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

Lecture - 23
Genetic Drift in evolution of microbial populations

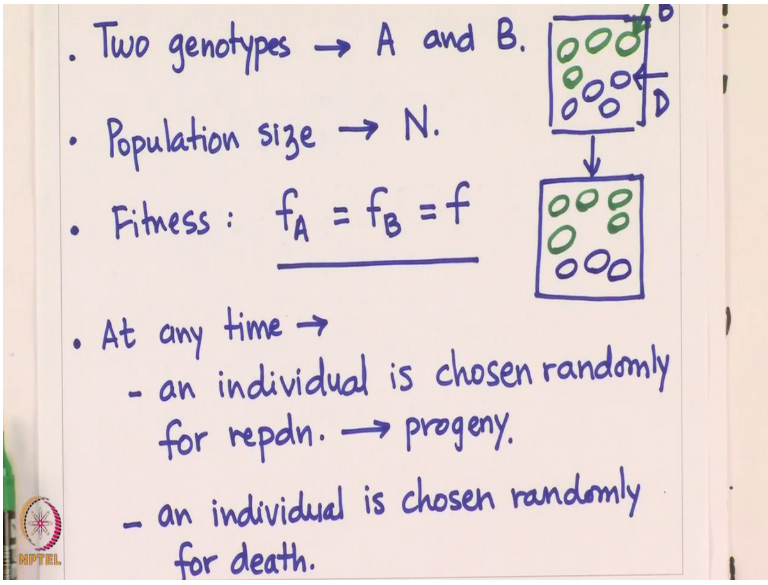
Hi everyone, in the last lecture, we had played this game called models in a jar game, where we had seen that the effect of drawing marbles from a jar randomly could have one of the 2 results eventual results for the game as we had defined the rules for. If we had started the game with red and blue marbles and we were drawing them randomly from the jar and adding the same marble of the same color to the consequent jar, then we had seen that eventually we would run into a situation, where either the blue marbles or the red marbles would become extinct and we would only be left with marbles of one color in the jar.

And from there on because marbles were not allowed to change color, if at one time step at N th jar in the game, if you are left with only blue marbles; then no matter how long we keep on playing the game for; we would only be left with that same color of marbles. It is important to draw the analogy of this game with bacterial dividing populations and what this is trying to communicate to us is; that randomness associated with the process of division and death is going to have implication in terms of what is the frequency of a type of individuals that are present in a population.

And what we are going to do today; is learn from the example that we did the last time around the marbles in the jar game and try and develop that quantitatively in terms of a bacterial division process and try and understand what effect does randomness associated with the process of division and death have in terms of dictating the dynamics of evolution and population frequencies in a bacterial population.

So, let us start our discussions and again will we will go back to the marbles in a jar game from time to time and try and draw analogies from that as we discussed our bacterial population drawing. So, as with the game that we had defined last time; we will start with a bacterial population that consists of 2 genotypes.

(Refer Slide Time: 02:33)



So, let us call the 2 genotypes that we have genotype A and B; we worked in an environment where the population size is constrained to be a constant and the invariant number of individuals in a population at any given time is equal to N. The 2 genotypes that we are talking about, the fitness is associated with those 2 genotypes; let us call f subscript A as the fitness of genotype A is equal to f subscript B; equal to fitness of genotype B and let this just be some number f.

So, to start off with we will discuss a case where there is no selection advantage that one genotype has over the other. Again, if we draw that parallel with the marbles in a jar game that we did in the previous lecture, what we should realize here is that this saying that the fitness of genotype A is equal to fitness of genotype B is equivalent to going into that jar drawing a marble out and in that example, the probability of drawing out a red marble was equal to that of drawing out a blue marble.

Hence both marbles of each color were equally fit, equally likely to be drawn out from the jar; where I was picking 1 marble out of the 20 marbles that I had in the jar. Exactly the same as that, the fitness of genotype a is equal to fitness of genotype B which means the chances that a replication event will happen in the next replication event that happens in the population; is equally likely to be a replication event associated with genotype A as with genotype B.

So, we will introduce selection advantage for one of the genotypes a little bit later on, but to start off with let us begin with an example, where there is no selection advantage and both genotypes have equal fitness and now at any time t and this is really the key at any time t; 2 things happen, an individual is chosen randomly for birth or for reproduction which gives rise to a progeny and an individual is chosen again randomly to represent death.

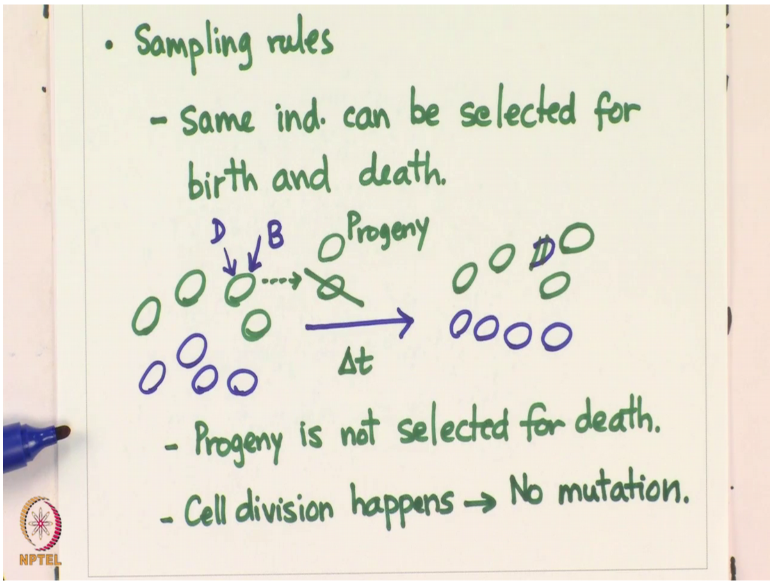
So, what is happening is that at any given point in time; if this is my system at time t and the population network of my system at time t looks as like this, I have a bunch of genotype A; represented by green and genotype B represented by blue. What I am going to do at this time step is; select an individual randomly and selection of that individual will mean that this is the individual which has given rise to the progeny in this time interval that I am talking about.

For instance, if this is the individual that selected for birth and simultaneously I select another individual which is marked out for death and let us imagine that this is the individual that is marked; that is selected for death. So, this is the one that selected for birth and this is the individual which is selected for death, then what happens in the system as I move from this time to this time is that because this individual has given rise to another progeny and that progeny will have the identical genotype as the parent.

So, in the next instant I have five green cells because this one has given rise to another one. So, the number of green cells goes from 4 to 5 whereas, the number of blue cells because the individual marked out for death happens to be a blue in color; the number of blue cells left in the system will remain; will come down to 3 from 4.

This at any given time we choose birth and death events to ensure that the number of individuals in the population at any given time does not exceed N. So, in the example that we had just drawn; the number of individuals in the population was it at 8 and then after that to represent a birth event that would lead to production of another progeny take the number to 9, but simultaneously we mark an individual to represent death and that individual gets eliminated from the population and the number again comes back to 8. So, hence the number of individuals present in the population at any given point in time does not change with time in this model.

(Refer Slide Time: 08:30)



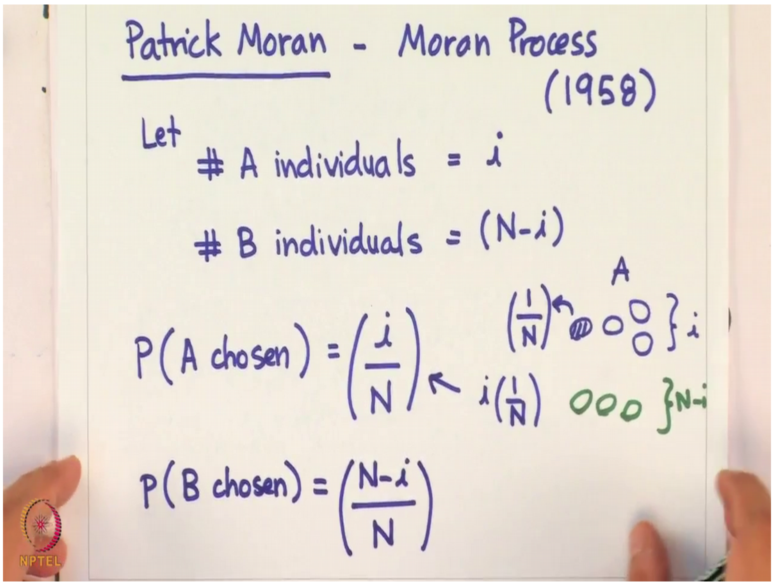
So, we have a few more rules that we need to define as we want to describe the dynamics of going from one time to another, where you want to capture that dynamics and those rules are such that we will call them the sampling rules and the rules are such that same individual can be selected for birth and death; that is the first rule. So, again in terms of our example what we mean here is that; if we have a bunch of individuals belonging to genotype one and these are the individuals belonging to genotype two; because I am choosing individuals for birth and death randomly, there is no bias associated with it the choice of individuals for birth and death is random in nature.

When I make thus, when I choose an individual out of this population for birth; let us say this is the individual that gets chosen for birth and when I am choosing an individual for death from among these individuals in the population, the same individual gets chosen for death. What happens is; what we are physically representing is the fact that in this time interval delta t; the other seven cells do not change, they remain the same and what happens is that this particular individual divides, this is the progeny and this is the parent and the parent cell dies in this delta t time and we have the progeny; I am sorry; it should be green, we have the progeny represented in the next time step.

So, the number of individuals belonging to a particular genotype does not change in this delta t time, you are still left with 4 green cells and 4 blue cells. So, the rules coming back to the rules is saying that in this delta t time when you are selecting randomly one birth and one death event, the same individual could get selected for birth and death and that is a permitted sampling strategy. The other rule is that the progeny is not selected for death; so, the progeny never get selected for death. An individual when it is selected for birth is selected from the N individuals that existed in the population prior to division and when an individual is selected for death is also selected from those and individuals that were present prior to the birth of the progeny.

So, the progeny is left out and the third rule is that when cell division happens; there is no mutation happens, which means that when I am selecting genotype A for reproduction, then the progeny that is produced from that genotype A will always belong to that genotype A. I am never going to get into a situation, where division of progeny A leads to production of an individual belonging to genotype B; that is not permitted.

(Refer Slide Time: 12:27)



So, with these rules we come to the algorithm that we want to discuss and this is a very famous algorithm which was proposed by somebody called Patrick Moran; who was an Australian geneticist and the process is called the Moran process and this was developed in 1958. So, we will discuss how do we frame the process that we have just described in words quantitatively and what insights does it lead me to when I am talking of evolutionary genetics and in particular what I am interested in understanding here is what is the role that randomness associated with drawing cells for birth and death, what implications do the randomness have associated with these processes in dictating evolutionary dynamics and the frequencies of individual genotypes that are present in the environment.

So, let us discuss this process and again we will start with giving a quantitative aspect to the steps that we have just discussed. Let number of A individuals; A is one of the 2 genotypes that represented my environment; let number of A individuals in the population be equal to i that automatically means that number of B individuals in the population is equal to N minus i; because the total size of the number of individuals present in the environment is always constrained to be equal to k.

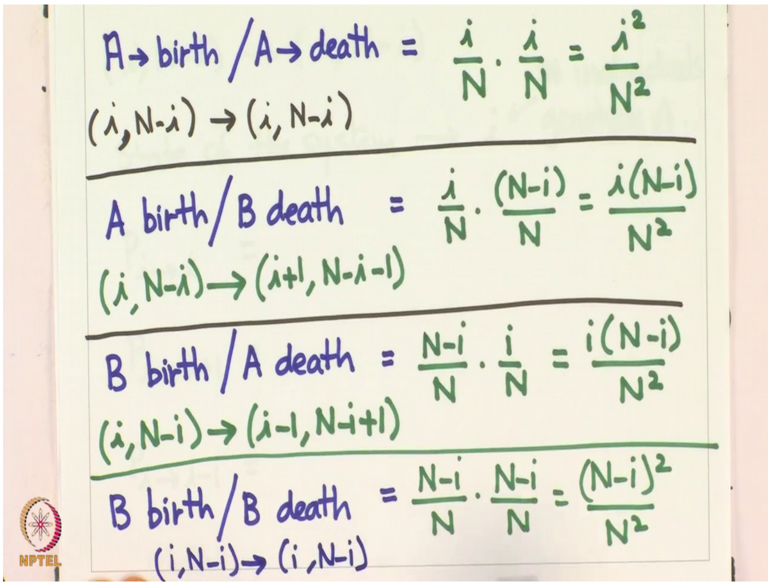
Now, when I am choosing these individuals randomly either for birth or death that choice is totally random, so, any one of the N individuals is equally likely to get drawn whether that drawing out process is for birth or death; irrespective of that its equally likely every individual is equal equally likely to get drawn. So, when I am drawing a cell for either birth or death, the probability that A gets chosen; will just be equal to i divided by N.

If i is 0; that means, there are no individuals in the population which belong to genotype A then of course, if there are no individuals of that particular genotype then there is no chance of drawing out that particular genotype and hence this comes to 0. On the other hand, if I is equal to N; that means, all individuals in the population belong to genotype A then I draw an individual and that is guaranteed to belong to genotype A because all individuals are that particular genotype; so, the probability automatically comes out to 1.

Another way to think about this is that if you have a situation such that the number of genotype number of individuals of genotype A are equal to I and number of individuals of genotype B are equal to N minus i, then because every individual in this population is equally likely to get drawn; the chance that I will draw this particular individual is just equal to 1 by N. Because when I am drawing an individual, I am drawing one of these individuals and the chance of drawing any particular is equal to everyone else's chance and hence drawing a particular individual out is equal to 1 by N.

And now when I am computing the probability that eventually I draw one of the blue genotypes which let us say is genotype A; then I just add 1, but the probability of drawing this plus the probability of drawing this cell, plus this cell, plus this cell; since there are i such individuals, the total probability will become i times 1 by N; which is just I by N. Similarly, the probability that I choose an individual of genotype B is just equal to N minus i upon N. So, I have these 2 probabilities that I am going to work with, now probability of drawing A and that probability that drawing could have happened for birth or death and probability of drawing B and that could have went drawing for A or for birth or death. These can be combined four different ways; to compute four different possibilities associated with one step of the Moran process.

(Refer Slide Time: 17:36)



So, when I am drawing; I need to; first thing I need to realize is that I need to do 2 draws; one to select which cell divides and the other one to select which cell dies. So, there could be four possibilities that I could choose A for birth and A for death or I could have A for birth; while selecting for reproduction, I draw out a genotype A and while selecting for death; I select a genotype B.

Or I could have selected an individual of genotype B for birth and an individual of genotype A for death or I could have an individual B for birth and an individual B for death. You should realize now, that these are the only four possibilities that are associated with this if I am dealing with 2 genotypes only; there is no other possibilities to represent birth and death process and what we will try and do now is write down the probabilities associated with each of these events happening.

So, when I am selecting A for birth, what is the probability? I have already seen that probability that I choose A is just equal to i by n. So, the chance that I select an individual of genotype A for birth is just equal to i upon N and the chance that I select another individual A for death is also i by N. This is also i by N because the same individual which was chosen for birth could also get selected for death and this comes from the sampling rules that we discussed a little bit back.

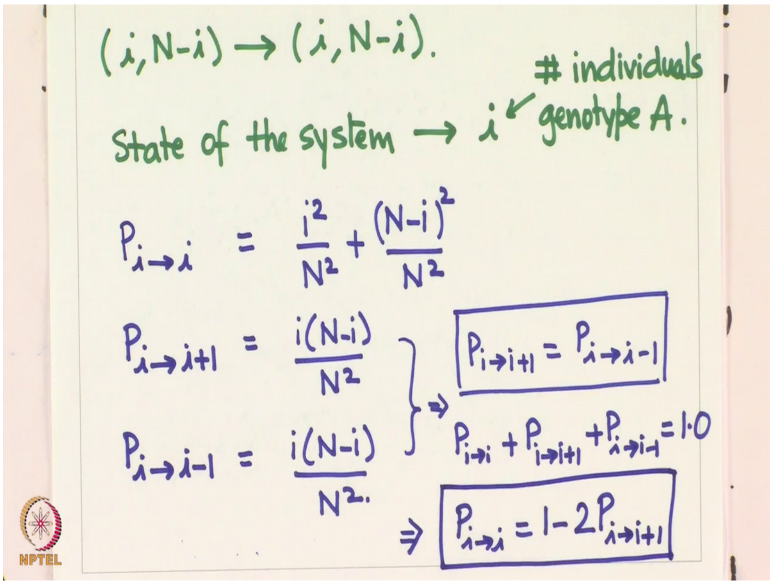
So, this probability that I choose A for birth and A for death is equal to i square by N square. What does this do to the number of the cells? So, I started my system at i and N minus i; this was the state of the system which represents the number of individuals of genotype A and number of individuals of genotype B. If I am choosing A for birth; that means, the number of a individuals goes up by 1 because the progeny is going to belong to genotype A, but since I am also choosing and genotype A individual for death; that means, number of A individuals comes down by 1.

So, if A gets chosen for birth as well as death; the number of; the state of the system does not change. I am still going to be left with i individuals of genotype A and N minus i individuals of genotype B. So, that is one of the four possibilities that can happen with the system; what about if A gets selected for birth and B gets selected for death. So, selection of A probability associated with that is i by N; selection of p; it is N minus i divided by N; this is just i times N minus i; divided by N square and the state of the system from i and N minus i; again this represents number of individuals of genotype A and number of individuals of genotype B. When this happens since A is giving birth; that means, number of A individual goes by 1 and B is associated with death; that means, number of individuals belonging to genotype B comes down by 1. What you are moving to is i plus 1 to N minus i minus 1, the total number is ensured to be held constant to N, number of individuals of genotype A goes up by 1, number of individuals of genotype B goes down by 1; so that is done.

What about the next one? B for births is just going to be N minus i divided by N and A for death is just going to be i by N; that gives me the same probability i into N minus i by N square and the state of the system here changes from i N minus i to i minus 1 into N minus i plus 1 because the number of; it is the number of B individuals now; which has gone up by 1.

The final case that we have; the final of the four possible cases that we have here is B birth and B death and; that means, I am selecting B for both the events that I am talking about here which is just equal to N minus i upon N dividing times N minus i upon N which is equal to N minus i whole square divided by N square and the state of the system is on another sheet, the state of the system associated with this fourth choice is B birth and B death; that means, B birth means number of B goes up by 1, but B is chosen to die; that means, number of individuals belonging to genotype B comes down by 1.

(Refer Slide Time: 22:57)



Hence the system moves from i, N minus i to i, N minus i; it stays at the same place because it is the same genotype which is selected for birth and death. Birth increases the number by 1 and choosing the same genotype for death in that time interval, brings down the number of individuals of that particular genotype by 1 and hence the state of the system does not change; it still remains that i and N minus i.

So, we discussed that these are the four only four possibilities that are associated with the system; there is no other possibility that is left out. So, one thing that you should do is sort of convince yourself that addition of these fourth probabilities equals to 1 because this represent all possible outcomes associated with the system, nothing outside of this can happen as per the rules we have defined.

So, these probabilities should equal to 1 and I leave that as an exercise for you to develop. The way we are going to represent this is; that we have seen that the state of the system can just be represented by a single variable i because this i will represent number of individuals of genotype A. Given an i, we automatically know that the number of individuals of genotype B is just going to be N minus i. So, we can define the state of the system by just one variable and we do not need to specify number of A individuals and B individuals.

Just by giving the number of A individuals, we automatically subtract that from N and we get the number of B individuals. So, we know that do not have to specify the number of B individuals and the way we are going to develop this forward is write three probabilities. What is the probability that the system starts from i and remains at i? What is the probability that the system starts at i and goes to i plus 1 and what is the probability that system starts from i and goes to i minus 1. To understand this we will write the expressions for each of these three. What, since these are the only four possibilities and let me just write down at the bottom here; this is the case where i N minus i remains i N minus i. So, now we are interested in probability that the system starts from i and remains at i.

Which is represented by 2 of the four options; which is A birth, A death and B birth, B death. At both these systems, we start with i and where we remain at i; that is true for both these cases. So, the probability that the system started from i and remained at i after delta t time is going to be represented by this probability plus this probability because both these options represent that the state of the system does not change.

So, this equals to; I am just going to write the 2 here, i square by N square plus N minus i square by N square. Similarly, I want to write an expression for the probability that system moves from i to i plus 1 and that is represented by this case, where A birth and B death happens; system moves from i to i plus 1 and that probability is equal to i into N minus i divided by N square and lastly I am left with this option and that takes the system from i to i minus 1 and the expression is equal to this, so I am going to write that here.

So, now we have probability expressions for each of these three transitions that can happen and we know that there are no other past transitions that are possible in delta t time. In that delta t time when we select one birth and one death event; these are the only three transitions that are possible. What you should also realize at this point; is that looking at these 2 expressions we get P i to i plus 1 is equal to P i to i minus 1; these 2 probabilities are equal, I also know that P i to i plus P i to i plus 1 plus P i to i minus 1 is equal to 1. I know this couple of ways; I know this because these four possibilities represent all possible outcomes of the system.

I could also add these probabilities and convince myself that the sum of these probabilities is equal to 1, which implies that I can write P i to i as 1 minus 2 times P i to i plus 1. So, we have developed these probabilities; we have developed these expressions and now we have sort of equipped ourselves to answer the question that we started this whole discussion with; that what is the role that randomness or genetic drift plays in dictating evolutionary dynamics and frequencies associated with genotypes. So, having developed these expressions, we will use this to develop our analysis further and answer that question in the next lecture.

Thank you.