

## **An Introduction to Evolutionary Biology**

**Prof. Sutirth Dey**

**Biology Department, Population Biology Lab**

**Indian Institute of Science Education and Research (IISER) Pune**

**Week 5 Lecture 25**

### **Finite population: Numerical Simulations**

Hi, so while we were looking at the derivation of the Hardy-Weinberg principle, one of the preconditions. What we said was required for the Hardy-Weinberg principle to operate was an infinitely large population size. Now, if you think about it, in real life, we are never going to get an infinitely large population. So, in today's discussion, we are going to see what really happens when the population size becomes finite. And in order to see that, what we will do is go to the simulation, the biointeractive simulation. And check out what happens when you reduce the population size from infinity. So, you already have the link. Let us go there. Here is the thing: here is the simulation. So, we click on the start tool and there we go.

So, we will start with the individual simulations. We will set the population size to infinity to begin with, just to get the baseline case and the number of generations. We will set it to about 1000, and starting allelic frequencies, we will keep them at equal frequencies for both alleles. And we say, "Run simulation." And what do we see? We see that there is no change in the allelic frequency, which is expected because we have clicked infinity for the population size.

And therefore, we are just getting back to the Hardy-Weinberg principle, which is that allele frequencies will not change. So, now what we will do is explicitly set the population size at something. And to begin with, let us set it at the highest possible size that this particular simulation allows, which is 10,000. And we have done that, and we

say, "Run simulation." And what do we see? We see that the allele frequencies, so in this particular case, the frequency of A1 is bouncing around a bit.

And if you look at changes from generation to generation, they are not too high; some are little, but overall it is going. Somewhat up and down, and after some runs in this particular case, this is coming down to about 0.381. So, what will happen if I run it again? Let us do that; we say add a new simulation, and again we see. That there is some amount of bouncing around in this green curve, but this bouncing around is not the same as the.

Previous times bouncing around, and honestly, after about 1,000 generations, this is now ending at somewhere around 0.502. We add one more, and we see the same thing. In each case, each of the three cases that we did the bouncing around. It is of similar magnitude, small magnitude, but where it is ending up varies in all three cases.

This is happening because, in a deterministic case where you have an infinitely large population, the moment you have reduced the population size to some finite value; even if it is very large, the simulations have become stochastic. And therefore, each run is now going to be different from one another even if you start from the same starting population frequencies. So, in order to contrast this, what we will now do is we will reduce the population size and let us say we reduce it by 10 times. So, the population size has now become 1,000, and we run another simulation. What do we see now? We see that two things, A, in this particular case, if you look at how much things are changing in the small time frame, you know, going from one generation to another, the magnitude of those changes has somewhat gone up, right? And that is why the line looks more jagged. If you look at it again, there is a lot of bouncing around. In this particular case, it has ended up settling at a value that is quite a bit away from 0.5. So, earlier the thing was going somewhere between 0.4 and 0.6, but now it has gone all the way close to about 0.9. So, let us add one more over here, and you can see again that the magnitude of the fluctuation has gone up. And it has now gone very close to 0. In other words, what is happening is that overall the amount of bouncing around has increased the moment we reduced the

population size.

And let us now change this even further and We make it 100, an even smaller population size. What happens in this case? Boom, after about, you know, how many generations? About 69 or 70 generations, the allele frequency has gone to 0; the allele frequency of allele A1 has gone to 0, which means that The A2 allele is the only allele that is present in the population. So, this situation where there is only one allele in the population, That allele is said to have been fixed, and the other allele is said to have been lost. So, in this particular case, When the allele frequency of A1 is 0, we say that A1 is lost and A2 is fixed; therefore, A2 is fixed here. What happens if we run the simulation one more time? Is A2 going to be fixed all the time? Let us see.

We do it one more time, and this time what we see is that A1 has been fixed; A1 is, you know, a frequency equal to 1. Let us do it one more time; again, A2 is fixed. One more time; again, A2 is fixed. One more time; this time. At least within 1,000 generations, nobody got fixed, but note again that the magnitude of the variation going from Generation to generation, that has changed quite a lot; therefore, overall, the graph has become even more jagged.

So, we can keep doing this. This particular simulator gives you a very easy way to do it. So, you know, instead of running each simulation one at a time, it has this thing called replicated simulations. So, we now go to replicated simulations, we set the population size to 100 and the generations to 1000. And starting frequency, we keep the same, and we say that we are going to look at 10 populations at a time.

We say, "Run simulation," and boom! Straight away, it gives you 10 simulations, and you can see that out of the 10, In one case, there has been no fixation after 1000 generations, but in the other 9 cases, about 4 have been 1, 2, 3, and 4. been fixed for allele A2, and the other 4 have been fixed for allele A1. So, if we ask the question, which allele is getting fixed? Is A1 getting fixed, or is A2 getting fixed? Then, at least as of this moment, it looks like we do not really have a bias. Roughly speaking, they are being fixed equally. Is

that always going to be the case? Let us run the simulation one more time, and again what we see is that one of them does not get fixed, and you know, 1, 2, 3, and 4 of them get fixed over here, which means that about 5 of them get fixed for the other allele. So, again, this is a stochastic thing, but roughly speaking, as long as your starting frequencies are similar. There is not much of a bias in terms of which allele is getting fixed; either one of them can get fixed. Now, does it mean that this has something to do with the starting frequency? What exactly happens if I change the starting frequency? So, let us assume that I change the starting frequency of allele A1 from, say, 0.5 to 0.9. So, the allele A1 is the most frequent allele in the population now. And now, if I run the simulation, what do we see? In two cases, I see that A2 has been fixed, but in eight cases, I see that A1 has been fixed. In other words, the fraction of times that A1 is getting fixed has gone up the moment I increased the starting frequency. Now, of course, this is just one set of simulations of 10. What happens if I do it all over again? I do it all over again, and I see that in this particular case, only in one instance, allele A2 has been fixed. In the remaining 9 cases, it is allele A1 that has been fixed. So, in other words, when I increase the starting allele frequency, the probability of The fixation of the allele with the greater starting frequency goes up. If this is the case, you know if I reverse the situation or if I make the starting frequency. of allele A1 equal to 0.1 then if things are unbiased we should get the opposite result.

In other words, we expect that allele A2 will now get fixed with a higher probability. And let us see what happens when we run the simulation; that is exactly what we see. In all ten cases in this particular situation, it is allele A2 that has become fixed. Let us do it one more time: again 9:1, with 9 for allele A2 and 1 for allele A1. Run it one more time. Again 9:1, 9 for allele A2, 1 for allele A1. So, of course, you know that in order to convince yourself you have to Run this many, many times, and you can convince yourself by going to this particular page and doing these simulations. But you will definitely see that on average, the moment you are there, there is one allele which is Having a greater starting frequency, the chances of that allele getting fixed actually go up drastically. However, if the two alleles are starting with equal frequencies, then the chances of their getting fixed are roughly similar. In other words, the chances of getting an allele fixed in

the population are roughly equivalent or is roughly related to the probability or to the starting frequency of that allele in the population. So, these are some of the main things that I wanted to show you. Let us get back to our PowerPoint. So, what are the main observations that we made? So the major observations were that the smaller the population sizes, the greater the fluctuation in the allelic frequencies from generation to generation.

Now why exactly is this happening? If you remember, we said that when we were looking at the derivation of the Hardy-Weinberg, Not all the possible gametes in a population are going to give rise to the next generation. There has to be some kind of selection. Of a sampling, and this sampling will always lead to a difference between the parental generation and the offspring generation. Now, in order for us to negate this when we were deriving Hardy-Weinberg, we said that the sample size had to be very, very, very large. So large that even though it is a sample, the sample is completely representative of the population from which it is being taken.

But the only way in which you can have a very, very large sample is if your starting population itself is extremely large. In the theoretical sense, infinitely large, which is why we had explicitly assumed that our population size was infinitely large. However, the moment that does not happen, the moment the population size becomes finite, the sample size becomes limited. The sample of gametes that you are taking from the population of all gametes itself becomes finite. And whenever the sample size is finite, there is going to be some difference between the original parental generation's allele frequency and the offspring generation allele frequency. Now, the larger the sample size, the smaller is the effect. The smaller the sample size, the larger the effect, which is why when we had a population size of 10,000, which means the sample that we are using for the offspring generation is very, very large, then The departure from, you know, generation to generation, on average, is much, much lower. However, the moment we reduce the sample size and the population size, the sample size gets reduced, and once the The sample size is getting reduced; the sample is becoming less and less representative of the population from which it is drawn. As a result, you remember I told you that the change

from generation to generation itself is greater.

As the population size decreases, the graph appears more jagged. And when that happens, the population allele frequencies change drastically from generation to generation. Now, one thing you have to note here is that in any given generation, the allele frequency can go up or it can go down, right? The only point at which the going up or down stops is when one of the alleles has become fixed or has been lost from the population. That is because once the frequency has hit 1 or 0, That means there is no other allele at that point; the frequency cannot change any further. So, those two  $p = 1$  or  $p = 0$ , or similarly  $q = 0$  and  $q = 1$ , those two are the so-called fixed states of this system.

Once the population has reached there, then barring mutation or migration, there is no other way for a lost allele to come Back into the population. Therefore, that fixed stage is a stage of equilibrium. After that, the allele frequencies are Not going to bounce around any further. The third very interesting thing that we saw was the fixation probability. The effect of an allele depends upon its starting frequency.

Now here I have to give you a slight explanation. If you are talking about the fixation probability of an allele in a population, then how are you going to figure it out? Remember, in a given population, an allele will either get fixed or it will get lost. So, the only way for you to figure out that probability is to run multiple simulations under exactly the same conditions. Starting frequency, etcetera, and then see in what proportion of those populations your allele A1 is getting fixed or not.

So, that is exactly what we did. If you remember, we ran 10 simulations, all with, let us say, a frequency of  $A1=0.9$ , and we saw that Roughly in 9 out of 10 of those populations, allele A1 is getting fixed, and therefore, we said that the probability is  $9/10 = 0.9$ . Similarly, when we changed the peak to 0.1, we saw that roughly 1 out of 10 populations the allele A1 is getting fixed.

As we will show you in one of our future discussions, the fixation probability The

frequency of an allele in the population depends on, or rather is equal to, the starting allelic frequency for that allele. The fourth very interesting point that we saw was whether we started multiple populations. replicate populations, as we were calling them, each with exactly the same starting allelic frequency. And then we just allowed them to run; then, due to this sampling error alone, the frequencies. Of these alleles, the alleles in these populations ended up diverging from each other. In other words, in terms of the genetic composition as well as in terms of the genotypic composition, they diverged from each other.

Now compare this with, let us say, the effect of selection. If you started all the populations with the same starting allelic frequency and the same genetic genotypic fitness, in all the cases, the frequencies would have gone in the same direction; they would have converged into the same outcome. Whereas in this particular case, when we are only looking at the effect of this sampling error, we see that they are diverging. And this is something to keep in mind: this sampling error can act like a very nice. An interesting evolutionary force that ends up making populations different from each other.

And this evolutionary force is what we call random genetic drift, also known as genetic drift, or simply drift. also called Wright's effect in the literature. So, many names, but all of them refer to the same phenomenon that When there is a change in allele frequency across generations due to random chance without any relationship, you know. With the fitness of the population, here are a few points that I would like to make in this context. First, any finite population, as long as the population size is finite, however large it is, will undergo some amount of drift.

Now, very often we think that large populations know drift; that is not really true. The effect of drift might be very small. In large populations, and that is why we might end up saying that we will neglect it, but that is always an approximation. In reality, any population, no matter how large, is facing drift.

Secondly, drift is a major evolutionary force. The strength of it entirely depends on the

size of the population. When the population size is large, it is a weak evolutionary force. But when the population size is small, it can become a very, very strong evolutionary force. And when I say strong, I mean that. It can lead to very significant effects, and it can lead to those significant effects very, very quickly in terms of how many generations it takes.

And even in the case when it is small, drift actually ends up interacting with other forces like selection and mutation. Very interesting ways lead to very rich patterns of, you know, evolutionary outcomes, some of which we are going to look at. In our subsequent discussions, this is again a very crucial point: drift can and does lead to changes. In the allele frequency, in the genotypic frequencies, and in that sense, it is an evolutionary force. However, the changes that it leads to are agnostic to whether it is increasing.

The fitness of the population or whether it is decreasing. And therefore, drift cannot explain adaptation. Now, this is very crucial to keep in mind. Due to drift, sometimes the population's fitness will go up, and sometimes it might even go down. That is not the point, but there is no guarantee that it is going to inevitably lead to. Increase in fitness, and in that sense, we cannot say that drift will lead to adaptation.

The only force that can lead to adaptation, under you know, specific circumstances, of course, is selection, right? And in that context, there has been some debate, you know, in the literature about what is a more important evolutionary force. Is it drift or is it selection? And the simple answer to the question is that it depends on the context. There are certain contexts in which drift becomes very powerful; there are certain other contexts in which Selection becomes very powerful, and there are several contexts in which drift and selection interact with each other. leading to interesting outcomes. So, some of those outcomes and a few other features of drift And of course, proper biological examples of drift are all that we are going to look at in our next discussion. See you then. Bye.