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: 07

Lecture-30: AFLP

Hello everybody!

Welcome to the SWAYAM NPTEL online course on the Basics of Crop Breeding and Plant Biotechnology, that is the types of molecular markers and application of molecular markers. Here, we will be discussing a few molecular markers like AFLP, ISSR, SSR, and CAPS. So, under this topic, we will be discussing about, few specific molecular markers, like AFLP i.e., an amplified fragment length polymorphism marker. Then we will be discussing ISSR marker i.e., inter simple sequence repeat marker. Then we will discuss, SSR markers i.e., highly used, very useful markers, simple sequence repeats markers, and we will be discussing the CAPS marker, cleaved amplified polymorphic sequence. Ok!

So, out of this marker, some other markers are also used like SNP, we cannot forget about that, single nucleotide polymorphism marker, it is mostly based on the sequence information, and we will not cover specifically SNP in our discussion. So, another important molecular marker, its name is the single nucleotide polymorphism marker. Ok! So, coming to AFLP, amplified fragment length polymorphism. So, AFLP is a restriction enzyme, as well as, a PCR-based marker, ok both types of things are available over there.

So, first I will describe the process of AFLP molecular marker, then I will discuss it in detail. So, what are the things needed to use this particular marker? First of all, we need to isolate the genomic DNA. Next one, we have to digest the genomic DNA with two

different restriction enzymes. So, two restriction enzymes are used in such a way, one restriction enzyme will be a frequent cutter and one will be a rare cutter. So, one frequent cutter and another one is a rare cutter.

I think that restriction enzymes which are commonly used in molecular biology, that is known to you. So, few restriction enzymes are there which can recognize 4 base pair sequence, and which can cause cleavage within that 4 base pair sequence. While, some restriction enzymes are there that recognize 6 base pair sequence on the DNA, and cause cleavage over there within that 6 base pair region. So, within a genome, the availability of a 4-base pair restriction enzyme definitely will be more. So, as a frequent cutter, we can use those restriction enzymes, that can recognize 4 base pair sequences like *Mse*I.

While, as a rare cutter generally those restriction enzymes are used, that recognize 6 base pair sequences, like *Eco*RI, or *Bam*HI, or *Hin*dIII, those sites could be used as a rare cutter. So, if we consider about AFLP process, first we need to isolate the genomic DNA, and then definitely we need to check the quality and quantity of the genomic DNA, not too much amount of DNA is needed, but the quality should be good. I mean lack of smear should be there, less degradation should be there, then we need to digest it with 2 different restriction enzymes, one frequent cutter and one rare cutter. The third point is very important in AFLP, and it is unique in AFLP here we need to design some adapters.

So, those adapter designing will be based on the restriction enzyme, we are using. Ok! So, it will be based on the restriction enzyme used, and we have to make our adapter in such a way that after adapter ligation, the restriction enzyme site will be abolished. So, then next part is primer designing. Ok! In AFLP, different types of primers are used, the first set of primers used, those are known as pre-amplification primers.

In pre-amplification primer, once we will use it then basically, we can get a smear kind of

thing, a lot of fragments will be produced, a lot of amplicons will be produced, and thereafter we have to use the selected amplification using a specific primer. Not specific basically, we have to design the selected primers in such a way that after the pre-amplification primer, we have to add different bases up to 3 bases, we can add consecutively base number 1, which could be A, T, G, or C anything. Then in base number 2 we can put A, T, G, C different bases. So, different permutations and combinations could be there and, in this way, we can make 3 additional bases in these selected primers. After this amplification, once the amplification will be done by using selected primers, then we have to load the PCR product on the gel, we have to do gel electrophoresis.

In most of the cases, the polyacrylamide gel electrophoresis is done for this type of analysis. So, let us start these things a little bit in detail. So, how actually at the molecular level, at the DNA level, the AFLP polymorphism works. So, suppose we have genomic DNA and within this genomic DNA, 2 restriction sites are available. So, this is the 5' to 3' strand.

So, suppose within the genomic DNA, the *Mse*I restriction site is available i.e., a frequent cutter, and a *Bam*HI site is available i.e., the rare cutter. Ok! So, let us draw another strand of the DNA it will be AATTCCTAGG. So, let us assume the genomic DNA has been isolated, this is the genomic DNA. Now we are starting, point 2, we are digesting this genomic DNA with 2 different restriction enzymes. So, what will be happening? In *Mse*I it will cause cleavage over here, and here, while in *Bam*HI it will cause cleavage over here and here.

So, after this double digestion means digestion of the genomic DNA with *Mse*I and *Bam*HI what will be getting? Let us assume, this is the upper strand, TAA different sequences are there up to G will be having over here, while in this strand, these things will be available. So, just I have done the cleavage over there. So, if you see carefully this part of the DNA has been cleaved, while from here this part of the DNA has been cleaved right? So, now we are considering only this part which is remaining over here.

So, in this way, once we will do the digestion of the genomic DNA with 2 different restriction enzymes, we may get this type of fragments.

In some fragments, only *Mse*I will be available at both ends. In some fragments like this *Mse*I and *Bam*HI sites will be available. In somewhere, maybe *Bam*HI and *Bam*HI might be available. So, in this way, different fragments will be generated. So, thereafter we have to design the adapter. Ok!

So, let us try to make the adapter for the *Mse*I restriction enzyme. So, *MseI*, AAGCAT suppose we have made this adapter as the *Mse*I adapter. Similarly, the adapter may not be so much small at least 12 to 15 bases will be available over there. This is just as an example; I am making this sequence. So, that you can understand how the adapters could be ligated over here.

So, that the restriction sites will be abolished. Ok! So, let us see how this adapter will bind over here. If you see carefully this *Mse*I adapter, the overhang is there AT overhang is there. So, what was our original sequence after digestion? We had the sequence like TAA different bases over here we had G in one strand, and in another strand, we had T over here and here CCTAG. Now this adapter can easily bind over here.

So, let us draw how it will bind over here. Over here AT overhang is available. So, this AT can bind in this region then we had CGAATGG. On the other strand here, we had GCTTACCG. So, this is the adapter that we have used and if you carefully see initially the site was TTAA, this one was the *Mse*I restriction site, sorry this one was the *Mse*I restriction site.

Now, this site has been modified after adapter ligation. So, it will not be attacked by *Mse*I further means, the restriction site has been abolished. So, now let us make the adapter for *Bam*HI. Ok? So, once we have to make the adapter we have to think about the overhang sequence. So, in the case of *Bam*HI, let us start from the overhang sequence we had GATC right?

Here we cannot put C base, because then GGATCC will be revived. So, other than C we can put any base. If you think about here GATC this overhang will be complementary to this sequence that is available over here. So, in this base we have to add any base other than C. So, that the restriction site will be abolished.

Thereafter, we have to put the other filler sequence which may be TCGCTC. So, let us make the other strand it will start from here TA then GCGAG. So, this could be used as a *Bam*HI adapter. Another base I can write over here ok? So, now I am just annealing this particular adapter over here.

So, we had GATC then A then TCGCTC. Over here we are having TAG CGAG and C. So, if you see GGATCC initial sequence was there, here a base has been modified. So, the restriction site has been abolished. So, up to this, we have done over here.

So, now I will go to the next slide, and I will tell how the pre-amplification primer and selected primers will be designed and finally, we can do the gel electrophoresis and those things to generate some polymorphism. So, we have a sequence GCC ATT CG. Well, so let us start with the adapter ligated product means once the DNA has been digested with two different restriction enzymes, one rare cutter and one frequent cutter then we have ligated the adapter. Now, I have written just the strand, 5' to 3', strand and we are having another strand 3' to 5'. So, this thing will be produced, GG TAA GC then AT then intermediate sequence will be available, then we will be having CC TAG and TAGC GA GC.

So, basically, this part has come from the adapter, right? This part has come from the *Mse*I adapter, and this part has come from the *Bam*HI adapter. *Bam*HI, sequence we had GGATCC, and *Mse*I sequence we had TTAA. So, once the ligation is done then we have to do the pre-amplification. We have to do the pre-amplification.

For pre-amplification, as we have used two different restriction enzymes, one frequent

cutter, and one rare cutter, we have to design a pre-amplification primer, for both restriction enzymes. Ok! So, till now we know about the PCR reactions, and how the primers are designed. So, let us make the primer for *Mse*I. Ok! So, *Mse*I pre-amplification primer what would be that? *Mse*I pre-amplification primer, or it is also known as +0 primer. Ok! If you carefully see this ligated DNA product this part could be used as *Mse*I primer.

So, the primer sequence could be GCC ATT CG TAA up to the last base of the restriction site, the restriction site was TTAA. So, up to A will use the pre-amplification primer, how it will work I will tell you later on. While we have to design the pre-amplification primer for *Bam*HI also. Ok! For *Bam*HI, basically, this sequence could be used as a primer. Ok! So, I am writing it *Bam*HI pre-amplification primer i.e., +0. Ok!

We will be having over here if we write from 5' to 3', the primers are written mostly in 5' to 3' direction. Here, it will be CGAG, I am taking from this side CGAG then CG then AT then GAT, and CC, Ok? This is the 3' end over here OH is available, here also I have taken up to the last base of the *Bam*HI restriction digested site. Ok! C is available at the 3' end. So, how it will work once the DNA, will be denatured during the pre-amplification process, here DNA will be denatured.

So, after denaturation one strand will be GCC ATT CG, then TAA other different bases will be there, then G then we will be having GATCATCGCTC, this is one strand, and another strand will be GGTAAGCAT, then we will have T here CC TAG then we will be having GGTAACGAGC. So, we have *Mse*I primer. So, if you carefully see this *Mse*I primer can easily bind over here like GCCATTCGTAA. Once the primer binds, the *Mse*I pre-amplification primer will bind, then it can start amplification in this way while over here once the *Bam*HI primer binds, where the primer will bind, a primer will be binding over here. Its sequence is CG means; it is starting from CGAGCGATGATCC. Ok!

This is the 5' end of the primer, *Bam*HI pre-amplification primer, and it can start amplification in this way. So, once the pre-amplification is done then we will see a smear

kind of thing there, after a few cycles of pre-amplification we need to use the specific primer. Specific means here those primers are known as selected primers. So, after pre-amplification, we have to do selective amplification.

We have to do selective amplification. Ok! Pre-amplification if you think about, a genomic DNA, where a rare cutter was used and a frequent cutter was used several fragments will be there, in some of the fragments, at its two ends only the frequent cutter site will be available. Some fragments will be there where on one end frequent cutter will be available at another end rare cutter will be available. In this way, different fragments will be generated. So, in pre amplification basically we will see a smear kind of thing.

Thereafter, after we have to do selective amplification. In selective amplification, again whatever the restriction sites, we are using based on that we have to plan it. So, let us make some selective amplification-related primers specific to *Mse*I. So, we have to use the same restriction same sequence available in the adapter as well as the restriction site. So, that sequence could be GCCATT, I am writing this sequence once again ATT then CGTAA thereafter we have to add any specific base. Ok! This N, this N may be either A or T or G or C any base we can add.

So, how it will work I will tell you later on. So, once we will add first base over here it could be A or T whatever then it is known as +1 primer. So, similarly, we can do the selective amplification by adding up to 3 bases, at the 3' end of the pre-amplification primer? So, in this way, +1 primer could be designed. Now we can make another primer where sequence, is GCCATTCGTAA then NN, 2 N are being placed. In the first N position suppose we have added any one of these bases, T we have added while in next N over here, we can add any one of this ATGC base.

So, in this way, different permutation combinations could be created in the last 3 bases. Ok! So, this is known as the +2 primer. Similarly, the +3 primer also could be made in this way the selective amplification primers could be designed for *Mse*I, and selective amplification primers for the *Bam*HI site, also could be created by adding CG AG CG

ATG ATCC. In this case, we have to add either 1 N or we have to add 2 N up to 3 nucleotides we can add gradually over here. Ok!

So, here I am mentioning this in red color. So, that you can easily understand which one is +1, which one is +2, or +3, I have not mentioned it yet. So, here this is the +1 primer, in +2 same sequence will be there, over here CCTAGTA CGAGC this will be the +2, sequence. Ok! In this way, up to +3, selective amplification primers we can make and once this selective amplification is used then from the smear whatever we got in the pre-amplification stage gradually, we will see specific bands. Ok! So, how we can see specific bands? So, let us take the example of two different varieties and I will try to explain how can we see the specific amplification using the selective primers. Suppose we have two different varieties, variety 1 and variety 2.

In varieties 1 and 2, in its genome, different digested fragments have been generated. Suppose we are considering only one frequent cutter, like *Mse*I we have used earlier and we are having a rare cutter site, like *Bam*HI we have used earlier. Suppose, this fragment is available while in variety 2 at the same position, the *Mse*I frequent cutter is available and *Bam*H1 rare cutter is available. Ok! So, within the restriction enzyme site, no variation is available between variety 1 and variety 2, but after the restriction enzyme site, only a single nucleotide variation is available. So, let us have a look into this suppose in variety 1, the sequence I am talking from here where the TTAA sequence was available.

If you recall on the genomic DNA, we need to have restriction enzyme sites. Suppose, over here the sequence was TTAA then, we had CGC and different bases and in case of the *Bam*HI site, we had the restriction enzyme site GGATCC and before that means, before that G suppose we had TT, this sequence, we are thinking that in variety 1 this sequence was available. So, let us think about its opposite strand, its opposite strand will be AATTGCG and here sequence will be AACCTAGG it was available. Let us make the sequence over here in variety 2, suppose over here TTAA was there then here sequence was CCC if you think here, it was G here it was C the variation is available at this

position after the *Mse*I restriction site. Ok! Then in-between sequence was available and over here suppose, the sequence was not altered TTGGATCC in its opposite strand the sequence will be AATTGGG define bases will be there then will be having AACCTAGG. Ok!

So, now if we carefully think about these two genomic DNA regions, in variety 1 and variety 2 once we will do the restriction digestion. So, digestion will take place over here and here, for *Mse*I and for *Bam*HI the cleavage will be done over here and here, right? Similarly, here we will see the cleavage, and here also we will see the cleavage. So, once the cleavage is done, what will be happening, is that this part will be removed and over here this part will be removed. So, an overhang will be created after restriction digestion this overhang will be created then we have to ligate with the adapter.

So, let us ligate with the adapter, what adapter we have used suppose we are using the same adapter, over here after TAA we had GCTTACCG, and in another strand now I am thinking about the ligated DNA. So, in ligated DNA in this strand will be having over here TA then CG then AA then TGG. Ok! So, after adapter ligation, this double-stranded DNA would be available. So, let us think about this sequence will be having: GATC then again ATC then GCTC. Ok! So, similarly over here also after G will be having TAGC GAG.

So, this one was the double-stranded DNA after adapter ligation. Now, carefully think about variety 1, the pre-amplification has been done, in pre-amplification this primer can easily bind before pre-amplification again, our DNA will be denatured then the pre-amplification primer which is starting from GCC. It can bind over here it can bind over here and it will bind up to AA means up to here, the pre-amplification primer will bind right. Similarly, over here the pre-amplification primer will bind up to this, sorry here the pre-amplification primer will bind up to this. Now once we are planning to do the selected amplification, once we are planning to do the selected amplification then what will be happening? In selected amplification, if we use the +1 primer, in +1 primer supposes, over here in N at the position of N, ATGC different sequence could be added if

we add C over here.

So, it can easily bind then think about the +2 primer, in +2 suppose we have used again C. If we use C, then it cannot bind now, think about the variety 2. Suppose, I am not writing all the filler sequences, let me write so that you can understand it properly. We had TAC, I am just talking about a single strand. So, this thing will be applicable in both the strands TAC, then G double A G double A then TGGTGG. Ok!

This thing we had in variety 2. Now during the selective amplification, during the selective amplification process our pre-amplification primer has bound up to this thereafter, suppose our first base was C as I mentioned there. So, it can easily bind, our second base means in +2 primer suppose the second position was also C. So, the +2 primer can bind variety 2, but it cannot bind variety 1, if we put it in +2 position C. Ok! So, in this way, if you compare these two varieties, variety 1 and variety 2 suppose once we are using the +1 primer and once we are using the +2 primers what could be the scenario once we are using the +1 primer, suppose over here we are getting this type of bands coming from a *Mse*I and *Bam*HI fragment available within the genome means wherever only after the pre-amplification primer, C is available that thing will be amplified. Next, once we are adding, once we are using the +2 primer, over here we have used +1 primer, once we are using the +2 primer means, two bases have been added after the pre-amplification primer sequence. Ok!

Over here in variety 1, we may see suppose, only 1, but in variety 2 two bands are available. So, we can interpret from here, that in variety 2 at this loci, this CC sequence is available while in variety 1, at this loci the sequence is C and something else thereafter some other bases were there, not C. So, it was not amplified in this case in this way using selected amplification, we can distinguish different varieties by using AFLP technique. Ok! Thank you.