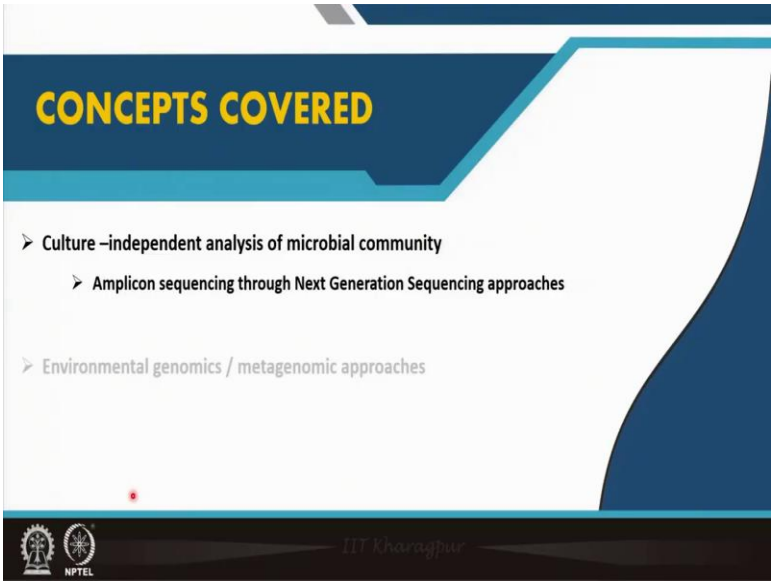
Environmental Biotechnology
Prof. Pinaki Sar
Department of Biotechnology
Indian Institute of Technology, Kharagpur

Lecture – 32
Methods in Microbial Ecology with Relevance to Environmental Biotechnology (Contd.,)

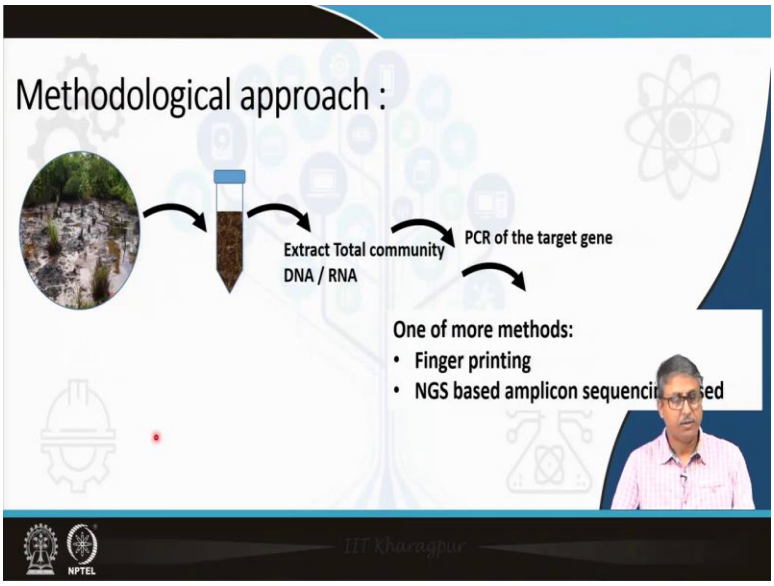
Welcome to the next lecture of this environmental biotechnology course. In this particular lecture we will continue our discussion on different methods which are applied in environmental biotechnology to understand the microbial community composition and their function.

(Refer Slide Time: 00:45)



So, in today's class we are going to address the following concept we will discuss about the the culture independent analysis method but particularly today we will focus on amplicon sequencing based methods which are carried out through next generation sequencing approaches.

(Refer Slide Time: 01:10)



So, the overall outline of the microbial community analysis within any environmental system particularly if we are adopting a culture independent method is basically the collection of the sample and it could be as many representatives as possible. However for simplicity only one sample is actually shown here but it could be as many as possible and then extraction of the total community DNA or RNA.

Now when it is said that the extraction of total community DNA and RNA it is to be understood that it intrinsically refers to the point that all the cells present within this sample are lysed and they are DNA and if required the RNA are obtained and a mixture of this nucleic acid is obtained. So, finally that mixture of the nucleic acid it may be DNA or it may be RNA as per the our experimental decision and then the PCR of the target gene.

I will come to come to this point that why do we need to have a PCR reaction over here and then once the PCR reaction is done we get a product and then one of the few methods that are available that is basically the fingerprinting methods or the next generation sequencing based methods are applied too. Identify and determine the relative abundance of the target gene the different target genes. Now as I mentioned one intrinsic assumption in this case is that the sample will have many cells.

So when we take this sample let us take an example that we are having the sample in a small tube

and here we have the sample. So, we are adding our reagents what reagent these reagents are basically for cell lysis and DNA recovery. So, we add chemical reagents into the sample. So, this is the sample. So, this is the sample and this sample is ideally consist of many cells. So, there are many cells as I mentioned.

So, all these cells are subjected to this chemical reagent like lysis reagent what you call which actually degrades the cell wall and membrane and allows this the cell cytoplasmic material and the nuclear material to be released into the solution. So, all these cells are broken and we get the DNA or the genetic material the DNA or the RNA and even the proteins and also the metabolites released into the solution.

Now this solution is processed. So, that we get the desired molecules now if we are desiring DNA. So, we will have according the required treatment. So, that only the DNA are recovered if we want to have RNA then RNA will be recovered if we want protein or any specific metabolites with that will be produced if we want carbohydrate or lipid that can also be produced. So, these are all based on the reagents that we select.

So, the kind of reagents, so, one part of the reagent is the basically the cell lysis reagent and another part of the reagent is the particular type of macromolecules that we are looking for but the ideal of the idea behind this the entire process is the extract the total. So, the importance of the total community DNA is very important. So, we must must understand this very clearly that here is no isolation of bacteria or no isolation of fungus.

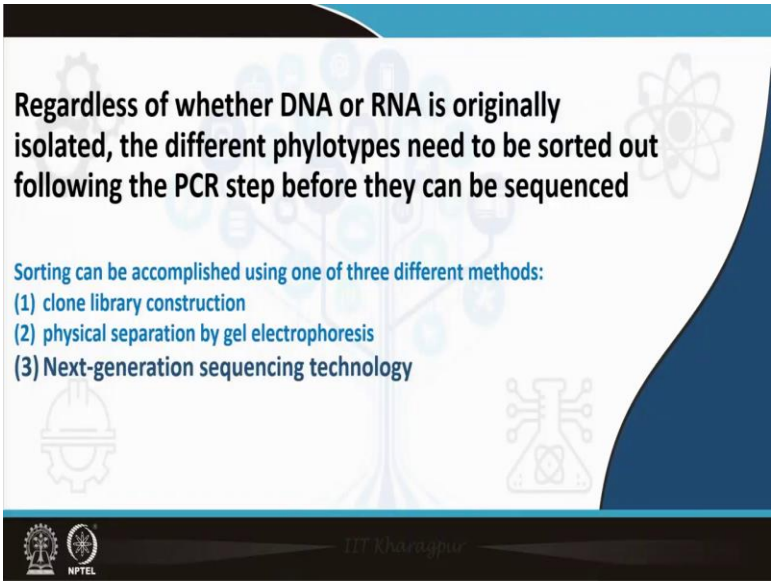
Not that we are going to have separate cells with us and then the cells will be subjected to treatment though that that could be one of the options but in general in situ lysis. So, this is basically called in situ lysis. So, we take out the sample and in the sample itself we add the reagent lies the cells and then we get the nucleic acids out of it. And then it is all about the chemical treatments and with those chemical treatments.

Basically allow the segregation of the desired molecule like the DNA from rest of the molecules and then once we have that desired molecule with us. So, that is the DNA or the RNA. So, as I

mentioned if someone wants to work on protein or carbohydrate that can be also done similarly but that will be a different process or experiment all together. So, if suppose we have the DNA or RNA then we will subject to a p subject this to a PCR reaction because we want to study a particular gene now once we have that particular gene PCR.

So, that the PCR product will be there and then the PCR product will be subjected to the one or the more methods.

(Refer Slide Time: 07:23)



Now here are; couple of questions now since this is having many cells. So, the sample is having many cells. So, sin cell samples are having many cells and these cells are of different species or phylotypes that we refer to them. So, we need to segregate or sort out these products. So, since we have we have actually taken the sample we have the sample with us from the sample we have extracted total DNA now this total DNA is subjected to PCR.

So, we have this amplicons we will call let us say we have amplified the 16s ribosomal RNA gene. So, we have 16s RRNA gene from since the sample is having many species. So, many species means many different 16s RRNA genes now interestingly all these 16s RNA genes are of same size why they are same size because we have used you we have actually use a set of a set of primers which are always going to going to bound to the a particular location and then will allow amplification.

So, we have same size but they are different in terms of their base composition that is their nucleotide based compositions are going to be different. So, basically they are different 16s RRNA genes. So, many different species, so, suppose if we have approx we just assume that 100 different species are there then there will be ideally 100 different 16s RRNA gene are supposed to be there.

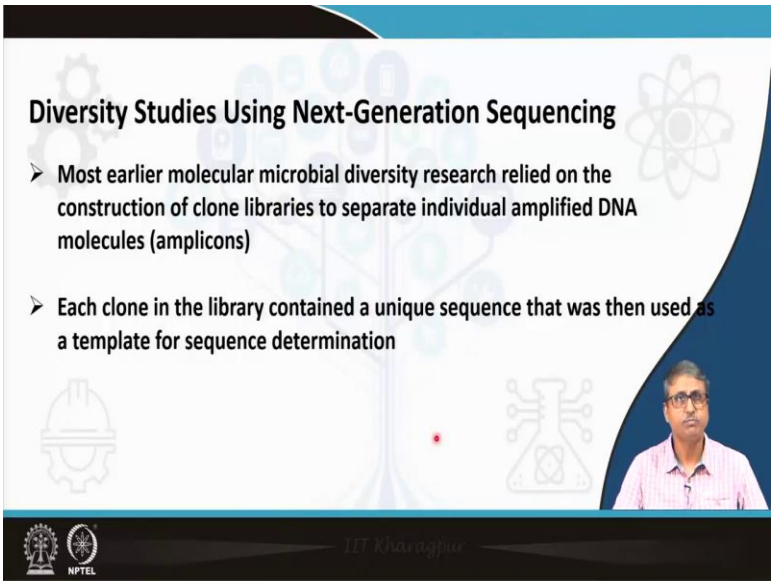
Now how do we segregate this? So, these are actually some kind of expression that we that we try to do that is the phylotyping. So, these 100 different types of 16s RNA genes are basically referring to 100 different types of phylotypes now how do we segregate those 100 different phylotypes. So, for segregating those 100 different phylotypes we know that there are a number of methods.

We have already discussed in our earlier lectures that these can be segregated by clone library construction or a method like denaturing gradient gel electrophoresis and then that is the physical separation of on the gel gel electrophoresis. And then there could be a very interesting method that we are very advanced rather that we are going to discuss today is the next generation sequencing technology.

So, all these methods are basically allowing us to decipher the different types of 16s RNA genes which are in this case because that is that is that is the gene that we are targeting in this case. But if someone wants to target some other gene for example the 18s ribosomal RNA gene in case of eukaryotes or may be some functional gene of any kind of functional genes which are very specific for any particular function that can be also be taken.

But we can we can take that assumption again that many different species will have many different version of that that particular gene that means the best sequences might be different. So, how do we segregate those different types of phylotypes in this case we are we are telling them.

(Refer Slide Time: 11:41)



Now conventionally diversity studies when we started with next generation sequencing method before that most methods that is the most molecular microbial diversity methods which are implemented in order to decipher the community structure in different environmental biotechnology processes they relied heavily on the construction of the clone libraries. So, till the next generation sequencing methods were available and they were affordable I will say.

The clone library based methods were very popular and till that many research groups like to use these clone library methods because they are easy to do people already know how to how to go about that you do not need the very sophisticated next generation sequencing machine rather a simple sanger sequencing machine which is more easily available in in different parts of the country. So, they are still being used very popularly.

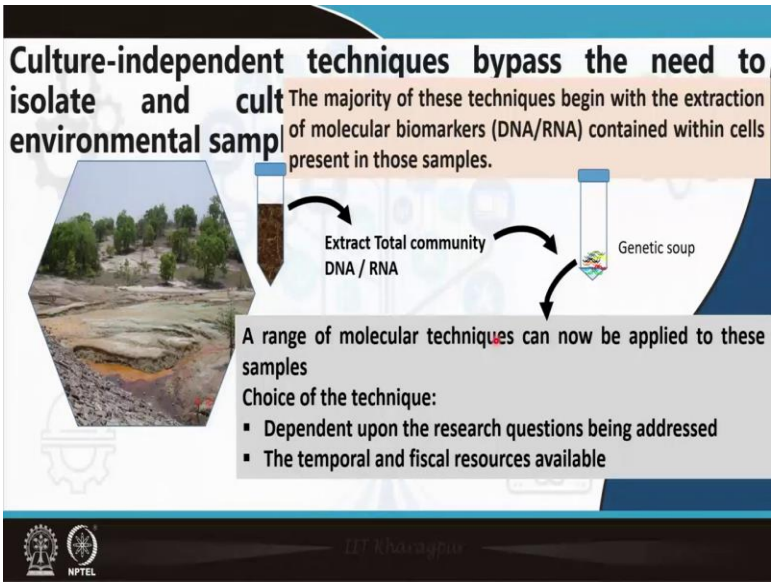
But we will see today we will discuss today that what are the drawbacks or what are the limitations of these clone library based methods or even the DGG like methods that which we cannot possibly recover or we cannot overcome those limitations. So, I ideally we should have some methods which are relying on next generation sequencing method which are actually next generation term is used only to indicate that they are of advanced generation.

Its generation means it is on only the chemistry the process the by how we are getting the sequence and the quality of the sequence and it also includes many other parameters. So, I will I

will discuss those issues that why it is called next generation sequencing but ideally it is a sequencing method that means if you have a DNA stretch you can you can get the composition of that DNA strain that is the ATGC based composition.

So, what I am referring here that molecular diversity research has used the clone library based method or the other gel electrophoresis based method like DGG for a long period of time and still we are using those methods. But now we realize that there are certain things that we must be care of. So, one basic assumption in case of the clone library based method was this that each clone in the library contained a unique sequence that was then used as a template for sequence determination.

(Refer Slide Time: 14:14)



Now most of the cases the culture independent techniques as we know bypass the need to isolate and culture microbes. So, we do not try to isolate any bacteria. Why we do not try to isolate any bacteria we can always isolate a bacteria and many times indeed that is being done for certain reasons but if we are interested to know what are the species what are the organisms present in an environment if that is our question that what are these species present in a system or a environmental process or environmental habitat.

Then isolation of bacteria may not be very good choice because isolation will lead to or give us idea about only handful of organisms or handful of or a limited number of members or the

bacteria present in the environment 99% of the bacteria or the species they will not be not be isolated and hence we will not be able to know anything about them. So, isolating and culturing them is not a very preferable way of doing it particularly when we know that this they have very serious disadvantages.

So, in this case we are again coming back to this acid mine drainage where a very highly acidic and heavy metal reach water is coming out and its I said it is a natural process and it is basically catalyzed by very specific type of microbes along with the physical chemical factors and it is contaminating the nearby areas unless and until they it is treated properly. So environmental biotechnology has a great scope to work on those type of environments.

Because in that case we can control the environmental deterioration as at the same time we can recover the useful metals from those effluents or the those water that is called acid mine drainage water. So, what is actually done in this case that basically we take the sample that is the sample again and then the extract the total community DNA and then this community DNA is subjected to some reactions which is called the PCR reaction.

Now as we possibly discussed earlier that almost all the methods we see that these techniques that culture independent techniques they begin with the extraction of the DNA because they DNA is the molecular biomarker the macromolecule is a very stable molecule inside the cell. So, and outside the cell also, so, we recover or try to recover the DNA from the sample and or we sometimes try to recover the RNA also if we have some specific questions that I will come back that.

But basically it is it is preferred that most of the time DNA but RNA can also be used. So, I write DNA or RNA. So, either of them can be or both of them can be extracted separately from the sample and these actually will include this DNA or RNA nucleic acid basically will include that the DNA and RNA from all the cells from all the cells present in this sample. So, if we have a sample. So, if it is having let us say 1000 or 4000 different species and each species having 10 or 20 or 50 numbers.

So, it is all together maybe 5 lakh cells are there or so, or may be maybe little more or less cells are there. So, we expect that all cells all cells are broken lysed and their DNA or the RNA is available to us that is called the total extraction of the total DNA. So, we have protocols which are very of well optimized protocols through which we can extract the DNA or the RNA that is again total DNA or the community DNA we call it from the sample.

So, it is it is expected that these protocols will allow us to get the DNA and RNA from all cells sometimes we need to modify the protocols if our environmental samples are of any particular category or particular type. So, that for that we need to consult the literature and see what type of modifications are to be are to be incorporated now after this extraction what exactly we obtain we what we up get actually is a mixture of all the DNA and this is the genomic DNA mostly and along with that the plasmid DNA will also be there.

So, all the DNA present within the cells and also the external or extracellular DNA because if the cells are dead then you have the extracellular DNA. So, all those DNA are basically extracted and that extracted DNA is called or the soup is called the we call it genetic soup now what exactly is genetic soup the genetic soup is basically comprising of all the genes because these are the short stretch of DNA which are functional may be producing the proteins helping in producing proteins or the ribosomal RNA or TRNA or different type of other RNA's.

So, genes, genomes transcript that is the RNA and transcriptome; that is the total pool of RNA from all the organisms whose cells were present in the sample. So, this is again a kind of a fundamental assumption that all the cells present in the in the sample are broken down lysed and all the genes or the genomes and all the RNA or the mRNA or different type of RNA's that is the transcripts and the transcriptome that is the collection of all the transcripts mRNA and RNA they are now they are in this genetic soup.

So, genetic soup is basically a mixture of all the different type of genomes and genes and other things if we are we have extracted the RNA then possibly it is the mixture of the transcript and transcriptome. If it is extraction of the DNA then we have the genes and the genomes extracted now what do we do with this genetic soup. Now a molecular number of molecular techniques

can be applied on this on these genetic soup.

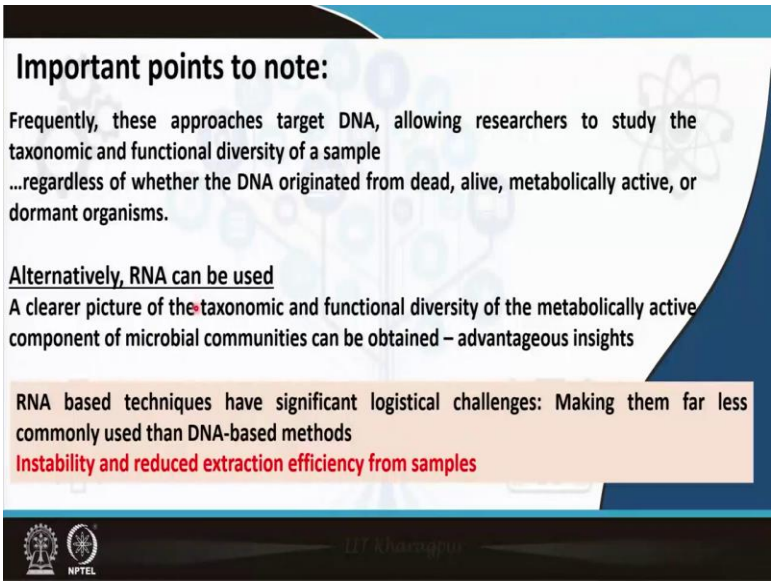
In order to decipher in order to answer the questions what questions the questions are to be framed by the researcher or the scientists who are working on this particular project that for example one of the most fundamental questions is basically the what are the organisms present there number one and number two is the what are the phenotypic or the what are the genotic potential of this organism or the what are the functional potential of these organisms and the more questions.

Like what are the actual functions carried out by these organisms and also the interactions between these organisms and their environment. So, there are 4 basic questions or fundamental questions that often we ask. So, now these questions may be more fine-tuned based on the environmental process that we are working on. Now what do we do with this soup. So, we need to actually choose the methods by which through which we can actually answer to those questions.

Now we have actually different type of techniques as we already discussed clone library DDG or some advanced methods. So, now the choice of the technique is basically depending upon the research question that is being addressed. So, I mentioned four general questions which are fundamental questions but these fundamental questions might be fine-tuned as I also mentioned in order to make them more particular.

And also it is also dependent on the temporal and the physical resources which are available that means physical difference means what are the funds the what is the how much money we have. Because some approaches may be more expensive. So, we need to actually take care of that fact also and the temporal resources that what are the resources available with the laboratory or the scientists at that point of time when they need the data and also the monetary resources that they have with them so, based on that also.

(Refer Slide Time: 22:19)



Now here are some important points to note frequently these approaches that the these approaches that we are going to going to adopt in order to decipher or get the answer of the fundamental questions of microbial ecology with respect to the genetic soup that we have just obtained. So, these approaches actually target DNA since if we are extracted DNA then they will target DNA.

Allowing the researcher to study the taxonomic and functional diversity of the sample. So, most often these are the two fundamental questions that what are the organisms present that is the taxonomy and what are the functional potential that is the functional diversity. So, we can actually target the taxonomic analysis and also we can target the functional gene analysis. Now regardless of whether the DNA originated from dead alive metabolically active or the or dormant organisms because earlier we have also learned that these there may be different type of cells present.

So, since we have adopted some methods through which all cells are lysed. So, all the DNA are available with us. So, it is a mixture of all the DNA and DNA will not tell us at this moment that whether the DNA is coming from a dead cell or a living cell or it was it a metabolically active cell or was it a dormant organism who has actually given the DNA in our soup. So, it is it is we have to take a decision and adopt appropriate method.

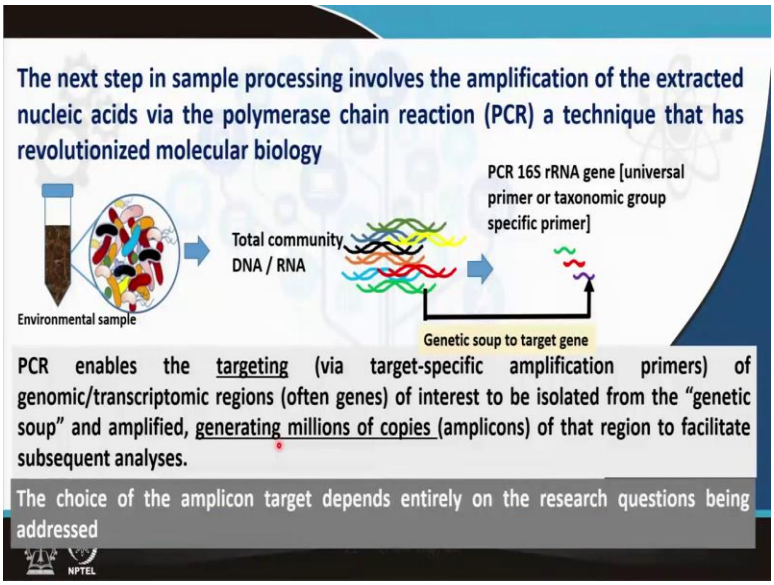
So, basically we need to target the DNA and then the other questions will answer slowly. So, there is another alternative that I mentioned earlier that RNA can also be used instead of DNA or parallel to the DNA. So, when we use RNA a clearer picture of the taxonomy and the functional diversity of the metabolically active components of the microbial communities can be obtained. So, this will provide of course lot of advantage to us because the form import from scientific insight that actually what is happening why.

So, because the RNA can be obtained only from metabolically active and living cells so, dead cells cannot give us RNA because there will not be any RNA present. So, if we are targeting RNA that means we are surely going to know that what are the metabolically active fractions of the community? So, it may be a wastewater treatment system or maybe a bioremediation field system. So, in any environmental setup.

If we want to know exactly which microbes are active then possibly the best way to do that is the addressing the RNA. However these RNA based techniques have significant logistical challenges what are the logistical challenges that they are instable. So, RNA molecules are very very instable and their extraction efficiency is also low. So, if we are planning to have RNA based experiments or RNA based investigation about answering the questions and get to know the what are the organisms and the processes going on there then we must be careful about that.

We must take appropriate care about their stability and also use a larger sample volume. So, that the extraction efficiency is taken care of.

(Refer Slide Time: 25:30)



Now once we address that issue the next step in the sample processing involves the amplification of the extracted nucleic acid via the PCR reaction that will that has actually the PCR has revolutionized the microbial or the molecular microbial ecology. So, as we have the basic understanding that the environmental sample is having multiple types of species and cells and we can actually extract the community DNA or the RNA from that sample very easily.

So, this is actually depicting the mixture of many different types of genomes and that is the part of what you call the genetic soup basically. So, numerous type of DNA and the genome and the genes are present over here. Now this is subjected to a PCR reaction. Now this is PCR reaction as you know that is basically with a help of a set of primers. So, depending upon the set of primers your amplification will go on.

So, suppose we have used a primer which can actually amplify only three types of 16s RNA genes then you will see only the red green and the purple one will amplify but if we use an use a set of universal PCR primer which can actually bind to any 16s ribosomal or energy. So, though there are primers which are called universal; 16s RNA gene primers. So, they can bind to any and you can have actually as many as.

So, if you have let us say 4000 or 5000 species. So, 4000, 5000 different type of 16s are energy. So, ideally four thousand to 5000 different 16s RNA gene products are going to be there. Now

this is basically very challenging stock task because from the genetic soup to the target gene because in genetic soup is basically mixture of all the genes and the genomes and the PCR primer will bind to the specific location and wheel amplification.

So, with as many as numbers are there in in the genetic soup we expect that. So, many numbers will be there in the PCR amplified product also now the PCR process since it is very specific it enables the targeting that is the target specific amplification primers of the genome of the transcriptome region that is the gene of interest to be isolated from the genetic soup. So ideally what we can do that the genetic soup might be having.

So, many different genes together all together if we assume that there are 4000 different bacteria or bacterial species each having maybe one thousand of different genes then 4000 multiplied by 1000. So, it is basically 4 into 10 to the power 6. So, it is 4 million different genes are there or some genes may be common. So, 4 million genes are there. So, out of 4 million genes or maybe if we take into account the genes which are common maybe there are let us assume that okay there are at least of 10000 genes which are unique, unique genes are there.

So, out of 10000 genes we want to amplify only one gene. So, that means it is possible. So, from a mixture of so, many different genes and genomes we will be able to pinpoint a particular gene of interest for example if I am working on let us say ammonia oxidation in an environmental system I want to remove ammonia. So, if I am if I am trying to remove ammonia I am targeting the organisms who are responsible for ammonia removal that is the amino oxidizing organisms.

So, in this case there is a gene called AMOA that is ammonium oxygenase or ammonia oxidizing gene. So, if we target that gene then out of. So, many genes I said there could be 10000 genes or therefore 15000 genes together. So, out of so, many genes I can easily amplify the AMO gene but again the AMO gene will be having different sequence variations because the AMO from different bacteria might have some sequence variation.

So, those will be incorporated. So, that is a different question but the point is that we will be able to have a targeted amplification. So, out of so, many genes our target we will be able to amplify

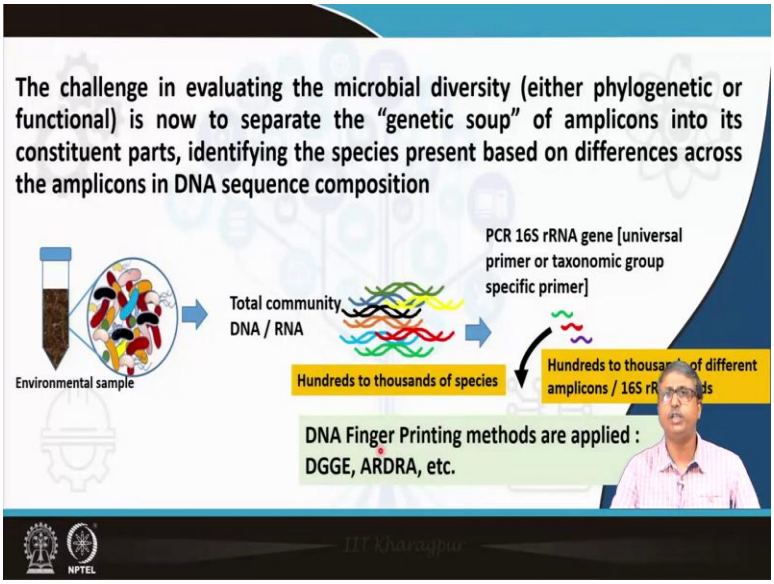
that is the beauty of this PCR reaction. Now it has another advantage the second advantage is that not only it will be very targeted but also it will create millions of copies that is called amplicons.

So, if I am targeting an AMO gene from a like might say activated sludge or a wastewater treatment environmental DNA then I will be able to amplify the AMO gene that is that is one purpose but it will be having millions of copies of that gene that is the AMO amplicon like am amplicon means will have millions of copies of or the reads of the AMO gene. Now that will help us to analyze the things further.

Now the choice of the amplicon target I have in this case taken for example the AMO because my system might be working on ammonia or oxidation but if I am working on let us say sulphate reduction then I will choose another gene which is called the DSR gene but if I am working on methane production in a biogas plant then I will select the gene which is called MCR gene but suppose I am working on just on the diversity point that what are the organisms present then I will simply select the 16s ribosomal RNA gene.

So, the target gene that we select the amplicon target is largely dependent on the research question that is being addressed.

(Refer Slide Time: 31:06)

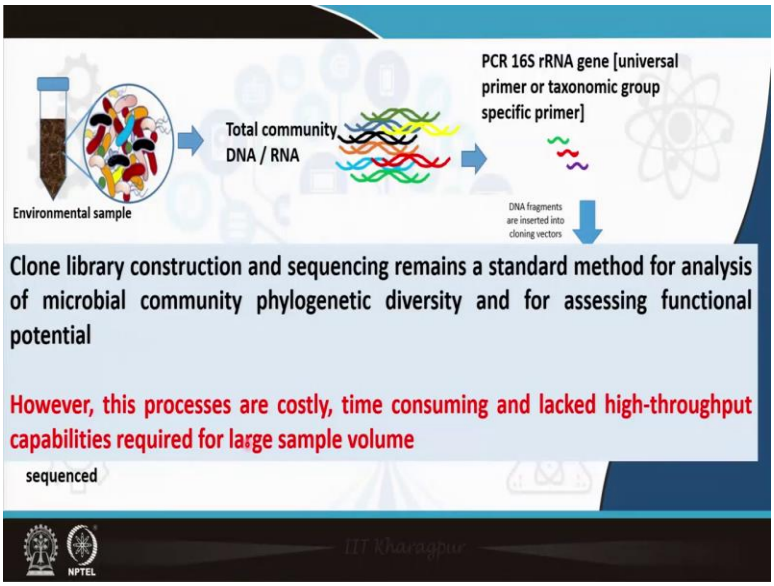


Now the challenge in evaluating the microbial diversity either phylogenetic or functional is now to separate the genetic soup of amplicon. Now since we have for example amplified the amogene from the genetic soup. Now AMO gene is not a single copy gene I said million copies of among and if we had for example 4000 different ammonia oxidizing species. So, there could be many different types of AMO genes.

So, the amplicons could be of different types. So, now we have to separate them now identifying the species. So, for example in case of a microbial diversity question, so, identifying the species present based on the differences across the amplicon of DNA sequence composition. So, how do we do that? So, suppose we have followed the same pattern that from the environmental DNA we could amplify and this amplification can be for the 16s ribosomal RNA gene in this case.

So, we could we can actually amplify using universal primer as many genes as possible. Now, 100s to thousands of different species will produce 100s to thousands of different amplicons or if we consider 16s ribosomal RNA gene 16s RNA gene.

(Refer Slide Time: 32:19)



Now DNA finger printing methods can be applied like DGG and all these things and if you are planning to have a clone library then possibly we can clone them individually they will be inserted and each of these clones will be subjected to further analysis by restriction analysis that is the ARDRA or maybe each clone can be sequenced directly.

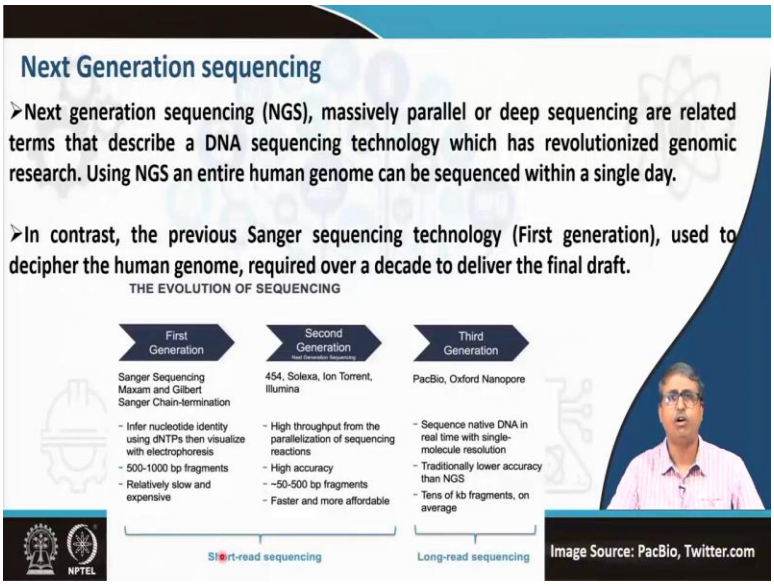
So, once we do that we will be able to identify each of these clones and will be able to identify most of these bacteria. So, far their DNA and their islands and come into this soup and they were actually amplified because later we will discuss about that point. So, clone library construction and sequencing remains a standard method for analysis of microbial community phylogenetic diversity and for assessing the functional potential not only for 16a.

But also for many functional genes that we have discussed already in some other lecture that a large number of genes are available which are very important or called marker genes for different environmental relevant process. So, based on or using those target genes. So, clone library based methods were adopted and it was very successful. So, however this process or these processes of clone library and then sequencing them are costly time consuming and lack the high throughput capabilities required for large sample volume.

So, they are costly because you have to first clone them and then you have to sequence them and also the time consuming how many clones you are going to sequence. So, that is a big question because if we assume that there are many minimum or 10 to the power 4 species or 10 to the power 5 species are there that means 1 lakh different species or 10000 different species are there then at least few 1000 clones are to be sequenced.

So, this is a very huge task because all of these clones need to be need to be processed separately their plasmid DNA need to be extracted and then they will be sequenced separately. So, they will be hugely costly and time consuming because you need to handle. So, many clones without having any contamination between them and for this process the high throughput capabilities are very limited. Generally they are not available for the common laboratories who are working on different aspects of environmental problems. So, with handling the large number of sample volume it becomes very inconvenient.

(Refer Slide Time: 34:38)



Now here comes the next generation sequence. Next generation sequencing is any of the several high throughput sequencing methods. So, there are number of high throughput sequencing methods are there whereby billions of nucleic acid fragments can be simultaneously and independently sequenced. So, there are two things. So, billions of nucleic acid fragments that means so, if your sample is having let us say 10000 different species and each species is maintained is represented by at least 10 cells.

So, that means total 1 lakh cells are there because 10000 different species each species having 10 cells. So, then the total number of cells are huge then that means when you extract them and you have a PCR amplification possibly you will have billions of reads which are called 16s RNA gene rates if you are adopting that now can that be sequenced simultaneously, yes. In next generation sequencing method we can sequence that amplicon the pool that which comes from the genetic soup can be sequenced directly.

No need to physically clone them one by one and take individual clones get their plasmids all these things are not required that is number one point number one. So, it is a simultaneous sequencing of the billions of nucleic acid fragments. So, it is actually two points one is the billions of reads are simultaneously done. So, its number is also very important it is not 1 or 2 or 3 or 4 its billions of reads are sequenced simultaneously and independently that means you can have actually multiple samples together that is also the case advantage is this is called

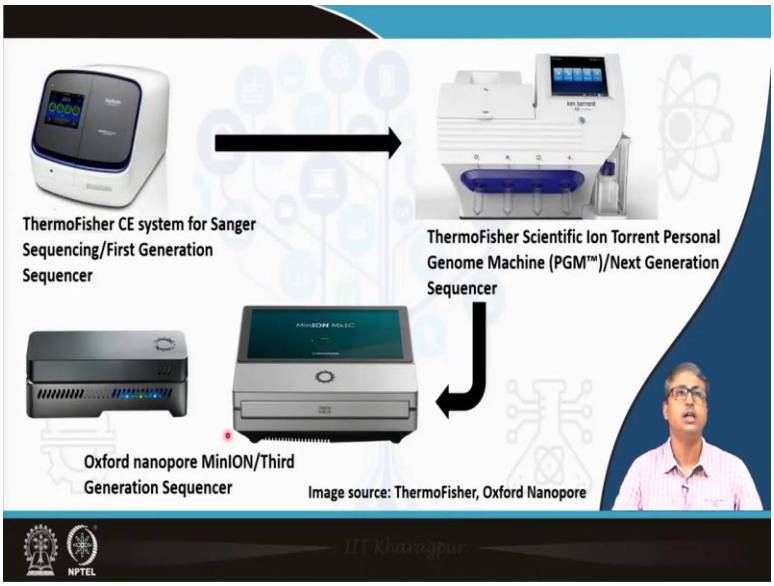
multiplexing.

So, if you are using some kind of barcoding. So, during your initial PCR reactions then possibly you can discriminate the sequence products from coming from sample one from that of sample 2. So those will be discussed maybe in little later in some other classes. Now that is something which is very interesting and it contrasts this technique the classical method such as Sanger sequencing method which processes only one nucleotide sequence at a per reaction.

During this Sanger sequencing of the first generation sequencing we are compelled to use a technology or a chemistry which handles only one nucleotide sequence at a time per reaction but here per reaction its billions of different sequences. And these are the some information about the different generations of sequencing like the first generation second generation and third generation.

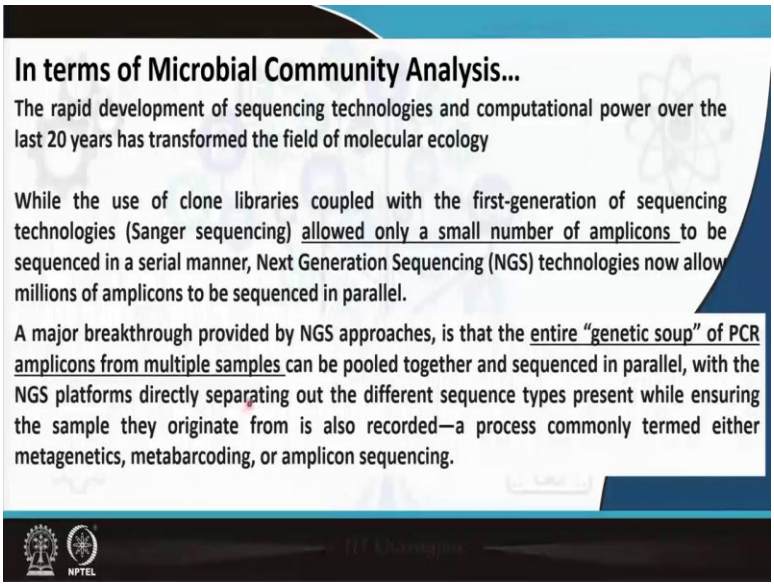
But some technologies are allowing short reads that means the read of the sequence rate could be like 500 to 1000 nucleotide or in some cases that could be large fragments like several thousand like KB, KB fragments means kilo base that means several thousand nucleotide at a stretch sequence can be obtained.

(Refer Slide Time: 37:39)



Now these are some of the photographs of the sequencing machines.

(Refer Slide Time: 37:43)



Now in terms of microbial community analysis this next generation sequencing is going to be very important and in the last 20 years or so. There is a huge development in the sequencing technology which actually transformed the field of molecular ecology. While the use of clone libraries coupled with the first generation of sequencing technology that is Sanger allowed only a small number of amplicons to be sequenced in a serial manner next generation sequencing technologies are now allowed million of amplicons to be sequenced in parallel.

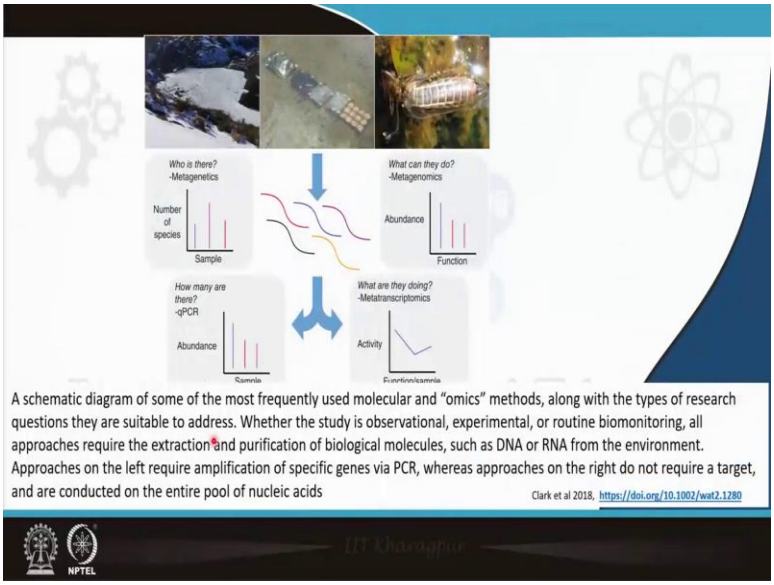
So, this what I was mentioning earlier in normal six Sanger sequencing based method you need to sequence one by one even when we have a multiple capillary Sanger sequencing machine that is a sequencer we can have a handful of samples can be run at a time compared to the next generation sequencing where millions of samples amplicons can be run together. Now a major breakthrough provided by NGS approach is that the entire genetic soup that is obtained after the lysis of the cells from an environmental sample of the PCR amplification from the multiple samples can be pulled together and sequenced in parallel.

So, you can actually sequence the entire PCR amplified product out of the genetic soup and also you can take multiple samples genetic soup. So, if you have let us say 10 samples. So, 10 samples genetic soups they are subjected to PCR and 10 now they are 10 amplicons those 10 amplicons can be run simultaneously in the this next generation sequencing platforms. With the

NGS platform directly separating out the different sequences type present while ensuring the sample they originate from.

So, they will give you read details that is the 16s ribosomal RNA gene if that is the target gene you can always select some other gene. So, the individual reads will be detected for each sample and if you suppose you have taken 50 samples. For 50 samples the these are the reads can be segregated.

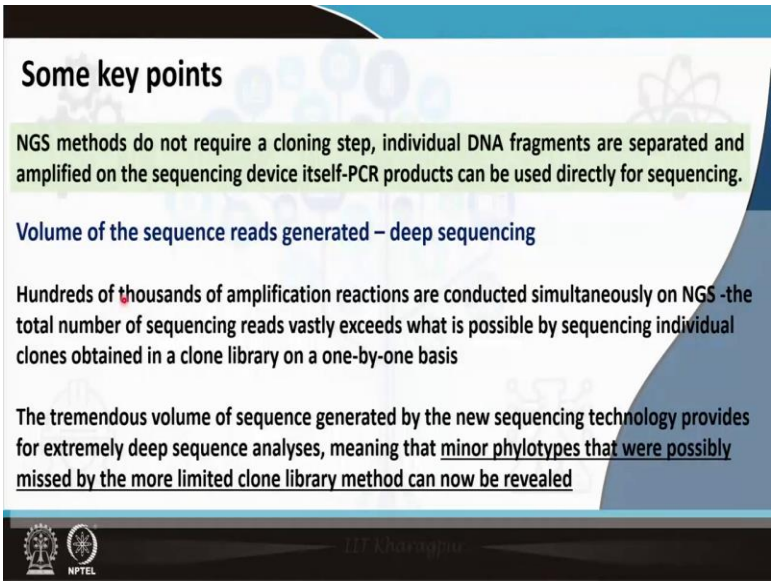
(Refer Slide Time: 39:48)



So, it is kinds of an elegant approach. Now there are something more interesting rather than this only. So, here is the picture that of the different environmental sample we get the DNA and then allow the next generation sequencing method. So, these are the different questions that often we get to answer actually. So, for example who is there that is the metagenetics and what they can do that is the meta genomics.

And how many are there that is the quantitative PCR and what are they actually doing the meta transcript of so, these three we will discuss later we are right now discussing the first point that is who is there answering that fundamental question.

(Refer Slide Time: 40:26)



So we will come to this particular part of the metagenetics. So, some key points before we go to this that the NGS methods do not require a cloning step individual DNA fragments are separated and amplified on the sequencing device itself. PCR products can be used directly for sequence. So, we use the genetic soup do a PCR and the PCR product which is having actually millions or the more billions of different type of reads of the total reads that is subjected to the sequencing machine number one.

Number two is the volume of the sequence reads generated. So, as I said the millions of reads are generated and that is also connected to something which is called deep sequencing. Since 100s of thousands of amplification reactions are conducted simultaneously on the NGS the total number of reads are vastly exceeding what is possible through sequencing individual clones obtained in a clone library one by one basis.

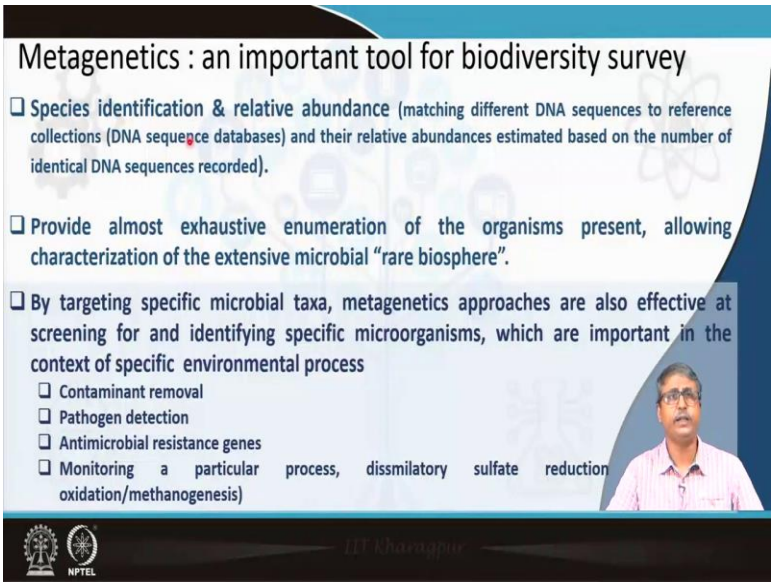
So, number one is that large number of entities truly large it is like close to a billion sequences sequence reads that can be possible and it is multiplexed. So, multiple samples together can be sequenced. Now, when we have several million reads coming out from a sample, so, one million means 10 to the power 6 reads which are equivalent to sequencing 10 to the power 6 clones otherwise if we think of a clone library based method and 10 to the power 6 means it is 10 lakhs.

So, now if we have a 10 million reads. So, it is a huge number of reads that we generate now

when we generate a huge number of leads it is expected that the organisms or the sequences which are very rare in the environment. Particular environment might have some bacteria or some archaea who are by number they are very less abundant but they might be playing very important role.

In a normal PCR and cloning procedure we might have amplified them but we might miss out them because their number wise they are so, less that unless you sequence a large number of reads you will not be able to catch them. So, the deep sequencing which is available through next generation sequencing it is basically allowing us to identify the organisms who are part of the rare microbiome. Maybe I will talk about the rare microbiome here in my next lecture.

(Refer Slide Time: 42:52)



So, the metagenetics that is that is an important tool for the biodiversity survey in environmental context. So, the issues that the metagenetics is going to address first the species identification and relative abundance. So, we will be able to identify each of the microbes present because the reads that we are going to generate will be matched to a DNA database there are beautiful database very huge database are there.

So, through the database match will be able to identify straight away that what is this read actually telling us which is the organism for which is this particular it is. So, taxonomic affiliation will be there. So, that is our species identification and the relative abundance because

when we count how many leads belong to a particular taxa will be able to able to actually pull that information that the relative abundance of the organisms.

Second to provide almost exhaustive enumeration of the organisms present allowing characterization of the extensive microbial rare biosphere. Rare biosphere means the organisms which are less abundant they may be more otherwise but they are abundance wise they are very less. So, if you try to count them they are not going to be counted because they are not generally present in so simple technique like a clone library or DGG. You will not be able to unless you do a deep sequencing you will not be able to catch them.

Third is the by targeting the specific microbial taxa metagenetics approaches are also effective at screening for the an identification of specific microorganisms. Like a specific taxon specific 16s ribosomal RNA gene you can always use during your amplification process. So, you can you know that what organisms you are possibly looking at and accordingly you could you can select the type of 16s primer which are important in the context of the specific environmental process.

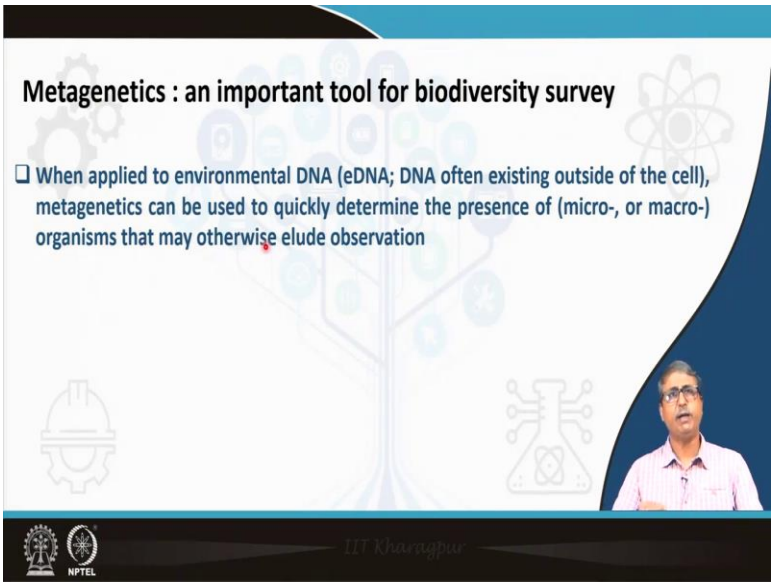
Or you can actually target functional genes also and this includes the contaminant removal that means any particular organism who is known to be involved in some kind of contaminant removal okay like the halococcoid is very important player in the hydrocarbon removal thing. So, you can actually target such organism or thyra or barcold area or pseudomonas type of organisms which are often found to be very effective in degradation of different hydrocarbon compounds.

Pathogen detection like vibrio and other pathogens you can you can detect in multiple samples and when you have large amount of data suppose is in a sample in a water sample you have a particular pattern is a deadly pathogen but the abundance of the pathogen is less but still you can detect it. So, that is called the sensitivity otherwise in normal method you would have said no this water is good for consumption but actually the water is not good because it has the pathogen maybe the pathogen concentration the load is low at this moment of time but you would never know it may increase.

So, pathogen detection particularly if the pathogen is at low level antimicrobial resistant genes. So, if you want to target gene specific NGS studies then you can actually amplify the antimicrobial resistance genes and for different other purposes the human health care purposes these are very important and finally for general environmental biotechnology processes monitoring a particular process like the disciplinary sulfate reduction ammonia oxidation methanogenesis and many other processes nitrate reduction denitrification nitrogen fixation and so, many processes which are very well known key enzymes are available.

So, you can monitor how the particular process is going on if you if you use such a primer in your amplification reaction.

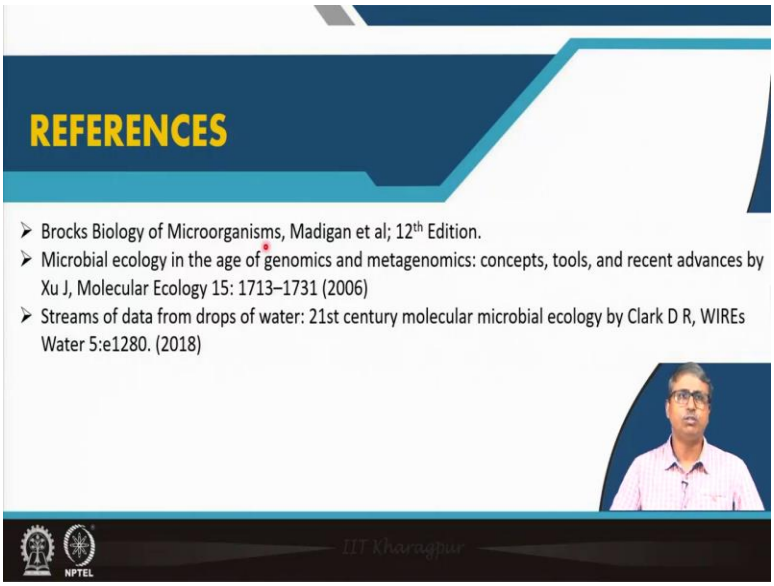
(Refer Slide Time: 46:57)



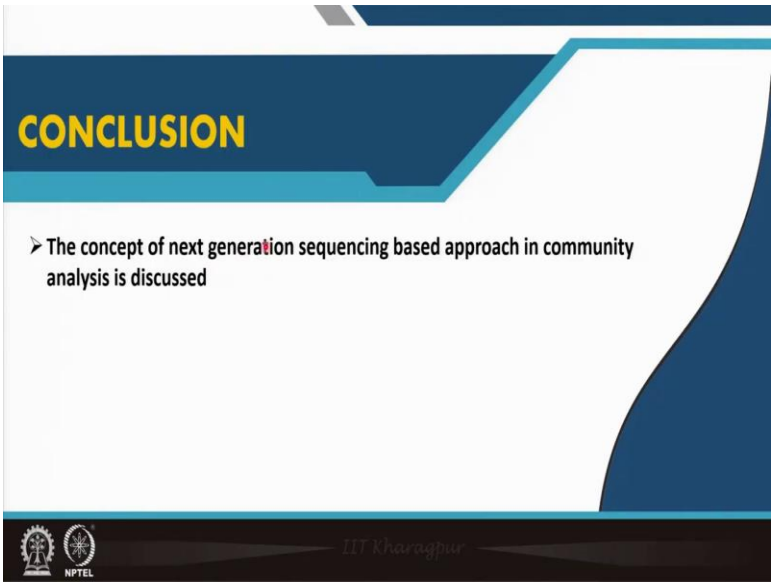
And it is also true that when applied to environmental DNA that is called eDNA. So, you totally start the total DNA. So, it can be targeted. So, that you can actually determine the presence of micro or macro organisms that may be otherwise eluded observation. So, you can use actually 16s and 18s primer both primers can be used separately you can actually. Then for the same sample if you use let us say if you use a bacteria specific archaea specific and 18s primer.

Then simultaneously you will be able to check that what are the diversity and the organisms present in terms of the archaea in terms of bacteria and in terms of different eukaryotic organisms.

(Refer Slide Time: 47:41)



(Refer Slide Time: 47:49)



So, for this lecture the following references will be will be useful. And in conclusion I will say that the concept of the next generation sequencing based approach in community analysis is discussed particularly we have we have tried to emphasize the need for the next generation sequencing and although we talked about the different aspect of the common cultivation independent process including the clone library based approach.

But clone library was just one of the methods in order to in order to address identification and the characterization of the all the different phylotypes or the other species who are present in the

environment. And what we have learned that the next generation sequencing based approach is one of the one of the best method in almost all sense. We will continue more discussion on this particular topic maybe in our subsequent lectures, thank you.