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 21

Dynamics of Selection 2: Analytical

So, in our last discussion, we simulated using Microsoft Excel how a one-locus, two-allele selection case behaves. And before we go forward, I would like to say that some of you might find it a little difficult to follow me. And create the Excel sheet the way we did it. So, for those people, you can actually, you know, go to this website. This is the one where we had gone previously to simulate Hardy-Weinberg, so I will quickly take you over there. So, you can go here, say "start," and then go to individual simulations. Remember, we are assuming that all other conditions of Hardy-Weinberg are operative. So, we will take the population size as infinite. The number of generations, as we saw, around 200 is good enough. So, we will keep that here, and then you have to go to the additional settings, and go to selection, and then you are going to have all these fitness coefficients for A1A1, A1A2, and A2A2. Set these, and you will be good to go. So, let us say we will put w11 at 0.6, and we will put w12 at 0.9. And let us say we put this w22 as 0.4 and just say to run the simulation. And you will be able to see, you know, the frequency of the alleles as well as the frequency of the genotypes, okay. So, everything that we did with our Microsoft Excel simulation, you will be able to do over here. I would still recommend that, if possible, you try to do the Excel thing.

Because that way you actually understand step by step how the process happens. And I mean while this kind of simulator is very convenient from a teaching perspective, it does not really give you The same level of intimacy with the subject the way you know hand-coded things like the one we did in Microsoft Excel does. So this is only for those people who are finding it difficult to work on a spreadsheet; they can look at this one. Okay, so let us get back to our PowerPoint presentation. So, in our last discussion, we said that, you know, we made a bunch of observations from our Excel simulations and we said that in order to understand those observations in greater detail, We will need to analytically figure out what's really happening in the one-locus, two-allele selection case. So I am going to analytically derive the expression, and I am going to, you know, draw various kinds of graphs. and mathematically show what the dynamics are likely to be under different situations. Now, quite frankly, this particular discussion that we are having today is probably, mathematically speaking, The most involved discussion in the whole 30 hours that we are going to spend together. This particular derivation is slightly advanced in the sense that most regular textbooks on evolutionary biology don't cover it.

It's only the slightly more theoretically oriented books, and not all of them, that show this particular derivation. It would have been very easy for me to just give you the formula and then ask you to take it on faith and go from there. But my personal experience of teaching this for many years at IISER Pune is that until and unless You actually see where the derivation is coming from and what kind of assumptions are being made in various places. You don't really get the feel for the whole thing, so I am going to do the derivation. This is going to require one or two very basic concepts in algebra, a little bit of calculus, and a very little bit of understanding of probability. I will try to explain it in as much detail as simply as possible. I am trying to target those people who might have taken a course in math in 12th grade. Or maybe they took their last math course in class 10. But they are still not afraid of the math thingy, and therefore, you know, they would like to follow what I am trying to do.

Just in case this derivation, you know, seems a little too tough for you somewhere in the middle. For too many concepts, you are not really figuring out how I am getting there. I would still request you to go ahead and see what the main results are that are coming. And in case that becomes too difficult for you, don't bother too much.

Okay. In the next discussion, I will actually start it with the results that I am getting today. And I am going to discuss the implications of those results without going into the math part of it. But as I said, evolution is one of the most highly formalized subjects, you know, mathematically formalized subjects in the whole of biology. So, understanding this particular derivation and where it is coming from will actually be very, very useful. and that is the reason I am going to great lengths to present it here.

So, in order to do this, we will head to different software. This is Microsoft Whiteboard, and okay. So, before we even start doing the derivation, there is one little concept that I have found people often misunderstand. You know, students, they have a little bit of an issue in comprehending that. People who have done a course in algebra do not find it really an issue, but for others, sometimes it is an issue.

So, I will just present this for you here. So, suppose you have, okay, let me slightly increase the, yeah, this is good. Let's say you have five numbers: 4, 4, 4, 5, and 5. Okay. So, 4 appears three times and 5 appears two times.

And I tell you to please take the average of this. What will you do? You are going to do 4 + 4 + 4 + 5 + 5, and since there are five numbers here in total, You are going to divide the whole thing by 5, and that is going to be your average, right? Now, I can write this thing as ((3*4)+(2*5))/5, right? (3*4)+(2*5), which I can write as ((3/5)*4)+((2/5)*5), right? Which I can further write, no, think about it: what is 3/5? 3/5 is the number of times 4 is present in this entire sample. In other words, 3/5 is the frequency of 4, and similarly, 2/5 is the frequency of 5. Therefore, I can write this as the frequency of a value multiplied by that value itself, and then this entire thing, This is, you know, the summation sign; this entire thing is summed over all values, right? In other words, if you have a bunch of numbers and you need to take their average, all you need to do is figure out The frequency of each number, and then multiply that number by the corresponding frequency, and, Add them all up, add in all these products, and that is going to be your average. Very, very simple concept, but this is a concept that is worth keeping in mind before we start this entire thing.

Okay, so the main derivation starts now. So, we are talking about one locus, sorry, two allele selection. Okay. And if you remember, yesterday I said that we are explicitly talking about the viability selection model: viability selection. So, what are our two alleles? Let us assume that the alleles are A1 and A2 and these alleles let us further assume that their frequencies are p and q, where $q = (1 - p)$. So, we will be going back and forth between q and (1-p). This obviously leads to three different genotypes: A1A1, A1A2, and A2A2. So, these are my genotypes, and remember these genotypes before selecting their frequencies: p^2, $2pq$, and q^2. So, these are the genotypic frequencies before selection.

Now, we have selection happening, and we have explicit viability selection happening. What does it mean? That means that each of these three genotypes has a different fitness. Let us call them w11, w12, and w22, and let me slightly increase the point of the pen here. Let us w11, w12, and w22, and these are the fractions of the genotype that are surviving, right? So, the progeny has been born, and now they are surviving as they grow into adulthood. Each genotype has a slightly different survivorship, and those are the ones that are being given by w11, w12, and w22.

So, these are our genotypic fitnesses. Now, if this is the case, what are the frequencies after the selection? So, after selection, the frequencies are p^2w11, $2pqw12$, and q^2w22. These are the genotypic frequencies post-selection. Now, as we saw yesterday during the simulation, these things are not going to sum to 1. Therefore, we need to do some scaling so that all three frequencies sum to 1.

What is that scaling? That scaling is essentially dividing each of them by this sum. Now, this sum we are going to, what is this sum? This is $p^2w_{11} + 2pq\ w_{12} + q^2w_{22}$, right? Now, what is this? As you can see, these are the values of the genotypic fitness multiplied by the corresponding frequencies, right? So, if that is the case, based on what we discussed just now, this is the average genotypic fitness in the population. In other words, $\bar{W} = (p^2w11 + 2pqw12 + q^2w22)$, and that is the average, sorry, average genotypic fitness. So, therefore, this is unscaled what we have over here, and the scaled fitnesses are $p^2w11/\bar{W}$, $2pqw12/\bar{W}$, and $q^2w22/\bar{W}$. Now, before we proceed, it is actually going to be very useful for us to Derive a certain quantity, a certain concept, and that concept is the concept of marginal allelic fitness.

Now, what the hell do I mean by that? Now think about it, fitness is actually never derived at the allelic level. Fitness is defined at the genotypic level: A1A1's fitness, A2A2's fitness, A1A2's fitness, and so on. So, what do I even mean by saying that I want to look at the fitness of an allele? So, here we borrow a concept from the subject of economics, and the concept is that of a marginal value. What this essentially means is that the marginal allelic fitness is the fitness of All the average fitness of all the individuals in the population who contain at least one allele A1.

In other words, let us see. So, marginal allelic fitness is the average fitness of all individuals who have at least one A1 allele. This is marginal allelic fitness for A1, okay, with at least one A1 allele. Now, how do I get to this? Which genotypes have at least one A1 allele? A1A1 and A1A2, right? A2A2 has no A1 allele, so forget about it. Now, what is the genotypic fitness of A1A1? That is w11. What is the genotypic fitness of A1A2? That is w12, right? Now, we have to calculate the average fitness.

For this, we somehow have to scale it with some measure of frequency. How do we get there? Now, think about the definition. The definition says the average fitness of all individuals who have at least one A1 allele. So, let us say that this A1 allele is fixed, right? Now, if one allele is fixed as A1, we have already picked it up. What is the probability, assuming random mating, of course? What is the probability that the other allele you pick is also going to be A1? Under the Hardy-Weinberg conditions, which is where we are operating over here, That is equal to the frequency of the allele A1; in other words, that is equal to p.

Similarly, assuming that one allele has already been picked up as A1, What is the probability that the other allele is going to be A2? That probability is simply q. So, W1*, which is the marginal allelic fitness of allele A1, is equal to $pw11 + qw12$. Similarly, let us think about the marginal allelic fitness of allele A2, which is $W2^* = ?$ What is going to happen? Which are the two genotypes that have A2? One is A1A2, the other is A2A2. So, in this case, the genotypic fitness is w12, and the genotypic fitness is w22, right? Now, given that one allele has been fixed as A2, what is the probability that the other allele is A1? That is equal to the frequency of A1, which is P. Similarly, given that one allele is A2, what is the probability that the other allele is also A2? That is equal to the frequency of A2, which is q, right? So, the sum, in other words, $W2^* = pw12 + qW22$.

These two quantities, although it are not entirely clear to you as of this moment, why we need these two quantities will become clear in a few minutes. So, in order to figure out what really is happening, we come back over here and start looking at this $\bar{W}$. So, we said this $\bar{W} = (p^2w11+ 2pqw12 + q^2w22)$. Now, I can write this as $\bar{W} = (p^2w11 + pqw12 + pqw12 + q^2w22)$. I will break $2pqw12$ up, just writing it in two parts, right? Equal to this, there is a p here; I will take the p as a common factor.

So, $p(pw11 + qw12) + q(pw12 + qw22)$, right? And if you just look at what we derived one second earlier, $pw11 + qw12 = W1^*$ and $pw12 + qw22 = W2^*$. So, if we come back over here, this simply becomes $\bar{W} = pW_1^* + qW_2^*$, right? This is my $\bar{W}$. This is another thing that we need to keep in mind, right? Now, let us keep a third thing in mind. We go back to this definition of $\bar{W}$ and now let us start writing it, you know, all in terms of p. So, what are you going to do? So, $W = (p^2w_{11} + 2pq\, w_{12} + q^2w_{22})$, right? So, 2pq I will write as $2p(1-p)w12$, and q^2 I will write as $(1-p)^2w22$, easy? I will expand it: $p^2w_{11} + 2pw_{12} - 2p^2w_{12} + \ldots$.. So, $(1 - p)^2 = 1 - 2p + p^2$. So, $...1^* w22 - 2pw22 + p^2w22$, simple algebraic expansion; nothing else. Yeah, okay, actually this is fine. Now I want to know how $\bar{W}$ changes with a change in p. Now, in order to know that from calculus, we need to look at the quantity $d\bar{W}/dp$. So we are going to differentiate this with respect to p.

Now, if you have forgotten differentiation, just take it from me that hopefully what I will do is correct. So $2pw11 + 2w12 - 4pw12 - 2w22 + 2pw22$. This w22 goes away. Now, if you remember your calculus, you can see that there should be a few extra terms that I am not putting over here. Why am I not putting them? Because I am explicitly assuming that the w11s, w12s, and w22s are all constant.

In other words, I am explicitly assuming that these things do not vary with P. If I had not made that assumption, then I would have to add three extra terms over here. We will come to that towards the end of today's discussion, but as of this moment, We are going to assume that all these w11, w12, etcetera are frequency independent. So what we are explicitly assuming is that there is frequency-independent selection happening. Frequency independent means that $\bar{W}$, the values of w11, w12, and w22, do not depend on p.

So if that is the case, let us go forward. This is my $d\bar{W}/dp$. Now I can write this $d\bar{W}/dp =$? See, I have a 2 everywhere, right? So I will just take this 2 outside, and this becomes $d\bar{W}/dp = 2(pw11 + w12 -$... So I have taken 2 outside. So I have 2pw12 inside. ||| $d\bar{W}/dp = 2 (pw11 + w12 - 2pw12$... So this 2pw12, I will write this as (-pw12), and I will again write it as (-pw12).

$d\bar{W}/dp = 2 (pw11 + w12 - pw12 - pw12$... 2 has gone outside; 2pw12 remains. I have split it into (-pw12) and (-pw12), and then 2 has gone outside. So I have -w22.

$d\bar{W}/dp = 2(pw11 + w12 - pw12 - pw12 - w22$... 2 has gone outside, so I have plus pw22. $d\bar{W}/dp = 2 (pw11 + w12 - pw12 - pw12 - w22 + pw22)$.

Bracket closed right? $d\bar{W}/dp = 2(pw_{11} +$ ... So this is w12 and w12 over here. So I can take the w12 common, $d\bar{W}/dp = 2(pw11 + w12(1-p)$... right? So this term is done. I will take the (-) outside, and this will become $d\bar{W}/dp = 2(pw11 + w12(1-p) - (pw12 + w22 - pw22))$, okay.

Now $d\bar{W}/dp = 2(pw_{11} + ...)$ and $(1 - p)$ is q. So, this is $d\bar{W}/dp = 2(pw_{11} + qw_{12} -$... I have pw12 over here, and I have w22; -pw22.

I can take $w22$ common over here. Then this will become $w22*(1-p)$, which basically means $qw22$. $d\bar{W}/dp = 2(pw11 + qw12 - (pw12 + qw22))$ In other words, I have $(pw12 + qw22) = ?$. I think you can already see where this is going. Remember, $pw11 + qw12 = W1*$. So, $d\bar{W}/dp = 2(W1* - W2*)$ right? This is the third relationship that we need to develop: $d\bar{W}/dp = 2(W1* - W2*)$ Assuming that our genotypic fitnesses are frequency-independent.

Again, I cannot stress this enough. This assumes frequency-independent selection. Great. So now that we have these things, let us return to our main topics. Remember, we started with trying to figure out the recursion of p due to selection across generations.

So these are the things that we have got. So now, allele frequency in the next generation among the offspring, what is that? We will call that $\bar{p}$; sorry, p' is equal to how you are going to get it? Remember, this is my P, this is my Q. Then $p' = (P + Q) / 2$. In other words, $p' = (p^2w11/\bar{W}) + (1/2)*(2pqw12/\bar{W})$. In other words, I will write it as $(pqw12/\bar{W})$.

Now, $\bar{W}$ is a common denominator. So, W bar is my common denominator, $p' = ?$ I will take the p outside, $p(pw11 +$ ... sorry there is a $w12$ over here, $p' = p(pw11 + qw12)/\bar{W}$.

Again, the stuff inside over here, you can see this is $W1*$. So, $p' = (p.W1*)/\bar{W}$. This is my next-generation allele frequency. Now suppose I want to look at the change in allele frequency. In other words, I want to look at Δp, which is defined here as $p' - p$. By how much has selection been able to change the allele frequency over one generation? Great.

So what is that value going to be? We just saw that p' is this. So $\Delta p = (pW1*/\bar{W}) - p = (pW1* - p\bar{W})/\bar{W}$, or I can take the p outside, $\Delta p = p/\bar{W}(W1* - \bar{W}) = (p(W1* - \bar{W}))/\bar{W}$. Now, if you remember, we at some point when we were looking at, you know, our expressions, saw that $\bar{W} = (pW1* + qW2*)$, right? We have already kept this one over here, so we will import that one over here. $p' = p/\bar{W}\ (W1*-pW1*-qW2*)$, right? $= ?$ Let us take it a bit forward, $p'=p/\bar{W}($... This is $(W1*-pW1*-$..., I can take the $W1*$ common; in other words, this will become $W1*(1-p) = qW1*$. So, $p' = (qW1* - qW2*) = ?$ I can take the q outside, $p' = (pq/\bar{W})(W1^* - W2^*) = (pq(W1^* - W2^*))/\bar{W}$; Now, if you remember when we were talking about $d\bar{W}/dp$, We derived that $d\bar{W}/dp = 2(W1^* - W2^*)$ under the frequency-independent condition. In other words, from this, we can see that $(W1^*-W2^*) = (1/2)(d\bar{W}/dp)$. So now I can take this relationship over here and put it over here. and say that $\Delta p = (pq/2\bar{W})\cdot(d\bar{W}/dp)$. There is a half, so I will put a 2 over here.

This is one of the most famous equations in evolutionary biology, known as Sewall Wright's equation for the fitness landscape. Sometimes people take it one step further, and what they do is let me do it over here. They write this as $\Delta p = (pq/2)\cdot(d(\ln\bar{W})/dp)$, where ln is a natural logarithm. The reason this works is that $d(\ln\bar{W}) = (1/\bar{W})\cdot d\bar{W}$. So, this is also another form in which the equation is often discussed, although this is the form in which we are going to analyze it.

So why is this equation so important? Why is this, you know, I am saying it is one of the most famous equations? To understand the implications of this equation, what we first need to do is just look at its structure. So see that it has p, it has q, and it has $\bar{W}$ over here. What is p? p is an allele frequency. So we know that it will lie between 0 and 1, which means it is positive.

Similarly, $q = (1 - p)$ will also lie between 0 and 1, which means it is positive. $\bar{W}$ is the average genotypic fitness, which basically means this is the average of lots of values, each one of which is between 0 and 1. The average of many positive values means that it will be positive. Therefore, this stuff over here, whatever is inside this circle, is always positive. That implies that the sign of Δp is going to be equal to the sign of $d\bar{W}/dp$. Now, what is Δp? Remember $\Delta p = (p' - p)$, which means by how much the allele frequency is changing in one generation due to selection.

So, what this is saying is that the sign of that will be the same as the sign of $d\bar{W}/dp$. Now, what exactly do I mean by that statement? What I mean by that statement is that if you end up having p on the x-axis and $\bar{W}$ on the y-axis, Then you are going to get some kind of graph, and we will talk about what kind of graph you will get in a few minutes. But that graph, you know, let us say you will get it, says something like this. Or let us assume that you end up getting something like this; whatever, you can get all kinds of shapes. But whatever you get for any value of p, the sign of Δp, in other words, whether p will increase in this direction, is unclear.

Whether p will decrease in this direction will depend on the slope of this particular curve at that point. If the slope is positive, Δp is positive, and p will increase. If the slope is negative, Δp is negative, and p will decrease.

Now let us further examine this $\bar{W}$. It is a fantastic you know thing. So we started by saying that this is our $\bar{W}$, right? Let me reduce the size a bit. Yeah, $p^2w_{11} + 2pq\ w_{12} + q^2w_{22}$. Now, as you can see, this is quadratic in p. What do I mean by that? What I mean is that the power of p, the maximum power that p can have, is 2, right? It cannot go beyond a square. Now we also know that if you have a quadratic, then that quadratic is going to have at most one maximum or one minimum, right? And that maximum or minimum is going to happen when $d\bar{W}/dp = 0$.

Now, what is the meaning of $d\bar{W}/dp = 0$? The meaning of $d\bar{W}/dp = 0$ is that the allele frequency is not changing; it has gone to an equilibrium, right? Now, what are the ways in which this equation can reach equilibrium? So, if you think about it, This equation can reach an equilibrium. So this is p, right? So p, you know, is going from 0 to 1, right? So when $p=0$, that means that $p=0$ implies that $q=1$, which means that the entire population is full of A2A2 individuals, right? There are no A1A1 individuals; there are no A1A2 individuals. So if that is the case, at this point, the average genotypic fitness of the population is just w22 because there are no other individuals. Similarly, when $p = 1$ and $q = 0$, what will happen? You only have A1A1 individuals in the population; there are no A1A2 and no A2A2. In other words, at this point, $\bar{W}$ is going to be equal to w11; here, $\bar{W}$ is going to be equal to w22, right? So, that point on this axis can lie anywhere.

Similarly, over here, it can lie anywhere. These are the two fixed points. Once the population has gone here, either fixed for allele A1 or fixed for allele A2, nothing can change. But barring those two points, what is going to happen in the middle for the intrinsic values of p? Now, suppose you have a scenario like this, where the slope of the curve is always positive. Then, as I said, Δp is always going to be positive. In other words, p will always increase, and you will get a fixed point here at $p = 1$. Similarly, if you have a scenario that looks something like this, you are going to have equilibrium here at $p = 0$.

It will go like this: I mean, these are the things that we saw yesterday, right? We saw that for many cases, the allele frequencies were either at equilibrium. were either going to 0 or going to 1, depending on who had a higher fitness. However, we did end up seeing certain situations where the allele frequency was reaching an equilibrium at an intermediate value of p.

In other words, some p was between 0 and 1. So, at these positions, $d\bar{W}/dp = 0$. Now, if $d\bar{W}/dp = 0$, we have already figured out that $d\bar{W}/dp = 2(W1* - W2*)$, which means that at equilibrium, $2(W1* - W2*) = 0$. In other words, $W1*=W2*$. Now, what is the formula for $W1*$ and $W2*$? So, remember let me just get this above. Remember $pw11 + qw12 = W1* = pw12 + qw22 = W2*$.

So, let me collect all the p's and q's on one side. So, $(pw11 - pw12) = (qw22 - qw12)$. So, I take p as common, $p(w11 - w12) = (w22 - w12)$. I take q as common, right? At equilibrium, remember that is the condition we are dealing here. Now think about it; p, as we just discussed, is an allele frequency, so it is always positive. q, as we just discussed, is an allele frequency, so it is always positive.

Therefore, forget about the magnitude; just think in terms of the sign. The sign has to be equal on both sides, right? So, this relationship in terms of its sign can be true, if and only if w11 and w22, Either both of them are greater than w12, or w11 and w22 are both less than w12. Why am I saying that? Because let's take the first condition: $w11 > w12$, so this is (+); $w22 > w12$, so this is (+); positive positive great. But suppose that is not the case; then you will have positive on one side and negative on the other side, which cannot happen, right? Therefore, this w11, w22 both being greater than w12 will ensure that you are having a positive sign on both sides. Similarly, if w11 < w12, you are going to get a (-) sign over here; if w22 < w12, you will get a (-) sign over here. However, if you have a scenario where w11 > w12 > w22 or the other way around, then this is not going to work.

You are not going to achieve an internal equilibrium. Now, great. So, we now know the conditions that will lead to these equilibria, but remember we are dealing with a quadratic form. In other words, our equation, or you know, graph is going to be a parabola. Now a parabola we know can either have a maximum or it can have a minimum. Now knowing this does not allow us to tell which one is right. So in which case we are going to get a maximum, and in which case we are going to get a minimum is not entirely clear.

So in order to figure that out, what do we need to do? We simply need to take a double differentiation, right? So we will start from this point, okay.

So we will differentiate this again, sorry, $d^2\bar{W}/dp^2$ =? We will take the 2 aside. $d^2\bar{W}/dp^2$ = (w11... This w12 will go; remember this is 2pw12. So, $d^2\bar{W}/dp^2$ = (w11 - 2w12... And this w22 will go means $d^2\bar{W}/dp^2 = (w11 - 2w12 + w22)$ So, this is my double differential. We know that when this double differential is negative, that is when we are going to get a maximum. and when this double differential is (+), that is when we are going to get a minima. This is what calculus tells us, right. When is this thing going to be? This thing is going to be when both w11 and w22, when both of them are going to be greater than w12. So, this case, right? So, when w11 and w22 are both greater than w12, we expect a minima, and similarly, when w11 and w22 are both less than w12, we expect a maxima, right? So, with this knowledge, let us now start plotting.

As I said, you can have three kinds of graphs over here, okay? Now in all these graphs, we have p, the allele frequency of allele a1 on the x-axis, and we have $\bar{W}$ on the y-axis. You can have a situation where the thing is either negative or positive throughout; the slope is negative or positive throughout. And in all those cases, irrespective of where you start your p, your $d\bar{W}/dp$ is always going to be (+) or always going to be (-). And therefore, p will either, you know, go in this direction and get fixed at 1, or it will go in this direction and get fixed at 0.

So, this is 0 and this is 1. If you have a situation that is like this, $w11, w22 > w12$, Then you will get a graph that will look like this. The minima will be somewhere in the middle. And in this situation where $w11, w22 > w12$, this means that the heterozygous form is less fit than both homozygous forms.

This is what is known as underdominance. This is the case where $w11, w22 > w12$. And this situation where the heterozygous has the maximum fitness compared to the two homozygous conditions. This is the case where you will get something like this, where you are going to get a maximum somewhere in the middle. Now, obviously, these points are the points at which $d\bar{W}/dp = 0$. In other words, these are the points at which p has stopped changing; these are equilibrium points.

However, there is a qualitative difference between the nature of these two equilibrium points. Why is that so? Let us see. Let us first start with this one. In this particular case, what is happening on the left side of the slope of this curve is (+) at all the points, right? So slope means $d\bar{W}/dp$.

Thus, $d\bar{W}/dp$ is positive to the left and greater than 0 to the left. and $d\bar{W}/dp < 0$ to the right. Therefore, what did I say about Δp? Δp will have the same sign as $d\bar{W}/dp$. So, on this side, $d\bar{W}/dp$ is (+), which means if the allele frequency is anywhere in this zone, it will go here. Whereas this side, it is (-), so if the allele frequency is anywhere in this zone, it will go in this direction. Now what happens when the allele frequency goes like this? This is $\bar{W}$, right? So, the average genotypic fitness of the population will go like this on this side, and on the other side, it will go like this.

Now what exactly happens here? As we said, this is an equilibrium point. But once you have reached this point, if you slightly perturb the system, if you slightly change the allele frequencies, It will come back to this, stating that this is, you know, a stable equilibrium point. I forgot to tell you that the situation in which the heterozygous condition is the most fit is what is called overdominance. These are technical terms: overdominance and underdominance.

Great. So, in other words, in this situation, the fitness is always increasing. Now, what is going to happen over here? In this case, again, this is an equilibrium point, but the slope of $d\bar{W}/dp$ is on the right-hand side. So if you slightly perturb the system, the system is going to go in this direction over here until it hits 1. Or if you start perturbing this on this side of the equilibrium point, it will go in this direction and it will hit 0. In both cases, what happens to the average fitness? The average fitness in both cases goes up, right? Even in this case, the average fitness is going to go up, okay. What happens here in this particular case? Here, if you see that this slope is positive over here, then any point where you start, the average fitness will go like this, right? Because Δp is always (+), p will go in this direction.

So it will go like this, and for the other one, this is (-), and therefore the average fitness will go like this. Δp will keep on reducing, and the average fitness will go like this. In other words, in all the graphs that we are seeing, we see that the average fitness of the population is always going to go up. And this is what leads to the thought process in the minds of most people that beauty is in selection. It is always going to increase the average fitness of the population; that is what selection does.

Note that that intuition is mathematically well supported, but it is supported only in this scenario. That we are talking about the 1 locus 2 allele scenario under the frequency-independent selection case. Now, as you can guess, if I am stressing this, harping on this so much, then obviously there will have to be a twist in the tail. What is a twist in the tail? The twist in the tail is related to this thing that we did over here. We said that because this is frequency independent, we are therefore dropping three terms.

What are the three terms that we dropped? So let us now go to that scenario. Let us go to the scenario where you have the frequency-dependent case: frequency-dependent genotypic fitness. Now we are looking at $d\bar{W}/dp$, and all the terms that we did over here, each one of these terms will all exist. And therefore, we are going to get the final thing that we got, which is $d\bar{W}/dp = 2(W1^* - W2^*)$. This part is constant; I mean the similar. However, on top of that, there will be three other terms. What are those three terms? Remember, now we are explicitly assuming that these three genotypic fitnesses are a function of p themselves. And therefore, the three terms are going to be $p^2dw_{11}dp + 2pqdw_{12}dp + q^2dw_{22}dp$.

This is the rule of differentiation. Now look at these three terms. What are they? These three terms are three differentials. The differentials of the corresponding genotypic fitnesses multiplied by their corresponding frequencies. Therefore, these three terms together form an average. An average of what? An average of the genotypic fitnesses; so this, or rather the average of the derivatives of the genotypic fitnesses.

So this can be written as $d\bar{W}/dp = 2(W1^* - W2^*) + \ldots$ Average in the statistical sense is also known as the expectation. $d\bar{W}/dp = 2(W1^* - W2^*) + E(dw/dp)$ Note this is not $d\bar{W}/dp$, $E(dw/dp)$, which essentially means the average of the derivatives of these three genotypic fitnesses.

So, if this be the case, then remember we need this stuff right. So then $(W1^* - W2^*) = ?$ $(1/2)$... goes this side, right. So $(W1^* - W2^*) = (1/2).((d\bar{W}/dp - E(dw/dp))$, okay, and this now we have to. You know, superimpose this stuff on this one. Okay, so yeah, this is where this gets superimposed.

So, $\Delta p = ?$ So let me write it properly: for the frequency-dependent selection $\Delta p = pq/2$, this part stays the same. $\Delta p = (pq/2\bar{W})\ .\ [(d\bar{W}/dp) - E(dw/dp)]$. Now here comes the first major problem. What is the problem? The problem is that it is not at all clear that these two terms will go to 0 at the same time.

In other words, when $dw/dp = 0$, it is absolutely not certain that $E(dw/dp)$ should also go to 0. In fact in many many cases it will not. Therefore, dw/dp going to 0 does not necessarily mean that your Δp has gone to 0. In other words, you have a situation where it is not entirely clear that you know you will definitely get an internal equilibrium. In many cases, what you actually end up getting is all kinds of complicated dynamics, including oscillatory dynamics. In other words, it does not settle into an internal equilibrium; the fitness does not always increase.

Sometimes it increases, sometimes it decreases. And under this situation, you can also have scenarios wherein Due to selection, the average fitness of the population will keep decreasing, finally leading to extinction. a phenomenon which is technically known as Darwinian extinction. Now just think about what we are saying. We are saying that when you have the 1-locus-2-allele selection case, then under frequency-independent selection, Sorry, frequency-independent genotypic fitnesses; selection can increase the average population fitness. However, even under the simplistic 1-locus-2-allele case, the moment the fitness itself, Genotypic fitness itself becomes dependent on frequencies; this is not going to happen.

And you can get a situation where selection leads to way more complicated dynamics. Oscillatory dynamics or even extinction, and this is with one locus. The moment you go to two loci or more than two loci, the interaction of selection actually becomes very complicated. This is why I wanted to show you guys this particular derivation: because if you talk to any normal person, Even biologists, not to mention non-scientists, even scientists, have this implicit notion in our heads. That selection will always improve the species; selection, or the population, will always increase fitness.

That entirely comes from, you know, this simplistic rights equation, which is this one. The moment you go out even slightly of that simplistic highly restrictive scenario, You are no longer guaranteed that selection will increase fitness. Selection can do many, many things. And people like us who actually work with selection and experimental evolution in our lab know. How complicated a thing selection can be! So, it is only under the simplest scenarios that selection improves fitness.

Under even slightly more complicated scenarios, selection does not really improve fitness all the time. It improves fitness many times, or at least some of the time, but definitely not all the time. So, we are going to stop here, and when we come back for our next discussion, We are going to ask what the implications of this are, and I am going to start that discussion with. Some of the insights that we are getting over here, and we will probably let me see if we have time. We will quickly do some Excel simulations to drive home some of the points that we are making.

And we will explicitly answer the three or four observations that we had last time. And we will explicitly answer why those observations are the way they are. In fact, if you just take today's discussion and go back to those observations, the answers should become very clear. But for the sake of those people who find the mass a little tough, I will end up discussing those explicitly in the next discussion. See you. Bye.

Please find the screenshots with the mathematical derivation explained in this lecture.

1-locus 2allele adaptive Landscape

A_1	A_2	
p	$1-p$	
A_1A_1	A_1A_2	A_2A_2
p^2	$2pq$	q^2
ω_{11}	ω_{12}	ω_{22}
$p^2\omega_{11}$	$2pq\omega_{12}$	$q^2\omega_{22}$
$\frac{p^2\omega_{11}}{\bar{\omega}}$	$\frac{2pq\omega_{12}}{\bar{\omega}}$	$\frac{q^2\omega_{22}}{\bar{\omega}}$

$$\bar{\omega} = p^2\omega_{11} + 2pq\omega_{12} + q^2\omega_{22}$$

$\textcircled{A_1}A_1 \quad A_1A_2$

$$\omega_1^* = p\omega_{11} + q\omega_{12}$$

$$\omega_2^* = p\omega_{12} + q\omega_{22}$$

$$\bar{\omega} = p(p\omega_{11} + q\omega_{12}) + q(p\omega_{12} + q\omega_{22})$$

$$\bar{\omega} = p\omega_1^* + q\omega_2^*$$

$$\bar{\omega} = p^2\omega_{11} + 2p(1-p)\omega_{12} + (1-p)^2\omega_{22}$$

$$= p^2\omega_{11} + 2p\omega_{12} - 2p^2\omega_{12} + \omega_{22} - 2p\omega_{22} + p^2\omega_{22}$$

$$\frac{d\bar{\omega}}{dp} = 2p\omega_{11} + 2\omega_{12} - 4p\omega_{12} - 2\omega_{22} + 2p\omega_{22}$$

$$= 2[p\omega_{11} + \omega_{12} - p\omega_{12} - p\omega_{12} - \omega_{22} + p\omega_{22}]$$

$$= 2[p\omega_{11} + q\omega_{12} - (p\omega_{12} + q\omega_{22})]$$

$$= 2[\omega_1^* - \omega_2^*]$$

$$p' = \frac{p^2 w_{11}}{\bar{w}} + \frac{pq w_{12}}{\bar{w}}$$

$$= \frac{p(p w_{11} + q w_{12})}{\bar{w}}$$

$$= \frac{p w_1^*}{\bar{w}}$$

$$\Delta p = p' - p$$

$$= \frac{p w_1^* - p\bar{w}}{\bar{w}}$$

$$= \frac{p(w_1^* - \bar{w})}{\bar{w}}$$

$$= \frac{p(w_1^* - p w_1^* - q w_2^*)}{\bar{w}}$$

$$= \frac{p}{\bar{w}}\left(w_1^*(1-p) - q w_2^*\right)$$

$$= \frac{pq}{\bar{w}}\left(w_1^* - w_2^*\right)$$

$$\boxed{\Delta p = \frac{pq}{2\bar{w}} \frac{d\bar{w}}{dp}}$$

Wright's eqn of Fitness / Adaptive landscape

Sign of Δp depends entirely on $\frac{d\bar{w}}{dp}$

Avg Fitness always increases

$\frac{pq}{2}$ = variance of probability of finding A_1 allele

∴ 3 components →
- Variance
- $\bar{w}$
- $d\bar{w}/dp$

$$\frac{d^2\bar{w}}{dp} = 2w_{11} - 4w_{12} + 2w_{22}$$

$$= 2\left[w_{11} + w_{22} - 2w_{12}\right]$$

This should be taught as For polymorphism, $w_1 \neq w_2$

This is going to be -ve when $w_{12} > w_{11}, w_{22}$

" " " " " +ve " $w_{12} <, w_{11}, w_{12}$