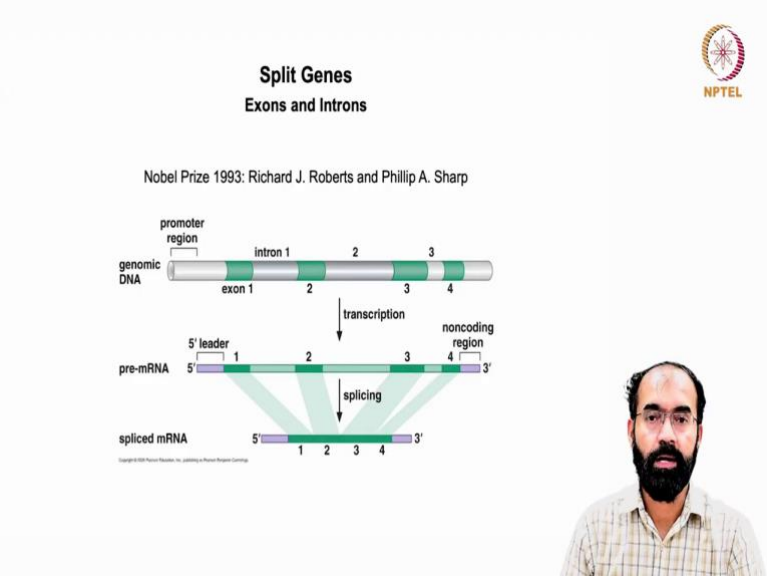
RNA Biology
Prof. Rajesh Ramachandran
Department of Biological Sciences
Indian Institute of Science Education and Research, Mohali

Lecture - 24
RNA Splicing, Export and Stability: Relevance of Introns

(Refer Slide Time: 00:21)

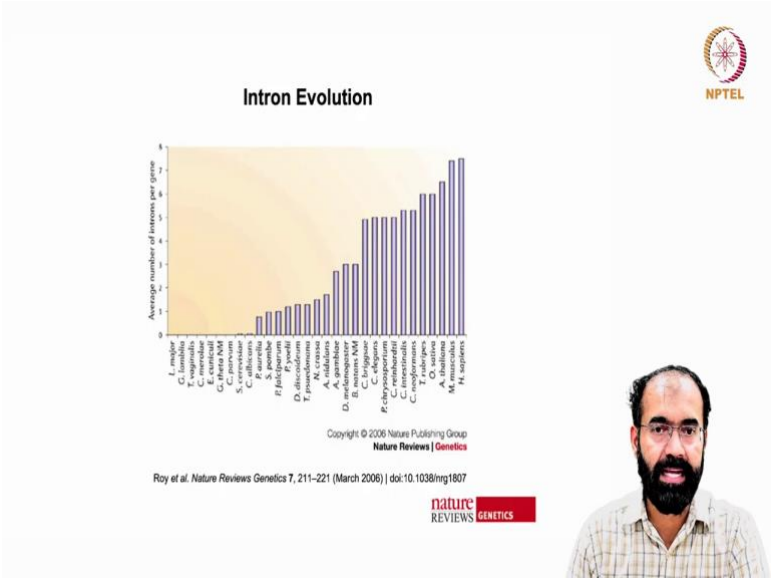


Hello everyone. Welcome back to another session of RNA Biology. So, we were here talking about the split genes. That means the genes are split in the genome and they are split in the pre mRNA and the split do not appear as split in the processed mRNA.

And we will see I am just giving some clue that future we will revisit this and to show that why this boundary or why this recognition and who is allowing this boundary recognition etcetera which proteins are binding onto the exon part? Which proteins are binding onto the exon-exon boundary? And what is their role etcetera.

So, keep that in mind. Before going into those direction, we will understand or we will try to follow the evolution of introns.

(Refer Slide Time: 01:15)



So, intron evolution is a very interesting topic. Because as the name itself indicates, intron is basically a intervening region or an intruder sequence which is introduced as a nuisance in the first appearance. But the actual truth is it is not nuisance. So, we will try to understand what is their relevance? How introns are evolved? Can we drive some information? So, can we inherit some information from the introns across different clades of animal.

So, like if you keep seeing different groups like in y axis you are having average number of introns purging and in the x axis you are having different species individual species. As you can see here from yeast onwards, *Saccharomyces pombe, Saccharomyces cerevisiae* etcetera you have got this is the origin from the yeast and these are all saccharomyces cerevisiae have got very minimal number of introns and *Saccharomyces pombe*.

This is *Saccharomyces cerevisiae* is the baker's yeast and saccharomyces pombe is fission yeast. So, these as you progress from yeast onwards you reach up to *Homo sapiens*. Different mammals are there, different stages of you know *Neurospora crassa* is there which is a fungus species.

And as you can see here in mammals, C. elegans is worm and number of the name of this animals are not important. But understand in the evolutionary time scale as the animal

complexity increases you always have a increase in the average number of introns per gene.

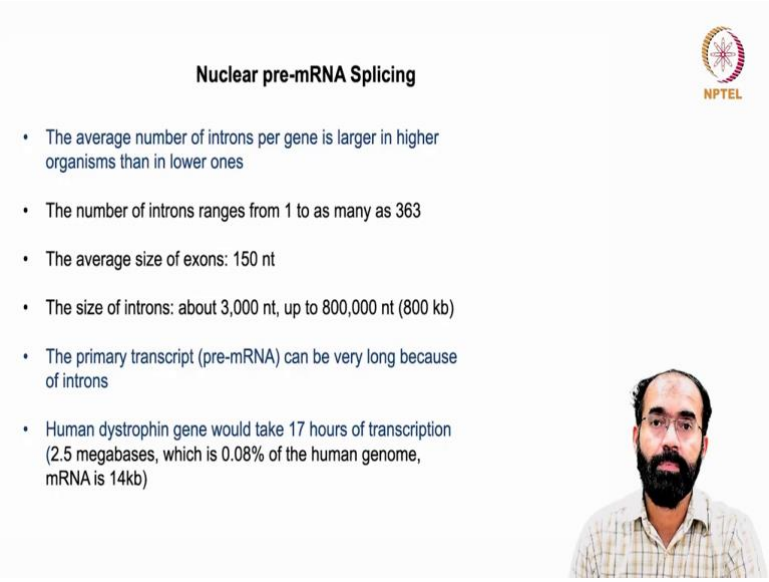
So, we can see here 1, 2, 3, 4, 5, 6, 7, 8. So, average number in human average number of genes with introns or the average number of introns per gene in *Homo sapiens* is 8. So, this number remains more or less same in the mouse and slowly decreasing. And if you see around here some 1 and half to 2 around this yeast do not have normally this unicellular eukaryotes. They do not have that many number of introns per gene.

But introns are there not that many this is of course, is an average number ok. So, total number of genes and you are looking for total number of introns divided by total number of genes. Say total number of intron say 10,000 is there and divided by total number of genes.

So, maybe around 15,000 or something or maybe 8,000, 9,000 something like that genes I am just giving a crude example. So, you have around 1 or 1.5 introns per gene average. But whereas, in human we have around 7 to 8 introns per gene. Of course, it is an average number there are genes in human with one intron or no intron and some genes with maybe some 20 or 30 introns average number is 7 to 8.

So, we should understand that as the evolutionary complexity is increasing the number of genes also increasing.

(Refer Slide Time: 05:01)



So, let us quickly see a little bit more in detail about the nuclear pre-mRNA splicing because their relevance is important to connect here. The average number of introns per gene is larger in higher organism than the lower ones. So, we know higher and lower is a relative term higher in terms of organismal complexity.

Yeast does not have liver, yeast does not have kidney, yeast does not have brain etcetera. But it may have some of the enzymes which are produced in kidney, liver, brain etcetera, but it is not organized into tissue structure; that is what we are referring to as higher versus lower ok. And we know as the organism becomes higher in complexity the number of introns also increasing and that has to have some relevance in introducing the organismal complexity.

So, number of introns range from 1 to as many as 363 in in a given gene. The average size of exons remains more or less same around 150 nucleotides. It does not change much. The size of introns is average size is around 3000 up to 800,000 nucleotide. 800,000 is a huge figure 800 kB a single intron. And interestingly we should also understand some genes, one intron can have multiple other genes also. So, that is something quite interesting to mention here.

When you say 800,000 nucleotide is there in one intron do not think that much area is a barren land or its only having an intron. Some introns have got a promoter a multiple exons and a gene and poly A signal etcetera. So, what you should understand genes can lie in an overlapping manner. So, this concept should be clear to you. So, intron is with respect to a given gene, but not intron is not a barren land or an unwanted piece with respect to a gene rich or gene poor point of view.

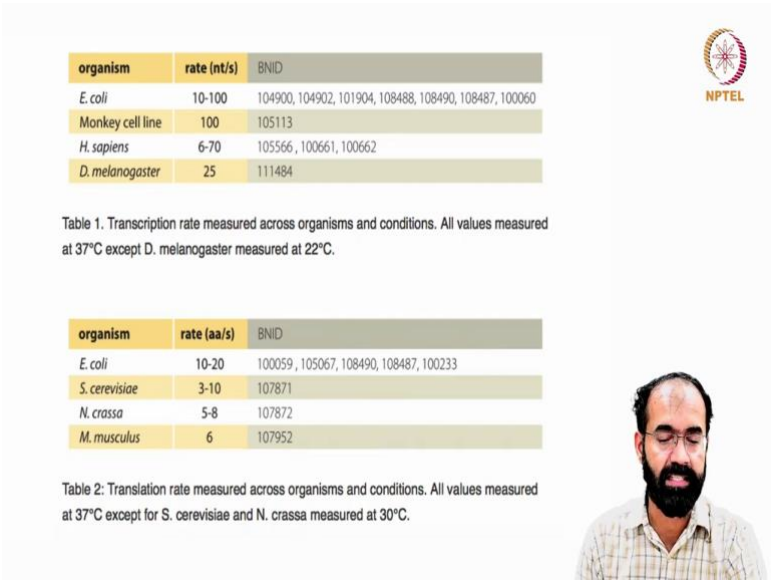
In drosophila several examples are there even I think in one gene in human also has similar example where I think it is expressed in liver, I do not really remember the name of the gene, but intron can harbour another full gene. So, the primary transcript that is the pre mRNA can be very long because of the introns and this takes longer time because the transcription takes longer and longer and longer.

Sometimes it can even take hours for the transcription like human dystrophin gene like Duchenne muscular dystrophy is caused in human males mainly because of defective protein dystrophin protein. So, dystrophin is an important gene in maintaining your musculature.

Human dystrophin gene would take 17 hours for transcription of 2.5 mega base. Remember 1000 base is 1 kilobase 1000 kilobase is 1 mega base. So, this you are talking about single gene dystrophin 2.5 mega bases which is about 0.08 percent of the entire human genome 1 gene is occupying that much space and the mRNA is only 14 kB.

So, after proper splicing mRNA so, the pre mRNA or the unprocessed mRNA will be 2.5 mega base which now after splicing it becomes 14 kB still it's a big protein, but it takes longer time it takes around 17 hours for the transcription of this gene. And the good thing is these proteins are quite stable and stay for long their life is pretty long. Once it is produced it stays for longer time because it is mainly structural protein.

(Refer Slide Time: 08:55)



organism	rate (nt/s)	BNID
E. coli	10-100	104900, 104902, 101904, 108488, 108490, 108487, 100060
Monkey cell line	100	105113
H. sapiens	6-70	105566 , 100661, 100662
D. melanogaster	25	111484

Table 1. Transcription rate measured across organisms and conditions. All values measured at 37°C except D. melanogaster measured at 22°C.

organism	rate (aa/s)	BNID
E. coli	10-20	100059 , 105067, 108490, 108487, 100233
S. cerevisiae	3-10	107871
N. crassa	5-8	107872
M. musculus	6	107952

Table 2: Translation rate measured across organisms and conditions. All values measured at 37°C except for S. cerevisiae and N. crassa measured at 30°C.

Now, let us see little bit more in detail about the organism and their you know rate of DNA, RNA replication, protein synthesis etcetera. In E. Coli the rate of nucleotide synthesis per second is 10 to 100 ok. So, that is the rate of nucleotide synthesis and some cell lines monkey cell line it is around 100 nucleotide per second and in humans on an average it is 6 to 70 nucleotides per second.

In drosophila it is averaging around 25 nucleotide per second that is the rate of synthesis. Transcription rate measured across organisms and conditions all values measured at 37 except for melanogaster because it is staying at around 22 degree Celsius.

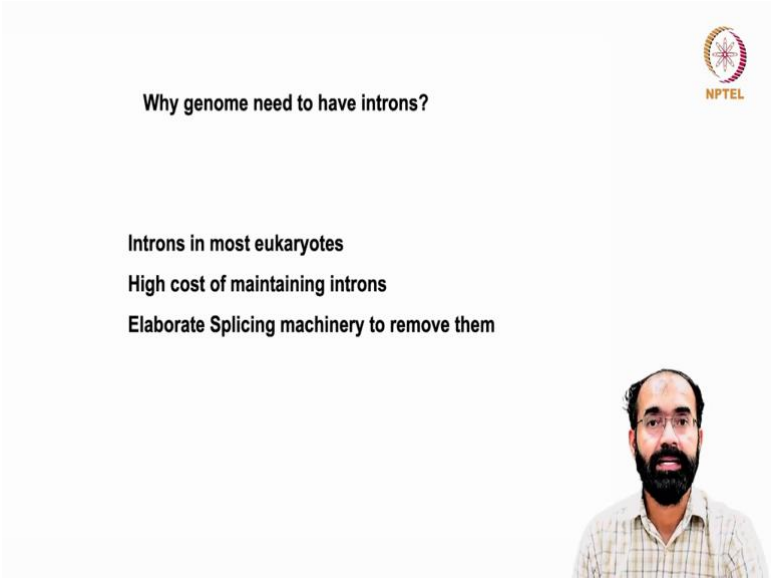
So, on an average you can see that E. Coli has the highest rate of transcription because its lifespan is around 20 minute. So, it has to perform the task much faster and cell lines also leave in a quite congenial condition with lot of nutrition, high temperature like that high I should not say high temperature because 37 is a normal temperature for the monkey's body also. But being cell lines they have nutrition quite easily available and they have a rate of production of 100 nucleotides per second.

On the other hand, if you look at the rate of amino acid or the protein production in E. Coli its around 10 to 20 amino acids per second and in yeast it is around 3 to 10 amino acids per second because E. Coli leaves in 37 degree, yeast leaves around 30 degree. So, this also could be one of the reason in neurospora another fungus its around 5 to 8 bases not bases 5 to 8 amino acids per second. This you are talking about the protein synthesis rate of production of protein from a given mRNA.

And Mus musculus it is roughly around 6 nucleotides per second and in human also more or less the same. Of course, it varies depending upon which cell you are talking about like whether you are talking about liver versus brain etcetera that can also influence the rate of production of protein from a given cell.

So, the translation rate measured across organisms and conditions all values measured at 37 degree Celsius in Saccharomyces cerevisiae, Neurospora crassa it is done at 30 degree Celsius.

(Refer Slide Time: 11:20)



So, now let us ask the most important question why genomes need to have introns? What is so much greatness? Or what is the importance of adding this extra complexity of introns on to the genes?

Introns are present in most of the eukaryotes. High cost of maintaining introns exist. Elaborate splicing and machinery needed to remove them. It is just like you create a trouble and solve the problem. You may wonder why? Just like you are climbing second floor of your house you are jumping from there breaking your leg and you are getting it fixed in a hospital and you may wonder if you would not have climbed down to the terrace would not have jumped you do not need to fix it you know.

So, this is what it may sound little baffling, but it is not done because of a solving a problem. It is done with a purpose. It has got huge importance huge relevance. One relevance I already told you that if you make a transgenic animal without introns, it is 100 times less efficient in getting expressed.

So, introns allow the expressivity of a gene. Of course, it does not various other functions say if you are talking about mutation. All of you would have heard about mutation right like if you get mutation of the p53 gene that can induce cause cancer because p53 is a guardian of the cell. It makes sure that no damage occurs to the DNA if there is any damage p53 helps in fixing it.

So, p53 does 3 main functions one function is immediately there is a DNA damage is detected p53 is turned on p53 first it will do stop the cell cycle. If it is a dividing cell, it makes sure that no it will not allow the cell cycle to progress. Why? If the cell cycle is progressing it will become a mad or bad or a defective cell is being propagated just like you are running a restaurant the food is gone bad you will not send out for eating right you will not send out that food; you will destroy it because food is bad.

If it is gone out it may cause so many people. So, that is what p53 will do. If p53 is turned on that there is a damage in the DNA. The first goal is to stop the cell cycle and second goal is it will turn on DNA repair genes. It has to turn on the DNA repair genes and they will try to fix it and p53 can monitor if the DNA damage is fixed if the DNA damage is fixed because p53 production will stop if the DNA damage is not there if DNA damage is there the p53 production will start.

So, the constant monitoring system will see whether the DNA repair happens. So, first it will stop the cell cycle by additional time and turn on the DNA repair genes. Still the DNA repair did not happen it will mark the cell for apoptosis that is programmed cell death. So, p53 make sure that no mad cell or no bad cell goes to the next stage.

But if p53 has gone defective then that can there is no quality control. It is almost like a school or a class where there is no teacher, there is no guardians like a country without police or country without army there will be full anarchy whoever has got muscle power he will win.

So, this is what happens that causes the cancer in majority of the cases. Like that, if elaborate machinery is there to remove it to fix it, it comes with a extra load of this introns also. So, if a gene is vulnerable to mutations like I told you about the p53. If a gene is vulnerable to mutation what if the gene is diluted with lot of introns; like a gene supposedly should have been in 1000 bases.

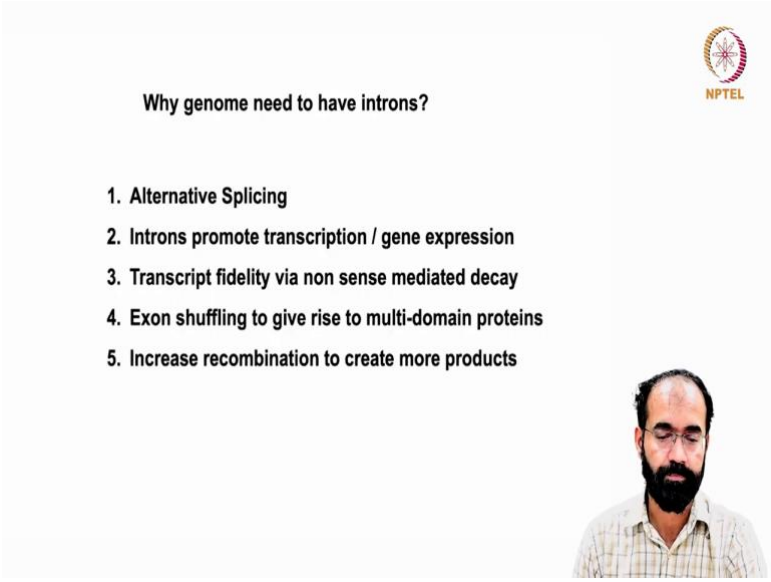
Now that 1000 base gene like you saw dystrophin example 14,000 bases only is the actual mRNA, but that is split into 2.5 mega bases or 2.5 million bases. Why you spread out so much? You are diluting the gene with the help of introns. So, whenever there is a UV damage or chemical mutations etcetera is possible in your genome. Introns can take it introns.

So, there is a good chance that is why you know earth surface is covered more than 70 percent is covered by water. So, that is why when a satellite is falling down or a meteorite is coming onto earth less likely it will hit the land because land mass is only less than 25 30 percent; very small 70 percent is ocean.

So, there is a good chance that it will go and hit the ocean rather than a someone's house. Same way when there is a mutation is possible because of UV radiation or chemical induced mutation. There is a good chance that it may go into an intron which anyway is going to be removed. So, it can act like a guardian.

So, this also you should understand gene expressivity is one thing and it can also act as a guardian and many more are coming one by one, we will see them in detail.

(Refer Slide Time: 16:36)



First thing is if you have introns, you have you can do alternative splicing; that means, if only one exon is there. How are you going to make multiple pieces out of it? Because one gene, if it is put into 10 pieces you can make use of piece number 1, piece number 2 and piece number 6, one product came.

Or you can use piece number here I am referring to exon ok. One gene is split into 10 exons. Exon number 1, exon number 2, exon number 3 and exon number 9 another variety came or exon number 1, exon number 8 and exon number 25 not 25 exon number 10 or some genes can have multiple exons.

The more number of exon we also saw DScam drosophila cell adhesion molecule we saw 38,000 different proteins can come from one by alternative splicing. So, having a gene split into multiple exons allows you or brings you extra freedom to bring in diversity.

So, alternative splicing is one option and introns promote transcription and gene expression like I already told you that 100 times more expressive if the trans gene is made or a gene is having introns and transcript fidelity via nonsense mediated decay. We will see more in detail about the nonsense mediated decay when we study. Because there are different types of RNA decay is possible, but keep this in mind transcript fidelity fidelity means reliability that is what fidelity means.

So, transcript fidelity is monitored via nonsense mediated decay. What is nonsense mediated decay? We will see in detail rather than just verbally saying right now. Because we are going to address them in the future classes and exon shuffling to give rise to multi domain proteins.

Many a times proteins have multiple domains. Like we saw in the case of nucleotide editing enzyme like cytidine deaminase, we saw some example they have an RNA binding domain and they also have a deaminase role like every protein will have a substrate recognition domain and also the catalytic action domain.

And many a times this multi domain proteins need to be rearranged many times you will see one particular domain is present in 10 different protein and there is nothing common to them. There are completely 10 different functions, but this domain is common in them. Like that what you should gain from or infer from this is that many times certain domains are essential for many proteins although their function is not common.

So, exon shuffling that is shuffling of the exon you can remember if you are playing cards or you are playing many games are there. You shuffle the cards right because you do not want the same pattern after playing once the playing is completed you shuffle them. So, sometimes these kind of shuffling brings in diversity in the distribution of the domains.

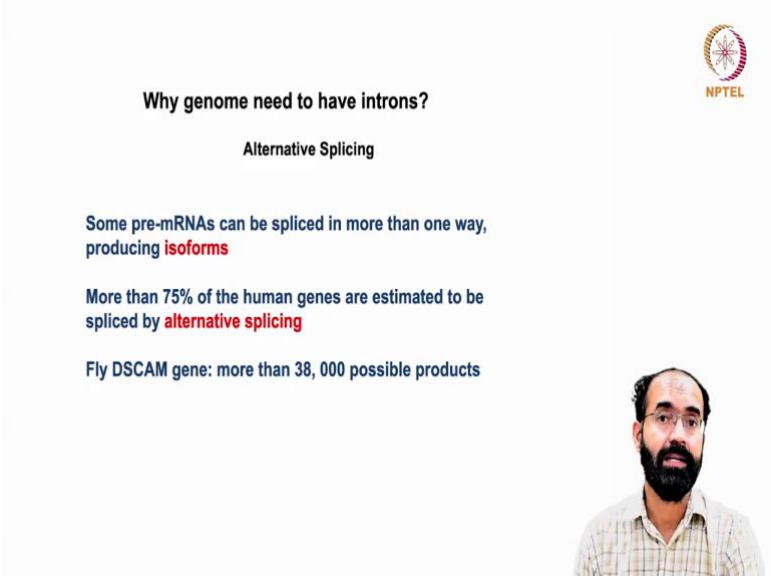
So, this is brought in in the case of multi domain proteins and many big proteins or many enzymes are multi domain. This minimum they will have two domain. One is substrate recognition domain and another is their functional domain where they are actually doing the job. And then other function is increase the recombination to create more products.

Because, if you have introns that can cause recombination; recombination means you can cut it and then get integrated elsewhere because, if you cut it in the exon there is a good chance that you disrupted the gene. Because we have seen when we studied this splicing that it has to be single nucleated precision.

So, it will not tolerate variation. So, it is very important that introns allow the recombination. It can also incorporate a portion of intron elsewhere. Sometimes it can be problematic also, but many a times this recombination during the early stages of evolution of organisms on planet earth; this recombination has helped a lot in bringing in

complexity. Because it can rearrange the genes much effectively. It can cause inversion like at chromosomal level by because of this recombination etcetera.

(Refer Slide Time: 21:05)



And coming back again continuing the same question why genomes need to have introns? One we already know causes alternative splicing and another few examples include some pre-mRNAs can be spliced in more than one way and this produces isoforms.

What are isoforms of a gene? Isoforms are basically the same gene, but in different issue. It can make use of a different stretch of fragment and it can bring in tiny changes. Sometimes it may be introducing a micro-RNA recognition element. Sometimes may be changing one amino acid.

It is not mutation; it is a slight alteration occurring onto a gene. Sometimes it may have a higher ability to perform. Like say for example, it is said that in Asian population the alcohol dehydrogenase people consume alcohol right, alcohol dehydrogenase is much less efficient in the case of Asian population compared to the Caucasian. So, it is same alcohol dehydrogenase it is not that they have a totally new gene.

But the isoform is such that it can perform, it can degrade alcohol in a much efficient and much better way. So, these are all some kind of variations and many enzymes are there. Some enzymes will perform better than the other and it is also you can see sometimes

Caucasian populations are more vulnerable to infection compared to people who are living in tropics such as Asians.

So, this is another example. So, that is how you are because they are very much vulnerable for serious flu infection. Whereas, in Asian flu is not a major thing they get you know few days of some uneasiness and it will just get cured by itself.

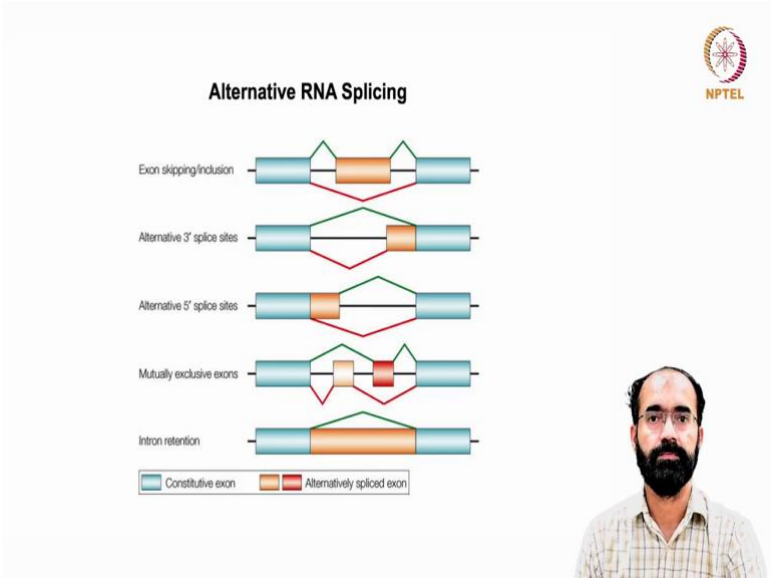
But this is because of variations in the gene isoform. It is not that Caucasians are another species and we are another species or Asians are another species no. They are all homosapiens, but there can be variations in the isoforms of the gene which in turn can bring in how they react to a given substrate or how they react to a given circumstance.

More than 75 percent of human genes are estimated to be spliced by alternative splicing. This I already told you that we have only 23,000 coding genes present in our genome which is far less than the required number of genes considering a complex organism such as human.

So, it is believed we should have around 1 lakh genes. It is a ballpark figure nobody knows that 1 lakh is a very you know it is a crude number, but 23,000 is no way close to one lakh. So, this extra balancing of 75,000 number of genes could have come from the alternative splicing only.

So, that is why more than 75 percent of the human genes are believed to be alternatively spliced and example we have seen the drosophila DSCAM gene more than 38,000 possible transcripts come out of a single mRNA.

(Refer Slide Time: 24:08)



What is this slide shows? How alternative splicing has implications in bringing in diversity? So, this first example is exon skipping or exon inclusion. Exon 1, exon 2, exon 3. So, in this bottom portion exon 1 and exon 3 are fused it has done the exon skipping that is exon 2 is skipped.

Whereas, this is inclusion exon inclusion 1, 2 and 3 are fused. This is one way another examples alternative 3 prime splice site; that means, it can make use of a particular splice site or not. Like if this exon after the branch point has done the first transesterification reaction of the splicing, then the free exon should decide whether it should pair here second transesterification reaction or here.

If it is doing here, the blue exon fuse with the blue exon. If it is doing here, it will include a part of the intron also. Whereas, in another scenario it is doing exon 1 and going to the correct exon. So, this is by inclusion of alternative 3 prime splice site and then inclusion of alternative 5 prime splice site; the same thing repeated, but the thing here the branch point is fixed and the branch point decides where it should go and release.

Whether it should release the first exon from here or whether it should release the first exon from here. Accordingly, it will decide whether it should fuse from here or whether it should fuse from here this is another variation. And then mutually exclusive exons; another way of alternative splicing. Here is exon 1 and exon 2, exon 3, exon 4. It skipped

exon 2, but included exon 3 and exon 4. Whereas, in another case it included the exon 2, but skipped exon 3 and that it caused variety that is called mutually exclusive exons.

So, intron retention; sometimes an entire intron if it is; if it is a huge intron it will know, but if it is a small intron, it can retain and this is how the alternative splicing can function in different ways to bring in complexity. We will study in detail about the diversification of RNA in the next class.

Thank you.