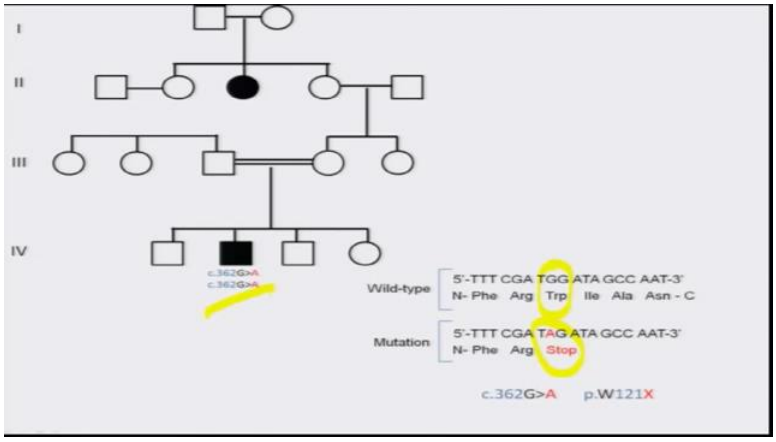
Human Molecular Genetics
Prof. S. Ganesh
Department of Biological Sciences and Bioengineering
Indian Institute of Technology, Kanpur

Module – 02
Lecture - 08
DNA Cloning and Hybridization Techniques - II

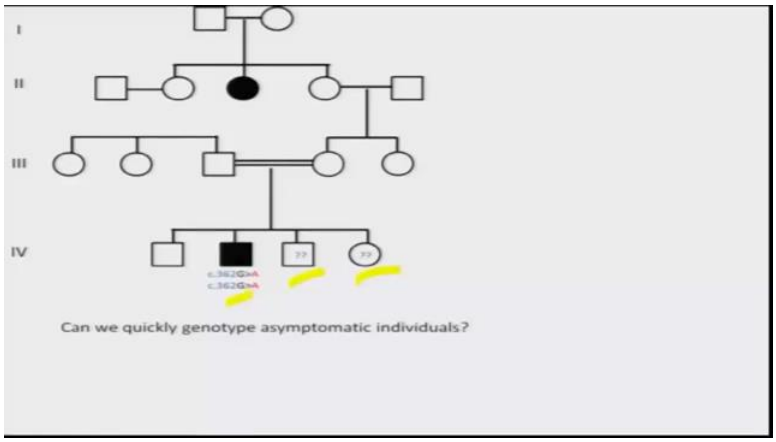
We will be discussing in this particular lecture which happens to be the fourth lecture for week II, how the basic molecular biology techniques that we discussed in the previous lecture can be used for diagnosis. Say, suppose you have a family, which is affected with a disorder; I want to diagnose, do a DNA diagnosis to understand whether an individual is having any mutation in the gene or not and whether that information can be used to predict whether his brother or sister who has not shown any symptom as of now will develop the disease later. So, that's, the focus of this particular lecture.

(Refer Slide Time: 00:58)



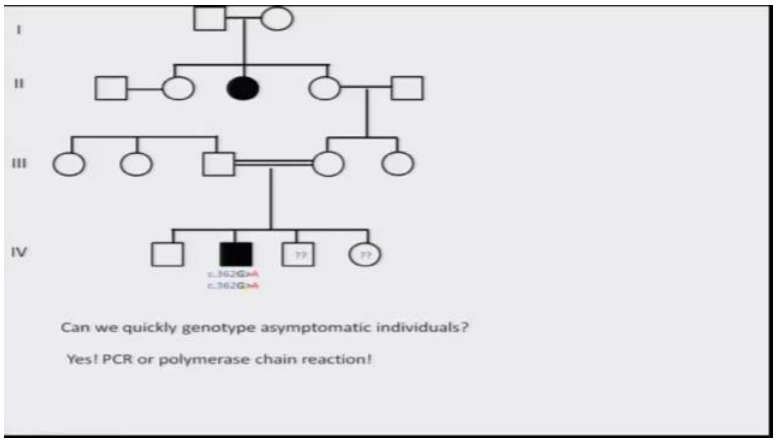
So, what you can see in this particular pedigree in the screen is that it is a typical autosomal recessive disorder in the family and you can see that affected parents are unaffected and there is a consanguinity and you have done a DNA cloning for a gene and then you have done a sequencing and you identified a disease mutation, something that we discussed in the previous class and the mutation is that you have a point mutation, resulting in the change in the codon. Now in the place of tryptophan amino acid, you have a stop codon. So, this individual is obviously homozygous for this particular mutation.

(Refer Slide Time: 01:41)



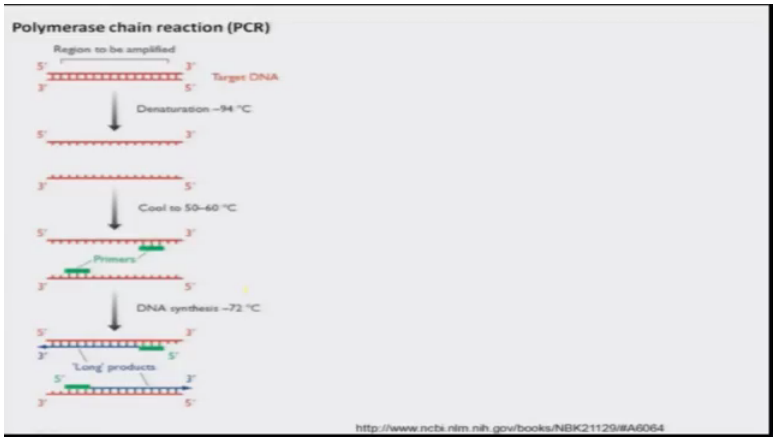
The question that you would like to ask is if that is the case what is the genotype of these two individuals? Assuming this disorder is little late on set, let's assume that the symptoms sets in around, when the individual is around 40 years or so, so, these individuals may be still they are young. They are 12, 13, 15, 20 years or whatever. What is the genotype, whether they will have the symptoms when they are reaching around 35, 40 years? So, you want to look at the gene sequence. Now if whatever we have discussed in the previous class that is going for making a library and pulling out the clone, sequencing is a laborious task, it takes years to make these advancement. So, what you are interested is to quickly understand within a day or two, whether these two individuals have this particular variant or not. That's your task. So, how will you really go about doing it? Can you quickly genotype these asymptomatic individuals?

(Refer Slide Time: 02:47)



The answer is ‘yes’. There's one powerful technique that we have which is called as PCR or polymerase chain reaction, many of you would have heard about it and, would have read, about this particular approach. So, what we are going to discuss is how this technique can be used to screen for mutations in the individuals.

(Refer Slide Time: 03:14)



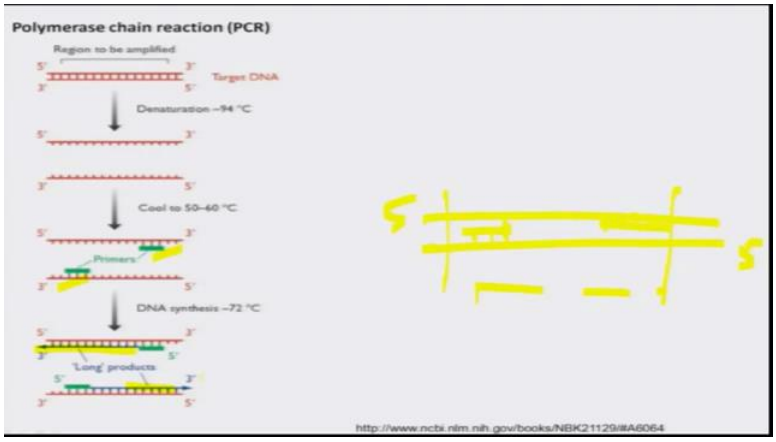
So, this is a diagram which pretty much tells you the principle behind polymerase chain reaction. Again,, the concept is very simple and but, it is so powerful a technique that the discoverer received Nobel Prize for this discovery. So, it uses again, the concept from our DNA replication machinery. So, how our cells divide? So, in our cell, whenever the cell divides, the DNA gets copied and you have a protein that mainly does the function of copying the DNA, which you call as DNA polymerase. Likewise, also in this reaction, we had a DNA polymerase, but the difference is this polymerase can withstand very high temperature, because these proteins are isolated from microbes that are living in hot springs. Therefore, the ideal temperature for them to survive is around 60 - 65 degrees. They are very happy in that temperature.

Therefore the protein also functions very efficiently at a very high temperature and even if you boil the protein to 95 degrees or 100 degrees, the protein, once it gets back to 60 - 65 degrees, it is able to regain its function and does whatever job it is supposed to do. So, that’s the advantage of such kind of proteins and people have used that protein to do this reaction.

So, what is this polymerase chain reaction? Polymerase meaning, you are making a polymer,

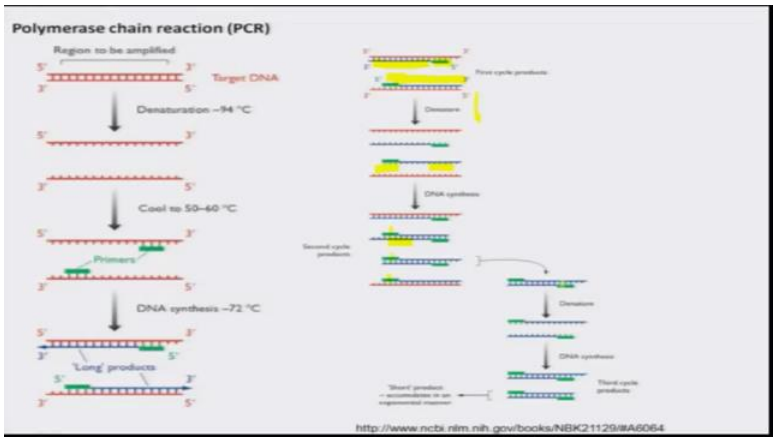
simply copying the DNA, a new strand is being made, bases are added based on the sequence that is available in the template strand. Why do you call as chain reaction, because you are repeating this similar process again and again. How does it really help you, something that you can see. So, what you need for a polymerase chain reaction? Of course, you need the template. So, if I want to see a given individual whether he or she carries a defective allele, what I need is the DNA from that individual.

(Refer Slide Time: 05:31)



So, extract the DNA, very little amount of DNA is good enough and then I need what is called as primers, the one that are shown here, these are called as primers. What they are? Say, suppose this is my DNA region that I want to be sequenced. So, my gene falls from here to here, somewhere there could be a mutation. So, what I need to do? I design a primer, for example, this is 5', and this is 5'. I design a primer that goes and binds to this sequence. Likewise, I design a primer that goes and binds here and then that's what is shown know, short oligonucleotide sequence that are synthetically made, they are DNA sequence, they go and bind, because they have the complimentarity. So, it will not go and bind anywhere else, because the sequences are so unique, because they design that way. Once they bind your DNA polymerase, extend these regions. So, that's what DNA polymerase does. It goes and sits and keeps on cropping all the strands that are present next to it. So, it extends. You allow that to happen for some time and then, what I do is, then again I denature the DNA. So, I heat the, you know, DNA tube, tube containing DNA.

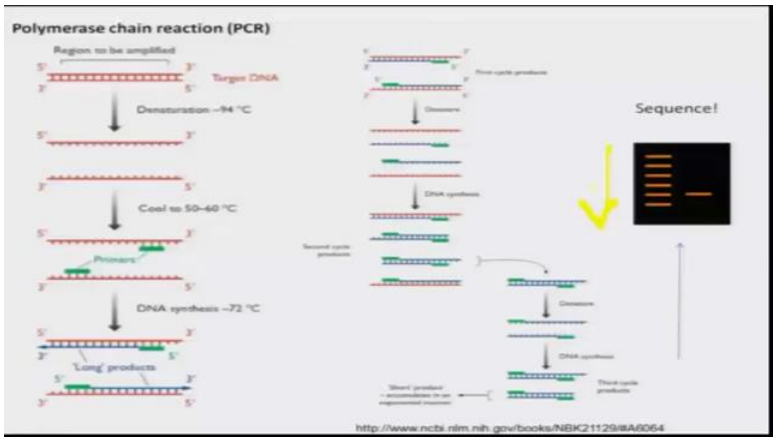
(Refer Slide Time: 05:31)



Therefore, these two strands separated and again I do the same reaction. So, what happens? So, I have completed, for example primers were added and then the synthesis took place, as I discussed just a while ago. Then, I increase the temperature. That is what is called as denaturation. You heat the reaction to say, 95 degrees. Then, all the hydrogen bonds are broken, so, again the DNA becomes single stranded. But you have excess amount of primers sitting in the tube. So, they are going to go and bind once again to the complementary sequence and the DNA polymerase is active again, when you bring it back to a normal temperature for the DNA polymerase, that is around 65 or 70 degrees or so on and then, it will extend again. So, it will copy. With this process, when again the primer gets extended you have made new copies of DNA, using the template that were already made.

Now, what do you do? Again you go and melt it. So, all the DNA becomes single stranded, again the primers are available; they will go and bind to your newly synthesised DNA, which has complementary sequence. Again the DNA synthesis happens. So, what you do is if you repeat this process for 30 or 35 times, you are going to make millions of copies of a small segment of the DNA, where the primer is able to bind and can make new sequences. In this process, you can make a small segment of your DNA multiple copies; millions of copies can be made in a matter of 4 hours, 3 hours.

(Refer Slide Time: 08:34)



So, at the end of that so called reaction, what you call PCR reaction, you can test whether you are able to make millions of copies of a small segment of the DNA, by running it in a gel. So, this is one way you can check the amplification, what you call as multiple copies of the DNA process. So, when you run it in a gel, so you can separate that DNA fragment according to the size and then you can add the dye and then you will be able to see them. That would tell you whether you are able to make these copies. So, that's one process by which you are able to make millions of copies of small segments. So, depending on which gene you are looking at, you can choose different primers, such that you can make copies of that.

(Refer Slide Time: 09:18)

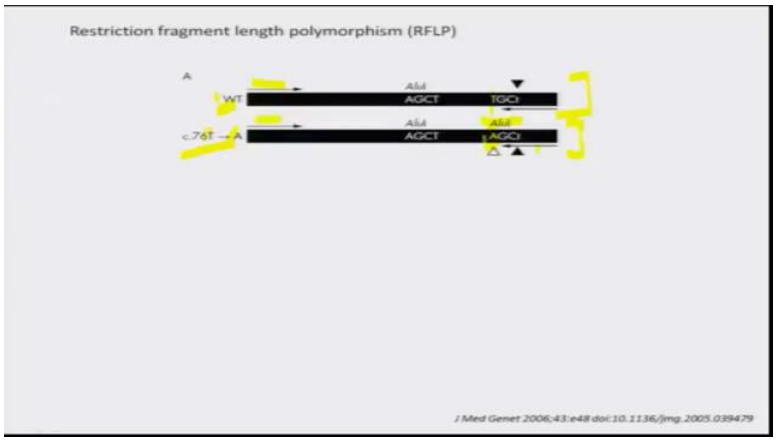
So, this is something, picture that I have shown you to explain how the electrophoresis work, rather how you separate the DNA. So, what you have is, here is, you have a matrix, which is made up of a gel, agar gel and this gel has got pores, small pores through which the DNA can

pass through, if you apply force. Here for example, you have applied the DNA solution here and then what you did is that you have applied electric field, here is negative, here is positive. So, since the DNA is negatively charged, it has to migrate from this direction to this direction and when it does, it has to go through the small pores that are present in the agar, agar gel. So, longer the DNA, it is going to take more time to pass through, navigate the pore; shorter the DNA, it will do relatively faster.

So, as you can see here, this particular segment, these are called as size markers. These are DNA fragments of known size. You can see that this is a smaller fragment and this is a larger fragment, we are able to separate them. You know the size that helps us to identify the size of other fragments. For example, these are the fragments that are either extracted from a library or you have a done a PCR to amplify and I can compare this band, this white band with this and I can say, I know this is the size of this particular band, therefore this should be the size. So, this is the way I am able to one, quantify the DNA, for example I have more DNA here, lesser DNA here or I am able to tell this is larger fragment as compared to this fragment and this particular sample has got two different fragments.

So, I can interpret this in various ways, depending on what is the condition I have used and what kind of information I am looking at. So, this is how you separate the DNA in agar- agar gel and then, you are able to come up with certain conclusions. If you have done a PCR, you will get a single band and single such kind of amplification, which you can go for sequencing to identify what is the variation that is present. So, that's one.

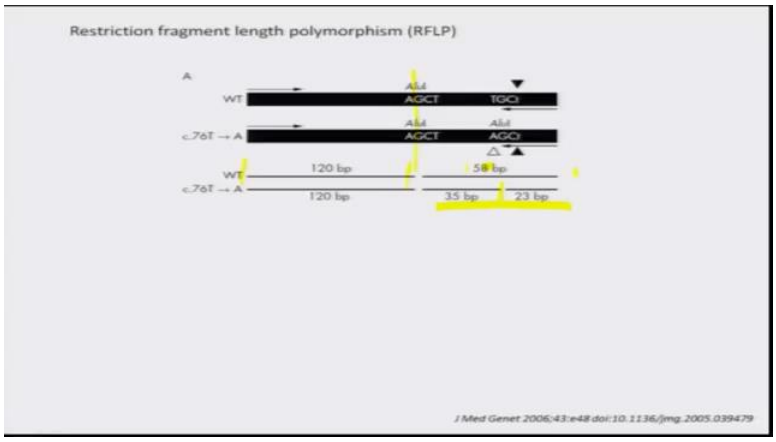
(Refer Slide Time: 11:35)



But, there are other various ways as well, to detect whether you have a base change. So, what I am going to show here is a method that more commonly used in diagnosis and this is called as restriction fragment length polymorphism. What does it mean? So, we have discussed about an enzyme called restriction enzyme. These are in the previous class that refers to enzymes that identify a particular sequence and cut, and you call as restriction enzyme, because they are unique to, they are so specific to their sites. They cut only when the site is present and we can use this property to identify whether there is a change in the sequence.

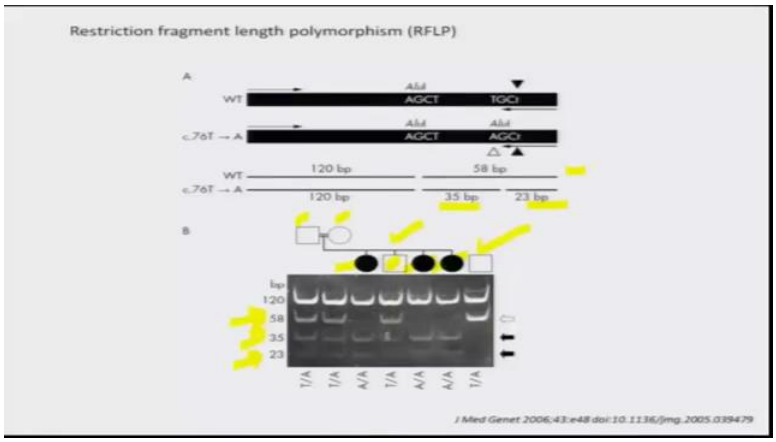
What is shown here is a schematic of some sequence, which can be digested or not digested by an enzyme. So, you can here, this is allele representing wild type DNA and this is an allele representing a mutant version of DNA. You can see here it is coding and then 76, that is the base from the start codon ATG and this T is converted to A, whereas the wild type will be T, in this allele it is A. So, what it is shown? So, you have used a primer as represented by an arrow, that are used for amplifying this segment of the DNA and then, you are looking at this particular site. This is T, because the T is converted to A. So, what is, we are trying to show here is, is the wild type has got T and the mutant has got A and because of the A coming in here, the mutation changes the base T to A, now this particular sequence gains a new site for an enzyme denoted as *AluI*. In other words, this enzyme would cut this mutant allele, but not the wild type allele, that particular region.

(Refer Slide Time: 13:49)



So, we can use this information to do a digestion. You do a PCR, amplify that segment, add this enzyme and then check whether it is able to cut there or will not be able to cut there. So, what is the sequence that you can see? Fragment size, see if it is a wild type, you will get, from here to here which is 120, and there is a natural site present here, both in wild type and the mutant allele. But this particular site is unique to this mutant allele; therefore this 58 base pair fragment will be made into two, in case of mutant. In other words if you have mutant allele, then the 58 base pair fragment will be missing.

(Refer Slide Time: 14:26)



So, let's see. This is in fact, a family suffering from Lafora disease. This work was done in our own lab. You can see, this is the father, mother and they have got 4 children, of which 3 are affected and we have done this PCR to detect and show the genotype. So, you can see here, this 58 base pair fragment which is present in the wild type, was seen father, mother unaffected and a

control, normal individual, but absent in the three affected. On the other hand, all the individuals of this family were having these two fragments, that is 35 and 23, these two fragments. They are present that means every individual of this family carries the mutant allele. So, father and mother obviously are heterozygous, even the unaffected brother, who is not showing symptoms, will not develop, because he is having a wild type allele or he is heterozygous for the mutant allele as well, whereas you can see here, this particular individual is a wild type, not having the mutant allele at all, do not show any such fragments.

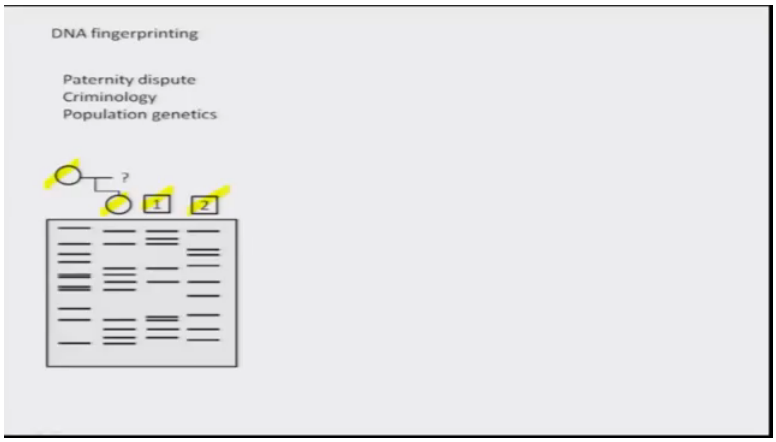
So, this is how really you can use the sequence information to come up with various ways to quicklytest whether a particular mutant allele is present in the other member of the family that we discussed, right, or not. So, you can either do a, after a PCR you can go and do a sequencing or if you know such kind of alleles present in a, in a given family, you can simply do a restriction, digestion and look at the banding pattern. So, you run them on a gel, look at the different fragments that they make and put the pedigree and you will be able to clearly show the segregation and tell, yes, this individual will not develop Lafora disease, because he is heterozygous for mutant allele. He has got a wild type allele, therefore he will not develop. So, this is some of the ways, by which you will be able to tell. That is about diagnosis. So, you know the gene, you know that that is the gene when it is mutated results in the disease and you can sequence it, you can look at the sequence variance there and then tell whether an individual, heterozygous, homozygous, whether he will develop disease and so on.

Let’s go beyond the diagnosis, we will look into some other aspects of DNA technologies or DNA molecular biology approach, which help in human molecular genetics. One of the powerful techniques that people use for many issues is DNA finger printing or DNA profiling, what they call. You must have heard about it in the newspaper, news channels and so on. Even, the suicide bomber of former Prime Minister Rajiv Gandhi and she was identified or even her, her mutilated body was identified by DNA fingerprinting. Even the Prime Minister's body was, people were unable to recognize and they have done a DNA fingerprinting to identify the body or DNA finger printing is being used in many different cases. For example, somebody is murdered and their body has decomposed. How will you know whose body it is? So, can that be, know, confirmed whether this is the body of this particular person? We will use DNA finger printing or there are

disputes for example, paternity disputes, two different males claiming that a particular child is their own, he is the father. So, how will you prove that?

Or it could be crime scene; somebody is murdered and you don't know who is the culprit, who did that and you can look into the DNA, the blood spots that are there and if this was a fight between two individuals, at the end one was killed, the other one would also have had some injury and so on. His blood also would be there in the spot. So, can we identify that individual using this, the DNA that is present in the blood and, and many such, you know, issues that one could really look at using DNA finger printing. Let's see what it is.

(Refer Slide Time: 18:52)



What is shown here is a banding pattern of DNA, shown here is schematic of a paternity dispute. So, you have a mother and her daughter. What we don't know is who the father, for this particular child is. There are two individuals who are claiming, individual male 1 and male 2. So, it went to the court and, the court ordered that let us do a DNA finger printing. So, they took the DNA from the mother, the daughter, the two suspects, suspect meaning who are claiming that they are the father for this particular daughter. So, this is the pattern that they got. Let's assume it is something like agarose gel and you have each band representing the DNA band.

So, how would you use this information to tell, who is father and mother? So, what you need to understand is that meiosis, so something that we discussed in the, in the previous week. Meiosis is that, all of us are resulting from the DNA that we receive from father and mother and there is a recombination as well. So, what happens is that for every DNA segment that we have in our

body, you can match it either with your father or with mother. The same applies to anything that you are part of it, which you amplify using a PCR or you do a RFLP, anything that you do, all these changes that you see in your genome is derived from either your father or your mother. So, if you look into this variation, you will be able to tell whether an individual could be a father or not. Certainly you will be able to tell this person is not likely to be father in this case. Let's look into that.

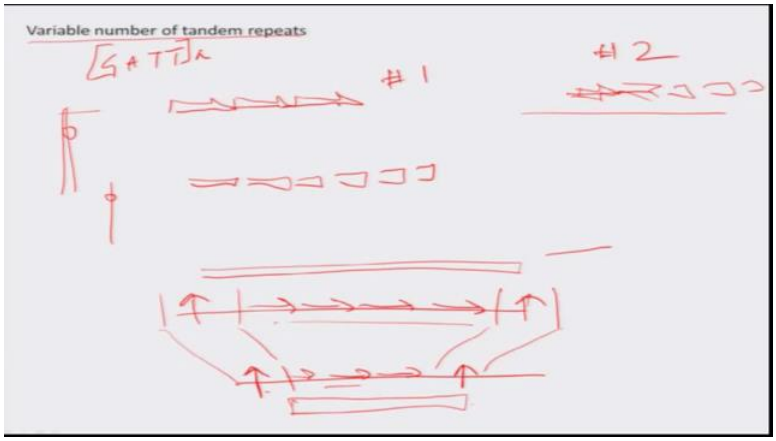
(Refer Slide Time: 18:52)

This is what it is. So, we know who is mother. So, what we can do is, we can compare the DNA profile of the daughter with the mother and as I told you, 50% of the DNA comes from mother; the other 50% comes from the father. So, at least let us match all those fragments that have come from mother. So, I have used the pink colour here to identify the fragments that are likely to be derived from the mother. So, there are unique fragments, like what is shown here, here, here, here; these are not present in the mother. So, definitely that must have come from father. Now, what we need to look at, whether each one of the fragments in the daughter that is not present in the mother, therefore likely to have come from father, is present in which one of the two individual.

(Refer Slide Time: 21:34)

So, this is what it is. So, you can see here all the bands that were missing in the mother, were present in one of the two males, male number 1. Some of them are absent. For example, if you look into, there were some fragments which are altogether missing. For example this fragment, this individual is not having. So, it could not have come from this individual. Likewise, this fragment missing in this individual 2, so could not have come from him. With this approach, we will be able to tell the biological father of this particular girl child could be male number 1 or at least you can say, this individual is not likely to be her father. Therefore, number 1, the male number 1 is likely to be. So, this is one simple example to show; the actual DNA finger printing will be more complex than this. But, what we are going to see is how really you get this kind of pattern. What is the genetic basis? What happens at the DNA, which help in seeing this kind of variation?

(Refer Slide Time: 22:43)



So, it all happens because of certain unique sequences that you have in your genome, which is

called as variable number of tandem repeats. What is that? What do you mean by repeat? For example, this is a repeat. It is a unit, which can be repeated many a times. At certain regions, you could have 20 of them; at certain regions you could have 25 of them and so on. So, what you see is that in a chromosome for example, if this is the chromosome, in this particular segment of individual 1, you may have, this is one repeating unit that is repeated for example, 3 times. But the same repeat in another chromosome, it could be repeated may be 6 times. So, likewise we are going to have different lengths of the repeating unit. That's why it is called as tandem repeat; one after the other, head to tail you have them.

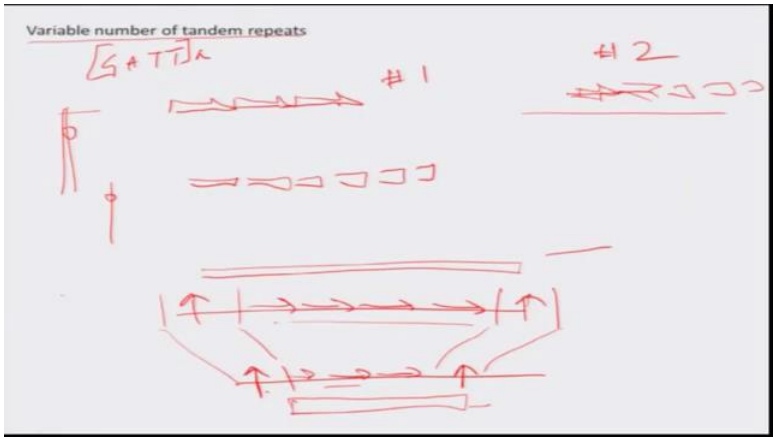
So, in different individuals, you are going to have different lengths of these repeats. The same region, for example in individual 1, you have 4 repeats, but in individual 2, it could be, for example it could be 5 repeats. So, how does it really, this kind of repeating pattern help? So, if you are looking at repeats for example, you have repeat 1, repeat 2, repeat 3 and repeat 4. You have another allele, which has got only, for example 3 repeats. So, the flanking regions that are seen here, the sequence that present on either side of the repeats are going to be common.

This is exactly the same like what you see, because, this is the same chromosomal region, but in different individual. So, if the flanking regions are common, you may have a restriction site here and here. Here also you are going to have the same restriction site. But, if you digest with an enzyme which cuts here and here, the resulting fragment that is generated between these two sites are going to be different, depending on the number of repeats that you have. More number of repeats here, larger this fragment is going to be. Fewer the number of repeats, the fragment is going to be smaller. So, that is exactly what makes the difference.

(Refer Slide Time: 25:03)

Perhaps this fragment that you are seeing here and the one above that represent the same region of the maternal and paternal chromosome. So, for example the maternal chromosome may have had fewer more repeats in that region, therefore it is little longer, therefore migrates slowly, whereas this one migrates, know, faster, because the size is smaller.

(Refer Slide Time: 25:34)



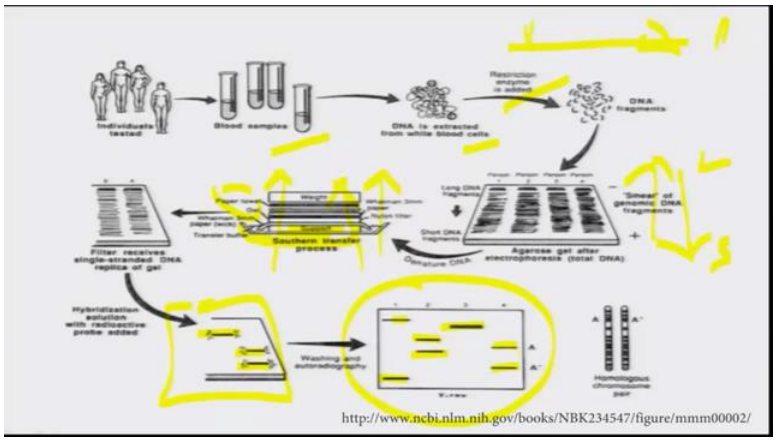
So, this is, this kind of variation that you see in the repeats, really brings in the changes. We are not just looking at one particular region of chromosome, you are looking at hundreds of different sites in the chromosome and when you do that, such kind of calculations tell the probability that an individual will have identical combination of various alleles of various regions of the chromosome is extremely low, unless you are identical twin, where the genetic makeup is identical. Otherwise you are not going to match. That's why it's called as fingerprinting, because your finger prints, your thumb prints again are unique to individuals. Very rarely, you will find

the thumb print patterns are identical in more than one individual. That's why it is called as DNA finger printing.

So, it is the sequence variation that you have in your genome, because of various alleles of the same region of the DNA that brings about difference in the migration pattern for the repeats and if you use a repeat as the region to detect the fragment, you are going to find bands like this and the bands are, patterns are unique to each individual. By mapping it to either father or mother, you will be able to identify which one had come from father, which one had come from mother. But, if you are unable to map it to one of the two parents, then that means that that individual may not be the biological parent.

So, the same principle can be used for forensic evidences. For example, you if you have a crime scene, where you have spots and each, blood spots you have collected separately, extracted the DNA and you are doing the DNA typing for the person who was killed in that encounter and then you map whether every blood spot that is present in that crime scene matches with the DNA profiling. If it doesn't match that means that is the blood of another individual that is present. If you find more than one such pattern, you probably will be able to predict how many individuals were there, when this encounter took place, and then you go with other, your investigations, you find a suspect and then you will be able to match his DNA with the DNA pattern that you were able to get from the blood sample. So, that would pretty much nail that this person was present in that site, when this death took place. So, this is how you are able to map it. So, that's about the DNA finger printing. How do you really do this? We will go to little more details.

(Refer Slide Time: 28:08)



So, basically you extract the DNA. This is something that is shown here in this schematic. We extract the DNA. For that, either you go for the blood samples or if the crime scene has got blood spots, you can go with that and extract DNA, because the blood has got white blood cells, which are having the nucleus, therefore you can extract DNA and then you digest with the enzymes. So, it cuts on either side of, for example, the repeat sequence. As I said, you have a repeat and you have a site on either side; it is going to cut and then separate them , you separate the DNA in a gel. They are separated according to the molecular weight. This is longer DNA, it is a shorter DNA, migrates faster and then what you do is, from the gel, the agar gel, transfer the DNA to a membrane.

This membrane is a membrane, that is positively charged, so the DNA goes and, sort of irreversibly bound to the positively charged membrane, because the DNA is negative and therefore the membrane would represent the agarose gel, because it is the same position where the DNA was present in the agarose, the same position it will be transferred to the membrane and we us the membrane to what you do as, what you call as hybridization, meaning you have the DNA strands separated and then transferred to the membrane in a single stranded form. So, once you have separated in agarose, you treat with certain chemicals, for example sodium hydroxide, which breaks all the hydrogen bonds between the two strands of the DNA and now you transfer the DNA to the membrane, that is kept over the gel and what you do in this case is that, you have a support, wherein you have the gel and then on the top of that you put a membrane and you allow the salt solution to pass through the agar to over the membrane and to the paper towel that are kept on the top.

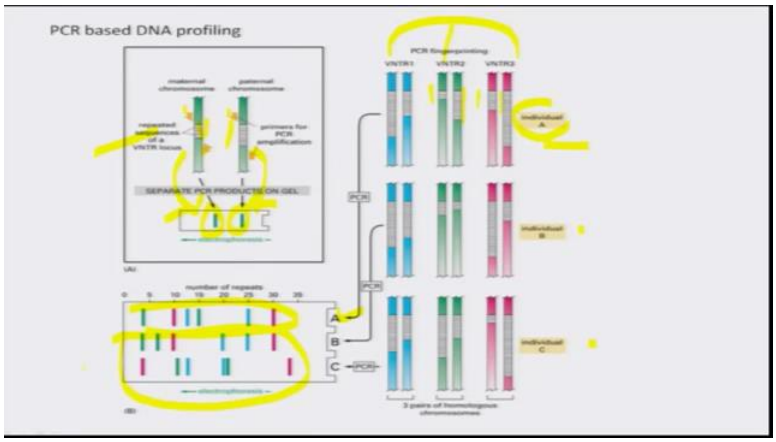
So, in this process, the salt solution would also carry the DNA and, but the DNA cannot pass through the membrane, therefore they are stuck there and they are single stranded, they are cross linked, meaning they are irreversibly bound to the membrane. So, once you have it on the membrane, so now what I said was that, in DNA finger printing, basically you look at the repeat sequence. Because the repeats are tandemly repeated, the units, the number of times they are repeated in a given region varies from individual to individual. As a result, if you digest that DNA with an enzyme, the fragments that would be produced because of, know, detecting the repeat would vary because, more number of repeats you are going to have, larger the DNA.

Now, you use the repeat as a probe. What is a probe? It is, you take the complimentary sequence of the repeat and then you label it with an isotope, radioactive isotopes you label and then you allow the labelled nucleotide to go and bind to the target. So, where are the targets? These are already size separated. So, it will go and bind, because these are single stranded, it will go and bind to the complimentary sequence and then since they are radioactively labelled, so you can put a film like what is shown here, extra film and then wherever the probe is bound, it would give you this kind of bands, something like, know, what is shown here. So, you can see, so in this lane you only found two bands, here it is two, here it is one, it is two and so on. So, it is simplistic; so you would basically get something like this. So, you will find multiple bands and then you will be able to match the banding pattern of every individual with the other and infer whatever you would like to infer.

So, this approach is called as Southern hybridization, because, the approach was developed by a person named Southern, therefore it is called as Southern hybridization, where basically you hybridize a DNA with a DNA. The difference is one of the DNA, the target is separated, size separated in agarose gel, transferred to a membrane and they are immobilised on a membrane. The probe is exact sequence that you are looking at. Where are such, for example GATTC? Where are such sequences present in your genome? So, use that sequence as a probe. So, wherever that homologous, the corresponding complimentary sequence is present, it will go and bind to. So, you can use various different types of sequences to profile. So, that is, is one such example of DNA profiling.

But, Southern blot is not the only way by which you will be able to get individual specific profiling DNA pattern.

(Refer Slide Time: 33:16)



People also use the robust quicker PCR approach. So, what is shown here is, so you have exactly the same thing. You have sequences, which are repeats. You could for example, in the maternal chromosome, the chromosome you derived from mother would have a shorter repeat and the paternal would have a longer repeat. So, if you amplify using a pair of primers that are binding to the regions that are present on either side of the repeat, it would amplify from both, but the resulting PCR product would be different. If it is longer, for example, it will migrate slower; if it is shorter it will migrate faster. But, if you do this similar kind of amplification from 40 different such regions of the chromosome or 40, or all the, each one of the chromosome, 22 pairs, then you are going to get a pattern which is unique to a given individual like, something like shown here.

So, you are going to, for example here you are looking at three different such repeat locus or repeat loci rather and then, you are looking at individual 1, 2 and 3 and depending on what is the repeat size, the alleles that are, each one individual is having and the combinations, you will be getting a pattern which is unique to each individual. For example, an individual has got a pattern, which is very, very different from the other two individuals. So, this you can do it in in about 4 hours, 5 hours, you will be able to type and tell whether these two individuals or the blood spots

that you found in a crime scene, whether they all come from one individual or different individuals. If it is different how many individuals were likely to be there, who, who had some injury and their blood spots are there or you can go and solve a paternity dispute and so on.

So, this is such a powerful technique which, which people use to identify the, match the DNA profile of different individuals or samples and give information, whether it represent the same person, different person, whether it is a biological father or related, even you can look at relatives, for example. The similarity is going to decrease with genetic distance decreasing. My pattern of DNA profiling will be more similar to my father as compared to, for example an unknown individual and so on. So, we can really use this to solve many questions. So that's that powerful and with that we will be ending the last lecture of this particular section, week II and we will be meeting again in the third week lectures, where we will be looking into various different types of mutations and how they may contribute to the disease pathology. See you again.