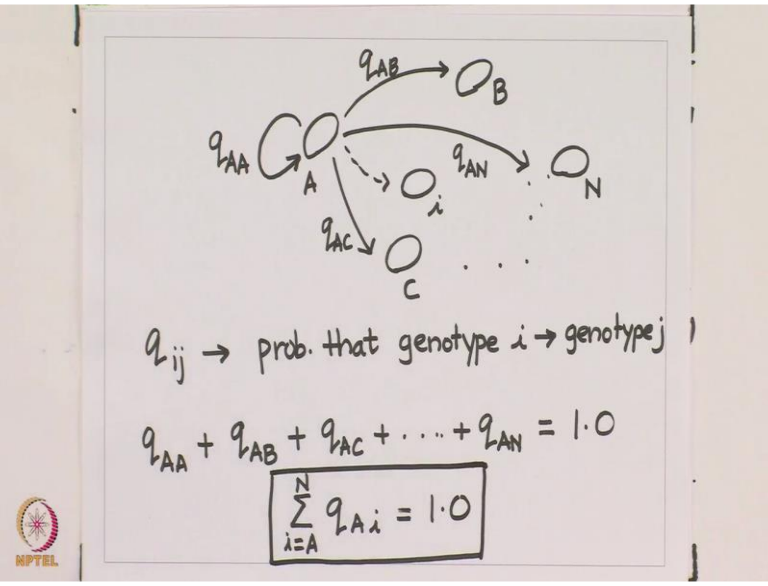
Introduction to Evolutionary Dynamics
Prof. Supreet Saini
Department of Chemical Engineering
Indian Institute of Technology, Bombay

Lecture – 10
Modelling Mutations – 2

Hi and welcome to the next lecture of the course, I may going to continue our discussion of mutations occurring in N species environment. We will just carry forward the framework that we had ended up in our last lecture and we had ended up and I want to spend a little bit more time on the variable q ij that we had talked about.

(Refer Slide Time: 00:42)



So, we have a bunch of genotypes genotype A, B, C, all the way going up to N and when a replicates it could either lead to a genotype A or via mutation it could either lead to genotype B or C or i or N.

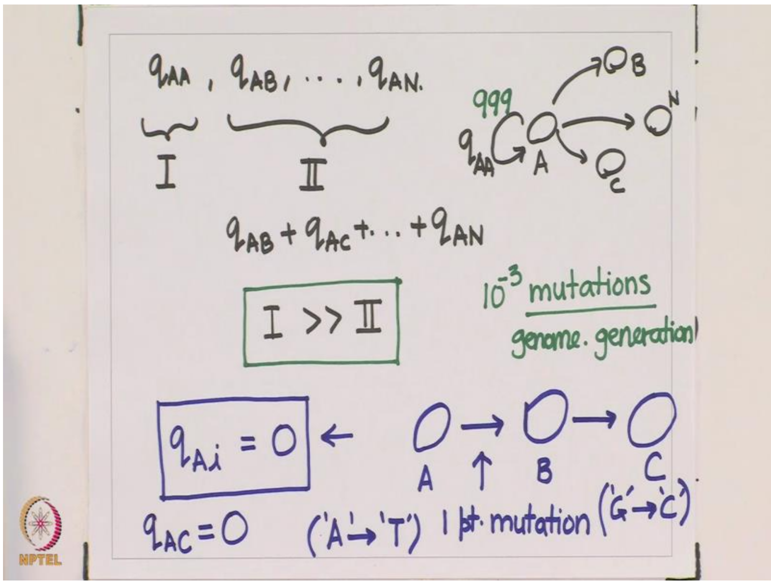
So, in this map, we are interested in 2 distinct type arrows that we have happening here; one is all the arrows which are leading out from a node such as this. So, remember q ij was defined as the probability that genotype i on division gives me genotype j. So, by that definition this arrow represents the probability that when A divides; I get back genotype A, this arrow tells me that when A divides; I get genotype B this represents that when A divides; I get genotype C and so on and so forth leading all the way up to q AN that genotype A divides and that gives me the resultant progeny is of genotype N via mutation.

This arrow here represents the case that no mutation is happened during the replication event and all these arrows here represents the fact that during the process of replication error had happened and the resultant the progeny is carrying the genotype which is distinct from the genotype of the parent one quantity that we should note here is q AA plus q AB plus q AC going all the way up to q AN, I would recommend that you pause the video for a few seconds think about what this sum should be and then we start the video.

In the representation that we had talked about this can we written as sigma i varies from A to N q Ai and this quantity should be equal to one and that is that comes from the fact that every time replication of A happens, it has to commit to one of these N arrows replication of A would lead to a progeny being generated and that progeny is going to belong to one of the N genotypes which are present in the environment, what is the chance that the progeny that has been produced belongs to genotype A is represented by this number the chance that the progeny has that is been produced belongs to genotype B is represented by this number and the chance that the progeny is of genotype N is represented by this number.

But the fact is that the progeny has to belong to one of the N genotypes which are present here and hence some of all the probabilities that the progeny belongs to one particular genotype has to be equal to 1. So, sigma q Ai; i varying from 1 to N is equal to 1 and this is going to come in handy as we move forward with our analysis; so now, when I am; when I am looking at an individual genotype.

(Refer Slide Time: 04:51)



Let me just drawback very quickly this just helps us keep in mind pictorially what is going on with the system that we are talking about I am drawing arrows leading out of only node a, but the same is going to be the case for every particular node that when B divides it could lead to progeny of type a or C or N for itself and so on and so forth.

Now, these arrows that are there I have quantities q AA the probability that progeny is of type a when division has happened with the parent A and q AB going all the way up to q AN, again I suggest that you pause the video for a few seconds and just try and think about the relative magnitude of these 2 quantities quantity one is the variable q AA and quantity B quantity 2 is the sum of q AB plus q AC going all the way up to q AN basically this sum includes all the q A terms accept for the q AA terms.

So, I suggest you pause the video for a few seconds and think about the relative magnitude of quantity one and 2 that should exist as you understand microbial physiology if you thought about this what should become apparent to you that what we have talked about. So, far suggest that one is much bigger than 2 and that is because if you remember the e coli mutation rate that we were talking about which was ten power minus three mutations per genome per generation.

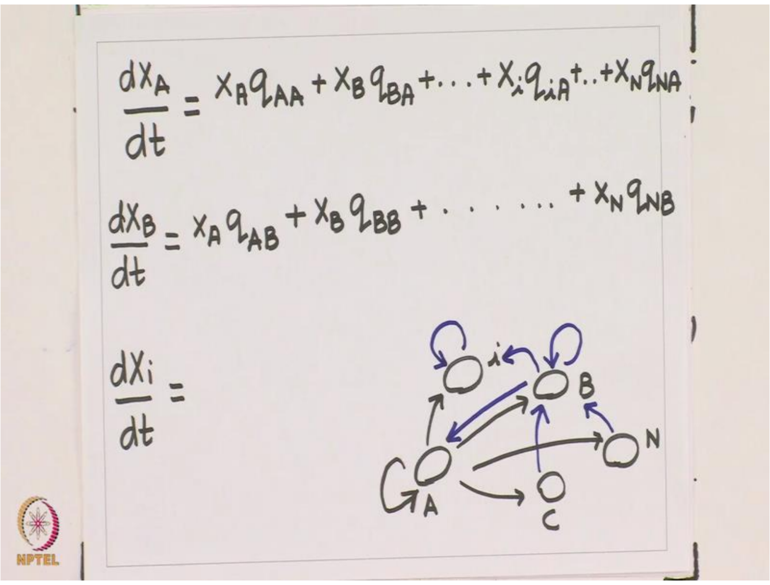
Which means if this was equal i replicating and you had node a as your parent equal i if you had thousand divisions thousand such replication events happening for division from node A, 99 out of them are going to be identical replicas of the parents genotype that was being copied and only one of these thousand replication events is going to be distributed to the remaining N minus one node N minus one edges which one of those N minus one edges is something that we do not know because mutations happened randomly it could be either one of the N minus one edges here, but the point is that this these numbers are. So, skewed that one quantity one is much much bigger than quantity 2 that something that is very important to realize.

We should also realize that some quantities q Ai maybe equal to 0, this is possible because so far we have discussing we are only talking of mutations occurring via point mutations which are single nuclei type errors which d N A polymerases making if that is the case then it is conceivable that genotype A gets mutated to genotype B this happens via one point mutation lets imagine this is an nuclei type A changing to nuclei type T, but B then mutates to genotype C and this change happens via g changing to C at another place on the chromosome. So, these 2 mutations that we are talking about are happening at distinct locations on the chromosome what; that means, is that when genotype A replicates there is no way that it can lead to a progeny of genotypes C provided that the maximum number of point mutations that we are allowing per replication event is equal to 1.

A cannot lead to C if only 1 nuclei type change is taking place in the progeny to lead to C A has to go via B it has to first lead to change of an A to A T at 1 location and then at another location on the chromosome G has to change to C. So, that genotype C is produced hence what that implies in this example is that q AC is equal to 0 and you might have several more quantities which are 0 depending on what are the precise mutations that link each of the in genotypes that we are discussing. So, 2 important things come out of this discussion that first the q AA the self replication arrow is the most likely arrow of division that is going to take place in the system that we are talking about and some of the other arrows in the system the way we are define them may turn out to be 0 if there is no way that genotype A upon replication and we are only permitting one nuclei type change maximum mutations to take place can lead to C genotype in one go.

So, having understood these let us try and formulate the dynamical equations associated with these with these end species let start with let start with d X A by dt d X A by dt would be imagine.

(Refer Slide Time: 11:16)



$$\frac{dX_A}{dt} = X_A q_{AA} + X_B q_{BA} + \ldots + X_i q_{iA} + \ldots + X_N q_{NA}$$

$$\frac{dX_B}{dt} = X_A q_{AB} + X_B q_{BB} + \ldots\ldots + X_N q_{NB}$$

$$\frac{dX_i}{dt} =$$

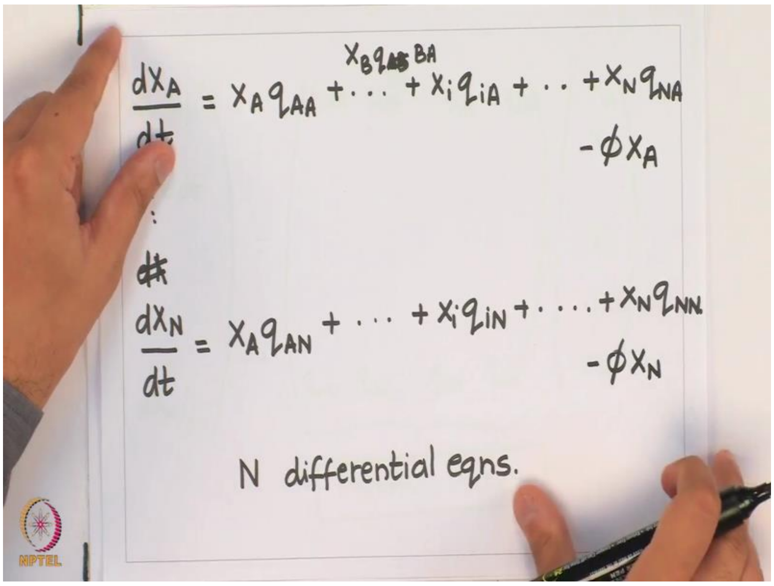
Again let us for the sake of clarity draw the picture once again we have genotype A, B, C and i and we have A leading to A, A leading to B, C and i, what influences the number of is that are present in the population that we are talking about here. So, to account for the number of as in the population we are going to be interested in all the arrows which are coming in to node a this is one arrow which is coming into node a which is every time a replicates and it leads to a progeny of type a what is this going to be equal to this is going to be represented by X 1 X A times q AA. So, important you realize this that q AA is the probability that when a divides the progeny is of genotype A and X A is the fraction of individuals which belong to genotype A at that particular instant. So, X A q AA represents how many progenies what fraction of progenies that are being produced are incoming as defined by this particular node.

Similarly, you are going to have division of progenies of genotype B and that via mutation may sometimes lead to a progeny which belongs to genotype A and that is going to be represented by X B which is the fraction of population which is of genotype B times q B A which represents what is the chance that whenever bacterium of genotype B replicates the progeny belongs to genotype A and so on and so forth we come to X i which is the fraction of population which belongs to the i X genotypes genotype times X i A representing the probability that when i replicates the progeny belongs to genotype A going all the way up to X N q N A make sure you spend a couple of minutes studying this equation and convincing yourself that this represents the change in the fraction of population which belongs to genotype A with time.

So, let us maybe try and write one more which is d X B by dt and now when you are talking of node B what we are interested in his all the nodes he is capturing the dynamics of all the nodes which are in coming to B where divisions are taking place of various genotypes in the environment, but the progeny been produced is of genotype B these are all the arrows that we are interested in. So, first of will start with this arrow that division of a parent of genotypes a leading to a progeny of genotype B and that is given by X A times q AB frequency of genotype A times the probability that when A divides the progeny belongs to genotype B plus X B times q B B, this represents this arrow where B divide and the progeny also belongs to genotype B and again going all the way up to X N times q N B.

Again I suggest that you take a couple of minutes and write an equivalent equation for d X i by dt what; that means, is now you are interested in all the nodes which are coming into any i-th genotype and account for its rate of change of frequency with time once we have this we will; obviously, have N distinct differential equations which captures the dynamics of this of these N coexisting species.

(Refer Slide Time: 16:32)



$$\frac{dX_A}{dt} = X_A q_{AA} + \overset{X_B q_{BA}}{\cdots} + X_i q_{iA} + \cdots + X_N q_{NA} - \phi X_A$$

$\vdots$

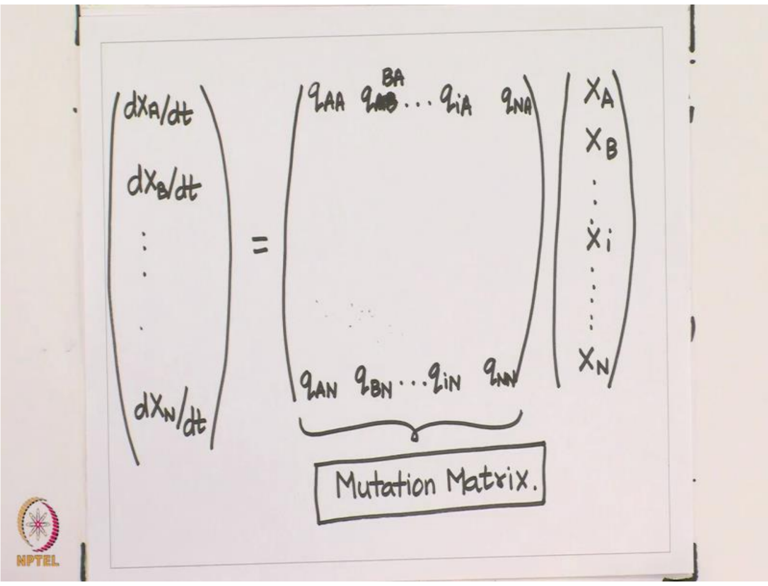
$$\frac{dX_N}{dt} = X_A q_{AN} + \cdots + X_i q_{iN} + \cdots + X_N q_{NN} - \phi X_N$$

N differential eqns.

So, and we can write them up fresh and I will write to which is d X A by dt equals X A q AA plus X i q i A going all the way up to X N q N A going all the way up to d X d X N by dt equals X A q AN plus X i q i N plus X N q N N in all these are N differential equations and in order to arrive at the steady state associated with this dynamical system these N differential equations have to be solved simultaneously that is what we would do to arrive at the steady state of this system and then we will have to analyze the stability of those steady states and decide where does what are the frequencies at which the system wants to stay here which is which is not a not a trivial task and what we are going to do now is take help of concepts from linear algebra to simplify the system further and see if we can get any insight from there.

So, linear algebra we are going to write this differential equation in the form of a matrix and if you look at the left hand side of the equations that we are just arrived you see even see that all we have on the left hand side are N particular derivatives for representing each of each one of the N species. So, the left hand side is just N derivatives and so the left hand side of our linear equation is going to be these N derivatives.

(Refer Slide Time: 18:42)



$$\begin{pmatrix} dX_A/dt \\ dX_B/dt \\ \vdots \\ dX_N/dt \end{pmatrix} = \underbrace{\begin{pmatrix} q_{AA} & q_{\overset{BA}{\cancel{AB}}} \cdots q_{iA} & q_{NA} \\ & & \\ q_{AN} & q_{BN} \cdots q_{iN} & q_{NN} \end{pmatrix}}_{\boxed{\text{Mutation Matrix.}}} \begin{pmatrix} X_A \\ X_B \\ \vdots \\ X_i \\ \vdots \\ X_N \end{pmatrix}$$

D X 1 X A by dt d X B by dt and so on and so forth going all the way up to d X N by dt that is my left hand side what is on the right hand side let me take a look at my equation I have a factor q AA multiplying with X A, but I also have a factor q AN multiplying with X A that is the first that the first entry for every equation for the i-th entry of every equation I have a factor multiplying with X I, I have a factor another factor multiplying with X i and for the last entry for each equation I have a factor multiplying with X N and I have another factor which is multiplying with X N.

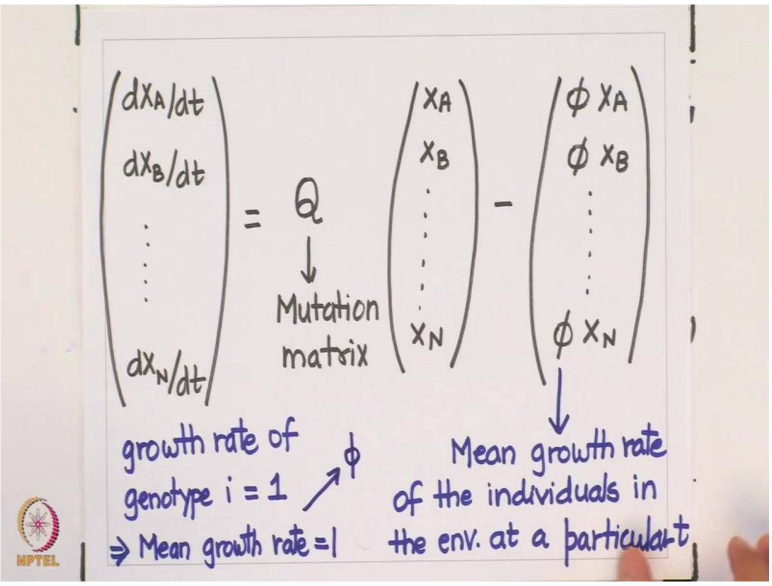
This tells me that this system can be represented as the product of a matrix and a vector and a my job is to find the that matrix and that vector which represents this particular dynamical system in a linear sense when we do that let me say this is the matrix that is still unknown and this is the vector that I am talking about and the vector can be sends from the fact that the first term the first time in each equation is multiplied with X A the i-th term in each equation is multiplied with X i and the last term in each equation is multiplied with X N that gives me a clue about what should be the nature of this vector and that should be X A X B X i going all the way up to X N that is the vector that I am talking about.

Now, what I want to do is find the elements of this matrix let me look at my first equation my first equation is d X A by dt which is this entry and the first term is that X A gets multiplied with q AA remember that when a matrix is multiplied with a vector the first row multiplies with the column vector and that gives me the first entry here and in that first entry X A is multiplied with q AA which should be the first term of the matrix which is q AA this is followed by the second term which will be X B into q AB, I have X B here; that means, the second entry here should be q AB going all the way X i should multiplied with get multiplied with q i A; I am sorry, this should be q B A because we are interested in whenever replication of B happens and the progeny belongs to genotype a. So, this entry here should be q B A i A and the last one here should be q N A which is division of genotype N happening and leading to a progeny of genotype A.

That is the first row similarly we can go on and we get the last row as q AN q B N q i N q N N which represents that whenever genotype i divides and leads to a progeny of genotype N that is represented by the probability q i N. So, that gives me my matrix structure and this matrix is called the mutation matrix q. So, this is my mutation matrix what am I missing? So, far you should ask this question at this point and realize that something that we will being talking about throughout is not been incorporated in this equation which is what is in every time we have talked about growth equation we have incorporated the variable phi, we have not done it at this point. And we need to take into account phi and remember the growth of each equation is growth of every particular genotype is represented by what is being added minus phi times the frequency of that particular genotype which is phi times X A.

The equation here will be phi times X N and so on and so forth. So, the complete equation which takes into account and which enforces that the total number of individuals gets constraint to K total number of individuals belonging to each genotype when added together does not exceed K has to be taken into account.

(Refer Slide Time: 24:29)



$$\begin{pmatrix} dX_A/dt \\ dX_B/dt \\ \vdots \\ dX_N/dt \end{pmatrix} = \underset{\substack{\downarrow \\ \text{Mutation} \\ \text{matrix}}}{Q} \begin{pmatrix} X_A \\ X_B \\ \vdots \\ X_N \end{pmatrix} - \begin{pmatrix} \phi X_A \\ \phi X_B \\ \vdots \\ \phi X_N \end{pmatrix}$$

growth rate of genotype i = 1 ↗ ϕ

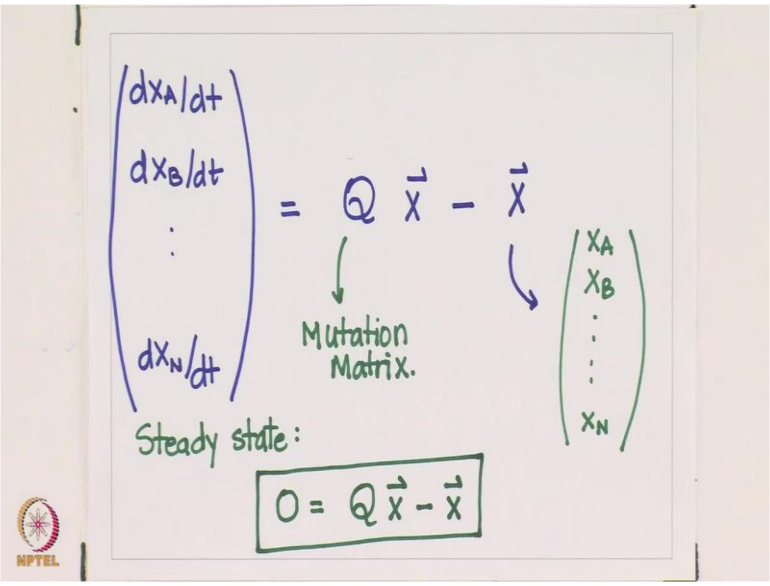
⇒ Mean growth rate = 1

Mean growth rate of the individuals in the env. at a particular t

And that leads me to my final equation that d X A by dt d X B by dt d X N by dt is my left hand side which represents the rate of change of frequencies of all N genotypes associated with the environment this is equal to q times q which is my mutation matrix times the vector X A X B going all the way up to X N minus phi times X A phi times X B and so on and so forth going all the way up to phi times X N and phi in all our discussions you have always found that to be equal to the mean growth rate of the individuals in the environment.

At a particular time that is that is what we have realize the significance of phi to be that it represents the means growth rate associated with the environment at any particular time d depending on the frequencies associated with each genotype and because similar to the 2 species mutation that we were talking about here also every single genotype is going with the growth rate A. So, since growth rate of every genotype is equal to 1; that means, mean growth rate is also equal to one and this growth mean growth rate equal to 1 is independent of what is the composition in terms of the vector X at any particular time T because no matter what is the composition of the population at any time T every single genotype is going at a growth rate 1 enhance the mean is always going to be equal to 1. So, phi for our case is equal to 1.

(Refer Slide Time: 27:51)



$$\begin{pmatrix} dx_A/dt \\ dx_B/dt \\ \vdots \\ dx_N/dt \end{pmatrix} = \underset{\substack{\downarrow \\ \text{Mutation} \\ \text{Matrix.}}}{Q}\;\vec{x} - \underset{\searrow}{\vec{x}} \quad \begin{pmatrix} x_A \\ x_B \\ \vdots \\ x_N \end{pmatrix}$$

Steady state:

$$\boxed{0 = Q\vec{x} - \vec{x}}$$

What is means our system reduces to is the following equation d X A by dt d X B by dt going all the way up to d X N by dt equals q which is the mutation matrix times vector X minus phi which is just one times vector X vector X here represents X A X B going all the way up to X N and mutation matrix is as we have defined in the previous slides and to arrive at the solution of this equation.

We have to first look at the steady state which would employ that all derivatives defined to the left hand side are equal to 0; that means, to solve the system we have to solve the following equation 0 equals mutation matrix q times vector X minus X. This is the equation that we need and solve to solve the mutation between N distinct genotypes at any given point in time and that is not a trivial task. And we are going to borrow help from concepts from linear algebra to be able to arrive at the solution and that something that we are going to pick up in our next lecture onwards. We will stop here for this one.

Thank you.