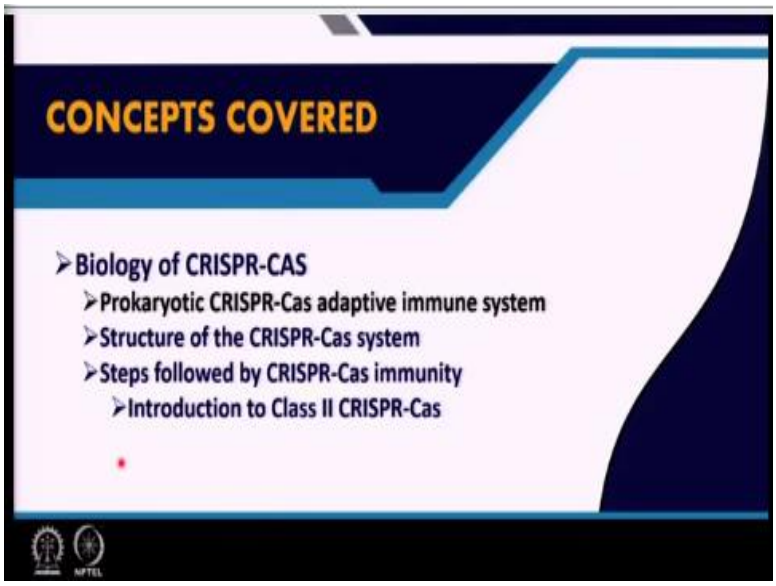
Metabolic Engineering
Prof. Pinaki Sar
Department of Biotechnology
Indian Institute of Technology-Kharagpur

Lecture-30
CRISPR-Cas System and its Application in Metabolic Engineering-Part II

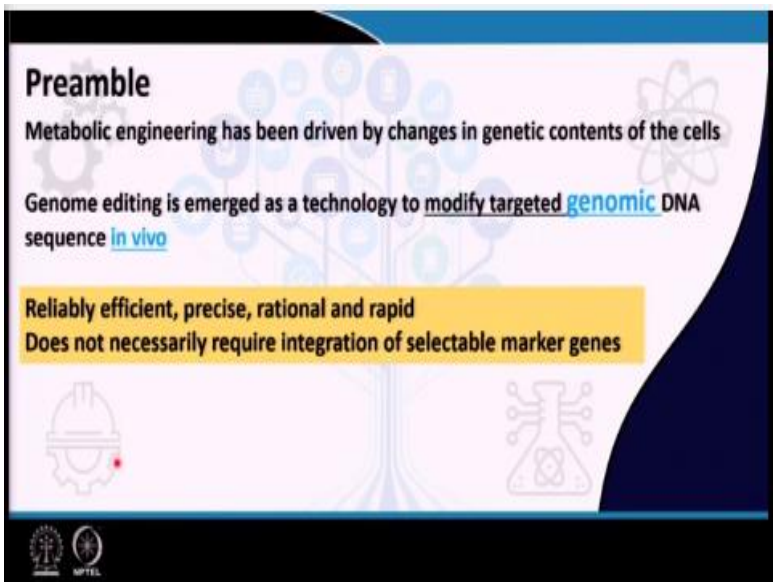
In today's lecture on metabolic engineering, we are going to discuss about the CRISPR-Cas system. And would like to emphasize upon the major points, and aspects about it is application in genome engineering.

(Refer Slide Time: 00:48)



So, the points that are going to be covered are mainly the biology of CRISPR-Cas, including the prokaryotic CRISPR-Cas adaptive immune system. The structure of the CRISPR-Cas system, the steps followed by CRISPR-Cas immunity and then the introduction to class II CRISPR-Cas, which is found to be widely used in genome engineering.

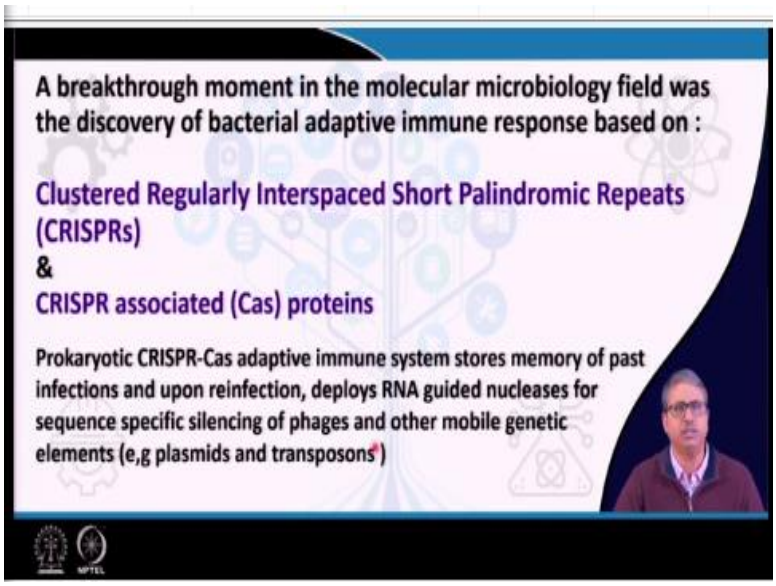
(Refer Slide Time: 01:15)



Now, metabolic engineering has been driven by changes in genetic content of the cells for various purposes, whether to produce new compounds or to produce the existing compounds or to change the quality of the compounds or even to change the physiological properties of the cells. So, one of the best methods evolved in the past decade is the genome editing technology. So, genome editing technology is achieved or performed by a number of methodological approaches.

And one of the most widely used and possibly accepted approach among them is the use of CRISPR-Cas technology. There are various reasons for identifying these particular methods. Now, one of the most critical and important aspects or advantages of these CRISPR-Cas based genetic or genome engineering methods, it is applicable in vivo. It is reliably efficient, precise, rational and rapid and very distinctly it does not necessarily require integration of selectable marker genes.

(Refer Slide Time: 02:44)



Now, introduction of CRISPR-Cas system which is found to be very useful and versatile has spawned various forms of derivative tools. So, subsequent to the discovery of CRISPR-Cas system particularly the elaboration of the class II system. A number of derivative tools have been identified and used for manipulating the gene function expanding the fields of application of metabolic engineering as well as the organisms in which these genetic perturbations or genetic alterations can be met.

Now, one of the very important aspect of is this particular event is the breakthrough moment in molecular microbiology, which was the discovery of the bacterial adaptive immune response, which is based on the CRISPRs. The full form of CRISPRs is the Clustered Regularly Interspaced Short Palindromic Repeats. And associated to these particular CRISPRs locus of the repeats, there are CRISPR associated proteins or protein coding genes.

So, these CRISPR systems are located within the genome of prokaryotic microorganisms like bacteria and archaea. And there are 2 distinct systems involved in these CRISPR; one is the CRISPR itself that is the regularly interspaced short palindromic repeats and the spacers that we are going to discuss today. And the associated or the protein coding genes which are called Cas protein system.

Now, prokaryotic CRISPR-Cas that is how it is actually used the term CRISPR-Cas system, it is considered as a adaptive immune system which stores memory of past infections. So, whenever prokaryotic cells, archaea and bacteria, they are exposed to phage or plasmid DNA invading other DNA, the history of infections are recorded, how? That we are going to see today that these infecting DNA molecules part of those DNA molecules are stored within the genome of the host organism.

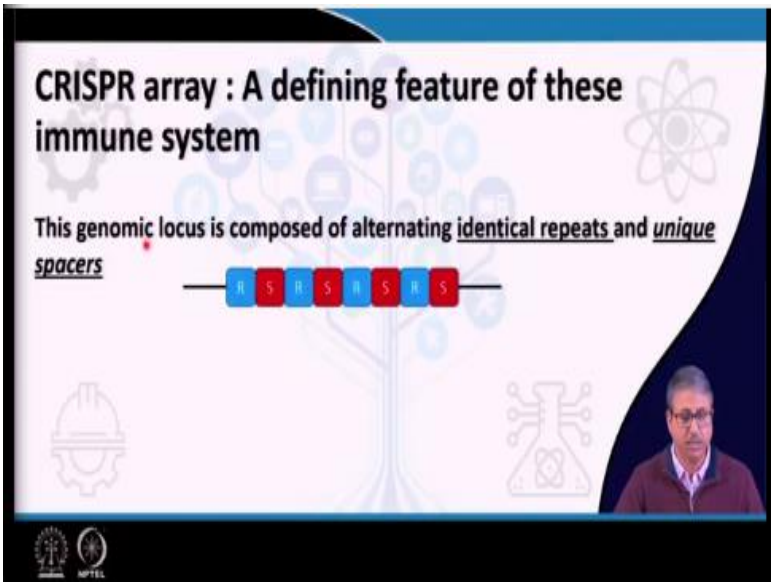
That is each of the bacteria and archaeal genome is found to harbor a number of the residual, which are called the kind of a memory of the past infection, the short of DNA which today we call them spacer within the CRISPR. So, these spacers are basically the remnants of the past infection by phage or plasmid. Now an evolutionary these are past from one generation to the others, very faithfully the genome is replicated.

And each time the daughter cells are formed, the entire genome is copied to the daughter cells during the genome replication. And the entire spacers which are there in the CRISPR as well are transferred to the daughter cells. And eventually upon reinfection, maybe after several years or several round of cell division some of the progenitor cells are exposed to again reinfection by the similar type of or a different type of phage or other virus.

Then there are systems like the CRISPR is that system which allows deployment of RNA guided nucleases. So, these are nucleases which are basically capable of cleave the DNA molecule but they are guided by RNA. Now, what is the source of this RNA? This RNA as we will discuss today, this RNA is coming from the CRISPR. So, CRISPR is getting transcribed and producing RNAs, short RNAs, those RNAs are called CRISPR RNA.

And they are guiding the entire nuclease complex, that is basically the Cas complex for sequence specific silencing of phage and other mobile genetic elements. So, it is very well applicable and used for any kind of invading DNA containing pathogens or mobile genetic elements even that is plasmids and transposons.

(Refer Slide Time: 07:00)

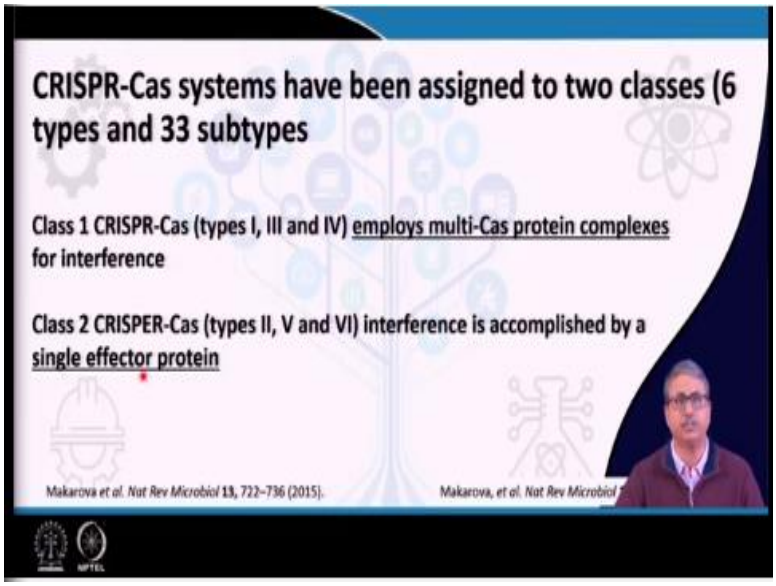


Now, one of the defining features of these immune system is the CRISPR array. So, as I said it is located within the genomic DNA, so CRISPR array is located within the genomic DNA, it is part of the genome. And this genomic locus which is referred to as the CRISPR array is composed of alternating repeats and spacers. So, here the repeats are marked as the blue box and the spacers are marked as the red box.

So, we have alternating repeats and spacers. So, you have a repeat and a spacer, then another repeat then another spacer, then another repeat and it goes on. Now, interestingly, these repeats are all identical. So, the sequences of these repeats are all identical for a specific organism or a specific genome, whereas the spacers are all unique, because, each of these spacers are actually derived from past infection by phage or mobile genetic elements.

So, each of these spaces could be or actually different, so S1, S2, S3, S4 and so on. So, all these spacer molecules are going to be different in terms of their sequence, so that is why it is called unique spacer. So, these alternating residues of repeats and then spacer is very characteristic property of this CRISPR array.

(Refer Slide Time: 08:29)



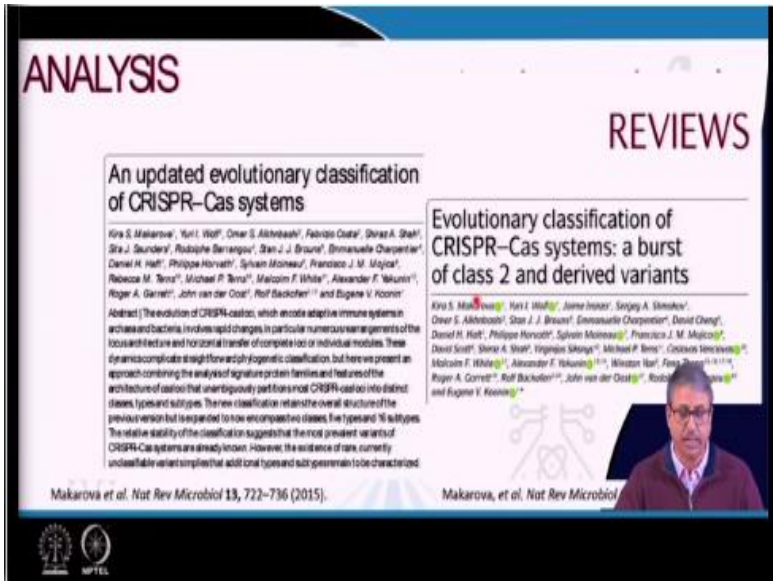
Now, the CRISPR-Cas system, that Cas refers to the CRISPR associated protein system that we are going to discuss very shortly. But as it is the entire system is called CRISPR-Cas system that have been assigned to 2 classes based on the different attributes or characteristic properties into 2 major classes, the class 1 CRISPR and class 2 CRISPR. So, altogether till 2020 scientist have discovered 2 classes of CRISPR-Cas system which together encompasses the six types and 33 subtypes depending on the molecular structure of the entire CRISPR-Cas system.

So, in the class 1 CRISPR-Cas which is basically represented by types I, type III and type IV. They basically recruit a multi-Cas protein complexes for their interference, interferences means for interacting with the invading phage DNA or mobile genetic element DNA. And they stop that invading DNA from function by cleaving it. So, we generally call it as interference. So, they bind with the guide RNA helps the Cas system to bind to the specific location of these invading DNA and then through the complementarity, it is defies the actual location.

There are certain other nucleotide marker motifs are there, so we will talk about those things a little later. So, in class 1, basically class 1 CRISPR-Cas system is characteristically containing a multi-Cas protein complex. So, it is a multimeric protein complexes, so multiple protein complexes are there in the Cas protein. Whereas, the class 2 CRISPR-Cas system, which is basically represented by type II, type V and type VI.

Interference is accomplished by a single effector protein, which is maybe represented by Cas9 more popularly and Cas12 also. So, class 2 is very characteristically found to be having a single effector protein molecule as a Cas molecule, which is the CRISPR associated protein molecule. So, in the class 1, it is multi complex protein, whereas, in class 2 it is a single effector protein and it is most illustrated with respect to Cas, class 2 is the Cas9 and Cas12.

(Refer Slide Time: 11:14)

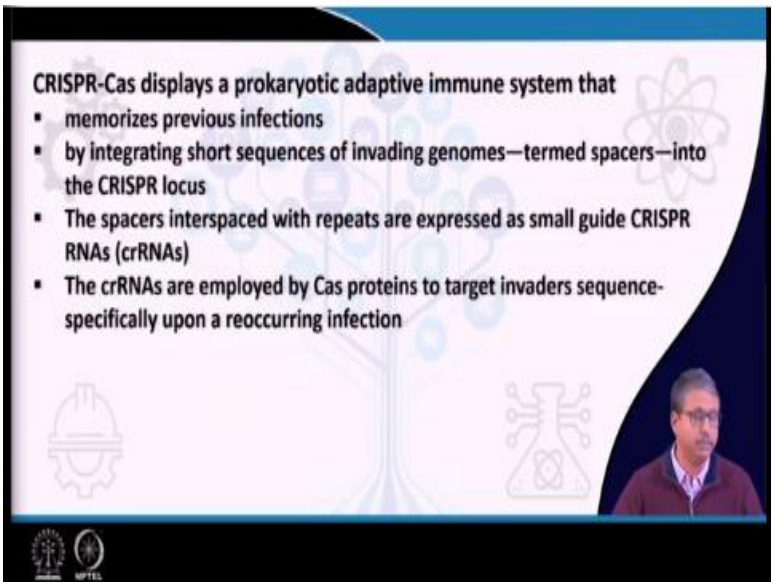


Now, there are a series of papers describing the updated evolutionary classification of this CRISPR-Cas system, which is basically a particular group by Makarova 2 important publications in these respects are an updated evolutionary classification of CRISPR-Cas system in 2015 and then 15 and then in 2020. Another updated publication came which is evolutionary classification of CRISPR-Cas, which covered almost all the 33 subtypes and 6 types.

(Refer Slide Time: 11:49)

Now, it is also very distinctive, that this CRISPR-Cas system is very abundant within the prokaryotic domain particularly within the archaeal genome. So, so far within the all the completed genomes available in the public databases, it is found that almost like 87 to 90% of the completed archaeal genomes are having this CRISPR-Cas system. Whereas, around 50% close to 50% of the completed bacterial genomes, they contain these CRISPR-Cas system.

(Refer Slide Time: 12:24)

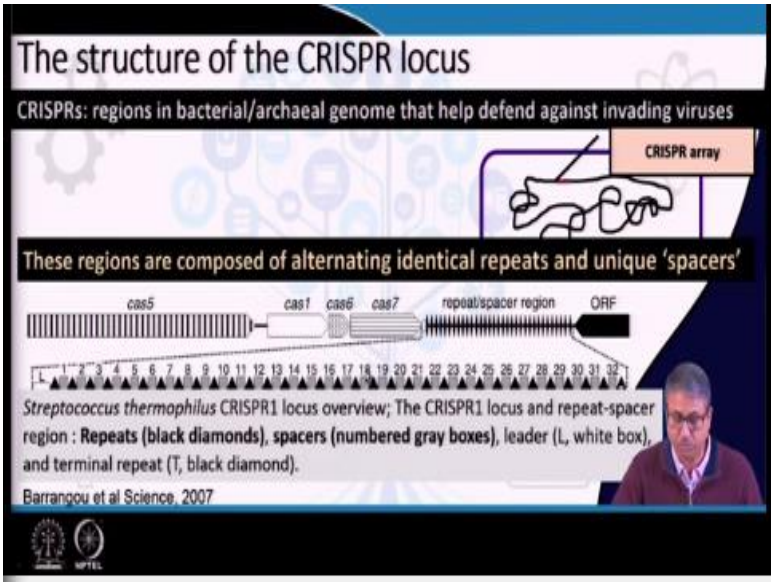


CRISPR-Cas displays a prokaryotic adaptive immune system. That memorizes the previous infections by integrating short sequences of invading genomes termed spacer into the CRISPR locus. So, any invading phage or other DNA elements, which are entering into the cell are converted into small spacer regions and they are integrated within the CRISPR locus. Now, the

spacers which are interspaced with repeats are expressed or transcribed as small guide CRISPR RNA or crRNAs.

And the crRNAs or CRISPR RNAs are employed by Cas proteins in the due step. So, there are steps by in which the Cas proteins are again in engaged Cas system is engaged to use these of CRISPR RNA to target the invader sequences specifically upon a reoccurring infection. So, if there is any second infection or a subsequent infection, the CRISPR-Cas system will be able to identify that this particular DNA was previously they are invaded and we have a spacer for that. And then will bind to the invading DNA molecule and then interfere with it.

(Refer Slide Time: 13:52)



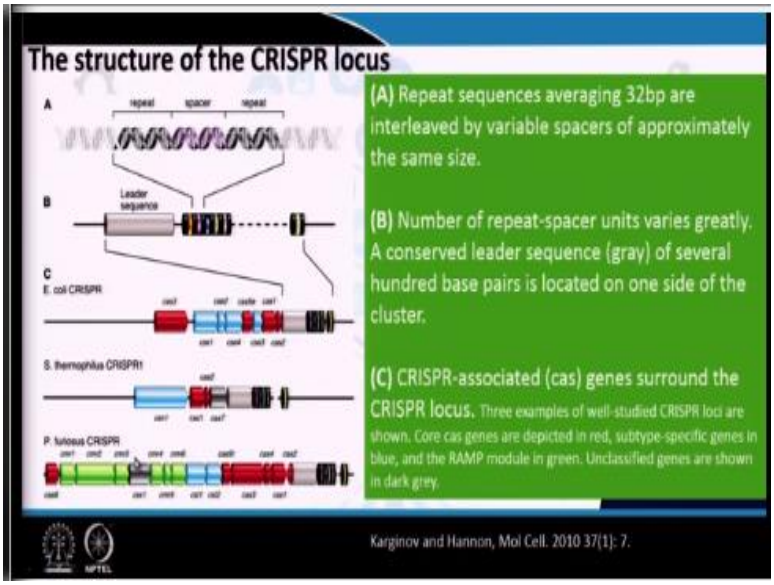
Now the structure of the CRISPR locus which is very important in order to understand the entire functioning of the CRISPR-Cas system. So, CRISPRs are the region in bacteria and archaeal genome that help defend against the invading viruses. So, these are present in the bacterial and archaeal genome and their purpose is to provide a kind of a defense mechanism.

Now, if we look at this simple cartoon where the bacterial cell and we can identify there is a small red coloured portion is there which is marked as the CRISPR array. And if we just enlarge this red colour portion, we would see something like this. So, this entire stretch is the CRISPR-Cas system. Now, if we look carefully, this part is the Cas system where we have the different Cas genes like Cas5, Cas1, Cas6, Cas7 etcetera, genes are there.

And then there is a small white colour box which is basically the leader portion and they we have these repeats. Now, if this repeats which are basically the repeat and spacer region, if these are then enlarged, we will find that there are number of repeats molecule and the repeats are all concepts. So, they have the similar sequences, whether the spacers numbered as 1 to 32 in this case are unique.

So, all these 32 spacers are unique, that means at least 32 different type of infections or infecting or invading DNA containing molecules or pathogens or viruses were encountered by these particular *Streptococcus thermophilus* which whose genome is actually mapped by in this case and shown as an example.

(Refer Slide Time: 15:43)



Now, there are certain features that we are going to highlight which are as below. That these the firstly the repeat sequences, so here are the repeat and spacer and repeat. So, repeat sequences are averaging around 32 nucleotide or 32 base pairs interleaved by variable spacers of approximately the same size. So, these are all short nucleotide region, so you have the repeats and the spacer. So, repeats are all same in sequences, but the spacers are unique sequences.

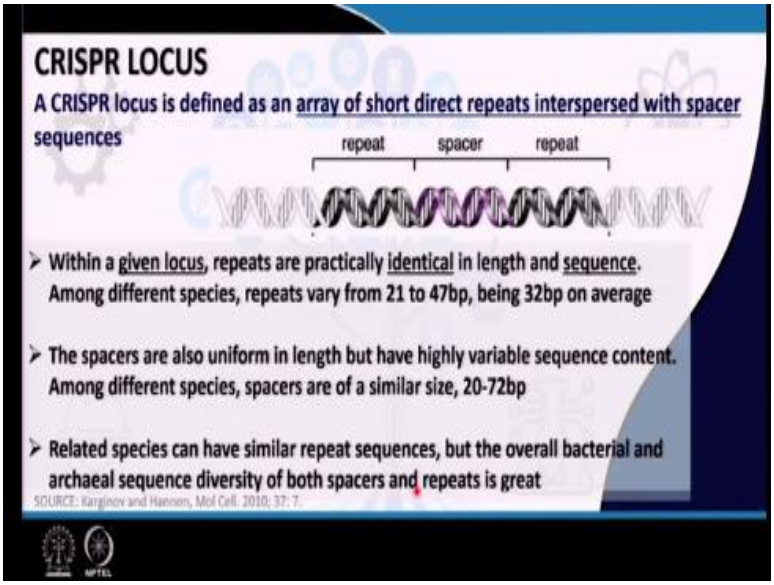
Second is the number of repeat spacer units varies greatly. So, how many repeats and spacers will be there in a particular genome of a particular species? We vary, so a conserved leader

sequence, which is coloured as grey over here. So, this is the leader sequence of several 100 base pairs is located on one side of the cluster. So, if we consider that this is the repeat spacer cluster.

Then on the one side of this cluster we have the leader sequences, which is several base pair, several 100 base pair long. And thirdly, the CRISPR associated genes which are called Cas genes**.** These Cas provides the catalytic characteristics of the entire system. So, the Cas genes are located after this leader, and you can see that there are a number of Cas genes, and in different bacterial systems, bacterial genomes like *E.coli, S.thermophilus and P.furiosus*.

We have actually discovered and numerous other organisms the bacteria and archaea, the entire CRISPR-Cas system has been elucidated quite thoroughly.

(Refer Slide Time: 17:36)



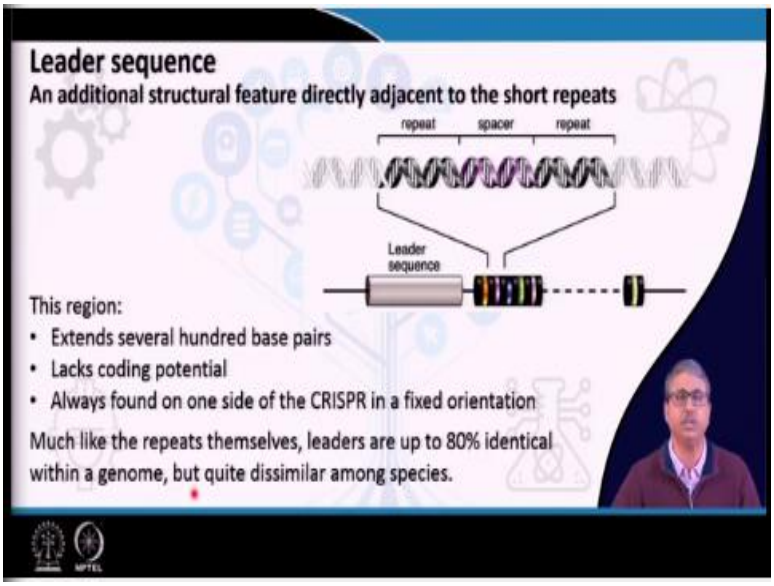
Now, if we want to discuss a bit about the CRISPR locus. This CRISPR locus is defined as an array of short direct repeats interspaced with spacer sequences, so repeats and spacers. Now within a given locus, so in a particular locus repeats are practically identical in length and sequence. So, both length wise and sequence wise the repeats are exactly the same and it varies between 21 base pair to 47 base pair, so far what have been found in different genomes and being 32 base pair as a kind of average.

The spacers are also uniform in length, like average of around 30, 32 and it varies between 20 to 72 base pair. But have highly variable sequence content, highly variable sequence content again because they are all derived from previous infections. And each of these infecting DNA molecules whose spacers are being integrated into case work must be separate, so that is why they are all unique.

So, in just few minutes ago, we were talking about the 32 different spacers which were found. So, related species, phylogenetically related species can have similar repeat sequences. But the overall bacterial and archaeal sequence diversity of both spacers and repeats is found to be great. So, in there are numerous studies which emphasize only on the diversity of these.

So, we have found that within a particular species of a particular organism, the repeat sequence may be conserved may be very much similar, but within other species related species, unrelated species, it varies enormously.

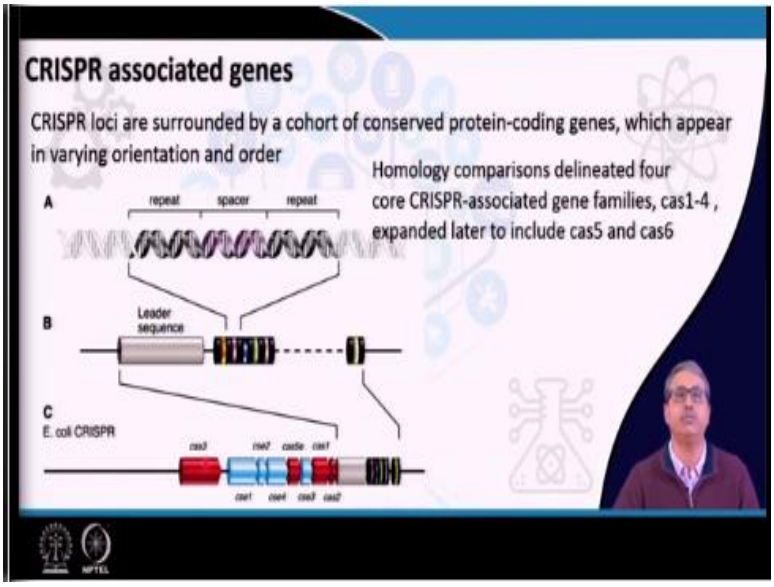
(Refer Slide Time: 19:31)



Now, the leader sequence is an additional structural feature, which is directly adjacent to the short repeat. So, you have the repeats and adjacent to the repeats, we have the repeats and spacer and adjacent to that we have the leader sequence, and this is extended several 100 base pairs and lacks a coding potential, so it does not code. And always found on one side of the CRISPR in a fixed orientation, it is not on the other side.

And much like to repeat themselves the leaders are up to 80% identical within a particular genome, but quite dissimilar among (()) (20:08) other species. So, within across the species they are different.

(Refer Slide Time: 20:11)

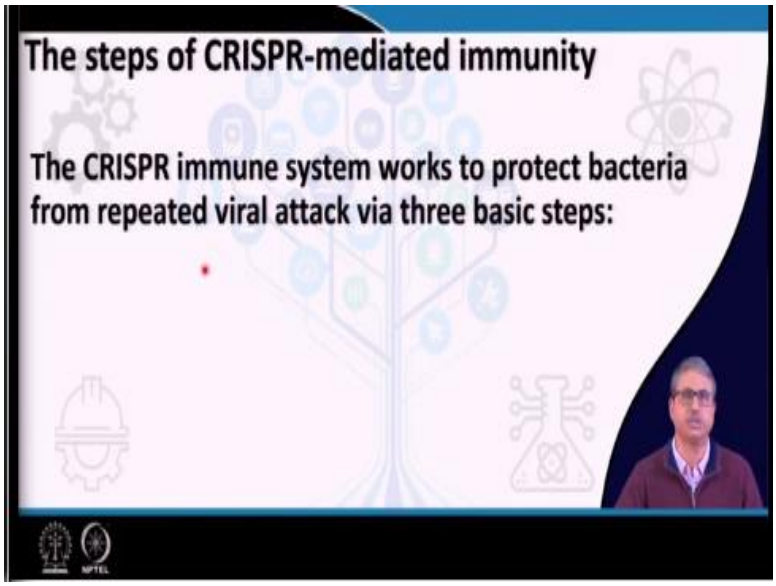


Now, the most important and most relevant component of the CRISPR-Cas system, which is enabling the catalytic function of the entire system. So, this is by virtue is called Cas system, which is CRISPR associated system. So, encoding proteins, genes, protein coding genes. So, CRISPR loci are surrounded by a cohort of conserved protein coding genes, so these proteins are highly conserved.

So, as we can see this is the repeat spacer location followed by you have the leader and then starts the different Cas genes, so this is the E.coli Cas genes. So, homology comparisons delineated four core CRISPR associated gene families Cas 1 to 4, which is expanded later into include Cas5 and Cas6 as well. So, we have Cas 1, 2, 3, 4, 5 and 6, so different type of Cas families are there, protein families are there.

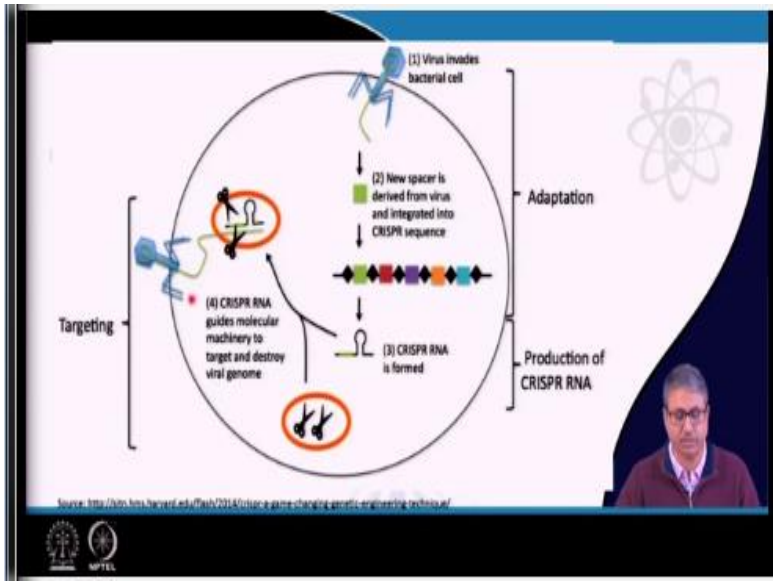
So, they are responsible for a number of catalytic function and the entire process of CRISPR-Cas mediated immunity is actually achieved through these catalytic functions of the Cas system.

(Refer Slide Time: 21:27)



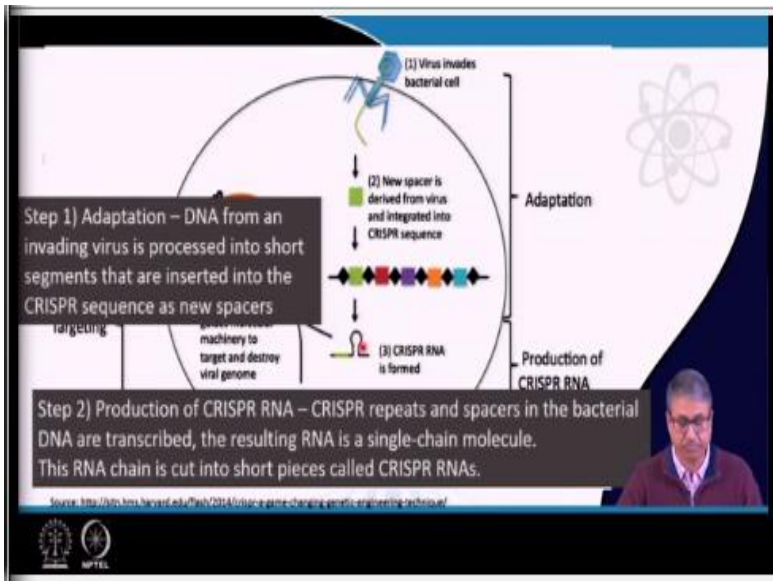
Now, the steps of the CRISPR mediated immunity. The CRISPR immune system works to protect bacteria or the archaea from repeated viral attack via three basic steps.

(Refer Slide Time: 21:40)



So, these steps are well illustrated like the adaptation, the production of the CRISPR RNA and then interference, which is basically called as the targeting and then achieving the interference.

(Refer Slide Time: 21:56)



So, as we can see over here, the adaptation which is again all are catalyzed by the Cas system. So, different catalytic domains are responsible for this entire processes. So, the first process is the adaptation where a new introduced or infecting virus, the virus DNA is processed to obtain the spatial molecule, this is catalyzed by Cas system. And then these small unit which is again the called as a spacer is going to be introduced into existing CRISPR.

So, already some part of the CRISPR is there in the genome. So, the newly invaded spacer part which is obtained from the infecting viral DNA is introduced, so this is called adaptation. So, DNA from an invading virus is processed into short segments that are inserted into the CRISPR sequence as a new spacer, so new spacers are added like that. So, now we have in this case, we can see in this cartoon five spacers.

So, in this next round of infection, if a new virus infects this particular bacterial cell, then another spacer will be produced and that spacer will be added. Provided the DNA is actually identified to be a new DNA. So, if existing known pathogens or known viruses are infected, then possibly the spacer is already there. So, the system will identify through their complimentary recognition system.

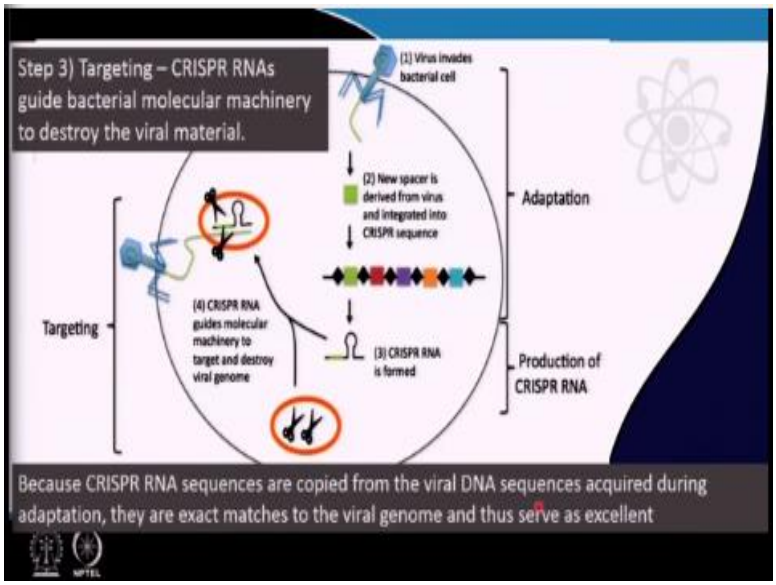
Next is the production of the CRISPR RNA. So, we have this CRISPR system which is if we assume that the entire spacer and repeat combination is existing. Now there is a fresh infection or

as soon as there is an infection that the CRISPR array is going to be transcribed normally it is transcribed. And so, the after the transcription of this entire portion, the CRISPR RNA is processed.

And finally the processed CRISPR RNA which we can see the entire spacer, a individual spacer molecule and the palindromic repeats structure is there. So, this is produced out of this CRISPR array, entire CRISPR array. So, if we have like a 5 spacers within this, so we can expect that five different CRISPR RNA will be formed. So, each will have a single specific spacer sequence, along with that the repeat sequence which is forming.

Basically the repeats are palindromic, so we will have this loop structure. Now, the CRISPR repeats and spacers in the bacterial DNA are transcribed and resulting RNA in a single chain molecule. And these RNA chain is cut into short pieces called CRISPR RNA, again the Cas system is responsible for carrying out the catalytic processes which are required for the, so from converting to the original CRISPR RNA which is transcribed. From that to produce the individual small the single guide RNAs are actually produced.

(Refer Slide Time: 25:04)



And now the third step, which is called targeting, which facilitates the interference basically by cleaving the most of the time, cleaving the invading DNA. Now the CRISPR RNA, which is already produced will guide the molecular machinery, again the Cas system. Now this Cas will

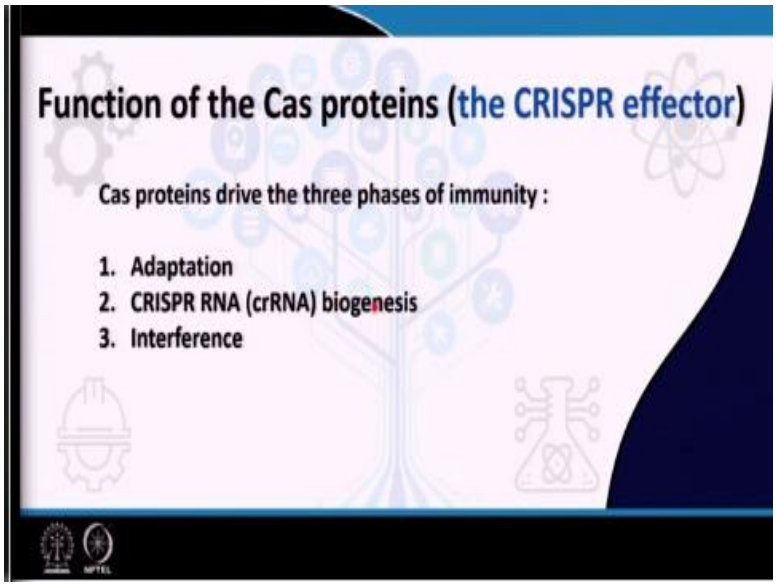
have a very significant nuclease activity, and very specific nuclease activity which can cut the specific region by introducing double stranded break.

So, this CRISPR RNA which is also called the guide RNA, which will be associated with the Cas system anyway because Cas is all the time associated with this in order to convert the main transcribed product into these the single molecule structure. So, the Cas is already associated with this. Now the Cas, the CRISPR RNA will guide it to the appropriate location of the invading virus.

And then allow the Cas system to cleave the appropriate region at the target site by introducing double stranded break. So, these three steps are very distinguishing which are all catalyzed by the Cas system. So, Cass has multiple catalytic domains and catalytic units. So, it facilitates the entire process of CRISPR mediated immunity, so that is how we call the entire process as CRISPR-Cas system.

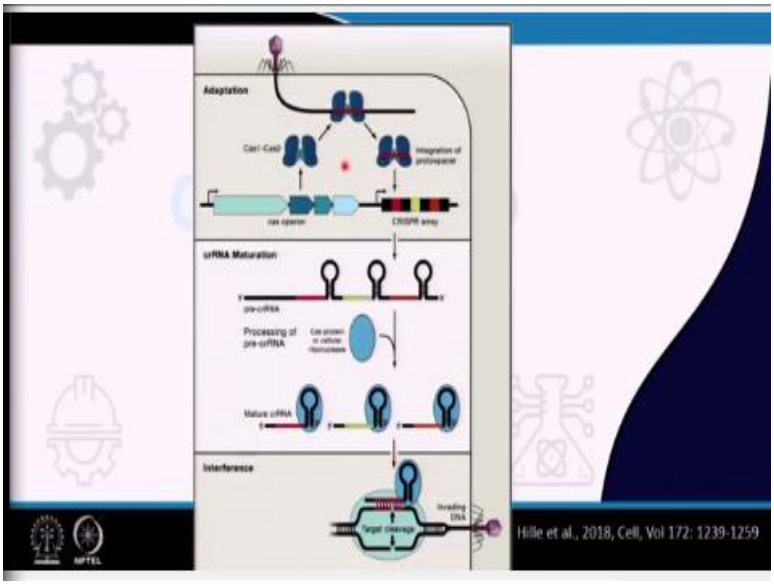
Now because the CRISPR RNA sequences are copied from viral DNA sequence. So, originally building the prior infections, the viral DNA sequences are already copied and they are stored as spacers. So, they are stored acquired during the adaptation process, they are exact match to the viral genome and thus serve as excellent tool marker for identifying the specific viral DNA, that it is invading virus and previously it was also invaded and the system has it is spacer.

(Refer Slide Time: 27:10)



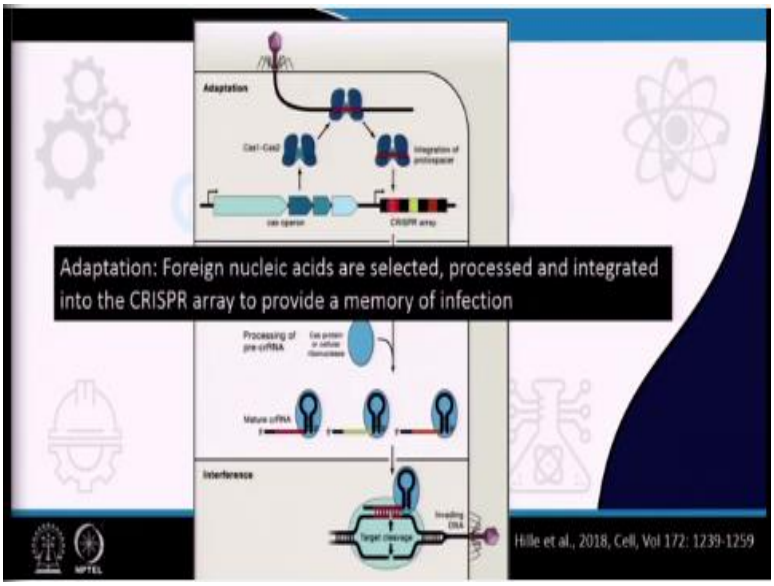
Now the three distinct function that we can identify are basically the adaptation, CRISPR RNA biogenesis. That is the transcription and followed by the formation of the single stranded guide RNA, single guide RNA formation and then achieving the interference.

(Refer Slide Time: 27:27)



Now if we see this entire process very carefully, so as I mentioned earlier that it is the Cas system, which is catalytically enabling the entire process starting from the adaptation to the interference to achieve. It encompasses a number of catalytic activities.

(Refer Slide Time: 27:49)

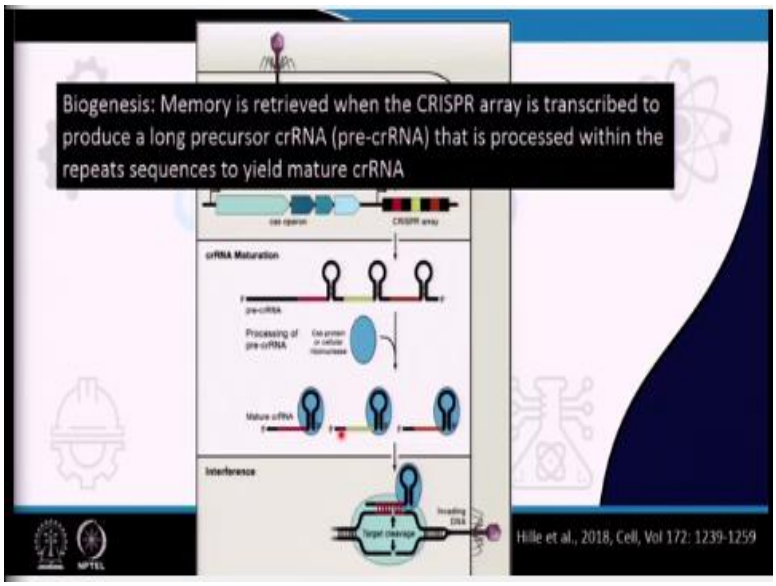


Like as we can see in the adaptation phase, the RNA foreign DNA molecule or the nucleic acids are selected the processed and integrated into the CRISPR array to provide a memory of infection. So, the Cas is expressed in the cell is a kind of a constitutive expression is there of this Cas1, Cas2. So, these Cas enzyme can actually identify if I invading a foreign DNA molecule.

And then utilize this particular short region that is very specifically identified by virtue of a motif that we will be discussing very shortly. So, it will identify that this is in the kind of invading pathogen and the DNA will be cleaved and then it will be integrated into the existing CRISPR system. So, the portion that is actually used by this Cas to cut out of the invading viral DNA is called protospacer.

So, protospacer is basically the region which is actually a cut and converted into cleaved and converted to spacer to be integrated into the CRISPR array.

(Refer Slide Time: 28:59)

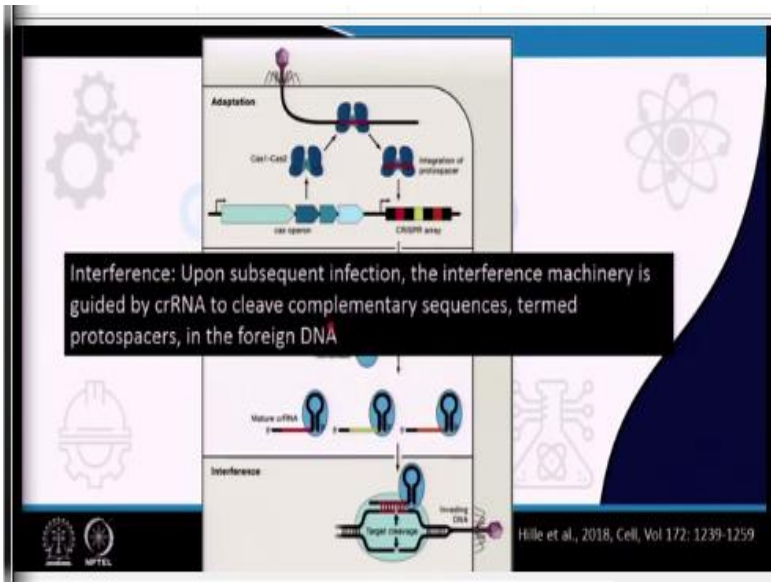


Now, the second part of this Cas system, once we have the CRISPR array is the biogenesis. So, in the biogenesis part the memory of this entire invading pathogens, invading viruses is retrieved when the CRISPR array is transcribed to produce a long precursor CRISPR RNA. So, CRISPR RNA is the long, suppose we have 32 spacers and then all the repeats together. So, it will be a long chain of RNA in which is called precursor CRISPR RNA.

And then this precursor CRISPR RNA or pre CRISPR RNA, which will be processed within the repeat sequences to yield the mature CRISPR RNA. Now, this is the precursor CRISPR RNA which is produced out of the transcription and this entire piece is subjected to the again the Cas activity. So, Cas would like to or we would identify each of the spacers and will cleave these into short pieces which are called processing of the pre crRNA or CRISPR RNA, which by cellular ribonuclease activities.

And then the mature CRISPR RNA, which will be produced. So, each of these mature CRISPR RNA which is coming out of this precursor CRISPR RNA will have a unique spacer region.

(Refer Slide Time: 30:26)

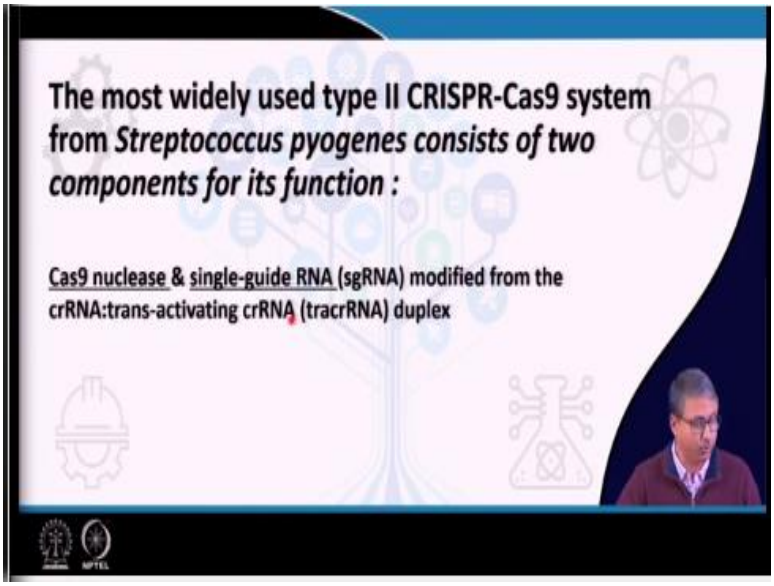


Now we are ready for the third interfering step. Now, upon subsequent infection, the interference machinery, that is the Cas system again, the Cas is already associated with the entire process. So, the Cas associated with this guide RNA, which is the spacer bound with the palindromic repeats structure is guided by this structure to cleave the complementary sequences within the invading DNA of virus or plasmid.

And this complimentary region within the invading DNA is called protospacer. So, any invading DNA will be searched for the presence of protospacer. Because protospacers are the regions which are complimentary to these spacers and these spacers are already over here. So, if in an foreign DNA protospacers identified, then the spacer will engage the Cas system to cleave or to introduce a double stranded break within it.

And which is clearly shown over here, that the protospacer region will be identified and we will be seeing shortly that there are actually other regions a particular region is a PAM region, which allows these single stranded guide RNA to identify this complimentary region perfectly. And then the cleavage is in double stranded cleavage, the double stranded break is introduced by the Cas system.

(Refer Slide Time: 32:00)



Now, class 2 CRISPR-Cas systems, we earlier discussed that there are 2 main classes class 1 and class 2. And class 2 is basically this large single effector molecule is there unlike the class 1. So, class 2 CRISPR-Cas system which is basically represented by CRISPR, Cas9 and CRISPR-Cas12 have been used widely for genome engineering. And CRISPR-Cas9 particularly which is originally found to be acting as a Cas9 CRISPR RNA.

And the strands acting RNA pattern is most extensively used. Now, the most widely used type II CRISPR-Cas9 system is from the Streptococcus pyogenes which is consist of 2 components for it is function. So, it is the CRISPR-Cas9, which is a type 2 system from Streptococcus pyogenes, which is extensively studied and found to be one of the best suited for genome engineering.

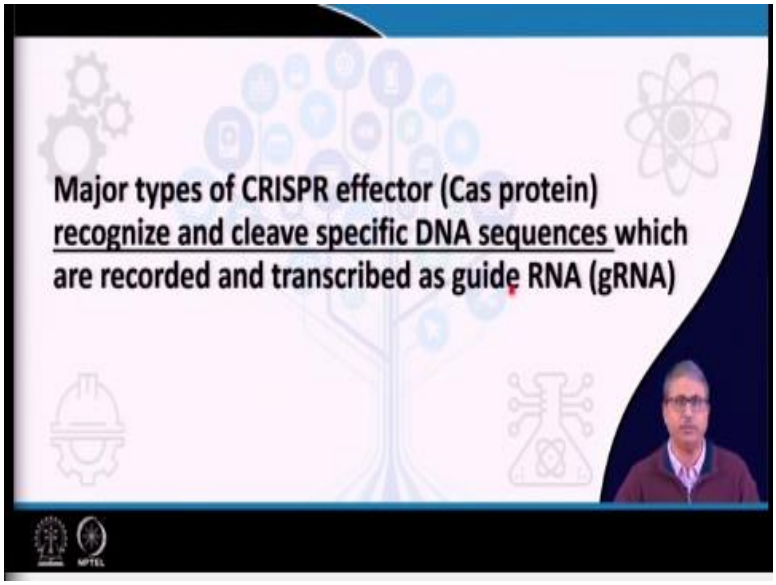
And let us see what are the components of this type II CRISPR-Cas9 system. So, as we mentioned type II, it is obviously kind of an having a single effector molecule as the Cas9 is the major effector molecule. So, it will have what it consists of the Cas9 nuclease which is the catalytically active portion of the enzyme or active characteristic property of the enzyme that is the nuclease Cas9.

And single guide RNA, now this single guide RNA which is modified from the CRISPR RNA and the trans activating CRISPR RNA duplex. So, the CRISPR RNA is produced out of the processing of the precursor CRISPR RNA, we have seen that. And trans activating CRISPR

RNA is a short piece of RNA which is also produced transcribed from a nearby locus and which is available for complimentary binding to this CRISPR RNA.

Forming the entire structure forms the single standard RNA guide RNA structure. So, guide RNA is having 2 components one is the CRISPR RNA having a very distinct spacer and the palindromic repeat part. And with this part, we have the trans activating RNA which is the trans crRNA forming a duplex structure.

(Refer Slide Time: 34:36)

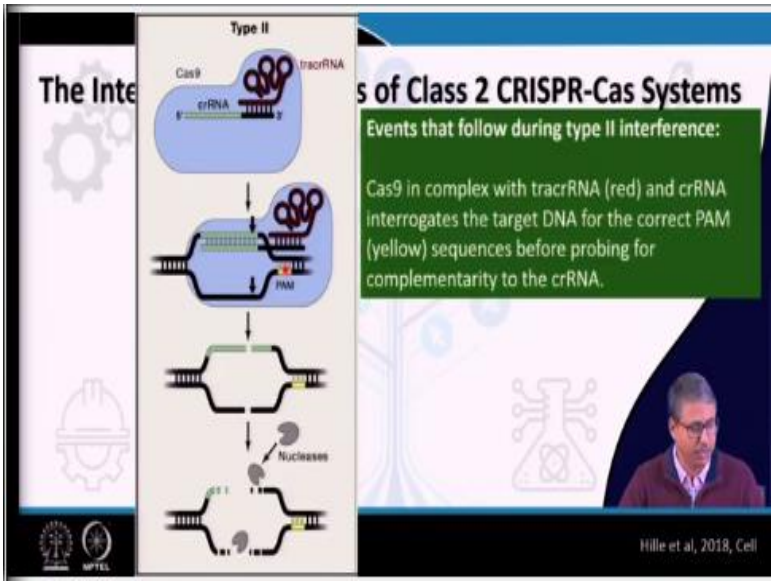


Now, major types of CRISPR effector that is the Cas protein recognize and cleave the specific DNA sequences which are recorded and transcribed as guide RNA. So, the guide RNA is responsible for guiding the nuclease which is the Cas system to identify the appropriate target site which is the protospacer within the protospacer it will identify. And will enable the Cas enzyme or nuclease of the Cas to introduce the double stranded break within it.

(Refer Slide Time: 35:10)

Now the interference pathway of class 2 CRISPR-Cas system is well elucidated.

(Refer Slide Time: 35:15)



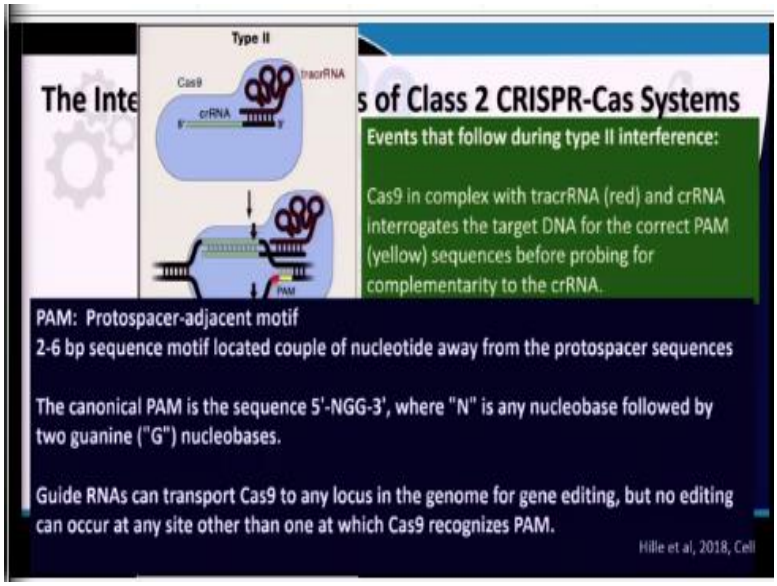
And if we see that within this, we have a very distinct components, which are basically the blue colour structure is the Cas system and we have these CRISPR RNA, which is the mature CRISPR RNA. And this is the trans activating RNA, which is complimentary to the part of this repeat part. And the entire 2 components duplex component like the CRISPR RNA and these are trans activating RNA together.

This is the duplex structure which activates or enables the Cas enzyme system to execute it is a nuclease activity very specifically. So, it can, Cas can now introduce a very specific double

stranded break. So, events that follow during the time to interference or the Cas9 in complexes with the trans activating RNA which is red in colour and the CRISPR RNA interrogate the target DNA.

So, it scans the target DNA, that is the invading phage DNA or the plasmid DNA is thoroughly checked by the Cas system. The Cas is the enzyme complex, so Cas will not respond unless and until the signal comes from these duplex which is the CRISPR RNA and trans activating CRISPR RNA. So, this duplex structure allows the Cas and it interrogates the entire target DNA for the correct PAM region. So, what is this PAM region? PAM region as it is marked over here.

(Refer Slide Time: 36:49)



It is a very specific protospacer and just in motif, we call it. So, any region which is within the invading DNA which is complimentary to the spacer is called the protospacer. So, the PAM region is a very short sequence motif 2 to 6 base pair long sequence motif, which is located only couple of nucleotide away from the protospacer sequence. So, if this is the protospacer, this green line is the protospacer.

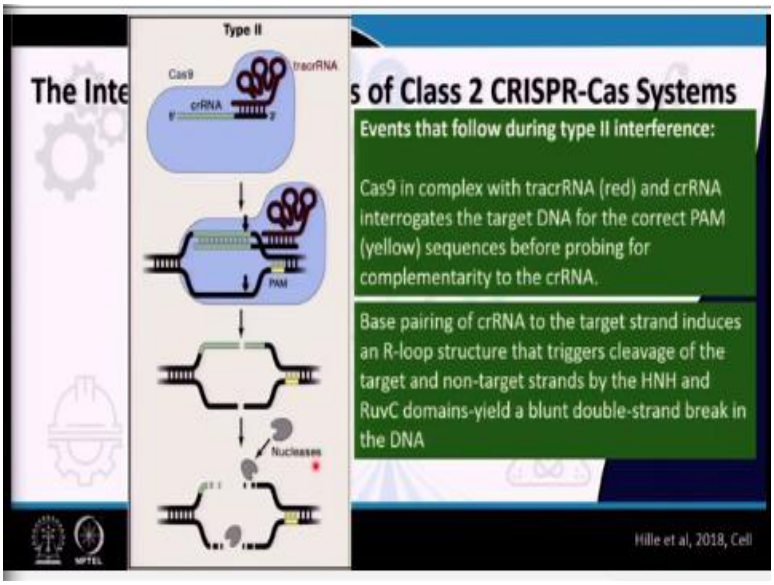
Then you can see that only few base away basis away here we have these protospacer adjacent motif. So, the entire duplex, this duplex which is the CRISPR RNA and the trans activating CRISPR RNA, this duplex will scan the entire invading DNA molecule to identify the PAM.

Now, this PAM is basically a canonical sequence content which is having NGG kind of sequence where N is any nucleobase followed by 2 guanine nucleobases.

So, the guide RNA these complex, the guide RNA can transport the Cas to any locus in the genome for gene editing. So, if we have the Cas9, which is catalytically the nuclease and we have the guide RNA. Then the guide RNA can take the Cas to any locus within the genome even for genome editing purpose also. But no editing can occur at any site other than one at which the Cas9 recognizes the PAM.

So, the PAM is so important, that unless the PAM is present and it is recognized the double stranded break or any kind of other activities or editing will not be achieved till the PAM is detected.

(Refer Slide Time: 38:46)

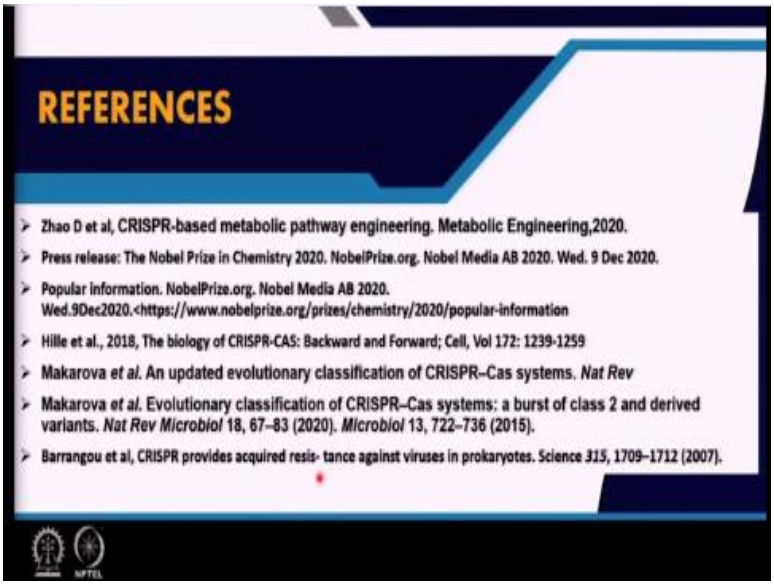


Now the base pairing of CRISPR RNA to the target strand which is the basically the spacer complimentary region of the spacer over there, induces an R-loop structure that triggers cleavage of the target and non target strands by 2 domains which are HNH and RuvC domains. And that these 2 domains are finally enable the entire system to yield a blunt double stranded break in the DNA.

So, you can see that the entire thing is recognized due to the PAM and then the protospacer to space or complementarity is achieved. And then finally, the Cas or nucleases are able to cleave these 2 by using 2 distinct domains of this enzyme. And then finally, the double stranded breaks are introduced. So, this double stranded break formation is the very beginning of the entire genome editing process.

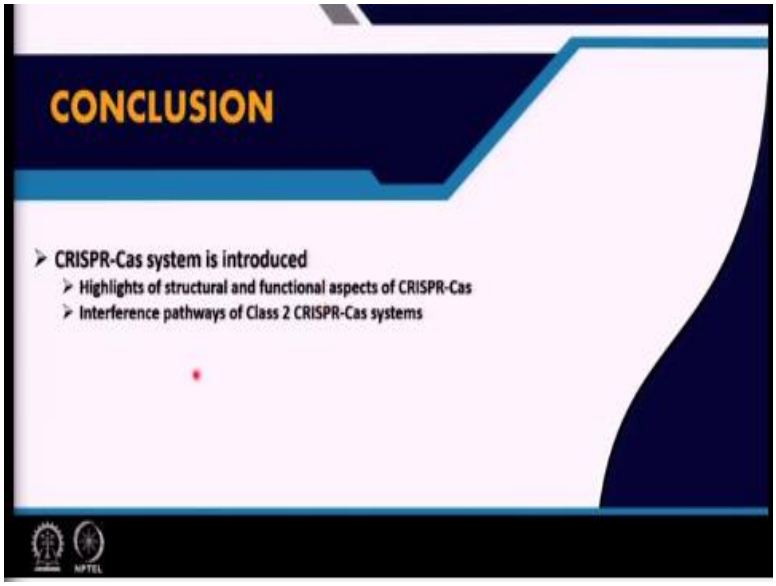
Because we can engineer that we are going to discuss maybe in other lectures that how the entire system can be engineered to identify a particular gene of interest or a particular location of interest or site of interest, and then can be utilized to introduce specific break, double stranded break within it.

(Refer Slide Time: 40:09)



So, in this portion we have used the following references and including 2 very important notes which are basically the press release from the Nobel prize in Chemistry 2020 and also the popular information released by the Nobel media.

(Refer Slide Time: 40:27)



And in conclusion the CRISPR-Cas system is introduced; we have highlighted the structural and functional aspects of the CRISPR-Cas system. And interference pathways of CRISPR-Cas system, Cas9 system of class 2 is also presented, thank you.