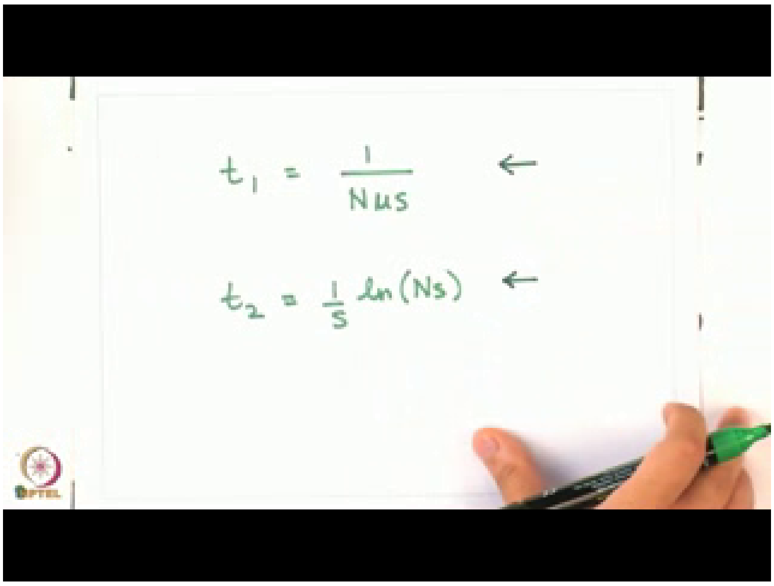
Introduction to Evolutionary Dynamics
Prof. Supreet Saini
Department of Chemical Engineering
Indian Institute of Technology, Bombay

Lecture – 30
Evolutionary dynamics when mutations are rare

Hi everyone, continuing our discussion from the last time we had derived expressions for t 1 and t 2 for a bacterial population evolving in an environment where populations tries is held to be constant equal to K. t 1 was associated with the time that I have to wait for the first beneficial mutation which does not get lost due to drift and gets established in the population. Once this beneficial mutation is established in the population selection takes over and its numbers go up to n and the mutation is said to be fixed in the population. t 2 is the time that it takes time that this beneficial mutation takes after getting established to reach fixation.

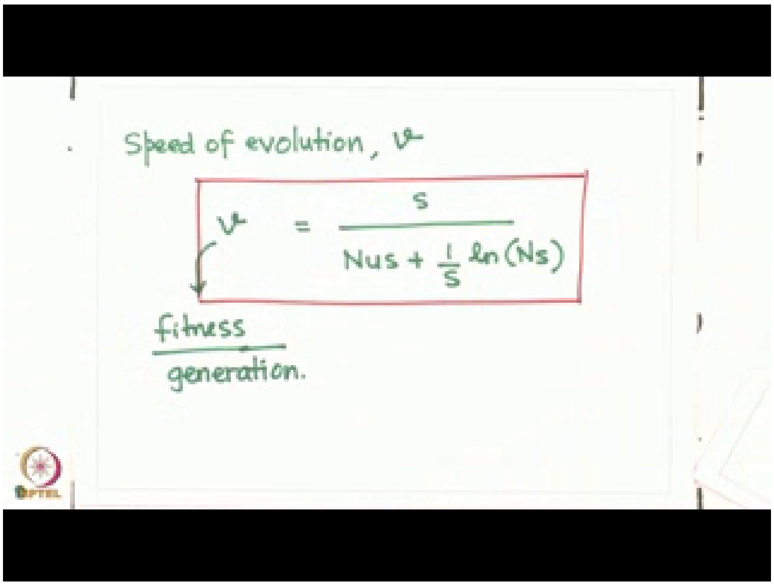
And in the last lecture we had obtained expressions which accounted for this t 1 and t 2 in terms of number of generations that I have to wait for each of these 2 events.

(Refer Slide Time: 01:19)



And now let us continue our discussion from there. So, this is the expression that we had received for t 1 and this is what we had got for t 2 and as we had discussed before that once a beneficial mutation is established and fixed in the population, fitness of every single individual in the population has increased from 1 to 1 plus s. Starting genotype every individual was at fitness one and now since the first beneficial mutation has become fixed in the population fitness of every single individual is equal to 1 plus s; that means, that mean change in the fitness that has happened is equal to s and that change s has happened in time t 1 plus t 2.

(Refer Slide Time: 02:04)



Speed of evolution, v

$$\boxed{v = \frac{s}{Nus + \frac{1}{s}\ln(Ns)}}$$

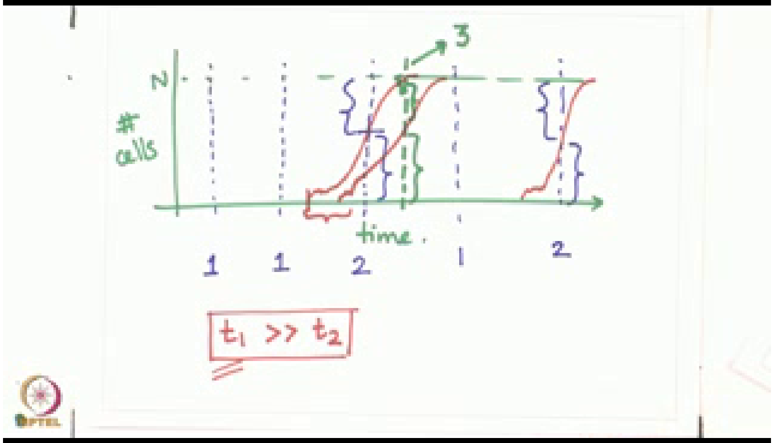
$$\downarrow \quad \frac{\text{fitness}}{\text{generation.}}$$

So, speed of evolution can be written as s divided by Nus plus 1 upon s l n of N s let us call this speed of evolution equal to v and that is equal to fitness per generation. So, that is the first result that we have in which we can quantify the speed of evolution in an analytical manner and come up with an expression that defines how fast is a bacterial specie evolving. But let us now take a look closer look at how does this experiment actually play out here in their physiological setting where this experiment is actually being conducted and try and understand some of the some of the implicit assumptions that have gone into development of this expression.

So, the way we have developed the system is if you just let us just take a step back and think about what we have assumed here the way we have developed this expression is that we have an environment e where there are n individuals and one beneficial mutation occurs in this environment this beneficial mutation might get lost or become fixed suppose it becomes fixed then this mutation is acted upon by selection and its numbers the number of individuals corresponding to this particular mutant genotype quickly increases to n, then we have to wait for the next beneficial mutation to become fixed in the population and so on and so forth.

The implicit assumption that we are making in this in this whole analysis is that at any particular instant in the system there is only a maximum of 2 genotypes that co exist the parent genotype and the mutants genotype because once a mutant genotype arises its numbers increase and go towards n.

(Refer Slide Time: 04:26)



So, the implicit assumption in this model that we are making is that this is time again the implicit assumption that we are making is that I am not going to draw the mutations that who are eliminated by drift, but only the once that gets established. So, and then the next one occurs somewhere over here and so on and so forth.

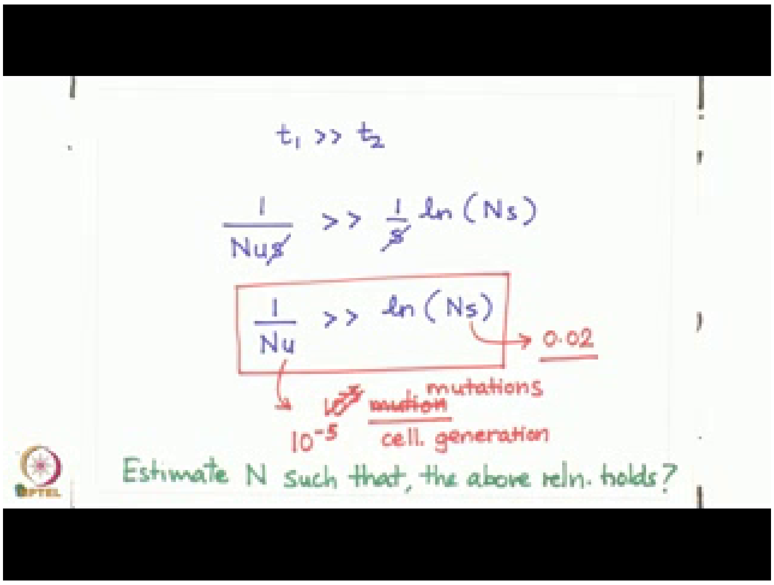
The implicit assumption is that if I were to draw an vertical line which corresponds to a particular time. So, this is like a snapshot of the environment at each of these times in any of these snapshots the maximum number of genotypes that exist in the population is equal to 2. So, here every individual belongs to the parent genotype. So, there is only one genotype again one genotype here there are 2 genotypes with this many number of mutant individuals and this many number of the parent genotype. So, this is 2 one and again mutant genotypes the parent genotype. So, this is 2. So, the maximum number of genetic variability that exists in the population is limited to 2.

We do not account for the fact that it is possible that while this mutation was becoming fixed and going to was going to fixed after establishment they could have been another mutation that also got established and was going towards fixation. And if that was the case then we have a scenario where at this particular instant we have some this mutant some individuals which are this new mutant this is the number of the first mutant and since some original parent genotype left in the population and then genotypic variability is 3 distinct genotypes. We do not account for the fact we do not account for the fact that the number of unique genotypes that are co existing in the environment in any particular time could go beyond 2 that assumption is valid if the t 1 expression that we have derived is much greater than t 2 remember t 1 is the time that I have to wait for the mutation to get established t 2 is the time that I have to wait for the mutation to get fixed once its established.

What this is telling us is that if on average t 1 is much bigger than t 2; that means, that the time that I have to wait for the first mutation to get established is much longer as compare to the time for it takes for the mutation to go from establishment to fixation then the assumption that I only see a maximum of 2 distinct genotypes is fair. Because since the waiting time for a mutation is very very large the chances that 2 mutations of beneficial nature will arise and get established in the population next to each other are going to be very very small because one beneficial mutations having becoming establish from here then the waiting time for the next beneficial mutation to get established is very large compared to the time that this mutation would take to go to fixation.

Hence the implicit assumption in all of our analysis is that t 1 is much larger than t 2. So, what we will do before we proceed is check this assumption and find out the physiological conditions under which this particular assumption holds.

(Refer Slide Time: 08:25)



$t_1 >> t_2$

$$\frac{1}{Nu\cancel{s}} >> \frac{1}{\cancel{s}} \ln(Ns)$$

$$\boxed{\frac{1}{Nu} >> \ln(Ns)}$$

0.02

~~10^{-5}~~ ~~mution~~ mutations

10^{-5} $\frac{\text{mutations}}{\text{cell. generation}}$

Estimate N such that, the above reln. holds?

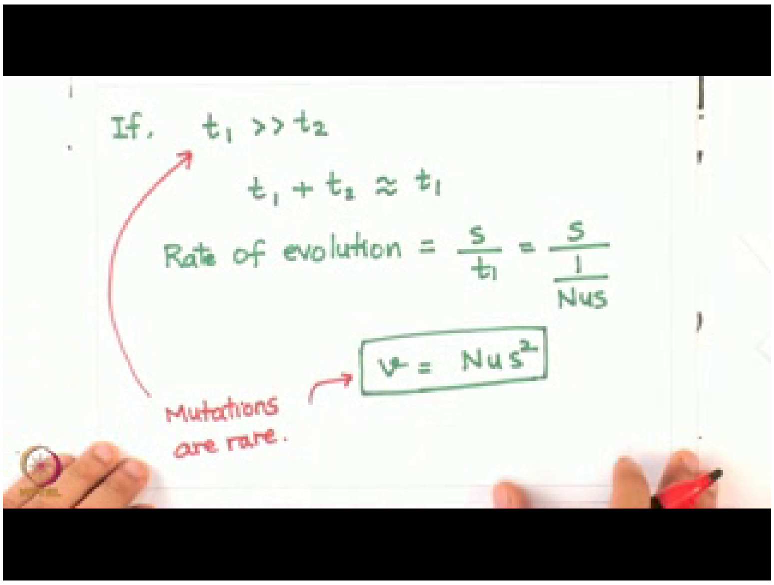
So, again all of our analysis what we have discovered holds if t 1 is much bigger than t 2. So, what are the constraints on the variables of our system which ensure that this is true? So, once again let us rewrite our expressions for t 1 and t 2 and we get one upon Nus should be much bigger than one upon s l n of N s we can cancel the s s from the 2 sides and the condition that we get is one upon nu is much bigger than l n of N s for our analysis to be valid.

Now, what does what does this tell let us try to think about this a little more what are these variables n is the population size associated with the environment that I am dealing with u is the beneficial mutation rate associated with the bacterium and s is the selection coefficient associated with each beneficial mutation that is occurring in the system. So, let us try and get realistic estimates of s and u the selection coefficient associated with each beneficial mutation and the beneficial mutation rate that is happening in this bacterium and try to get estimates of n such that this relationship holds that will give me an idea of what is the size of the environment that I am looking at in terms of number of individuals that that environment should support such that this assumption holds and all of the previous analysis that we have just completed is also valid for this environment. So, what could be a good represented value for s large beneficial mutations we know confer a advantage of 4 to 5 percent, but those have beneficial mutations are relatively rare. So, s of perhaps 2 percent is a realistic representation of average benefit that a beneficial mutation confers to an individual.

So, s let us take that as 0.02 u is the beneficial mutation rate the total mutation rate of equal is 10 power minus 3 mutants mutations per cell per generation I am sorry, this is 10 power minus 3 mutations per cell per generation. Now we understand that these this mutation rate encompasses all mutations deleterious lethal neutral and beneficial. So, the beneficial mutation rate is bound to be lower than this and if we imagine that one in a hundred mutations is beneficial in nature; that means, this mutation rate can be estimated as 10 to the power minus 5 mutations per cell per generation which are beneficial in nature.

Now, once we have these 2 numbers I leave this as an exercise for you to come up to solve this inequality and come up with an estimate; estimate n such that having plugged in these values of s and u the above inequality holds the above relationship holds. So, you should try and solve this and find out that what is the constraint on n that we are putting when you are saying that this in equality must be obeyed what is the constraint on n that that automatically imposes itself. So, solve for n and comment on what is the nature of the environment e where you can do this experiment with this constraint on n such that t 1 is much bigger than t 2 which means that a picture like this that we drew where one beneficial mutation occurs goes to fixation this one does not exist then we wait a long time before the next beneficial mutation gets established this picture holds true depending on what is the answer that we get from this inequality. So, that is there.

(Refer Slide Time: 13:08)



If, $t_1 >> t_2$

$t_1 + t_2 \approx t_1$

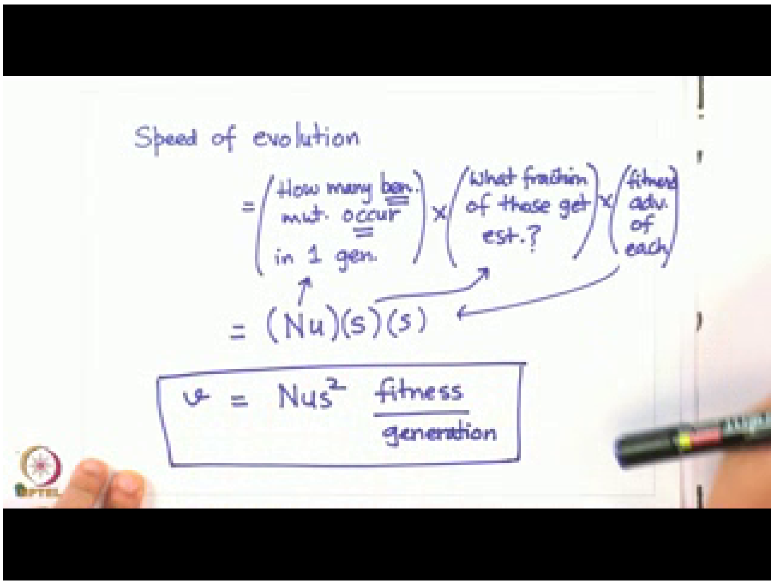
Rate of evolution $= \frac{s}{t_1} = \frac{s}{\frac{1}{Nus}}$

$\boxed{v = Nus^2}$

Mutations are rare.

But what that also means is that if t 1 is much bigger than t 2 then we can write that if t 1 is much bigger than t 2 then we can write that t 1 plus t 2 is approximately equal to t 1 which just means that the rate of evolution is just equal to s upon t 1 which is s upon one divided by Nus which is equal to Nus square again this relationship is valid in a regime where mutations are rare which is essentially what saying t 1 much bigger than t 2 means because t 1 is the waiting time the that you have to wait before the first beneficial mutation gets established if t 1 is very large compare to other times in the process. That means, the waiting time is the determining factor which decides speed of evolution and this will only happen another way of same the same thing is that mutations in the system are very rare. So, that is the speed of evolution associated with the condition that mutations are very rare.

(Refer Slide Time: 14:55)



Speed of evolution

$$= \begin{pmatrix}\text{How many ben.}\\ \text{mut. occur}\\ \text{in 1 gen.}\end{pmatrix} \times \begin{pmatrix}\text{What fraction}\\ \text{of these get}\\ \text{est.?}\end{pmatrix} \times \begin{pmatrix}\text{fitness}\\ \text{adv.}\\ \text{of}\\ \text{each}\end{pmatrix}$$

$$= (Nu)(s)(s)$$

$$\boxed{v = Nus^2 \ \frac{\text{fitness}}{\text{generation}}}$$

Another way to think about this is that if we try our goal is to try and estimate an expression for speed of evolution. So, another simple way to think about this is speed of evolution is just equal to if I were to ask you that in one generation time how many beneficial mutations occur. So, we are going to write this as how many beneficial mutations occur not fixed, but occur in one generation we need an expression for this times what fraction of those get established we want an expression for that times what is the fitness advantage of each of these mutations. So, what we are saying is that in one generation how many mutations occurred times what fraction of those got fixed. So, this gives you the total number of mutations that got established in one generation time.

So, if this number is let us say this number of mutations that occurred is hundred then each gets established by a probability of point one; that means, a total number of mutations that get established in one generation is equal to hundred times point one which is 10 and if each confers a fitness advantage of 2 percent 0.02 then you have 10 into 0.2 as the 0.2 fitness increase per generation as the speed of evolution.

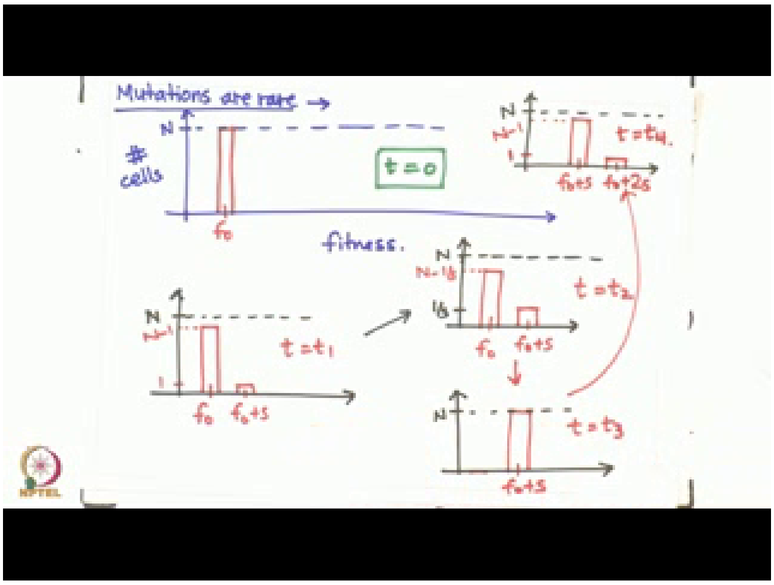
So, if we want an expression for each of these quantities. So, number of mutations that occur and again when we say mutations we are only talking of beneficial mutations here. So, the number of mutations that occur is equal to N u the fraction of these that get established is going to be equal to s because we already know that beneficial mutation gets established with the probability equal to its selection coefficient this is s. And each of these fitness each of these beneficial mutations confers a fitness advantage equal to s as we have assumed in our model that all beneficial mutations are equivalent in nature and each confers a selection coefficient s and that is the benefit that this mutation upon fixation confers to every individual that is present in the environment.

So, this quantity will again be s this just gives me speed of evolution is Nus square fitness per generation which is v which is which is what we had obtained in the last slide when we had assume that t 1 much bigger than t 2 and hence t 1 plus t 2 is approximately equal to t 1. And we substituted rate of evolution as mean change in fitness divided by the total time in which that change in fitness has taken place which is equal to t 1 plus t 2, but under these conditions can be approximated to be just equal to t 1 and hence you get the same expression from either of these 2 ways to think about this.

So, this was analysis considering that mutations are rare beneficial mutations you have to wait long time before a beneficial mutation occurs and gets established in the population. But in a physiological setting if you if you work out the N u and s inequality that we had developed and find the appropriate condition where this inequality that waiting time is very large compared to fixation time holds you will find that that is that this scenario is not really relevant for a large number of physiological settings because N u that is the total number of mutations that occurring in any in one generation time for any bacterial system. And because n associated with a system is lightly to be very very large N u is often a very large number that is a total number of beneficial mutations happening in this system if N u is large; that means, mutations are not rare and waiting times are not as large as compared to the what was going on in the picture that we had just drawn.

So, mutations are actually felly rapid and what we are going to do next is derive an expression for speed of evolution in a case where mutations are happening rapidly may be for we try and develop that expression let us try and intuitively understand what would the picture look like in a setting like that. And to understand that better we go to one of the other representations of output from an evolutionary experiment because I think that gives a better insight in to what is happening in the system.

(Refer Slide Time: 20:21)



So, this is the case where mutations are occurring rapidly now before we do that its perhaps useful to understand what will be the corresponding picture in this representation that we are going to talk about now and understand what would this look like when the mutations were rare. So, this is representing what we have already done mutations are rare and I this representation y axis is number of cells this is n and x axis is fitness. So, in what we just did when mutations are rare you have all individuals at fitness f naught this is at t equal to 0 then at some point you have one beneficial mutation occur

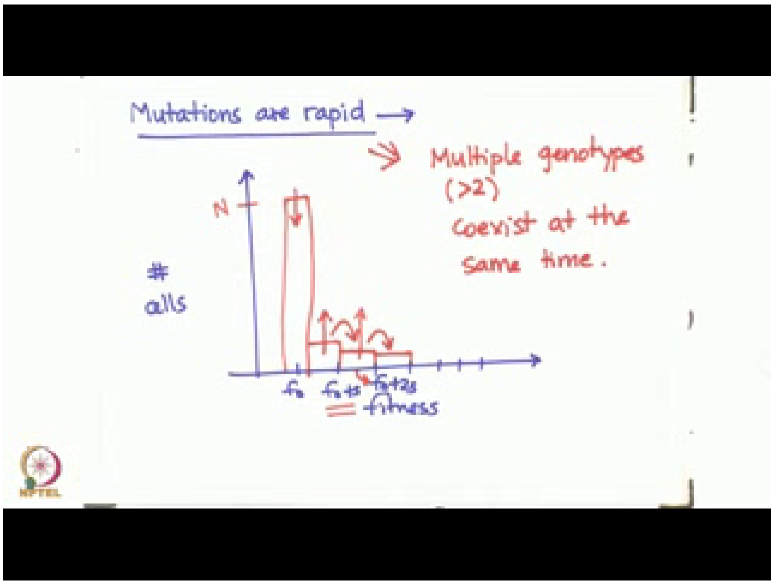
So, this is at t equal to 0 all individuals at fitness f naught as we go forward in time one beneficial mutation occurs and that gets established not labeling the axes, but the same axes labels carry here as well. So, at this time now one beneficial mutation has occurred and you have n minus 1 individuals of the parent genotype with fitness f naught and we have one individual with fitness f naught plus s this is one this is t equal to t 1. Now this mutation lets imagine that it does not get eliminated due to drift and is able to establish itself if that is the case then at t equal to t 2 what the picture is going to look like is that the number of individuals corresponded to the parent genotype at fitness f naught is going to be n minus 1 by s and this one mutant has increased in numbers and gone up to one by s number of individuals which correspond to this particular genotype.

And now the mutation has gotten established this is t equal to t 2 now selection takes over and fixation of this mutation happens and fixation of mutation happening means that this particular genotype is going to out compete the parent genotype and eventually you will end up with all individuals at f naught plus s and 0 individuals at the original fitness f naught this happens at t equal to t 3. Then under this regime that we just discussed fitness mutations are relatively rare you have to wait a long time before the next mutation occurs at t equal to t 4 and gets established.

And in that case at t equal to t 4 now you have n minus 1 individuals at f naught plus s and you have one mutant individual that has arisen which is f naught plus 2 s and the whole story repeats itself and this is equal to t equal to t 4 and the idea been thought once the population reaches t equal to t 3 it takes a long time for the system for the first mutation thereafter to get established and t 4 is much larger as compare to the other time steps associated in this process. So, if this is the setting that we have for rare mutations.

In an equivalent setting in an equivalent setting what do you imagine is going to be happening when mutations are no longer rare, but are much more rapid in nature? So, N u is a very large number which is the total number of beneficial mutations occurring per generation and u s is the total number of beneficial mutations which are going to get established by generation if this number is much larger than one then what you have is that more than one mutation are getting established per generation. So, in if we draw the equivalent picture in these regime then mutations are rapid what would that look like.

(Refer Slide Time: 25:09)



So, now, we are interested in a regime where mutations are rapid what is this going to look like the way it is going to play out is the following that we have fitness here f naught f naught plus s f naught plus 2 s and so on and so forth and we have number of cells on the y axis and to start with if this is n there were n number of individuals at fitness f naught and as I go forward in time the cells are growing eventually what is going to happen is that one mutation is mutations are going to keep on happening and some of them will be lost because of drift. But one mutation will eventually get established when that happens number of individuals which correspond to fitness f naught will come down and you will have some mutation some number of individuals which correspond to a fitness of f naught plus s now because this mutation has gotten established the numbers associated with this particular genotype will increase numbers associated with the parent genotype will decrease.

But before this can reach fixation that is before this genotype can totally eliminate the parent genotype from the environment another mutation from this pull happens and you have another genotype which corresponds to fitness f naught plus 2 s this event has meant that now you have 3 distinct genotypes in the population at the same time and we had precisely seen that mutations are rare implies that there are not more than 2 genotypes at any given time in the environment. So, we have 3. So, now, this particular genotype is going to out compete both the genotypes the frequency associated with the middle genotype will keep on increasing as long as this fitness f naught plus s is more than the mean fitness of the individuals in the environment.

So, if the frequency is were as represented this will go up this will go up this will come down, but before this individual has gone to fixation or this has been completely remove from the system another mutation happens and we have 4 genotypes. So, what mutations are rapid this regime implies that multiple genotypes and by multiple its greater than 2 genotypes co exist at the same time because we are not limiting beneficial mutations occurring in the environment beneficial mutations are constantly available to individuals of the environment the fitness will keep on increasing, what we are interested in is that what shape does this graph take when system takes on when we wait for a very long time and observe this observe this graph take a particular shape what is that shape going to be when we waited for a very long time and the system has a time to accumulate accumulations and that something that we will start the next lecture with.

Thank you.