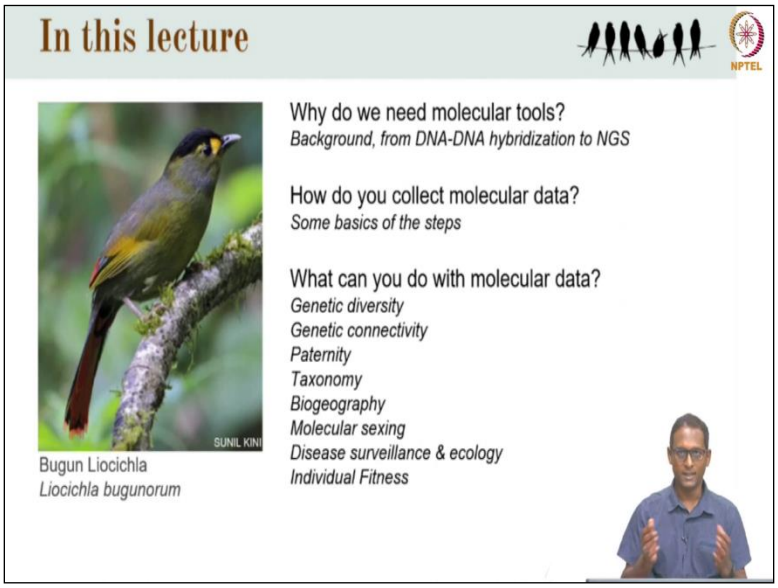
Basic Course in Ornithology
Dr VV Robin
Indian Institutes of Science Education and Research-Tirupati

Lecture 40
Molecular Techniques Part 1

Hello! I am VV Robin or Robin Vijayan and I am a faculty at IISER, Tirupati. Today we will talk about molecular methods and various aspects of molecular methods and how they can be used for ornithological research.

(Refer Slide Time: 00:36)

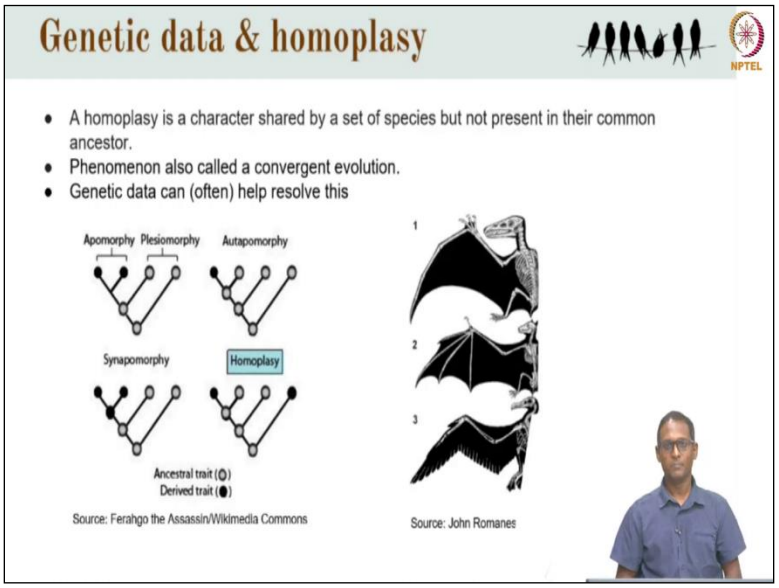


First, the outline of this talk of this lecture is that we will first just talk about why we need these molecular tools. I will give you a little bit of background. Primarily based on genetic relationships between different species and there we would talk about DNA-DNA hybridization studies that were pioneering at the time to next generation sequences which are becoming increasingly common today.

We will very briefly touch upon how one collects molecular data because if you are an ornithologist and are interested in collecting molecular data, you probably need some information on how to do that. So, we will very quickly summarize these basic steps. We can we will then get to the main part of the of this lecture which is what you can do with molecular data. And here, we will talk about genetic diversity how you measure genetic diversity,

how one can look at genetic connectivity. You can examine paternity of different organisms, how you use molecular data for taxonomy, very little about biogeography, but we will just touch upon that we will also talk about molecular sexing. This is particularly relevant for monomorphic birds, disease surveillance and ecology and finally about individual fitness.

(Refer Slide Time: 02:16)

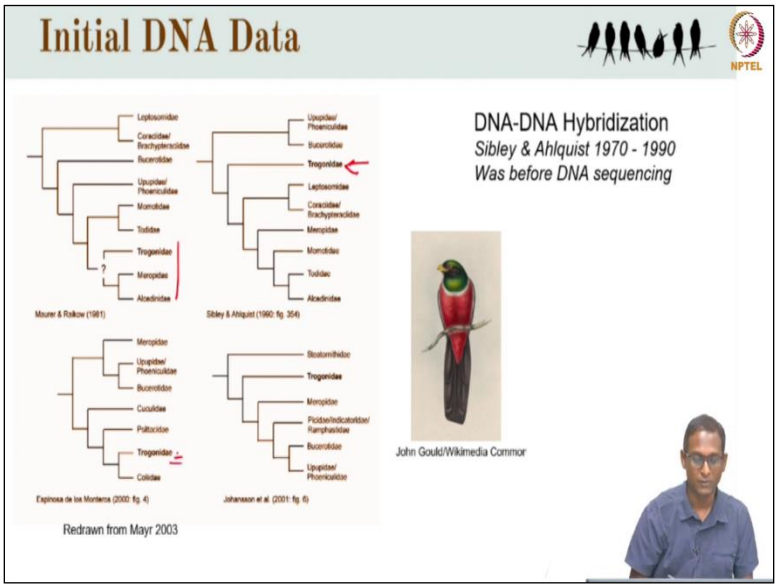


So, my particular focus especially in this lecture is on using molecular data to understand relationships between different species. And here, one important point that people have to keep in mind is about homoplasy. Homoplasy is nothing but a shared you know a shared character and it is shared by a set of species which does not have a common ancestor. For example, this you see here, there is a trait which is found in a species which is here and another that is here and this kind of a trait which kind of originates independently a derived trait that is essentially considered to be convergent evolution or homoplasy.

A famous example of this is the flight you know the flight of different organisms; birds and mammals and dinosaurs all shared the same kind of plain for flight and that is convergent evolution essentially because they are not related to each other although the dinosaurs and birds are. But in this case, you can consider that as a convergent evolution. Now, how you can trace or you know identify such convergent evolution you cannot sometimes rely on morphology alone and that is where molecular data (genetic data) comes in comes into play.

So, while we will highlight different aspects of molecular data we will also try and tell you a little bit about the drawbacks of that. So, it is not that any method in science is without its own drawbacks and limitations. And you must view molecular data also in the same light that it has its own limitations and you must use the appropriate molecular data for the appropriate question.

(Refer Slide Time: 04:39)



So, the initial days of understanding species of course, if you look at the DNA-DNA hybridization studies which were famously created by Sibley and Ahlquist between 1970 to 1990, they looked at relationships between different organisms. And this was based on a very simple idea that. So, you must remember that this was before DNA sequencing was actually you know invented or it became popular.

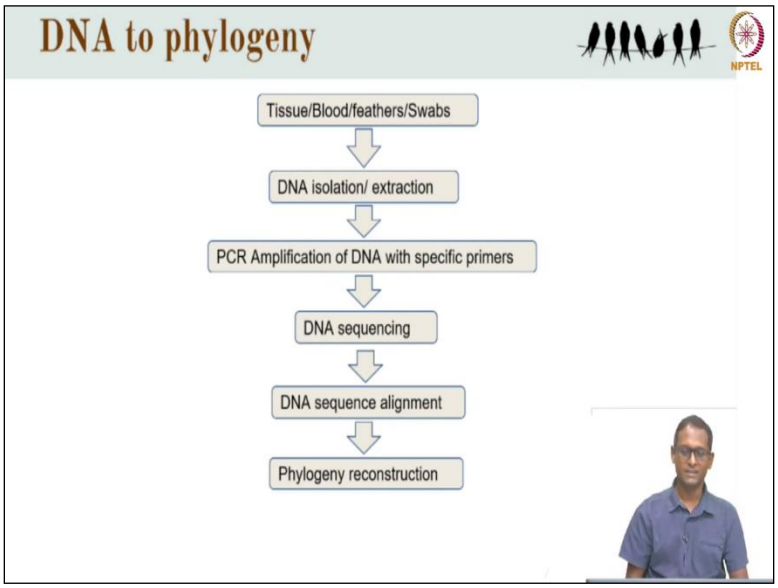
So, what they did was they took the DNA of a particular species and another and then they heated these two strands up and when they formed when they hybridized that temperature was noted. And they used this data to kind of look at similarities and differences between different groups. And as an example I have picked out relationships of Trogons.

And you know initially trogons were thought to be closer here, but Sibley and Alhquist they placed Trogons here. And today with more and more data you find that that is you know it was challenged depending on different kinds of data, the position moved around. But you will also notice that

some other groups had interesting relationships. For example the Upupidaeand Bucerotidae today relationship really first came up with Sibley and Alquist data and that has kind of more or less stayed consistent through other investigations also.

So, DNA-DNA hybridization although is very preliminary provided the first kind of insights to some convergence and relationships that were not otherwise evident by just looking at morphology of species.

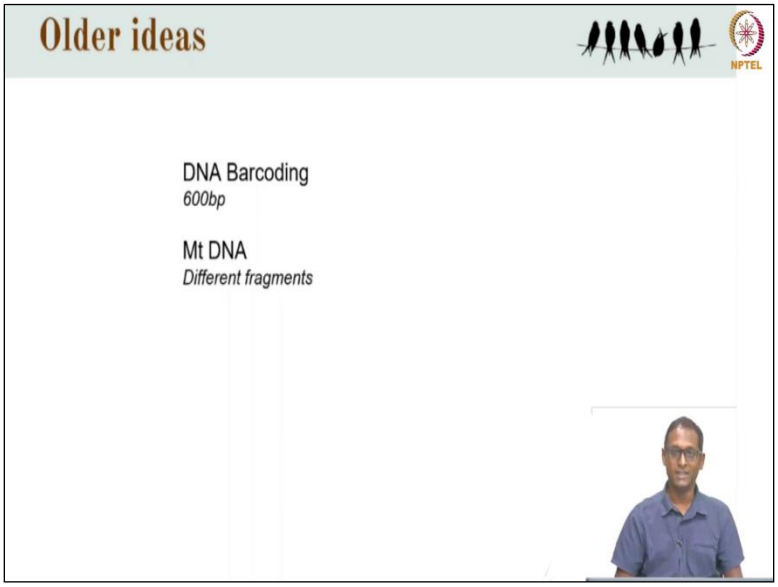
(Refer Slide Time: 06:51)



Now, we go from you know the bigger picture to we move down into the nitty gritty of this we try and understand how does this actually happen. So, how do you get to this phylogeny? How do you get genetic data? And this is through a process that I will walk you through. But essentially, the first step is of course collecting the tissue or the blood sample. This can be done by various ways- tissue, feathers, swabs, buccal, swabs and so on.

And then DNA is isolated and from that isolated DNA, you can amplify parts or specific parts of the DNA with specific primers. If you do not completely understand this, we have lot of details coming up for you. Once that DNA is amplified, it can be sequenced, the sequences are aligned and then phylogeny is constructed. So, this is the entire this is the flowchart of what happens you know from when you start your study on two species that you think are similar or different and then you create a phylogeny to examine that.

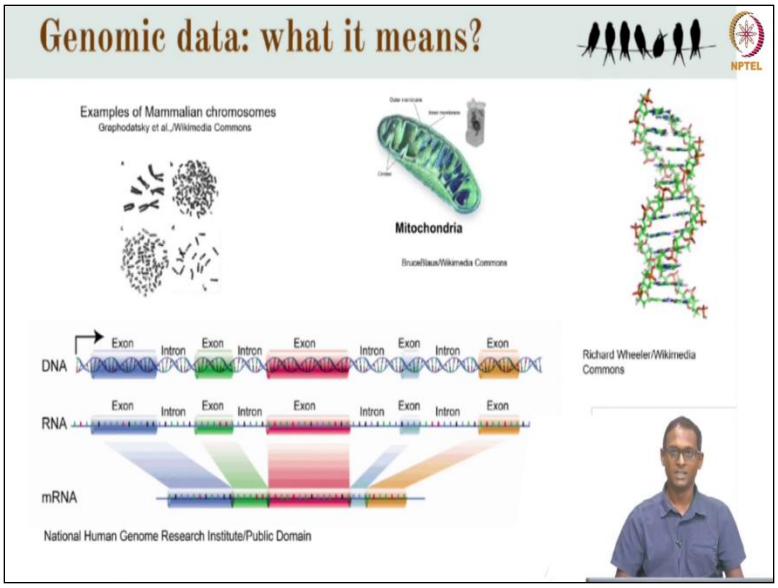
(Refer Slide Time: 08:11)



So, there were a lot of different ideas after DNA-DNA hybridization. A few other universal markers universal ideas were proposed. One of them was of DNA barcoding and this was a particular region of mitochondrial DNA about 600 base pairs that people started using to understand species level differences. So, the simple question was, is this these two groups that you see, are they the same species or not? And a universal kind of cutoff was proposed and so on.

This was popular especially at a time when there was no other genetic data that was available. So, this was the first step towards understanding species, you know to understanding the boundaries of species divergences you could say. Later people have used mitochondrial DNA, different genes or different parts of mitochondrial DNA more extensively to look at relationships between organisms. So, this was all sequences that were used unlike the DNA-DNA hybridization studies.

(Refer Slide Time: 09:30)

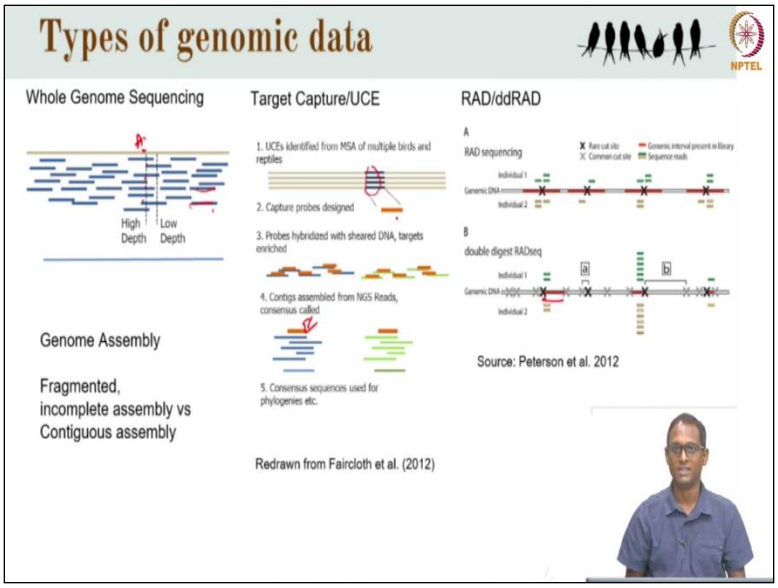


So, when we are talking about molecular data, what does it actually mean? Of course, we know that DNA itself is a double helix and so on. But, actually people can use different parts of the genome and you have mitochondrial DNA which has a maternal line of inheritance and then you have nuclear DNA with several chromosomes and so on. And of course, DNA's have exons and introns.

And exons are the ones that do get translated into the mRNA and introns do not and but you will find that to understand relationships, people can use both exons and introns, they could use either. Introns tend to have a lot more diversity than the exons because exons translate to something specific. So, depending on your questions you can target specific parts of the genome. So, it is important to understand which part of the genome the data is coming from.

And in fact, that has a relevance for the question that you ask, you know the biological question that you are asking. Your choice of the part of the genome that you want to get a sequence for depends on the question itself.

(Refer Slide Time: 04:39)



And when we talk about whole genome data, which is with the next generation sequencing, there are different kinds of data again that people use and I am this is by no means an exhaustive survey, but I will give you three examples. One is when people do whole genome sequencing, there are very nice examples of bird phylogenys which are created now with the entire genome. But here, you have to remember that actually it is not like you get the sequence of the whole genome just in one shot.

What happens is that you have little fragments like this that you get sequences for and all of these fragments are kind of put together and here there is an idea called depth which means that at a particular site, how many different you know times do you sequence that area. And this gives you the power of how reliable your estimate is of a base being a certain base at that particular point. So, that is the depth.

So, you can have high depth or low depth these are things that you should that you should look for and you can have fragmented incomplete assembly versus a contiguous assembly. So, essentially, depending on the depth and the coverage your whole genome sequencing can be very detailed which is very expensive or it can provide you a little bits of information from all over. So, this is a trade-off here.

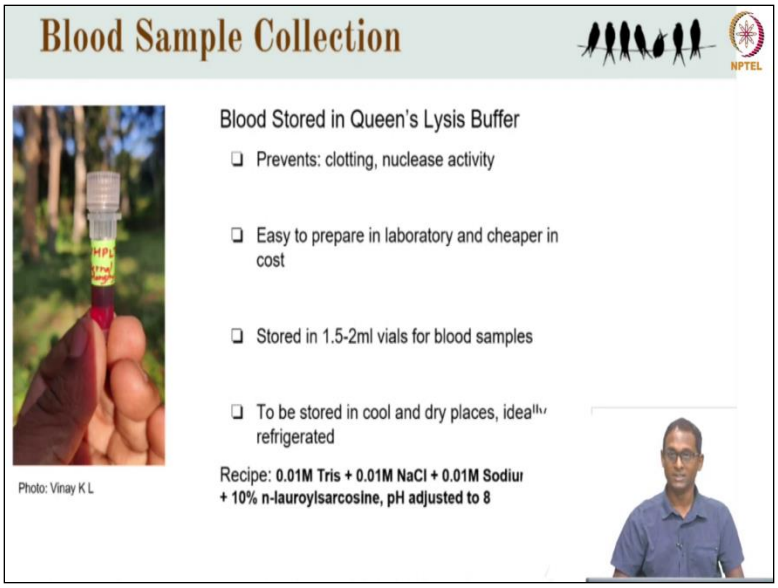
One of the popular methods in birds especially is a target capture method that particularly the one that is called the UC, Ultra Conserved elements. Ultra Conserved elements are parts of the genome which are like this, which are conserved across different vertebrates. So, Faircloth et al they identified some parts of the genome that remain conserved across Tetrapods and they designed these probes, probes are specific primers that kind of attach to that part of the DNA.

So, because it is conserved they were able to design something that works for a variety of species and through various steps where the DNA is sheared and then it is kind of enriched with some probes and so on. Basically, what they are able to do is read the sequences around these Ultra Conserved elements (these parts) and then you get some reads just like what you get here but except these reads are around just those probes that we saw here. yeah

And you put them together in the same way as you put these together and you get consensus sequences which you can then use for creating phylogenies to understand relationships of species. You also have another technique called RAD seq or the ddRAD and this one is based on a set of restriction digestion enzymes which cut the DNA, they share the DNA at specific sites uh. So, the regular RAD seq sequencing has just one cut side, but the double digest actually has two. It uses two enzymes and then the sequences that you get are between those two. So, it is between the dark and the light gray but as opposed to the RAD seq where you get the sequences from across there.

So, you get the sequences and then you put them together and then you get the genetic data that is then further used for examining relationships of different organisms.

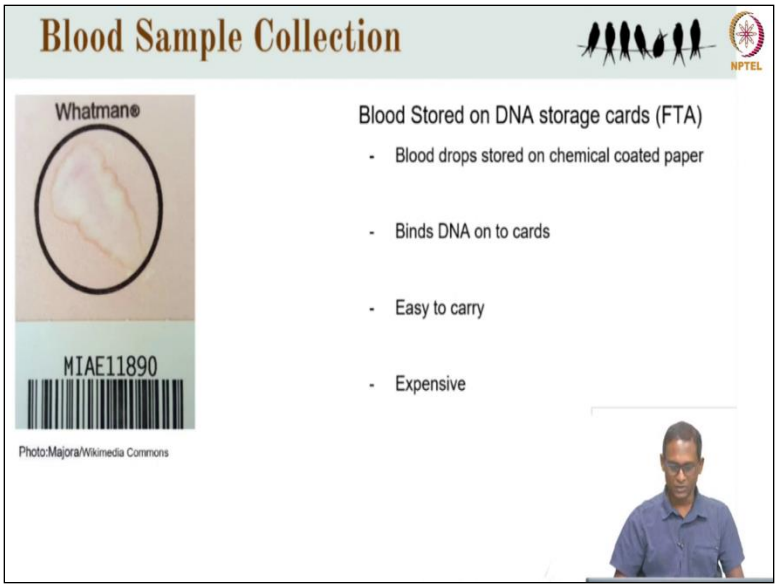
(Refer Slide Time: 15:27)



But how do you start the whole process and you start by collecting samples (blood samples), we are not talking about the field protocol here. But let us assume that you have captured a bird using mist nets or some similar something similar and you collect a drop of blood and this drop of blood can be stored in different ways one of the popular ways of storing blood especially on field is in what is called the Queen's Lysis Buffer.

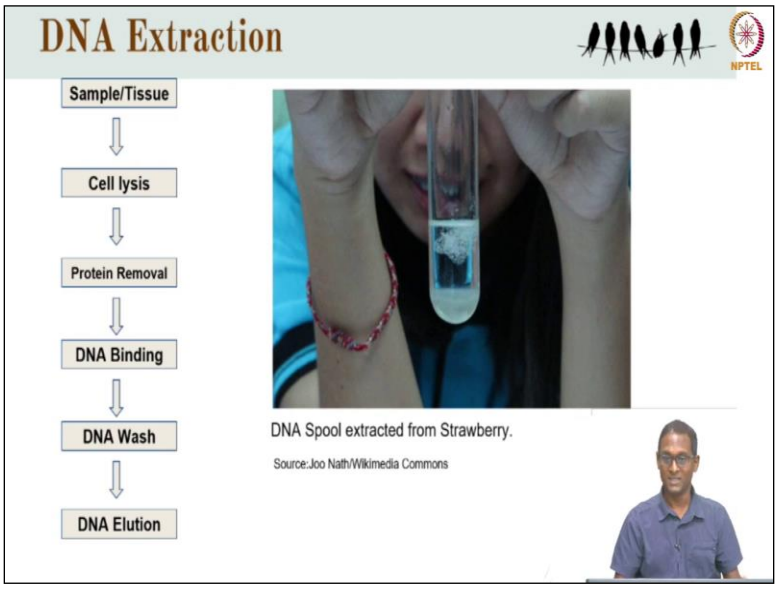
This essentially prevents clotting and it is easy to prepare in the laboratory and it is fairly cheap and you can you can store them at room temperature. They should be ideally refrigerated but it can it can preserve in room temperature until you get back to your laboratory. So, they are very useful for field work in remote locations.

(Refer Slide Time: 16:22)



Blood samples can also be collected on FTA cards which are essentially like a special filter paper, these are filter papers. But these are specifically created for storing blood samples and this is fairly simple all one has to do is put a drop of blood on the on the Whatman filter paper card and they are easy to carry because there is no liquid it is fairly straightforward. However, there is a trade-off here which is that it is fairly expensive. So, that is something to kind of keep in mind.

(Refer Slide Time: 17:00)

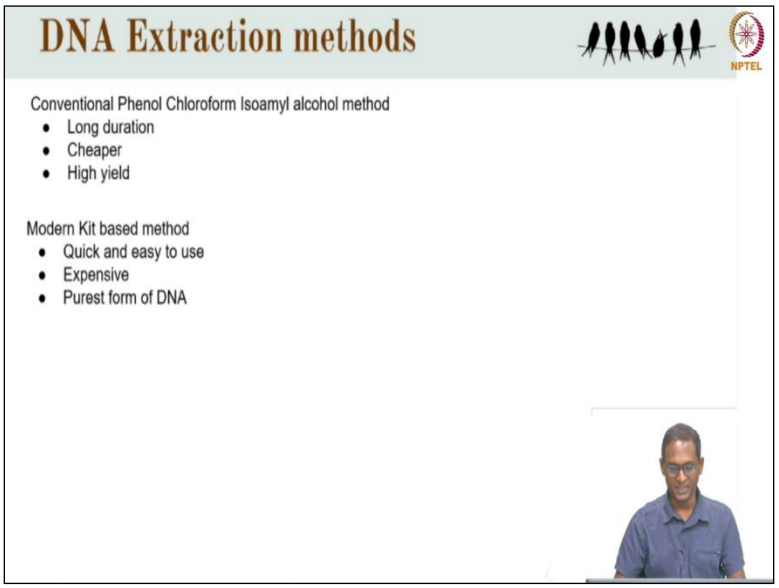


So, what does one do once you have the you have the blood sample you bring it to the lab you have to still extract the DNA. It is not enough if you have the sample, you need to get to the actual DNA. And that for that there is a process of DNA extraction. You start with the sample which is

the blood or tissue sample, there are various procedures. But essentially, it involves removing the you know you lyse the cell you break the cell wall.

Then you remove the protein and then the DNA is bound to a special kind of filter and then you wash and then you elute the DNA (which is you take only the DNA out). This is again I mean it is I am just putting this in words I think it is better understood when one actually practices it in the lab. So, it is it is like *kung fu* or *swimming* I can describe it, but you need to kind of practice it to get the feel of it. By the way I do not know *kung fu* or *swimming* but that is a different matter.

(Refer Slide Time: 18:12)



So, DNA extraction methods can be conventional the Phenol Chloroform method which is longer in duration, but it is cheaper and it is thought to provide higher yield. So, a lot of people still use it. You also have modern kits which are very easy to use it just says you know pour liquid A spin it do you know add liquid b and. So, on and it is very, very simple anybody can do it unfortunately it is a little bit more expensive but it is thought to give the purest form of DNA.

So, once you have the DNA then you reach the next point which is the PCR you need to kind of get to a specific fragment which you then want to analyze. So, how do you get there and for that you have to understand how the PCR works. So, PCR is Polymerase Chain Reaction and I am going to play you a video.

(Video Start: 19:21)

So, essentially what the PCR does is, it goes from you know from the sample DNA to copies of the target DNA which is then used for your research. And let us go a little bit deeper into that. The PCR technique has a thermal cycler which is what you see here and that for that to work. Essentially, you have in your tube, your DNA, you have primers you have nucleotides and you have a polymerase which is an enzyme.

And all of them are mixed in a tube and that is what goes into the into the thermal cycler what the thermal cycler does is that it moves the temperature between different specified points and. So, you have the target DNA that is shown here and it goes through various cycles. So, the first part of the cycle is something called denaturation where the temperature is increased to a very high degree.

And that kind of separates the two strands of the DNA and then your polymerase which is a heat tolerant enzyme comes in the primers and the free nucleotides come in. The primers are specific sequences which target parts of your the DNA and they bind there. And the polymerase start acting by adding these nucleotides to the DNA and they create a copy essentially. And it goes through a phase of extension which is an intermediate temperature.

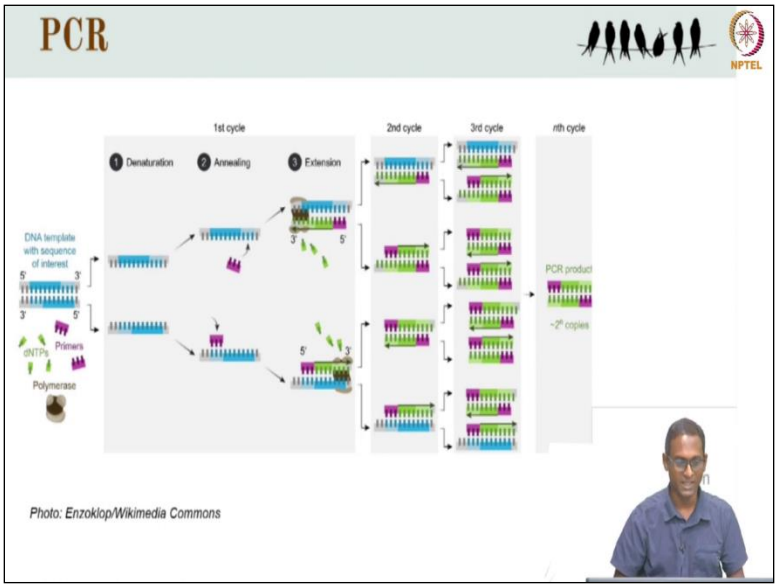
And this process then repeats and the template from the previous step becomes what you use. So, essentially that creates more numbers again creating the polymerase adding the nucleotides and creating another complementary strand and this is then kind of the process is repeated and this goes into another cycle. So, what we just saw was cycle two as it goes to cycle three.

You can imagine that again the strands are split and more complementary strand is produced and goes through the same extension phase. But what is important to understand is that we are starting with one strand of DNA one template. So, to say and which is your target DNA and then from that we are kind of getting multiple copies of that. And yeah, just focus on the cycle which is the number of cycles that we kind of discussed.

And as time goes by the copies of the target DNA kind of keep increasing and within about let us say four hours if you if you run about 30 cycles or 40 cycles, you get from a single strand you know so many copies of the DNA. So, this is a truly remarkable idea so, to say.

(Video End: 22:45)

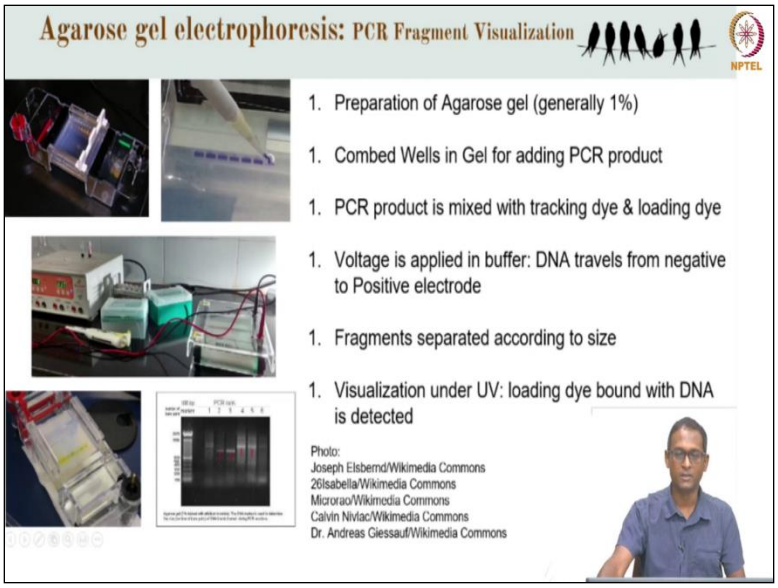
(Refer Slide Time: 22:46)



And this is what is very popularly used by a lot of molecular biology labs. Again, if you kind of missed anything essentially you go through these different cycles which includes denaturation. annealing and extension and the same thing kind of repeats in multiple cycle and at the nth cycle essentially you have two power n (2^n) copies of your DNA. And this is what then goes into your sequencing.

Because now you have only the target area which is amplified and you know found in a lot of different copies now.

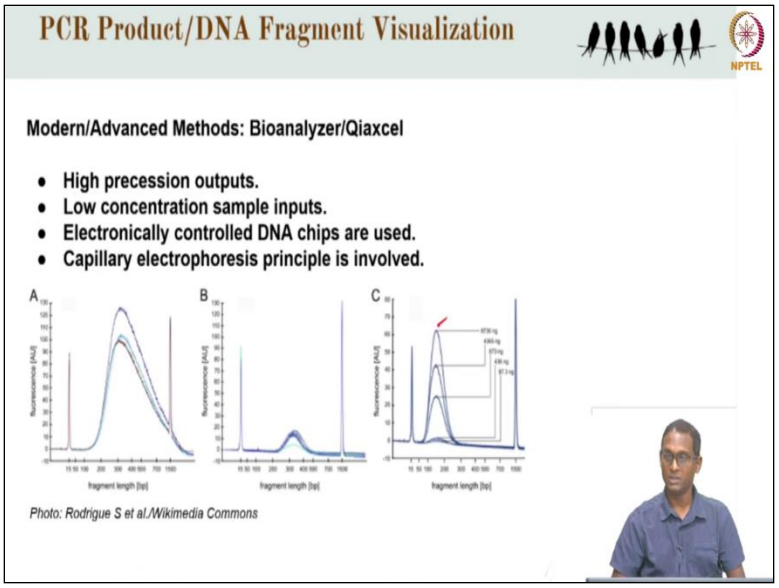
(Refer Slide Time: 23:23)



You can you can do if you want to visualize this one way to visualize this is with gel electrophoresis. You prepare agarose gel which looks like this usually about let us say 1%. And in the gel there are these little wells there are these holes where you add the PCR product and it is mixed with a tracking die and a loading die. So, that it kind of sinks in then you put it in a buffer and apply a current to it.

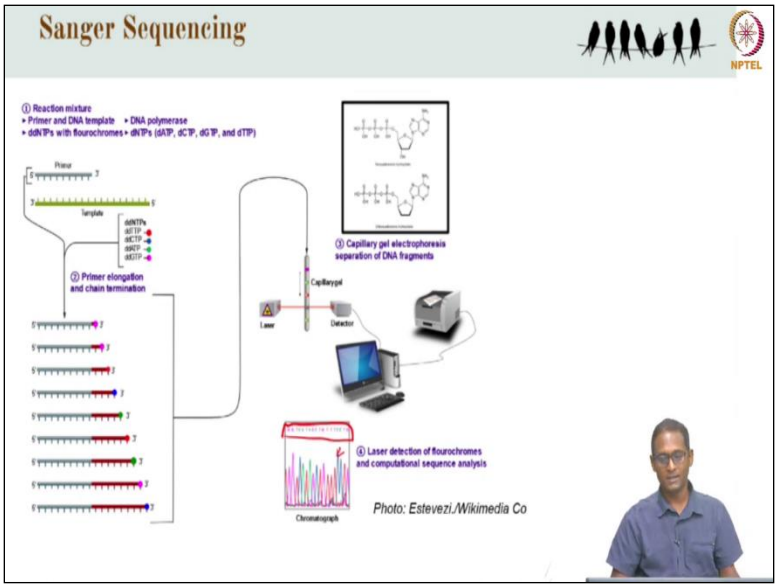
So, voltage DNA travels from negative to positive and the fragments then separate out according to size DNA travels from negative to positive and the fragments separate out to size. You can then visualize this under UV you can see the different this is a ladder, you can see the different sizes of DNA that get detected.

(Refer Slide Time: 24:23)



There are more modern techniques that are there which use some form of the same idea which is electrophoresis but you the technique is more complex and you can actually get digital outputs of how much of the fragment (what is the length of the fragment and you can quantify the signal strength which is how much DNA there is in nanograms that is what that is what is shown here). So, these are some more advanced techniques unfortunately it is more expensive. So, it is usually used when you need much greater precision for your data especially for next generation sequencing applications.

(Refer Slide Time: 25:17)

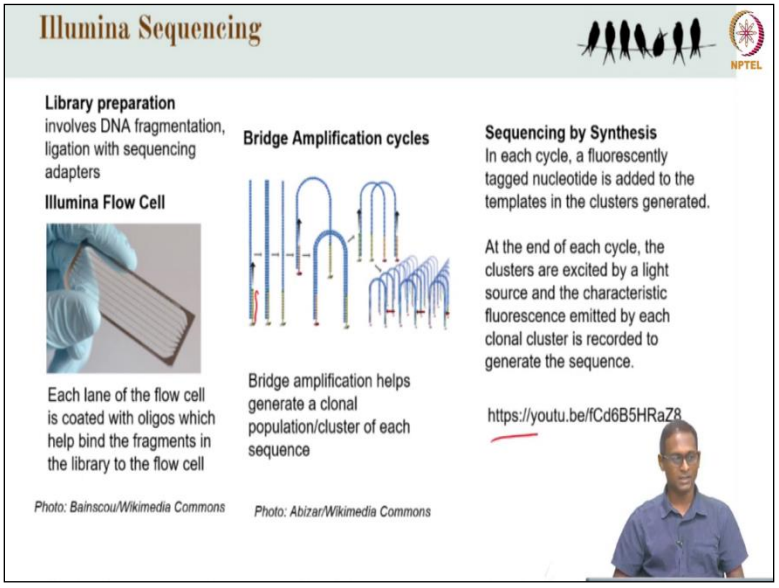


So, you go on from let us say visualizing your output of a PCR to sequencing which is that you want the sequence data the ATGC you want to kind of get that data and for that you run it through a machine called the sequencer and the popular one is the Sanger Sequencer which is the workhorse of a lot of laboratories across the world. How this works is that it is very similar to it is a combination of a PCR and a gel in some ways.

So, you go first you add a primer and the DNA template which you got from your PCR and this time when you add DNTPs it has a fluorescently labeled DNTP's which is you know red, blue, green and pink and so, on and when these specific bases get added because they are fluorescently tagged and as it runs through this capillary gel there is a laser that actually detects it and you can look at the intensity of that and you can say OK

so, that was C because that is it has a blue tag to it and then you actually get to visualize this as a sequence and then. So, you get the sequence out from this process. So, this essentially tells you how you go from samples DNA that is collected on field until you get the sequence.

(Refer Slide Time: 26:53)



There is the next generation sequencing one of the most popular version is Illumina Sequencing and this uses not exactly the same but somewhat similar an idea but they have at the core of it a flow cell and the Illumina flow cell has a lane which has it is coated with oligos which are like this

here. And it helps to bind the fragments of a library that you have to kind of prepare earlier and this gets amplified.

So, the difference the key difference here is that this process happens massively in parallel as opposed to you know one single strand that is getting detected and getting called uh. So, this is this idea of bridge amplification is revolutionary in some ways. And it creates millions of reads which is then later using bioinformatics they kind of put together and assembled in pieces like the example that I showed a few slides earlier.

I would recommend that those who are interested to know a little bit more about the process you can watch a video that is created by Illumina which is really fascinating. It will provide a lot of insights. So, that was everything about how the data is generated from free from the field it gave you a little bit of idea about different methodological procedures that go into it just an overview.