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: 08

Lecture-32: Backcross breeding through molecular marker (part-I)

Hello everybody. Welcome to SWAYAM NPTEL online course on Basics of Crop Breeding and Plant Biotechnology. So, today we will discuss that is on Marker Assisted Selection in Plants. And, in this lecture mostly we will be discussing about, the back cross breeding through the molecular marker. How molecular markers could be used in different breeding strategies, especially in the back cross-breeding will be discussed over here. So, these are the concepts that will be covered in this particular under this particular topic.

First of all, the suitable marker for marker-assisted selection will be discussed means which type of marker will be mostly suitable, and mostly useful for marker-assisted selection or marker assisted breeding. Then we will be discussing, about the application of molecular markers in hybrid detection. How can you tell that if two plants are crossed the hybrid is a true hybrid, or it is just the self-seed from the female plant? So, using molecular markers, how can you know that these things will be discussed?

Then we will be discussing about foreground selection and background selection, along with the recombinant selection process. Then the integration of disease-resistance genes in plants will be discussed. Then the marker-assisted selection process, overall will be discussed under this particular topic. So, first of all, we will be discussing about the most suitable marker for marker-assisted selection. So, far we have discussed about various molecular markers like RFLP, RAPD, ISSR, SSR marker, then SNP, AFLP different

types of markers have been discussed, EST those different types of markers have been discussed.

So suppose we are planning to do marker-assisted selection, if we identify some genic molecular markers that will be highly useful. So, what is a genic molecular marker? Genic molecular markers are mostly of two types. They basically, target the genic region, within the DNA, within the chromosome. In the chromosome different regions are there centromere region, and telomere region. In the centromere and telomere regions, mostly the repetitive sequences are available.

Even throughout the chromosome, at different positions the repetitive sequences are available. So, they are from SSR markers that could be designed easily SSR, ISSR types of markers, mostly can work over there. And, if you think about the chromosome constitution of different plants, then the non-genic part is maximum. The chromosomal part, the chromosomal segment available between two genes is huge, while the genic part is very less. So, if we identify molecular markers from the genic region then it could be easily transmitted with the target gene.

If our gene is transmitted from one plant to the next plant, the genetic marker that is available close to that gene or within that gene that marker will be transferred also. So, our selection process will be better. So other than genic molecular markers, different random DNA-based markers are available. Those are available randomly on the chromosome. It does not depend on the gene part.

It may be available in the repetitive sequence, it may be available within two genes in those regions, and it may be available within the gene also, but genic molecular marker is exclusively present within the gene. So there are two types; one is gene targeted marker, that is GTM and another one is functional marker it is commonly known as FM. Now, if you see the gene structure of a eukaryotic gene. So, in eukaryotic genes basically, if this is the CDS the coding DNA sequence of the gene, then it will have a 5′ UTR, it will have a 3′ UTR region and with that for expression of that gene, definitely a promoter region is

needed. So, these things are associated with most of the eukaryotic genes for each and every gene, in most of the cases single promoter is there.

So, let us see how the gene-targeted markers and functional markers work. So, if we identified some marker which is available within the CDS region the coding DNA sequence region then it is known as functional marker. Suppose a gene had ATG and other different sequences at last sequence was TAA within the CDS, which means this is the start codon and this is the stop codon of the gene. While in another allele, the gene has some mutation, has been occurred within the CDS. So, the sequence has been ACG different bases are there, and at last TAA is available the stop codon.

So, if you see about this allele suppose, this one was the capital 'A' allele and this one was the small 'a' allele. Ok! So, over here as the ATG is not available, the start codon is not available the protein formation is stopped, is hampered for this allele. So, this type of marker if we can develop i.e., totally dependent on the coding DNA sequence, then it is known as a functional marker. Ok! Means over here the gene may not function, but over here the gene will function properly. Ok! So, the marker is directly associated with the function of the gene.

Now the other parts are there within the pre mRNA, where the 5′ UTR will be there the 3 ′ UTR will be there. If you think about the total genic part of a particular gene, the promoter will be available. If you identify some molecular marker from that region from the promoter, UTR, 5′ UTR, or 3′ UTR part, those are known as the gene targeted marker means those are tagged with a particular gene. So, there is also a high chance that once this gene is transferred this marker will be also transferred with the gene, because this is part of this gene, right? So, in this way, the most suitable marker for marker-assisted selection is a functional marker and thereafter next suitable one is the gene-targeted marker, but if these two things are not available still, we can do marker-assisted selection using other different molecular markers.

So, now we will be discussing about the application of molecular markers in hybrid

detection. So, I think you guys know, that how hybrids are formed by crossing two different plants right? So, two different plants have different genotypes if the genotype is the same then it will not be a hybrid. Ok! Although one chromosome will come from the maternal plant, one chromosome will come from the paternal plant, but it will not be a proper hybrid because if the genetic constitution is the same, no variation will be available in the F_1 or no special type of genetic combination will be obtained there. So, let us try to discuss, about the hybrid development by crossing two different plants and then how molecular markers could be used to detect the true hybrid.

Suppose, we have a particular plant P_1, one pair of chromosomes, which is available in the P_1 plant, and in the P_2 plant also one pair of chromosomes is available. Suppose, we have another plant that is P_3, these are three different varieties of a particular crop, let us assume. So, all the plants we are assuming they are in homozygous condition. In the P_1 plant, you have seen this chromosome suppose different genes are there, capital A, capital B, capital C, and capital D. In its homologous chromosome, the same gene sequence is available, because I am telling one pair of a chromosome is available in each of the plants, we are assuming and all the plants are homozygous in nature, for each and every gene it is in homozygous condition.

In P_2, suppose we had capital A, small b, small c, and capital D this is the genetic constitution of this plant while in P_3, suppose the genetic constitution is capital A capital B small c and small d. So, if we see carefully these three plants the variation is there in gene content. So, if we cross them, we can get a hybrid suppose if we cross these two plants. So, one chromosome will come from here and one chromosome will come from here in the F_1 right. So, in F_1 we will be having capital A capital B capital C and capital D and from here we will be getting capital A small b small c and capital D.

For these two gene, we can see earlier all the chromosomes were homozygous means these two genes were homozygous in parent 1 as well as in parent 2, while over here for these two genes, B and C gene we can see heterozygosity. Ok! So, in this way, the hybrid is formed. So, now through molecular markers, how can we trace it? Ok! Let us start

that discussion. So, for this discussion, I am considering only two plants.

Let us think one plant is having, capital A capital B small r and capital D; capital A capital B small r and capital D. So, it is P_1, parent 1, and we have P_2, parent 2, where the scenario is like small a small b capital R and small d both the plants are homozygous over here also. Suppose, if we are trying to attempt some hybrid development definitely some purpose will be there. Suppose, A gene is responsible for the number of tillers in a plant. These are the genetic constitutions of some rice varieties maybe this is variety 1, this is variety 2. Ok!

Where A gene is responsible for tiller number, capital A or small a. If capital A is there then the tiller number will be more, if small a allele is there the tiller number will be less, just assume. Suppose, the B gene is responsible for the panicle length, if capital B is there we can get a longer panicle, and a larger panicle, if a small b allele is there we will see a shorter panicle. Suppose the R gene is responsible for resistance to a particular disease. Capital R, is responsible for the resistance, and small r is responsible for susceptibility.

This is the resistance to bacterial leaf light, and suppose the D gene is responsible for the grain weight. If the capital D allele is there the grain weight is more, and the 100 seed weight is more, if the small d allele is there the 100 seed weight is less. So now, if you carefully see these two plants P_1 and P_2. In P_1, most of the better traits are there it is having a high tiller number, it can produce more, number of tillers, the panicle length is higher in this plant, and the grain weight is also higher in this plant, but it is susceptible to a particular disease right. Similarly, if you think about P_2 it is an inferior variety, which means most of the traits are not suitable not agronomically suitable not accepted by the consumers, but it has some resistant genes.

If you think about the back cross or other different crosses, breeding methods that have been discussed earlier, there you can recall in some of the wild germplasm, and wild rice these types of resistant genes are available. Ok! So, just assume it for most of the traits it

is not up to the mark, but some resistant allele is there. So, if we make the hybrid, then this resistant allele will come in the F_1 ok because what will be happening, what will be the F_1, if we cross these two plants one chromosome will come, and from here this parent and one chromosome will come from this parent, capital A capital B small r capital D, and here small a small b capital R and small d. So, for all the genes we are getting heterozygote condition and as the capital R allele is coming over here, we may see some type of tolerance in this case. So, now our target is to identify some molecular marker using that we can detect this hybrid. Ok!

So, A B R D, these are different genes and for each and every gene some specific DNA is available. Now what is a molecular marker, those are some specific DNA sequences that are available on the genome, and we can easily detect it and we can trace it also from which parent it is going to which progeny that is the purpose of molecular marker. Now, suppose we have identified some SSR markers we will be discussing, about some SSR marker. So, for doing this type of analysis first we need to identify any specific markers, that are highly tagged with this particular gene because we are considering about 4 genes over here A B R and D. So, first, we need to identify molecular markers for each and every gene, if we can identify a genic marker that would be best otherwise, we have to identify some other markers that are available in close proximity of that gene.

Now, let us assume that capital A. I am discussing this part let us assume the capital A allele it has ATG repeat, within or within the promoter region or within the UTR region ATG repeat is available 30 times. While in small a allele it is a different form alternate form of this particular gene there ATG repeat is available 20 times. So, it is a molecular marker because this simple sequence ATG, it is a tri-nucleotide sequence it is repeated 30 times in the case of capital A allele while it is repeated 20 times in the case of small a allele. Ok! Now, if we just think about suppose we have designed the SSR primer by considering this repeat sequence, how we have to design SSR primer we need to design the primer from the flanking part of this repeats right? So, if we design this primer suppose, SSR1 has been designed for A gene and we are analyzing P_1 and P_2, P_1 from here, and P_2 from here. Ok!

Suppose, we are analyzing, we have isolated the genomic DNA, we have used the SSR1 primers from the flanking region and we are amplifying these 2 plants, these 2 plants genomic DNA. Over here in $_{P1}$, as capital A is there, we should get a band at 90 base pair right because ATG repeat is available 30 times. So, at least 90 or more than 90 base pairs will be available over there, while where small a allele is there here band will be close to 60 base pair. This is a kind of polymorphism means this molecular marker is showing different banding patterns in these 2 plants. So, using this type of polymorphism we can detect hybrid.

We are discussing the application of molecular markers in hybrid detection. So, in this way, if we identify any polymorphic marker among the parents, that are being used in the crossing program, then we can detect the hybrid. So, suppose we have taken the P_1 plant, we have taken the P_2 plant, and we have attempted some cross and we have generated some seeds. We are assuming that those seeds are supposed to be hybrid seeds. Ok? Suppose cross number 1, cross number 2, cross number 3, cross number 4, different crosses have been made and this is the P_1 and this is the P_2 plants.

Once we get the putative F_1 lines, we need to isolate the genomic DNA from those plants at the seedling stage. Then using this molecular marker, we can screen them. So, in P_1 we are supposed to get 90 base pair bands, in P_2 we are supposed to get 60 base pair bands. Suppose in line number 1, we are getting only this band, in line number 2 we are getting these 2 bands, in line number 3 we are getting these bands, in line number 4 we are getting these 2 bands. So, here we can tell that a true hybrid has been formed because if these 2 chromosomes come together in F_1, the capital A allele will be there as well as small a allele will be there.

So, we should get both of these bands. In this way, if we identify any polymorphic molecular marker, we can easily detect the F_1 using that particular marker. Now, coming to the foreground selection and background selection along with the recombination selection. Ok! So, to discuss that, I need to draw the chromosome structure once again for

$_{P1}$ and P_2. Suppose this is the P_1 we had and this is P_2. In P_1 we had capital A capital B small r and capital D.

While in P2 we had small A small b capital R and small d. For your understanding let us make some shade in parent 2 so, that our discussion will be better. This means you can easily trace that the green part is coming from parent 2. And I am drawing another thing this is the capital R gene that is responsible for the resistance to sheath blight of rice. So, if you think about this particular cross what is our target? Our target is to generate a particular plant where this resistant trait will be coming.

So, if you recall this type of thing is done through back cross-breeding. Because we have one variety that is agronomically superior and highly accepted by the farmers by the consumers it has very good yield attributing traits, but it is lacking for some traits some traits are lacking, which means it is susceptible to a particular disease. So, we are trying to transfer this resistant gene from another plant through a backcross breeding method. So, in back cross breeding method first we need to identify the recipient parent and the donor parent right. So, from the donor parent here we are trying to transfer the capital R gene into the genome of the recipient parent.

So, if we cross them first, we will get the F_1, in F_1 this type of thing will occur, one chromosome will come from this parent and one chromosome will come from this parent in F_1. So, its constitution will be capital A capital B small r capital D, and it will be small a small b capital R and small d. This chromosome is coming from parent 2, I am shading it and this is our target gene the resistant allele that is supposed to come in F_1. Now let us start discussion, about the foreground and background selection over here what is our target to transfer the resistant gene? So, our foreground selection is done based on the target trait.

So, here target trait is transferring the resistant gene. So, for capital R gene we need to identify the polymorphic marker means that is either from the genic region of R gene or it is very closely associated with the R gene. So, suppose here we had capital A small one

set of gene capital B small b capital C sorry capital R and small r and capital D and small d. We are discussing about 4 different genes over here. Suppose for A gene the ATG sequence is repeated for 30 times, in capital A allele while over here ATG sequence is repeated for 20 times.

Suppose for the B gene, we found another molecular marker that is AT repeats for 50 times, while in its another allele that is also polymorphic here, AT repeat is available 30 times. For R gene the resistant gene that is the trait for which we are selecting. Ok! Suppose if capital R is there means those plants which is showing resistance to the sheath blight of rice or to the bacterial leaf blight. Suppose, over here TAA this sequence is available 30 times in the resistant one while in the susceptible one the TAA sequence is available 10 times close to the promoter region while or it is close to the 3′ UTR region. While, in the D gene, we have the CT sequence it is another repeat that is available for 60 times while if small d allele is there CT sequence is available for 80 times.

Suppose, for all these genes we have identified some polymorphism. Ok! Then using these markers how can I do foreground selection and background selection we will be discussing over here. So, first, we have identified P_1, and P_2 we have isolated the genomic DNA from them, and using these molecular markers we found that for each and every gene some diversity is available between P_1 and P2 then we got the F_1 also. So, suppose we are making gel pictures for each and every SSR. We are doing SSR1, we are using SSR2 for the B gene, we are doing SSR3 for the R gene and we are doing we are analyzing SSR4 for the D gene. Ok!

This is for the A gene, this is for the B gene, this is for the R gene and this is for the D gene. Ok! So, we had P_1 and P2 we have found F_1. So, in P_1 what should be the banding pattern using this SSR1, ATG repeat is available 30 times, which means here we should get a 90 base pair band and here we should get a 60 base pair band right? While in F_1 if it is a true F_1, we should get both of these bands. In SSR2 for B genes, if capital B is there 50, AT repeats is available means we should get a band at 100 base pairs.

While, over here 30 repeats are available if small b allele is there this is P_1, this is P_2. So, here we will get a band of 60 base pairs. If F_1 is a true F_1 both of these bands should be available for SSR 3 that is the important gene for which we are planning to do the back cross-breeding. Now, here in parent 1 and parent 2 what will be the banding pattern? In parent 1 if we carefully see the small r allele was there, right? So, a 30 base pair band will be there while, in parent 2 a 90 base pair band will be available and if the F_1 is true, F_1 both of these bands should be available over there.

In the case of SSR 4 in parent 1, parent 2, and F_1 what will be the scenario? In parent 1 capital D was there.. sorry in parent 1 capital D was there. So, the 120 base pair band is supposed to be there, while in parent 2 the repeat is 80, which means the 160 base pair band will be there right? If the F_1 is true 160 and 120 both bands should be visible in F_1. So, this is the initial scenario. So, through foreground selection, first, we need to identify for our target gene.

Suppose, we have F_1 plants, first we need to identify whether our desirable capital R allele has come in F_1 or not. Ok! Then if you recall the back cross breeding method, what happens? first 2 plants are crossed we get F_1, then F_1 is crossed with the recurrent parent again, and again right? F_1 is crossed with the recurrent parent. Then, we will get some individuals that will be crossed with the recurrent parent once again. In this way, after a couple of generations, we have to do backcrossing and then we need to do selfing, thereafter we have to do selfing and finally, we will get to develop an isogenic line that will be almost similar to P_1, but our targeted R gene, capital R gene will be transferred over there that is our target right. So, not only in the F_1 generation, but from next onwards in each and every generation we need to screen these plants, whatever the plants we are obtaining we need to screen them whether our capital R allele is going to that plant or not because that is our target.

So, if we screen those BC_1, and BC_2 plants, using the particular molecular marker associated with the R gene, then it is known as foreground selection. Well, simultaneously our final target, if we think about the back cross-breeding process what is

our final target? Our final target is to develop a plant like this where capital A capital B capital R and capital D will be the genetic constitution that is our final target right. All the things all the genetic constitution will be like parent 1 because most of the good genes are available in parent 1 just, we just need to transfer the capital R allele which is our final target. So, other than this particular gene we have to screen the plants in BC_1, BC_2, and subsequent generations for rest of the molecular markers, which means for those molecular markers which are available across the genome other than this target trait that is known as background selection. Ok!

So, later on, we will be discussing it once again. So, in this way foreground selection and background selection is done and another thing I was mentioning that is the recombinant selection. Ok! Here we are discussing, the molecular marker-based selection process. Our target is to transfer the gene, but we are screening it through a molecular marker that should be tagged with that particular gene. If there is some difference suppose it is a chromosome our gene is here this is our target gene and we have developed a marker from here that is a little bit away from the gene. Ok? So, during the process once the gametes are formed crossing over may occur between this place, in between the non-sister chromatids of homologous chromosomes.

So, we may lose our gene suppose, we are getting an individual where our marker is there, but the target gene has not been reached there that thing may have occurred. Ok! So, we need to identify at least two different markers that are available close to that gene in two different orientations, one in the proximal part of the gene, and one in the distal part of the gene. So, that we can confirm it whether our gene part, total this part is being transferred or not whether any recombination is being taken place over there or not. Ok! So, we should not lose our genes that is our major objective, and the molecular marker is the aid through which we are screening those things in the gene transfer process.