the order of mutations has been changed, let us imagine that the order in which these mutations were acquired in the blue path is as follows. Let us say the first mutation to occur was C, the next one was E, then B, then D, and the last one was A. Now, what you should realize here is that this genotype is C, this is CE, this one is CEB, and this is CEDB. And the last one—the last one is the same in every path because the last one is the same genotype where all five mutations are present. What you should realize here is: let's focus on mutation B. Let's see what happens when mutation B occurs on the green path.

On the green path, mutation B occurs on a genotype that already carries mutation A. So, mutation A is already present in the genotype when B happens, and on this genetic background, the new mutation B takes place. On this background, B takes place, and as you can see here, going from genotype only A to genotype AB, fitness increases from this level to this level. So, the acquisition of mutation B on a background where mutation A was already present led to an increase in fitness. Hence, on the green path, mutation B—small b—was a beneficial mutation. But that's not the case for the blue path.

Let's take a look at mutation B on the blue path. Now, in this case, mutation B occurs here when C and E are the two mutations that are already present. So, in this case, let's say C is already there and E is already there. And in this context, mutation B takes place on this genome. But the CE genotype had a fitness given by this number.

But on this background, when mutation B occurs, fitness drops significantly to this level. As a result, when mutation B occurs in a background that already contains mutation CE, fitness decreases dramatically, and in that context, this mutation is deleterious. Deleterious mutation. So what you should note here is that the effect of mutation B on fitness is context-dependent, and that context is dictated by the precise genotype in which mutation B occurs. If it happened in one, it was beneficial, but if it happened in the blue path, it was deleterious—the same mutation.

And this is, of course, the definition of sign epistasis: a mutation is beneficial in one context but becomes deleterious in another. And it is sign epistasis that leads to these local peaks. So. So, that is the reason why we have paths like these blue paths, which have this local peak, and the population would get trapped here. Back to our study of Cooper and coworkers, of the 120 paths, we are interested in knowing how many paths belong to this category, the green category, and how many belong to this category.

So, as with the previous study of Hartle and coworkers, this study also quantifies all 120 paths and tells us how many belong to the green category and how many to the blue. And what they record is that out of 120, 86 paths belong to the green category, and the remaining 34 paths belong to the blue category. Which means that in this landscape, most of the paths are adaptive in nature. That means they are permissible in a Darwinian sense, whereas a much smaller number—roughly 30% of the paths—are not accessible via Darwinian fitness. Compare this with the previous study, where more than 80% of paths were not accessible and less than 20% were accessible.

So that number of accessibility has increased to over 70% and that's a huge change that has happened and the discrepancy that we note between the two studies. So let's try and summarize this discrepancy here that We have looked at two studies. Each of them had five mutations. So this was the zero mutation variant.

This was the five mutation variant. In both these studies, the zero mutation variant was low fitness. Five mutation variant was high fitness. Fitness was quantified differently in the two studies. In the first one, it was MIC.

In the second one, it was growth rate. And MIC was Heartland co-workers and growth rate was Cooper. And we know that in these two, there were two types of paths. One was strictly increasing paths, such as this. This is a permissible path.

Alternatively, we had a qualitatively different type of path which had a local peak. This is a non-permissible, a non-permissible path. And the relative fractions in Hartle, this was greater than 80%. And in Cooper, this was less than 30%. And by the same token, for the permissible paths in the Hartle study,

This was less than 20%. And in the Cooper study, this was greater than 70%. So the question then arises: how do we generalize this process, and how do we understand what has happened here? In one case, we observe numbers which are like this. They tell us that evolutionary processes, these evolutionary dynamics, proceed in a way such that a huge majority of paths are non-permissible.

But in the other case, we get an answer which is the exact opposite of what we saw, which is that in a huge majority of cases, paths are permissible in a Darwinian sense. So how do we reconcile this? In some sense, it's a very fortuitous comparison where both studies are doing the exact same comparison of five-point mutations, 120 paths, and everything is the same organism and so on and so forth. But you get these diametrically opposite results. So the way we think about this issue now is the following.

That one big difference in the two studies was that in the first one, which was by Huddle and group, what was studied was one gene that acquired five mutations. All of this happened in one gene. In the second study, Cooper and group, the mutations that occurred were in five different parts of the genome and in five different genes. Five different genes. And since then, this problem has been studied in different contexts.

This model system was done here in one gene, but we have also studied other genes to examine this particular phenomenon of the Huttel group. And what we now know is that when it comes to studying epistasis within a gene, sign epistasis is rampant. There is a lot of sign epistasis, which means that when a gene is evolving, the genomic context in which a mutation occurs becomes very important.

The same mutation in one context could be beneficial and could be deleterious in another context. And the way we should understand it is that this entire piece of DNA is going to be transcribed and translated. So we will get the mRNA, then we will get the amino acid chain, and this amino acid chain will fold to form the protein. However, if a mutation occurs, this mutation in a folded state will interact with neighboring amino acids. And it is these interactions that matter.

This interaction might be a good interaction. But if another mutation changed the amino acid right here to red, then this green may not want to interact with the red amino acid. And in that case, it becomes deleterious. So these interactions between amino acids in a protein because all these mutations are happening in the same gene, the gene when transcribed and translated has to, the resulting protein has to fold and these amino acid residues have to interact with each other. That's not the case that is happening here because each of these mutations is in five different genes.

And these genes, at least in this context, are not coming together and working together as a protein complex. They are all doing their independent things, not directly associated with each other. So, sign epistasis is very prevalent when we are talking of epistasis within a gene, but not that prevalent when we are talking of epistasis between genes. Instead, what we note in the five gene study is that negative epistasis dominates when we are talking of between genes.

Remember negative epistasis means that this is the fitness of the background. So, let us say there are two, this is one background and this is another background, this one carries this one mutation, this fitness is F 1, this fitness is F 2 and on these on both these backgrounds the same mutation happens. Let us say on both these backgrounds this particular mutation happens. Now, when this mutation happens, this is a beneficial mutation and on the y-axis, we will say effect of mutation.

The same mutation happening—negative epistasis says that it will have a greater benefit in a background of lower fitness compared to the benefit it confers on a genome with a higher background. So, even though the mutation happening is exactly the same in both of them, their fitnesses are different. Hence, the fitness benefit that this same mutation confers on the two will be greater in the background with lower fitness compared to the background with higher fitness. And that's negative epistasis.

These issues of epistasis—these properties associated with negative epistasis—are what dominate epistasis. What is the dominant feature of epistasis when it comes to between-gene epistasis? Through these two studies, we want to understand—we saw that epistasis plays out differently within a gene and between genes. In the next video, we will continue with the third example, which is a more recent study published last year. Thank you.