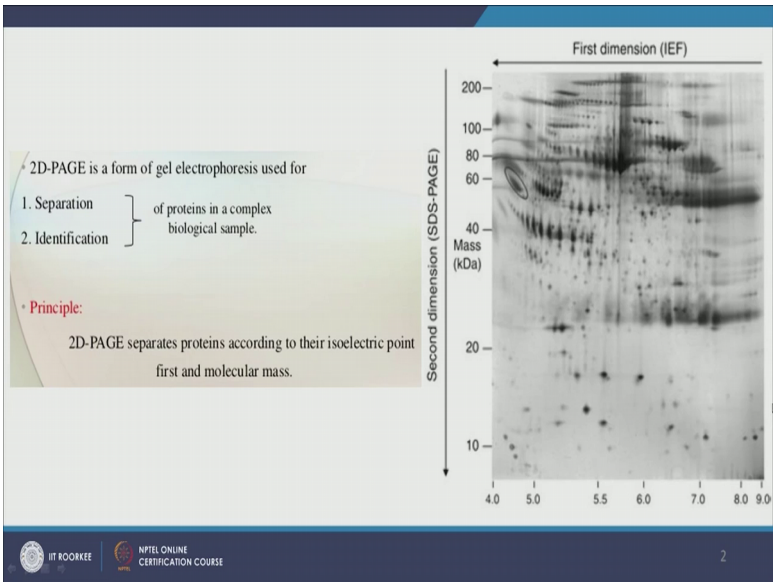
Applied Environmental Microbiology
Dr. Gargi Singh
Department of Civil Engineering
Indian Institute of Technology, Roorkee

Lecture – 23
Environmental Genomics III

Dear students, in this lecture for environmental genomics, we will look at how when microbial communities or microbes undergo succession, that is, when they replicate from parent cell to daughter cell and at each stage they undergo certain differences in their genetic material. It leads to evolutionarily distinct microbes or microbes with distinct characteristics. So, this is we can refer to it as evolutionary aspect of environmental genomics.

So, to begin with as microbes evolve and as they have developed they develop communities of multiple populations, I want to start with expressing that they have very different protein signatures and this is something I wanted to cover earlier, but I want to include it with this lecture.

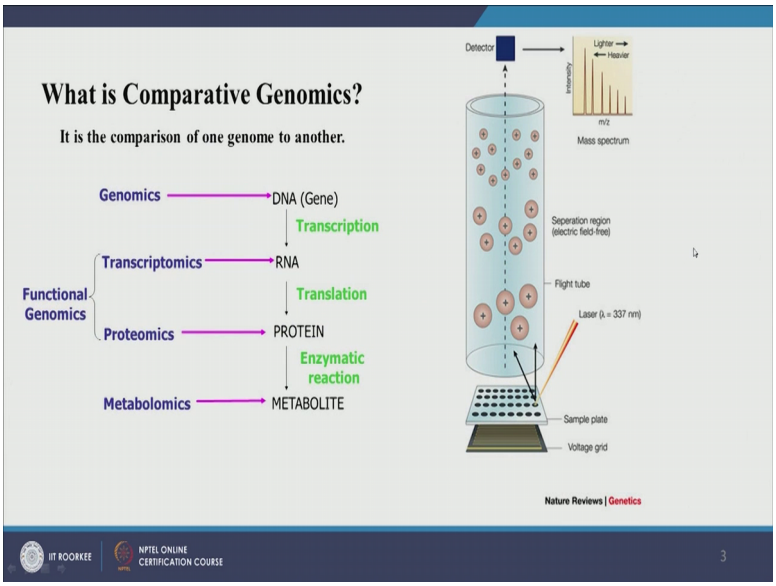
(Refer Slide Time: 01:12)



So, on the right panel you have a plate that is showing you that has differentiated different proteins on basis of their isoelectric point and their mass. So, this is a 2D-page, 2 dimensional page, it is a form of gel electrophoresis; first proteins are resolved on x axis in the first dimension on basis of their isoelectric points which are given here. So, I said electric points

are the pH at which they have neutral charge and then they are resolved around second dimension according to their mass in kilo Daltons. So, the smallest proteins will be here, the large proteins would be here. So, this way we have isolated proteins and then we can compare and we can see how proteins are different for different samples.

(Refer Slide Time: 01:52)

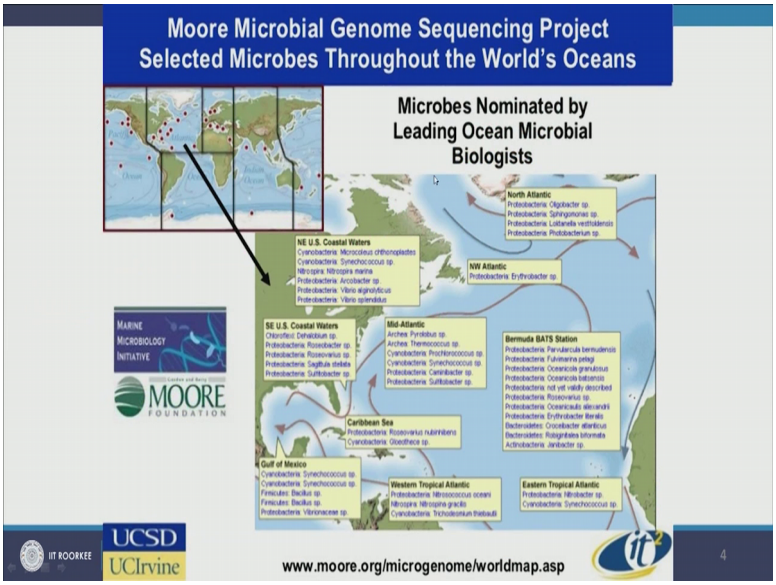


Then, question is what comparative genomics is. So, I started this lecture by talking about how the microbial communities succeed or when microbial populations replicate they undergo differences. So, one of, because this is the fact that they undergo differences when we generate lot of sequences we do a comparative genomics analysis which is we see in comparison to other genomes that we have in our database how is a different, how is it similar.

So, the similarities would give us an idea about similar functions, the dissimilarities will give us idea about dissimilar function. For example, let us say E. coli; one strain of E. coli is totally harmless, it is (Refer Time: 02:31) commensal with it. It is found in our bodies and it does not have let us say some genes that another E. coli has which is highly pathogenic, highly virulent. So, we can say that the difference between the virulent E. coli and the commensal E. coli, the genetic difference is the reason why they have different virulence, different pathogenicity. So, this is comparative genomics and I can do it at all levels. I can do it on the genomic level, DNA level, RNA protein level or metabolite level.

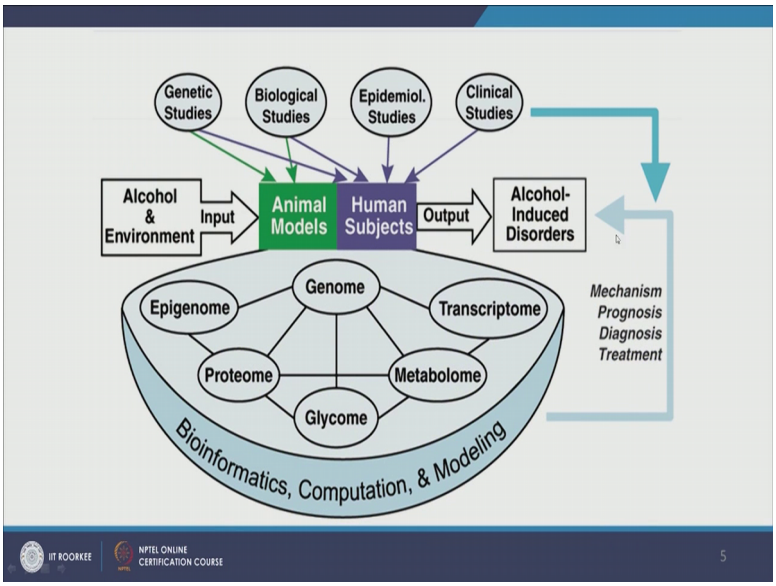
Now, here is a picture of metabolomics, metabolite analyses where what I do is, I have my sample I put a charge a voltage and I allow my metabolites all the acetate, leukocytes etcetera to go through this column and this column separates them on basis of mass by charge ratio. I am applying charge which is forcing it to run away and then the weight and the speed with which they run away is dependent on both mass and charge. So, based on the mass and charge ratio they will reach at detector the different times and then I can separate them on m by z ratio.

(Refer Slide Time: 03:38)



So, what we have done now is that over across the globe we have sampled our environmental microbial communities, we have done metagenomics, we have done proteomics whenever we could and we are moving towards metabolomics transcriptomics, interactomics for select environments of interest. So, this picture is moving microbial genome sequencing project and they have sequenced the selected microbes throughout the world's oceans. So, they have Bermuda BATS station, Northwest Atlantic, North Atlantic, Mid-Atlantic, Caribbean Sea, Southeast U.S. coastal water, Northeast U.S. coastal water, Gulf of Mexico where there was an oil spill in 2010, Western Tropical Atlantic, and Eastern Tropical Atlantic. So, this entire region is being characterized by the Moore foundation.

(Refer Slide Time: 04:28)



So, why would we care? Why would we want to? Last lecture I talked about earth microbiome, I talked about how we are sequences or have been completely sequenced and now, we have the Moore project, Moore foundation which is characterizing the microbes in the ocean, in Atlantic region and nearby seas why would you want to do that? because having all this information about the genetics, about the proteomics, about the metabolomics gives us an idea about what is really happening and it helps us solve our public health environmental health challenges. For example, if we know the sequences of microbes in different parts of North Atlantic Ocean, if there is an oil spill we know how we can then characterize it again and see how things have changed.

We shall give us an idea about what microbes are eating? Are they eating the petroleum? The oil? Are they eating something else? Are they dying? Are they not dying? What genes are getting up regulated? What are getting down regulated? all this information can be gathered which will help us device interventions. Now, let us say, I devise an intervention and I decide to put a surfactant which will disperse the oil in the marine area here, it will disperse oil in the North Atlantic Ocean and the we expect the microbes will eat it, but we do not know if microbes are eating it or not or just getting dispersed and diffused or it is getting stuck to certain surfaces and it is not being detected easily by us. So, in that case we can again characterize microbial community and we can see if they have up regulated down regulated

different genes and by that we can get and by looking at the metabolites and the proteomics we can get an idea of how microbes are responding to our interventions.

Now, one example here is alcoholism. So, public health challenge, how we can use genetic study, biological study, epidemiological study, and clinical study all together to get an idea about alcohol induced disorders. So, this is how it looks like, we have alcohol and we have the environment. The recent studies and alcoholism talk about alcoholism is not just a very individual decision or individual affliction, but it is a result of the environment, what are the microbial communities inside the human being, outside the human being, what are the behavior pattern, social patterns, economic patterns and lifestyle patterns around the human being and the history of the human being lot of things are happening here.

So, we need to do all these kind of studies. So, we do epidemiological biological genetic study and clinical studies on the human subject. We can even do periodical and genetic study on the animal subjects. They are introduced to alcohol in environment. Some of them have alcohol induced disorders. Now, we can do genomic study in which we can see how alcohol induced disorders if they have a genetic backup or not, if they have a genetic signature or not. They can do epigenetic genomic studies, proteomics studies, glycomics, metabolic transcriptomic studies to understand how alcohol is processed in the body, how it affects different parts of the body and how as a result we have different alcohol induced disorders. So, we get a very much clearer picture about the mechanism of alcohol induced disorders and thus we will be able to understand.

For example; in same environment under same conditions some human beings are more likely to undergo alcoholism to be affected by it and some are not and even those among even among those who have alcoholism. So, only few are prone for alcohol induced disorders and some are not. So, the questions that scientists ask now are their genes that are responsible for alcohol induced disorders or which make people more susceptible for alcoholism and alcohol induced disorders. So, in that case we want to find out the genetic signature. So, for example, if I get my genes type and I find out that I had the genes responsible for or which make me more susceptible for alcohol induced disorders and alcoholism then I know better how to take the right steps so that I do not fall prey to them.

So, doing my genetic characterization will help me choose a better lifestyle and know where my limits are. The other place where this is very important is, for example, cancer studies.

Earlier it would be like this if my mother had the cancer, my sister, my brother and blah blah had this cancer. So, I am more likely to get the cancer because cancer is considered to be hereditary disease, but we do know scientifically is not in popular knowledge still a lot that it is cancer is not just hereditary disease I might have this susceptibility for getting a particular kind of cancer because let us say it runs in my family, but unless the environmental factors are right, unless my up regulation and down regulation of the genes are right, it is not necessary that I will develop the cancer.

In that case I have definitely want to do my genomics analysis, metagenomic analysis to understand, what genes are present, how prone I am for my genetic point of view to particular disease, to particular cancer, how are the other microbes in my body. For example the bacteria, the virus, the fungi are behaving, what are present and how they are affecting my genes, for example, there is a human papillomavirus which is transmitted via cats and other animals and it can result into cancer. It can cause cancer in humans mostly females, well, anybody. So, doing my metagenomic is very important to get an idea and then what I can do is, I can do my proteomics or transcriptomics which will give me an idea of which of the genes are being up regulated which are being down regulated. For example, I want to know if I eat chocolate am I protecting myself against the cancer that I am prone to or I am rushing towards the cancer.

So, in that case I cannot eat chocolate for a few days see how our genes are being up regulated, down regulated and then I can over eat chocolate and I can see how things are changing and then I will get an idea whether I should consume chocolate or not. The other place where this is very helpful is obesity. We all know from our personal experience, some of us are prone to gaining weight; some of us are not prone to gaining weight. So, in order to understand that will it be laughing more, will laughing more help me lose weight, will doing more exercise help me lose weight? Will eating greens help more lose more weight, with eating high-protein diet, high fat diet help me lose weight or no sugar diet help me lose weight. So, in order to understand these lifestyle changes their effect on public health this genomics epigenomics, proteomics, glycomics, metabolomics, transcriptomics together they provide us a very good picture, together they help us understand what will up regulate our bad genes what will down regulate our bad or good genes and this gives us insight into how to be healthier.

Now, coming to environmental systems, let us I am talking about waste water treatment plant and now the influent for waste water treatment plant especially in the ASP part of the plant are changing. I want to know, how the microbial communities are changing? I want to know that when the sludge is floating which microbe is growing, which microbe is not growing. So, I can do simple typing to find that out, but if I can do metagenomics, I will know how the whole microbial community will allow this particular microbe to grow so that the sludge will not settle it will float.

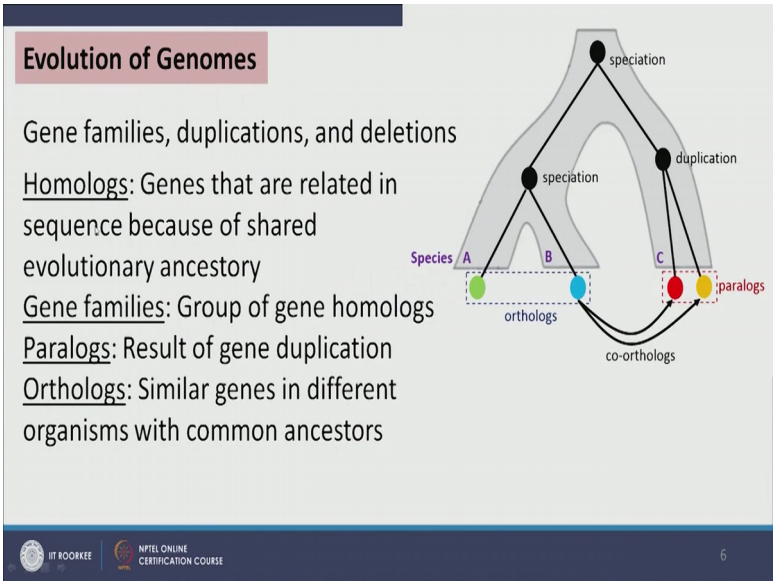
So, this, an inside into the entire microbial community will help me understand what are the factors. What kind of food was not present? What kind of food was present? What temperature was wrong, what was wrong that allowed this microbe to grow? Yes, I can add some chemicals inhibit its growth and come back to good well behaved sludge that settles on time, but it is even better if I can flip the problem in the part. So, for that metagenomics, metaproteomics this kind of analysis are very helpful.

Now, let us compare evolution of genomes the evolutionary environmental genomics. So, as microbes replicate the genes often undergo certain mutation, certain changes and over a period of time some genes might be entirely cut off. So, this is deletion. So, we result we get a smaller size of genetic material now. So, earlier, if let us say 200 base pairs get deleted; now the genetic material is less by 200 base pair. We can have inversion. So, if we had a sequence of genes such as a, b, c if this inversion it will go c, b, a which might change a lot of things for microbe because they are usually the transcription of DNA the translation of MRNA is very detectional. So, if they are in the same operon that would change how in what sequence the proteins are translated, expressed.

Then, the other thing is that sometimes what microbes do is they have the ability to duplicate the gene. So, when they replicate they know do not make only one copy of their genome or particular genes they made 2 copies. So, of the 2 copies, the duplicate copies one copy can serve the essential function, you know it can serve the cell and encode the protein that are required the other is a spare. So, in this the spare one can afford to undergo faster changes, faster evolutionary changes. It might be completely useless because the changes are harmful or useless to the cell or might be very useful. For example; it might evolve very fast into and in efflux pump coding protein that will allow the xenobiotics the heavy metals to leave the cell as soon as possible without harming it.

So, the spare copy is the reason why we have very fast rate of evolution even in healthy communities because one gene copy is fulfilling it is purpose. So, cell is in most cases not suffering a loss the other is evolving and trying to figure out experimenting. So, think of it this way, we had twins in a family; one of them will serve the family, get the money for the family, feed everybody and the other will experiment, do science and maybe come up with a great technology, maybe come up with nothing very special just doing incremental research here and there maybe will cause an explosion and the whole house will die. So, either of them is possible, but gene duplication is a very important mechanism.

(Refer Slide Time: 14:31)



So, in genes we let us look at some of the terminologies. Homologs are genes that are related in sequence. So, they have sequence similarity, because they have the shared evolutionary ancestry. For example, these green, blue, red and orange genes they are from their homologs because they have some similarities with each other, because they had the last common ancestor.

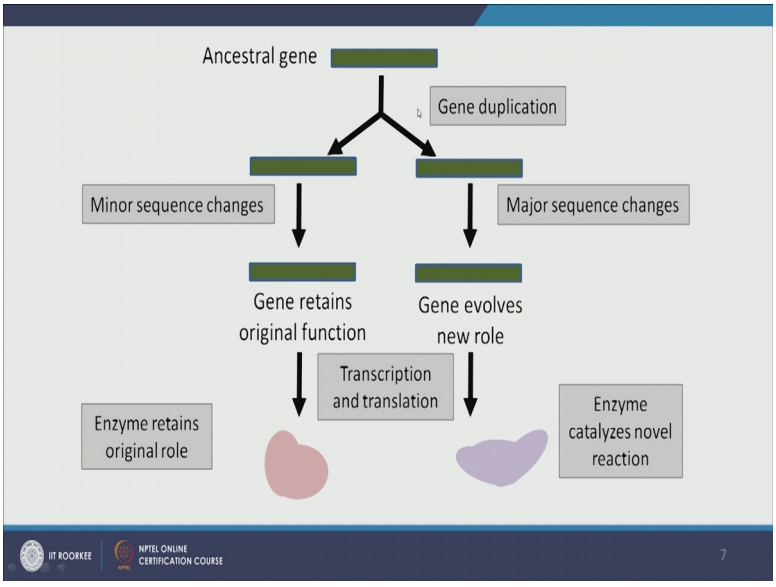
So, if you remember in one of the first lectures I talked about LUCA, L U C A Last Universal Common Ancestor because all life forms that we know of now, whether it is bacteria, archaea or eukaryotes all of them have certain similarities with each other and that is the reason we expect that because their similarities we all are homologs and we all share one Last Universal Common Ancestor from which the life split into different types. Now gene families are group

of gene homolog. So, these all colorful genes on the right panel, the green, the blue, the red and orange are gene families.

And, then paralogs; paralogs are a result of gene duplication. So, here this last common ancestor form 2 gene copies; one of them is serving the purpose the red one, the orange one is undergoing experimentation. So, they will have some difference with each other because one is undoing the differences the other is staying same as much as it can. So, these are paralogs. They usually found in the same stream. So, equalite K12 may have paralogs.

Orthologs are similar genes in different organisms because they have a shared ancestor. So, initially they had same ancestor, but then they underwent speciation. So, speciation is when they split into 2 different species and then one of them had gene duplication, so, made paralogs and then they may be the underwent another speciation and so on and so forth and they made orthologs. So, now, it is possible that clostridium thermocellum has the same gene as one of the equaliters, but they are orthodox.

(Refer Slide Time: 16:27)

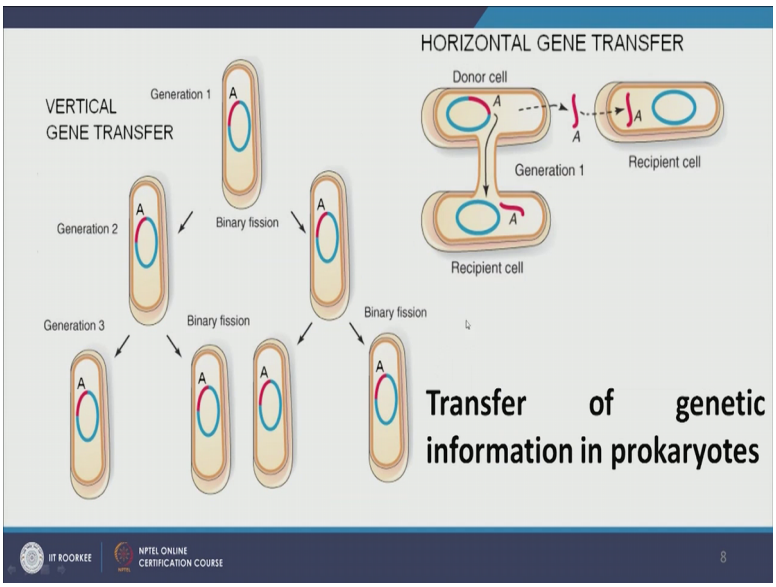


Now, let us look at this whole in enzyme changing business at a closer way. So, this is an ancestral gene alright. So, this is a gene that was present in our let us say common ancestor, it underwent gene duplication. So, it is very important to understand the difference between duplication and replication. So, when we replicate the cell, let us say it is undergoing binary fission, equal binary fission. They, in turn make 2 copies of the genome before it undergoes

binary fission and each daughter cell will get one genome. However, at times the cell undergoes gene duplication or genome duplication. Some of the microbes can actually undergo the genome duplication. So, think of it this way; some bacteria of one kind in a population have 6 mega base pair and some others have around 12 mega base pair. So, that is a case of gene duplication.

So, when they undergo gene duplication, the duplicate copy will undergo major sequence changes whereas, one of the copies will undergo minor sequence changes. So, the minor sequence changes will allow the gene to retain its original function, the cell should survive. So, this is the original function of the gene or inner shape of the gene. The enzyme retains original rule after transcription and translation. So, this is enzyme, this is not the gene. The duplicate gene will undergo major sequence changes; it will evolve to a new role. The gene will evolve after transcription translation, it will catalyze novel reaction. So, what would earlier catalyze only glucose may also catalyze lactose degradation now.

(Refer Slide Time: 17:58)

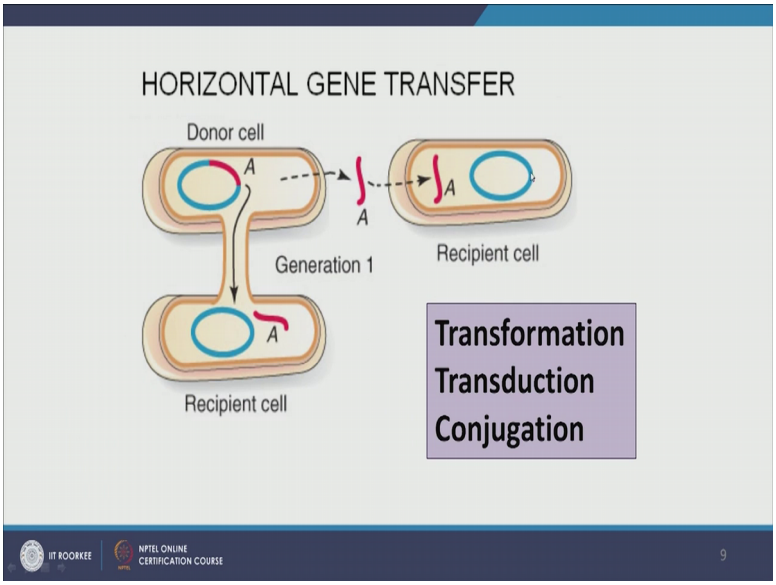


But, it is not always orthology or paralogy that results into differences in gene or similarities across different species, but it is also horizontal gene transfer. So, let us look at how is vertical and how is horizontal gene transfer. This is how prokaryotes share the genetic information. When the genetic information is transferred from parent to daughter and daughter to its daughter by a binary fission this is vertical gene transfer. So, from one parent

to 2 daughters 2 daughters to 4 daughters so on and so forth, all of them are getting similar genetic information, this is vertical gene transfer.

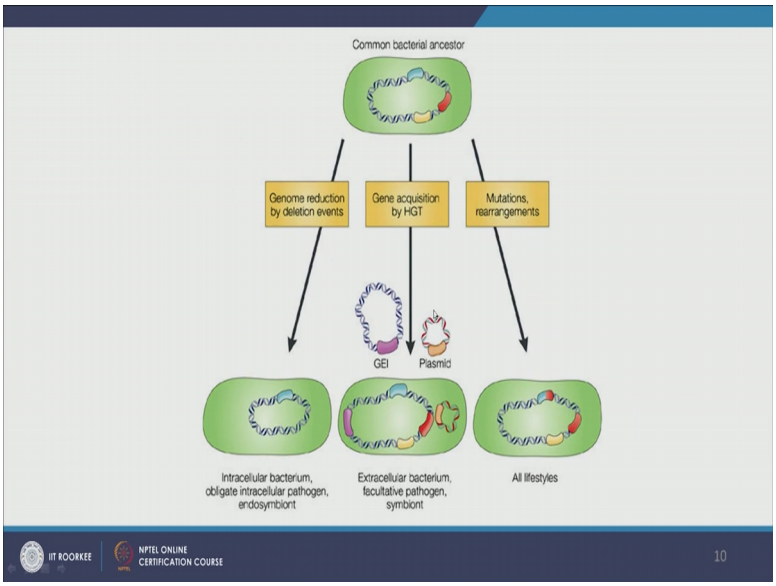
Now, let they are more interested in the gene a, by the way. So, it is highlighted in red. Now, horizontal gene transfer is when 2 microbes that are close to each other they share genes. Now, it does not have to be microbe one of them could be virus as is the case with conjugation, one of them they can either just pick up the gene that is lying in the environment. For example, here a is lying in the environment the recipient said initially does not have a, but it picks up it transforms a it allows the a to come in and makes it part of it is genetic material. It might survive here as a plasmid, it might actually become part of the chromosome. Either case, this is transformation or they can do this where they actually form a bridge to microbes and this has a this does not have a. So, they form a bridge, it gives it a copy of a. Now, both of them have a and they destroy the bridge after the sharing of genetic material.

(Refer Slide Time: 19:27)



So, we notice that microbes can share genes horizontally by undergoing transformation, transduction and conjugation.

(Refer Slide Time: 19:36)



So, to summarize we have a common bacterial ancestor, it can undergo deletion. So, it loses the red and yellow genes. So, this happens a lot with intracellular bacterium. So, there are hints that suggests there is evidence actually that suggests that our certain organelles in some eukaryotes most eukaryotes; for example, mitochondria, plasmids they were initially bacteria that lost many of their genetic elements due to deletion and because they do need it because they were having a very nice symbiotic relationship with a bigger organism microorganism. So, they did not need the genes and they underwent deletion.

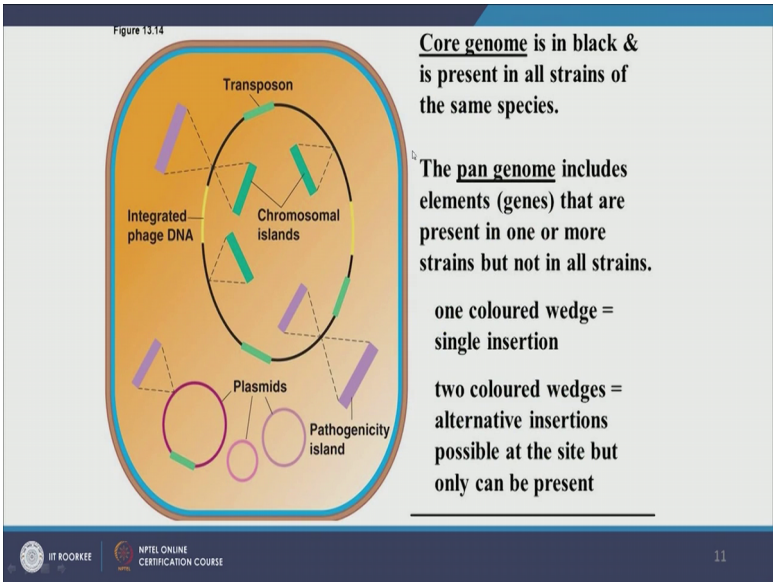
So, they have a smaller genetic size than a typical bacteria, but they are commensal and they are symbiotic with other microbes. So, these are intracelular bacterium. They can be also obligate intracellular pathogens endosymbionts. So, endosymbionts would be example of mitochondria would be an example, obligate intracellular pathogens would be pathogens they need to survive inside somebody yeah.

So, some cells need to invade that we have talked about parasitic cells. So, these parasitic microbes need to invade another host in order to replicate. So, they do not need to carry out all the functions that are typical cell needs to carry out to replicate and to survive. Thus, they survive deletion really well and definitely intracellular bacterium. The other possibility is that they can acquire new genes by horizontal gene transfer. So, this is a gene mobile genetic element is carrying blue purple gene it and this is a plasmid that is carrying orange gene and then this cell, I am sorry, this cell can uptake them via transformation. So, this will become

actually part of the chromosome and the plasmid will freely float around. And, then the other possibility is mutational rearrangement. So, earlier it was blue, green, yellow, but now it can be blue, red, red, yellow. So, they can rearrange and they can have all kind of lifestyle. The one of the example is inversion that I talked about a, b, c will become c, b, a or they might cut copy paste with each other. So, these are mobile elements is a very important term.

So, when you talking about mutational rearrangements they are mobile elements in prokaryotes and eukaryotes that can skip. So, this is more often the case with eukaryotes. So, a particular mobile element can skip from one part of chromosome and go and join some other part of chromosome insert itself there or a plasmod used to be a plasmid can now be part of chromosome. What used to be an extracellular DNA material, now becomes part of chromosome. So, these are all transposon for example, we look into them. So, these are extra chromosome, but these are mobile elements in their chromosome.

(Refer Slide Time: 22:27)



So, let us look at this figure it will make it more clear what you see in black here, this part what you see in black here is your core genome and it is present in all the species all the strains in this species. So, for example, if this is E.coli and all the E.coli will have the black ones. All the multicolored stuff you see the one in yellow the chromosomal islands in blue and green and purple all these are extra stuff that may or may not be present another E.coli. The pan genome includes everything. So, the plasmids, the islands chromosomal, islands are

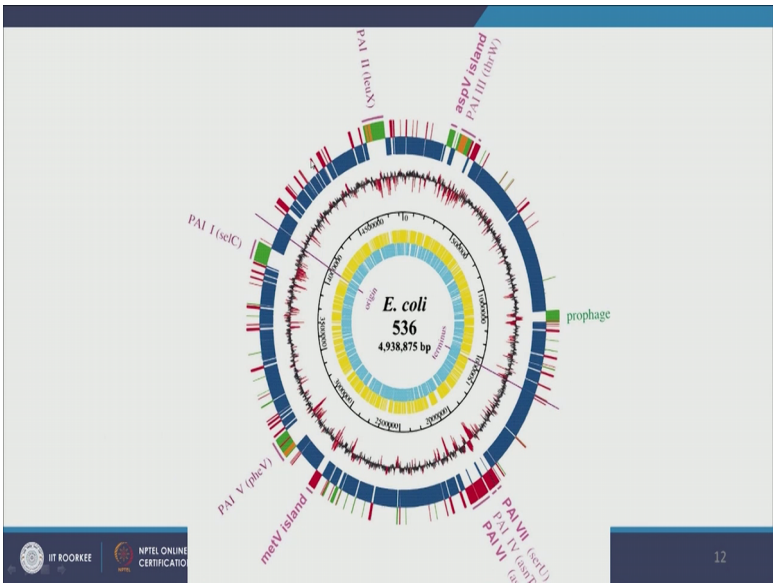
integrated phage DNA, phage (Refer Time: 23:04) viral DNA. All of them are a kind pan genome together. Some of these colorful elements may not represent another E.coli.

So, if you see one colored wedge here, this is single insertion. If you see 2 wedges here then it is alternate insertion, either the purple will insert or the blue will insert, but only one of them will insert. Then, they have transposons that have made themselves part of the chromosome material. Now, this is very important especially when we are talking about pathogenicity of virulence. For example, I want to know I have mentioned this example earlier in the class today; I want to know which E.coli is going to be pathogenic, which E.coli is not going to be pathogenic.

So, I can compare all the E.coli and I can identify the black parts. The core genome; the core genome is the one that all E.coli will have. So, I can say the core genome is not the reason why K-12 is not pathogenic, but some other E.coli is highly pathogenic. Once, I identified then I can look for the mobile genetic material, the chromosomal islands, the transposons, the plasmids and the integrants and once I notice them I can identify pathogenicity island. The E.coli that have this particular insertion all of them are pathogenic.

So, this might be the reason, this is the island of genetic material when it sticks in here and we have pathogenicity.

(Refer Slide Time: 24:24)



So, this is an E.coli picture for you. Now, in this E.coli you notice that them some core is a core genome and then beyond the core genome we have many genetic islands. So, some of them are the reason why we have pathogenicity and this pathogenicity causing. So, what advantage we get from this is if we notice a new another microbe, for example, staphylococcus aureus which have a similar genetic sequence that gives pathogenicity to E.coli and I can say all rightly this staphylococcus aureus might be pathogenic.

Now, here is a cool neat feature. If we look here, we have integrated phage DNA the yellow one. So, this is this used to be viral DNA that has been integrated into the chromosome of this particular equalize this particular bacteria. Now, in case of staphylococcus aureus some staphylococcus aureus are pathogenic, some are not. And, we notice that the ones that are pathogenic have these pathogenic islands in them which make them virulent, which make them harm human beings, public health.

Now, if now these pathogenic islands for staphylococcus aureus are highly mobile. So, they can very well remove themselves from there, make a copy or without removing make a copy and then the copy can insert itself into a staphylococcus aureus that had no pathogenicity island and thus, in this way staphylococcus aureus is notorious for picking up pathogenicity. So, for example, in my nose, in fact, 3 out of 1 human being sorry 1 out of 3 human beings have staphylococcus aureus in our nose, they are just commensal, they are staying there.

Now, these staphylococcus aureus can pick up virulent genes very easily integrate them into the chromosome and become viral and become pathogenic and cause infections really nasty infections. The other reason why this is very important to study this core genome and the pan genome is because there are because it is not just pathogenicity that transfers, but there is a very challenging problem of antimicrobial resistance. The antimicrobial resistance we will talk about it in a little more detail later, but it is the ability for microbes to resist antibiotics to resist antimicrobial drugs.

Now, many of these antimicrobial resistance genes are actually present in these not in the core genome because core genome will have what the bacteria need to survive. So, they are not present in the core genome, but they are present in the pan genome minus core genome, the mobile elements of the genome. So, when a bacteria that has them present it is exposed to antimicrobials, it will survive, it will populate itself and when it populates because others are

dying it has a higher chance of transferring that mobile element it is giving its resistance to other susceptible microbes.

For example, few decades earlier all pathogens or most pathogens were susceptible to antimicrobials expose them to antibiotic antimicrobials and they died, but now the soil microbes that were usually or the environmental microbes that were resistant to some of these anti microbials, they managed to transfer their antimicrobial resistance genes into the pathogens using this mobile elements.

So, this is a typical case of horizontal gene transfer and this is very important when we are talking about our environmental challenges such as antimicrobial resistance, how it spreads to environment and also when we are talking about degradation of certain pollutants. Let us say a microbe does not require a pollutant, it does not require to degrade a pollutant to survive, but if it gets the genes to do it, it will do it. It will degrade it. So, or it will throw it out, it will have an efflux pump. So, in that case, we know that it is very important for us to understand the genetic mobile elements that are encoding the genes that are in case they have the genes which allow or make a microbe to degrade the pollutant.

Another example in one of the previous lectures I talked about mer gene which gives microbes resistance to methyl mercury. The methyl mercury is very toxic. What mer gene does is it encodes the protein that will divide, break down methyl mercury into mercury. So, if I have lot of methyl mercury environment and I know what the gene is and I can transfer that gene to the environmental microbial community that is present in environment and I can make them degrade this very fast. So, I will have more mercury and less methyl mercury. Thus, I will reduce the toxicity of environment. So, mobile elements are very important and very relevant when we talk about both environmental challenges and public health challenges.

Alright, dear students, this is all for today. In the next lecture, we will explore a little bit more about genomics and then henceforth, we will start moving into environmental challenges one by one as a given in the syllabus and this is all for today.

Thank you very much.