An Introduction to Evolutionary Biology

Prof. Sutirth Dey

Biology Department, Population Biology Lab

Indian Institute of Science Education and Research (IISER) Pune

Week 4 Lecture 22

Features of selection: Insights from theory

So, in our last discussion, we derived the one-locus to allele model for viability selection. And I said that for those people who are not very comfortable with math, Who are not able to follow some of the calculus concepts or some of the probability concepts; for them, I am going to repeat the primary messages from that derivation, and this lecture is where we are going to do that. So, for all those people who were able to follow the math and who understood, you know, what was happening, my apologies. But I would still recommend that you go through the contents of this lecture. Because right towards the end, there will be one or two extra things that were not discussed in the previous case. Also, you know it is always a good idea to repeat your understanding a few times because that is what gets stuck in your brain.

So, anyway, without further ado, we are looking at certain features of selection. Primarily from the insights that come from theoretical modeling. So, in the last class, we saw that the average genotypic fitness is given as $\bar{W} = p^2w11 + 2pqw12 + q^2w22$. Where p is the frequency of allele A1, q = (1 - p) is the frequency of allele A2, and w11, w12, and w22 are the genotypic fitnesses.

So, we explicitly talked about the viability selection case, which means that w11, w12, and w22. They represented the survivorship of the genotypes. Then we said that assuming that these three things w11, 12, and 22 are constants, we said that $\Delta p = (pq/2\bar{W})(d\bar{W}/dp)$. where $d\bar{W}/dp$ gives me the way $\bar{W}$ is changing with a change in the

allele frequency p. And I said that this is a very famous equation named after the great population geneticist Sewall Wright.

It is called the equation for fitness landscape, also known as the equation for adaptive landscape. And as I said, this is when the survivorships are constant, when the survivorships are frequency dependent. That is, the survivorship values change with the frequency of each genotype; then we get an extra term over here. this term, and you know there is a [-E(dw/dp)] that enters into the picture. $\Delta p = (pq/2\bar{W}).[(d\bar{W}/dp)-(E(dw/dp))]$ Now, what are the implications of this? The main message that I want to give you in this context is that, in the frequency-independent case, You can understand the dynamics of selection, how p is changing, simply by studying how $\bar{W}$ is changing with p. The entire evolutionary dynamics can be understood in this context. So, I am giving you a graph of $\bar{W}$ against p. Now, here you can see that this is p = 0, and this is p = 1 over here. Now, what happens when p = 0? That means there are no allele A1 in the population, which means all individuals are A2A2.

Now, if all individuals are A2A2, then what is the average fitness of the population? The average fitness of the population is going to be the genotypic fitness of the A2A2 genotype, which is w22. So at this end of the graph, you are going to have when p = 0, $\bar{W} = w_{22}$. What happens at the other end of the graph? When this is equal to one, p = 1, which means that you only have the A1 allele in the population. When you only have the A1 allele in the population, that means all individual genotypes are A1/A1. Which means that the average genotypic fitness is going to be $\bar{W} = w_{11}$, right? So, these two points at the two ends, p = 0 and p = 1, are fixed at w22 and w11, respectively.

Now the rest of the function, how $\bar{W}$ changes with p, is going to be between these two, you know, extreme points. And the steeper the slope of $\bar{W}$, the faster allele frequency will change. That is the main thing over here, and the faster the population is going to evolve. Now, obviously, if you know this vertical distance, that is, the distance between w11 and w22, this distance is very large. Then the chances that this graph is going to be steeper at some point are much larger.

So, roughly speaking, that is why if you have this distance to be large, then The chances that at least during some part of evolution it is going to be very fast increase. Now, with this information in our heads, we are going to look at the three points. that we had, you know, figured out when we were playing with the simulations. So you know this was the lecture before the last one. So the first point that we saw was when the survivorships are close to each other.

When the genotypic fitnesses are close to each other, selection is slow to change the frequencies. When they are very different, the change in frequencies is fast. Now, to explain that, we will quickly go to Excel. So, let us assume that we are keeping our starting frequency at 0.1. And let us say we are now going to have these very different from each other, say 0.9, 0.7, and say 0.1, right? So what I have done over here is show you the change in allele frequency over time; this is a graph we have already seen. I have added an extra graph here.

What graph is this? Here I am putting the frequency of A1, p on the x-axis, and I am plotting the average fitness $\bar{W}$ on the y-axis. Where am I getting this graph from? I am getting this graph from this material over here. So what I have done is, here I have taken p equal to 0, 0.01... So, at a step size of 0.01, I have increased it all the way up to 1. Then I took q = (1-p) and just put it over here. And then in this column, I have put the formula of $\bar{W} = p^2w11 + 2pqw12 + q^2w22$. Where am I getting the values of w11, w12, and w22? I am picking them up from here.

So, if we change things over here, we can see the effect of that in this graph. Now, one quick thing to note is that here I have time on the x-axis, and here I have the frequency of A1 on the x-axis. So, that is why if I change the p-value over here, you know, Things will change in the left graph, but it will not change in the right graph. Why? Because in the right graph, I have already, you know, hardcoded the values of p, which is the frequency of A1, right? So, that is something you have to keep in mind when you are looking at these two graphs. Now, look at what I have put over here.

Remember, I said that you know this side is going to be W22 and this side is going to be W11. I am sorry, but not on this graph; on this graph, this side is going to be w22. So, this is 0.1; that is why I have 0.1 over here. This side is going to be w11. So, you can see this is 0.9; I have 0.9 over here, and the rest of the line is just going to connect these two points. Now, obviously, the greater the difference, the more significant the difference will be between this and this, the vertical distance.

And therefore, the chance of getting a higher slope is greater. This means that the change in allele frequency when the slope is high will be greater; Δp will be greater. And, more importantly, what we saw in the simulation was that the sign of Δp would be the same as the sign of $d\bar{W}/dp$. So, in this particular case, you can see that the slope of the graph is always positive, and therefore, Δp will always be (+). which means Δp will increase in this direction, and the frequency of A1 will keep going up And that is precisely what you see over here.

Now, you make these points come much closer to each other. So, let us say I make this 0.85 and I make this, say, 0.8, and what do you see? You see that starting from the same starting point $p0 = 0.2$, this is taking much longer to hit, you know, the frequency of 1.

And you can see over here that the slope has become much, much shallower; it has become much flatter. So, this tells you that the greater the difference in the relative frequencies, the faster things will happen and vice versa. Okay, so that explains our point 1, and now we go back to our PowerPoint. Where is it? Yeah, okay. So, what was our second point? We saw that sometimes one of the alleles was getting fixed.

In fact, most of the time one of the alleles was getting fixed. However, there were certain situations in which both alleles were present in the population. And also, we saw that in terms of what is happening in the long-term outcome, the equilibrium. Sometimes the allele frequencies matter, and sometimes the allele frequencies do not matter. So, what was going on? So, again going back to our Excel, what happens is like this.

Suppose we have a situation like this, where both homozygotes have lesser fitness than the heterozygote. When you have a scenario like this, your equilibrium frequency is much lower than 1 in every case. It will be sorry, not much lower; it will be somewhere between 0 and 1. So, when that happens, the relationship between $\bar{W}$ and p takes on some kind of convex shape. with a maximum somewhere in the middle.

Whenever you see this thing, this is when you know that The equilibrium frequency is going to be somewhere between 0 and 1. Now, what is the implication of this? The implication of this is that at equilibrium, both A1 and A2 alleles are going to be present in the population. Until now, in all the cases we observed, one allele was going to fixation while the other allele was going to extinction, right? This is the situation where that will not happen. Now, if you have a situation like this, does the initial allele frequency matter? Now, you can satisfy yourself that it actually does not. I can put whatever value I want over here; the equilibrium frequency is always going to be the same.

Now, this was a situation where, as I said, the heterozygote is fitter than either homozygote. This is what is technically known as overdominance, but we have another scenario. Where the heterozygote is less fit than the two homozygotes, let me just do that. So, let us assume that this is 0.9, let us assume that this is 0.8, and let us assume that this is 0.4, say. Now, if you have a scenario like this, look at the shape of the $\bar{W}$ graph. It has now become concave. Now, it has become concave, which means that there is going to be a minimum somewhere in the middle, somewhere about here.

Now, if you see the graph like this, you immediately know that one of the two alleles is still going to go to fixation. But which one it will be depends on where your system is. So think about it. Here you see that the minima are somewhere between 0.4 and 0.5. We can figure out the precise value, but it is somewhere over here. Now, in the system, the frequency of allele A1 on the right-hand side will go up like this, and on the left-hand side, it will go like this. In both situations, $\bar{W}$, the average genotypic fitness is going up. But in one case, allele A1 is going extinct; in the other case, allele A2 is going extinct.

And just to show you in the simulation, here, if we look at it, our equilibrium point is somewhere between 0.4 and 0.5. So, we took our p0 as 0.3, which was to the left of this; therefore, it went in this direction. So, if what I said is correct, if I take it greater than 0.5 or 0.45, then I should have the opposite of this.

In other words, p should go to 1. So, let me make this 0.8, and there you go. So, sometimes you get fixation of allele A1, sometimes you get the fixation of allele A2. So, let me just play with this a few more times, just to show I am doing it as 0.2, which means I am putting it over here. So, what should happen? p should go to 0. Does that happen? Yes, it does. Let me put it this way: let us say 0.75, or let me make it 0.55, okay, a little larger. Let us see what happens. Okay, 0.55 goes to 1. You can see it is going in the other direction. So, if I were to summarize the whole thing, this is what the summary would look like.

If I have a scenario where the fitnesses of the three genotypes are going monotonically, in the sense that $w11 > w12 > w22$ or the other way around $w11 < w12 < w22$. In that case, the consequence for allelic frequency at equilibrium is fixation. One or the other allele is going to have a frequency equal to 1. The allelic frequency, the starting allelic frequency will not matter at all, and which allele will get fixed will depend on.

Whether $w11 > w22$. If so, A1 will be fixed. If $w11 < w22$, A2 will be fixed. If you have a scenario where $w12 >$ both w11 and w22, Heterozygote is superior to both homozygotes; we call it overdominance. In this case, The polymorphism will be maintained, which means that at equilibrium, both alleles will be present in the population. And again, in this case, it does not really matter where you start your allelic frequency. In the case where the heterozygote is less fit than both homozygotes, There you are going to have unstable polymorphism, which means that there will be a single point.

Somehow, if the population reaches there, it might remain stable for some time. But if there is even a little bit of a perturbation, then it will fix for one allele or the other. And in

general, if you are having no noise or anything, the starting frequency will matter in this case in a simulation. So, when you start to the left, then the allele on the left of the equilibrium point is an unstable equilibrium point. Then allele A2 is going to get fixed; when you start to the right, allele A1 will get fixed.

And this is what is called underdominance. Now, why am I making such a big point out of this? This will become clear in one of our later discussions, but realize one thing here. Normally, we say that selection reduces the variation in a population, but that is not what is happening in the middle case. You can see that there is a situation where, even though there is selection, you are going to have. Both alleles are maintained in the population, and this happens when a particular combination of genotypic fitnesses is realized. Basically, when the heterozygote is superior to both homozygotes.

Remember, everything that we have discussed until now is true when the genotypic fitnesses are constants. They do not change based on the genotypic frequencies. What happens if that is not true? That is when you get the frequency dependence of genotypic fitness. If that happens, as we saw in the discussion yesterday, I mean in our last discussion, all kinds of dynamics of p will become possible. You can have the kind of dynamics that we saw for the density-independent case and for the frequency-independent case.

You can even see situations where the dynamics of p become totally chaotic. And here, "chaotic" is a technical term, which means that it never repeats itself. And what kind of dynamics you will end up getting will depend crucially on precisely what. how the genotypic fitnesses depend on the genotypic frequencies.

You have one relationship; you get chaos. You have another relationship; you will get some nice, monotonic kind of stuff, and so on. So, just to give you one random example, here is a situation where, under a certain kind of frequency-dependent selection, This is what is happening to the change in p, you know, over time. As you can see, it is not settling down, and it is not that this is a transient phase or anything. I can simulate it for

thousands of generations; it will always oscillate, you know, very chaotically like this.

And if you look at the corresponding figure for the $\bar{W}$ vs. p graph, you can see that it is a very complicated one. You can straight away see that you have, you know, one minimum over here, another minimum over here, one maximum over here, and so on. So, the basic message I am trying to give you here is that the moment fitness, genotypic fitness, becomes frequency dependent, The predictability of the dynamics becomes much less, and it all becomes very, very context-dependent. Now, what are the implications of this? What I will give you is the implications in detail. This is kind of a summary of whatever we have discussed so far, and once I am done with the detailed discussion, I will give you a one-sentence implication of the whole thing.

So, when the genotypic fitnesses are constant and frequency independent, Then average population fitness will always increase, irrespective of the starting allele frequency. This is what leads to the common notion in the minds of most people that. Selection increases the fitness of the population all the time. That comes out of this derivation. However, as I showed you, the moment you go from frequency independence to frequency dependence, this relationship will break down.

In some cases, fitness will still increase due to selection and average fitness. But in many cases, average fitness will actually go down due to selection. And in many cases, you won't even be able to predict it. It will move around very unpredictably. And the special case where, due to selection, the average fitness of the population can go down, That case is what is known as Darwinian extinction, also known as evolutionary suicide.

Unfortunately, I did not find a very good example of evolutionary suicide in the literature. However, as far as the theory is concerned, it has been studied very thoroughly, and people have shown that under many conditions, A population can actually go down that path and commit evolutionary suicide. Now, everything that we have said so far is if you think of selection in terms of viability, in terms of survivorship. But remember, even Darwin himself, when he composed the thing, talked about some genotypes leaving more

babies than others. And when we were deriving the Hardy-Weinberg principle, we explicitly said that you can have Selection in terms of fertility, and you can have selection in terms of viability.

If you think in terms of fertility, then that entire process actually becomes way, way more complicated. And, for example, one of the things that can come out of a fertility selection scenario is that The allele frequencies can become constant across generations, and yet the genotypic frequencies can keep changing. Just think about that. That is something that does not happen in the viability selection model.

That is why we were looking at only the allele frequency being constant. And when that happened, we said the population had reached equilibrium. But in this particular case, you can have equilibrium in terms of allele frequency. But you cannot have equilibrium in terms of genotypic frequency. This is something to which we will come back; you know this particular phenomenon can happen in another scenario. This is important because often you will see that evolution is simply defined as a change in allele frequency.

Many books actually do that, and that is something that many modern evolutionary biologists are taking grave objection to. And that is why I am explicitly showing you cases where allele frequency change does not. necessarily mean genotype frequency change or does not necessarily mean; I am sorry, where Allele frequency changing does not necessarily lead to genotypic frequency changes or the other way around. Genotypic frequency change does not necessarily lead to allele frequency change. Therefore, when you are talking about evolution, you should be very, very careful and Are you talking in terms of the allelic level or the genotype level? This is again something that is not really present in the literature, but I am sorry, it is not really present in the textbooks.

But if you look at the literature, then people are becoming very, very conscious of this particular thing. And everything that we have said until this point is when we are thinking about selection acting alone. But in any real-life scenario, selection is not acting alone.

Selection is acting in conjunction with a bunch of other evolutionary forces, for example, mutation and genetic drift. about which we will be talking at some point in the next module, and this selection, the way it operates, is Very much dependent on, or rather the dynamics that it leads to, is very much dependent on the genetics of the whole thing.

Whether the alleles being considered are dominant, recessive, or dominant to different degrees? All that changes the way the dynamics proceed, and also a huge elephant in the room is what is known as the transmission fidelity. Now, if you are talking in terms of alleles, then barring mutation, an allele is passed faithfully from parent to offspring. But if you are talking in terms of traits, remember that everything we are discussing can also be thought of in terms of traits. If you are thinking in terms of traits, then traits are not passed as they are from the parent to the offspring.

Traits are passed to various degrees. Some traits are passed very faithfully, while some traits are passed less faithfully, and One can show that the effect of selection is very crucially dependent on the transmission fidelity of the trait being considered. If the transmission fidelity is very high, then it is relatively easy to model it. If the transmission fidelity is not very high, then modeling the effect of selection becomes a very big pain all over the body. So, the simple point that I am trying to make is that, although conceptually, selection is a very simple thing, Three conditions being met, selection will happen, and when selection happens in most cases. Unless you are at an equilibrium or something, the frequencies of the trait or the frequency of the, you know, allele.

All of those are going to change; all that is fine, you know Darwin said it in just a few words. However, when you actually look into the biology of the system and are trying to figure out, you know, What will this lead to? What will be the effect of selection and other forces on the evolutionary trajectory? Then it becomes pretty complicated in most cases to predict it. And whether you see that complication as a mess or as a challenge to grapple with, to figure out. That is what distinguishes you as a non-evolutionary biologist, If you see it as a mess versus something interesting, something superb, if you are an evolutionary biologist. And the people who are doing it, well, there are many,

many people who are studying this problem in great detail.

The population geneticists are obviously studying this in great detail. The quantitative genetics people have a lot to say on this, and experimentally, the people. The class of people who are studying this problem is the evolutionary ecologists to some extent. And, more importantly, the experimental evolutionary biologists. For the experimental evolutionists you know, this is basically their bread and butter. These are the kinds of problems that they study to see if selection is interacting in different ways.

With different evolutionary forces, what exactly is it leading to? Now obviously this is complex, but the course of which you and I are a part is an introductory course. So, as far as we are concerned, we are not going to get into all those complexities. We will keep things simple; we will by and large deal with those examples. Where selection increases the average fitness or some kind of measure of fitness, and so on.

These are the examples that you are typically going to find in most textbooks. Safe area will that is where we will mostly hover from this point onwards. However, I want you to keep in mind, while you are reading all those simple examples and those simple things, that all this is. Just a prelude, just an introductory note for a process that is way more beautiful, way more complex, way more magnificent. And if you forget all the math that I did in the last discussion, along with all the graphs and all the slopes, and this and that, All those relationships, if you forget all that stuff and you remember just this one thing. That although simple, the effect of selection can be complicated; my job in the last lecture and this lecture is actually done.

So, in the next discussion, we are going to look at certain other features of selection. But instead of looking at it from the mathematical point of view, from models, etc., we are going to bring in more experimental insights. See you at the next discussion. Bye.