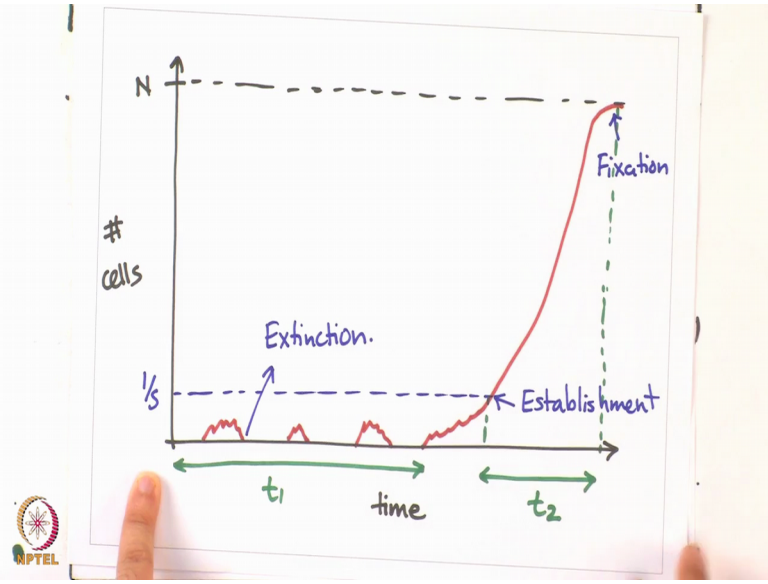
Introduction to Evolutionary Dynamics
Prof. Supreet Saini
Department of Chemical Engineering
Indian Institute of Technology, Bombay

Lecture – 29
Estimating the speed of microbial evolution

Hi everyone, in the last few lectures we had talked about representations of data that we obtained from evolutionary experiments and we had an mission different scenarios in which we can depict how evolution in a microbial population can be represented graphically. Today what we are going to do is try and try and develop expressions for rate of evolution in these microbial populations taking in to account the fact that there is a significant role which is played by genetic drift or randomness associated with which beneficial mutations survive and which a go extent.

And again at the back of our mind we should always keep in our mind the very important result that we had derived that should there be a beneficial mutant that arises in the population which holds a selection coefficient S over the rest of the population then the chance that that mutant is able to go to establishment and eventually fixation and not be eliminated via genetic drift is equal to S. So, a beneficial mutation which confers a 5 percent advantage over the rest of the population only has about a 5 percent chance of going to fixation in the population and there is a 95 percent chance that that mutation is eliminated via genetic drift.

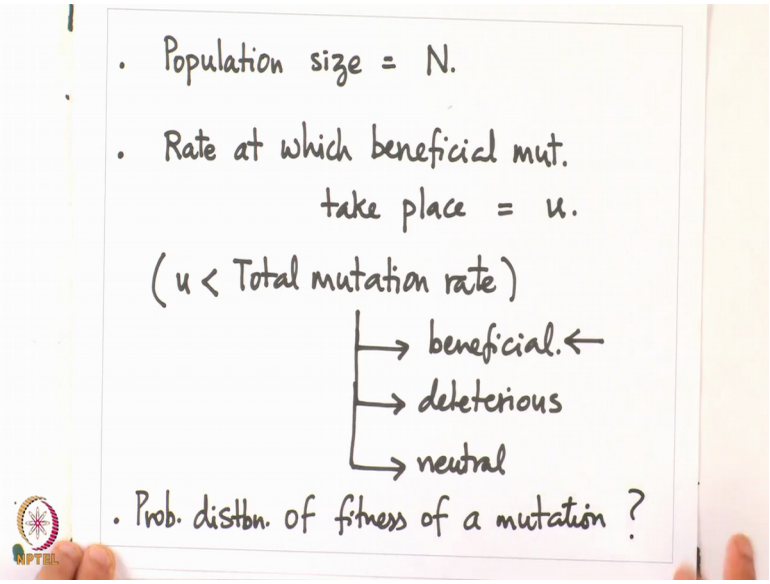
(Refer Slide Time: 01:40)



So, if you just start with the graph that we had drawn last time we know this dynamics of a mutation whenever a beneficial mutation arises it could go to extension which is going to be the case most of the times, but eventually there will be one beneficial mutation which randomly increases to enough numbers such that it gets established. And establishment is said to occur when the number of individuals corresponding to that particular genotype is equal to 1 by S. So, if this is a beneficial mutation which confers a 5 percent advantage or in other words S is equal to 0.05 then this mutation gets established when there are one upon 0.05 number of individuals which is 20 individuals which belong to this particular genotype.

After that the role of genetic drift is minimal the chance that this mutation is now going to be eliminated via drift is very small and then selection takes over and this mutation goes to fixation and all individuals in the population correspond to this particular genotype at this point and there are no individuals which belong to the parent genotype. So, we had defined these 2 times t 1 and t 2, t 1 is the time it takes for the first mutation that is able to escaped drift and go towards establishment. This is sort of the waiting time associated with the process that I have to undergo before I see a mutation get established in the population, and t 2 is the time which a mutation takes to go to fixation once it is established.

(Refer Slide Time: 03:32)



So, we will try and develop a model which goes towards explaining these results quantitatively and if we can come up with what is the speed of evolution that is associated with this particular frame work.

So, to define our frame work we will say that the population size in our model as always remains fixed and is equal to N we say that the rate at which beneficial mutations are taking place mutations take place is equal to small u. You should keep in mind now that this u is less than the total mutation rate because total mutation rate encompasses what are called beneficial mutations these are mutations which lead to increase in fitness, total mutation rate also includes deleterious mutation rates, these are the mutations which lead to decrease in fitness and they also include what are called neutral mutations which do not cause any change in the fitness associated with the bacteria.

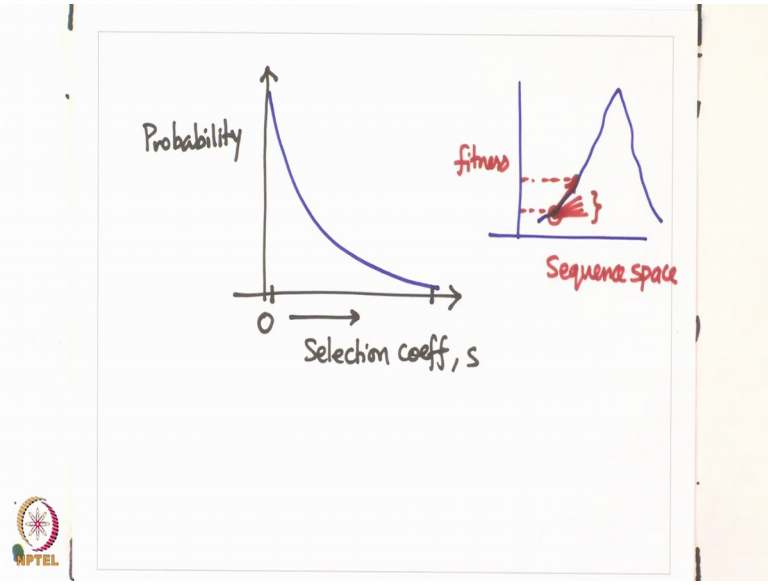
So, we are only interested in beneficial mutations because those are the mutations which are going to get acted upon by selection leading to fixation. Deleterious mutations are going to be acted against by selection and these will be removed from the population very rapidly and we are not going to be discussing neutral mutation in our model. So, the rate at which beneficial mutation takes place is equal to mu which is less than the total mutation rate associated with the population.

And the third very important factor that we have to take in to account is the probability distribution of fitness of a mutation, now what does this mean and that is where we had

ended in our in our last lecture. That the idea being that when a beneficial mutation occurs it might confer a fitness advantage of 1 percent to that individual which is carrying that mutation, but after some time if I wait a certain amount of time another beneficial mutation occurs in the population and this beneficial mutation might confer an advantage of 2 percent to the individual that is carrying it.

Similarly, there might be lots of beneficial mutations that arise in the population and each one of them is associated with a fitness advantage which is unique to itself. So, in our modelling effort when we are trying to understand the dynamics of evolution here; however, we too comment on that when a beneficial mutation occurs what is the appropriate selection coefficient that that mutation confers on the individual that carries it knowing that different beneficial mutations are going to carry different selection coefficients which are associated with them.

(Refer Slide Time: 06:43)



And the graph that we had stopped at last time hopefully you got a chance to think about that is on the x axis is the selection coefficient S and this S is 0 here which means we are only talking of positive S which confers to beneficial mutations and y axis is the probability. And we are interested in the probability distribution associated with this graph that am I what is the likelihood of me witnessing a mutation of selection coefficient this magnitude as compared to new witnessing a beneficial mutation with the selection coefficient of this magnitude.

Now, hopefully you got a chance to think about this, but intuition should tell you if you try to imagine these mutations occur in a process intuitively you should get an image that less beneficial mutations which is beneficial mutations whose S is relatively smaller would be available to the cell in a larger number as compared to beneficial mutations whose selection coefficient is very large. For instance if a cell is at a particular sequence space on a fitness landscape there will be lots of mutations which confer a very small advantage to it, but there will be only very few mutations which confer a vary large fitness advantage to it.

And since mutations occur randomly the chance when that genetic change is made and a beneficial mutation is acquired by an individual, the chance that the selection coefficient associated with that mutation is bound to be smaller as compare to larger. So, there is a greater chance that a beneficial mutation of a lower magnitude happens then a beneficial mutation of a very large magnitude which are comparatively rare in occurrence. So, this graph would look something like this. Again all this graph is trying to tell you is that there is a greater probability of a beneficial mutation of a lower selection coefficient and as you go towards higher selection coefficient the mutation associated with these higher selection coefficients are going to be relatively rare.

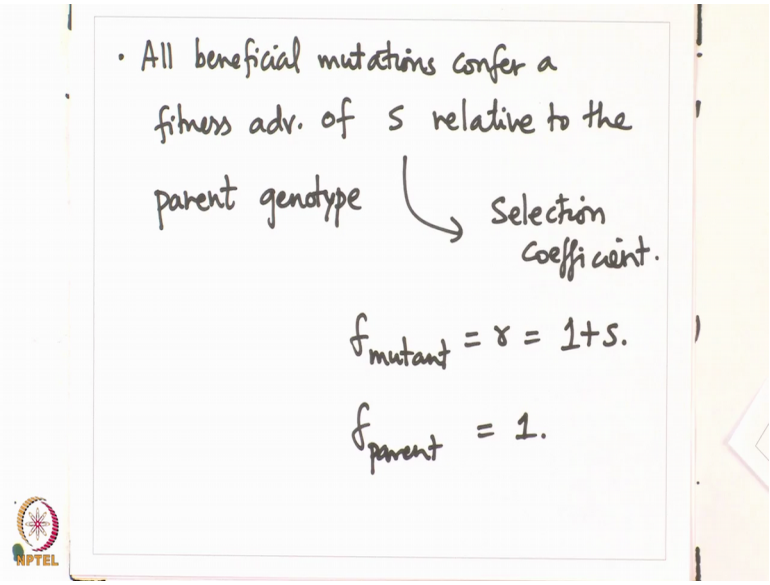
If you imagine this on a fitness landscape, imagine this is the fitness landscape and the sequence where the population is at this point then all these mutations and similarly because this is a three dimensional hill you will have equivalent steps on the other side of the hill. So, you have a large number of mutations which confer only a small benefit amount. So, remember on a fitness landscape this axis represents fitness and this axis represents sequence space. So, all these mutations confer a relatively small benefit to the individual, but there is only one mutation which goes up the hill which confers a relatively larger benefit to the individual.

So, there are more mutations available which confer a smaller fitness advantage and there are fewer mutations available which confer a large fitness advantage to the individual and that is essentially what is represented here in the diagram. So, this is what we have drawn here is understood to be a realistic representation of the frequency of mutations which are available to an individual. We are not going to; again remember that we are only talking of beneficial mutations in the population we are not discussing

deleterious mutations which are going to acted against by a selection and removed from the population immediately.

So, as for as beneficial mutations are concerned this is the distribution that I am interested in, but the problem with this is that in cooperating this in to any model which will make it very very complex and difficult to analyze. So, for our purposes we are going to make a simplifying assumption here that all beneficial mutations conferred the same selection coefficient to an individual which means all beneficial mutations have the same benefit, have the same benefit associated with them relative to the wild type which is which is a big assumption that we are making, but in order to progress with this analysis. We need to make this assumption because otherwise the model would come very unwieldy and difficult to analyze.

(Refer Slide Time: 11:31)

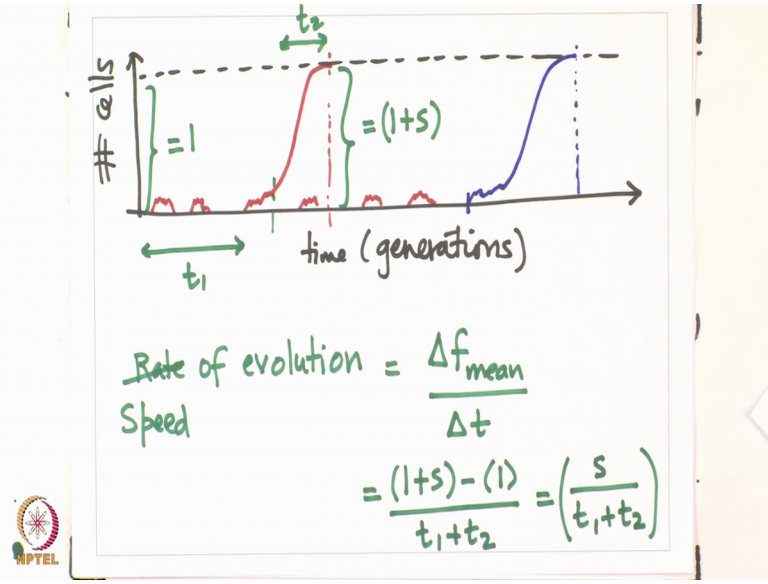


So, we will say that all beneficial mutations confer a fitness advantage of S relative to the wild type wild type or the parent genotype and S, remember is the selection coefficient which means that the fitness of the mutant is equal to r which is equal to 1 plus S and fitness of the parent genotype is just equal to 1. That is a big assumption that we are in cooperated in to our model, but we will see that it leads us to very interesting analysis as we developed this further.

So, with these things what next now. Now we have assumed a population size which is constant equal to N we have a mutation rate associated with beneficial mutations

occurring in this population which is u and we know that every time a beneficial mutation occurs the individual has a selection coefficient S associated with it. Now how is this going to play out?

(Refer Slide Time: 13:06)



Again let us try to imagine the dynamics of this process on the x axis here let us imagine this is time which is again measured a number of generations and on the y axis is the number of cells and let us imagine this is N. So, we are only interested in the region in between.

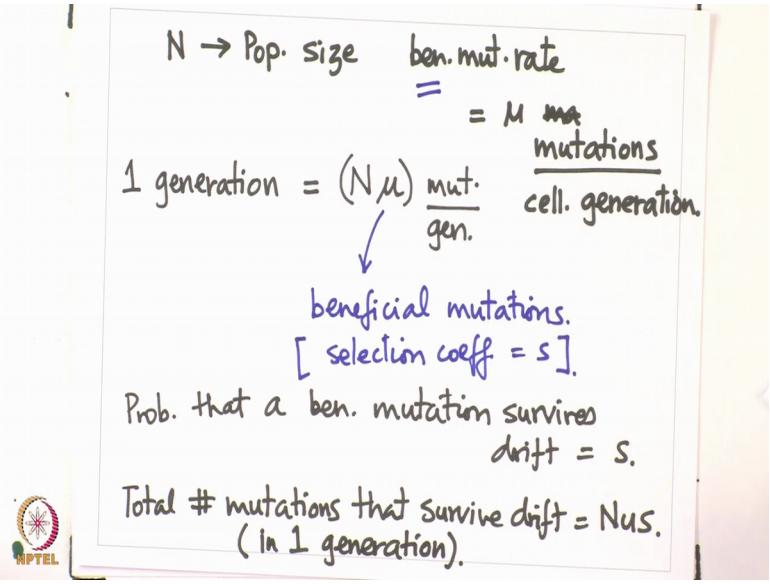
And now from what we have seen what will happen is that a mutation comes and it goes extent another one comes goes extent remember these are all beneficial mutations, we not even representing deleterious mutations here. Eventually a mutation comes which gets established and upon establishment it goes to fixation. At this point now every individual in the population is carrying this mutation and since this has happened other mutations are cropping up in the system and eventually another mutation gets established and that is goes to fixation and now at this point every individual in the population is carrying this particular mutation because all individuals at this point are carrying this mutation and the individuals here also carrying this mutation which got established and fixed. Because the mutation that happened at this point is carrying this red mutation already every individual at this point is carrying the red mutation as well as the blue mutation.

And in this process we note that there are 2 times associated with the dynamics and the time points are first one is t 1 which is the time I have to wait before a mutation gets established and the second time is the time associated with fixation of a mutation . And since the fitness of every individual at this point is equal to 1, fitness of every individual at this point is equal to 1 plus S because the mutation that occurred increased the fitness of the mutant from 1 to 1 plus S and that mutants spread through the population now every individual is carrying that mutation hence fitness of every individual is equal to 1 plus S.

So, the rate of evolution or I should not call it rate speed of evolution which is equal to delta mean fitness of the population divided by delta t it can just be represented as the delta f mean at the end point since every individual has a fitness 1 plus S the mean fitness is also 1 plus S. So, this is 1 plus S minus the initial fitness since every individual was at a fitness level 1 the mean fitness at the beginning of the experiment is 1 and this change happened in time t 1 plus t 2. So, the speed of evolution is just given by S divided by t 1 plus t 2, now what is. So, the that is the speed of evolution in a context such as this, but what I still do not understand are what are the values associated with t 1 and t 2 and that is what we are going to find next.

I will keep this picture here and let us try and understand, let us try with t 1 first remember t 1 is the waiting time associated with a mutation occurring in the population and establishing itself. So, let us try and get, let us try and get an expression for t 1 first. How much is t 1? So, if we look at our graph here t 1 is this waiting time, I know the populations size associated with my system is equal to N and the beneficial mutation rate is equal to u mutations per cell per generation.

(Refer Slide Time: 17:20)



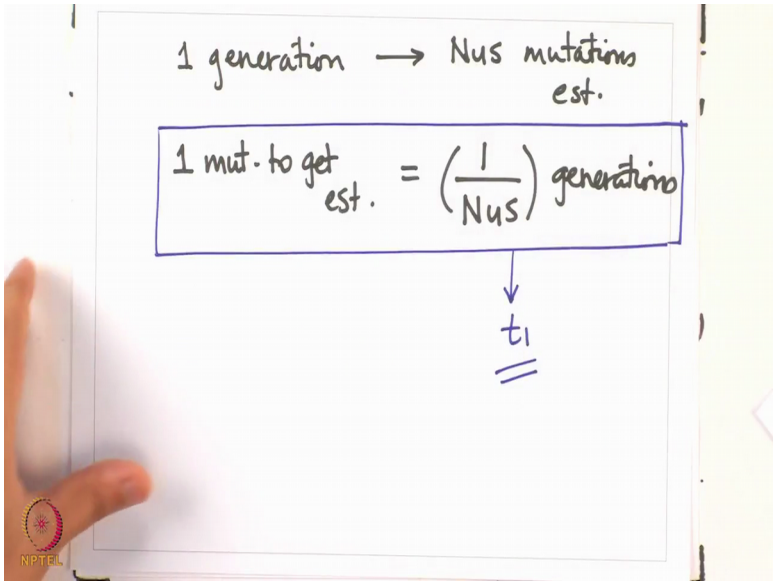
What that means is that every time one cell divides once which is one cell one generation the likelihood that a mutation arises is mu, which means in one generation if there was one cell dividing you have mu mutations occurring in the population, but since you have N divisions taking place in one generation you have N mu mutations per generation.

So, you have N mu generations occurring per generation and remember these mutations are not comprising of all mutations this is the mutation rate which is associated with beneficial mutations only. So, these are N mu beneficial mutations occurring per generation. Each of these mutations has a selection coefficient S associated with it. So, of the total N mu mutations that are occurring in this one generation time what we want to find out next that how many actually survive in that one generation time because as we have seen most beneficial mutations are not going to be able to survive drift and are just going to go extent and lost from the population.

So, of the N mu number of mutations that have occurred what fraction survive, but we already we already know this result because we know that the probability that one beneficial mutations survives drift is just equal to S. So, if one beneficial mutation survives with a probability S how many are going to survive if the total number of mutations that are occurring is N mu. So, in one generation time total number of mutations that survive drift is just equal to Nus and I should re emphasize that this is in one generation time only. So, in one generation time I have Nus number of mutations

which survive drift and get established, but what I am interested in my analysis is that how much time do I have to wait before the first mutation gets established whereas, what I found here is that in one generation time how many mutations got established what I want is how much time in terms of number of generations do I have to wait for the first mutation to get established.

(Refer Slide Time: 21:16)

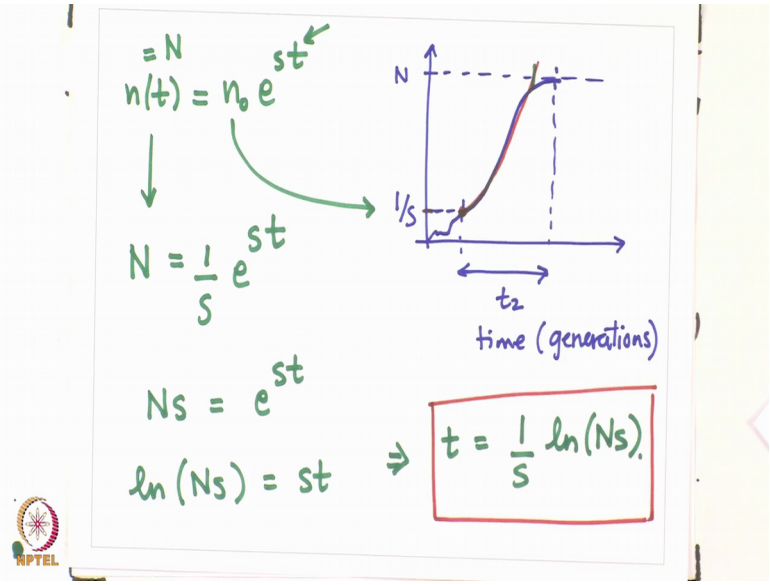


That is petty straight forward because what I found is that if I wait one generation I get Nus number of mutations which get established; that means, what I am interested in is how much time do I have to wait for 1 mutation to get established that is just going to be equal to 1 upon Nus number of generations. So, that is the time I have to wait for the mutation to get established for the first mutation to get established an important point to keep in mind at this point is that these are average times this is the time you have to wait for the first mutation to get established on average.

If you were to do this evolutionary experiment sometimes the first mutation might get established, on the other hand sometimes one out of 20 mutations. None out of 20 mutations could get established all might get lost due to drift and you have to wait 40 beneficial mutations for the first one to get established. But this particular number represents that if you were to do this experiment a very large number of times what would be the average number of generations that you have to wait to see the first mutation go towards establishment.

So, we have an expression for one of the times and if we compare it with the graph that we do earlier this is the waiting time before the first mutation gets established. So, this is an expression for t 1. So, I have t 1 now. How do I get t 2? That now after this mutation gets established it goes towards fixation that is what we will work towards next. So, a mutation getting established is as we have already discussed if the selection coefficient associated with that mutation is S then it is said to be established if the number of individuals belonging to that particular genotype is at least 1 by S.

(Refer Slide Time: 23:40)



So, what I am interested in now is this understanding this dynamics that once a mutation is established; that means, its numbers are 1 by S its going to increase deterministically and then reach 1 by S and I am interested in this time where x axis represents time in generations. So, how do I do that? Again we are going to realize on some simplifications and assumptions and always keep this at the back of your mind that we are interested in, we are interested in order of magnitude estimates of these numbers. We are not interested in precise values, but I just a ball part estimate because we want to understand the overall behavior associated with these phenomena. So, if we look at our graph carefully what we can see that after establishment we are right here we are at this point.

And from this point onwards the graph follows a more or less exponential trajectory corresponding to the increase in the number of mutants number of individuals which belong to the mutant genotype. So, if I were to assume just an exponential growth that is

not too bad in estimate that I have when I am trying to compare time t 2 and we already done in one of the first few lectures we had talked about differential calculus and we had talked about growth equation in an exponential growth in an exponential fashion for a very long time and we saw that very soon this exponential growth cannot hold down because there is no environment on this planet which can support exponential growth for a very long time.

But for the system that we are talking about just reaching the number of mutants equal to N exponential growth is not a bad estimate in the sense that we are only interested in order of magnitude estimates. So, using exponential growth we know that numbers vary according to this relation n t is equal to n naught e to the power st where this is number of individuals at any time t this is the starting number of individuals and s is the growth rate or in this case selection coefficient associated with the particular genotype.

In my case I am interested in finding out the time at which n t becomes capital N that is the time that I am interested in. So, I am going to plug capital N here, n naught is the starting number of individuals in my exponential curve and it is easy to see that N naught is just equal to 1 by S. So, you have 1 by S, e to the power s and t, s is the selection coefficient associated with this mutant and t is the variable that I am interested in finding out.

So, now, if I have this and I process this further I take S over to the other side and I get ns equal to e to the power st then I take log of both sides. So, I get natural log of N s equal to st which implies the t is equal to 1 upon S times ln of N s right that is my estimate for time t 2. Again I do this with the knowledge that this is not exact because the actual growth rates slows down a little bit as I approach N, but again I aware that we are only interested in order of magnitude estimates of these quantities and now precise values hence this is a good enough estimate.

So, we stop here in this class and next we compare t 1 and t 2 and we calculate the rate of evolution associated with this whole process and we will see that what this graph tells us and more importantly what this relationship does not tell us in terms of how physiologically a real evolutionary experiment is being played out.

Thank you.