Identification of a new allele of GS3 gene in an aromatic Indica rice cultivar Badshabhog

Sheuli Roy¹, Tirthartha Chattopadhyay¹,², Mrinal K. Maiti¹,³*

1. Advanced Laboratory for Plant Genetic Engineering, Advanced Technology Development Centre, Indian Institute of Technology Kharagpur, Kharagpur 721302, India
2. Department of Plant Breeding and Genetics, Bihar Agricultural University-Sabour, Bhagalpur 813210, India
3. Department of Biotechnology, Indian Institute of Technology Kharagpur, Kharagpur 721302, India

*Corresponding Author email: maitimk@hijli.iitkgp.ernet.in, mrinalkmaiti@gmail.com

ABSTRACT: The GS3, a major quantitative trait locus (QTL) for grain length and weight and a minor QTL for grain width in rice has been reported to encode a transmembrane protein that acts as a negative regulator of grain size. On the basis of different mutations, 4 different types of GS3 alleles have been documented so far. In this study, we report the evidence of a long genomic DNA deletion in the 5th exon and 3′ untranslated transcribed region (UTR) of the GS3 gene present in an aromatic Indica cultivar cultivar Badshabhog. The deletion results in a frameshift mutation that predicts to a truncated GS3 protein with only 2 domains in comparison to the 4 domains present in the wild type GS3 protein. The small-grain phenotype of the Badshabhog cultivar could be well-explained on the basis of the previously established inter-domain interaction of the GS3 protein. Based upon our findings, we claim the presence of a new 5th GS3 allelic version in this Indica rice cultivar Badshabhog.

Key Words: Frameshift mutation; Genomic deletion; Grain size; Inter-domain interactions; Point mutation.

INTRODUCTION

Rice (Oryza sativa) is the principal staple food crop for the majority of the world population. Among the several parameters, grain size has been focused as an important one in breeding programmes for rice improvement, as the long-grain Indica rice is preferred over the short-grain one in USA, China and other Asian countries (Unnevehr et al., 1992; Juliano and Villareal, 1993). Availability of the complete genome sequence and development of several molecular markers have been explored to identify different quantitative trait loci (QTL) for important agronomical, biochemical, morphological and physiological traits in rice. In the recent past, a major QTL for grain length and weight and a minor QTL for grain width and thickness has been identified in the pericentromeric region of chromosome number 3 in rice (Fan et al., 2006). This locus, named as GS3, has been reported to explain 80-90% of the variation for grain weight and length in the advanced backcross progeny of a cross between a long-grain and a short-grain rice cultivar. Further, the locus has been delimited to a ~7.9 kb DNA fragment between the cleaved amplified polymorphic sequence (CAPS) markers GS63 and SF19, and has been found to code for a transmembrane protein containing 232 amino acids (Fan et al., 2006). Bioinformatics analysis of the predicted GS3 protein sequence has revealed the presence of a phosphatidylethanolamine-binding protein (PEBP)-like domain, a transmembrane domain, a putative tumor necrosis factor receptor (TNFR)/nerve growth factor receptor (NGFR) family cysteine-rich domain and a C-terminal cysteine-rich part similar to the von Willebrand factor type C (VWFC) module. Later, the PEBP-like domain has been found to be no longer predicted, and hence has been renamed as organ size regulation (OSR) domain (Mao et al., 2010). The GS3 gene has been reported to contain 5 exons with the 5th exon being the longest in size (Fan et al., 2006). The 2nd exon of the GS3 gene of long-grain rice cultivars has been documented to contain a C to A point mutation which results in a premature termination codon (TGA in place of TGC) in the open reading frame (ORF), and thus to code for a truncated non-functional GS3 protein. Considering the GS3 as a negative regulator of grain size in rice, the long-grain phenotype of rice has been reported to arise due to this non-sense mutation resulting in the loss-of-function GS3 allele from the functional GS3 allele of small-grain rice cultivars. Afterwards, this point mutation has been reported to be highly associated with grain length of rice, and a linked functional CAPS marker (SF28) has been developed to study the GS3 allelic diversity in rice (Fan et al., 2009). In a similar approach, a PCR-based SNP marker system (DRR-GL) has also been developed to identify this C to A point mutation in rice cultivars and explore the same
through marker assisted selection (MAS) in rice breeding programme (Ramkumar et al., 2010). Apart from this point mutation in the 2nd exon, polymorphism has also been reported in the 2nd intron, 4th intron and the 5th exon of the GS3 gene, and corresponding functional markers (SR17, RGS1 and RGS2) have been developed (Wang et al., 2011). Depending upon the different mutations of the GS3 gene, at least 4 different alleles of the gene (GS3-1, GS3-2, GS3-3 and GS3-4) have been proposed (Mao et al., 2010). Among the 4 alleles, GS3-4 has been reported to be associated with very small grain size of the cultivar Chuan 7. Logically, it is interesting to uncover the allelic version of GS3 gene present in rice cultivars having smaller grain size than Chuan 7. In the present study, we report a new GS3 allele in an aromatic Indica rice cultivar Badshabhog having smaller grain size than that of Chuan 7.

MATERIALS AND METHODS

Plant materials

In the present study, 4 locally available rice cultivars, namely Badshabhog, Gayatri, Supershyamali and Lalsita were selected. Among these 4, Badshabhog and Gayatri come under short-grain class whereas Supershyamali and Lalsita are in long-grain class. Furthermore, Badshabhog is the only aromatic cultivar considered in this study.

PCR amplification and cloning

Genomic DNA isolation from the leaf tissues of rice seedlings was carried out using the standard cetyltrimethyl ammonium bromide (CTAB) method described previously (Doyle and Doyle, 1990). The 447 bp (excluding the primer regions) genomic fragment of the GS3 gene containing the part of the 1st intron, the complete 2nd exon and the part of the 2nd intron was PCR-amplified with the help of gene-specific primer pair GS3PMS FP 5’-GTGACAGATCTGTGCTAG-3’ and GS3PMS RP 5’-GCACGATACTAGTATTAATG-3’ using genomic DNA isolated from all the afore-mentioned cultivars. Amplification of the 429 bp region (encoding the 91st to 232nd amino acid part of the wild type GS3 protein) of the 5th exon of the GS3 gene was attempted using gene-specific primer pair GS35E FP 5’-TGTTTGTGCAGAACGTC-3’ and GS35E RP 5’-TCACAGAGGGGGGACGAGC-3’ taking Badshabhog genomic DNA. The specific primer GS33U RP 5’-AGCAATCCTACCTACATGCAGCAGC-3’ designed on the basis of the reported 3rd untranslated transcribed region (3’UTR) of the GS3 gene was used with the GS35E FP primer to amplify a genomic fragment of expected 631 bp length. All the PCR reactions were carried out in 25 µl volume containing 150 ng of genomic DNA, 0.2 mM of each dNTPs, 0.4 µM of each primers and 0.5 U of Phusion® High-Fidelity DNA Polymerase in 1X PCR (HF) buffer (New England Biolabs). The PCR reaction was programmed as follows: initial denaturation at 94 °C for 4 min, 35 cycles of 94 °C for 30 s, 61 °C for 40 s and 72 °C for 1 min followed by final extension at 72 °C for 10 min and cool-down at 25 °C for 2 min. Following PCR, amplicons were analyzed through 1.5% agarose gel electrophoresis and ethidium bromide staining for visualization under UV transilluminator. Respective amplicons obtained from all the 4 cultivars were separately cloned in pTZ57R/T vector (Fermentas) through TA cloning, as per the guidelines provided by the manufacturer.

Sequencing and sequence analysis

Following cloning of the respective amplicons, representative clones were subjected to custom sequencing. Sequences were analyzed through Jellyfish 3.3.1 software and multiple sequence alignment was carried out in ClustalW (Thompson et al., 1994).

RESULTS

In the present study, 4 local Indica rice cultivars were selected for analysis. Among these, Badshabhog and Gayatri were short-grain in nature, whereas Supershyamali and Lalsita had long grains (Fig. 1a). Furthermore, grains of Badshabhog were shortest among these four (Table 1) and smaller than the Chuan 7 grains, as reported earlier (Mao et al., 2010). It has been previously reported that the GS3 allele present in the long-grain rice cultivars contain a C to A point mutation in the 2nd exon. Due to the presence of this mutation, the wild type GS3 allele is often referred as ‘C type’ allele whereas the mutant allele is referred as the ‘A type’ allele. The C nucleotide, site for the point mutation is also part of the PstI restriction site (CTGÇAG) present in the 2nd exon of the GS3 gene. Naturally, the C to A point mutation in this position abolishes the PstI restriction site (CTGÇAG changed to CTGAAG). Presence of this C to A point mutation in the presently selected 4 rice cultivars was examined through a PCR based strategy. For this purpose, a 447 bp (excluding the primer regions) genomic fragment of the GS3 gene containing the part of the 1st intron, the complete 2nd exon and the part of the 2nd intron was PCR-amplified with the help of gene-specific primer pair GS3PMS FP and GS3PMS RP using genomic DNA isolated from all the afore-mentioned cultivars. Respective amplicons were cloned in the pTZ57R/T vector through TA cloning. The pTZ57R/T vector contained single PstI and BamHI sites in the
In accordance with the previous observation regarding the exon identity (Fig. 2c). This implied exon and thus encodes a protein of 232 amino acids; the site (MCS) region (Fig. 1b). Moreover, the GS3 2′UTR, TM, TN elements alone. Hence, we made attempts to amplify and subjected to PCR, finding the expected 631 bp length. Interestingly, in this case we were able to amplify a genomic fragment of expected size (429 bp, as per the reported sequence, GenBank: DQ355996). Therefore, we designed the specific primer GS33U RP based on the reported sequence, GenBank: DQ355996). Consequently, the Badshabhog genome was used as a template for PCR using gene-specific primer pair GS35E FP and GS35E RP from Badshabhog genomic DNA. However, even after several attempts, we failed to generate a PCR amplicon of expected size (429 bp, as per the expected amplicon size for the wild type GS3 allele (C type allele), as presented in Fig. 2b of 311 bp length from the Badshabhog genome, which was found to be significantly smaller than the expected amplicon size (631 bp) from the wild C type allele. The amplicon was cloned in pTZ57R/T vector through TA cloning, as described earlier, and subjected to sequencing. Analysis of the sequence revealed a very interesting observation in this regard. Alignment of the reported sequence (GenBank: DQ355996) with the sequence of the cloned fragment amplified from the Badshabhog genome (GenBank: KF056798) revealed that the first 129 nucleotides of the cloned fragment shares 100% sequence identity with the reported one (Fig. 2c). However, the following 342 nucleotides were found to be absent in the cloned fragment and replaced by a stretch of 22 nucleotide sequence having low level of sequence similarity with the reported sequence. Furthermore, the last 160 nucleotides (part of the 3′ UTR) shared 100% sequence identity (Fig. 2c). This implied a very significant genomic deletion in the 5th exon and 3′ UTR of the GS3 gene in Badshabhog. The result also justified our repeated failure to amplify the GS3 5th exon from Badshabhog genome using GS35E FP and GS35E RP primer pair. From sequence analysis it was clear that the region, on the basis of which the GS35E primer was designed, is absent in case of the Badshabhog GS3 allele (Fig. 2c, black arrow).

The deletion of nucleotides in the 5th exon was found to result in a premature termination codon (TTA) due to frameshift mutation in the ORF of this new GS3 allele. Consequently, the Badshabhog genome was observed to contain a truncated 5th exon (141 bp) in comparison to the corresponding 5th exon part (429 bp) present in the wild type GS3 allele of short-grain rice cultivars (Fig. 3), leading to a truncated short GS3 protein of 136 amino acids, in comparison to the wild type GS3 protein of 232 amino acids, as revealed through in silico analysis (Fig. 4). Study of the domain architecture indicated that the Badshabhog GS3 protein contains the complete OSR domain as well as the TM domain, but lacks most of the TNFR and complete VWFC domains (Fig. 5).

**DISCUSSION**

The allelic variation of GS3 and its association with grain length in rice has been studied extensively in recent past (Mao et al., 2010). On the basis of mutations in the GS3 gene sequence, existence of at least 4 types of GS3 alleles (GS3-1 to GS3-4) has been proposed. The GS3-1 allele (C type allele), as present in the cultivar Zhenshan 97 codes for a protein of 231 amino acids; the GS3-2 allele (C type allele) present in the cultivar Nipponbare contains a 3 bp addition in the 5th exon and thus encodes a protein of 232 amino acids; the
GS3-3 allele, found in the cultivar Minghui 63 (A type allele) contains the C to A point mutation in the 2nd exon and thus codes for a truncated non-functional GS3 protein. The GS3-4 allele (C type allele), present in the cultivar Chuan 7 has been reported to contain 1 bp deletion in the 5th exon, and encodes a truncated protein of 149 amino acids as a consequence of the frameshift mutation due to the deletion. Among these 4, Nipponbare and Zhenshan 97 have been reported as medium-grain, Minghui 63 as long-grain, whereas Chuan 7 as short-grain cultivar (Mao et al., 2010). Using transgenesis in the background of a long-grain rice cultivar (Minghui 63), the authors have established the function of the different domains present in the GS3 protein. It has been established that the OSR domain is the negative regulator of grain size, whereas the TNFR and VWFC domains negatively interact with the activity of the OSR domain. The grain size of Chuan 7, a small-grain rice cultivar has been associated with a novel GS3 allele (GS3-4), where a point mutation results in a truncated GS3 protein with intact OSR domain but deleted TNFR and VWFC domains. Furthermore, transgenic expression of the OSR and TM domains has resulted in the maximum reduction of grain size, even more than the grain size reduction due to the transgenic expression of Chuan 7 GS3 allele (GS3-4) in the long-grain rice cultivar Minghui 63 (Mao et al., 2010). In our study, we find the grain length and width of Badshabhog (Table 1) to be even shorter than those reported for Chuan 7, which implies the existence of a new type of GS3 allele in Badshabhog.

We document here a significant 320 bp genomic DNA deletion in the GS3 gene of Badshabhog through PCR-mediated amplification of the 5th exon and part of the 3' UTR of the gene. The frameshift mutation resulting from this deletion leads to a mutant GS3 protein in case of Badshabhog, where the most part of the TNFR domain and the complete VWFC domain are abolished (Fig. 5). The 1 bp deletion and subsequent frameshift mutation in case of the GS3-4 allele of Chuan 7 also results in a mutant protein lacking the aforementioned domains; however the GS3 protein of Badshabhog is predicted to be smaller than the same of Chuan 7 (136 amino acids in case of Badshabhog and 149 amino acids in case of Chuan 7). On the basis of the effect of the TNFR and VWFC domains on the grain size reduction activity of the OSR domain, absence of the TNFR and VWFC domains from the OSR domain containing GS3 protein of Badshabhog (small-grain cultivar) is well-justified. Furthermore, reduced grain size of Badshabhog in comparison to Chuan 7 can also be explained from the domain deletion studies carried out previously (Mao et al., 2010). The authors have documented that the extent of grain size reduction through expression of the OSR and TM domain of GS3 is greater than the extent of grain size reduction through expression of the GS3-4 allele of Chuan 7 in the background of a long-grain cultivar Minghui 63. On the basis of their results, the authors indicated the effect of the frameshift C terminus part of the GS3-4 allele on grain size regulation. In corroboration to that work, we found a natural genomic DNA deletion in the GS3 allele of Badshabhog, which can explain the small grain size of this cultivar due to the presence of only the complete OSR and TM domains in the protein. On the basis of our findings, the GS3 allele of Badshabhog (GenBank: KF056798) should be rendered as the 5th GS3 allele (GS3-5).

Previously, the GS3 gene has been reported to contain point mutations as well as small (not more than 3 bp) insertion or deletion mutations (Mao et al., 2010). The present study documents a large genomic DNA deletion from the 5th exon part and part of the 3 UTR of the GS3 gene in the aromatic Indica rice cultivar Badshabhog. The genomic deletion causing the frameshift mutation is capable of producing a truncated GS3 protein (containing 136 amino acids) with the complete OSR and TM domains only. The small-grain phenotype of Badshabhog can be properly justified in the light of our findings and the inter-domain interaction studies on GS3 protein reported earlier (Mao et al., 2010). It is pertinent to mention here that the same genomic deletion in the GS3 gene has been identified very recently in a few other aromatic rice cultivars collected from India (Takano-Kai et al., 2013). Taken together, our findings strengthen the hypothesis that the GS3-5 allele and its product, i.e. the truncated GS3 protein might be associated with the small-grain (and aroma?) trait(s) in case of Indica rice cultivars, and claims further research in this regard. Moreover, the information provided in this study will help in developing gene-based functional markers for marker assisted selection of grain size (and possibly aroma) character(s) in future rice breeding programmes.

ACKNOWLEDGEMENTS

Authors thank Prof. S.K. Sen and Dr. A. Basu of Advanced Laboratory for Plant Genetic Engineering, Indian Institute of Technology Kharagpur for their cooperation. TC and SR acknowledge Indian Institute of Technology Kharagpur and INSPIRE programme of Department of Science and Technology, Government of India, respectively, for providing research scholarships. Financial supports from Department of Biotechnology and Department of Science and Technology, Government of India are acknowledged.
Figure 1. Badshabhog contains the C type GS3 allele. 

a) Comparison of grain sizes among the selected rice cultivars. BB-Badshabhog; GA-Gayatri; SS-Supershyamali; LS-Lalsita. White bar indicates 1 cm. b) Ethidium bromide stained 1.5% agarose gel showing restriction fragments generated through BamHI and PstI digestion of respective clones harbouring specific amplicons (containing the part of the 1st intron, the complete 2nd exon and the part of the 2nd intron) of the GS3 gene derived from Badshabhog, Gayatri, Supershyamali and Lalsita genome. Schematic diagram of the cloned amplicon is shown by arrow shape drawing. PstI* indicates the PstI restriction site, to be present in the wild type (C type) allele of GS3, only. MCS indicates multiple cloning site of the vector. c) Nucleotide sequence alignment of the cloned amplicons derived from Badshabhog genome (BB), Supershyamali genome (SS) and the corresponding sequence reported in the public database (GenBank: DQ355996). Position of the C to A point mutation is shown inside the red box.
Figure 2. Badshabhog GS3 allele contains a large genomic deletion. a) Autoradiogram generated after Southern hybridization using GS3 gene-specific probe, showing the genomic organization of GS3 gene in Badshabhog rice cultivar. b) Ethidium bromide stained 1% agarose gel showing the amplicon (5\textsuperscript{th} exon and part of the 3\textsuperscript{′} UTR) generated from the Badshabhog genomic DNA using the GS35E FP and GS33U RP specific primers. M denotes HinfI digested pUC18 DNA as standard molecular weight marker. c) Nucleotide sequence alignment of the cloned amplicon derived from Badshabhog genome (BB) with the corresponding sequence reported in the public database (GenBank: DQ355996). Position of the GS35EFP and GS33U RP primers are indicated by grey-colored arrow, whereas the position of the GS35ERP primer is indicated by the black-colored arrow. Positions of stop codons are indicated by red asterisk marks.

Figure 3. Badshabhog GS3 allele contains a truncated 5\textsuperscript{th} exon. Nucleotide sequence alignment of the cloned 5\textsuperscript{th} exon part derived from Badshabhog genome (BB) with the 5\textsuperscript{th} exon part from the sequence reported in the public database (GenBank: DQ355996). Position of the premature termination codon in the 5\textsuperscript{th} exon of Badshabhog GS3 allele is underlined.
Figure 4. Badshabhog GS3 allele codes for a truncated GS3 protein. Amino acid sequence alignment of the translated coding DNA sequence (CDS) of Badshabhog GS3 allele and the corresponding amino acid sequence reported in the public database (GenBank: DQ355996). Identical amino acids are highlighted in red.

Figure 5. Schematic diagram of the different types of GS3 alleles and domain architecture of their corresponding deduced amino acid sequences. E1, E2, E3, E4 and E5 denote exon 1, exon 2, exon 3, exon 4 and exon 5 respectively. OSR- organ size regulation domain; TM- transmembrane domain; TNFR- tumor necrosis factor receptor family cysteine-rich domain; VWFC- von Willebrand factor type C domain.

Table 1. Grain dimensions and types of the selected rice cultivars

<table>
<thead>
<tr>
<th>Cultivar</th>
<th>Grain length in mm (mean ± SE)</th>
<th>Grain width in mm (mean ± SE)</th>
<th>Grain thickness in mm (mean ± SE)</th>
<th>Grain type</th>
</tr>
</thead>
<tbody>
<tr>
<td>Badshabhog</td>
<td>6.2 ± 0.17</td>
<td>2.00 ± 0.05</td>
<td>1.43 ± 0.03</td>
<td>Short</td>
</tr>
<tr>
<td>Gayatri</td>
<td>6.8 ± 0.16</td>
<td>2.86 ± 0.02</td>
<td>1.90 ± 0.05</td>
<td>Short</td>
</tr>
<tr>
<td>Supershyamali</td>
<td>10.13 ± 0.24</td>
<td>2.27 ± 0.07</td>
<td>1.73 ± 0.03</td>
<td>Long</td>
</tr>
<tr>
<td>Lalsita</td>
<td>10.07 ± 0.11</td>
<td>2.36 ± 0.05</td>
<td>1.57 ± 0.05</td>
<td>Long</td>
</tr>
</tbody>
</table>

*Values represent random samples of 25 grains from each cultivar. SE = Standard error

REFERENCES

Doyle JJ, Doyle JL. 1990. Isolation of plant DNA from fresh tissue. BRL Focus 12:13–15
Juliano BO, Villareal GP. 1993. Grain quality evaluation of world rices. International Rice Research Institute, Manila