Received: 05 Sep 2021 | Revised: 20 Sep 2021 | Accepted: 28 Sep 2021

Electronic Journal of Plant Breeding

Research Article

Haplotype analysis of *SCM2* loci for identification of donor for lodging resistance in rice

K. Madhuri[1] , S. Manonmani[1*], T. Kalaimagal[1], R. Veera Ranjani[2] and M. Raveendran[2]

[1] Centre for Plant Breeding and Genetics, TNAU, Coimbatore
[2]Centre for Plant Molecular Biology and Biotechnology , TNAU, Coimbatore
***E-Mail:** manonmanitnau@gmail.com

Abstract
Lodging resistance is one of the vital traits in improving the yield and sustainability of production in rice (*Oryza sativa* L.). A quantitative trait loci (QTL), *SCM2* for enhancing the culm strength and increased spikelet number has been reported. In the present study, a diverse collection of 150 accessions from the rice 3K-RG panel were assembled and evaluated for the *STRONG CULM2* locus. All the accessions under study showed a normal phenotypic distribution for culm diameter and the allelic diversity analysis for *SCM2* resulted in four INDEL's and 17 SNP's in both coding and non-coding regions of the genome. Haplotype analysis grouped the entire population into three significant haplotype groups with 114 accessions in H1, 19 accessions in H2 and 17 accessions in H3. Haplotypes were formed using two significant SNP sites viz., 27480778 and 27481339. Dunnett's test disclosed a significant variation between H1-H2 and H1-H3 haplotypes. Phenotypic evaluation for culm diameter in all the three haplotype groups revealed that the genotypes BR IRGA and PERUNEL in H1, CICA 9 and MOSHI in H2 and KITRANA in H3 possess a maximum culm diameter. H2 and H3 were found to be superior to H1 hence can be utilized for lodging resistance in further breeding programs.

Key words: Lodging resistance, *SCM2*, culm strength, allelic diversity and haplotype analysis.

INTRODUCTION

Rice is a staple food for more than half of the global population and it provides 15 % of per capita protein and 21% of global human per capita energy (Keerthiraj and Biju, 2020). The rising world population and the effects of global climate change pose a serious threat to the future food crisis (Brown and Funk, 2008). Increasing the yield potential and decreasing the yield gap of rice crops has been one of the critical challenges to the agricultural research scientists of the state (Bidhan *et al*., 2021). One remarkable success has been encountered through the development and the adoption of semi dwarf varieties in the 1960's. As part of the green revolution, there was wide usage of fertilizers for high yielding semi dwarf varieties, which promoted an increase in yield along with stem and leaf elongation. With the application of fertilizers, the traditional rice varieties grow tall making them susceptible for lodging whereas the dwarf varieties, because of their short stature, were resistant to lodging with more tillering capacity and harvest index (Ookawa *et al*., 2010).

During the Green revolution, the extensive usage of gibberellin (GA) related semi dwarfs is because of its improved bending type lodging resistance. When compared to normal plants, short statured plants possess a lower centre of gravity which make them resistant to bending pressure (Kashiwagi *et al.,* 2004). When compared with traditional long culm varieties, the biomass production potential of semi dwarf varieties was lower due to high leaf area density in the canopy and poor CO2 diffusion efficiency because of their short

culm nature (Nomura *et al*., 2019). One of the genes which have prominent role in inducing semi dwarfs in rice is *SD1*. These mutants have been reported to show negative effects not only on biomass production but also on grain weight when compared with wild types (*SD1*) (Okuno *et al.,* 2014).

All these reasons reflect a need for a different strategy to develop varieties that possess high biomass and harvest index with an increased yield. To attain this objective, it is necessary to develop varieties with improved lodging resistance. Major genes and other quantitative trait loci (QTL) controlling culm strength need to be identified and isolated.

A potential quantitative trait loci (QTL) for culm strength has been identified using chromosome segment substitution lines (CSSL) called *STRONG CULM2* (*SCM2*). Through positional cloning, it was revealed that *SCM2* was similar to *ABERRANT PANICLE ORGANIZATION1* (*APO1),* a gene that is reported to control panicle structure. *SCM2* is a gain-of-function mutant of *APO1,* but it doesn't show any negative effects that are reported for APO1 over expression mutants such as decreased panicle number and abnormal spikelet morphology (Ookawa *et al.,* 2010).

For identifying the superior version of genes in the germplasm, the haplotype assembly concept can be used (Bevan *et al.,* 2017). In the following study, allelic diversity and haplotype analysis has been performed for the diverse 150 accessions from the rice 3K-RG rice panel for the target gene *SCM2*.

MATERIALS AND METHODS

A diverse collection of 150 accessions from the rice 3K-RG panel was assembled and evaluated for culm strength along with the allelic diversity and the haplotype analysis for the *SCM2* gene. The field experiment was conducted during *Rabi* season, 2020 at Paddy Breeding Station, Tamil Nadu Agricultural University, Coimbatore, India in a Randomized Block Design (RBD). All the standard agronomic practices were followed and data were documented for three plants selected randomly from each accession for the target trait stem girth (mm). Digital Vernier callipers were used to measure the stem girth at the fourth inter-node at 20 days after heading (Merugumala *et al.,* 2019). Descriptive statistics and frequency distribution of the whole population for stem girth was performed using Minitab statistical software (Minitab, Inc. 2010)

Allelic diversity in the entire population was performed by downloading their genomic sequences from Rice SNP Seek Database (Mansueto *et al*., 2017) and the multi sequence alignment was done using Bioedit software (Hall *et al*., 1999) using Nipponbare as a reference genome.

Grouping of the entire set of accessions into significant haplotype groups was performed using Haploview software (Barrett *et al.,* 2005). For carrying out the analysis, the PLINK file of the target gene was downloaded from the rice SNP seek database and converted to Haploview format using PLlink software (Weeks, 2010). By utilizing the allelic variation in the population, haplotype groups were formulated using Haploview. Using Dunnett's multiple comparison test, significant differences between the haplotypes were analyzed.

RESULTS AND DISCUSSION

Lodging has been a major barrier in increasing rice productivity. A quantitative trait loci (QTL), designated as *STRONG CULM2* (*SCM2*) showing an enhanced effect on culm strength has been reported in rice (Ookawa *et al*., 2010). *SCM2* is similar to the *ABERRANT PANICLE ORGANIZATION1* (*APO1*) gene, which was known to encode an F-box-containing protein that is involved in controlling the number of rachis branches in a panicle. In this study, a subset of 150 accessions from the rice 3K-RG panel was evaluated for stem girth.

A subset of 150 accessions from the rice 3K-RG Panel was analyzed for culm diameter using Digital Vernier callipers. For understanding the distribution of all the 150 accessions, a histogram and box plot was generated for the stem girth using Minitab 19 statistical software. The resulted histogram revealed a normal frequency distribution for the target trait (**Fig. 1 & 2**). Descriptive statistical analysis performed on the whole population set indicated that the mean value for culm diameter was 3.65 mm and coefficient of variation (CV) was 19.11.

The allelic variation for the target locus *SCM2* was analyzed in all the 150 accessions. *SCM2* is positioned on the sixth chromosome occupying from 27480082bp to 27481387bp (Mansueto *et al*., 2017;

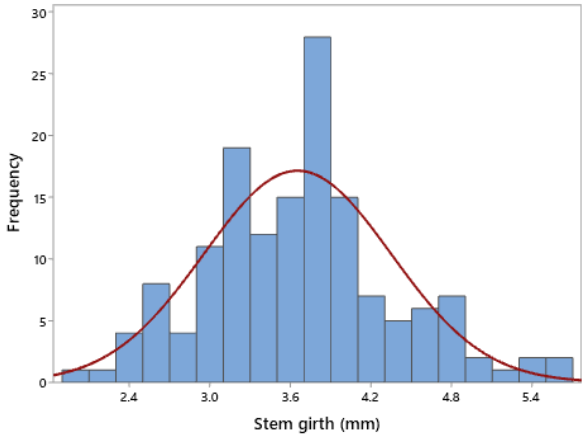

Fig.1. Histogram of Frequency distribution for stem girth

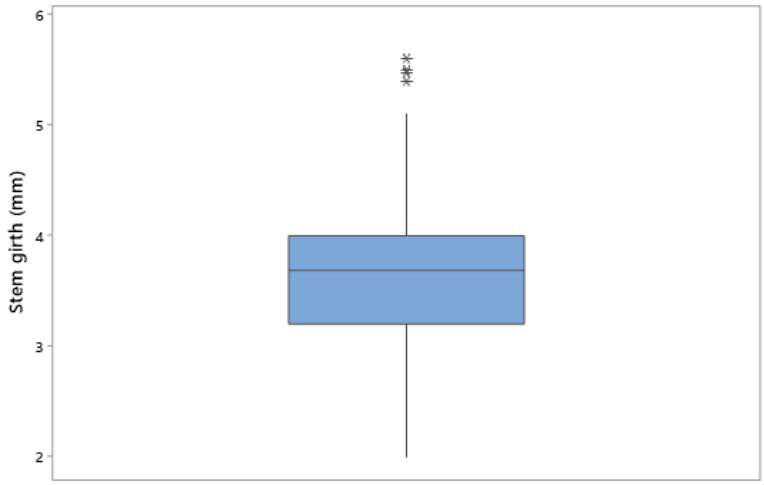

Fig. 2. Box plot for stem girth

http://snp-seek.irri.org/_download.zul). For performing the analysis, the genomic sequences of all the accessions were downloaded from the rice SNP seek database and compared with the reference genome Nipponbare. All the genomic sequences for the *SCM2* loci (LOC_Os06g45460) along with the reference genome were aligned using Bioedit software. In comparison with the reference genome, accessions showed allelic variations in both coding and non-coding regions of the genome. Exon1 harbours six SNP's and Exon2 possess eight SNP's and four INDEL's whereas, three SNP's were seen in the non-coding regions of the genome (**Table 1**).

The substitution of a single nucleotide at a specific position in the genome may lead to changes in the amino acid coding (non synonymous) or sometimes without any change in the coding pattern (synonymous). The variations in the SNP sites 27480787, 27480856 and 27481348 of the Exon1 region revealed synonymous SNPs showing no change in amino acid coding whereas single nucleotide variation in the SNP site 27480917 leads to a change in the coding pattern where Valine was coded instead of Alanine. In a similar way, variation in the SNP site 27480926 leads to the production of Alanine in place of Glycine and variation in the position 27480836 leads to

Table1. Allelic diversity observed in *SCM2*

Region	SNP site	Allele	Type	MAF
EXON2	27480200-27480205	GCGCCG	INDEL I	2.67
EXON2	27480298	G/C	SNP	0.67
EXON2	27480311	T/C	SNP	0.67
EXON2	27480374	- /G	INDEL D	1.33
EXON2	27480376	C/G	SNP	6.00
EXON2	27480379	A/G	SNP	0.67
EXON2	27480430	T/A	SNP	10.67
EXON2	27480433	T/G	SNP	10.67
EXON2	27480447-27480455	CGCCGCCGC	INDEL D	32.00
EXON2	27480535-27480537	CGC	INDEL I	8.00
EXON2	27480586	T/C	SNP	12.00
EXON2	27480616	C/T	SNP	12.00
INTRON	27480684	G/C	SNP	12.67
INTRON	27480704	A/G	SNP	10.67
INTRON	27480720	A/G	SNP	11.33
EXON1	27480787	C/G	SNP	74.67
EXON1	27480836	T/G	SNP	12.00
EXON1	27480856	A/C	SNP	1.33
EXON1	27480917	A/G	SNP	12.67
EXON1	27480926	G/C	SNP	12.00
EXON1	27481348	C/T	SNP	87.33

MAF – Minor Allelic Frequency

the production of terminating codon in the Exon1 region. Correspondingly, Exon2 showed non synonymous SNPs at five positions. Variation in the SNP site 27480376 leads to coding of Glycine instead of Alanine. Similarly, changes at SNP sites 27480379, 27480430, 27480433 and 27480616 lead to the production of Valine, Glutamate and Cysteine instead of Alanine, Valine, Alanine and Tyrosine, respectively; whereas, variation at 27480586 SNP site leads to the production of a termination codon. Besides, a 6bp insertion was found in the Exon2 region at 27480200bp position and a 3bp insertion at 27480535bp position. A single base pair deletion was observed at 27480374 positions and a 9bp deletion at 27480447 positions in the Exon2 region (**Table 1**). A similar investigation performed by Rashid *et al.* (2016) for 795 accessions belonging to both *indica* and *aponica* subspecies revealed an allelic diversity for *SCM2* having 12 non synonymous SNPs and eight synonymous SNPs in both coding and non-coding regions of the genome.

The entire population was grouped into three different haplotypes (H1, H2, H3) using Haploview software. The Haplotype1 (H1) contained 114 accessions whereas Haplotype2 (H2) and Haplotype3 (H3) contained 19 and 17 accessions, respectively. The resulted haplotypes were grouped using three allelic combinations *viz.,* CC, GC and GT (**Fig. 3**). Two significant SNP's (27480778 and 27481339) are responsible for haplotype grouping had an LD r^2 value of 40 (**Table 2**). Correspondingly, the findings of Rashid *et al.* (2016) indicate seven diverse haplotype groups, *viz., SCM2*-1 to *SCM2*-7 based on four SNPs for 378 rice cultivars.

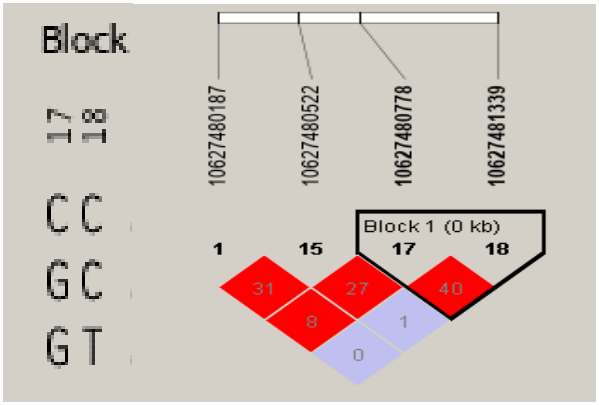

Fig.3. Haplotype formation and Linkage disequilibrium (LD) plot for *SCM2* loci. Numbers in the block indicates LD r^2

Table 2. List of significant SNPs from *SCM2* for the whole population.

Marker No.	SNP sites	Position	HWpval	MAF	Alleles
1	10627480187	27480187	3.1391E-08	0.028	C:G
15	10627480522	27480522	2.3055E-18	0.081	T:C
17	**10627480778**	**27480778**	**3.5592E-36**	**0.24**	**C:G**
18	**10627481339**	**27481339**	**2.671E-23**	**0.113**	**C:T**

HWpval – Hardy Weinberg p value
MAF – Minor Allelic Frequency
Highlighted SNPs were involved in Haplotype formation

Table 3. Descriptive statistics for different haplotypes

Statistics	Whole population	H1	H2	H3
Range (mm)	2.00-5.60	2.00-4.07	3.87-5.50	4.03-5.60
Mean (mm)	3.65	3.35	4.58	4.56
Median (mm)	3.68	3.43	4.6	4.5
SE mean	0.05	0.04	0.1	0.11
Variance	0.48	0.22	0.2	0.2
CV (%)	19.11	13.99	9.86	9.95

Table 4. Dunnett Simultaneous Tests for Level Mean – Control Mean indicating significant difference between haplotypes

Difference of levels	Mean difference	T-value	Adj. p-value
H2 - H1	1.21	10.47	0.00001
H3 - H1	1.227	10.13	0.00001

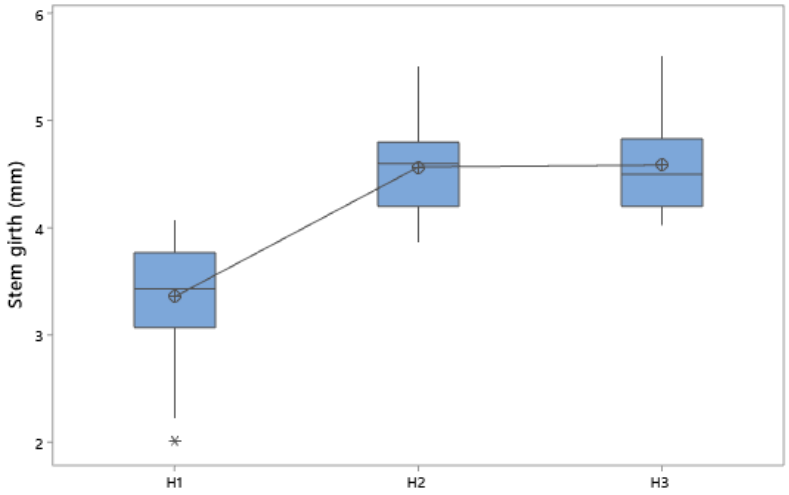

Fig. 4. Box plot of different haplotypes. Line indicates significant difference between haplotypes based on Dunnett's multiple comparison test

Descriptive statistical analysis was performed on individual haplotype groups (**Table 3**). Haplotype H2 had the highest overall mean (4.58 mm) for stem girth with a phenotypic range of 3.87-5.50 mm. The extent of variability in the population was analyzed using CV and H1 was observed to record the highest CV (13.99) (**Table 3**).

The significant difference between the three haplotype groups was analyzed by performing Dunnett's Multiple Comparison test using Minitab 19 statistical software. The test result revealed a significant difference between H1 and H2 haplotypes and H1 and H3 haplotypes (**Fig. 4**). The haplotypes H1-H2 and H1-H3 showed a p-value less than 0.001, indicating a highly significant difference between them (**Table 4**).

In Haplotype1 (H1) with a total of 114 accessions, the genotypes BR IRGA and *PERUNEL* showed the maximum stem girth of 4.07mm whereas, in Haplotype2 (H2) with 13 accessions, the genotypes CICA 9 and MOSHI showed the maximum stem girth of 5.5 mm and 5.4 mm, respectively. In Haplotype3 (H3), with the total accessions of 11, the genotype KITRANA showed a maximum stem girth of 5.6 mm. Among the three haplotype groups, H3 had the most accessions with maximum range for stem girth (4.03–5.60 mm) indicating the superiority of the H3 allele combination over H1 and H2. This provides a reason for considering H3 as a superior haplotype and can be explored for allele mining in future breeding programs.

Lodging in rice production leads to a decrease in grain yield either by breaking or bending of stems, which limits the usage of mechanized methods of harvest. Imparting strong culm confers lodging resistance thereby reducing the yield losses and aids in mechanization. In the present study, haplotypes H3 and H2 had maximum range and mean values for stem girth which can be used for allele mining and superior donor selection in further breeding programs.

REFERENCES

Barrett, J. C., Fry, B., Maller, J. D. M. J. and Daly, M. J. 2005. Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics*, **21**(2): 263-265. [Cross Ref]

Bevan, M. W., Uauy, C., Wulff, B. B., Zhou, J., Krasileva, K. and Clark, M. D. 2017. Genomic innovation for crop improvement. *Nature*, **543**(7645): 346-354. [Cross Ref]

Bidhan Roy, Vikash Kumar and Das, B. K. 2021. Uttar Sona: A medium-slender grain rice variety suitable for cultivation during Boro-season. *Electronic Journal of Plant Breeding*, **12**(2): 298-307. [Cross Ref]

Brown, M. E. and Funk, C. C. 2008. Food security under climate change. *Science,* **319**: 580-581. [Cross Ref]

Hall, T. 1999. Bio Edit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *In Nucleic Acids Symp. Ser.,* **41**: 95-98.

Kashiwagi, T. and Ishimaru, K. 2004. Identification and functional analysis of a locus for improvement of lodging resistance in rice. *Plant physiology*, **134**(2): 676-683. [Cross Ref]

Keerthiraj, B. and Biju, S. 2020. Genetic variability, heritability and genetic advance of yield and lodging-related

traits in rice (*Oryza sativa* L.). *Electronic Journal of Plant Breeding*, **11**(04): 1093-1098. [Cross Ref]

Mansueto, L., Fuentes, R. R., Borja, F. N., Detras, J., Abriol-Santos, J. M., Chebotarov, D. and Alexandrov, N. 2017. Rice SNP-seek database update: new SNPs, INDELs, and queries. *Nucleic acids research*, **45**(D1): D1075-D1081. [Cross Ref]

Merugumala, G. R., Satyanarayana, P. V., Narne, C., BNVSR, R., PV, R. R., Pavani, L. and Deepika, V. 2019. Molecular breeding of "Swarna," a mega rice variety for lodging resistance. *Molecular Breeding*, **39**(4): 1-14. [Cross Ref]

Minitab 17 Statistical Software. 2010. [Computer software] State College, PA: Minitab, Inc. (www.minitab.com)

Nomura, T., Arakawa, N., Yamamoto, T., Ueda, T., Adachi, S., Yonemaru, J. I., Abe, A., Takagi, H., Yokoyama, T. and Ookawa, T. 2019. Next generation long-culm rice with superior lodging resistance and high grain yield, Monster Rice 1. *PloS one*, **14**(8): 0221424. [Cross Ref]

Okuno, A., Hirano, K., Asano, K., Takase, W., Masuda, R., Morinaka, Y., Ueguchi-Tanaka, M., Kitano, H. and Matsuoka, M. 2014. New approach to increasing rice lodging resistance and biomass yield through the use of high gibberellin producing varieties. *PLoS One*, **9**(2): 86870. [Cross Ref]

Ookawa, T., Hobo, T., Yano, M., Murata, K., Ando, T., Miura, H., Asano, K., Ochiai, Y., Ikeda, M., Nishitani, R. and Ebitani, T. 2010. New approach for rice improvement using a pleiotropic QTL gene for lodging resistance and yield. *Nature communications*, **1**(1): 1-11. [Cross Ref]

Rashid, M. A. R., Zhao, Y., Zhang, H., Li, J. and Li, Z. 2016. Nucleotide diversity, natural variation and evolution of Flexible culm-1 and Strong culm-2 lodging resistance genes in rice. *Genome*, **59**(7): 473-483. [Cross Ref]

Weeks, J. P. 2010. plink: An R package for linking mixed-format tests using IRT-based methods. *Journal of Statistical Software*, **35**(12): 1-33. [Cross Ref]