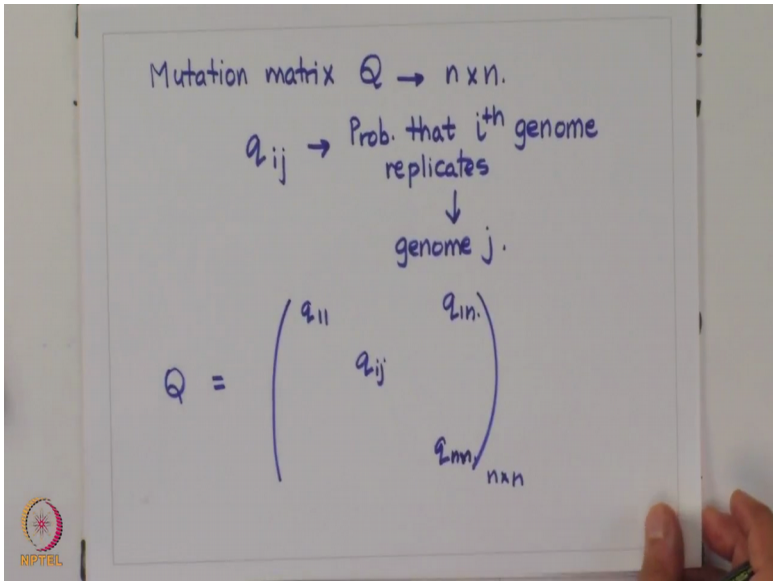
Introduction to Evolutionary Dynamics
Prof. Supreet Saini
Department of Chemical Engineering
Indian Institute of Technology, Bombay

Lecture - 16
Integrating Reproduction, selection and mutation

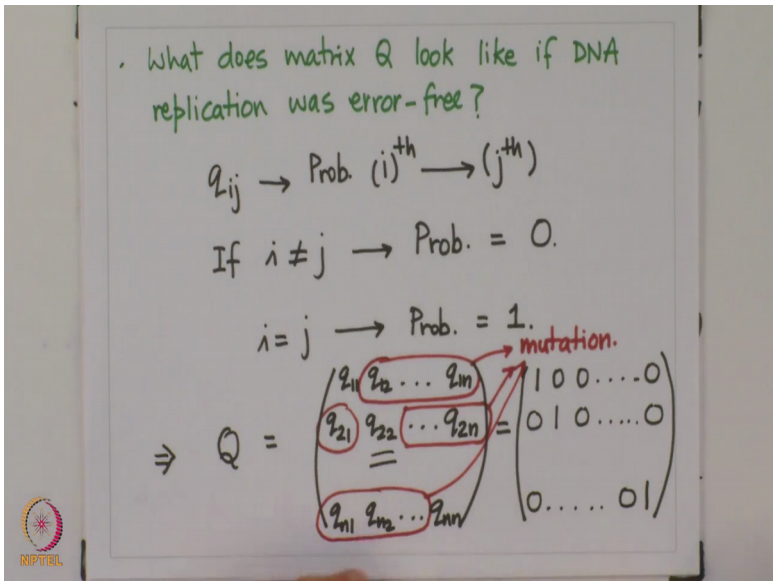
Hi everyone. Let us continue our discussion of the development of the cause of species model, where we are incorporating selection reproduction and mutation into account and in this mathematical frame work.

(Refer Slide Time: 00:36)



So, let us try to think about this, we talked about mutation matrix Q which is n cross n matrix and q ij represents the probability that when i-th genome replicates the progeny up on replication belongs to genome j. That is what this represents and we saw that this Q matrix is a n cross n matrix q 11 q ij and the last entry is q nn and so on and so forth.

(Refer Slide Time: 01:48)



So, we have this as a structure of the matrix Q; the how do we think about this first thing that I would like you to think about is what does matrix Q look like if DNA replication was error free which means that whenever a bacterium is defending its copying its DNA to be passed on to the progeny the DNA replication machinery the DNA polymerize does not make any error so what so ever and it is 100 percent faithful. Just pause the video for a few second, think about it for 30 seconds that what would this mean in terms of giving us the structure of matrix Q.

So, if this was the case then you have q ij as probability that i-th genome up on replication gives me the progeny which belongs to the j-th genome now because replication is error free if i and j are distinct; that means, if i is not equal to j; that means, these are 2 distinct genotypes then the probability associated with this transition is equal to 0 because there is no way my i-th genome up on replication is going to give me the j-th genome when the DNA replication is error free.

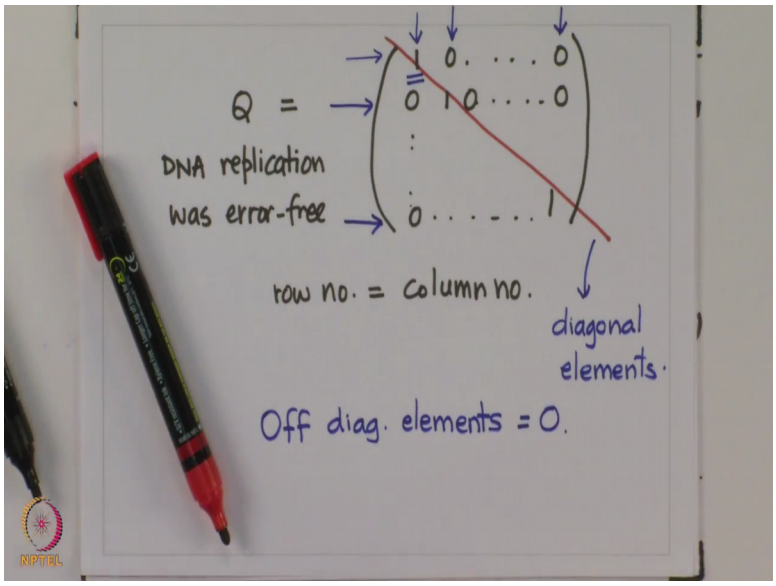
On the other hand, if i is equal to j what; that means, is the probability of this event happening is equal to 1 because every time a genome replicates the progeny is passed the exact same genotype as was there in the parent because DNA replication is error free hence the progeny and the parent are always carrying the same DNA and the probability of this event is equal to 1. And what does this tell us about my mutation matrix this tells me that the mutation matrix q whose elements are q 11, q 12 going all the way up to q 1n

for genotype 2, this will be q 21, q 22 going all the way up to q 2n and for the last one q n1, q n2 going all the way up to q nn. What this means is that in the first row all of these terms these terms represent that there was a mutation that happened when DNA was being copied, because the starting genotype and the ending genotype are distinct from each other. But that is not permitted if DNA replication was error free if this mutation even did not happen, but also note that the sum of that elements on this row i is equal to 1 what; that means, is that q 11 element is equal to 1 and all other elements in this row are equal to 0.

So, there is 0 probability that the replication event let to a genotype which was distinct from the parent genotype and the probability that the progeny DNA was identical to that of the parent DNA is equal to 1 for the second row this the first element and third to the n-th element are the elements which represent that there was the mutation event that happened up on replication again these are not permitted. And hence all these entries are equal to 0. And the only entry which is nonzero is q 22 whose probability of happening is 1, because every time a genome 2 divides the progeny will have genome 2, because DNA replication is error free hence the elements here are going to be 0 1 and then the rest of the row will have just zeros.

Similarly, the last one will have all of these n minus 1 entries representing mutations hence all these elements will be 0 and q nn will be the only nonzero entry in this and you will get 0es all the way and the last one will be equal to 1. So what that means, is that my matrix Q.

(Refer Slide Time: 06:36)



If DNA replication was error free would just be equal to what is called an identity matrix which has one as the one as the diagonal entries and all other off diagonal entries are equal to 0 and by diagonal entries you mean these entries whose row number is equal to the column number.

So, this entry for instance is called a diagonal entry because this corresponds to row number one column number one this is q 11 entry first row first column this is a diagonal entry because this is second row second column 2 two entry second row second column and so on and so forth going all the way up to q nn entry which is n-th row and n-th column.

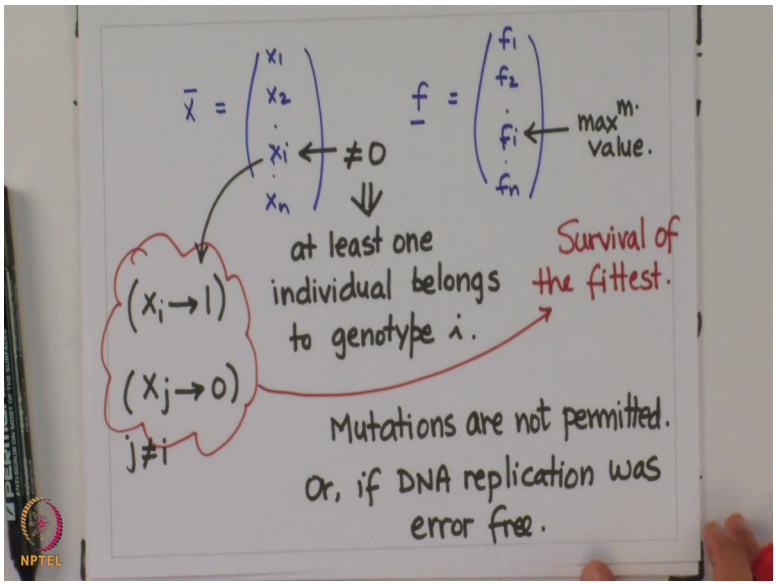
So, these are all the diagonal elements and the off diagonal elements associated with this matrix are all equal to 0 if DNA replication was error free. So, if replication is error free let us visualize our system for a minute and try and understand if we can intuitively guess as to what will happen to the process as we play this forward we have indistinct genotypes starting from one to n each growing at its particular growth rate f 1 to f n. And mutation events are not being permitted which means that whenever genotype 1 replicates the progeny is guaranteed to be of genotype 1.

On top of all of this we have constraint that the total population size does not exceed the carrying capacity of the environment which is k in such a scenario what should happen what should be the eventual outcome of this population. I mean, I would like you to just

pause the video for may be a minute or so think about this problem and try to come up with an intuitive picture of how the dynamics of a process such as this where mutations are not permitted will play out when we have a vector x defining the structure of the population and a vector f defining the fitness associated with each of the genotypes that existence in the population and mutations are not permitted, what happens.

So, what should happen here is that because fitness of each genotype are distinct; that means, that means fitness are unequal among the genotypes there is bound to be one genotype which has the maximum fitness from among the vector that defines fitness of all genotypes and irrespective of what the starting fraction of that population is.

(Refer Slide Time: 09:56)



Let us imagine that the structure of population is defined by this vector x which is x 1 x 2 going all the way up to x n and the fitness vector is just equal to f 1 f 2 going all the way up to f n its f i here and x i here.

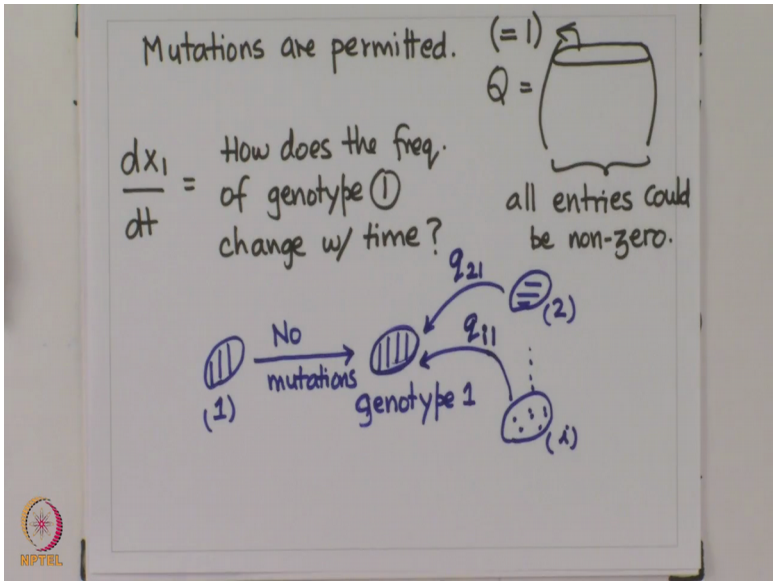
Let us imagine that, of all the n values here this one corresponds to the maximum value and this x i is not equal to 0 which means what you what this implies is that act of beginning of the experiment there is at least one individual which belongs to genotype i. So, it is not as if there is no individual in the population which belongs to genotype i then in that case this one or more individuals which belong to the i-th genotype which has the maximum fitness rate then that these individuals are growing to grow faster than every other genotype that exist in the population and eventually selection will ensure that this x

I, this genotype out competes all other genotypes in the in the environment and this eventually x i approaches one and all x js where j is not equal to i approach 0 as time moves forward.

So, this is going to be the outcome of the evolutionary dynamics of these n genotypes if mutations are not permitted. So, again this is the scenario when mutations are not permitted or if another way another way to say the same thing is if DNA replication was error free and this here what this means is you have one genotype going to frequency one all other genotypes going to frequency 0 and that one genotype that goes to frequency one is the one that corresponds to the maximum fitness this is an example of sort of survival of the fittest in this environment.

But let us now generalize is this was the special case in which we were talking that there are no mutations permitted. Let us now generalize this and try and understand that how would this play out in reality when there are mutations taking place and my mutation matrix that is associated with growth (Refer Time: 13:23) growth dynamics of this and genotypes is not an identity matrix where only the diagonal entries are equal to 1 and everything off diagonal is equal to 0, but the mutation matrix is far more general in nature and each entries allowed to be nonzero.

(Refer Slide Time: 13:44)



So, what happens then? So, now, we want to generalize this case and now mutations are permitted which is a much more realistic representation of what you would expect to

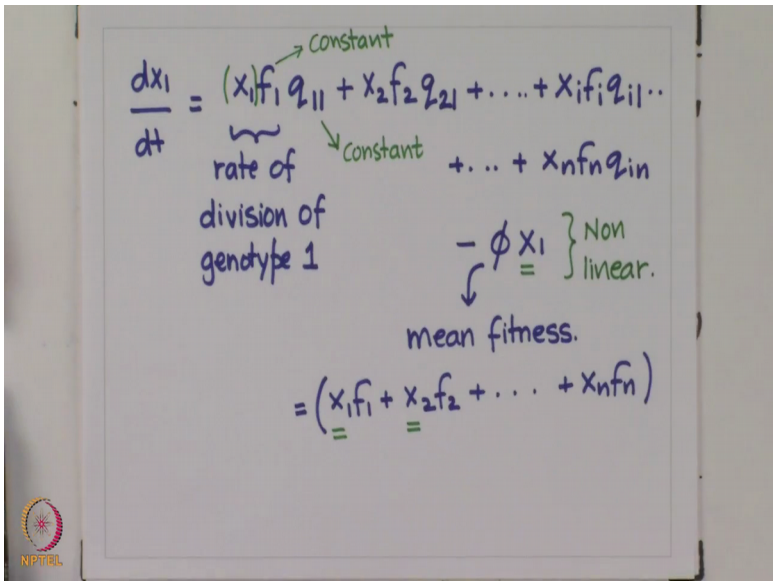
happen in nature if that is a case Q matrix is much more general and we cannot make any statement about any of the entries. So, all entries could be nonzero the only thing that we can say about certainty about this Q matrix is that the row sum has to be equal to 1 that is the only thing that we can say, but other than that and also the fact that every entry on this in this matrix is a is a positive number because these represent probabilities and you cannot have negative probabilities of events happening if an event is impossible you have probability equal to 0, but not negative probabilities.

So, now when we have this let us try and understand that how can we write d x 1 by d t which means how does the frequency of genotype 1 change with time what is the rate of this of this process happening with time in such an environment. So, to understand this what we want to see is that if we have genotype if we have for what is happening to individuals to genotype 1 this is an individual of genotype 1, I am interested in the rate at which this frequency is changing, but what all are the processes which could lead to a progeny which belongs to genotype 1 let us try and identify that.

So, a individual of genotype 1 could be produced if there is another individual of genotype 1 and its replication occurs and there are no mutations this process would give me a progeny which belongs to genotype 1 on the other hand I could also have an individual of genotype 2 and when this replicates mutation occur with probability q 2 one that the second genotype replicated and the progeny belong to the first genotype and that could also give me a genotype 1 this goes on, and if I have genotype i, when this replicates, it could give me a genotype 1 with the probability q i 1 that genotype i replicates and I get the progeny of genotype 1.

So, these are all the all the physical events the division events that could give me a progeny which belong to genotype 1. So, how do we know our job is? So, first of all you should identify that how many arrows or how many such distinct processes are there which would give me a genotype which could give me a progeny which belongs to genotype 1 there are n such distinct type of division events which gave a progeny one. And now our job is going to be to come up with an expression for each one of these n events and that gives me an estimate of at what rate is the frequency of genotype once changing with time.

(Refer Slide Time: 17:45)



So, let us try and write those expressions now. So, d x 1 by d t the first one let us try to write down the expression for this event that you have progeny of genotype 1 replicating no mutations happening I am sorry, the parent of genotype 1 replicating no mutations happening and the progeny also belongs to genotype 1 if that is the case then this expression can be written as x 1 f 1. This represents the rate of division of genotype 1 which also takes into account what is the current frequency of individuals belonging to genotype 1 in the environment and multiply that with q 11 which is the probability that should this replication event happen the progeny also belongs to genotype 1.

So, this gives me a quantitative estimate of how many genotype 1s are been generated via this arrow. Let us now try to do let us not try and now do the same for the second one and which gives me x 2 f 2 this is the division events that are happening per time for the individuals which belong to genotype 2, but I am only interested in those events which give rise to a progeny which belongs to genotype 1. Hence this will be multiplied by q 2 one and going all the way the i-th expression is going to be x i f i q i 1 and there will be n terms and this and the final term will be x n f n q i n this gives me an expression for all the events which give 2 a progeny which is of genotype 1.

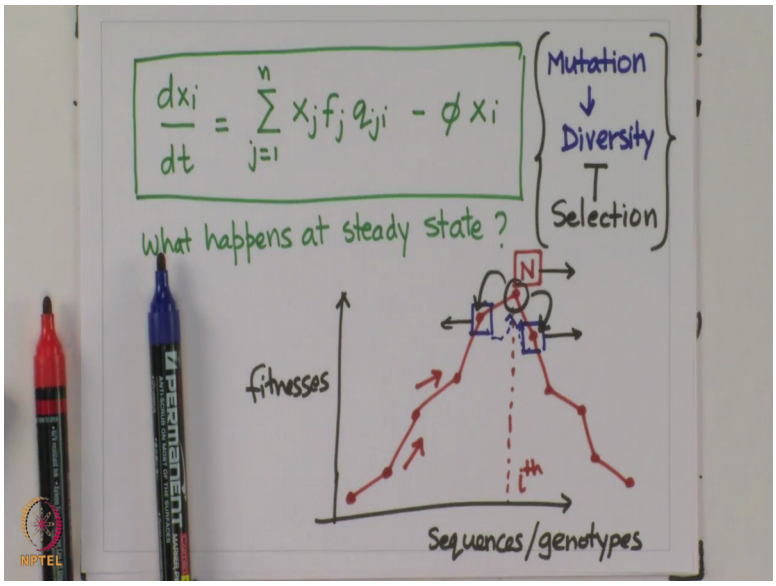
But remember; this setup has been done in an environment where the total number of individuals is constrained to be equal to the carrying capacity of the environment hence there has to be a dilution terms which accounts for a individuals being eliminated such

that the total population size is constant and that is why the variable phi comes back in to picture and we have done this derivation in the earlier parts of the class. So, this expression in addition to all the generation terms we will have phi times x 1 as the rate of elimination of individuals which belong to genotype 1 where phi is the mean fitness mean fitness taking in to account all the individuals which are present in the environment at this particular time and this we have already said is x 1 f 1 plus x 2 f 2 going all the way up to x n f n.

The first thing that you should note here is that this equation is a non-linear equation my variables the fs are fixed let me just write that down this f all the f is are constant because for every particular genotype the fitness rate of fitness or the growth rate is fixed it does not change with time the mutation matrix q is fixed once I define it once I say that q that genotype i can change to genotype j with the probability point one that mutation the that probability does not change with time. So, this 2 are constant the only variables in this setup are my x is which are the frequencies of every genotype and these are my variables if you look at the first n terms each one of those terms is linear with respect to x i there is only 1 of the x is that comes in each term and the power associated with that x is 1 hence these are linear.

But because of this term which is there to ensure that the population size remains constant there is nonlinearity in the system because phi contains and x 1 there is also an x 1. So, there is an x 1 square term then the second term in phi is x 2 that multiplies with x 1. So, there is x 1 x 2 term. So, this makes the equation non-linear which makes to solving this equation a little non trivial. So, it is a non-linear differential equation and of course, will have one such equation for every genotype that we are talking about. So, we are we are talking about solving n non-linear differential equation simultaneously to track the dynamics of this process.

(Refer Slide Time: 22:51)



So, this equation can be some simplified a little bit and we can write this as d x i by d t is equal to sigma x j f j q ij varies from one to n minus phi times x i, I leave that you to sort of try and convince yourselves that if I want to plug in i equal to 1, I get back this equation if I want to plug in i equal to 2, I would get the corresponding equation for rate of generation rate of change of frequency associated with genotype 2 and so on and so forth, this is the eventual equation that we have for such a system where we have n genotypes which are coexisting with each other.

Now, the question is that what happens at steady state is what we are interested in. So, what we want to find out is what happens at steady state and when we discuss the simpler of these 2 cases where there were no mutations being permitted it was easy to see that the structure of the population moves in a direction such that the fittest genotype eliminates all the other genotypes that exist in the population and the population is constrained to that particular genotype and it represented a case of survival of the fittest would that same thing apply here as well to understand that and these are non-linear equations whose solutions is not so trivial. So, we are going to try and use our intuition a little bit to try and understand can we figure out that what happens at steady state when we are talking of these systems.

So, let us draw a fitness landscape where this n sequences are represented. So, we have these sequences or genotypes and these are it is the corresponding fitness of each of the n

genotypes. And now let me say that these look something like this and you could acquire a mutation and move in a particular direction and so on and so forth. So, increase in fitness means that you are a genotype which is growing at a faster rate and this is the fittest genotype which is the i-th genotype associated from the n that I am talking about here. So, if selection were acting by itself we would have the entire population move and all n individuals will come to this particular genotype.

So, all individuals move to this particular genotype if selection were acting just by itself, but what is happening is that there are mutations happening in the system an at any particular time if you have n individuals and mutation events also taking place you would have this genotype mutating into some of the other genotypes the other genotypes also mutate back in to the n-th genotype that is true, but because there are no individuals to begin with at this particular genotype there is there are no individuals which are mutating into the n-th genotype. Similarly individuals here could mutate into this, but because selection has ensured that there are no genotypes at this at this particular sequence this arrow does not contribute towards any mutational events.

So, if the entire population was constrained at the top of this fitness landscape then you would have mutational events which would diversify which would create diversity and lead to other genotypes which are non distinct from which are distinct from the fittest genotype that we have around there, but once you have these genotypes being produced these now selection takes place and these individuals get eliminated because of selection and you have again have individuals belonging to this particular sequence which are being selected for. So, what is happening is that mutation is aiding diversity because individuals here are mutating into another genotypes mutation is trying to create more genotypes in the environment, but at the same time selection once to get rid of diversity and only select for the fittest individuals that are present in the environment which in this case correspond to those of sequence i.

So, eventually at the steady state of this process the balance between mutation and selection plays out and you get you get a spread of individuals about the mean where the maximum number of individuals are present at the fitness peak. But in the neighboring genotypes such as this and this you have a nonzero number of individuals being present and that plays out because of this balance that we are talking about between mutation and selection if there was only mutation you would have every possible genotype exist in the

population it was only selection you would only have the fittest genotypes survive in the population. But because both coexist there is fight between the 2 and the eventually at equilibrium is established where you have some spread about the population about the peak on the fitness landscape and that represents the steady state frequencies associated with the population of this n genotypes.

So, we will continue with this discussion and its implications on movement of populations on landscapes in our next lecture let us stop for this lecture.

Thank you.