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-29: RAPD

Welcome back. So, we will continue again. So, what are the advantages of RFLP? First of all, prior knowledge of DNA sequence is not required. We have just taken the probe from rice in our previous example, we have just taken the actin specific probe from rice. We did not have any information, any genetic information or DNA information of the touch-me-not plant, right? So, prior knowledge of DNA sequence is not required.

Next one, it helps in study of genetic diversity and polymorphism. Definitely the polymorphism could be detected, the genetic diversity could be analyzed through this. Ok! Then it may act as co-dominant marker, if the specific restriction site, restriction enzyme site is mutated. So, that is a new term coming co-dominant marker.

Just let us try to start our previous discussion. Suppose, we had the Indian plant and European plant. In Indian plant we are getting a band at around 7 kb, in European one we are getting a band at around 8 kb. Now, if you just think about the chromosomal constitution, what will be the Indian chromosomal constitution? Suppose, this is the chromosome number 2 we are talking about. Ok! In Indian case an *Eco*RI site was available over here.

These are homologous chromosome… in chromosome 2 of Indian origin. While, in case of the homologous chromosome of European origin, the *Eco*RI site was mutated, right! This one was the scenario. Now, suppose we have attempted some cross between these two

plants, the Indian plant and European plant, this two has been crossed. So, if we cross them what will be happening in $F_1$? One chromosome will come from here, sorry, one chromosome will come from here and one chromosome will come from here in $F_1$. So, this will be the chromosomal constitution of the $F_1$ right!

In one chromosome which is coming from the Indian origin, here *Eco* site is available, while in another one *Eco* site is not available. So, suppose in the next generation we have got the $F_1$ again, we are planning to do the Southern hybridization. There, the Indian sample is available, European sample is available and $F_1$ sample is available. In Indian sample, will be having the band of 7 kb, in European sample we are having the band of 8 kb, while in $F_1$ both the chromosomes will be there. So, both of these bands will be highlighted right, the 8 kb, as well as 7 kb, this is the co-dominant nature of the marker.

So, if the restriction specific restriction enzyme site is mutated among the targeted individuals, then it can be easily used as a co-dominant marker. Then, it may provide insights into epigenetic modifications since some restriction enzymes show methylation sensitivity that is a very important thing. Ok! So, once we will discuss about the epigenetics, then we will be discussing about different modifications, different DNA modifications. So, some of the restriction enzymes might be non, means some restriction enzymes may not work properly due to DNA methylation. Ok! So, if epigenetic modification is taken place between two varieties and if that modification is associated with some restriction enzyme, then also we can see variation through RFLP.

Then it helps to analyze the number of transgene integration copies, transgene integration copies in polyploid species and transgenic lines that is very important things. Once we will be discussing about the agricultural biotechnology part, integration of transgene within the plant, those parts then we will be discussing these things or related problems once again in detail. So, it is widely used in DNA fingerprinting. So far, whatever we have discussed, in RFLP we have used a gene specific probe. We know that in eukaryotes, a lot of repetitive DNA sequences are there, right!

So, if you use a repetitive DNA sequence as a probe then we can see specific banding pattern between the plants of Indian origin and the plants of European origin. Their banding pattern may be changed because the repetitive sequence will be available number of times, right, and the results are reproducible that is another important advantage of RFLP. Now, what are the disadvantages of RFLP? First of all, high quality and significant amount of DNA is required. I was telling earlier for doing RFLP for doing Southern hybridization at least 10 microgram of genomic DNA sample is required for each of the sample. Ok! And the DNA quality should be good, because if the DNA quality is not good, if the DNA is sheared already then, a number of bands could be obtained right! means we may not see the clear picture.

Then it depends on the development of specific probe or probe libraries, definitely we need to have some probe or probe library, otherwise we cannot do RFLP. Until and unless we have any related probe, like as we have used the actin specific probe or we can use some GAPDH specific probe, until and unless we have any specific probe or probe library, we cannot play with RFLP. Next one, very important radioactive probes are used in most of the cases in RFLP the radioactive probes are used. Nowadays, the non-rad methods have been discovered, it is being followed in different laboratories also, but radioactive probes were mostly being used in RFLP and that is hazardous to health. Then, it is time consuming, expensive and labor intensive, a lot of training is needed to do RFLP analysis and it is expensive also, a lot of chemicals are needed, different sets of buffers have to be prepared. Ok!

And another important thing, low level of polymorphism why is it so, because here we are dealing with a specific probe. If that specific probe is available 2 times, 3 times or 5 times within the genome then we may see variation among those 5.. among those 5 numbers. Ok! And limited resolution and difficulty in multiplexing as different steps are there, multiplexing means different plants, their DNA will be isolated together, gel will be run properly, it will be transferred to membrane, hybridization will be done, all things could not be done easily. So, each and every probe has to be used separately and the option of multiplexing is not at all possible in RFLP. Now, coming to another molecular marker that

was very popular, initially that is RAPD, Random Amplified Polymorphic DNA.

So, earlier one, we have discussed RFLP that one was the hybridization based, now we are coming to the PCR based molecular marker; RAPD is a PCR based molecular marker. So, few salient features of RAPD; first of all, in RAPD, in PCR, primers are needed right! That thing is known to us, here basically 10 mer primers or 10 nucleotide long primers are used with GC 55 to 60% and generally, but at least 45 % GC is must, at least 45 % GC should be there in RAPD primer. 10 mer primer, means 10 nucleotide long primer sequence should be used. Next one, 45 to 50℃ is used as annealing temperature as the primer length is small. So, it's annealing temperature will be also small, generally 45 to 50℃ is used as annealing temperature.

Now, primers are random as the name itself random amplified polymorphic DNA. So, random primers are used, means, we can just tell A T C C G C A C T A, it may be RAPD primer 10 mer are there 3, 3,3, 10 mer are there. So, in this way, we can design RAPD primers means some companies have already developed their libraries, means different sets of RAPD primers are available in those libraries. So, if anyone is willing to do the molecular marker related work using RAPD, he can order from there. Then another important thing, a single primer is used for detection that is true for RAPD molecular marker.

Now, a lot of things are coming primers, then polymorphism, PCR. So, before discussing RAPD, we need to discuss how PCR reaction is done, PCR is polymerase chain reaction. Through polymerase chain reaction, suppose within a genome, through this particular reaction, we can amplify a specific region from the genome, we can amplify a specific region from the genome. So, let us assume that this is our genomic DNA, double stranded genomic DNA, this is the 5' to 3' end and this is the 3' to 5' end. Let us assume, only one chromosome is available in a particular crop.

So, in this chromosome, we are trying to amplify this particular region, suppose, we are trying to amplify this particular region. Ok! So, using PCR or polymerase chain reaction,

we can amplify this particular region. How, first of all in PCR 3 steps are there, first one the first step is denaturation, next one is annealing and third one is extension. These 3 steps are there denaturation, annealing and extension. What are the things needed to start a PCR? First, to start PCR, what are the things needed, I am talking about generalized PCR not RAPD.

To start normal PCR, first we need template DNA, we need Taq DNA polymerase, basically Taq DNA polymerase is an enzyme, it is a DNA polymerase enzyme, using this enzyme, from one strand of DNA, complementary strand is being produced. And it was initially isolated from *Thermus aquaticus*, a particular bacterium, available in hot spring. So, the name Taq is given isolated from *Thermus aquaticus*. And the scientist found that this bacterium can replicate itself easily within the hot spring the temperature close to 100°C. It means, this enzyme it's DNA has to be amplified.

So, it's polymerase enzyme will be highly suitable for such type of reaction. So, this enzyme is basically used. Then template DNA, Taq DNA polymerase, then we have to use specific buffer for conducting PCR, then we need primers. In most of the cases, we have to use forward and reverse primer for conducting PCR to amplify a specific region of the DNA. We need a forward primer and a reverse primer, later on, I will explain once again.

Then we need dNTPs, we know that on DNA, ATGC different bases are there ATGC. So, until and unless we supply these bases from one strand, how another strand will be produced. For incorporation of this bases, dNTPs has to be added and next, we need the $Mg^{+2}$ ion. So, those things are basically needed to start a PCR reaction. So, I have mentioned about three steps initially denaturation, then annealing and third is extension.

So, how the PCR reaction is taken place? First during denaturation, the double stranded DNA is denatured into single stranded DNA. Let us assume this is the 5' to 3' strand and this is the 3' to 5' strand. The DNA has been single stranded now. So, denaturation is done at 96°C generally at 96°C. I will tell about the temperature and time period later on once again.

Then, our next step is annealing, during this step, anneal means something is attaching with something else. So, over here, these two primers, forward primers and reverse primer, they bind on the single stranded DNA based on the sequence complementarity. Ok! So, primers are single stranded DNA, they will bind with another single stranded DNA which has been denatured just earlier. So, over here, in this way, a primer can bind because primers will be having 5' to 3' direction, it is a single stranded DNA molecule and a 3' OH should be available at the last base. While, in this strand also, another primer may bind based on the sequence complementarity that is the annealing step.

And third one, is extension step, during extension step what happens DNA polymerase come over here and one by one bases will be added means different dNTPs will be added, and in this way, the 5' to 3' extension will be done of this strand and 5' to 3' extension will be done of this strand. Ok! In this way, from one double stranded DNA after completion of one PCR cycle, two double stranded DNA will be produced. Ok! And, if you see carefully, our target region is being amplified, in this way, if we do PCR for 30 cycles or 35 cycles, this specific region, our target region for which or specific to which we have designed the forward primer and reverse primer, that region will be amplified enormously and we can see it on the gel through agarose gel electrophoresis. So, briefly, I am mentioning the final setup of a PCR reaction. First, we have to have a initial denaturation, initial denaturation is done at around 96°C, it is done for 3 minutes almost.

Then again, we have to do denaturation at 96°C for 30 seconds then annealing, annealing temperature will depend on the primer sequence. If the primer sequence is smaller, then, we have to use less annealing temperature. If the primer sequence is larger or if more GC is available, G or C is available within the primer our annealing temperature will be higher. You know because in AT, two hydrogen bonds are available while in GC, three hydrogen bonds are available. So, to break this bond more temperature is needed, to break AT bond less temperature is needed.

So, in this way, the annealing temperature depends on the primer sequence. So, generally

45 to 70°C could be used as annealing temperature for different PCR reactions and it is generally done for 40 seconds. Thereafter extension, extension is done at 72°C for 1 minute, 2 minutes, it depends on the length of the amplicon size. Our length of this amplicon size if the length is 1 kb, then 1 minute extension is fine because the processivity of DNA polymerase or Taq DNA polymerase is generally like that, in 1 minute it can amplify 1 kb region. If the size is more, we have to increase the temperature of extension.

So, these three things: denaturation, annealing and extension; these three things are repeated for 30 to 35 cycles during each PCR. And thereafter, the final extension step is there that is done at 72°C for 7 minutes. If any bases are not added during the PCR reaction, during those 30 cycles those things, those small bases could be incorporated during this final extension process. So, this is the overall PCR process. So, now, we will gradually move into the RAPD part once again.

In RAPD, it is a PCR based molecular marker, Random Amplified Polymorphic DNA. Here, single primer is used; in normal PCR, two primers are used while, in case of RAPD, single primer is used that is of 10 mer means 10 nucleotide long. Its annealing temperature is less because the primer length is smaller, and primers are random in nature. So, let us see how does RAPD work?

So, suppose this is our genomic DNA. Ok! Suppose, within this genomic DNA a sequence we have ATGG CATA GC. Suppose, it is a sequence it is available on the DNA. So, it is complementary this is the 5' to 3' standard the DNA and let us assume, this is the 3' to 5' standard of the DNA. So, in its complementary strand, sequence will be TA CC GT AT CG right! Now, suppose over here, another sequence is available ATGG CATAGC and in its opposite strand TA CC GTAT CG.

While, in another place on the genome, this sequence is available TA CC GTAT CG. So, in its opposite strand, the sequence will be ATGG CATAGC right! So, now suppose this is the constitution of a particular genome. Let us assume, we are talking about a single chromosome, a single chromosome is available, within that chromosome somewhere, this

sequence is available, somewhere this sequence is available and somewhere this sequence is available. If you carefully see this sequence and this sequence are same, right!

So, suppose we are trying to plan, we are trying to do RAPD and in RAPD, we are using this single primer ATGG CATAGC. Suppose, this primer we are using in RAPD sorry. So, we are planning to do PCR, during denaturation process, the double stranded DNA will be denatured into single stranded this is the 5' to 3' strand and this is the 3' to 5' strand. So, here, we had ATGG CATAGC here, we had ATGGCATAGC here, we had TACC GTA TCG.

Here, we had TACCGTATCG, TACCGTATCG we had ATGG CATAGC. Ok! So, denaturation has been taken place, now our next step is annealing, in annealing step this specific primer will try to find out its complementary region, it will try to bind out its complementary region this is the 5' to 3' end, the primer is single stranded in nature and it has 3' OH group also. So, that primer can bind over here if you see ATGG CATAGC this is the 5' to 3' end the primer can easily bind over here, it can start amplification right! Another primer can bind over here, and it can start amplification, but what will be happening until and unless our amplification is not done, in this way, we cannot see a amplified product right! If the amplification is taken place in single direction taking a single stranded DNA, we will not see $2^n$ type of amplification. So, if you carefully see about this sequence, here, can our primer bind? Our primer cannot bind over here also.

So, for properly doing RAPD, we have to have this specific sequence in forward and reverse orientation. So, this sequence has to be placed in forward orientation as well as in reverse orientation. So, let us try to think about what will be the sequence in reverse orientation. Suppose, I am extending the DNA part ok! suppose I am extending the DNA part somewhere G C T A T G C C A T. Suppose, somewhere on the genome, this sequence was available and its complementary strand, the sequence was C G A T A C G G T A. Ok!

This is, this strand the 3' end and this is the 5' end, the extension of this strand. So, if you see carefully over here, this particular primer can bind, what is this sequence from 5' to 3',

its sequence was A T G G then C A T then A G C. So, the primer can bind over here and herefrom, it can start amplification. So, basically through RAPD we can get the amplicon from this whole region.

Our primer may bind, single stranded primer. So, it can bind over here, it can bind over here. So, we have to remember the RAPD will work if the single stranded primer binding site is available in forward and reverse orientation. So, the sequence is available in forward orientation and reverse orientation in between a span of amplifiable distance. The distance should be amplifiable upto 2 kb to 3 kb maximum we can consider if the sequence is available in reverse orientation, beyond 5 kb it will not be amplified through RAPD. So, this is the process through which a single primer can do the amplification and based on that, the polymorphism could be detected.

Now, these are the advantages and disadvantages of RAPD what are the advantages? First, small amount of DNA is required as PCR reaction is done, small amount of DNA generally 50 nanogram of DNA is needed per PCR reaction, even less than of that can be used also. Now, it can be automated, means, a number of samples could be amplified together within a single PCR machine using multiplexing, we can do, we can set the reaction of a number of reactions together, then once optimized, then it is very less laborious and convenient. Because in RFLP, you have to do different processes; hybridization, then gel preparation, gel transfer, here it is a very simple process, in 3 to 4 hours your process will be over, you can see whether they are different or not, whether they are polymorphic or not, less time consuming and no hazardous substances are used, mostly like radioactive which is used in RFLP. Now, disadvantages; first, lack of reproducibility here, the reproducibility is very less. Suppose, different laboratories may get different banding pattern in ….., in RAPD means random amplified polymorphic DNA, why? Because, if their DNA isolation process is different, their PCR master mix composition is different, then, they may get different results because it depends on the binding of only 10 nucleotide DNA.

So, those binding may be hampered it, ok, mostly act as a dominant marker means, it cannot discriminate capital A, capital A with capital A, small a, the dominant homozygote

could not be distinguished with heterozygote individuals using this. Then, homology among siblings or similar types of plants may not be obtained through this one. Ok! Then, absence of specificity, here specificity is very less as it is not reproducible in different labs due to the difference in DNA isolation procedure, different PCR master mix composition used. Ok! So, the specificity is very less and the primers are random. Ok! If some parts can bind properly, the Tm is less, it can start amplification.

So, we may get different types of result in different times. Then, parental bands from which progeny is obtained cannot be identified, earlier we have told that the Indian origin and European origin, those two plants if crossed together in $F_1$, we can tell from which parent which band is coming, but here the parental bands from which the progeny is obtained cannot be identified. So, these are the references of this particular class. Thank you.