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-31: ISSR, SSR, CAPS

Welcome back. So, we will continue again. So, let us see what are the advantages and disadvantages of AFLP marker. First, it is highly reliable and reproducible. No DNA information is required, which means we can start with any unknown plant because we have to isolate the genomic DNA, we have to digest it with two different restriction enzymes. It is understood that in most of the genomic DNA, the restriction enzyme sites will be available so, you can choose any two different restriction enzymes.

It is highly information-rich due to its ability to analyze a large number of polymorphic loci at the same time. This means, if you think about the restriction digested fragments *Mse*I, *Bam*HI fragments. These fragments will be produced, a number of times within our genome because several *Mse*I sites and several *Bam*HI sites will be available. So, a number of loci means a number of positions on the chromosome could be targeted, and based on that the polymorphism could be detected at the same time.

Then co-migrating AFLP amplification products, are mostly homologous and locus-specific. Means, if we are obtaining that suppose we are analyzing three varieties, variety 1, variety 2, and variety 3. In varieties 1 and 2 we are getting this type of band. While, in variety 3, we are getting this type of band. It means we can assume that from a particular locus in variety 1 and 2 we are getting this type of amplification, while in variety 3 we are getting this type of amplification. Ok!
Those are mostly homologous and locus-specific. Now, let us discuss, the disadvantages of AFLP. First, multiple steps are required and the process is labor intensive. Here you

have to isolate the DNA, you have to digest it, you have to design an adapter, and different processes are there. Then, template DNA free from inhibitors is required, because if inhibitors are there then the restriction digestion will not be possible. Ok!

Then complicated techniques like PAGE, i.e., polyacrylamide gel electrophoresis, as well as silver staining need to be performed. Because, sometimes a large number of bands are obtained. So, by normal polyacrylamide gel electrophoresis, we may not see it. So, we have to do silver staining. So, that the smaller fragments could also be visible or closer fragments could also be visible.

The process is costly, because of restriction enzymes, adapter designing, primer designing, lot of PCR reactions. So, those things will be involved and it acts as a dominant marker, mostly until and unless the restriction enzyme site used for digestion is mutated. So, suppose within some genotypes, as I was giving the example of variety 1 and variety 2, what did I say there that no changes were available, no changes were available within the restriction enzyme site? In between the restriction site, some variation was there right, but if the variation is available within the restriction enzyme site, then it can work as a co-dominant marker, but otherwise, mostly it will work as a dominant marker. Now we will be discussing the inter simple sequence repeats. Ok!
Now, before discussing that what is simple sequence repeats? Those are simple sequences that may be a mononucleotide, dinucleotide, trinucleotide, or tetra nucleotide. These number of nucleotides are repeated several times within a genome. Those repeats are very common in most organisms, especially in eukaryotes and those types of repeats are very common. So, let us assume, a repeat may be AT, AT, AT say AT sequence may be available for 50 times. Ok!

So, it is a small nucleotide dinucleotide repeat so, we can write it as $(AT)_{50}$. Somewhere the sequence may be ACG, ACG, or ACG this sequence may be available on a particular position on the genome for 17 times. So, we can write it as $(ACG)_{17}$. Ok! So,

in this way, different repeats are available on our genome. So, using this particular type of molecular marker, inter simple sequence repeats, we can distinguish two different genotypes or we can distinguish different varieties based on this repeat sequence and their binding things and those things. So, first of all, it is a molecular marker, here single primer is used it is a PCR-based molecular marker.

So, we need to have a single primer. Now, you think can a single primer cause polymer means cause amplification in PCR; whatever we have discussed so far in PCR two primers we are using forward primer, and reverse primer. Ok! So, can a single primer make amplification, during RAPD we have discussed, it yeah the single primer can cause amplification if the single primer can bind in forward and reverse orientation within a genome. So, the single primer should bind on the forward and reverse orientation within a genome, within a span of 2 kb, to a maximum of 3 kb, because in PCR generally, we set the reaction in such a way that the amplification product should be obtained within 3 kb. Ok! So, let us think about this particular type of molecular marker, suppose this is the genome over here we had the AT sequence available for 15 times, and the 15 times AT sequence is available.

Thereafter, a few bases, we have the TA sequence is available for 20 times. Suppose this is the 5' to 3' end of the DNA, and this is the 3′ to 5′ end of the DNA, we had here it will be $(TA)_{15}$ here it will be $(AT)_{20}$. So, this is our original DNA strand. Now suppose we have designed an ISSR primer i.e., $(AT)_7$ if you recall RAPD, in RAPD generally 10 mer primers are used 10 nucleotides long, in ISSR the primer sequences are relatively larger, and primer sequences are larger. So, its T_M is also relatively higher than the RAPD primers.

Suppose we have used an $(AT)_7$, ISSR primer let us see how it will bind over here. So, during PCR our 2 strands will be separated here AT, AT, AT, this AT is available for 15 times, is available while we have TA, TA, this is available for 20 times. This is one of our strands, 5' to 3' and we have another strand TA, TA, TA, TA; AT, AT, AT, AT; 2 strands have been separated, during the course of denaturation then our primers are coming into

the picture because primer has to anneal. So, what is $(AT)_7$? $(AT)_7$ means AT, AT this sequence is available for 7 times it has a 5′ to 3′ end, 3' OH will be there. So, our AT, AT can bind over here 5' to 3', and it can cause amplification in this way.

While, if you see that sequence our AT can start binding from here also within the TA the AT is also available. So, in this way the primer can bind over here also, this is the 5' to 3' end, AT, AT it can bind in this way it can start binding from here. So, in this way, if this region is available in between 2 kb or maximum within the 3 kb span, then this things could be amplified easily by such type of ISSR markers. Ok! So, in inter simple sequence repeats how does it work? This primer basically, binds to the repeat sequences, on the genome and those repeat sequences will be available in inverted orientation, the repeat sequence has to be available in inverted orientation. And, another thing is that it should be available within a span of 2kb to 3kb maximum. Ok!

And, the region between 2 repeats which are available in inverted orientation is amplified i.e., the way in which ISSR works. Now, if you think about this particular sequence whatever the primer we have used $(AT)_7$ and over here we had $(AT)_{15}$, over here we have $(TA)_{20}$ means the primer binds can bind over here, can bind over here also. So, if you see $(AT)_7$ it can bind at this proximal part of the DNA at this distal part of the DNA also right? Hence, the primer can start binding from here because $(AT)_7$ is our primer, it can start binding from the middle part it can start binding from the end part also. While over here also here the sequence is repeated approximately 20 times.

So, our primers can bind from the far end, it can bind from the middle it can bind from the forward part also. So, at different positions, the primer may bind. So, it may not be very specific. So, how can we improve its specificity? We can improve its specificity by using a 5' anchored base or a 3' anchored base. Suppose we are using $(AT)_7$ G this is the 5' end of the primer. This is the 3' end.

So, at the 3' end suppose we have put G. So, within this AT, AT, AT, AT, over here our

primer cannot bind, our primer will bind only this region if C is available, because we are adding $(AT)_7G$ primer then only our primer could bind over here, right? In this way, we can improve the specificity of the primer by adding either 5' anchored bases, or by adding 3' anchored bases. So, now let us see what are the advantages and disadvantages of ISSR primer. In advantages, first, no prior sequence information is required like AFLP we can play with ISSR marker in any unknown genome because the repeats might be available in most of the plants maybe in spite of using $(AT)_7$ if we use $(ATG)_5$, that may show some polymorphism, but the repeats will be available. Ok!

So, prior sequence information is not required and then displays high polymorphism because, if we use different sets of ISSR primers definitely different levels of polymorphisms could be detected. The technique is simple and cost-effective, we have to isolate the DNA, a small amount of DNA is required because it is a PCR-based technique and we have to design the primers then through PCR amplification we can design the polymorphism. Then multiple loci can be amplified simultaneously. Ok! This means suppose in chromosome number 5, 7, and 9, in 3 different chromosomes, suppose in chromosome number 5, ATG and its reverse sequence are available maybe at 1 kb distance while in chromosome 7 the ATG and its reverse sequence ATG repeats and its reverse sequence is available maybe within a span of 1.5 kb. So, we can get bands from here, also as well as, from here also after PCR and so, multiple loci can be amplified simultaneously if the same types of repeats or its reverse orientation is available.

The technique can be applied across various organisms. Suppose, we have designed some ISSR primers for screening 100 rice genotypes those ISSR primers could be applied in lentils and wheat in other genotypes and also in different organisms it can work. So, what are the disadvantages of this ISSR primer? First of all, ISSR is a dominant marker. Ok! What is a dominant marker, a co-dominant marker? I have discussed earlier in the dominant marker, that basically, the homozygous dominant and heterozygotes could not be distinguished. Ok! But if we use a co-dominant marker, then these two could be distinguished. So, ISSR is a dominant marker then no direct sequence information is obtained just getting the PCR amplification, we are getting the band over there, but the

sequence information is not obtained because the primers may bind at the start of the repeat it may bind in between of the repeats and it may bind at the end of the repeats.

So, at different times maybe our amplicon size might vary. Hence, lack of reproducibility is there, sometimes we are getting, a 190 base pair band using ISSR sometimes we may get a 170 base pair band. So, the reproducibility will be less, because the repeat our primer can bind anywhere within the repeat then ISSR results may not be directly comparable across different species, due to variability in repeat motifs i.e., also true whatever the band we are expecting in rice, in different rice germplasm that things may not be expected in other plant species, because there the number of repeats might be more, there the other variations might be more. Ok! So, these are the advantages and disadvantages of ISSR primers. Now, we will be discussing about simple sequence repeats. Ok!

So, in ISSR we were playing with a repeat sequence, we do not need any sequence information just based on the repeat sequence we can design a primer. Suppose this is our genome, as I was talking about AT, AT this AT is available for 50 times while this TA is available it is available for 70 times. So, this is available within our genome suppose it is a variety 1. Ok! Now in variety 2, we are having the AT, AT this sequence here it was 50 times over here suppose this repeat is 39 times while this TA, TA this repeat is available 64 times. Suppose we have this type of variety 1 and this is the variety 2 this is our genetic constitution the variation is available within the repeat numbers.

But if you consider the region between these 2 repeats, suppose they are unique suppose this is the same, the region available between 2 repeats suppose, they are the same. So, if we use ISSR primer, if we use $(AT)_7$ ISSR primer then our $(AT)_7$ can start binding from here, or it can start binding from here over here also either here or here it will bind. So, we may get some amplification, but effectively ISSR primers will amplify this region right, it will amplify the region available, between these 2 repeats along with some parts that may come from here. But in the case of ISSR primers, first of all, we need to know the genome

sequence, we need to have the sequence information because the name itself is simple sequence repeats here the simple sequence repeat parts, will be amplified using the primer, that will be designed from the flanking region of the repeats. In SSR, we have to design the primer from the flanking regions of the repeats.

So, the SSR primer for this loci and the SSR primer, for this loci will be different because this sequence will vary this green colored arrow this sequence will vary this is one loci and this is another loci. So, suppose we have the genomic information of rice, in NCBI database. So, therefore, we can know that in chromosome number 7, maybe this repeat is available maybe in chromosome 5 these repeats are available. So, based on that by using the flanking sequence or the border sequence from each and every repeat we can design SSR primers. Ok! So, each SSR primer will amplify those particular repeats.

Now, let us see if we use the SSR for loci 1, and SSR for loci 2 how can we see some difference? Suppose, over here what is the process, just we have to isolate the genomic DNA it should be free from any garbage and those things thereafter specific SSR primers are needed for each and every loci then we have to do amplification using SSR primer. Ok! Suppose SSR 1 and SSR 2 these two primers are being used in variety 1. So, in SSR 1 what will be observed over here, the primers have been designed in this part from here. So, based on that we can get a 100 base pair product right because AT repeats are available 50 in number.

So, 100 base pair amplification will be observed once we will be using the SSR 2 or primer from this region, then what will be obtaining, will be getting a band of 140 base pair. We are talking about in variety 1, in variety 1 two different SSR primers are being used and we are getting two different amplicons. Ok! If you just recall the ISSR part, the ISSR will amplify this region, in the, in-between part along with some repeat sequence, but here distinct SSR primers are needed for each and every loci for each and every repeats. Now in variety 2, what will be obtaining SSR 1 SSR 2, in variety 2 here 39 repeats were there right in loci 1. So, here we will get a band of around 78 base pairs because 39 repeats were available, while over here in SSR 2, here 64 repeats were there, right?

So, here we will get a band of 128. So, now if we screen these two varieties, variety 1, and variety 2 using SSR 1 primer what we may get in SSR 1 over here, we are getting a 100 base pair band over here we are getting a 78 base pair band this one is 100 this one is 78 base pair. If you compare variety 1 and variety 2 using SSR 2 primers. Ok! What difference we may get over here in variety 1, we are getting 140 base pairs over here we are getting 128 base pair right, and over here it was 140. So, in this way, we can differentiate or we can distinguish different varieties using various SSR markers. It may be true that as we are discussing about variety 1, variety 2 may be once we will try with SSR 3 our banding pattern might be the same, it may amplify the similar sequences. If there no variation is there within the repeat numbers then we may not see any variation in the amplicon product also.

So, through SSR we can identify different varieties easily in a more perfect way and it is one of the more robust systems, different alleles can be distinguished easily, different allele means suppose a particular gene let us assume this is a gene A within this gene A, the AT repeats 17 is available in a particular variety. So, over here AT repeats within the gene A in another variety, AT repeat is available 7 times. It means within this allele some mutation has taken place, this is the same allele of the particular gene A, and some mutation has been taken place. Maybe in some other one $(AT)_{27}$ is available i.e., also another mutation here deletion has been taken place here duplication has been taken place. So, in this way, different alleles could be formed and if within those alleles the variation is available with the within the repeat sequence then using SSR we can easily detect it. Ok!.

So, now let us discuss about the advantages and disadvantages of SSR markers. First, it is a co-dominant marker as I was discussing earlier, we are discussing about two varieties variety 1 and variety 2. In variety 1, we have a band of 100 base pairs, in variety 2 we have

a band of 78 base pairs using SSR 1 primer. So, now suppose we have crossed these two variety we have got F_1.

So, what will be happening in F_1? One chromosome will be coming from this one parent, one chromosome will come from this parent.

So, in F_1, we are supposed to get both of these bands the 100 base pair as well as the 78-base pair, both of these bands should be obtained over here. So, this is the co-dominant marker, because if variety 1 is considered as AA and variety 2 is considered as aa. What is this amplification, is this is a different allele of that particular gene right? So, let us assume in variety 1 we had the genotype AA, in variety 2 we had aa. So, in F_1 we are having Aa, one chromosome will come from this parent one will come from that parent.

So, the capital A allele and small a allele both are showing its impact or its presence in F_1. So, i.e., a co-dominant marker through which we can easily identify the true hybrid. Ok! So, it displays a high level of polymorphisms, the results are highly reproducible to then SSR marker, and can be more easily transferred across related species facilitating comparative genomic studies. Ok! So, if you try to understand this particular thing means it can be transferred across related species. So, basically within a related species the number of repeats may change easily, but the regions available in the flanking part of a repeat they generally try to be unique, means those sequences are not too much changed.

So, it can be easily transferred across the species, suppose, someone has designed some SSR primers for mung bean, you know that mung is a pulse crop. So, those SSRs could be utilized in lathyrus, in lentils, in another pulse crop easily, which means some of the SSRs might work over there then the technique can be applied across various organisms also and what are the disadvantages? First of all, costly, and time-consuming process, why costly because we have to have the sequence information, until and unless we know the flanking sequence of repeats, we cannot design the primer. So, we need the sequence information there are a lot of costs involved there. Then nontransferable between distantly related species, it is obvious that a limited portion of the genome can be covered. Suppose in ISSR using a single ISSR primer, if that repeat is available multiple times in forward

and reverse orientation within a genome we can get different amplification, but in case of SSR for each and every, SSR only single loci will be targeted. Ok!

So, a limited portion of the genome can be covered then the analysis of data is complex, once we have to screen a large number of plants using a large number of SSRs then the data analysis becomes complex and the marker is sensitive to mutation basically, based on mutation it works. So, suppose we are trying to play with a couple of rice varieties variety 1, variety 2, and variety 3 using a particular SSR, suppose in variety 1 and variety 2 we are getting the SSR amplicon of 100 base pair, and 78 base pair. While in variety 3 we are not getting any product that may occur because within the flanking sequence if some mutation occurs within this variety the primer may not bind properly. Ok! So, these are the advantages and disadvantages of SSR. Now we will be discussing another thing i.e., CAPS marker is the cleaved amplified polymorphic sequence.

So, the polymorphic sequence is there means, the sequence must have to be polymorphic and the amplified word is available. Ok! So, first of all, we have to do some amplification if we have to use this particular marker then we have to do cleavage and we have to analyze the sequences. So, suppose we have 3 different varieties, we have 3 different rice varieties. So, from each of these rice varieties let us assume we are trying to characterize a particular gene *OsCaM1* i.e., Oryza sativa (Os) calmodulin 1 gene we are trying to characterize. We are trying to characterize this particular gene because this gene is found to be associated with different stress salt stress, drought stress.

So, we are trying to characterize this particular gene sequence from these 3 different rice varieties. So, this is the OsCaM sequence available in 3 different rice varieties. So, how can we amplify the gene? Definitely, we need to design the primers of this particular gene from NCBI database or from other literature. So, based on that we have designed the primers and we have amplified this particular gene from 3 different varieties. And once we have run the gel, agarose gel after this amplification in variety 1, variety 2, and variety 3 we are getting a band of 2.

5 kb. Suppose, we have amplified the total genic part of this particular gene means including exon, and intron some parts of UTR are available and we have designed the primers, we have amplified the 2.5 kb region. Now out of these 3 rice varieties, suppose 1 is Swarna I am just giving an example. Suppose 1 is Swarna it is a mega rice variety i.e., popularly grown in most of the part of India. One is Sonamasuri i.e., another popular rice variety of India and another one is Pokkali.

So, these 2 varieties the Swarna and Sonamasuri those are mostly grown in most of the conditions, while, Pokkali is suitable for salt tolerant condition. Suppose under extreme salt conditions where EC is more than 5 there the Pokkali can grow easily. Ok! So, initially once we have isolated the genomic DNA, we did not find any difference. Once we have done the sequencing of these 3 PCR products, we found that in variety 1 over here, and over here, 2 *Eco*RI restriction sites are available. While in variety 2 over here and over here 2 *Eco*RI restriction sites are available at the same position.

While in variety 3 only this restriction site is available while the restriction site over here is not available after sequencing, we have confirmed it. So, now suppose we have done this PCR product again, and thereafter we have digested the PCR product of these 3 different rice varieties using *Eco*RI. So, in variety 1, variety 2 and variety 3 what may be the scenario.

Suppose this band is 1 kb, this band is 1.5 kb, sorry, this band is 0.75 kb this is 1.5 kb and this is 0.25 kb. Ok! So, in variety 1 we will get 3 different bands after digestion. So, we have to do PCR and then digestion right?

Let us see, what we will see over here, we will get a 1.5 kb band, 0.75 kb band, and 0.25 kb band. In variety 2, also, we will be getting these 3 bands while in variety 3 we will be getting 0.25 kb band, and 1.5 plus 0.75 means, 2.25 kb band here the band size will be larger these 2 bands will not be available over here.

So, in this way now if we try to attempt some crossing between Sonamasuri and Pokkali and if we have to analyze the F_1, then easily, we can identify the F_1, because in F_1, these bands will be available as well as these bands will be available. So, in this way first, we need to identify some sequences, we have to design the primers, to amplify them after sequencing if we find any restriction site variations then using that restriction site, we can develop this type of markers, cleaved amplified polymorphic sequences means once we will cleave the amplified sequence then we can see the polymorphism.

So, these are the advantages and disadvantages, first of all, precise detection of genetic variation through PCR amplification and cleavage analysis can be employed for genome-wide diversity studies, because for different regions if we do some sequence analysis, we can identify some restriction-based variation and we can utilize it. High specificity, robust tools for detecting single nucleotide polymorphism and other genetic variations, then it is highly accurate in terms of different results because it is a restriction enzyme based right? Once the PCR will be done then through restriction enzyme, we must have to get those specific bands with no requirement for specialized equipment.

What are the disadvantages, first dependency on restriction enzymes. So, a little bit of cost will be involved there, then the inability to detect mutations like insertion and deletions, if a small insertion or deletion takes place, a couple of bases 3 to 4 bases we cannot detect it because based on digested product we can see the difference. Ok! The limited resolution in detecting multiple nucleotide changes in close proximity if the nucleotide changes, but the fragment size remains the same we may not detect the variation. Then labor intensive, and time-consuming, and limited multiplexing because a lot of enzymes could be utilized the sequencing has to be done. So, multiplexing is not done properly. So, these are the things for today's class these are the references you guys can go through these journal papers. Thank you.