Course Name: Basics of Crop Breeding and Plant Biotechnology

Professor Name: Dr. Joydeep Banerjee

Department Name: Agricultural and Food Engineering

Institute Name: Indian Institute of Technology Kharagpur

Week: 01

Lecture-04: Concept of Gene and Experiments on Plant Hybridization

Hello everybody. Today, we will start discussing module number 1. So, in module number 1, we will discuss the Concepts of Genes and Mendelian Genetics. Today, we will discuss the concept of genes and some experiments on plant hybridization. So, the ideas that will be covered in this particular area, i.e., first, we will discuss chromosome packaging, then the definition of a gene will be discussed here, then what are the basic structure of DNA and RNA will be discussed, then we will go through the central dogma. Finally, in brief, we will discuss the replication, transcription, and translation process.

So, in the future, in this course, we will be able to understand different things easily. Next, we will discuss the structure of a eukaryotic gene. In the course, we will mainly be discussing the plant system. So, an understanding of the eukaryotic gene structure is needed.

So first, we are discussing the concept of genes. So, to understand the idea of a gene, we first need to know how the genes are available in the prokaryotic and eukaryotic systems. In prokaryotic system, defined chromosome are there. So, on the naked DNA, different genes are available. While in case of eukaryotic system, the genes are available within the chromosome, mostly within the chromosome.
So, we will discuss those chromosomes' segregation and those things, earlier also we have discussed. So, in the chromosome how the genes are packed that thing we need to know. Well, so now, we will discuss how the double-stranded DNA is packed inside of a chromosome, and the chromosomes are mostly available in eukaryotic system. So, how the DNA is tightly packed over there let us see that. So, this is the double-stranded DNA, it’s

diameter is close to 2 nanometer, it will be having a 5' to 3' strand and another strand will be 3' to 5' strand.

So, initially this double-stranded DNA is packed within the nucleosome. So, what is available? The chromatin at the simplest level is the double-stranded helical structure of DNA. Its diameter is close to 2 nanometer and it is wrapped within a nucleosome. So, what is available within this nucleosome? So, if you see carefully, here 4 types of histone proteins are available H_2A, H_2B, H_3 and H_4. 4 types of histone proteins are available in two copies.

So, a histone octamer is formed, and around this histone octamer, the double-stranded DNA is wrapped in this way. As you can see, the white color double-stranded DNA is wrapping around this particular histone octamer. And H_1 is another histone protein that basically seals this nucleosome. This is the entry part of the DNA and this is the exit part of the DNA. These two parts are sealed by H_1 protein.

So, basically these things are available within the nucleosome and in this way, the double-stranded DNA is packed using different nucleosome structure, in this way. So, over here, the nucleosome molecules are available, over here nucleosome molecules are available in this way the double-stranded DNA is packed. So, its diameter is close to 11 nanometers. So, nucleosome core having 8 histone molecules around which the double-stranded DNA wraps 1.6 times and H_1 seals two DNA ends.

So, what is the further level of chromosome packaging? It has been found that these histone molecules which is available as histone octamer, they basically form a cylindrical structure. Ok! Means if you see this cylindrical structure over this, is a particular nucleosome; this is another nucleosome. Ok! In this way, in each of this nucleosome, histone octamers are available, H_1 is sealing the DNA incoming and DNA outgoing strands and the DNA is wrapped around it. And those nucleosome molecules, basically, they are found to be available in this structure,
so that they make a barrel like structure where the DNA become further constricted or the condensation becomes more within the DNA. So, the structure basically looks like in this

way and its diameter is 30 nanometers. The diameter of this particular coiled- like structure. Ok! Where 6 different histone, 6 different nucleosome molecules are available and the stack of nucleosomes are present over here. So, let us see the further level of condensation. So, over here, 6 nucleosomes further folds into 30 nanometers fibre what we have discussed already, and thereafter those coiled things, those cylindrical structure basically folded in this way.

It is folded in this way and if you see, just this particular part over here, in this way, you can see this barrel- like structure. Ok! This is the further level of condensation. So, let us see what is the diameter, here, diameter is close to 300 nanometers of this fold like structures. Each loop contains super-coiled DNA having 300 nanometers diameter and thereafter, also this super-coiled DNA is further coiled. So, the further coiling take place like this, this coiled like structure is further coiled and its diameter is close to 700 nanometers.

And in this way, the super-coiled DNA condensed in chromatid having diameter of 700 nanometers and i.e., mostly available within the chromosome. We know that in a particular chromosome, 2 chromatids are available, this is a particular chromatid and this is another chromatid. So, in each of this chromatid, this 700 nanometers coiled DNA is available and here from you can know that from a single double-stranded DNA, how DNA is packed around initially across the nucleosome molecule thereafter the nucleosomes are stacked together to make a 30 nanometers fibre. Then 30 nanometers fibre is further coiled into 300 nanometers folding, it's again coiled into 700 nanometers folding and finally, it is available in each chromatid available within a chromosome. So, now we would like to discuss what is a gene? So, at genetic point of view, gene is considered as the basic unit of inheritance. Ok!

We know that the DNA is transferred from one generation to another generation. So, in Mendelian genetics also we have seen that. So, within the DNA, different small units are there those are considered as gene. Ok! It means, it is a basic unit i.e., transferred from one generation to next generation. And basically, it is a part of DNA that mostly determines a

single trait such as leaf color, but sometimes few genes can control multiple traits also.

So, in this way, if you see the double-stranded DNA is available. So, the double-stranded DNA will be having a 5' to 3' end and another end will be 3' to 5' end. So, if you see about this particular DNA little bit in detail, you can see different bases are available. Ok! Let us assume this is the 5' to 3' direction and this is another strand of the DNA. In 5' to 3' A G T C G C A T A G A G C T it is available on one strand.

While in its opposite strand the complementary DNA are available. In DNA mostly A pairs with T, G pairs with C. Ok! So, in this way, the opposite strand of the DNA will be like this. And eventually from each and every gene we can see some particular product, later on, we will be coming to that and some of the gene will show certain features like here, if this gene is transcribed suppose, we can see the leaf color. So now, let us define a gene at molecular level.

How can we define a gene at molecular level? So, a gene is a stretch of nucleotides, earlier we have seen A T G C... different sequences were there. So, it is a stretch of nucleotides which codes for a functional product. Ok! So, this is very important point functional product. Ok! It may be protein or RNA. So, you guys should recall it because not all the genes code for proteins some genes are there, they do not code for any protein.

So, it is a DNA sequence, therefrom, in most of the genes we can see mRNA formation. mRNA is messenger RNA and if mRNA is formed, therefrom through translation process finally, we can see the polypeptide formation. And if polypeptide is formed, thereafter these polypeptides, it means, it folded properly to make a particular protein. While some genes are there, they code for tRNA i.e., transfer RNA, rRNA i.e., ribosomal RNA, snRNA, small nuclear RNA and microRNA they do not code for any protein. Ok! So, these are protein coding genes, they basically make mRNA while the non-protein coding genes, they make these things. Ok!

So, in this way, any gene can be a stretch of nucleotide which codes for a functional

product, it may be protein or RNA. Now, let us discuss about the nucleotides. Ok! Because DNA stands for deoxyribonucleotides and RNA stands for ribonucleotides i.e., known to us. So, in nucleotides, basically, it is composed of three different things it contains pentose sugar means 5 carbon sugar, it contains nitrogenous base and a phosphate group. So, this is one nitrogenous base, ribose sugar, sorry this is a ribose sugar a 5-carbon pentose sugar i.e., available on RNA. While, it is another pentose sugar i.e., deoxyribose it is available on DNA.

If you see over here, in deoxy the oxygen group is not available here, oxygen group has been removed. So, in DNA it is available as deoxyribose sugar while in case of RNA at this position at the 2' position the OH is available. Now, what are the different nitrogenous base available on DNA and RNA let us see. So, basically, two types of nitrogenous base are available, either they belong to purine group means two heterocyclic compounds are there or it may be pyrimidine group. So, in purine, two types of purines are available adenine and guanine, both are available on DNA as well as RNA.

So, in purine, we are having adenine and guanine, both are available on DNA as well as RNA. Within pyrimidine three bases are available in pyrimidine i.e., cytosine, thymine and uracil. So, out of these three cytosine and thymine they are available on DNA while in RNA no thymine is there, in-spite of thymine the uracil is available. So, in RNA we will be having cytosine and uracil, while in case of DNA, cytosine and thymine are available among the pyrimidine base. So, let us see the structure of DNA and RNA.

So, we know that DNA is mostly available in double-stranded nature and it has two polarity, one end will be having 5' and another end will be having 3' end. So, in 5' end of the DNA, the phosphate group is present, you can see a phosphate group is available over here in the 5'end of the DNA, phosphate group is available. So, in nucleotide if you just recall, what is available in nucleotides? There, a phosphate group will be available, a pentose sugar will be available and a nitrogenous base will be available. So, this is the pentose sugar i.e., available on the DNA, it is deoxyribose in nature. So, no oxygen is available and this one is the nitrogen base, nitrogenous base available on the DNA.

So, on the DNA in this way, different nucleotides stay together and once the next nucleotide comes, a phosphate group basically form a bond between this CH_2OH and this OH available in the 5-carbon sugar and a hydrogen molecule and a water molecule is released. And finally, a phosphodiester bond formation is taken place. In this way, different nucleotides join together on the DNA. So, in this way, a single DNA strand is formed and eventually another DNA strand which will be available in a different polarity means 5' in this, in another strand the 3' will be there close to the 5' end. Ok! So, one strand will be 5' to 3' and another strand will be 3' to 5' in this way double-stranded DNAs are available. And if you see in between these two strand the hydrogen bond formation is taken place in between A and T these two nitrogenous bases, two hydrogen bonds are formed while in between G and C, three hydrogen bonds are formed.

So, this bond is more stronger, the GC bond is more stronger compared to the AT bonds. Ok! So, in this way the double-stranded DNA is available in most of the organisms while the RNA is mostly in single stranded nature, it is mostly in single stranded nature. So, here also, nucleotides are available and here you can see the OH group is available in-spite of H group which is available in DNA. So, in this way the single stranded RNA is also available and RNA has also a polarity in 5' to 3' end. Ok! Let us start our discussion on central dogma. Ok!

What is central dogma? In central dogma, basically how DNA to DNA is formed, DNA to RNA is formed, RNA to DNA is formed those things are basically discussed. So, let us start with a single double-stranded DNA we have a double-stranded DNA one strand is 5' to 3' and another strand is 3' to 5'. So, we know that in double-stranded DNA, A pairs with T with two hydrogen bonds; while G pairs with C with three hydrogen bonds. So, for here if we know the sequence of a particular strand the 5' to 3' strand, then we can easily tell what is the sequence available in another strand. So, over here the sequence will be T, it will be A, it will be C, it will be G, it will be G T A C G C… and so on.

So, in central dogma, first thing we will be discussing the DNA replication. DNA

replication means DNA to DNA formation. So, through DNA replication what is the DNA available in a particular cell, it is replicated, it is multiplied and during cell division one segment of DNA go to the parental, to the offspring cell. So, during mitosis this type of DNA replication is taken place in each and every cells. So, this is done by a particular enzyme known as DNA polymerase in eukaryotes, as well as in prokaryotes, the DNA polymerases are available.

So, the next thing in central dogma is transcription. Transcription is the process through which from DNA, RNA is formed. Ok! It is the process of RNA formation DNA to RNA is formed, through transcription, the mRNA, rRNA, tRNA all types of RNAs are formed. So, this is done by a particular enzyme i.e., known as RNA polymerase. So, in this way, once the transcription is taken place then within the DNA strand one strand acts as a coding strand, basically, the 5' to 3' strand acts as a coding DNA strand while the 3' to 5' strand act as a template DNA strand.

So, using the template DNA strand the RNA polymerase basically synthesize new RNA, it synthesizes new RNA. So, the new RNA what will be produced? It will be produced, it will be similar to the coding sequence, similar to the coding DNA strand except in case of T there will be uracil residue, in case of thymine there will be uracil residue. So, if from here the transcription is taken place, the sequence will be A, U, G, C in this way. So, this is done by RNA polymerase. Now, another process is there through which we can prepare DNA from RNA i.e., known as the reverse transcription process, this is just opposite to the transcription.

Through reverse transcription RNA, basically mRNA is converted into cDNA the mRNA is converted into cDNA through reverse transcription process and the enzyme involved over here is reverse transcriptase. Then in some of the virus, we can see RNA to RNA formation i.e., replication RNA to RNA formation is available in some of the virus and i.e., done by the enzyme RNA dependent RNA polymerase or RdRp and in this way, from plus (+) strand of RNA i.e., sense strand of RNA, the antisense strand is formed and vice versa, from antisense RNA also sense strand could be formed. So, this is done by RdRp enzyme.

Then another thing available in central dogma, i.e., the translation, it is the process through which from mRNA, protein is formed, from mRNA, protein is formed through translation process. Ok! And based on the codon available on the mRNA different polypeptides gradually will be formed on the protein and finally, the polypeptide chains will be formed.

So, this is the overall central dogma, the DNA replication is there DNA to DNA formation, then transcription i.e., DNA to RNA formation; then reverse transcription i.e., RNA to DNA formation, basically, mRNA to cDNA formation; then RNA to RNA formation is done by RdRp, RNA dependent RNA polymerase and translation is there, in translation ribosome is involved. So, now, briefly we will discuss about the DNA replication. So, it is the process by which cell doubles its DNA before cell division. If you recall the mitosis and meiosis those things, there at the end of the mitosis, at the end of the meiosis, we are having either two cells, at the end of the mitosis, from one cell we are getting two cells; in each cell, single chromatid of each chromosome is available right? So, those chromatids have, each chromatid is having single double-stranded DNA.

Now, through DNA replication, those single double-stranded DNA is converted into two double-stranded DNA formation and finally, two chromatids are formed i.e., a full-fledged chromosome is formed that will be eventually used in next cell division. In case of meiosis also, it occurs single chromatids are available in the sister, in the daughter cells and thereafter, it is replicated and another double-stranded DNA is synthesized. The DNA replication is semi-conservative in nature and it was demonstrated by Matthew Meselson and Franklin Stahl. So, what is semi conservative DNA replication? Let us see that, in DNA it is double-stranded in nature two strands are there, one is the 5' to 3' strand and another one is 3' to 5' strand. So, during replication process, these two strands are separated and finally, once the newly synthesized strands are produced in each of the cases, one parental strand will be there and using that strand as a template, another newly synthesized strand is formed. Ok!

This is true for both the DNA strands. So, in this way, over here you can see the parent strand i.e.,5' to 3' strand is here and this is the newly synthesized strand. Over here also, we

can see the 3' to 5' strand i.e., coming from the parent and this one is the newly synthesized strand. So, in this way, the DNA replication occurs in semi-conservative way. Now, briefly we will discuss about the transcription parts. Ok! What is transcription? We have, discussed earlier transcription is the conversion of DNA to RNA. Ok!

So, what are the things we will discuss over here, we will discuss about some consensus sequences, we will discuss about the coding and non-coding strand of the DNA, we will discuss about the promoter, terminator and the upstream and downstream sequences and what will be the structure of primary transcript those things will be gradually discussed. Ok! First of all, consensus sequence. Ok! So, this is a double-stranded DNA, it is a double-stranded DNA. So, for transcription, first thing we need to know for each and every gene in most of the cases the transcription start site is same, transcription start site each and every gene will have a particular transcription start site. Ok! And mostly it will be available, means the transcription start site will be same for that particular gene in different species. Ok!

So, this is the transcription start site, means that particular base on the DNA wherefrom first RNA will be synthesized or if we discuss about the protein coding genes, from this particular base, the first mRNA will be synthesized. Ok! So, before the transcription start site, the sequence which is available before it i.e., known as the upstream sequence. While the sequence which is available after the transcription start site it is known as the downstream sequence. If we think about the upstream sequence, how those sequences are designated? The previous base of the transcription start site is known as -1 site. It's previous base will be -2, in this way different bases are designated in the upstream region those are mentioned as minus (-). Ok!

While from the transcription start site, different bases will be there, in this way, +1, +2 in this way. So, in case of bacterial transcription, few upstream sequences available in -10 region, -35 region they play major role; the -10 region and -35 region in case of bacteria or prokaryotes. While in case of eukaryotes, approximately in between -25 to -30 region, in between this region some TATA box sequence is available, some sequences those are

known as TATA box binding protein. So, those TATA box binding protein sequences, they basically play major role in most of the eukaryotic gene sequence, eukaryotic gene expression or eukaryotic gene transcription. So, this -10, -35 in case of prokaryotes or -25 to -30 in case of eukaryotes, those are the part of the promoter region.

So, another term is coming i.e., promoter. So, what is promoter? Promoter regions are available in the upstream part of a gene, in the upstream part of a gene the promoter region is available. So, basically in this promoter region, the RNA polymerase binds, the major enzyme i.e., responsible for DNA to RNA conversion that RNA polymerase binds to the promoter region. Where does it bind? It binds to -10 or -35 region in case of bacterial gene expression and major binding will be in these -25 to -30 regions in case of eukaryotic gene promoter. So, in this way, a gene is transcribed. Gradually, we will discuss it little bit in more detail. Then for each and every gene, a terminator region is also needed. Ok!

The promoter is needed for initiation of the transcription, while once the transcription is done, a terminator sequence is available at the end of the gene. Once the terminator sequence is available, thereafter, only the RNA polymerase will leave the double-stranded DNA. So, then mRNA formation is stopped. So, now another thing i.e., available here, i.e., upstream and downstream sequence. We have mentioned about few upstream sequences like -10 or -35 regions in case of bacterial gene expression, while -25 to -35 region in case of eukaryotic gene expression.

So, these are the upstream sequences, along with these upstream sequences, some other upstream regions are also needed which, in which different transcription factor binds. Ok! Different transcription factor binds and they basically after binding, regulate the gene expression. In some of the eukaryotic genes, even after the transcription start site either in the UTR region or within the intron or some other part of the gene, some important downstream sequences are available those play some role in gene expression also. So, I am telling it once again, in case of prokaryotic gene, mostly the upstream part plays major role in transcription in gene expression. While, in case of eukaryotic gene, along with the

upstream part, some downstream sequences also play some role in gene expression for certain genes. Thank you.