Evolutionary Dynamics
Supreet Saini
Chemical Engineering
Indian Institute of Technology Bombay
Week 05
Lecture 25

Thank you. Hi, everyone. Welcome to the next lecture. We'll start our discussion on fitness landscapes and how experimentally we can decipher them. The first one is from a paper that came out about 20 years ago in Science.

This was one of the first exhaustive surveys at the time of a fitness landscape of a particular gene in E. coli. This particular paper discusses antibiotic resistance. As you might know, antibiotics were discovered in the first half of the 20th century. When they were discovered, there's a famous example of Fleming discovering penicillin on his Petri plates. This was in the earlier decades of the 20th century, but as the century progressed—during World War II, for instance—antibiotics were a great savior because soldiers had wounds.

Because of that, they would get infections. Antibiotics were a tool used to control infections and have saved numerous lives then and since then. What was thought at the time was that now that we have access to these wonderful chemicals called antibiotics, whenever there is an infection or unwanted growth of bacterial species, we can simply take these antibiotics and apply them to that area. That could be any environmental niche or even my body—I can just consume these antibiotics, and these bacteria would die. It was thought that this is such a wonderful tool, and the problem of bacterial infections and all the health-related issues that arise with them are now taken care of because we have these antibiotics at hand.

What was not realized at the time and which is now abundantly clear in the last few years and decades is that the bacteria would adapt to the environment that we are placing them in and any mutation that makes their survival or increases their fitness in that particular environment, increases their fitness, will be selected for. In this case, since we are talking of antibiotics, When we are applying antibiotics rampantly, then we are selecting for bacteria to acquire mutations which make them resistant to antibiotics. Because when we are applying antibiotics, any bacteria which is susceptible to the antibiotic would simply die. And only those which are carrying these mutations which make them resistant to antibiotic would not just be able to survive, but also replicate in that area.

And as a result, there is spread of antibiotic resistance among bacteria. In fact, one of the hotspots where antibiotic resistance strains are seen most are hospitals because hospitals are constantly using antibiotics for various purposes. And as a result, any bacteria, all surfaces there, they are under tremendous selection pressure to survive in cleaning agents, antibiotics and so on and so forth. And hence, resistant mutations are most frequently found from hospital settings. So from the context of this paper, we want to see how antibiotic resistance can be seen to evolve in this particular gene in E. coli.

So let's imagine that we have two variants of E. coli. Both these variants have the exact same DNA except some differences in one gene. Remember that E. coli genome is about five into 10 power six and the number of genes in E. coli is about 5,000. The only difference that exists between these two variants of E. coli is in one gene. Let's say this is that gene, this is that gene.

So, let me, for reference, let me just write these numbers down. So, E. coli genome is 5 into 10 power 6 nucleotides. Number of genes in E. coli is 5,000. And length of each gene is about 1,000. So these are some broad ballpark back-of-the-envelope type, useful for back-of-the-envelope type calculations numbers that we should take.

So let's say there is this gene X. In these two species, the DNA is identical everywhere else except for five mutations that occur in gene X. So, let us say in this variant A, the mutations are marked in black and in variant B, these nucleotides where they are different is marked in red. So, first thing we should realize that Hamming distance between A and B is 5. What having this does is that it makes one particular species resistant to antibiotic.

And this particular variant of the gene makes it susceptible to antibiotics. When we say antibiotic, we don't mean to say that this particular variant is resistant to all antibiotics and this particular variant is susceptible to all antibiotics. Antibiotics typically target processes which are central to cellular function. So if this is a cell, then and if this is DNA, then we have central processes associated like transcription, translation.

This is done by ribosomes. This is done by RNA polymerase. DNA itself is copied by DNA polymerase. We have cell membrane. which maintains the integrity of the cell.

And all of these different antibiotics target different processes. Each of these processes is a key process in cellular functioning. And each of these is targeted by different antibiotics. So in the case, although we won't get into the names and the biochemical details associated with it, in the case that we are discussing here, the antibiotic that we are talking of particularly has to do with synthesis of cell membrane. So this antibiotic targets and destroys cell membrane and the cell dies.

However, if you are resistant, you will survive the presence of antibiotic. Okay. So with this we move forward and so far we are saying that if I have a bacterial species, it is either susceptible to antibiotic, which means that if antibiotic is present, bacteria dies or you could be resistant in which case if antibiotic is present bacteria survives. However, that's not a classification that we'll use.

We will define something called the minimum inhibitory concentration of an antibiotic. Minimum inhibitory concentration, which is also referred to as MIC. So, MIC can be thought of as the following. Let's imagine we have a gradient of antibiotic.

So this is the concentration of antibiotic. Antibiotic very low, zero here, low here, and very high here. Now, let's go back to our two variants, these two variants, and ask the following question: If I have variant A, this is variant A, I grow variant A in this concentration, And it's able to grow because the concentration at the point that we have marked is extremely small, and the bacteria should be able to grow. In the next experiment, I try to grow variant A at a slightly higher concentration. And I find that it's still able to grow.

And then I increase the concentration even more. And then I can barely register growth, but there is still some growth that takes place. Eventually, I will come to a concentration where bacteria, this A variant bacteria, does not exhibit growth. At this concentration, A does not exhibit growth. And as I move to subsequent higher concentrations of antibiotic, it also does not exhibit growth.

So the question that we ask is: What is the minimum concentration of antibiotic? What is the minimum concentration of antibiotic such that growth will stop? And this concentration—the minimum concentration at which growth stops—is called MIC. So let's call this concentration CA. So the MIC for strain A of E. coli variant A is just CA.

Now I do the same experiment with variant B. And I find that variant B is able to exhibit growth at a far higher concentration of antibiotic. In fact, it's only at this concentration that variant B does not exhibit growth. Let us call this concentration CB. So, the MIC of variant B is simply equal to CB.

So this MIC is of great clinical relevance, obviously, because it determines the antibiotic dose at which growth will stop if you have a bacterial infection. Because anything below that will not stop growth, and hence this is of clinical relevance. So what Hartl and coworkers report in this paper is the following: There are two variants of E. coli that they isolated from nature. They both carry the same gene.

However, one variant of the gene carries five mutations. Let's say Let's say the nucleotide present in the original sequence in one of the strains was A at each of these five positions. It wasn't A, but we're writing A just to make the analysis a little bit easier. And let's say it changed to G at each of these five positions.

So, And if we measure, and the rest of the genome is identical, so we don't worry about it from this point on. And what we note is that if the MIC of this strain is some concentration C naught, MIC with respect to a particular antibiotic, Then the MIC of the other strain was 1,000 times C0. And these were two strains that they isolated from different environmental niches.

As I mentioned, a hospital is a great place to select for strains that have antibiotic resistance. So the question that they ask is that if I make these, let's label these positions from one to five. Let's call them one, two, three, four, five. The question that they ask is the following. That if I draw the following graph

Where the x-axis is, let's call this ancestor. Let's call this ecological strain. So the ecological strain is carrying five mutations. The ancestral strain is carrying zero mutations. So this, in some sense, represents the genotypic space—the possible genotypes.

On the y-axis in the fitness landscape, we have been seeing fitness, and fitness has been measured with a proxy, which is growth rate. But in this case, the proxy for fitness is going to be MIC. How much antibiotic can a strain sense in the environment and still be able to grow? That is a measure of growth.

That is a proxy for fitness in this particular environment. So, as we have seen, when you have zero mutations, the MIC is really small, and it is just C naught. C naught. However, when you have all five mutations, such as this version, then the MIC is really high. Let me redraw this.

The MIC is really high, and it is actually a thousand-fold. 1000 times C naught. Now the question is, what happens if, to this variant, I go to a Hamming distance of one neighbor, which will be GAAAA, then GGAAA, then GGGAA? Next, G, G, G, G, A. And finally, the sequence of the ecological strain. This represents one path that the population can traverse when going from the ancestral strain to the ecological strain.

There are other paths also, and we'll come to that in a second. But this is one possible path. The question they are asking is, if I go from zero mutations to one mutation, to two mutations, to three mutations, to four mutations. The first question we have before us is, how many paths are possible? How many paths exist?

To go from the ancestor strain to the ecological strain. How many paths exist? Let's see. So the ancestor strain is AAAAA. The ecological strain is GGGGG.

This is 0 mutations. This is 5 mutations. From 0 mutation, I have to move to 1 mutation. But that one mutation, I have several options available. That one mutation could be at the first place, in which case the genotype will be GAAAA.

Or it could also be AGAAA. Or it could be AAGAA. Or, let's get rid of this. Or it could be AAAGA and lastly it could be this is g a and lastly it could be four a's and then g (AAAAG) so this is zero mutation this is five mutations

And as I move from 0 mutation to 1 mutation, there are 5 possible options for me. 5 possible options. It could go to this, to this, to this, to this, or this. Note that each one of these is having distance one from ancestor. Each one of these five sequences is having distance one, and hence that move is permitted.

So, similarly, Now, we are going to quicken this up and say that to go from, so since there are 5 mutations, to go from 0 to 1 mutation like this, there were 5 options available. To go from 1 to 2, there are 4 options available because of the 5 mutations, I have already used up 1 in the first mutation that I picked. Similarly, from going from two to three mutations, there are three options available because I've used two up in going from zero to one and one to two. Similarly, three to four is just two mutations and four to five is no choice because depending on which four I picked up before, there is only one choice left.

So the total number of paths that are there to go from AAAAA to traversing some path and arriving at GGGGG is simply equal to 5 into 4 into 3 into 2 into 1, which is 120. This, as some of you I'm sure will remember, is referred to as 5 factorial (5!). So we have 120 paths to go from the ancestral sequence to the ecological strain sequence. Okay, so what we have so far is the following, that I have ancestral sequence and I have ecological strain sequence.

Fitness is measured via a proxy, and that proxy is MIC. And ancestor is right here, and ecological strain is very high up. And to go from one to the other, I have one mutation, two, three, four. And then this is all five mutations. Now, there are 120 paths to go from ancestor sequence to the ecologically isolated sequence.

And in these 120 paths, if I measure the MIC associated with each variation that I'm each variant strain that I'm creating. So there is. So let's imagine that we have a particular path, which is like this. In this particular path, this is the strain which carries one mutation. This is the strain which carries two mutations, this one and on top of this one more and on top of that one more and then one more and finally to the five mutation sequence.

And corresponding to each strain, I have the MIC values. So what we are seeing in this particular path, so shown here is just one path out of, 120 possible paths. In this particular path, what you should notice is that with every subsequent addition of a mutation, the MIC associated with the strain is increasing and that's extremely important. that in this particular path, with every additional mutation, fitness is increasing, which means movement along this path is permissible by a natural selection.

Because if I have a population of these individuals, if I have a population of these individuals and this variant comes up, this variant gets selected for. And then the population is here. But then this variant comes and this happens and so on and so forth. And eventually, following this particular path, the population can reach the five mutation ecological sequence. As a result, we say that this path is

Permissible as far as natural selection is concerned. So this path is permissible, it makes sense, and it's also a likely explanation of how the protein could have evolved. As the antibiotic concentration went up, increasingly higher MIC strains were being selected for. However, on the other hand, we could also have a scenario such as this. This is an ecological strain.

This is the ancestral sequence. And I could have this particular scenario play out. This is the same; this is fitness, MIC, first mutation, second mutation, third mutation, fourth, and then fifth. This path is not permissible by natural selection. That is because natural selection will only permit paths which increase fitness and not decrease fitness.

So when you are starting from here and there is a population of bacteria which exists here, One mutation happens, and this is fine. This bacteria should be able to replace this one. And then another one happens, but then the population is stuck here. Because now the next subsequent mutation decreases MIC and decreases fitness, which is not permissible by natural selection.

So the population in this case is stuck in what is called a local peak. Its neighbors on either side are of lower fitness, and hence the population cannot move from here. So the question that Hartl and coworkers ask, again going back to this graph, is that in this graph, ancestor, ecological, fitness as a measure of MIC, This is the ancestor MIC. This is the ecological strain MIC.

The question that Hartl and coworkers ask is, of the 120 paths, how many are permissible like the first one we saw? And how many are not permissible. And in this study, they constructed all 120 paths and studied how many of the 120 In how many of the 120 can evolution take us through a path that is permissible by natural selection? And of the 120, they found that 18 were permissible and 102 were not permissible.

And this is what epistasis does. We'll continue our discussion about this result and what it means in the context of predictability of evolution and epistasis in the next video. Thank you.