Evolutionary Dynamics
Supreet Saini
Chemical Engineering
Indian Institute of Technology Bombay
Week 05
Lecture 24

Hi everyone, welcome back to the next video. We will continue our discussion on epistasis. So we saw last time that if the contribution made by a particular locus is independent of what is present elsewhere, then movement on that adaptive trajectory leads you to the same final conclusion—the same final sequence, which has the maximum fitness. In the toy example that we discussed in the last video, that sequence was GCA, irrespective of what the starting sequence was. That is a case when epistasis is absent.

So, if we look here, more often than not, what is going to be the case is that the fitness contribution—let us imagine again this three-base sequence—and we could have two scenarios. In scenario 1, the fitness contribution of A at position 1 is simply 0.2 and is invariant; it does not depend on what is present at position 2 and position 3. On the other hand, we have scenario two, where the fitness contribution of A at position one is actually dependent on what nucleotides are present at position two and position three. This phenomenon, which is that the fitness contribution of a locus—the fitness contribution of a locus— is dependent on another locus in the genome.

It could be just another one locus or multiple loci in the genome. This phenomenon is referred to as epistasis. In the case of no epistasis, the fitness contribution of A at position one is simply 0.2, and this is independent of what is present at other loci—at other loci, just positions 2 and 3—and this is a case when there is no epistasis. We also saw in the previous video that in the case of no epistasis, the implication here is that irrespective of the starting point, you end up at the highest fitness sequence.

The highest fitness sequence is reached. One of the questions of great interest in evolution is: How predictable is evolution? In one of the popular manifestations of this question, a renowned paleontologist and evolutionary biologist, Stephen Jay Gould, who is no longer with us, asked a question in the late 1980s. The rhetorical question that Gould asks is this: The planet we live on is about four and a half billion years old.

The rhetorical question that Gould asks is: If we were to go back four and a half billion years ago and restart the formation of the planet and everything that has happened since, if we let Earth evolve once again, would evolution take a similar trajectory to the one it has already taken? Would bipeds like human beings be the dominant species on the planet, or is evolution an extremely noisy process that could end up in any other direction, making it very rare for bipeds like us to be the dominant species? This is also a question of chance versus determinism: Which is the more important factor that dictates evolutionary trajectories? If we rerun the Earth experiment a hundred different times and every single time it's Homo sapiens that come to dominate the planet, that means there was some necessity in the evolutionary process itself for species like us to dominate. However, if in these 100 runs the evolutionary outcomes are completely different and organisms of different shapes, sizes, and morphologies emerge every time, that means evolutionary processes are dictated by chance events to a very large extent.

What no epistasis does is say that, irrespective of your starting point, you reach the same peak every single time, which makes the whole process of evolution predictable. So that's one consequence of epistasis not being there: it makes evolution predictable. In the context of fitness landscapes, this can be viewed as follows. Let's imagine that this is the sequence space. Let's imagine that this is the sequence space.

So this two-dimensional space is the sequence space where I'm drawing all the nodes. And this third dimension is the fitness space. And here I have nodes, and each node is pulled up to an extent which is proportional to its fitness. What lack of epistasis would do is give the fitness landscape a structure like this. A single peak, which is a smooth structure.

So let me complete this again. This would be what the structure would look like. What this means is that if I have a starting population here in the sequence space, this is the node that the population starts from. Its fitness is very low, but it will keep acquiring mutations and move up this fitness landscape.

Because every move up is preferred by natural selection and is selected for. And eventually, I will end up at this particular fitness level, which corresponds to a node here on the sequence space. However, if I started from another place in the landscape, let us say somewhere here, then natural selection will still force me up this fitness peak, and I will end up at exactly the same position as before, irrespective of where I start. And I can choose any starting point on this landscape, and the end position would be the same fitness peak, which is the only peak present.

So if there is no epistasis, then I have a single peak, a single peak, and any starting point leads me to this peak. As a result of this, the evolutionary process in the absence of epistasis is predictable. Because I know that it doesn't matter where I start; the endpoint where the sequence is going to end up is already defined.

But it's not as simple as this because epistasis is pervasive. This phenomenon of epistasis that we are discussing is just pervasive in biology. So epistasis is everywhere. That means evolution is dictated by epistasis and hence is not as predictable as it is when there is no epistasis. So let us see what epistasis does.

For the same structure, if epistasis were present, the fitness peaks may look like this. This is the sequence space, and this is fitness. So now what has been shown in this figure is that there is not one but actually two fitness peaks in this landscape. So if I start from here, then I am going to move up and will end up on, let us say, peak 1. If I start from here, I will end up

At peak two. But there is some degree of unpredictability involved, depending on where I start on the fitness landscape. The final destination—the peak that I'm going to end up in—is determined. So just to label this fully, we have this as sequence space. And this as fitness. But even more surprisingly, what happens if I start here?

If I start in the valley between two peaks, then if the first mutation takes place in this direction, I move towards this peak. But if the first mutation takes place towards this direction, then I move towards this peak. So what these two cases illustrate is that, depending on the starting position, The population ends up on peak 1 or 2. If the starting position is this one, you end up on peak 1. If the starting position is this one, you end up on peak 2. However, this particular case represents another This represents another case altogether where your starting position is here, but despite the starting position being here, it's not clear where you will end up.

Sometimes you may end up on peak one, and sometimes on peak two. So from the same starting position, The population could end up at peak 1 or 2. Which is really surprising because now you're saying that it's the first mutation that dictates the evolutionary trajectory. First mutation towards peak two, and then you have started climbing.

So, hence, you can't come down because natural selection will not permit it, and so on and so forth. So, epistasis—the net result of all of this—means that epistasis, makes fitness landscapes more rugged. Ruggedness is a concept that's used to define fitness landscapes

and is a measure of the number of peaks that exist in the landscape. So, this structure is rugged because it has two peaks.

So, epistasis makes fitness landscapes more rugged, and as a result of that, evolution becomes more unpredictable. And this unpredictability comes from two facets. One is that, depending on the position a population occupies on the fitness landscape, it could go to peak one or peak two. That is a function of where one is located on the landscape. And the other unpredictability is that there might be some places where sometimes you will end up on peak one and sometimes on peak two.

So, this is sort of the concept of epistasis, which brings about its own idiosyncrasies while defining evolutionary processes. More formally and mathematically, epistasis is defined as follows. Imagine a DNA sequence. This DNA sequence has no mutation. This DNA sequence might acquire one mutation.

Let's call this mutation A, which brings about a change of delta F1. What that means is that if the fitness of the ancestor was F naught and the fitness of mutant 1 is F1, then delta F1 is simply F1 minus F naught. This could be beneficial, this could be deleterious; we are not making any distinction at this point. Let us also imagine another mutation. This is mutant B.

Mutant B has a fitness F2, which means the effect of that particular mutation, delta F2, is just F2 minus F0. Lastly, we have another sequence, which is a combination of both these mutations: mutation B and mutation A. Suppose we have a genotype which has both these mutations. Let us call the fitness of that genotype F12. So, delta F12 is going to be F12 minus F naught.

That is the scenario we have. If there was no epistasis, and these mutations were making their effects on fitness independent of each other, then you would imagine that delta F12 is simply the sum of delta F1 plus delta F2. That having both these mutations would simply add up the effects of the two mutations by themselves. But that is the case of no epistasis.

Sometimes, so epistasis can play out in the following fashions. Let's imagine a scenario where only mutation A happens. Mutation A. And without any loss of generality, let's assume that mutation A and mutation B by themselves are both beneficial mutations. Up here, we are saying what is the effect on fitness. So let's say mutation A has this much effect on fitness.

This is delta F. So delta F1 is this much. Similarly, we'll do the same for mutation B, which is also a beneficial mutation in the case we are discussing. And let's say its benefit is slightly bigger than delta F1. And this number is delta F2. That's mutA plus mutB.

Now, that's these two. And these two in these genotypes are the only mutations that are present. But now we'll try to pose the different possibilities of what might happen in a case like this when both the mutations are present. So, when... mutation A and mutation B are both present then one of several things could happen.

The simplest case that could happen is that the effect is simply the addition of the two effects. This is the simplest thing that can happen. This is just delta F1 plus delta F2. This is a case of no epistasis. On the other hand, you could also have a case where the effect of the two mutations is greater than the sum of the two individual mutation effects.

You could have this scenario play out. In this case, delta F12 is bigger than delta F1 plus delta F2. So, the actual effect of In the latest case that we have drawn, the actual effect of having both mutations is greater than the sum of the two mutations that we started with. This scenario is referred to as positive epistasis.

In another scenario, the effect of these two mutations could be something like this: on the whole, having these two mutations confers a benefit that is greater than delta F2 but less than the sum of the benefits observed in the no epistasis case. In such a case, epistasis is said to be negative epistasis. In this case, delta F1, 2 is less than delta F1 plus delta F2. The net benefit of having both mutations present in the genome is less than the sum of the two benefits conferred by these mutations when they are present singly in a genotype. And the final case of epistasis that we'll look at here is that you can also have cases where something like this might happen.

This tells you that even though having mutation one or mutation A was beneficial because it increased fitness, having mutation B was also beneficial. It also led to an increase in fitness. But having mutation A and mutation B in the same genotype was actually deleterious. For some reason,

having both these mutations led to the detriment of the cell. And as a result of this, the individual that was carrying both these mutations was less fit compared to individuals that were only carrying either one of the two mutations or even the ancestral one. Having both has a deleterious effect on fitness. This particular manifestation is known as sign epistasis.

And there are more intricacies and complications in each one of these cases. But by and large, these are the three major divisions that we'll be interested in when we are discussing epistasis. That in the case when there is no epistasis, the fitness effects in the double mutant are just the sum of the two individual mutation effects. In positive epistasis, the double mutant has an even greater benefit than the simple sum of the two individual mutations. In negative epistasis, the benefit is less than the sum, but it's still a benefit.

And in sign epistasis, the nature of the mutation flips as compared to the two individual ones where they were found to be beneficial but now are deleterious. Combined together, it actually turns out that this is a deleterious combination. As a result, in sign epistasis, this individual is of much lesser fitness. So the net result of this discussion is that epistasis is a ubiquitous phenomenon in biology. It dictates evolution.

It dictates cellular function in ways that we are still just beginning to understand. And from the context of fitness landscapes, it makes these landscapes multi-peaked, as we saw. And in these multi-peaked structures, because of the structure that we have come up with, evolution becomes less predictable. And it is this unpredictability that is going to define the complications associated with understanding evolutionary processes. So in the next video, we will start our discussion of three experimental fitness landscapes that have been constructed.

And these three studies are all done in E. coli. The first one is a paper published by Daniel Hartl's group in 2006 in the journal Science. The second one is a paper by Tim Cooper's group in 2013, also in Science. And the third one is a paper from Andreas Wagner's group.

This was published late last year, also in Science. This will also give you a sense of papers that describe fitness landscapes in E. coli, of different scales from different cellular contexts, but they will also tell you how much progress we have made in the 17 years between these three experiments toward understanding and experimentally deciphering these structures called fitness landscapes. So, this is what we will start with in the next lecture. Thank you.