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: 09

Lecture-38: Gene Cloning

Hello everybody. Welcome to SWAYAM NPTEL online course on Basics of Crop Breeding and Plant Biotechnology. And today we will start discussing on the preparation of gene constructs and promoter reporter constructs. This is the lecture 1, here we will be discussing the gene construct preparation and promoter analysis. So, these are the concepts which will be covered under this particular topic. First of all, we will be discussing about gene cloning, how eukaryotic genes look like, what are the things available within a gene and what is promoters will be discussed gradually.

Then we will be discussing the over expression construct preparation, then we will be discuss we will be discussing the gene silencing or hp-RNA, that is hairpin RNA construct preparation. Then different promoter reporter gene construct preparation will be discussed and finally, we will discuss some numerical questions on promoter analysis. So, let us start our discussion on gene cloning. So, before gene cloning or before any type of cloning, first we need to know the gene structure in a eukaryote.

So, in a eukaryotes, you know that in most of the genes in eukaryotes, the exon, introns are available. Ok! So, let us think that this box region are the exon part of a gene, while this line region is the intron part of a gene and the shaded area is the UTR region of a particular gene in most of the eukaryotic gene these things are available. So, a 5′ UTR is available untranslated region, a 3′ UTR is available that is also untranslated region available in the 3′ end part of the gene. Then different exons are there, let us assume this is exon 1, this is exon 2 and this is exon 3. If you think about the gene sequence which is

available in the genome on the chromosome, there within the DNA those exon, intron, UTR all sequences are available.

And at the end of most of the eukaryotic gene, the terminator region is available, while for controlling a particular gene for proper expression of a particular gene, a promoter sequence is available at the upstream region of the gene part. This is the promoter region. So, in this promoter region, basically the RNA polymerase binds and it start the transcription process of the gene. So, once the transcription is taken place, once the transcription is taken place first from this particular DNA the premature mRNA is formed, if it is a protein coding gene. So, from the gene the pre mRNA is formed and thereafter different processing is taken place in eukaryotes and during the course of those processing the intron portion, this is intron part, this is another intron part, some intron parts are here and here also those intron parts are removed.

So, let us assume this is the 5′ UTR part, this is the exon 1, this is the exon 2 and this is the exon 3 part. So, from this pre mRNA upon processing, the mature mRNA is formed, in mature mRNA these things are available means the UTR, 5′ UTR, 3′ UTR and exon 1, exon 2, exon 3 these parts are available while the introns are removed. After 5′ capping, 3′ polyadenylation is taken place and mature mRNA is formed. So, then those mature mRNA or mRNA that is translated during means for protein synthesis and once the protein synthesis is taken place, the protein synthesis will start from the exon parts, the UTR parts are not involved in the protein synthesis. So, in this way, different amino acids are synthesized based on the codon available in the exons and finally, the polypeptide is formed.

So, initially we have discussed about the transcription process, then this is the processing, pre mRNA processing and this is the translation process. Now, as we are discussing about the gene cloning or any type of cloning taking the eukaryotic model. So, first we need to think that what we have to clone and where it will be available. So, suppose we have to clone a particular gene, I am just giving an example its name is *Os-GLP1*, OS stands for *Oryza sativa* and *GLP1* is *germin like protein 1* gene. So, for

cloning of this particular gene means 2 options might be there means 2 objectives might be there.

First of all, some researchers may clone a particular gene from different cultivars of a particular crop, different variety of a particular crop to find out whether there is sequential variation or not. Ok! For doing such type of work we can clone the gene from the genomic DNA region, but if we have to express that gene through *Agrobacterium* mediated gene transformation, if you have to express that gene into some heterologous system it may be in bacterial system or it may be in other plant system, then we have to clone it from the matured mRNA, we have to start our cloning from the matured mRNA part. So, the gene what we have discussed *OsGLP1*, in this gene basically no intron is available. So, for any type of cloning first we need to get the sequence information of that gene, in case of rice most of the sequence information is available in NCBI database and for some other crops like *Arabidopsis, Brassica* several sequences are also available, but it is not true for all the crops. So, anyway we will take the example of rice and I will describe a couple of process for cloning different genes.

So, suppose one gene we have *OsGLP1*, it has no intron. If no intron is available in this gene, then we can clone this gene from the genomic DNA, as well as, we can clone it through the mRNA. But if we have to clone another gene let us assume *OsCAMTA1*, Os stands for *Oryza sativa* and CAMTA is calmodulin binding transcription activator, it is a transcription factor protein. This type of protein is involved in different signaling cascades, once a plant encountered any types of stress, this type of transcription factors play crucial role in other different protein synthesis and protein-protein interaction also. So, if you see the gene sequence of *OsCAMTA1* in NCBI database, there you can know a number of introns are there, means this gene has intron.

So, if intron is available, we have to start our cloning from the mRNA, I will describe the process later on. So, for gene cloning, basically, we can take any one of this approach based on the availability of intron. Ok! Next, if you have to clone the promoter region of a gene, if you have to clone the promoter region, what is the promoter region of a gene?

Later on, we will discuss once again, but promoter region is the upstream regulatory region of a gene, where mostly RNA polymerase binds and in addition to that other different transcription factor binding sites are available. So, suppose if we have to know about a particular gene that in which tissue those genes are expressed in which tissue the protein of that gene will be active, then we have to characterize the promoter region of that gene. Because from that promoter maybe in a plant suppose different tissues are there, seeds are there, seed tissue, then leaf tissues are there, root tissues are there, flowers are there.

So, all genes are not expressed throughout the plant, the genes which are expressed throughout the plant in all types of tissues that are known as housekeeping gene. So, housekeeping genes are expressed throughout the plant, it is expressed in all the tissues, in all developmental stages, while most of the genes available in eukaryotic system, they show tissue specific or stress inducible expression. Most of the gene, they show either stress inducible or tissue specific expression. So, for those genes which either show stress inducible expression or tissue specific expression, those genes might not be expressed in all the time. So, if we characterize the promoter of those genes, then we can tell that in which particular tissue those genes are expressed.

Suppose, some genes which is active in the root tissue. So, its promoter will be active in the root tissue, in the root tissue, it will form the mRNA, it will form the mRNA, therefrom protein will be produced and it can show its features. While in the leaf tissue, in the genome in the chromosome, the gene, promoter everything will be available, but it will not be expressed as some tissue specific expression is associated with that particular gene. It's true for stress induction also. So, if you have to clone the promoter region of any gene, we have to plan the experiment by considering the genomic DNA or the chromosomal DNA, because in mRNA the promoter is not available.

The promoter is available on the DNA sequence what I have drawn earlier. So, three things we are discussing, one the gene cloning where no intron is there, another one is gene cloning where some intron is there and third one is the cloning of the promoter

region. So, in case of gene cloning where no intron is there, we can start the experiment from the genomic DNA. First, we have to design gene specific primers, first we have to design gene specific primers. So, the promoter region cloning suppose, this is the point number 3.

So, in case of 1 and 3, how should we plan the experiment? We need to isolate genomic DNA, then we have to design gene specific primer. Now, suppose this is the gene, at its 5′ to 3′ end some sequences is there ATGC, while at the 3′ end TAAC this sequence is available. While in the another strand, the sequence will be TACG and this sequence will be ATTG. So, we are talking about the gene specific primer designing. So, in case of ideal primer which are used for cloning at least 18 mer or 18 nucleotide is taken from the gene part at least.

So, we can extend it up to 24 mer also. Ok! So, just for an example we are considering 4 bases over here. So, if these 4 bases are available in the 5′ end of the gene and in the 3′ end of the gene, how can we design the primer? So, herefrom if we just recall the PCR process, these 2 strands will be separated one will be 5′ to 3′ and 3′ to 5′ strand. And DNA synthesis in PCR or during DNA replication is always taken in 5′ to 3′ direction. So, our primer will move from here and another primer will move from here right?

So, one of our primer sequence should bind over here means, it should bind this strand the 3′ to 5′ strand. So, the primer sequence will be ATGC …. if we have to take the 18 mer, the other bases will be taken. So, in this way we can design the gene specific primer from one end while from this end we have to design the primer in such a way that will be complementary to this strand right? So, this will be our forward primer while our reverse primer will be here sequences from 5′ to 3′ will be GTTA, GTTA that sequence will be available over here. So, that it can start polymerization by taking this DNA strand as a template.

So, this is the 5′ end of both the primers and in this way, we can extend the primer sequence up to 18 mer. So, in this way, we can design the gene specific primer. Now, if

we have to clone a gene where intron is available, in that case, we have to start our experiment or we have to start our cloning by taking the mRNA sequence first. First, we will take the mRNA sequence because in the matured mRNA different exons are available together, no intron is available and we have to design the primer from this part and from this part. So, first we have to convert the mRNA into cDNA that is complementary DNA by using which enzyme we have mentioned earlier that is reverse transcriptase enzyme.

By using this particular enzyme, we have to convert mRNA into cDNA, then using this cDNA and gene specific primers, we have to amplify our target gene. We have discussed about PCR earlier. So, in this way, first we have to design the gene specific primer to amplify our target gene or targeted promoter region. For certain case for some cases, we have to start with the genomic DNA, for some cases if the introns are available we have to start from the mRNA, mRNA to cDNA then primer designing, PCR amplification and finally, we will see the amplified band of that particular gene.

Now, I will continue this process, how can we clone a gene sequence within a plasmid? Earlier we have discussed about different plasmids. So, now suppose in one primer the forward primer what initially we had their sequence was ATGC, while in the reverse primer we had initial sequence GTTA. So, if we use these two types of primers we can amplify the gene. How can we clone it? Once we can amplify the gene, we can initially clone it in TA cloning vector because in most of the PCR reaction different companies make the PCR cocktail, PCR master mix. In most of the PCR reaction at the end of PCR, some poly some A bases are added poly ‘A’s are added. Ok!

So, using this TA cloning vector, we can clone our PCR product easily in those vectors and once we can do cloning in TA cloning vector, through sequencing we can confirm whether our desired gene has been inserted or not. So, it is a type of cloning vector, not expression vector, it is a type of cloning vector here, we can clone our PCR product easily and then we have to confirm it by sequencing. Next, we are trying to gradually we have to move into the construct preparation right. In TA cloning vector our PCR product has been cloned over here, but we do not know at which direction our start codon is

available at which direction our stop codon is available, we do not know. So, for proper orientation, we need to put some restriction enzyme for subsequent cloning.

So, let us think that we are adding two different restriction enzymes over here. We have to again design the primer by taking the forward sequence of the gene and the reverse sequence of the gene, the forward primer and reverse primer sequence, then we have to add the restriction enzyme sites. Let us assume over here, in the forward primer GAATTC this restriction site has been added, while in the reverse primer CCCGGG this restriction site has been added. So, it should be added in the 5′ end of the specific region, this is the gene specific region, at it is 5′ end we have to put the restriction enzyme sites and before that, at least 4 filler bases should be added, at least 4 filler bases should be added. This is due to the fact that until and unless we keep some bases over there, once the PCR product is formed, the restriction enzyme cannot bind the DNA properly because the restriction enzyme has to sit over there, thereafter, it can cause cleavage on the DNA.

So, now let us assume we have designed new forward primer and new reverse primer by adding *Eco*RI restriction site in the forward primer and *Sma*I restriction site in the reverse primer and then we have PCR amplified the DNA, this is our PCR amplified DNA. So, at its one end, this sequence will be available GAATTC ATGC and some other bases right? While in the reverse primer we will be having CCCGGG then ATTG different bases, this one was the 5′ end. Ok! In this way, we can amplify our initial gene which was cloned in TA vector, in TA vector what we cloned earlier, we can amplify it using this type of primer where the restriction enzyme sites are incorporated. So, let us draw the other strand of the DNA, here it will be GCAT, then GAATTC the filler bases, here it will be CAAT CCCGGG and filler bases.

So, this PCR product now we got this is our PCR product. We are discussing the gene cloning, now we have to clone it in some vectors for doing the directional cloning. Ok! So, let us for such type of cloning mostly we used pUC based vector, pUC18 or pUC19 those vectors are very popular. So, let us assume, this is our pUC vector, it is a plasmid that is double stranded circular DNA, it has a origin of replication, it has selectable

marker gene and it has multiple cloning sites. So, let us assume in the multiple cloning sites, different restriction sites are supposed to be there, suppose one site is there *Eco*RI and at another end or somewhere over here, the *Sma*I site is available.

So, in *Eco*RI, the sequence GAATTC was available GAATTC in both the strand of the DNA while, in case of *Sma*I, CCCGGG sequence was available. Ok! Suppose these two sequences are available in the pUC based vector. Now, once our PCR product is ready and this vector is available, we have to digest the vector as well as the insert with specific restriction enzyme. Here two restriction sites have been used *Eco*RI and *Sma*I. So, we will digest it, basically we have to do double digestion means we have to digest the PCR product with two different restriction enzymes while we have to digest it also.

So, if we digest it, digest the PCR product what will be happening? The restriction enzyme will cause cleavage over here, the *Eco*RI restriction site will cause cleavage over here, while the *Sma*I enzyme will cause cleavage over here. So, after double digestion what will be the status of our PCR product? Here we will be having AATTC …. from here I am putting …. So that, it may not be too much complicated and at the end of that we will be having CCC sequence. While in the other strand, the G will be available over here, sorry, I have made some mistake over here basically this one was the G AATTC site. So, cleavage will be done over here.

So, over here G will be available and here from I am adding …. and at last, we will get this sequence. So, basically this part of the DNA will be removed, will be having this type of staggered end at one side of the PCR product. While in another side, this part will be removed after digestion we will get a blunt end over here, means our DNA will be like this. Because two types of restriction enzyme has been used one can cause sticky end and one can cause blunt end. Now if we digest the vector what will be the scenario? In the vector, cleavage will be taken place over here and here.

So, at one end G will be available in another strand CTTAA will be available. Then over here, we will be having CCC and because the *Sma*I will cause cleavage in between and we will be having GGG. This will be our pUC vector digested part after this

digestion. Because this small part will be eventually removed during the purification of the DNA digested product. So, then earlier we have mentioned about restriction enzymes.

So, in this way restriction enzymes are used, it is a part and parcel of the recombinant DNA technology or plant molecular biology work. Then we have to seal these two things using ligase enzyme. So, once we will put this vector, sorry, this vector part, digested vector part and this digested insert part together along with T_4 DNA ligase and buffers. Then you can see the sticky end is available over here.

So, it can stick in this region. So, eventually we can generate a plasmid where it will be like this CTTAA. Here, CCC GGG this part came from the vector, while AATTC GGG and over here, we will be having CCC. So, these things will come from the PCR product. In this way, the digested plasmid will be circularized again and if you recall the pUC vector, we had the selectable marker gene, we had the origin of replication right! Once this ligation process is done, then we have to use this ligation product for bacterial transformation. Ok!

Then we have to do bacterial transformation and we can find out the recombinant bacterial colonies where this particular plasmid will be available. And the selection will be done based on this selectable marker gene on the media we have to put a particular antibiotic maybe kanamycin or ampicillin based on this selectable marker gene. And those bacteria which will be harboring this particular ligated plasmid, they will survive over there. Because in ligated product, our PCR in the ligation mixture, our digested PCR product and the digested plasmid was available. So, if it ligates properly, then only it can circularize, then only it can replicate.

So, it is mostly obvious that whatever the colony will be getting under the selection media all of this bacterial cell, will be having this type of ligated plasmid or recombinant plasmid. So, in this way, we can clone a particular gene, either we have to start from the

genomic DNA or from the mRNA, even through this process, we can clone a particular promoter region of a gene, if we know the sequence information for designing the primer.