An Introduction to Evolutionary Biology

Prof. Sutirth Dey

Biology Department, Population Biology Lab

Indian Institute of Science Education and Research (IISER) Pune

Week 3 Lecture 12

Revising Mendel and introducing population genetics

Hi, so before we go ahead with today's discussion, we need to refresh our memories about a few terms and concepts. So, the first major term is that of the genes. Now, nowadays, when anyone thinks about the gene, what they think about is a stretch of DNA. which is transcribed and either codes for RNA or codes for some kind of polypeptide, which either has a known function or a presumed function. Now, this is a very mechanistic way of looking at a gene, and although this is what it is, you know, Mostly molecular biologists prefer this; this is not how evolutionary biologists typically think of a gene. Or, for that matter, even the transmission genetics people do not think about genes this way.

The way they think about it is what we call a Mendelian gene. which is a genetic unit that is transmitted from parents to offspring This is detected through phenotypic differences associated with different alleles at the same locus. In other words, if you have multiple forms of that gene, Then they are defined in terms of having their effects on the phenotype, not simply having a difference in the sequence. Now, in this particular course, we are going to use the term "gene", both as a unit of inheritance and as a mechanistic unit of some kind of transcription leading to a protein or RNA. So, how will you know in which way I am using it? So, whenever I say "gene" in general, I will be talking about a Mendelian gene. But suppose I tell you the names of specific genes, for example, the serotonin transporter gene that we discussed in the last discussion. When I give a specific

name, that is when we talk about a molecular gene. But when I say gene in general, it is the Mendelian gene in the context of a transmission unit; that is what we are going to talk about.

So, this definition has two more terms: alleles and locus. So, what are they? So, an allele is simply one of the several alternative forms of a gene that is present in a population. So, for example, in our last discussion, we talked about two alternative forms of the serotonin transporter gene. One was the L form and the other was the S form. So, these two are alleles.

And what is a locus? So, you know that the genes are arranged on the chromosomes and have specific positions on the chromosomes. And these specific positions of a gene or a genetic marker are what is known as a locus. Two other important terms are genotype and phenotype. What is a genotype? Simply the genetic makeup of the organism for a particular locus or for multiple loci. So, for example, in the context of Mendel.

Remember, he said that he started with true-breeding tall plants and true-breeding short plants. And he said that the true breeding tall plant has the TT allele at both positions on the homologous chromosome. So, at the same locus, it has two copies of the T allele. So, this thing, this TT or tt, is what indicates the genetic makeup of the organism at that particular locus. Hence, it is the genotype of the pea plant.

Now, what is the phenotype? The phenotype is simply an observable trait or characteristic of an individual or organism. So, for example, in the context of height in the Mendel example, tallness is a phenotype and shortness is a phenotype, right? Now if you remember, we said that the rediscovery of Mendel's laws was exceedingly important for evolutionary biology. So before we go ahead, we need to quickly have a look at what exactly Mendel said. So if you look at Mendel's three laws, the first law is the law of dominance. which says that in a heterozygote, the phenotypic effect of the recessive allele is masked.

In other words, if you have, let us say, the TT plants being tall and the tt plants being short, Then, when you mate the TT plants with the tt plants, in the next generation, the F1 generation, The genetic makeup of all the individuals will be Tt. So, they are going to be like this: the heterozygous form, yet phenotypically, all of them are going to be tall. In other words, the effect of the t allele will be completely masked by the T allele. Now remember, this is what was happening at the time of Mendel. Today, we know that this is not always the case.

So many times when you have a scenario of a heterozygote like this, The phenotype of the heterozygote is actually intermediate. So nowadays, the specific thing that Mendel saw, we call complete dominance. When the dominant form is able to completely mask the phenotype of the recessive form. However, nowadays we know that you have different degrees of incomplete dominance. And in all those cases, the phenotype of the heterozygote is actually some kind of intermediate between the phenotype of the dominant homozygote and that of the recessive homozygote. So, this is the Law of Dominance. Now you have the Law of Segregation. So suppose we are talking about a one-locus scenario, a diploid organism. So the organism has two copies of that particular gene.

Now, what Mendel says is that during the process of gametogenesis, That is when the sperm cells or the ova cells are formed. Then the cells have to undergo what is known as a reduction division, right? So, each gamete actually has only 50 percent of the genes that the original organism has. And in the process of forming the gamete, Each gamete will get only one copy of the gene, whereas the parent would have had two copies, right? So, this is the Law of Segregation. And finally, suppose you are talking about multiple loci; then you have, you know, genes at these loci and alleles at these loci. And what Mendel said was that during the process of gametogenesis, The alleles at different loci are going to assort independently, which means that they are going to go or divide themselves into the gametes irrespective of what allele is present at the other locus. So, the allele at any one locus is going to get into the gametes without any effect from what allele is present at the other loci. Now, today we know that this is true only when the loci are on different

chromosomes. or the loci are so far away from each other on the chromosome, Let us say that a crossing over event is highly likely at two ends of the chromosome. If that does not happen, if the loci are on the same chromosome and are relatively close to each other, We get a different phenomenon called linkage.

However, Mendel did not talk about it. Mendel explicitly looked at those loci that were either on different chromosomes or were very far from each other. So, this is where our revision ends, and now the discussion on this particular topic begins. Now, if you look at all that I talked about, Mendel's work is in the context of the transmission of genes within a family. Why am I saying this? Remember that what Mendel was doing was taking true breeding varieties.

And then, you know, he was taking other true-breeding varieties. He was crossing them and trying to see what was happening in the next generation. And then he would, you know, self-cross them, and he would look at what was happening in the next generation. So, in all the cases, he has full control over who is mating with whom. And he is simply looking at what the offspring look like, right? However, if you think about a population—a real biological population— We do not really know, nor do we really have any control over who is mating with whom, right? So, if you are thinking about looking at evolution from the perspective of changing genetic variations, Remember that is what I said: you can think about evolution as how variation changes across generations.

So, if you are trying to do that, then it is not very difficult conceptually. Conceptually, what do you have to do? Conceptually, you have to go out there and measure every single individual and look at their genotypes. And then you have to look at all their babies, and in the next generation, look at what their genotypes are. And you can simply compute what the change in genotypic gene frequencies is. Now, of course, empirically, this is going to be very expensive.

This is going to be very troublesome, but conceptually there is not much of an issue. However, when you try to think about what is happening under the hood, Then you

realize that, biologically speaking, this is not a straightforward process as it looks. Why is that so? Because underlying what is happening is something like this. You have the male; I am talking about sexually reproducing diploid organisms. So, you have the male individuals, you have the female individuals; the male individuals are undergoing meiosis.

So, during the process of meiosis, two things happen. A crossing over is happening, because of which the chromosomes are all getting jumbled up. The homologous chromosomes are all getting jumbled up. And then B. reduction division is happening because each gamete is getting only 50% of the genome that the original organism had, the parent organism had And that stuff that it is getting, that genetic composition, is actually all chopped up thanks to crossing over. This is happening in the males; this is also happening in the females, right? So, now you have lots of gametes, but the organisms are mating. And they are mating, you know, with whomever they please. And even when they are mating, not all the gametes are forming the zygote. Only a subset of the gametes is forming the zygotes.

And then this subset of the zygote, A small fraction of that is actually, you know, being born and then growing up to form the next generation. And you sum over all the mating pairs, and only then do you get the next generation, right? So, biologically speaking, this process is a lot more complicated and therefore very difficult to track. If you are going to track family by family, So, what you need to do is come up with a way for this entire process that is happening here. At the level of the family, this entire process can be, you know, somehow abstracted out. such that you can go directly to the population level from generation t to t+1.

And the branch of evolutionary biology that does that, which takes you from generation t to generation t+1 at the population level. Somehow taking care of all the nitty-gritty, gory details of the biology at the lower level, That branch of evolution is what is known as population genetics. Simply speaking, it extends the concepts that you get from Mendel's laws from the family all the way to the population. Now, how does one do that? How

does the subject do that? Now, in order to do that, what it does is create certain abstractions, and these are very, very powerful abstractions. Conceptually, they look trivial, but once you see the kind of insights they can give you, You realize that you know it is an absolute marvel.

So, what is the nature of the abstraction? So, first we will talk about the nature of the abstraction, and then in the subsequent discussions, We are going to see what comes of it. So, what is the nature of the abstraction? So, suppose we are talking about a one-locus scenario, right? Now, if you are talking about the organism, an organism obviously does not have just one locus. It has many, many loci and many, many genes, all interacting with each other in a very complicated way. But population genetics says that when you are thinking about a one-locus scenario, Then just think about that one locus; forget about everything else, right? Now, if you have a diploid, sexually reproducing organism, then in that population, If you are thinking of each individual and concentrating only on that locus, So you know that for each locus you will have two homologous copies of the gene, right? So, you think of each individual as some kind of pair like this, right? And here, let us say, you know we have two alleles; we can have more. It does not really matter, but for the sake of discussion, let us assume that we have two alleles.

I am representing one as shaded and the other as white. So, these are the various genotypes that we are going to have. Of course, I am showing you just three individuals, but there can be many more such individuals. So, what is the implication of this? So, if you think of the organism like this, this is the parent, let us say, then what will the gametes look like? Remember, the gametes are simply coming into being because of reduction division during meiosis. So, according to Mendel's laws, the gametes are just these genes separating from each other.

And some of them in this particular case will be dark because we have two alleles, right? So, the gametes are simply the genes, these genes. Now, you have this happening for the males and this happening for the females, correct? So, you have all the gametes produced by the males; you have all the gametes produced by the females. Some of these will be

dark like this one. Now, at the level of biology, mating is happening, okay. But at the level of this particular locus, how do you conceptualize mating? At the level of this particular locus, mating is simply taking one gamete from the male pool which is this one: taking one gamete from the male pool and uniting it with one gamete from the female pool, right? So, when you are looking at the next generation, essentially all you have to do is To pick up one gamete from here and one gamete from there and join them to form the next generation, right? Something like this, it all goes back. So, this is the offspring generation. Now, why should I think about it this way? What benefit do I receive? The benefit I get is that if you think about the whole process like this, Then mathematically, all these things are going from parents to the gametes. Going from the male gamete and female gamete to the next generation's offspring, all these essentially become processes in sampling. You are simply sampling some gametes from the overall pool of gametes.

You are simply sampling some male gametes and female gametes and putting them together. And the moment you enter the domain of sampling, Then we have the entire mathematical tools of the science of probability, you know, available to us. And because we know a lot about how probabilities work, We can actually deal with all these things using the mathematics of probability. So, this is precisely what population genetics does. It takes a problem that is in the domain of biology, and by making these assumptions, Making these, you know, mental constructs actually translates them into stuff in the domain of probability.

And, of course, it is much easier to handle them in the domain of probability. So, the entire process of the transmission of genes over a large number of mating pairs in the population, that just gets abstracted down into sampling from a bag full of gametes. Now, at some level, many people find this somewhat discomforting because, you know, biology is complicated. So many genes, you know, behavior, and everything is determining who is mating with whom: behavior, culture, etc. And yet, at the level of the locus, you are just putting all those complications under the carpet, and you are just focusing on sampling.

And that is why some people very derisively call population genetics "beanbag genetics." It is simply sampling beans from a beanbag. But wait until you see how powerful this conceptualization is and the kind of things. The kind of results that it leads to, which can then be tested both empirically in the lab and in the field. So, a lot of what we are going to do over the next 2 to 3 weeks is going to be in this framework.

Trying to first derive what the prediction is and then explicitly checking it using simulations. and explicitly looking at real-life data to see whether that prediction holds or not. But before we get there, there is a very simple result that we need to derive. It is a very simple result, but again, it is a very powerful and important result. So, let us assume that we have a one-locus, two-allele scenario and that the locus is autosomal.

It is a sexually reproducing organism, and the generations are discrete, which means that parents and offspring never mix. Now, let us further assume that the two alleles we have in hand are called A1 and A2, respectively. And the frequencies of A1 and A2 are p and q. Let us then understand that if you have two alleles A1 and A2, then you are going to have three genotypes: A1A1, A1A2, and A2A2. And let us assume that the frequencies of these genotypes The proportions of these genotypes in the population are represented by P, Q, and R, as I am showing you.

Great. So, if that is the case, then certain things are very obvious. What is obvious? We have just three genotypes in the population, right? So, the sum of the frequencies of the three genotypes, that is $P + Q + R$, has to be equal to 1. There is nothing else that can be, right? Similarly, if you have two alleles, A1 and A2, then the frequencies of these alleles have to sum to 1; $p + q$ is equal to 1. So, this is kind of obvious. What is a lot less obvious is how we are getting to this result that $p = P + Q/2$ and $q = R + Q/2$.

Now, there are many people who think in terms of ratios, proportions, and simple algebra. They will take one look at this relationship and say, "Yeah, it is very intuitive." However, there are many more people, and I am one of those for whom it is not really obvious at

first glance how p = P + Q/2. So, what I will do is derive it very easily for you, and that hopefully will make things very clear. So, let us assume that the number of individuals of the A1A1 type in the population is N11.

The number of individuals of the A1A2 type is N12, and the number of individuals of the A2A2 type is N22, right? So, if that be the case, then the total number of individuals, we will call it as N is equal to sum of these three, right? So, N11 + N12 + N22, right? Now, we want to figure out what the frequency of the A1 gene is. We want to figure out what p is. Now, this p is equal to the total number of A1 allele alleles divided by the total number of alleles. Now, what is the total number of A1 alleles? Remember that the A2A2 folks do not have any A1 allele, right? So we can disregard them. The A1A1 folks and the A1A2 folks are the only ones that have the A1 allele, but they have it in different numbers.

So, each A1A1 individual has two copies of the A1 allele, right? And we have N11 number of A1 individuals. Therefore, the contribution of the A1A1 homozygote is 2 * N11, correct? And what is the contribution of the A1A2 heterozygotes? Each one of them is giving me one copy of the A1 allele, and there are N12 individuals. So, the total that we are getting here is 2 * N11 + N12. What is the total number of alleles? Remember, we are talking about diploid organisms. So, each one of them has two copies of the alleles of any allele, and therefore, the total number of alleles is 2N.

Now, this thing I can write as (2N11/2N) + (N12/2N), correct? So, these 2 and these 2 are cancelled out. So, this becomes (N11/N) + (N12/2N). Now, here is what (N11/N)? N11 is the number of A1A1 individuals, and N is the total number of individuals. So, this (N11/N) represents the frequency of the A1A1 individuals in the population.

In other words, this term is equal to P. Similarly, here, if I just look at this part (N12/n), there are N12 individuals of A1A2, and N is the total number. So, this is the frequency of the heterozygotes (N12/N), and this is Q, and I have this extra 2. So, this is P + Q/2, right? This is what we derived over here. It is very simple once you think about it, but

this is actually a very important result. And as we will see in the next discussion and the one after that, many times students end up ignoring this particular result.

And therefore, they end up getting completely, you know, wrong answers to certain kinds of questions. We are going to deal with those questions explicitly. But what I want you to do is satisfy yourself using the same logic that we used here, that q is indeed equal to R + Q/2. So, with this result, we are now in a position to extend Mendel's laws to the entire population level. But that is what we are going to do in our next discussion. See you there. Bye.