Introduction to Developmental Biology

Prof. Subramaniam K

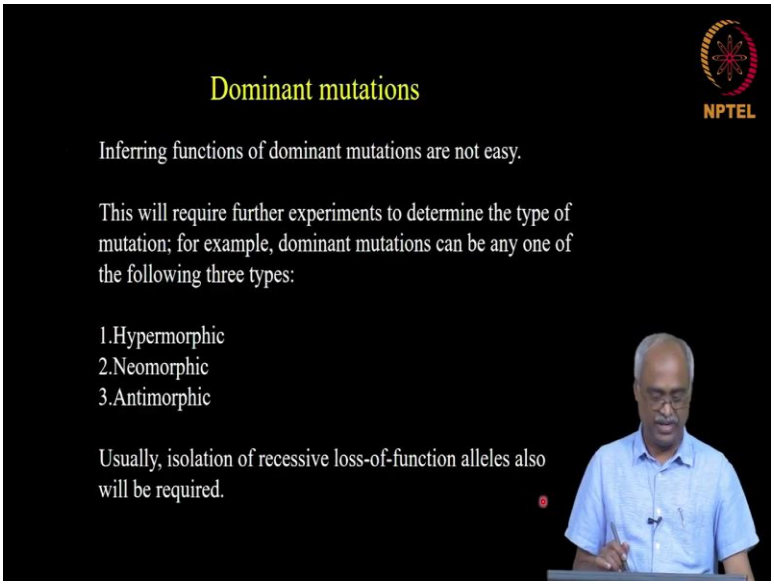
Department of Biotechnology

Indian Institute of Technology - Madras

Lecture-11

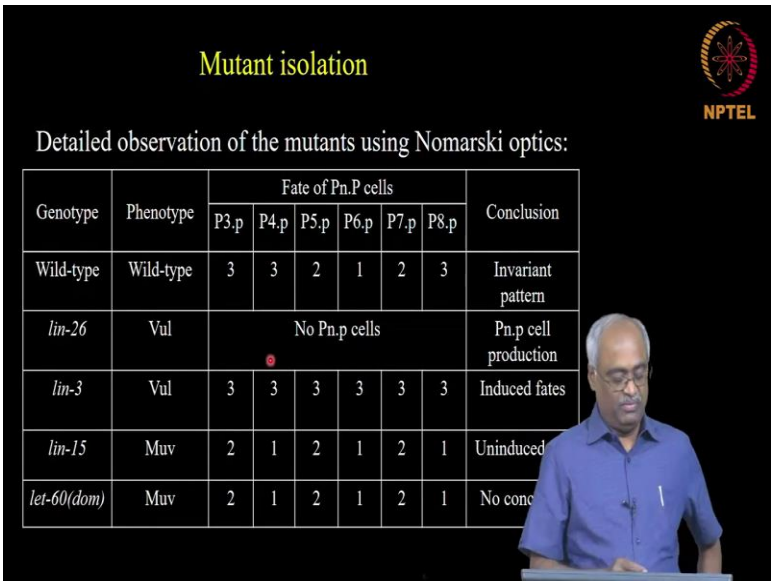
Genetic basis (Part 4 of 5)

(Refer Slide Time: 00:16)



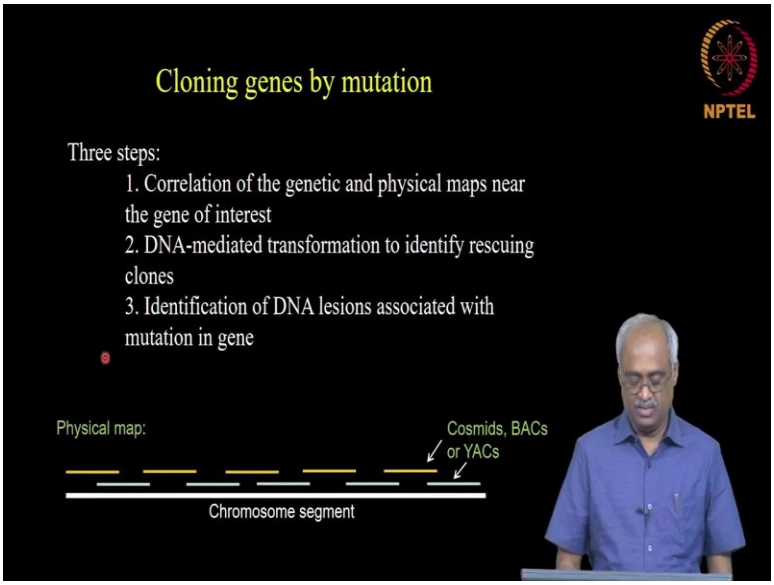
So, yesterday we discussed the mutagenesis screen and genes that are involved in vulval development. One of the genes was dominant, and therefore, we discussed that we could not immediately conclude the normal function of that gene for that particular tissue development. Suppose you want to understand how a specific biological system functions, then that particular gene function should be removed. So only a loss-of-function allele can reveal the function of a gene. To understand the development of an organism knowing the roles of specific genes is essential.

(Refer Slide Time: 02:47)



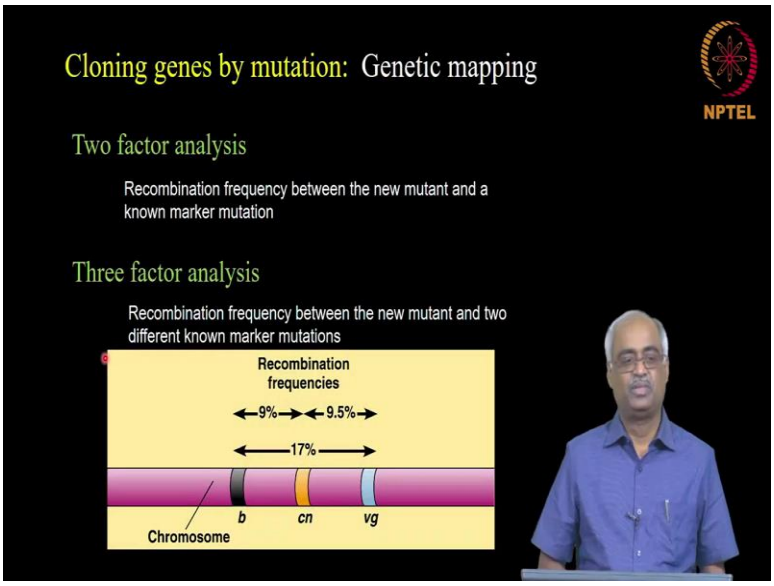
Now let us discuss the mutants that were isolated from the screen. We found genes that are lineage defective or *lin*, and one such is *lin-26*, it is the 26th gene in the order in which it was identified. But this does reveal its role in vulval development. So to understand its function, firstly, we need to identify the chromosomal location of the gene. So that process is called a genetic mapping, so mapping a given mutant. So, now let us focus on how that is done.

(Refer Slide Time: 03:37)



There are multiple ways of mapping; I will go through an the most common way that is done in standard genetics.

(Refer Slide Time: 03:53)



So, first, we need to find out on which chromosome that particular mutant allele is located. So, for that, we can take a known mutant allele of a gene; whose chromosomal location is already known. So, in our particular example, since we have six chromosomes, we take six different markers, each one well known for its position on let us say, chromosome 1, 2, 3, 4, 5, and X.

Now using Morgan's experiment, we can try to see the recombination frequency between the new mutant allele and the known mutant allele. So based on the recombination frequency, if they are independently assorting, then they are not on the same chromosome. If they show linkage, Let us say *lin-26* shows tight linkage to a gene located on chromosome 1 and independently assorts with the rest of the five, then you know *lin-26* is probably on chromosome 1. So this is called two-factor analysis, where you find the recombination frequency between the new mutant and the known marker mutation. It is the same as what we had already discussed when we learned about Morgan's experiment with the black and vestigial. Now we know that the chromosomes have millions of base pairs. *C. elegans* has about 100 million base pairs divided into six chromosomes; each one of them is having more than 10 million base pairs. An average gene is under 2 kb, so you need to know out of the 10 million bases which 2 kb is your gene. So, there you should try to narrow down within the chromosome to a particular genetic interval, meaning you can select two known positions. The distance between the two positions is called as the genetic interval. So, for that, the three-factor analysis should be done. This is a very useful analysis; it gives a very specific position for your gene at the end of the three-factor analysis. It is the same as the two-factor analysis except that you are determining the recombination frequency

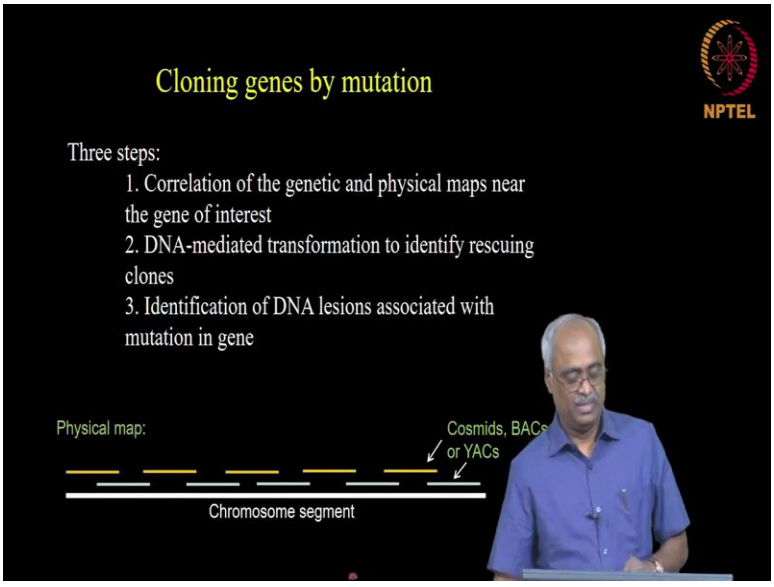
between your new mutant allele and two different known markers located on the same chromosome at known intervals.

Now let us take Morgan's example, here we determined the recombination frequency of black and vestigial is 17%. Now, this 17% can be on either side of black. One percentage is considered as one centimorgan in memory of Morgan. So recombination frequency does not tell the side of a given reference point. Now let us take **cn**, the recombination frequency of black is 9%. Now, this 9% could be on the left or right of **cn,** and the recombination frequency of vestigial is 9.5%. This tells that **cn** is between the two, and vestigial is on the right of black.

So, you give a specific position for the known new mutation, and here we are assuming black and vestigial as known markers, and the **cn** is the unknown. So this is called three-factor analysis. So, the three-factor analysis gives a very specific position, but the resolution and the accuracy of your conclusion depend on the number of progeny you observed. So higher the number higher the accuracy. So, once that is done, then we need to do a giant leap that is earlier, they had to continue this.

Like what do you do is first, you take markers that are far apart and therefore, your allele is somewhere in between; then you select newer known markers that are closer, by repeating it then you narrow down to a small region. So, that is very tedious, and that is why gene mapping took a few years to map a gene. But nowadays, we do it in a month or so if you are entirely on it.

(Refer Slide Time: 10:32)

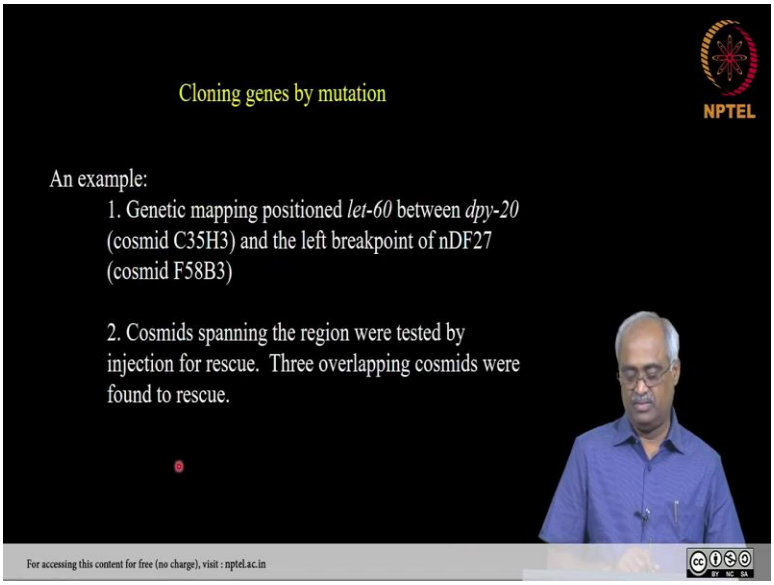


So, before we go into this, I need to introduce what is called a genetic map and a physical map. So, a genetic map is purely an imaginary map based on recombination frequency. There is no physical basis for it whereas the actual DNA sequence is called the physical map.

So historically, people developed methodologies to determine the *C. elegans* genome sequence. And one of the methods was breaking down the chromosome into smaller pieces and cloning them in vector. It is the same concept as the usual plasmid except that in the plasmid, you are having few 100 base pairs to a few 1000's of bases. So, the biggest insert you might clone is probably about 10 KB to 12 KB size. But there are vectors called bacterial artificial chromosomes, yeast artificial chromosomes, and then the cosmids. So these are large vectors, and large size insert can be cloned in these vectors. For example, you can have a 100 KB range in yeast artificial chromosomes or BAC's; in cosmids, usually about 40 KB to 50 KB in size can be cloned. So, the broken chromosome is cloned in these vectors, and transformed into the respective organism like; for example, if you take BAC and cosmid, you can use bacteria, and then you get the colonies. Now each colony will have the required vector. Let us say if you isolate a cosmid and since the cosmid sequence is known, you can design primers using the flanking sequence, and we can use those Primers as a template for sequencing. So, you do not need to know the insert sequence in the vector portion, using those vector-specific primers you get some sequence information about the insert. Then you could either go by sequencing or take smaller fragments from that insert, radiolabel it, and screen the library again to find newer cosmid clones that light up with that probe. Like for example, here in the slide, if you take the first orange one and below light blue one; now randomly I pick the orange piece, and in that, I took one of the ends to say the right end, and I made a radiolabeled probe and screened again that cosmid library. Then I find a new one that matches with the probe and then I go and take the other end of that new one and search for another piece and so on. Since they are overlapping I will be able to find all the DNA pieces and this is called chromosome walking. So, you walk along the chromosome in terms of identifying the cosmic clones that are sequentially arranged along the length of the chromosome. And then you make a map where cosmid 1 is followed by cosmid 2 followed by cosmid 3 and so on and that is a physical map. So, what is shown here in the slide is a physical map.

So, now in our mapping approach, we know that **cn** 9 map units right of black, so now you need to use the physical map to find in which cosmid that map unit 9 is present. So, for that again you need to know the known markers physical positions. And once you know that you jump from the genetic map to the physical map. So the first step, correlation of the genetic and physical maps near the gene of interest. For that essentially you need to know the physical map. If you know those two cosmids, so let us say black is in one cosmid and vestigial is in another cosmid. Then you know your gene is within these cosmids. Then you can do DNA mediated transformation where you are going to take each one of these cosmids and inject, introduce into the organism, let us say here in *C. elegans* you can inject those cosmids. And then see which one rescues the mutant phenotype because these are made from the wild type organism so wild type copy of that gene is present in those cosmids. Now you introduce the cosmids and then see which one rescues.

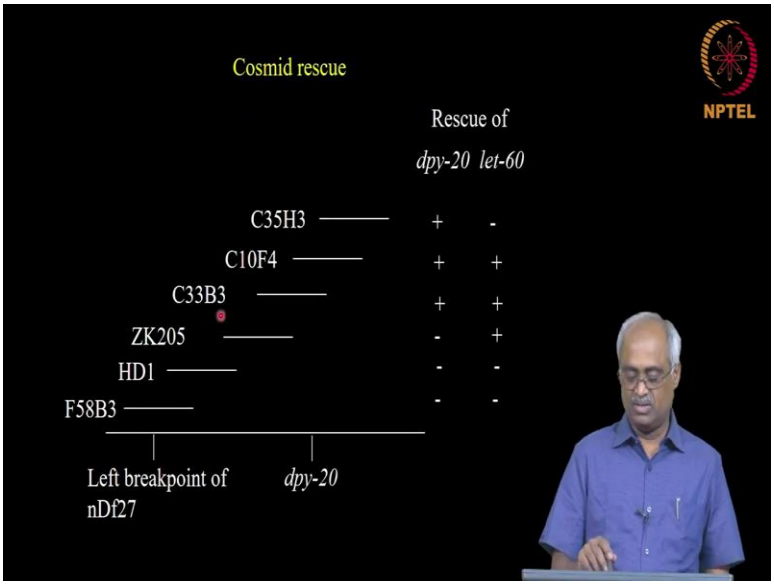
(Refer Slide Time: 17:26)



So now let us take an example that we saw earlier, so we know the *let-60*, the dominant allele that was found in the screen. So, the genetic mapping that is the two-factor, three-factor analysis put *let-60* between *dpy-20* (*dpy-20* is a known marker and its physical map is known, the cosmid number is C35H3) and the left breakpoint of nDF27. nDF27 is a deficiency; deficiency means larger deletions of chromosomes instead of point mutations or a few 100 bases of deletions. They are often used to determine whether an allele that you got is a null allele or not. So you can

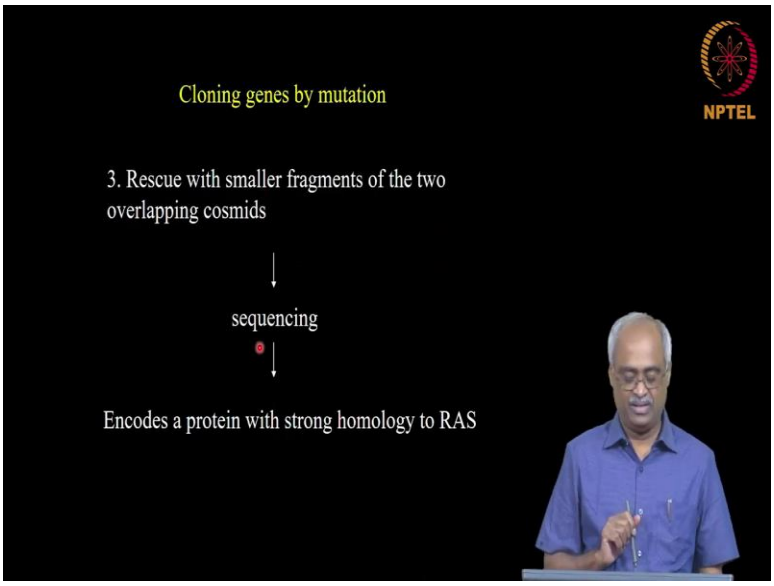
consider it as another known gene like *dpy-20* its genetic position is also known, this cosmid number is F58B3.

(Refer Slide Time: 18:33)



Now you take all the cosmid between *dpy-20* and nDf27. So in the slide it just the representative list of cosmids and all the cosmids numbers between the two are not mentioned. So you inject the cosmids and see which one of the injected one rescues *dpy-20* and *let-60*. So when you inject cosmid C35H3, you find it rescuing *dpy-20* but not *let-60*. So, somewhere in that cosmid, *dpy-20* is covered but not *let-60*. Then inject the next one that is C10F4 and then you find it rescues both. This tells that *let-60* is present left to *dpy-20*. And that gets confirmed when you go to the next one C33B3 which covers that full region and that also rescues. And then when you go further that is ZK205 you lost the *dpy-20* portion but then you have the portion where *let-60* is present and therefore *let-60* is rescued. So, then you get to know where the *let-60* gene sequence is present. So, now you take C10F4 cosmid and look at its left side and check what kind of open reading frames are there, how many are there, and then you inject each open reading frame and then you will find the actual *let-60* portion.

(Refer Slide Time: 19:45)



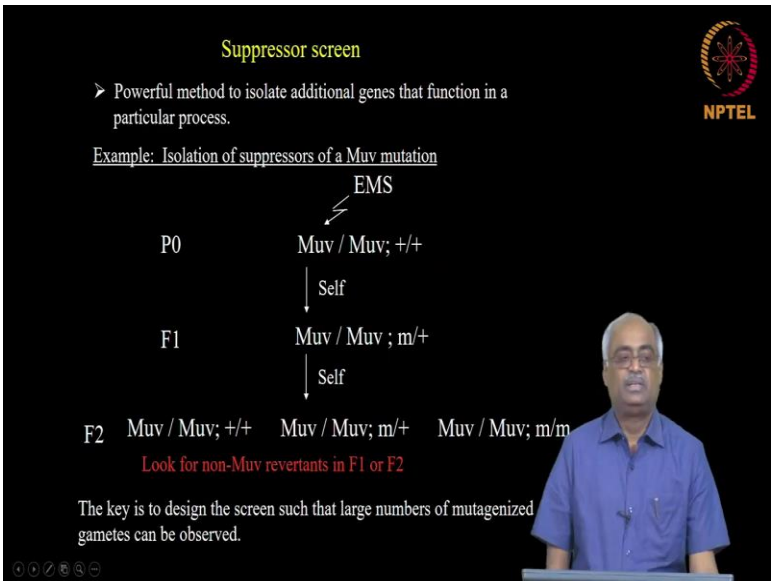
So in this particular case when they sequenced the open reading frame that rescued, they found it encoded a protein that was known already as RAS. Earlier RAS was known only as an oncogene where its expression caused cell proliferation but people did not know what it normally did. And this finding revealed that RAS is involved in normal growth and development. So the loss-of-function allele of *let-60* was lethal. So it is a recessive lethal and it is essential for embryonic and larval development. So therefore RAS is not going to cause cancer all the time, without it you are not even going to live; only when its expression is out of control you get cancer. So, that is how they discovered the growth and development role of RAS. So this gives you an idea of how you would map a gene. So, this is not the only way of identifying the function of a gene; there are multiple ways.

I will give you an example, what if during the cosmid rescue you find that the DNA injection for whatever reason is not rescuing and the fragment of the chromosome that you have taken as an open reading frame or as a cosmid insert need not always express when taken out of the chromosome context. So, how do you proceed further? So, one approach is you can use an antisense RNA. So, whatever is the copy of normal mRNA now you are suspecting as a particular open reading frame, for that particular open reading frame you make antisense RNA and inject it. So during the 90's period, this was a common practice. So, in plants even they made commercially grown crops where antisense RNA expressed against a particular polysaccharide hydrolyzing enzyme allows the tomato to mature but not to ripen because when the polysaccharide is degraded you get monosaccharides which make it sweet and then it becomes

softer and ripened tomato. So they used antisense RNA to block that particular enzyme's expression, the polygalacturonase and it worked, it is proven and it is commercial. So, in *C. elegans* if you remember the embryonic division, the first division is asymmetric; one cell is larger and another cell is small. So, there was a lab that focused on finding out genes involved in this asymmetric division. They mutagenized the worms and looked for mutants, like the way our mutagenesis screen was done, only difference is instead of looking for a bag of worms they were looking for dead embryos. Once they found dead embryos, they did a detailed observation, like what we did by looking at the Pn.p fate whether they become primary secondary, etc., they looked at when does the embryo die, at how many cells stage, then they were looking for whether that dead embryo allele was having problem in the asymmetric division. So they found one gene that when mutated led to symmetrical division, the first division. So, they did this standard experiment, using the cosmid to look for rescue and it did not work. So, they suspected an open reading frame and they decided to make antisense RNA and inject it and it worked. So, in the wild-type, blocking that one particular ORF by giving antisense led to symmetrical division. So they were convinced they mapped the actual gene. But then usually people do a control, control is sense. So, if you inject the sense RNA you should not get that effect. So, if you get that effect then the effect is nonspecific. Maybe just injuring the gonad caused that effect, it could be anything or any RNA sequence could do that. So, they did the sense injection and sense also had the same phenotype. So, they were worried and that is when Andy Fire, who worked in Phillip Sharp's lab(the lab that won Nobel Prize for discovering splicing), who knew very well about RNA chemistry reasoned out that when you are making single-strand RNA you are probably making trace amounts of double-stranded RNA and that is probably somehow interfering. That is how RNAi was discovered. So, while simply mapping a gene and while trying to fix an experimental artifact they discovered a new phenomenon. Then they decided to make double-strand RNA and inject it intentionally instead of having trace contamination and it worked. So this happened in 98.

So, this tells there are multiple ways of mapping a gene, you could either block it or you could rescue it.

(Refer Slide Time: 25:45)

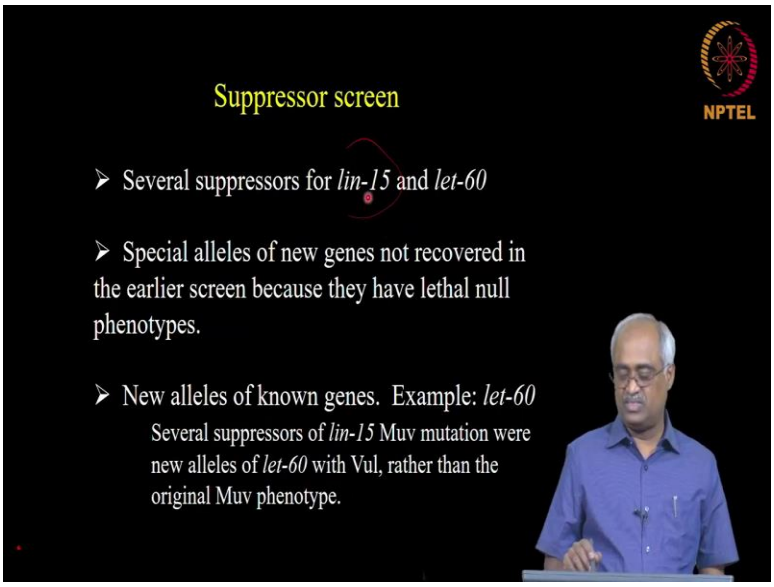


So, now we are moving away from mapping. So, we will revisit a little bit of mapping to the very end when we have a complex screen. So, the next is once you find the loss-of-function allele of a gene and its role in a particular function, like for example, in this example, we are considering multi-vulva(Muv), now you want to find out additional genes involved in it. So, you still do not have the confidence that by crushing the worm and making it into a lysate you will be able to purify the protein that is involved in this function. So, you do not still put a lot of weightage on that technology and you want to still rely on genetics. So now what you do is, you take the mutant and mutagenize and screen for new mutants in which the old mutant phenotype is lost. Meaning can I make a new mutation and rescue the old mutant phenotype back to wild-type and they are called suppressor screens. So, here you do not do just 10000 gametes like you do not simply screen 5000 mutagenize worms, here you go large number because you are looking for a specific genetic interaction, a new mutation that rescues or suppresses the original phenotype. This will show that these two genes must have an interaction. The mechanism is not known at this point but they somehow interact. So that interaction in terms of the phenotypic outcome is called genetic interaction. So, you want to find such genetic interactions. So, this method helps you to find additional genes and that is why this is a powerful method.

So, here you take Muv/Muv, it is homozygous for that loss of function. The new mutation, the new gene that you are looking for let us say it is wild-type here +/+. Now you mutagenize and allow it to self by cloning single worms and repeat the same thing and look for in F2 where some of them have the phenotype being suppressed. Since you are starting with worms that are

homozygous for Muv, and not doing any cross with wild-type, this Muv/Muv must remain always, which means you should always get multi-vulva and the moment you do not get multi-vulva then you found the new mutation that is suppressing multi-vulva. So it should be made sure that the new mutation for this phenotype follows the Mendelian inheritance pattern like 3:1 for the single phenotype segregation and so on to ensure that you have hit a single gene that is suppressing it. And here the key is to screen a large number. And if you are going to screen a large number, the mutant phenotype should be readily visible. The difference between the rest and the mutant should be obvious then only you can identify it. It cannot be like you had mount of worms on a slide and go and sit in a microscope and look at the staining pattern of something, then you will not be able to screen millions of them, in a plate if I have 10000 worms and if I look at it, the mutant should distinct. So that is the key.

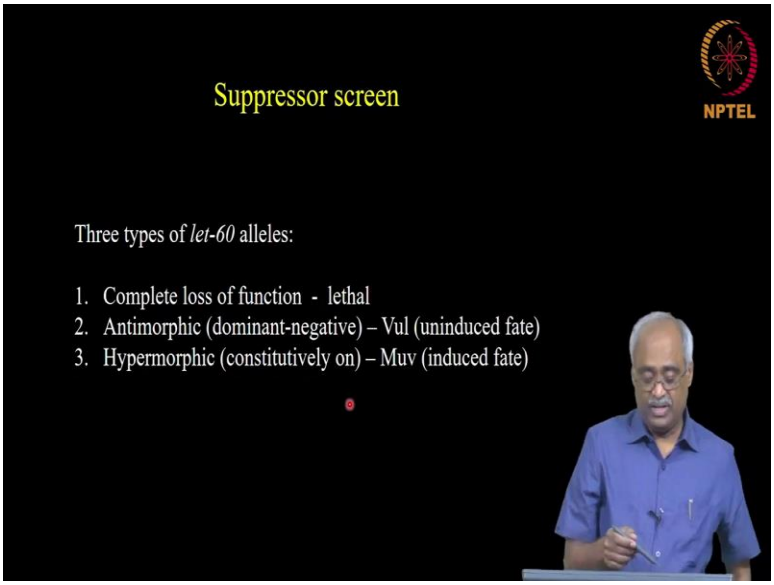
(Refer Slide Time: 29:43)



So, several suppressors were found from the screen among them are the two multi-vulva genes, *lin-15* and *let-60*. So, special alleles of new genes not recovered in the earlier screen because they were lethal or null phenotype, have been recovered now. So, like people in our lab know this because this is the only kind of screen we have done and we found a lot of alleles where if you cause a total loss of function they would be lethal and a specific partial loss-of-function would have allowed its function during the early development. Then in a special specific situation, you find a defect particularly in the background of another mutation. So, you find those kinds of different alleles having different effects. So you can imagine that an amino acid

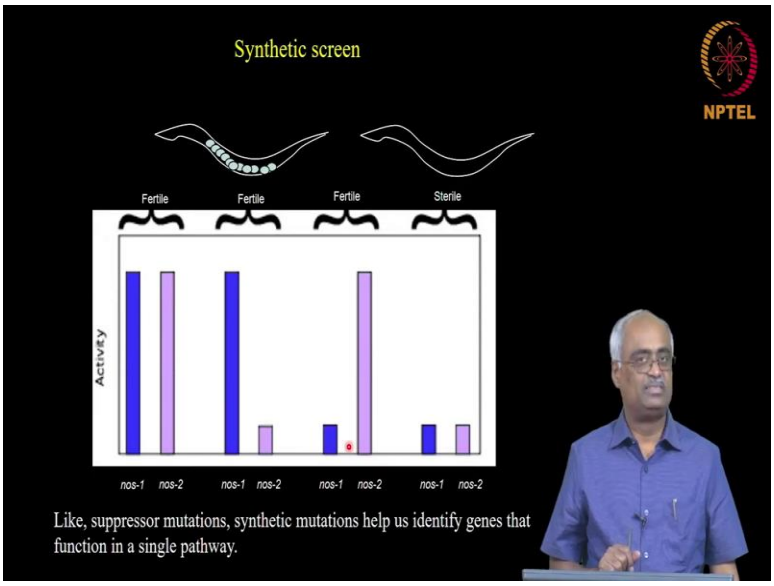
substitution in one region versus another region might have different consequences on the protein function. So that is how you get partial loss of functions. And sometimes you get new alleles of known genes for example several suppressors of *lin-15* multi-vulva mutations were new alleles of *let-60* with the vulva less phenotype. So, worms do not make vulva only when they lose *let-60* function during vulval development. If they lose it in embryonic development then it results in larval lethal and that is why the term *let*, larval lethal is used. So they ended up finding that some special alleles of *let-60* made it not doing its job during vulval development so therefore whether *lin-15* is there or not it did not matter, *let-60* stops the vulva from developing and that is how it ended up suppressing *lin-15*. So, these are the kind of alleles that you get when you do the suppressor screen. So, these kinds of variant screens there are many, so we are not going to go through all of them because we just want to get some good idea of genetics to continue on our development. So, this is just one example of a complex screen.

(Refer Slide Time: 32:03)



So after this sort of a screen they ended up discovering this for *let-60*, a complete loss means it is lethal and antimorphic means dominant-negative like vulva less it does not allow other cells to be in the uninduced fate and hypermorphic, if it is continuously on so it makes multi-vulva.

(Refer Slide Time: 32:32)



So that is a suppressor screen. So, here the original mutant gave a weak phenotype and not a strong phenotype so we want to mutagenize and look for new mutants that gives a stronger phenotype; or sometimes the original mutant did not have a particular phenotype, so we discover because of some other phenotype defect. But when you combine with another mutant mutation in another gene the double mutant gives a new phenotype. So those are called synthetic phenotypes. And a genetic screen looking for that kind of phenotypes is called a synthetic screen. So, here I am using an example that did not come from a synthetic screen. This is something that I did and therefore it fits as a good example, so I am telling this. So here *nos-1* and *nos-2* are closely related genes. So, in the wild-type when you have both of them it is fertile and it is totally fine. And when you have a mutation in *nos-2*, let say the *nos-2* expression is extremely low, the activity in y-axis and *nos-1* is fine it is wild-type and the worm is again fertile. Now you do the reverse *nos-1*'s activity is reduced and *nos-2*'s activity is high and it is still fertile. Now you bring both of them down, it is sterile. So this is an example of a synthetic phenotype. So in this situation what you are going to say is, these two seem to function redundantly. So, like suppressor mutations, synthetic mutation helps us identify genes that function in a single pathway. So, redundancy is what you are discovering in this kind of an example. So, we stop here, and in the next class we will go on to find the hierarchy or the sequence in which a given set of genes function.

Like for example, we found 20 genes involved in vulval development. Two of them were multi-vulva and out of which one was the dominant allele. So let us say we ignore those two. Then we

have 18 of them giving vulva less, now we do not know which one of them acts first and which one of them acts as the 18th and which one of them is in a parallel pathway and so on. So we want to find an order, that is doable using genetics and it is not always done using biochemical methods.

In biochemistry, the order of genes, for example, we know A gets converted to B gets converted C, and so on. That order was determined using epistasis analysis. So that is what we are going to discuss in the next class, that is one and second to be fair there are non-genetic methods like radioactive labeling of a specific atom. Like you label the particular carbon and then you see where that carbon goes. Like for example in glucose if we label the 6th carbon, now in pyruvic acid we see what is there at the end of glycolysis. So, this radioactive tracer as well as this epistasis analysis is how all the biochemical pathways worked out but of course, biochemical pathways were not worked out by doing genetics in *C. elegans* or Drosophila. They were done mostly using E. coli bacterial genetics. But the basic rules apply in all organisms the same way. So, we will stop here and continue on epistasis analysis in the next class.