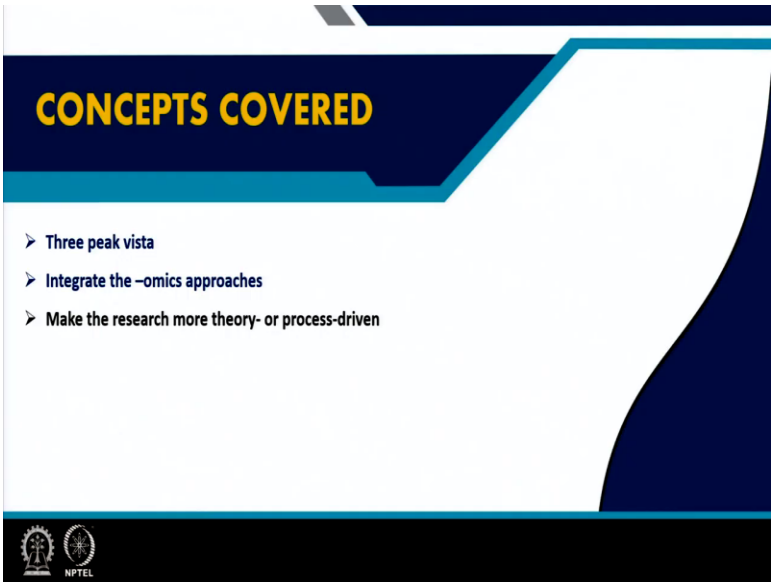
Environmental Biotechnology
Prof. Pinaki Sar
Department of Biotechnology
Indian Institute of Technology, Kharagpur

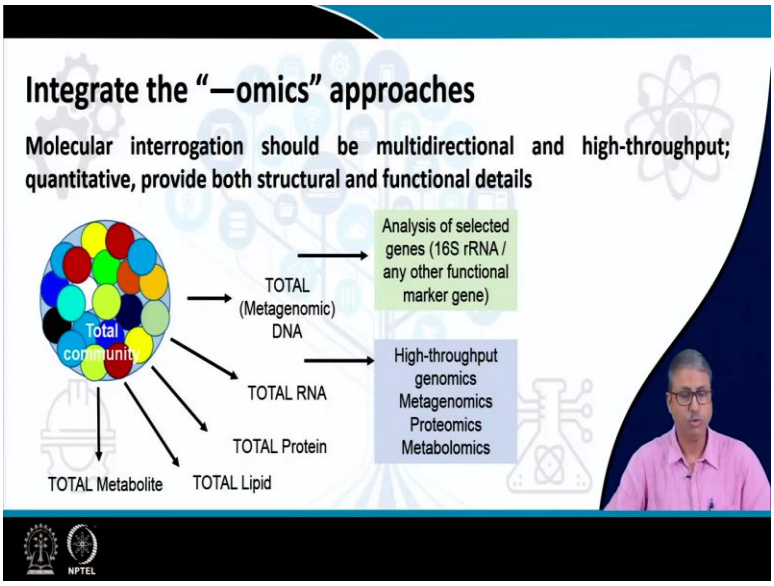
Lecture – 16
Microbial Ecology and Environmental Biotechnology - Part C (Contd.,)

(Refer Slide Time: 00:35)



Welcome to this lecture on microbial ecology and environmental biotechnology wherein the concept of this three peak vista or the major themes will be discussed and completed. We will continue and we will discuss more on the integration of the omics approaches in order to make the maximum application of microbial systems into environmental biotechnology. And finally we would discuss about how to make the research more theory or process driven rather than it rather than tool given.

(Refer Slide Time: 01:15)



So, with respect to the different omics approaches we have previously noticed that the total community present within any environmental system wherein we are expecting some services from the environment may be remediation may be waste treatment may be energy generation or resource recovery maybe sequestration of carbon or generating useful molecules useful chemicals.

So, we have found that there is a strong preference towards using an approach which targets mostly the DNA from the samples. So, we need to understand that of course the total DNA or the total metagenomic DNA and its analysis could be a very important approach very important method in order to understand the basic ability or basic characteristic of the community in terms of its structure and as well as in terms of its function.

But we need to also understand that our molecular interrogation, interrogation in the sense that we would like to interrogate the community microbial community which are used in a biotechnology setup to gain the maximum understanding maximum insights which can be useful for gaining the services out of the community. These should be multi-directional and high throughput they should be quantitative and very importantly they should provide both structural and functional details.

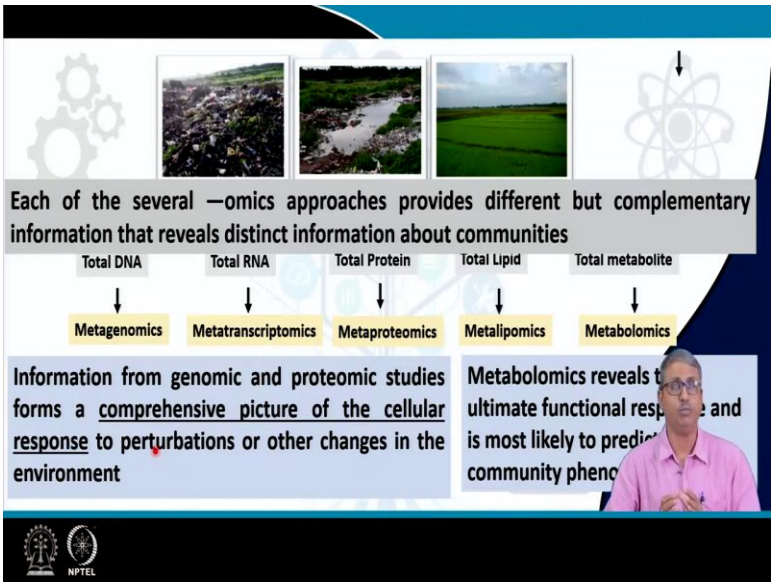
Now having an analysis done solely on metagenomic DNA may not be a very good idea. So,

other molecules which are present in the cells including the RNA particularly the mRNA which gives us the sense of the gene expression. The total proteins which are capable of providing us the enzymatic repeater present in a system the total repeat which provides us the ability of the membranes to respond to the different functions going on.

And lastly but not the least the total metabolites that means the all the small molecules which are produced inside the community which are actually the workhorse they are interacting they are the substrate they are the products of different reactions. So, we need to analyze and need to understand them as well and then integrate all the details in into a common platform to understand how the community is working with reference to a particular aspect of our interest.

So, here again the total community DNA community RNA protein lipid or metabolites can be analyzed with reference to any specific marker molecule like a marker gene or a marker metabolite or it could be subjected to a high throughput analysis as we know we do we do like genomics metagenomics or proteomics and metabolomics analysis. And we will learn about those methodologies and approaches in the due courses.

(Refer Slide Time: 04:49)



Now let us look into this situation that we have a different sets of the contaminated environments or environments where like a rice paddy field which where we want to understand how the paddy soil microbiome is actually working towards the crops the quality and the yield of the crops

which are being grown there. So, we can obtain the environmental samples as we often do from these sites or these fields following all kind of standard protocols.

And then we can extract the total DNA or the total metagenomic DNA. Now here the word total refers to the total it is not that we are going to isolate a bacteria or a fungus and then extract the DNA no. We are not going to do that we are going to take the sample, sample means the tots a subset of this soil or the water collected from the particular environment and we can have a replicate environment where we can we can collect as many as sample as possible and required.

Because depending upon the nature of the contaminated site a site like the landfill site might require more number of sampling because it is more heterogeneous. So we should have as many number of samples as as representative as it is possible. And then instead of isolating any specific organism we will isolate or we generally extract the total DNA which is called the metagenomic or the environmental DNA.

Similarly the total RNA is extracted total proteins can be extracted total lipids can be extracted and total metabolites can be extracted. Now as we extract the total DNA or the metagenome we subject it to analysis which is called metagenomics. So, this metagenomics that we are referring here is not based on any single gene it is a kind of a holistic approach where the entire pool of DNA that we collect is subjected to sequencing mostly and generally without any PCR reaction involved into that.

Because we are not in particularly focusing on any particular gene the gene focus or the target gene identification can be done on the later stage where we will when we will analyze the sequence data but initially we are not going to incorporate any bias into this. Bias in the sense the incorporation of a particular set of PCR primer might lead to a bias. Because as I mentioned earlier all PCR primers may not work for the genes they are they are made to. So, metagenomics means it is a PCR less reaction.

So, the total DNA is collected fragmented and sequenced and the sequence data that is the sequence of nucleotides are analyzed to identify the genes which are present. Now during this

analysis we can have our own set of genes that these genes we are interested to and some other genes we may not be interested to study. So, that is up to us but the entire metagenome should be sequenced. Similarly the total RNA can be sequenced that is called metatranscriptomics.

The total protein pool or the collection of all the proteins can be sequenced or analyzed this is called metaproteomics. Total lipids can be analyzed total metabolites can be analyzed that is called the metabolomics. Now each of these omics approaches like metagenomics metatranscriptomic, metaproteomics etcetera. They provide us different but complementary information how? For example if we look at the metagenomic data.

Now from the metagenomic data we can we can surely identify who what are the micro organism present and what are the major genes responsible for any particular compound degradation or metabolism. So, we can identify those genes let us say we are interested in carbon assimilation carbon fixation in a particular system. So, we can easily identify the all the genes which are involved in multiple carbon fixation pathways.

And then predict a model that based on this metagenomics data we provide a kind of an hypothesis that this system is capable of doing carbon fixation by this mechanism although the genes for others are also present and maybe multiple carbon fixation pathways are working. Then how metatranscriptome or metatrader, scriptomic data is going to help you or help us metatranscriptomic data will provide us the real time data about out of the different carbon fixation genes what are the genes or what pathway is actually predominated at any particular time, time in the sense when we have sampled the RNA.

So, you may have the all six or seven pathways of carbon fixation operating as far as the total metagenomic information is concerned. You may find out that all the genes responsible for different carbon fixation pathways are present there. But it is not it may not be the case that all the carbon fixation pathways are operating within the system. So, you may find out that you have the the CBBL pathway or the Wood Lung Gel pathway or the three hydroxy propionate carbon fixation cycle reverse TCA cycle genes for all are present.

But when we look into the metatranscriptomic data you may be surprised to see only the wood lung gel pathway is operating the other like rubisco based or the three hydroxy propionate cycle or the reverse TCA cycle are not actually operating because the mRNA's for those pathways are not present. Next is the metaproteomics which will offer you actually the functional protein which will carry out the enzymatic process.

So, if we believe that this particular pathway would long gel is operating then the wood look gel enzymes must be present in the total protein data. So, metaproteum will actually validate us. So, that is what is very important because they may be very providing us very discreet set of information the metagenome will provide the total genomic DNA based pool the transcriptome will provide us the the expression level data the metaproteome will provide us the total protein pools data but these are complementary.

And when we try to connect these data connect this metaproteum data to the metatranscriptome data metatranscriptome data is fitted into the metagenome data we are able to find out the distinct characteristic and their function of the community. Now if we look very carefully that the information that we gain from the genomic and proteomic studies including the metatranscriptomic studies they forms a comprehensive picture of the cellular response to the perturbations or other changes in the environment.

So, for example we have a landfill site here another contaminated site over here. So, if we have couple of sides like this some of them might be more contaminated some of the sites may be less contaminated. How do we know that when we analyze the contaminant concentrations using some sensitive method we will be able to find out that some of the sites are highly contaminated. Now so, that contamination is a kind of a perturbation to the community.

Now when we have this metagenomic, transcriptomic or metapotemix data across these different sites which are actually representing a gradient of contamination we are going to have a comprehensive picture about the cellular response to the perturbations. These response could be seen as a kind of a change in microbial community composition number of species present the abundance abundant species or less abundant species or more abundant species or it could be the

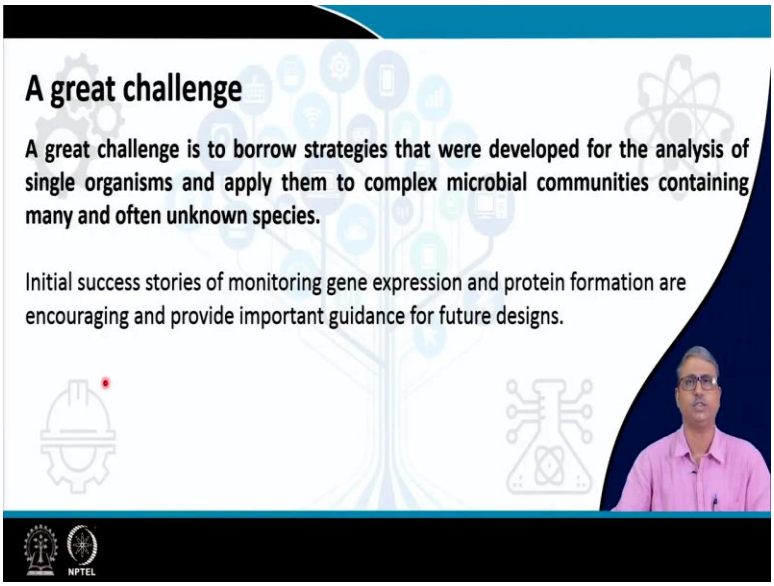
set of genes they harbor.

The set of genes that those are expressed under different conditions and the set of enzymes which are produced under different conditions because each of the sites as I mentioned are having different levels of contamination this is an example. On the other hand the metabolomic studies or the sequence of the or the or the lipids or the all the metabolites will reveal the ultimate functional response.

And it is most likely to predict the community phenotype. So one hand you have a cellular response which is a internal affair of the cell or the cells in a community. If we have one lakh species how one lakh species is behaving or how one lakh species together they are behaving itself in a cellular context? What is happening inside their cells? What are the genes which are being expressed? What are the enzymes which are being produced by these cells?

This is one part but how they are affecting the function that will come only when we do the metabolomics and when we will connect this metabolic metabolomic data with the metagenomic and metatranscriptomic and metaproteomic data that will give us the nearly complete picture about how this community or communities are functioning?

(Refer Slide Time: 14:34)



Now we will talk briefly about the challenge. Now one of the one of the challenges which is

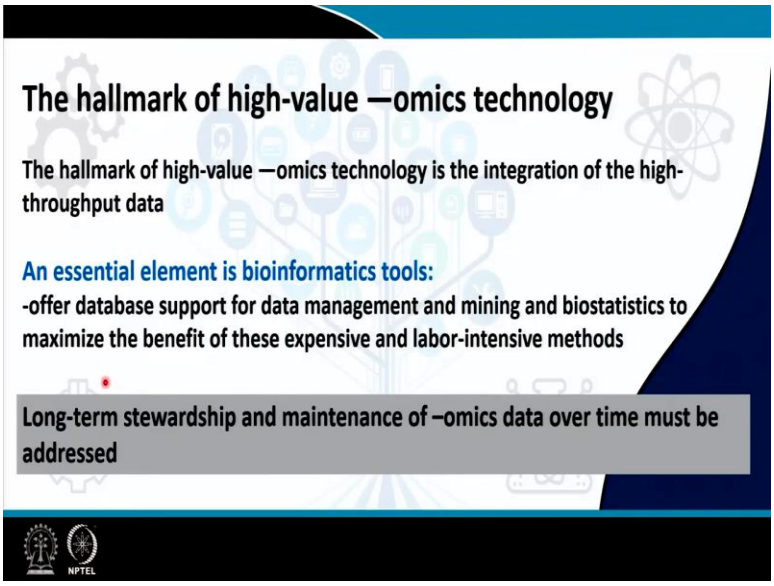
called a great challenge is to borrow strategies that were developed for the analysis of single organisms and apply them to complex microbial communities containing many and often unknown species. Now what is this actually? This is a very, very important and interesting fact that for microbial ecology microbial metabolism we have relied on individual species ability for a long period of time and there are valid reasons for such things.

So, previously we talked about the halococcis, the dialycoides are involved in tetrachloroatrichloroethylene degradation that means if we have the halococcodies we will instantly rely on the or infer that the community has some kind of potential towards tetrachloroethane degradation. On the contrary if we don't have the halococci this type of organism we may infer that the community might not have the ability to degrade the tetrachloroethane.

But sometimes that may be true or come true but sometimes it may not be. So, we are often relying on because so, many cultures. So, many pure cultures are already available and they are well characterized and many of the microbial metabolisms are very well understood using those type cultures or those are isolated bacterial strains of bacterial organisms. Now when we think of a complex microbial community as I mentioned where we have tens of thousands of species and if you look at their cultivability will find only a handful of them are cultivable.

And in any kind of database you will have only few of them are actually previously isolated from some other environment or a similar environment and have been characterized with respect to a metabolic process. Now whether the same metabolism will be carried out by the same species in your environment or not that is a very important question and that is the challenge. Now initial success stories of monitoring gene expression and protein formation are encouraging and provide important guidance for future designs.

(Refer Slide Time: 17:01)



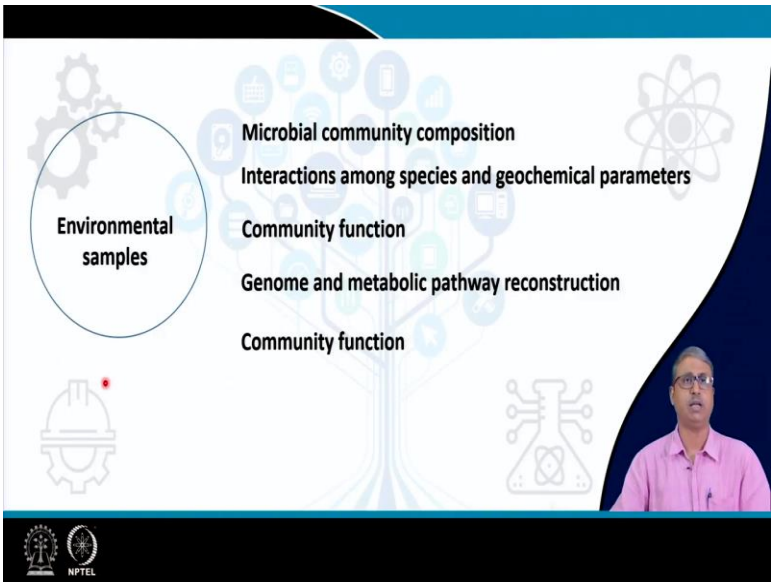
Now the hallmark of high value omics technology is the integration of this high throughput data. So, we are now going to emphasize on that that we need to actually integrate the data that we get metagenomic data to transcriptome, transcriptome to metaproteum and finally the lipid and the metabolic profiling data and that there is a very important all essential role of the bioinformatic tools.

Now bioinformatic tools will provide you the database support because each and when you we are we are generating the sequence data or the metabolic profiling data or metaproteome data either we are publishing it or we are we are submitting into different database. So, our databases are getting enriched. So, bioinformatic tools are being developed or need to be developed further in order to have that database support for data management and data mining.

The moment I get a sequence or an information about a particular species I should be able to find out maximum information about that particular species or that particular gene sequence that what are the other environments from where this particular sequence have been previously sequenced or previously obtained and then apply different biostatistics or bio statistical tools to maximize the benefit of those expensive and labour intensive methods.

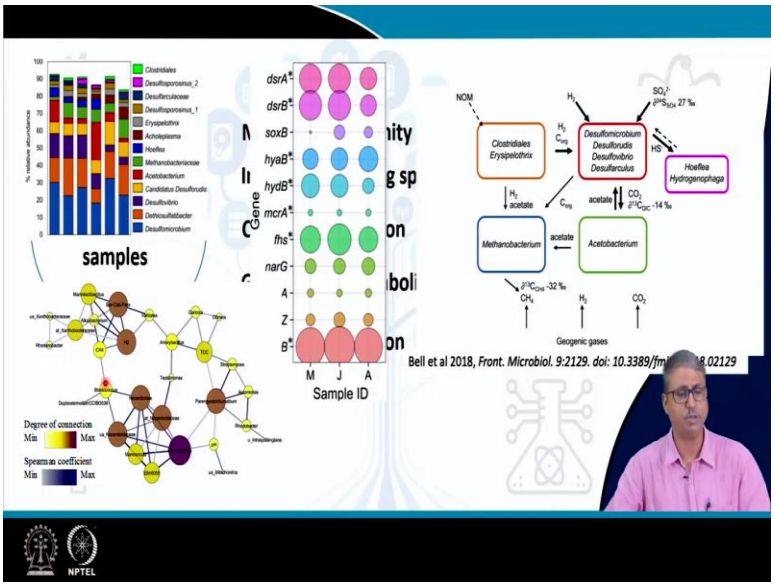
And in this regard the long term stewardship and maintenance of omics data over time must be addressed. So, the maintenance of the data and their curation is found to be very important.

(Refer Slide Time: 18:40)



So, ideally we have environmental biotechnology system where some environmental samples are to be to be implemented or characterized. So, we have the community composition goal we have the interaction of species between the species and between the geochemical parameters we have the community function through metagenome. We have the metabolic and genomic reconstruction processes and eventually we have the community function.

(Refer Slide Time: 19:06)



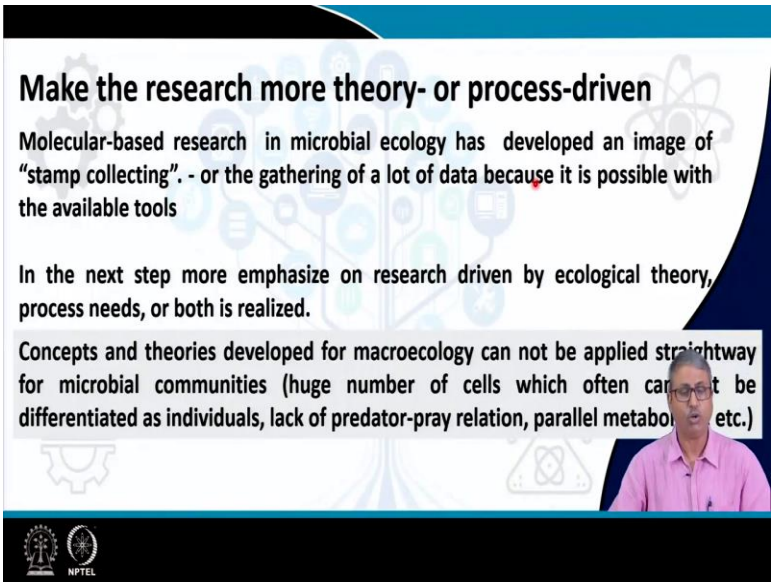
So, in order to give some examples that how it looks. So, from the environment we can analyze the composition of the taxonomic composition we can have the metagenomic data pooled with respect to specific processes like the metabolism of sulphate or metabolism of sulphur or

sulphide or hydrogen metabolism or methane metabolism or nitrate metabolism for example. So, from the entire metagenome of numerous genes enzymes encoding those we can we can fetch out all the required other genes which we feel that we are interested into.

So, and then we can look into this and then we can establish the networks between the different type of parameters like with hydrogen and methane or organic carbon what are the organisms which are interconnected more with respect to hydrogen or methane and similarly with respect to any other kind of parameters or within the microorganisms or the species themselves. And finally we can have a kind of a metabolic pathway.

Wherein we can try to decipher how methane hydrogen carbon dioxide for example in this case we are we are trying to elucidate or these authors like Bell et al they have they have studied that how the metabolic processes are actually controlled using a in set of metagenomics, metatranscriptomic and metaproteomic study or even metametabolic study. We can actually infer and develop such models and then validate those models that how actually the community is functioning.

(Refer Slide Time: 20:40)



The next most important point is make the research more theory or process driven. Now molecular based research in microbial ecology has developed an image of what Rittman said kindly correctly that is a kind of a stamp collection or gathering a lot of data because it is

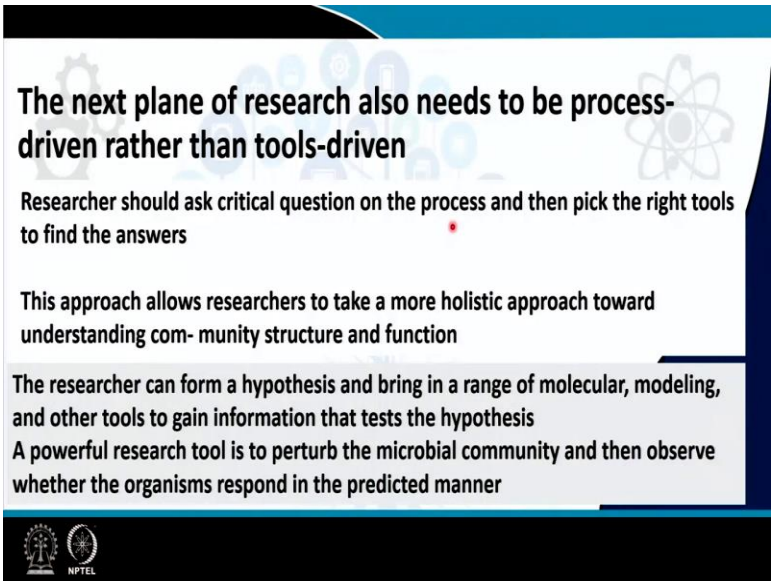
possible with the available tools. So, as soon as you go for a metagenomic analysis or any kind of other high throughput analysis you get a large number of data.

So, are we just collecting those data like a community composition or something like that perhaps not. So, what we need to do in the next step more emphasis on research driven by ecological theory, process needs or both realized. Concepts and theories which are developed for macro ecology very well placed with respect to macro ecological system these theories are very well placed.

But unfortunately most of those theories which are very well suited for explaining the macro ecological processes predicting the functions of the macroecological processes cannot be applied straight away for microbial community or microbial ecological systems some of them could be but not all. One of the reasons or some of the reasons are the huge number of cells which are often cannot be differentiated as individuals which is done in case of a macroecological landscape.

Lack of predator prey relationship or the presence of a parallel metabolism in case of micro microbial system nonetheless the proper application of ecological concepts both diversity, stability, competition, redundancy and allied tools will be essential to provide us a kind of intellectual framework for designing research and understanding in the meaning of the of such results.

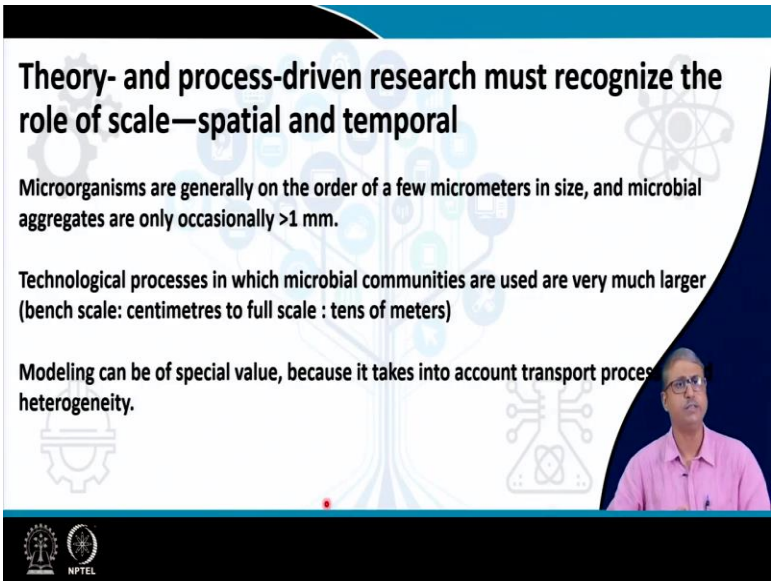
(Refer Slide Time: 22:35)



Now the next plane of research also needs to be process driven rather than tool driven. So, we are not developing tools only. So, it must be hypothesis driven approach the question leads the selection of the tool. So, it depends on what are the questions the researchers are or the engineers are asking depending on that the tools must be selected or developed. So, researchers should ask or the engineers environmental engineers must ask critical question on how the process and then how about the process how and why this particular thing is happening.

And then pick up the right tools to find the answers. This approach allows the researcher to take a more holistic approach towards understanding the community structure and function. The researcher can form a hypothesis and bring in a range of molecular modeling and other tools to gain information that test the hypothesis. A powerful research tool is to parter the microbial community and then observe whether the organisms respond in the predicted manner.

(Refer Slide Time: 23:40)

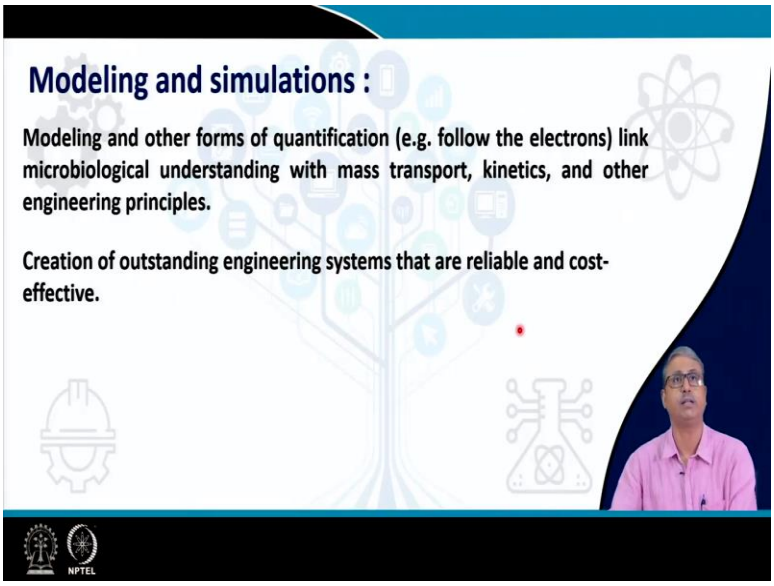


And eventually this theory and process driven research must recognize the role of scale both special and temporal. Because the microorganisms are very small even when they are aggregates they are slightly more than one millimeter. Whereas the technological processes where these microorganisms are actually applied they are large centimeters in the lab scale to tens of meters in the in the application scales.

Similarly on a temporal scale also microorganisms in the lab they perform or they they are generally grown for a few hours or a day. Some experiments with respect to utilization of substrate or microcosm or mexico studies are often done for weeks months or max or sometimes years but rarely for years it is months or several months or weeks. But in reality most of the environmental processes are to be operating continuously round the clock with variations in the input concentration of the different substrates the temperature fluctuations, pH fluctuations and many other variations.

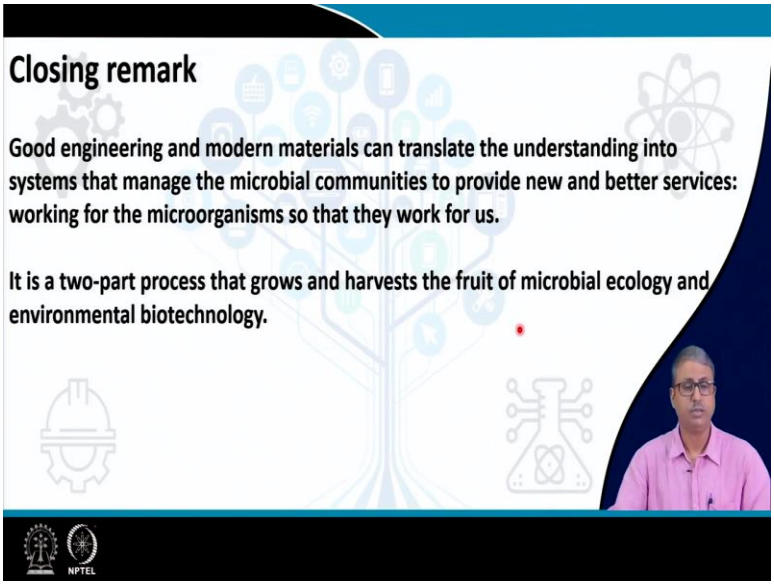
So, modeling is identified to be a very important tool in this case because it takes into account the transport process heterogeneity and all these parameters.

(Refer Slide Time: 25:11)



And finally we need to understand the importance of integration. It is essential for creating successful processes in environmental biotechnology to integrate the knowledge of microbial community with several other factors like the modern materials or the mathematical processes and simulation. Creation of outstanding engineering system that are reliable and cost effective would necessarily require the successful modeling and other forms of quantification.

(Refer Slide Time: 25:47)



Now the cloth as a closing remark I would like to emphasize that good engineering and modern materials including the modeling techniques, powerful molecular tools everything can translate the understanding into systems that manage the microbial communities to provide new and better services. So, that we reach to a situation where working for the microorganisms are favourably

achieved.

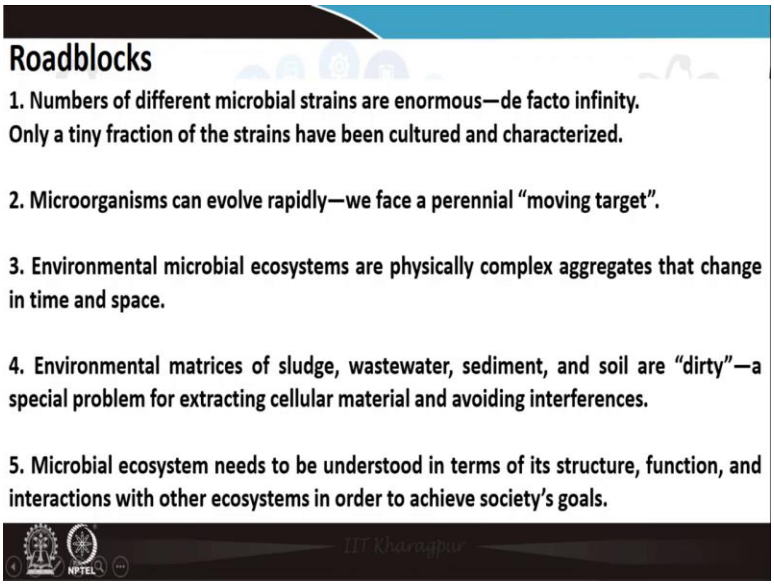
So, that they can work for us it is a two-part process that grows and harvests the fruit of microbial ecology and environmental biotechnology.

(Refer Slide Time: 26:28)

However it also suffers with some roadblocks despite the impressive achievements and developments that happened in the last 30 years or. So, there are that that are beginning to emerge and we see lot of publications and lot of other inventions and discoveries and we appreciate we start appreciating the high potential for those partnership between the microbial ecology and their translation into successful translation into application domains.

Road blocks are inevitable because of the inherent complexities of environmental micro biology system.

(Refer Slide Time: 27:09)



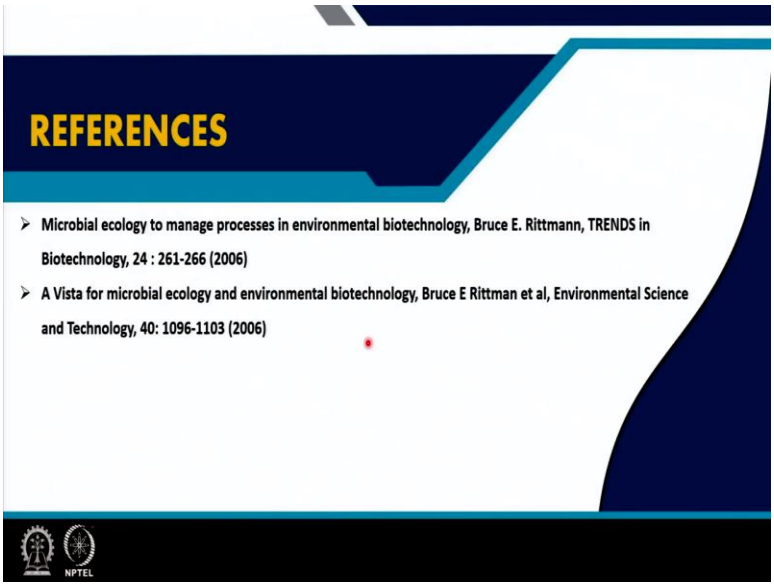
Now with respect to the road blocks the most important one is that the number of different microbial strains are enormous as I mentioned throughout my lecture it is tens of thousands to lakhs in a given soil one gram of soil or. So, it is considered to be de facto infinity. Only a tiny fraction of the strains have been cultured and characterized. So, we have a very limited information it is only in the lately like last 10 years or so, after the metagenome assembled genomes are in place.

So, we are able to reconstruct the genomes of the uncultivable organisms we are able to identify the abilities of some of the species which remain absolutely uncultured and uncharacterized so long. Microorganisms can evolve rapidly. So, we face a perennial moving target. So, it is all the time changing. The environmental microbial ecosystems are physically complex aggregates that change in time and space.

So, this they are self assembled structure. So, this nature of self assembly is subjected to different interactions and different perturbations and other factors. Environmental matrices of sludge waste water, sediments loyal or dirty. A special problem for extracting cellular material and avoiding interference it is easy to say that we will extract the protein or extract the total mRNA or extract the total metabolites that when we extract the total metabolites from its dirty sample like a wastewater or a soil full with different kind of many type of waste.

So, those derivatives or those contaminants are going to be extracted. Now getting the actual data about the metabolites produced by microbes and the contaminants present in the environment is a very tough problem. Microbial ecosystems need to be understood in terms of its structure and function and interactions with other ecosystems in order to achieve the society's goal.

(Refer Slide Time: 29:08)



These are the references those are used in this particular lecture. And in conclusion the importance of the Multimix approaches are elaborated. We also highlight the fact that making the research more theory or process driven is one of the priority areas. Roadblocks and possible direction of current and future R and D are highlighted, thank you.