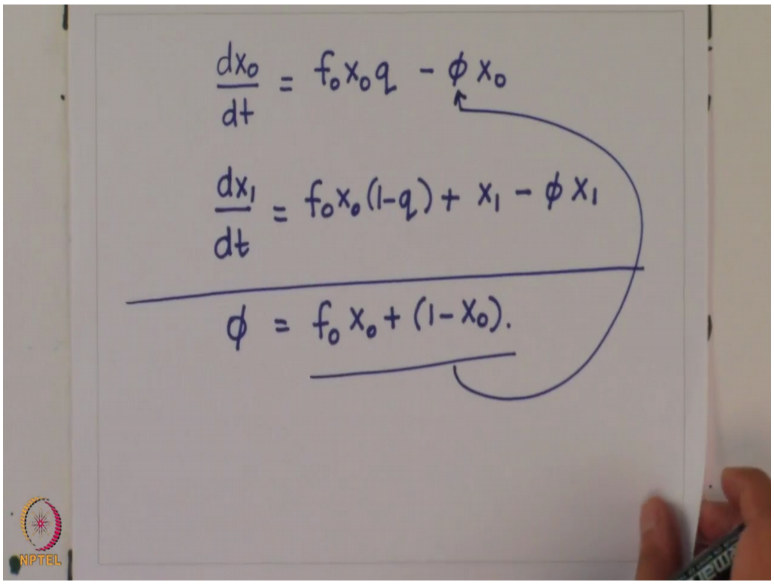
Introduction to Evolutionary Dynamics
Prof. Supreet Saini
Department of Chemical Engineering
Indian Institute of Technology, Bombay

Lecture - 21
Modelling evolution on fitness landscapes – 3

Hi and let us continue our discussion of localization of Quasi Species on a Fitness Peak. When we are changing the mutation rate associated with the process and where we had stopped our last lecture was we had developed equations for d X naught by d t and d X 1 by d t, where X naught represented the frequency of individuals which belong to the fittest genotype and X 1 represented the frequency of individuals of all other genotypes, and this one was in these equations were in conjunction with the fitness landscape. We have developed where we say that the fittest in sequence corresponds to a fitness of f naught and all other sequences on this landscape are have a fitness equal to 1 and this highly idealized representation of a fitness landscape in a binary DNA sequence is sort of necessary to be able to develop this analysis further which will lead us to hopefully valuable biological insights.

So, let us take a look at the equations that we had developed and see how we can get some insights from them.

(Refer Slide Time: 01:25)



$$\frac{dx_0}{dt} = f_0 x_0 q - \phi x_0$$

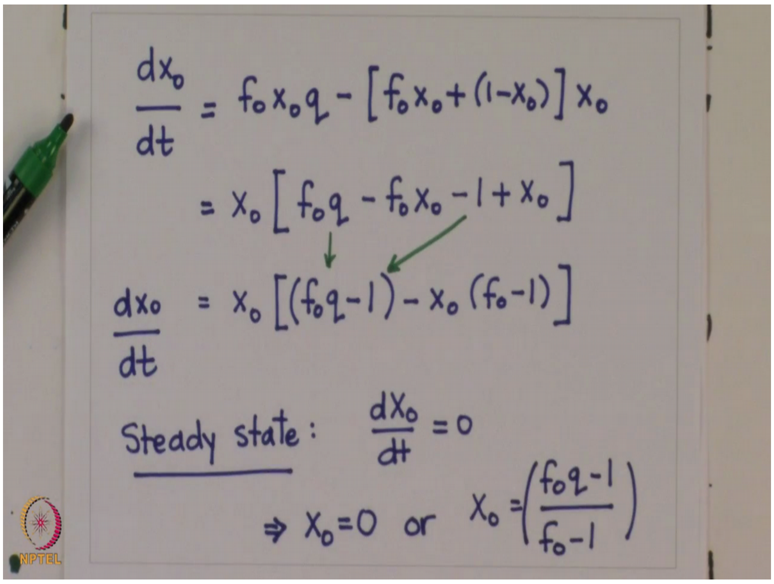
$$\frac{dx_1}{dt} = f_0 x_0 (1-q) + x_1 - \phi x_1$$

$$\phi = f_0 x_0 + (1 - x_0).$$

We have 2 differential equations d X naught by d t which represents the rate of change of frequency of the fittest individual and d X 1 by d t which is the rate of change of frequency of individuals which belong to all other genotypes except the fittest 1, and phi which is the mean fitness can be represented via this relationship. That is just using the fact that X naught plus X 1 is equal to 1 at any given point in time because the population size is fixed.

Now, what we are going to do is use this definition of phi and substitute this definition here and see what the equation tells us.

(Refer Slide Time: 02:11)



$$\frac{dx_0}{dt} = f_0 x_0 q - [f_0 x_0 + (1-x_0)]\, x_0$$

$$= x_0 [f_0 q - f_0 x_0 - 1 + x_0]$$

$$\frac{dx_0}{dt} = x_0 [(f_0 q - 1) - x_0 (f_0 - 1)]$$

Steady state: $\frac{dX_0}{dt} = 0$

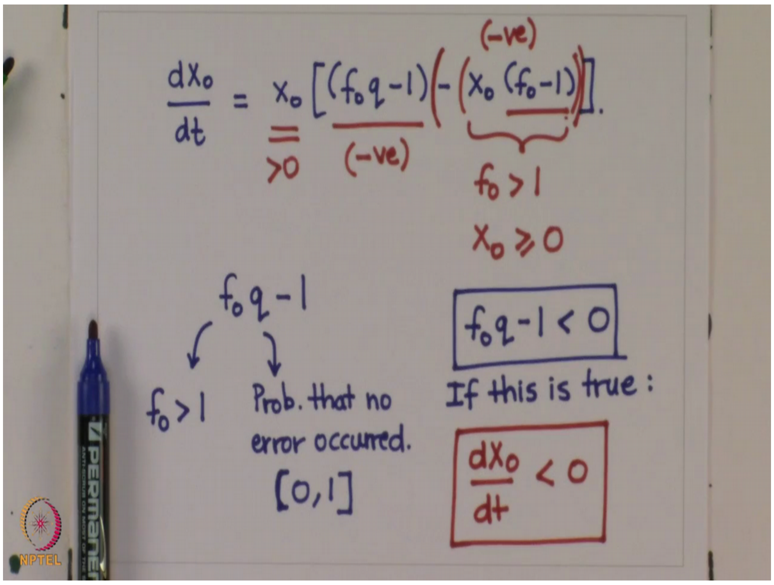
$\Rightarrow X_0 = 0$ or $X_0 = \left(\frac{f_0 q - 1}{f_0 - 1}\right)$

So, when we do that, we get d X naught by d t equals f naught X naught q minus f naught X naught plus 1 minus X naught that is equal to phi times x naught. The three terms involved here all have a factor X naught common to them. So, I can take X naught out and what that leaves me is f naught q minus f naught X naught minus 1 plus x naught. That is what I am left with when I take the X naught factor out of this equation and then, I can write this as X naught times f naught q minus 1. I am doing this taking these 2 terms together and I want to combine these 2 because they carry another factor common which is X naught and arrive this as minus X naught f naught minus 1. So, that is my equation for d X naught by d t.

So, again we will go ahead with our analysis consistent with the approach that we have been taking earlier on in the course and the first thing that we can do is analyze the steady state associated with this equation and that steady state can be obtained when we put d X naught by d t equal to 0 which gives me that this expression is equal to 0 when either X naught equal to 0. So, X naught equal to 0 immediately leads me to 0 times this whole quantity which is 0 or the quantity within the brackets is equal to 0 which is true when X naught is equal to f naught q minus 1 upon f naught minus 1. So, that is system has 2 steady states and I leave this as an exercise for you to comment on the stability associated with these 2 steady states after assuming certain values for f naught and q which are constants in the system of equations that we have developed here.

So, what else can we say about these systems? Let us take a look at the d X naught by d t equation that we have developed and see if that tells us something more.

(Refer Slide Time: 05:20)



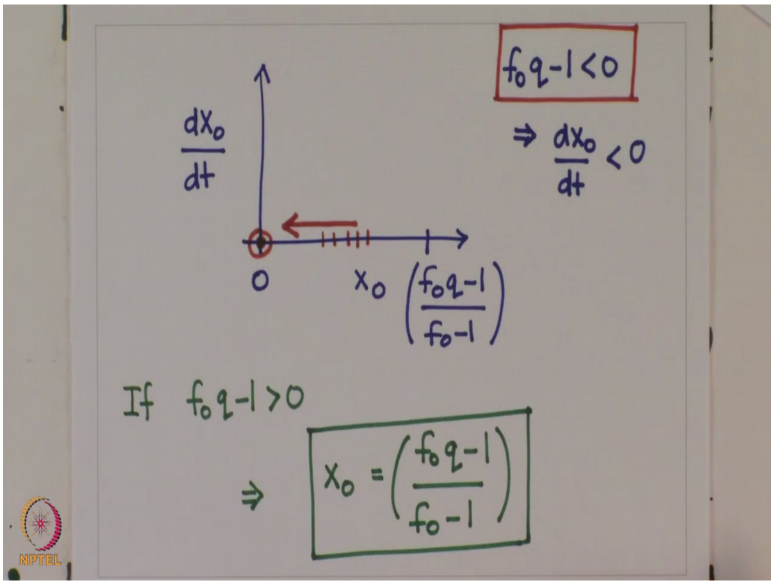
Again we just write that equation down quickly, d X naught by d t is equal to X naught times f naught q minus 1 minus X naught times f naught minus 1. If this is the case, if this is what we are dealing with here, what is the sign of this particular term? What is the sign in whether it is greater than 0, equal to 0, more than 0 of X naught times f, naught minus 1? The way we have defined our system, we have already said that f naught is always greater than one. So, this quantity is always positive. X naught is frequency of individuals of a particular genotype. So, X naught cannot be negative. X naught is either 0 or it is positive. So, this is greater than equal to 0. What that tells us is that this quantity, this is a positive quantity and this is the positive quantity. So, the relative contribution which is coming from here is always going to be negative. This whole quantity is positive; hence the relative contribution of d towards d d X naught by d t which comes from this entire expression is negative.

Now, if on the other hand what is the other term that is equal to f naught times q minus 1? Can we comment on the sign of this? This term we cannot at this point because we know that f naught is more than 1, but we do not know q and remember q is the probability that when a genome is replicating, when this DNA sequence of length L is replicating the progeny is the same, is the identical sequence as the parent and no mutation happened in the replication process. So, this is a probability that no error occurred.

If this is the probability; what then? That means, this is the number between 0 and 1 which means f naught f naught q minus 1 could be greater than 0 or could be less than 0 depending on what is this value q because f naught for sure is more than 1, but let us imagine this scenario such that f naught times q minus 1. If this is less than 0, if this holds true, if f naught times q minus 1 is less than 0, what that means is that in the d X naught by d t equation, this quantity is positive. This quantity is negative because we are dealing with the situation, where f naught minus q f naught q minus 1 is less than 0 and this entire quantity is starting with this negative sign going up to the end of the expression is also negative.

So, what that means is if this condition is satisfied, then the sign associated with d X naught by d t is less than 0. What does that mean in terms of the system moving in a particular direction? What that means is, remember recall from the discussion of our stability analysis of dynamical systems if d X naught by d t is less than 0, what that tells us is that as time increases, the value of X naught decreases and this d X naught by d t is less than 0 independent of what is the value of x naught. X naught could take on any value that it chooses to between 0 and 1. So, for any value of X naught as long as f naught q minus 1 is less than 0 and this expression does not have anything to do with x naught.

(Refer Slide Time: 10:23)

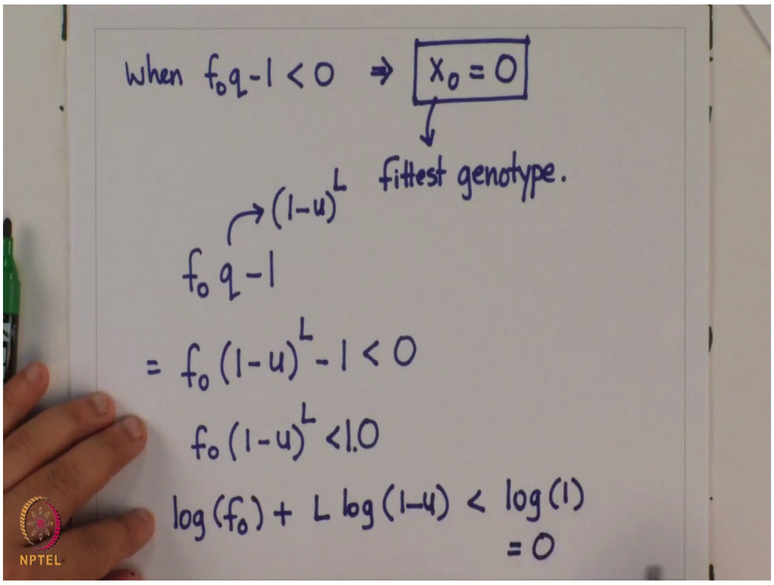


So, X naught can take any value as long as this satisfied d X naught by d t is less than 0 and as we had seen in the discussions of stability that if you have a system, this is x naught, this is d X naught by d t 0 and the other steady state that we had derived was f naught q minus 1 divided by f naught minus 1.

What this results tell us is that as long as f naught q minus 1 is less than 0, this implies that d X naught by d t is less than 0 which means you could be anywhere, you could have any value of x naught, but as time increases, X naught decreases because d X naught by d t is less than 0 and hence, the population moves in this direction and it only stops as X naught approaches 0 because when X naught at the time X naught is equal to 0, d X naught by d t automatically goes to 0 because then you have a number which is 0 times and negative numbers which is equal to 0 d X naught by d t in that case turns out to be 0, and the system stops at the value X naught equal to 0 which makes biological sense too because X naught values, X naught cannot keep on decreasing even after it has achieved the value of 0. X naught is frequency of individuals which belong to a particular genotype after you once you say that X naught is equal to 0. That means that there are no individuals belonging to a particular genotype, that is that species is g 1. Now, X naught cannot have negative values and keep on decreasing further. So, X naught equals to 0 as long as this condition is satisfied, that is the stable steady state associated with the system when this quantity is met and on the other hand, if this quantity is greater than 0, then the steady state.

The second case is if f naught q minus 1 is bigger than 0, then the system stabilizes at X naught equal to f naught q minus 1 divided by f naught minus 1. So, we have derived these results, but if you want to understand this, this is mathematics and we have derived some results relating to X naught depending on f naught and the values of q associated with the process, but we want to understand the biology associated with these results.

(Refer Slide Time: 13:35)



When $f_0 q - 1 < 0 \Rightarrow \boxed{X_0 = 0}$

$\downarrow$ fittest genotype.

$f_0 \overset{(1-u)^L}{q} - 1$

$= f_0 (1-u)^L - 1 < 0$

$f_0 (1-u)^L < 1.0$

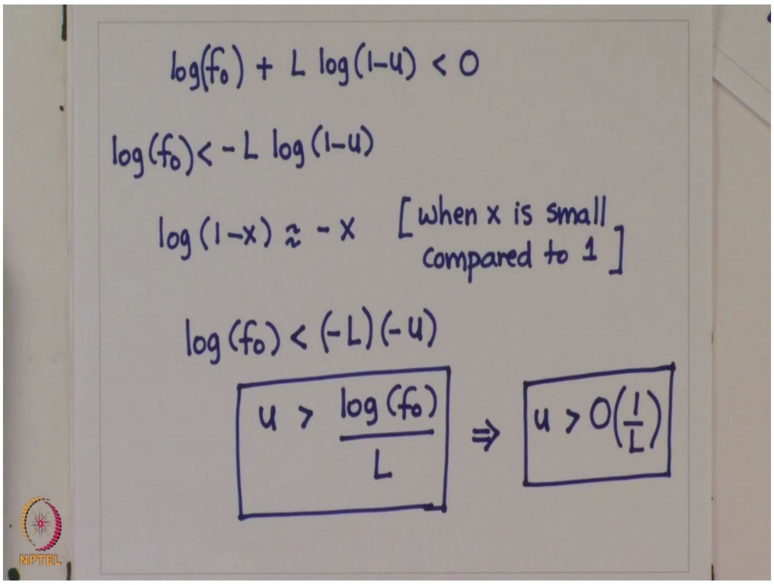
$\log(f_0) + L \log(1-u) < \log(1)$

$= 0$

To do that let us take a look at what is happening here. We have f when f naught q minus 1 is less than 0. So, let us talk about the case when f naught q minus 1 is less than 0, what is happening in this case is that mathematics is telling us that the steady state associated with this system, the stable state associated with the system is X naught equal to 0 which is a surprising result because X naught remember is the fraction of individuals which belong to the fittest genotype. What we are saying that when this quantity, when this relationship holds the number of individuals which belong to the fittest genotype on my fitness landscape goes to 0, that particular genotype is no longer around which is surprising because you would expect that the fittest landscape is present in the genome that should act and ensure that the fittest genotype is always present in the environment. What is happening f naught times q minus 1 can be written as f naught into what is the value of q or the expression associated with q. Q is the probability that when replication of an individual’s DNA occurs, the progeny has the identical DNA and if u is the error probability of making an error at a single nucleotide, the value of q is 1 minus u to the power L. So, this relationship tells us that f naught into 1 minus u to the power L minus 1 should be less than 0 for this result of the dynamics of the system. If that is the case, we can rewrite and simplify this further and say that f naught into 1 minus u to the power L should be less than 0 and then, we take logarithm on both sides and we get, sorry this is 1 because I am taking the 1 to the other side. This should be less than 1.

Logarithm of f naught plus L times logarithm of 1 minus u should be less than logarithm of 1 which is a quantity equal to 0.

(Refer Slide Time: 16:26)



$$\log(f_0) + L\log(1-u) < 0$$

$$\log(f_0) < -L\log(1-u)$$

$$\log(1-x) \approx -x \quad \left[\text{when } x \text{ is small compared to } 1\right]$$

$$\log(f_0) < (-L)(-u)$$

$$\boxed{u > \frac{\log(f_0)}{L}} \Rightarrow \boxed{u > O\left(\frac{1}{L}\right)}$$

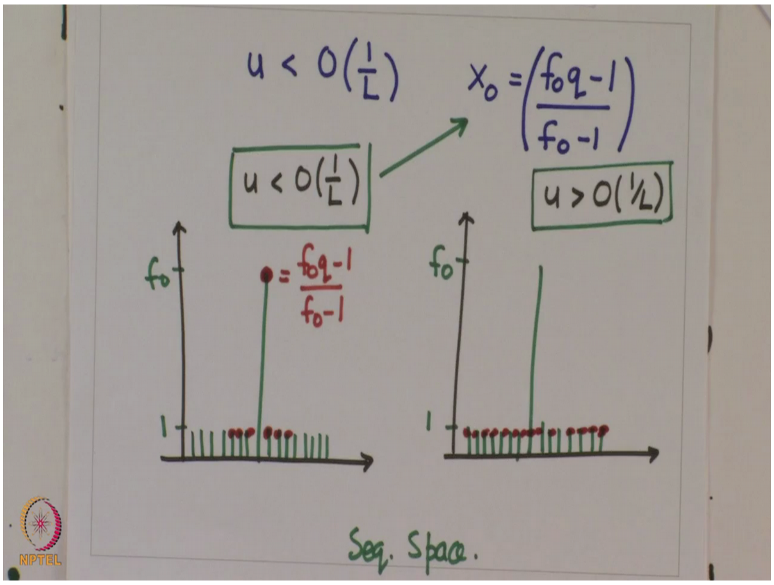
We can simplify this a little bit more and get to a result which says log of f naught plus L times log of 1 minus u should be less than 0 or L times log of 1 minus u like this is log of f naught should be less than minus log of 1 minus u and now, I am going to use a very simple property of logarithms which is that log of 1 minus x is approximately equal to minus x, provided when x is small compared to 1. So, as compared to 1 if x is very small, this is roughly equal to minus x and that gives me log of f naught minus L times minus u which means u is greater than log of f naught divided by L which implies that u is greater than order of magnitude 1 by l.

This is extremely interesting and important result that we have derived here. I am dropping the term log of f naught and just writing u as greater than the order of magnitude because that is what we are interested in the derivation that we have derived because we have already made so many assumptions and we just want to get an estimate of how does this population behave when moving on a fitness landscape, and we are not interested in precise mathematical answers. So, this expression says that u when bigger than 1 by L and remember what u is. U is the probability that when 1 nucleotide is dividing, it is not replicated correctly and error is made at a single nucleotide position. So, if u is bigger than 1 by L, that is when you have collapse of population happen and you no longer get localization of population around the peak which was the premise which we had started with.

Remember the question. The original question that we had started this entire discussion was based on the fact, based on this understanding that as u increases, quasi species should get localized, should be localized around the concept, around the fitness peak of the landscape, but as u increases, their distribution around the peak should get widened and we wanted to understand if there exists a value of mutation rate u beyond which that localization collapses and we get a flat distribution everywhere in the genome and that is what this expression is telling us.

What this is telling us is that when u increases beyond a certain threshold, then the fraction of individuals which belong to the fittest landscape, which belong to the fittest sequence approaches 0. That means, there is no localization around the sequence which corresponds to the maximum fitness in the landscape and you get a flat distribution of individuals for all genotypes because mutation is happening at such a rapid rate that it is dominating over selection and you do not get any localization around the fitness peak.

(Refer Slide Time: 20:46)

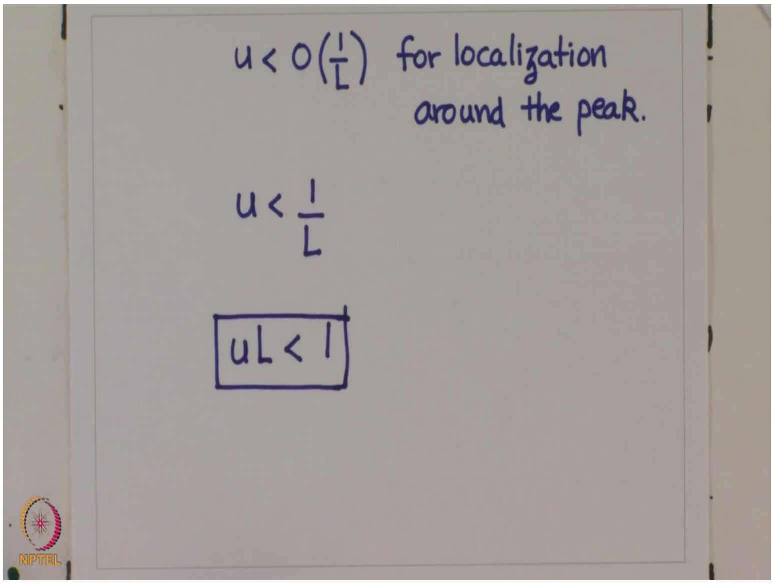


On the other hand, if u was less than this quantity, if u is less than order of magnitude 1 by L, then we have the result that we just derived that X naught is equal to f naught q minus 1 divided by f naught minus 1 in terms of the landscape that we had drawn this picture for.

Let me represent these 2 cases of high and low mutation rate. This is the case, where u is less than order of 1 by L mutation rate and this represents the other case, where mutation rate goes beyond 1 by L order of magnitude. What happens can remember our fitness landscape is identical. Its fitness is 1 everywhere and you have the fittest sequence and then, the fittest sequence corresponds to a fitness of f naught. Everything else is at fitness 1.

The x axis represents the sequence space. What is happening in this case when error rate is below the threshold is that you get a finite X naught distribution. The number of individuals which correspond to this particular sequence, the fittest sequence, the fraction of individuals which belong here is equal to f naught q minus 1 by f naught minus 1 and then, there is a distribution around this peak as in you have some individuals here and the low mutation rate ensures that the population is centered around the sequence which is corresponding to the maximum fitness on the landscape. However, when the error rate crosses this threshold and you have a very high error rate, this localization disappears and you get equal distribution of individuals spread throughout the landscape and the number. You do not get any localization of population centered around the peak of this fitness landscape. So, that is the transition that has happened as the mutation rates associated with a species cross a particular threshold and it is very interesting to note that threshold that we have come up with having talked about a highly simplified version of fitness landscape that we have defined.

(Refer Slide Time: 23:45)



$u < O(\frac{1}{L})$ for localization around the peak.

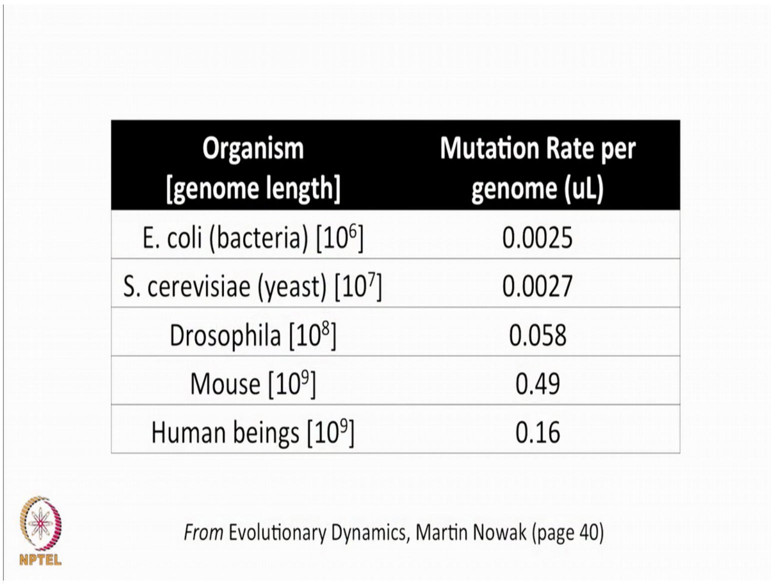
$u < \frac{1}{L}$

$\boxed{uL < 1}$

Let us look at that definition of the threshold and see what does that tell us. The threshold for mutation rate that we have come up with is u less than order of magnitude 1 by L for localization around the peak. Again we are just interested in order of magnitudes. This can be written as u less than 1 by L. This is an assumption. In reality there could be a factor of 2, there could be a factor of 5 or 0.5, but we are interested in some very engineering type approximations and trying to understand the behavior of the system and what it is trying to tell us.

So, I am going to assume that u less than 1 by L is the threshold, where we ensure that localization of population around peaks of fitness landscapes are still happening. If that is the case, then what the threshold tells us is that u times L should be less than 1 for localization. That is a constraint that population should maintain, should respect if they do not want collapse of populations from fitness landscape and population should represent quasi species centered around genomes which correspond to maximum fitness in them.

(Refer Slide Time: 25:20)



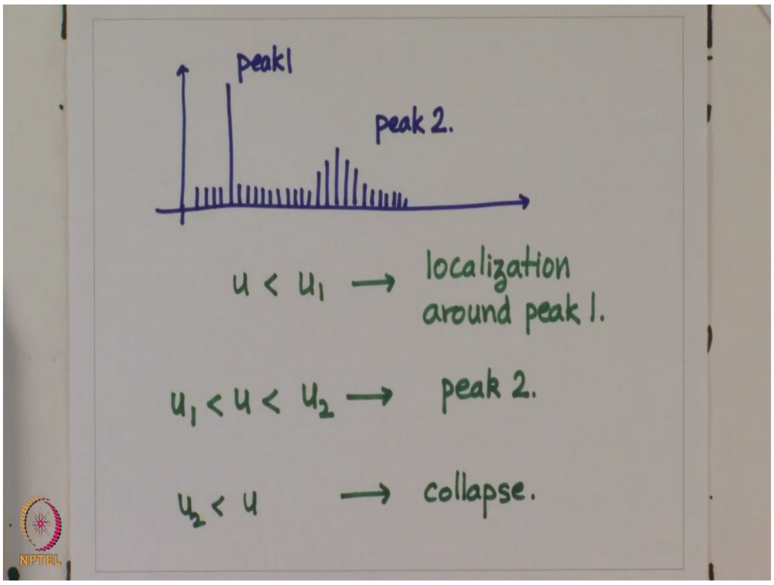
Organism [genome length]	Mutation Rate per genome (uL)
E. coli (bacteria) [10^6]	0.0025
S. cerevisiae (yeast) [10^7]	0.0027
Drosophila [10^8]	0.058
Mouse [10^9]	0.49
Human beings [10^9]	0.16

From Evolutionary Dynamics, Martin Nowak (page 40)

If you switch to this slide that we have picked up from our textbook, what we show here is five different species. We have representations from bacteria yeast and some eukaryotes and given here are the corresponding lengths of their genomes. These are again order of estimates of genomes and what the column on the right hand side represents is what is the value for u times L for each respective species that we have listed in this table here, and what can be seen here is that for all five species and this table can be extended to incorporate other species to what consistently emerges that u times L for each of these species is less than 1 which is something remarkable compared to, if you compare this result let us go back to our notes. This result that we observe for pretty much all life form that u times L turns out to be less than 1 is remarkable when you consider that having assumed such a simplified version of a fitness landscape, we are able to get arrived at that result and that is so strictly adhered to by nature in all life forms. They do exist a couple of exceptions to this rule. Some viruses have u times L values associated with them which are greater than 1 and the precise reasons for those existences of such exceptions to this rule are not quite well understood.

So, again u times L represents a threshold below which, well as long as the error rate is below that quantity, you get localization. If the error rate were to cross that threshold, you do not get localization and you get collapse of population.

(Refer Slide Time: 27:14)



Another result because fitness landscapes are really single peaked and we also know it can be shown if you have a fitness landscape which has a narrow peak and then, a smaller peak, but broader in nature. So, let ud call this peak 1; let us call this peak 2. If that is the nature of your fitness landscape which is more represented of the fitness landscapes that you would actually encounter in biology, then it can be shown that if the error rate associated with replication is less than certain mu1, then localization happens around peak 1. If the error rate however is more than u 1, but less than a certain value u 2, then localization happens around peak 2.

So, if replication is very faithful, then the species would center around the peak which offer some higher fitness. If the error rate is high, then a collapse could occur as far as this peak is concerned, but this peak is broad, so it is able to contain the quasi species within it and you get localization around the lower peak, but if error rate were to exceed the value u 2 as well, then you would get a collapse of the population and see no localization around any of the 2 peaks here. The result that u L is less than 1 is something that has very important biological implications and there are a couple of different ways that we can think about this result in terms of the nature of mutations that happen when replication occurs and we will revisit that in a few lectures time, and we will try to explain that why should this be expected if we were to think about the nature of mutations that happen in an intuitive sense.

So, this closes our discussion on fitness landscapes. We will start our next lecture and will start with the topic of Genetic Drift and Randomness in Biological Population.

Thank you.