Introduction to Evolutionary Dynamics
Prof. Supreet Saini
Department of Chemical Engineering
Indian Institute of Technology, Bombay

Lecture - 09
Modelling Mutations - 1

Hi, and welcome to the next lecture. Of the course I am today having completed our discussion on reproduction and selection, what we are going to talk about is start our discussion on mutations. Now in bacteria there are many modes. Mutation in general refers to any process which leads to change in the genotype associated with the organism, and in bacteria there are multiple mechanisms by which that changed can happen.

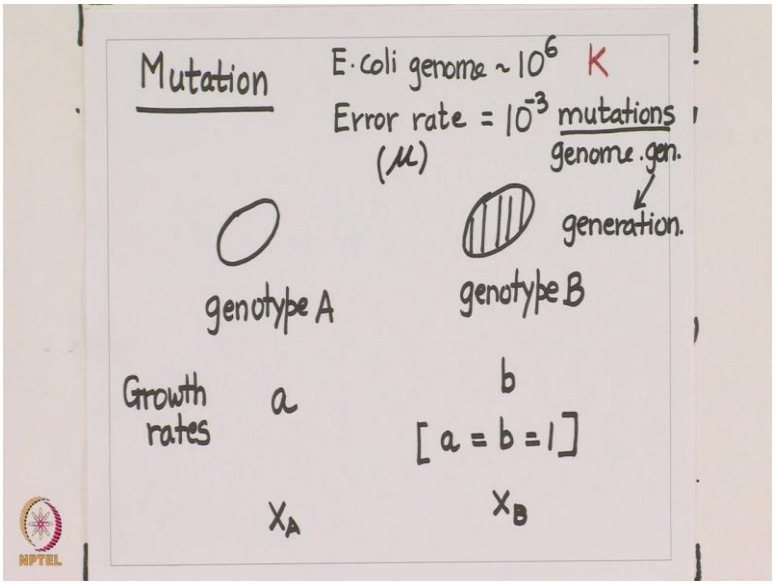
One of the mechanism the simplest one is when the machinery that is responsible for copying DNA to be divided to be given. So, that one copy can be given to the progeny and one retained with the parent makes an error while make while it is copying the parent DNA. There is a natural error rate associated with the DNA polymerase, which is the machine which is doing the copying of DNA is make makes at a natural error rate in every organism, that we know. And that error rate is different depending on which particular species where talking about.

In vision we can have events such as duplication of parts of DNA or certain regions of DNA might just get deleted from the organism while the DNA is being copied. And in bacteria of course, there is a process called horizontal gene transfer where transfer of DNA is not just happening from one generation to the next one which is a vertical mode of DNA, but. In fact, bacterial species are continuously exchanging DNA with each other they pickup DNA fragments from the environment and try and integrate into their genomes. So, there is a lot of horizontal gene transfer where not just exchange of DNA is taking place from a parent to a progeny, but also between 2 bacterial cells which are in existence independently of each other.

So, for our discussion for today we are going to rely on the most simplest form of mutation that can occur which are called point mutations, where change in a nuclear tight associated with one genotype leads me to approach any being produced which of which

is of another genotype and so again as always let us let us start our discussion with 2 distinct genotypes and let us call them a and b.

(Refer Slide Time: 02:42)



So these are mutation, and let us say we have genotype A and genotype B. Now what we are going to imagine is the natural growth rates associated with these 2 genotypes. So, the growth rates at equal to a and b. And in the most general sense a and b are going to be numbers which are unequal one particular type is going to go faster than the other one, but as we have seen from our discussion in the last lecture, when we have unequal growth rates the one that is going faster will take over the entire population and the one which is going slower will get eliminated from the environment.

So, for a discussion on mutations we understand the behavior of mutations occurring in these 2 specie model better, we are going to assume that both our genotypes are going at the same rate and to make things even more simpler we are going to assume that not only is a equal to b, but a is equal to b is equal to just 1. So, we are going to assume that a is equal to b is equal to 1. This is just so that we understand the process of mutations taking place better. And a is equal to b helps us eliminates selection from our discussion and only focus on changes in frequency is which are taking place only because of mutations happening in the environment.

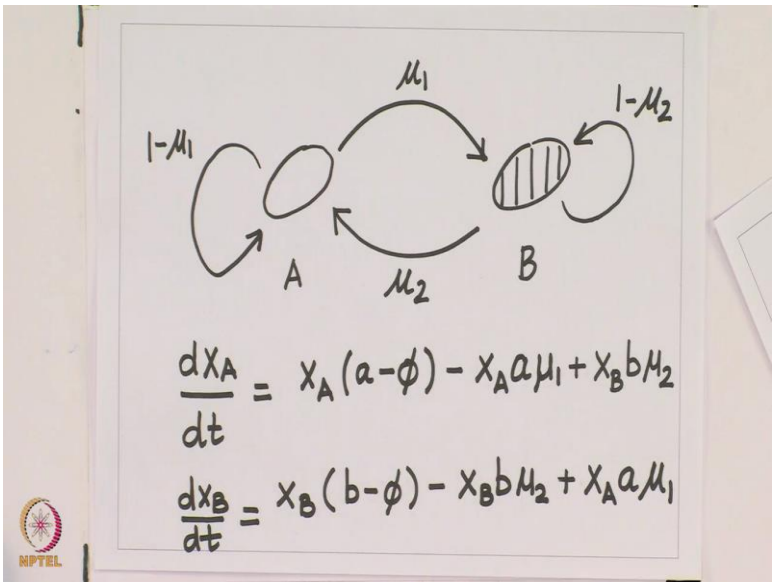
So, now again we have an environment where the total number of individuals is constrained to be equal to K. So, carrying capacity K and the fractions of individuals of

each type are going to be referred to as XA and Xb. So, XA is the fraction of the population which belongs to genotype A and XB is the fraction of population which belongs to genotype B. And we are interested in monitoring that how do these fractions change with time if there are mutations taking place where genotype A gets converted to genotype B and vice versa. Because of the errors in the; I mean replicating machinery associated with the growth of these organisms. So, we had said that every organism has a natural error rate associated with it is DNA copying machine which is DNA polymerase. Just to give you a sense of the numbers which are involved, equalize the common bacteria that we know the most about has a genome of the size roughly about of the order of 10 to the power 6.

So, E-Coli genome is of the order 10 to the power 6, and the error rate of E-Coli DNA polymerase the rate at which it makes error while it is copying gets DNA to be given to the progeny of the bacterium dividing is equal to 10 to the power minus 3 mutations per genome per generation. This word that this enwrited a full generation. What does this number mean error rate is typically represented by the variable mu in literature, what does this number 10 power minus 3 mutations per genome per generation actually mean. What this means is that when the bacterial DNA polymerase is copying the entire 10 power 6 bases of the equalize genome, there is a 10 power minus 3 chance or there is a chance equal to 0.001 that and error is made at one location while the DNA was being copied.

So, one whole genome 10 power 6 basis copied there is a chance of 0.001 that there was one error while the copying was being done. Which means if you have a thousand equalized cells dividing which is 10 power 3 cells dividing then there is a chance that one of the thousand progenies resulting from these thousand division processes has one error inside them the other 999 have the exact same faithful replication of the DNA that was contained in their parents. So, DNA polymerase are extremely fit well they do not make errors very often and as we will see later on in the course the accuracy of DNA polymerase is typically increases as the genome size associated with the organism increases. So, our DNA polymerase is actually more accurate than the one associated with E-Coli, which itself is working at a very high accuracy.

(Refer Slide Time: 08:36)



So, now that we have a sense of these numbers let us go back to our 2 genotypes and try and understand what is going on? So, we have genotype A and genotype B. When A is replicating most of the time it produces a progeny which is also of genotype A right. So, this arrow represents that when A divided the progeny was also of genotype A, but in our 2 specie model whenever an error is being made we will say that the progeny is of type B and this represents the rate at which error is happening while this replication is going on. So, we will say that the rate at which a might lead to a production of progeny of type B is mu 1; that means, the rate the probability that division of genotype A leads to progeny of another genotype A is 1 minus mu 1.

So, one division event happening there is a mu 1 chance that the progeny is of genotype Bm and there is 1 minus mu 1 chance that the progeny is of genotype A, looking at the E-Coli numbers that we just did you should try and convince yourself at this point that mu 1 is going to be a very small number and more often than not division of progeny A is going to lead to another a type A genotype bacteria more often than not.

Similarly, let us assume that b one replicates leads to another bacterium of genotype B most of the time, but every time an error happens genotype B leads to approach any belonging to genotype A and this error is happening at an unequal rate mu 2. We can put mu 2 equal to mu 1 and that. In fact, makes our analysis easier, but we want to take the most general case mu not equal to mu 1, and mu 2 is the error which genotype B makes

while replicating. So, when genotype B is replicating and it leads to a progeny of genotype B that is happening with probability 1 minus mu 2. So, that is that is the setting that we have. And now we want to understand how can be dynamically capture the frequency is associated with these 2 genotypes as this process moves forward in time and what are the steady state steady state fractions of populations belonging to genotype A and genotype B.

So, let us try and write down the dynamical equations associated with these processes and we will start with genotype A. So, just as we had done last time dXA by dt can be written as XA times a minus 5, and you should remember the significance of the variable 5 from a 2 lectures pack that phi is the constraint that we have placed on the growth rate which ensures that the total number of bacteria in the environment does not exceed K. The exact value or the expression of phi is unknown at this point, but we are going to use the fact that dx K by dt plus dXb by dt is equal to 0, because the total number of individuals in the population is constraint 2 K, and that constraint is going to give us the value or the expression associated with the quantity phi that is being defined at this point. So, that is a variable phi this is a natural growth, but then we have this happening that some individuals which divide and belonging to genotype A lead to actually production of genotype B and not A.

So, we subtract that amount which is XA times a times mu 1, mu 1. So, this is this is over counting that has happened in this expression, because some of these replication events do not lead to production of genotype A, but actually it to production of genotype B. So, those have to be subtracted, but at the same time while this genotype is replicating some genotypes are also produced which have not been taken into account so far. So, to do that we have X b in to b into mu 2 right. This represents the total production of genotypes A which are happening without taking into account mutation, but in this expression we have over counted because we have also included the ones which actually lead to genotype B and not genotype A.

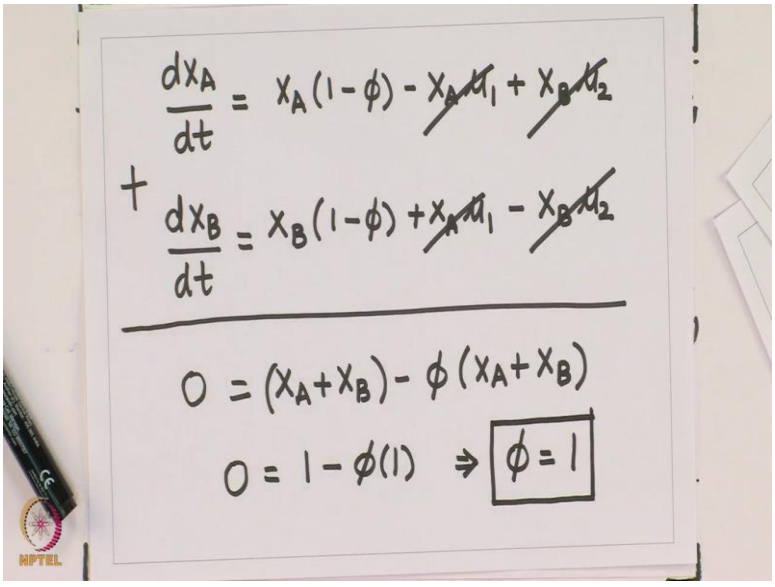
So, we subtract them from the expression that we have, but we have under counted the ones the genotypes A is which are being produced because of division by genotypes B and the error rate mu 2. So, we incorporate that. So, this gives me a dynamical equation for dxk by dt. Similarly, dXb by dt can be written as XB into b minus phi minus XB b mu 2 this is because we have complete all divisions that are happening with genotype B

here, but we have to subtract those divisions which actually lead to production of genotype A, which is given by this arrow we subtract them, but with this we have to add all those divisions which are happening with genotype A, but lead to progeny of genotype B, which is this expression.

And that is XA times a times mu 1. So, we will note that what has been subtracted from the dxk by dt equation is exactly the same as what as been added to the dx by dXb by dt equation, and which makes sense because these are these are divisions, which if the DNA polymerase was not making an error should have led to production of genotypes A individuals, but is actually leading to production of individuals which are carrying genotype B. So, that is our dynamical equations which represent what is going on here.

Now, the first thing that we have to do is find appropriate expression for the quantity phi, but we had initiated this discussion with the assumption that we are not going to bring selection into the equation at this point in the lecture today, because we want to understand mutations better hence we are going to put a equal to b equal to 1 in this equation and that is simplifies our equations a little bit.

(Refer Slide Time: 15:49)



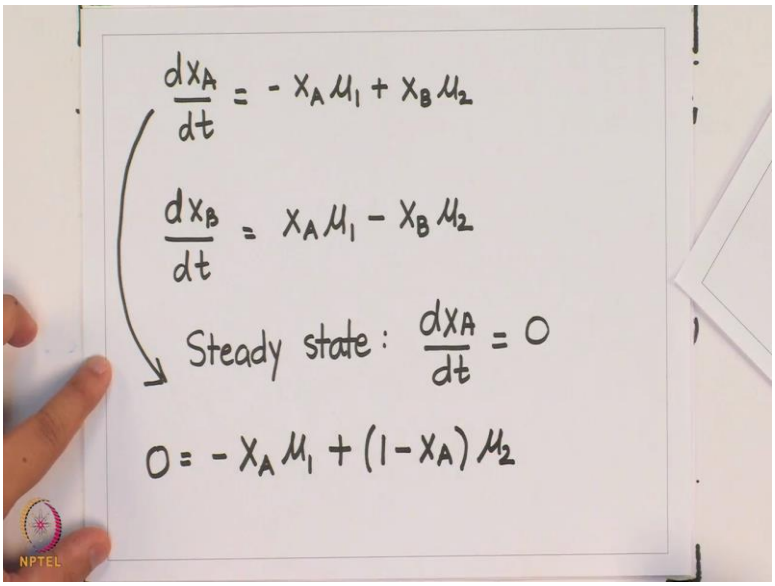
And let us let us rewrite them again with a equal to b equal to 1 we get dxk by bt equals XA 1 minus phi minus X K into mu 1, plus X b into mu 2 and dx b by dt is equal to X b 1 minus phi plus XA into mu 1 minus XB into mu 2. And now I am going to use the constraint that a total number of individuals in the population is equal to K and hence

that gives me the relationship that is sum of derivatives dx K by dt plus dXb by dt is equal to 0. And when I do that I just have to add these to get that I add them, but I know that this sum is equal to 0. This and this, then when I add these expressions is to cancel when I add these 2 is to cancel and I end up with XA plus ex minus phi times X K plus b which, but I know that xk plus XB which is the fraction of individuals fraction of the population belonging to genotype A plus fraction of population belonging to genotype B, this sum of fractions is equal to 1 and that gives me 0 equal to 1 minus phi times 1 which implies phi is just equal to 1.

At this point you should go back to the lecture where we had defined the expression of phi and realize the biological implication of the variable phi that we had talked about. We had said that phi represents at any particular composition of the population the mean fitness of the 2 genotypes that are present in that population. Which is exactly what is it is happening here it is just that phi is a constant because the mean fitness of both the genotypes that we have in this population is equal to 1.

So, irrespective of composition of the genotype the mean fitness associated with the environment e that we are talking about is going to be equal to 1 and hence this math naturally leads us to the relationships that phi is equal to 1. So, phi is equal to 1 that simplifies our equations a lot because when we plug phi equal to 1 in these equations 1 minus phi this term becomes 0 1 minus phi this term becomes equal to 0. So, what do our dynamical equations reduced to let us rewrite them?

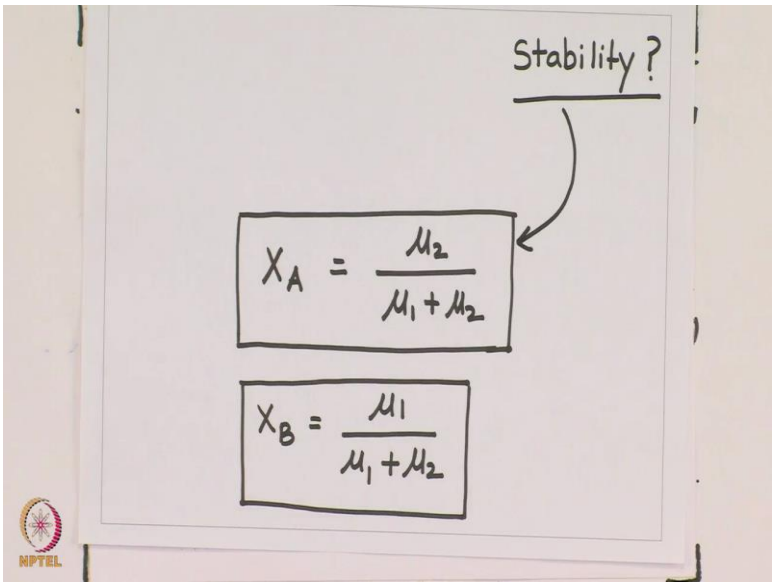
(Refer Slide Time: 18:56)



So, now we have dx K by dt equals minus xk mu 1 plus XB mu 2. And dXb by dt is equal to XA mu 1 minus X b mu 2. Both the equations are selected I am only going to work with this one. So, now, if you want to understand the steady state, the state or states of this particular equation, will do what we have been doing. So, far put the derivative dx a by dt equal to 0. And find out the values of X here for which this relationship holds.

And to do that I work with the right hand side of the equation and say 0 equals minus XA mu plus when it comes to XB I should write 1 minus XA, because XA plus X b is equal to 1 times mu 2. Just to wait on the next slide that gives me the relationship at steady state when dXA by dt is equal to 0, this gives me to solve this in terms of XA, I am working with this equation. Now that gives me the relationship that XA at steady state is just equal to mu 2 divided by mu 1 plus mu 2.

(Refer Slide Time: 20:29)



You can simplify this relationship to arrive at this just simple manipulation and that is the steady state fraction of the population which belongs to genotype A. Similarly, you can work with the second equation that we have dXb by dt and analyze it steady state what that will take you to is XB is equal to mu 1 divided by mu 1 plus mu 2.

Once you arrive at these relationships one thing that you should convince yourself is that XA plus X b in the relationships that we derived is equal to 1 which makes sense again because sum of frequencies of the 2 distinct genotypes is equal to 1. That is well and good. Another thing you should look at it when we look at these expressions we have XA the denominator and both is the same, but the numerators are different. The fraction of individuals belonging to genotype A XA is actually dependent on mu 2 in the numerator and mu 2, if we go back a couple of slides represents at what rate is genotype be making an error when it is replicated right. So, higher the rate which genotype B is making violet us replicating higher is the frequency associated with individuals which are which belong to genotype A.

Similarly, the frequency of individuals belonging to genotype B is actually dependent in the numerator on mu 1 which is the error that is being made by genotype A, individuals when they are copying their DNA and eventually ending up as genotype B. So, this again makes a lot of physiological sense. Suppose you have mu 1 equal to 0; that means, there is no error that is being made by genotype A which means that genotype A will never

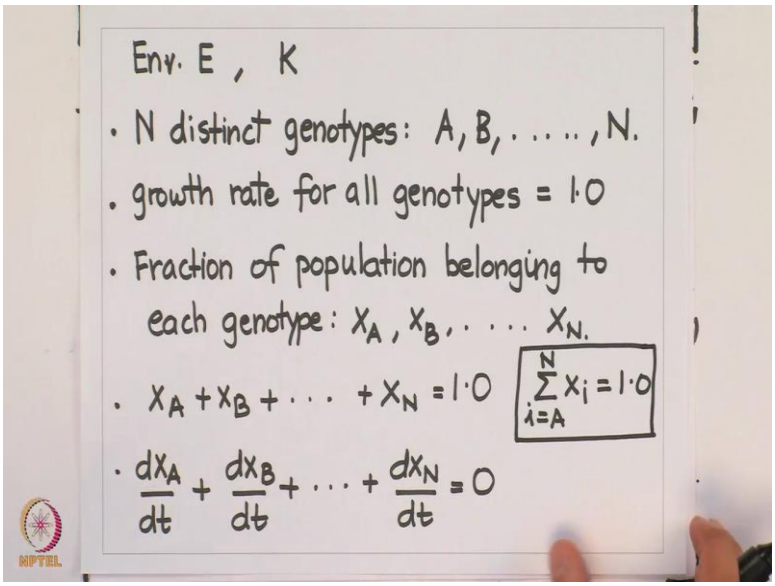
give rise to genotype B. So, this error term becomes 0, but mu 2 is not equal to 0; that means, genotype B is replicating and every once in a while it makes the error and resulting in genotype A.

In the scenario it is easy to imagine that this genotype will continuously keep rising to genotype A whereas, this arrow is equal to 0. And hence eventually all individuals will belong to genotype A, which is what our result shows that XA is equal to mu 2 divided by mu 1 plus mu 2 and XB is mu 1 divided by mu 1 plus mu 2 and should be 1 be equal to 0, XB would automatically come out to be 0 and XA would automatically come out to be mu 2 divided by mu 2 which is equal to 1 right. So, all of this makes a lot of sense I leave this as an exercise to analyze the stability of these 2 steady states for the 2 genotypes that we have.

So, each genotype has only one solution in this case, and I leave this as an exercise to you to comment on of the steady state XA equal to mu 2 divided by mu 1 plus mu 2. While you are doing you can make the assumption that mu 1 and mu 2 are both positive quantities because you cannot have negative communication, rights their positive quantities you can take them to be unequal or equal that should not affect the analysis associated with this problem. So that is the discussion on mutation rates when we have 2 distinct genotypes, but in reality it is going to be a lot more complicated than that. And even in an experiment as simple as when we have bacteria growing in 1 millimeter tube we are going to have distinct genotypes coexisting in that tube with 1 millimeter culture.

So, how do we deal with cases like that? So, the we will develop our next what, we are going to do is develop a analysis for a situation like that you are not too, but n distinct genotypes are coexisting with each other. This is going to be slightly involved, but it turns out we get a very simple result at the end of our analysis. So, let me just lay out what is it that we are going to be talking about.

(Refer Slide Time: 25:14)

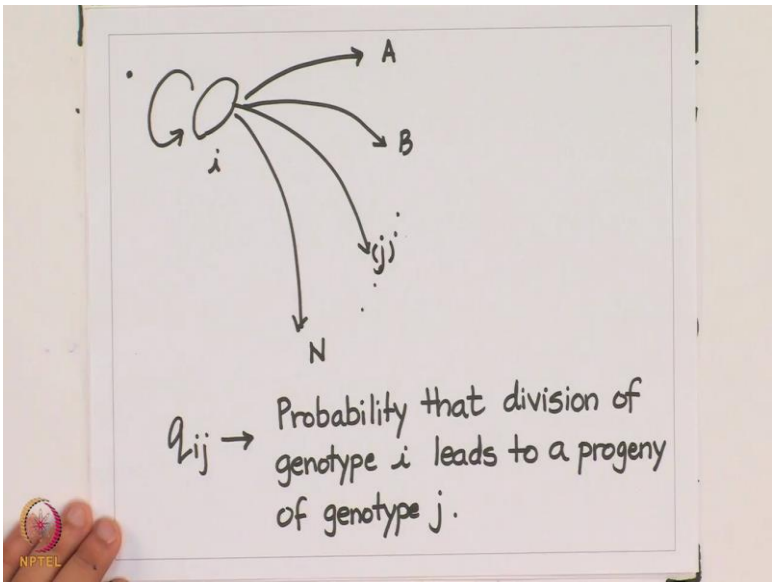


We have an environment e whose carrying capacity is K, hence all of this remains the same as what we been talking about, but now we have n distinct genotypes in this environment. Let us call them A B going all the way up to N. So, we have these n distinct genotypes again for the sake of simplicity. And because we only want to talk about mutations and not selection at this point growth rate for all genotypes is equal to 1.

So, we at this point we just want to lay the framework out which helps us develop the analysis for that. Next we want to define; what is the fraction of population belonging to each genotype. And that is going to be X subscript a will be the fraction of population that belongs to genotype capital A X subscript b will be the fraction of population belonging to genotype capital B going all the way up to XN, you should note that XA plus X b going all the way up to XN is equal to 1 rights, because sum of fraction of individuals belonging to each particular genotype has to be equal to the carrying capacity K.

And hence this comes out to be equal to 1. And when I differentiate this dx a by dt plus dXb by dt going all the way up to dxN by dt this quantity should be equal to 0. One thing that you should also note which is a matter of representation is that when you are summing quantity form XA to X n, xl this can also be written as sigma i varies from a to n X i equals one these 2 statements are equivalent and the last thing that we want to define is that of all these genotypes A to N.

(Refer Slide Time: 28:21)



What is the chance that one when genotype i divides it could lead to it is own genotype the progeny could have the genotype of the parent or the progeny could have any of the other genotypes associated in the environment?

It could have genotype A it could have genotype B and so on and so forth going all the way up to genotype n to represent this mutational effect we define a variable q ij which is the probability that division of genotype A division of an individual which is of genotype i leads to a progeny of genotype j. So, this quantity q i j represents what is the probability that when i divides, I end up with the genotype individual which is of genotype j right. This definition and the previous slide give us everything that we need to go ahead and analyze coexistence of n distinct genotypes in an environment e and calculate the frequency is associated with their numbers at steady state and that something that we are going to try and address in the next class.

Thank you.