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: 12

Lecture-56: Characterization of *OsGLP1* Gene from Rice (Part I)

Hello everybody, welcome to the SWAYAM NPTEL course on, Basics of Crop Breeding and Plant Biotechnology. Today, we will continue the last week of this lecture and in this lecture, we will be discussing Genetic Engineering in Crop Improvement. Some selected case studies will be discussed and, in this lecture, we will be discussing the characterization of the *OsGLP1* gene from rice. So, these are the concepts, which will be covered under this particular part. First of all, we will be discussing, germin-like protein genes. Those are a member of cupin family members. Ok! A large number of germin-like protein genes are available, in different plant systems, as well as, in some fungus also, *GLP* genes are available.

And, we will discuss, the characterization, cloning, and subsequent characterization of a particular rice germin-like protein gene. So, we will discuss the primer designing and cloning part in detail, and then how should we plan the experiment. Before cloning, and those things in vector, we need to analyze the restriction sites available within the gene part. So, it will be discussed here, then how the constructs will be prepared, for the characterization of a gene.

This construct preparation and subsequent study will be the part of functional genomics study of a particular gene. So, two types of constructs will be prepared over here, one gene-silencing construct, and one over-expression construct will be prepared. Later on, we will be discussing, and once the constructs are prepared then plant transformation will be done, and subsequently the putative transgenic lines or transformed lines will be

screened, through Southern blot analysis, northern blot analysis, western blot analysis, and other different molecular techniques. And, then we will discuss the functional significance of this particular gene. How can we know the functional significance? Because, from this molecular analysis, we can tell whether our transgene whether our transgenic plants are up to the mark or not, whether the over-expression has been done properly, and whether the silencing has been done properly those things will be understood, from these molecular studies.

But for functional significance, means this particular gene has any physiological impact or not for understanding that several studies, has to be conducted. So, those things will be discussed over here, in relation to, a particular rice *germin-like protein 1* gene. So, let us start our discussion. So, first, let us see what are, *germin-like protein* genes. Ok! Germin-like proteins, in brief, are known as GLP.

So, GLPs or germin-like proteins are widely distributed glycoproteins found in plants, that are members of the cupin superfamily, which means several beta-geraolyl structures are there, in cupin family members of protein. They have a variety of functional roles in various plant species. It has been identified in tobacco, in *Gossypium*, in rice, in different crop species, and they were found to have various functional roles. So, in rice, at least 40 GLPs are available. So, each of them might be having, a different role.

So, in our present discussion, we will try to focus on a particular *GLP* gene from rice. We know that in rice, more than 30,000 genes are there. We know the sequence of those genes, we get that information from the NCBI database, but if you have to characterize a particular gene then we have to follow this procedure. So, those germin-like protein 1, may act as structural proteins in some of the plant systems, or aid in signal transduction through their receptor role. They were found to have one or more enzymatic functions in different plant species.

Basically, 3 enzymatic functions were observed in different GLPs, either they were found to have oxalate oxidase property, or they were found to have SOD activity i.e.,

superoxide dismutase activity, or they were found to have ADP glucose pyrophosphatase or phosphodiesterase activity AGPPase activity. These 3 enzymatic activities are mostly found in different GLP members, across the plant species. So, here in this particular experiment, we are trying to find out the role of a particular rice GLP which is *OsGLP1*. So, we are trying to characterize this particular gene from *indica* rice because at NCBI in most of the cases, you will get the information from *japonica* rice. Ok! But there from, we can design the primer, we can clone our target gene from the *indica* rice genome then we can characterize it. So, we will be discussing those things.

So, here *OsGLP1* gene, it will be cloned from an *indica* rice genome, and eventually, it will be characterized. To understand the functional role of this gene, basically, 2 transgenic approaches are being taken here. One is the gene silencing approach i.e., endogenous gene silencing approach because *OsGLP1,* from the name itself it is known that *Os* stands for *Oryza sativa* means the gene is available in *Oryza sativa*. So, as the gene is there, there might be mRNA production, and protein production, within the plant system. So, through endogenous gene silencing, we will try to silence that gene we will try to silence the expression of that gene through post-transcriptional gene silencing mechanism or PTGS mechanism.

We will try to do the gene silencing in rice, the endogenous gene silencing, while we will try to over-express, the gene in a heterologous plant system over here we have used tobacco plant. We will do the over-expression, because in tobacco the *OsGLP1* will not be available that, gene will be a foreign gene over there. So, we will try to over-express it in a heterologous system, to see its features. So, let us see how should we proceed? First of all our experiment started with the CDS of *OsGLP1,* which was available at NCBI database, CDS stands for Coding DNA Sequence. Ok!

So, this is that part of the gene that eventually codes for the protein. Ok! So, the coding DNA sequence, it was initially available at the NCBI database, and there from we started our experiment. So, this one is the accession number of the coding DNA sequence of *OsGLP1* gene and then, I will show how should we start our experiment from here. So,

this one was the accession number, let us go to NCBI, in the NCBI database, we can search about this particular accession number and you can see the Oryza sativa *OsGLP1*, gene i.e., germin-like protein 1 complete CDS, coding DNA sequence is available over here. So, let us see what is available inside here you can see the complete CDS of this, gene is available its 642 base pair ok!

So, it is 642 base pair and it was available from the *Oryza sativa Japonica* group, means from *Japonica* rice cultivar this sequence was initially available in the NCBI database and this is the sequence. So, first, we will try to find out through which way we should clone the gene, I have mentioned about two approaches earlier during the cloning process. If no intron is there within the CDS, then we can clone it from the genomic DNA, but if intron is available then we have to clone this particular gene from the cDNA which will be prepared from the mRNA. So, we know the CDS sequence. So, first, we need to check whether intron is available or not.

So, we have to take the FASTA sequence of this gene, we have copied the FASTA sequence. Now we will BLAST, at the NCBI database once again. At the NCBI database, once we will go to the BLAST,, for knowing whether intron is available within the CDS or not. We are testing our gene, over here the CDS sequence has been tested, then we have to check this database over here, different options are there nucleotide collection, reference sequence selected RNA sequence. We have to search in the reference sequence of the representative genome.

We will see in the genome whether the CDS is hitting at a particular stretch or not. If any gaps are there, it means intron might be available, and here we can put the organism's name it was taken from *Japonica*. So, we are initially finding it *Japonica* database. So, BLAST is going on, and we got a hit in *Oryza sativa Japonica*, you can see over here in chromosome number 8, Ok? Let us click over there here, we can see our query 1 to query 642 is hitting at a stretch, which means that the CDS part is not hitting in different positions within the genome or within the chromosome.

It means those are intact, which means no intron is available within this particular CDS. So, we can clone it from the genomic DNA from this study, and we can understand through which approach we should clone either from the genomic DNA, or we can clone if intron is not available, or we can clone from the cDNA prepared from mRNA, if intron is available within the CDS, means if intron is available within the gene. So, now this part is done, thereafter I am going back over here. So, this one was our original sequence of this particular CDS, no intron is there. So, easily we can clone it from the genomic DNA.

Now for cloning purposes, only single-stranded DNA is written here, right? So, only single-stranded DNA is written here. So, basically in a double-stranded DNA 5' to 3' strands, and 3' to 5' strands are available normally, only one strand is shown in most of the websites. So, this is that strand means 5' to 3' strand. So, based on that we have to design the primer.

For primer designing first, we need to identify some parts from this region of this particular gene, means we need to clone the gene from the start codon i.e., ATG, while we need to clone this gene up to the stop codon also, TAA the stop codon is available over here. So, first, we need to identify some specific sequence from here ok. Suppose we have taken a sequence up to this over here while we have to identify some sequence at the 3' end of the gene also. Now if you just recall over here, the primer designing ok, how does primer work? The forward primer will bind over here in a 5' to 3' direction and will synthesize accordingly, while the reverse primer will bind over here and it will synthesize in this way. The reverse primer will be the 5' to 3' end in this way it will do the synthesis.

Now in the case of forward primer, if you think about let me take a different color. If you think about this forward primer, this forward primer sequence will be exactly similar to this sequence right, because it will be complementary to this strand the 3' to 5' strand. So, the primer sequence will be exactly similar to this one. So, as a forward primer, we can take this sequence along with that for cloning purposes, we need to add some basis, I

am coming to it. While for the reverse primer as it will be similar to this strand the reverse primer will be similar to this strand ok.

So, we need to make the reverse complement sequence of this one, like over here the reverse complement sequence will be 5' to 3', it will be TTA then ACC in this way gradually we have to go up to this, here from we have to start TTA ACC in this way the reverse primer has to be prepared. While the forward primer, I have mentioned will be just similar to this sequence, means it will, it may start from ATG and then GCC with different bases. Ok! In this way, the initial primer position, we have to identify for cloning a particular gene from genomic DNA or from cDNA, we need to have at least 18 mer primer from this genomic region. Ok! In addition to that, we have to add a few bases for cloning purposes, few restriction sites are used for cloning purposes. Restriction sites, mean restriction endonuclease sites will be used.

So, over here we can clone our gene in the *Bam*HI and *Hin*dIII sites. So, the *Bam*HI restriction enzyme site will be added over here, while the *Hin*dIII restriction site could be added over here, and before those restriction enzyme sites, we have to put some filler bases. So, that our restriction enzyme can sit over there properly. So, once those things are done, let me tell you, how the amplification will be taken place. Suppose, this one was the part of chromosome 8, in chromosome 8 the *OsGLP1* gene is available, and we have designed two different primers. Ok!

We have designed the forward primer, and we have designed the reverse primer. If you carefully see in both the primers, the forward primer and the reverse primer we have added some restriction sites. So, initially, as those restriction sites are not available within the genome it will be placed in this way, means it will not bind with the denatured DNA, means after denaturing once the annealing, will be started then our primer will bind with the denatured DNA. So, at that time the restriction sites, part will be like this green portion it will not anneal initially. So, in this way first couple of cycles, of amplification will be done, and thereafter gradually this part will be amplified, and

restriction parts will be amplified.

Within this first couple of cycles, it may not bind properly, but thereafter those parts will be amplified and eventually, we can get an amplicon, at its one end, will be having *Hin*dIII restriction sites, and at another end will be having the *Bam*HI restriction sites. In this way, in our PCR product, those two restriction sites will be incorporated. This is the *OsGLP1* gene, which we are trying to amplify the 642 bases per region, and the BamHI and *Hin*dIII, those are coming from the primer part. So, in this way we can design the primers of this particular *OsGLP1* gene, the GGATCC, *Bam*HI restriction on one site, have been added over there, and AAGCTT the *Hin*dIII restriction sites, have been added over here. A few buffer bases have been added so that the restriction enzyme can do the cleavage later on, and thereafter PCR amplification has been done using *indica* rice genomic DNA.

So, in this particular study, the *OsGLP1* gene was initially amplified from the Badshah bhog rice cultivar. So, Badshah bhog is a small-grain aromatic *indica* rice cultivar. So, this is the gel picture means after doing the PCR, 1 % agarose gel electrophoresis, was done and you can see approximately at around 642 base pair, we are observing the PCR amplified product of the *OsGLP1* gene, but until and unless we sequence it we cannot confirm whether it is, *OsGLP1* gene or anything else. So, it was loaded with a molecular weight DNA ladder and basically, pUC18 DNA was digested with the *Hin*fI restriction enzyme, and these different fragments, are available, and based on that we confirm that the size was close to the 642-base pair. So, thereafter this part was initially cloned and it was sequenced through Sanger sequencing, and after Sanger sequencing, we found this particular sequence that was, belongs to the *OsGLP1* gene, from the Badshah bhog rice cultivar means earlier the sequence was available in the NCBI database from *Japonica*.

Here from *indica* rice, we have cloned it and that sequence we got. Now for subsequent cloning, because we have to make different types of constructs using this gene. So, first, we need to check what are the restriction enzymes available within this gene? If you see

carefully this sequence, we have trimmed the restriction enzyme sites available in the primer. So, no *Bam*HI is available here, no *Hin*dIII is available over here just we have placed the gene part from the start codon to the stop codon.

Now we will check whether, means, the restriction enzyme, and restriction sites available within particular this gene. So, the gene sequence was taken, and we had to go to a different site i.e., known as NEB cutter. Ok! So, in NEB cutter, we can put our gene sequence, we can put the sequence over here, and we can find out, what are the different restriction enzyme sites available over there. So, if we see carefully no *Bam*HI site, no *Hin*dIII sites are available over here, and no *Eco*RI sites are available over here, but for the enzymes, that are commonly used in molecular cloning, one restriction enzyme site is available over here i.e., known as *Pst*I. Ok! You can see over here *Pst*I restriction site is available. Ok!

So, for checking it once again, we can do custom digest. In custom digest, suppose we are putting *Eco*RI means no match found in the table, which means *Eco*RI site is not at all available in this particular gene. Let us think about *Pst*I. So, *Pst*1 site is available one *Pst*I site is available over there. So, in this way, we can find out what are restriction sites, available within this particular gene.

So, that those sites could be, or may not be utilized in the cloning purpose. Now we are going back to our original part. So, far we found that within our gene *Pst*I restriction site is available. Now this is the sequence of this gene and based on that different amino acid sequences have been written over here. Ok! Like ATG is the start codon, methionine amino acid will be produced here.

In this way, different amino acids will be produced, from this particular gene, and at last, TAA the stop codon is available over there. And if you, see carefully the *Pst*I site is available in this position. Next, we have to find out the basic information of this gene. If we are trying to, characterize this particular gene, the first part is done, the cloning of this gene and sequencing is done, then we need to know how many copies of this particular

gene are available in the rice genome, otherwise, we cannot plan our experiment.

So, for copy number determination of the *OsGLP1* gene, the southern blotting was conducted. You know about, the southern blot technique, through southern blot first we need to isolate the genomic DNA, and then the genomic DNA will be digested, with different restriction enzymes or a specific restriction enzyme. Then, it has to be transferred from gel to membrane and subsequently, it will be hybridized with a particular gene-specific probe. And, then we can see some signals in the autoradiogram, based on that we can tell how many copies of that gene are available. So far here I will show you a case study on *OsGLP1* to understand how many copies of this gene are available in rice could be identified.

So, here basically the genomic DNA was isolated from Badshah bhog rice because there from we have cloned this particular gene. So, you have isolated genomic DNA from Badshah bhog rice and thereafter, we digested the DNA with *Bam*HI enzyme, with *Eco*RI enzyme, with *Hin*dIII enzyme, and *Pst*I enzyme. And, along with that the undigested DNA was also loaded on the gel and 0.8 % agarose gel and electrophoresis was done.

So, let us see, what did we get over there? So, we will see only this part we will focus on Figure A only. If you see, in Figure A, in undigested DNA, in the southern blot a signal came it was more than 21 kb. Ok! While in the case of *Bam*HI digested, in case of *Eco*RI digested, we got a signal at close to 20 kb or 18 kb at a higher part higher size. In case of *Hin*dIII, we got a signal over here at around 1.4 KB. While, in case of *Pst*I, we found 2 signals one at around 2.2 kb, and one at around 0.83 kb, 2 signals were detected. So, if you see, means if you think about the previous information that we got from the NEB cutter. We found that in the NEB cutter, the *OsGLP1* gene possesses a *Pst*I restriction site in between.

As the restriction site was available inside, once genomic DNA was digested with *Pst*I. So, it was cleaved our particular *OsGLP1* gene has been cleaved over here, and as we have used a 642 base pair, *OsGLP1* specific probe. Some part of the probe bound over

here, and some part of the probe bound over here. Hence, we got 2 signals in the case of *Pst*I, but in the case of *Bam*HI, *Eco*RI, and *Hin*dIII, only a single band was detected only. Because no internal *Bam*HI, *Eco*RI or *Hin*dIII restriction sites were available.

You might see some smear-like things, over here in the top part in *Pst*I digested DNA. Because *Pst*I cannot digest the genomic DNA, rice genomic DNA uniformly. The *Bam*HI, *Eco*RI, and *Hin*dIII, they basically digest the genomic DNA uniformly. In *Pst*I, the *Pst*I sites are not available too much. So, some undigested DNA might be there and those are giving these unwanted signals.

So, here from, we concluded that the *OsGLP1* gene is available as a single copy in rice genome. Because in *Bam*HI, *Eco*RI, and *Hin*dIII, we are getting a single signal, while in *Pst*I we are getting 2 signals because within the gene within the probe, the *Pst*I restriction site was available, within the gene. So, it will be fragmented. Thereafter, coming to the next part coming to figure b through this analysis we try to find out in which tissue our *OsGLP1* gene is expressed. Ok! So, far we understood that in Badshah bhog rice, one copy of the *OsGLP1* gene is available.

Then, we are trying to find out that in, which tissue it is being expressed. For tissue specific expression, we have to do northern blot analysis. So, we isolated RNA samples from different rise tissues, means the root tissue, the sprout, the leaf tissue, the developing endosperm tissue, the husk, as well as mature endosperm tissue. So, from these different tissues, the RNA was isolated, and finally, the RNA was loaded in northern gel and the gel finally, was hybridized with *OsGLP1*, gene-specific probe. And, we found that in the leaf tissue as well as in the sprout tissue i.e., in the green part of the plant, leaf tissue and the sprout over there, the strong, *OsGLP1* specific transcript band is available at around 1900 base pair, at around 1900 bases, RNA are single-stranded right?

At around 1900 bases, we observed signals in these two tissues only. While in other tissues no specific band was detected. To check whether an equal amount of RNA was added in different samples or not like in root, sprout, leaf, developing endosperm, husk,

and mature endosperm. We have hybridized the same blot, we have hybridized the same northern blot, using actin gene-specific probe, using actin gene-specific probe you can see almost similar type of loading was done for all types of RNA.

But over here, clearly we can see that in the green tissue, in the sprout as well as, in the leaf tissue our *OsGLP1* was highly expressing, means the tissue-specific expression of this gene was established from here. So, then we are coming back once again to our original discussion. So, once we understood that the gene is available as a single copy and it is expressed mostly in the green tissue then we planned the functional genomics study. So, what is functional genomics study? I have mentioned earlier in functional genomics study, we need to know the function of a particular gene, through this study ok. To understand the function, we can take different strategies over here, two strategies have been taken, one is the endogenous gene silencing, and another one is heterologous overexpression, over-expression in a heterologous system.

So, in endogenous gene silencing, if we just recall in the normal rice cell, the *OsGLP1* gene is available. So, from this gene, the *OsGLP1*-specific mRNA will be produced. So, if we do RNAi construct preparation, if we make RNAi construct specific to the *OsGLP1* gene then, what will be happening from that RNAi construct through successful transcription finally, the siRNA will be produced the small interfering RNA will be produced. And, small interfering RNA this sequence will be similar to the *OsGLP1* mRNA sequence. So, those siRNAs can bind over here, because over here the RNAi construct will be made using *OsGLP1* specific gene right?

So, those specific siRNAs can target, those mRNA and once those mRNA will be targeted, eventually it will be cleaved. So, in our transgenic plant, the *OsGLP1* mRNA would be reduced, if we do the gene silencing would be reduced. So, the protein production will be reduced also, we may see some features, and we may see some functions while in the over-expression strategy, we are attempting in heterologous system. Suppose, in rice sorry, suppose in tobacco, in tobacco different cells are their different proteins are there, but OsGLP1 protein is not available. Ok! So, if we over-

express, OsGLP1 protein, that mRNA will be new over there, and there from OsGLP1 protein will be there also.

So, here we can see, OsGLP1 protein over-production, it can show some features also. So, in this way through functional genomics study, we can characterize a particular gene.