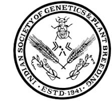
Indian J. Genet. Plant Breed., (2024); 84(1): 38-45
https://doi.org/10.31742/ISGPB.84.1.2
ISSN: 0975-6906



www.isgpb.org

RESEARCH ARTICLE

# Allele mining for *Granule Bound Starch Synthase1* (*GBSS1*) gene governing amylose content in aromatic rice (*Oryza sativa* L.) germplasm

**Deepshikha Dixit[1], N. Siddiqui[1], Haritha Bollinedi, Gopala Krishnan S., Ankit Malik, P. K. Bhowmick, R. K. Ellur, M. Nagarajan[2], K. K. Vinod and A. K. Singh***

## Abstract

The current research investigated sequence polymorphism and allelic variation in the *GBSS1* gene responsible for determining the apparent amylose content (AAC) in an aromatic rice germplasm panel. Analysis of 271 germplasm accessions from India revealed a wide range of variation in AAC, from 4.81% (Manipur Black Rice) to 35.88% (IGSR -3-1-5). Re-sequencing and sequence analysis of a representative subset of 28 accessions unveiled a nucleotide diversity (π) of 0.00374, identifying 72 sequence variants. Two novel non-synonymous SNPs were identified: a G/A transition in exon 2, resulting in the substitution of glutamic acid at the 91$^{st}$ position with lysine, and a C/T transition in exon 11, substituting alanine at the 468$^{th}$ position with valine. Additionally, two previously reported SNPs were validated: A/C in exon 6 (Tyr224Ser) and C/T in exon 10 (Pro415Ser). Furthermore, a 23 bp exon 2 duplication was found in Manipur Black Rice (4.81%), Bhogali Bora (6.20%), and NJ-72 (7.36%), indicating the prevalence of this allele in the very low AAC category (3–9%) in addition to the glutinous or waxy rice category with 0–2% AAC. These findings are significant for manipulating starch content in aromatic rice germplasm.

**Keywords**: Amylose, granule-bound starch synthase I, rice, SNP, waxy allele

## Introduction

Rice (*Oryza sativa* L.), a member of the Poaceae family, stands as one of the foremost cereals in human nutrition and consumed by two-thirds of the global population. It is typically consumed as whole grains after cooking, constituting approximately 40 to 80% of the total caloric intake (Thomas et al. 2013). Aromatic rice represents a distinct subset within the rice family, with grain having characteristic fragrance and shapes ranging from short to long and slender varieties. This subgroup encompasses Basmati types, renowned for their delightful fragrance, long slender grains that remarkably elongate on cooking, having superior eating and cooking quality (ECQ). Elite Basmati cultivars such as Pusa Basmati 1121 and Pusa Basmati 1509, dominate the export market (Singh et al. 2011). Additionally, there are short-grain aromatic varieties like Kalanamak, Jeerega Sambha, Tulai Panji, Govind Bhog, which enjoy niche markets and are equally favored by consumers for their exceptional aroma and grain quality. In recent years, India has witnessed a remarkable surge in foreign exchange earnings from Basmati rice exports, skyrocketing from 0.2 million metric tons in 1990-91 to 4 million metric tons in 2021-22. This unprecedented growth led to a record-breaking foreign exchange earning of Rs. 38,524 crores during the fiscal year 2022-23 (APEDA).

Division of Genetics, ICAR-Indian Agricultural Research Institute (IARI), New Delhi 110 012, India;

[1]Amity Institute of Biotechnology, Amity University, Noida 201 313, Uttar Pradesh, India

[2]Rice Breeding and Genetics Research Centre, ICAR-IARI, Aduthurai 612 101, Tamil Nadu, India.

***Corresponding Author:** A. K. Singh, Division of Genetics, ICAR-Indian Agricultural Research Institute (IARI), New Delhi 110 012, India, E-Mail: aks_gene@yahoo.com

**How to cite this article:** Dixit D., Siddiqui N., Bollinedi H., Krishnan G.S., Malik A., Bhowmick P.K., Ellur R.K., Nagarajan M., Vinod K.K. and Singh A.K. 2024. Allele mining for *Granule Bound Starch Synthase1 (GBSS1)* gene governing amylose content in aromatic rice (*Oryza sativa* L.) germplasm. Indian J. Genet. Plant Breed., **84**(1): 38-45.

**Source of support:** NPTC Project, ICAR-IARI, New Delhi, India.

**Conflict of interest:** None.

**Received:** May 2023 **Revised:** Nov. 2023 **Accepted:** Dec. 2023

*© The Author(s). 2024 Open Access This article is Published by the Indian Society of Genetics & Plant Breeding, NASC Complex, IARI P.O., Pusa Campus, New Delhi 110 012; Online management by www.isgpb.org*

Starch, the principal constituent of rice grains, consists of two glucose polymers: amylose and amylopectin. The relative proportions of these polymers determine the ECQ of rice, with the apparent amylose content (AAC), playing a pivotal role. Rice cultivars are classified into five groups based on their AAC: glutinous (0–2%), very low (3–9%), low (10–19%), intermediate (20–25%), and high (>25%) (Kumar and Khush 1986). Higher levels of amylose result in cooked rice that is firmer upon cooling, less tender, and exhibits a drier, more separate texture, whereas low amylose varieties yield sticky, soft, and glossy cooked rice. Intermediate amylose content (20–25%) produces rice with a softer, fluffier texture that is somewhat sticky. Amylose content significantly influences consumer preference based on firmness and stickiness of the cooked rice texture primarily (Bollinedi et al. 2020). To elucidate the genetic underpinnings of amylose content variation, identifying allelic variants of the gene responsible for amylose synthesis is imperative. Amylose synthesis in developing rice seeds is primarily governed by the *Waxy* (*Wx*) gene, located on chromosome 6, which encodes the Granule-Bound Starch Synthase I (*GBSS1*) enzyme. The level of amylose in rice grains is directly correlated with the abundance of GBSS1 in the endosperm (Mikami et al. 2008).

An array of allelic variants at *GBSS1* locus including $Wx^{a}$, $Wx^{b}$, $Wx^{in}$, $Wx^{op,\ hp}$, $Wx^{mq}$,$Wx^{mp}$, *wx*, $Wx^{lv}$ and $Wx^{la}$ have been identified previously and these alleles have contributed to the tremendous natural variation in rice AC and ultimately affected consumer preferences (Sano 1984; Wang1995; Mikami et al.2008; Larkin and Park 2003; Mikami et al.1999; Liu 2009; Sato et al. 2002; Yang et al. 2013; Wanchana et al. 2003; Zhang et al. 2019; Zhou et al. 2021). $Wx^{a}$, (G/T) generally found in non-glutinous *indica* varieties and wild rice, confers relatively high amylose content (Umemoto and Terashima 1999), whereas $Wx^{b}$ resulted from a point mutation at the 5′ splice junction of the first intron with a mutation from GT-TT, causing the reduction in the mature transcript that confers lower amylose content in glutinous *japonica* varieties (Wang et al. 1995; Hirano et al. 1998; Isshikiet al. 1998). A Cleaved Amplified Polymorphic Sequence (CAPS) marker developed based on this SNP is largely being used to discriminate between high (G allele) and low (T allele) amylose varieties (Ayres et al. 1997).Another diagnostic molecular marker widely utilized is the RM190 $CT_n$ microsatellite, located in the 5′ untranslated region (UTR) of *waxy* exon 1 (Bligh et al. 1995). There are seven known variations of this marker, allowing for the differentiation of genotypes with low, intermediate, and high AAC levels (Ayres et al. 1997; Bergman et al. 2001; Tan and Zhang 2001). Another novel *Wx* allele *i.e.* $Wx^{in}$ carry an A/C mutation in exon 6 is responsible for intermediate levels of amylose content (Larkin and Park 2003). Another variant of *Wx* allele is *Waxy opaque* ($Wx^{op,hp}$) allele having A/G transition in exon 4 resulting in the opaque seed with very low AC (Mikami et al. 1999; Liu et al. 2009). The accessions with $Wx^{op}$ allele have <10% amylose content (Mikami et al. 1999, 2008). Similarly, $Wx^{mp}$ allele evidently originated from the $Wx^{b}$ allele contributes to lower amylose in *japonica* rice due to mutation (G/A) in exon 4 (Yang et al. 2013). Zhang et al. (2019) identified another unique allele termed $Wx^{lv}$ responsible for low viscosity and high AAC. Recently, a novel *Wx* allele known as $Wx^{la}$ that contained combination of both $Wx^{b}$ and $Wx^{in}$ mutations was reported by Zhou et al. (2021).

Finding superior and new alleles to employ in breeding programmes is made necessary by the rising demand for rice varieties with improved eating and cooking qualities. In order to study the diversity and develop allele-specific molecular markers for marker-assisted selection (MAS), the allele mining approach, which is based on the sequencing of various alleles of a single gene in various genotypes within a species, has been largely applied. In this context, in the present study we have analyzed the allelic diversity at *GBSS1* locus in the aromatic Indian rice germplasm accessions including both Basmati and short grain aromatic rice accessions.

## Material and methods

### *Experimental material and field evaluation*

In the present study, a diverse set of 271 aromatic rice germplasm (including both basmati and non-basmati) was grown atresearch farm of the Division of Genetics, ICAR-Indian Agricultural Research Institute, New Delhi (Latitude 28°38′ 23″N, Longitude 77° 09′ 27″ E and Altitude 228.61 m above mean sea level). The seeds were sown on a raised seedbed and 30-days-old seedlings were transplanted in the main field. Each genotype was grown in a single row of 4 m length with a row-to-row spacing of 20 cm and plant to plant spacing of 15 cm. Standard cultural practices for rice were followed to raise a good crop.

### *Sample processing*

Five random plants were selected per genotype and harvested grains were threshed, cleaned and dried to the moisture percentage of 14%. Samples were stored at room temperature in plastic bags for further analysis after proper labeling. Brown rice was obtained by dehulling rough rice with a paddy dehusker using SATAKE™ testing husker (Model THU 35B, Satake Corporation, Hiroshima, Japan) and further polishing of brown rice was done using a milling machine SATAKE™ testing mill (Model TM05, Satake Corporation, Hiroshima, Japan). The milled rice thus obtained was ground into a very fine powder using a motor and pestle. This fine powder was then used for the estimation of AAC.

### *Estimation of apparent amylose content*

The AAC of rice cultivars was estimated according to the protocol of Juliano et al. (1981). One hundred milligrams powder of milled rice was weighed into a test tube to which

1 ml of 95% ethanol and 9 ml of 1N NaOH were added. The sample was warmed up for 10 min in a water bath maintained at 100°C to gelatinize the starch. Then, it was cooled and transferred into a 100 ml volumetric flask and the volume was made up with water. Two millilitres of iodine solution (0.2% I2, 2% KI) was added to 5 ml of the starch solution. The solution was made up to 100 ml with water, shaken well, and let it stand for 20 min, covering with black cloth. The solution absorbance at 620 nm was measured using a spectrophotometer. The amylose-iodine complex has the maximum absorbance at the wavelength of 620 nm. The absorbance value of the amylose-iodine complex is proportional to the amylose content of the ground rice. Therefore, a calibration curve was plotted based on the rice flour standards (Sigma-Aldrich-Amylose, CAS-No. 9005-82-7) with known amylose content to cognate absorbance values. This calibration curve then used to determine the amylose content of a rice sample with unknown amylose content.

### *Selection of genotypes for re-sequencing of GBSS1 gene*

Based on AAC, a subset of 28 genotypes representing cultivars of very low, intermediate and high amylose classes was constituted from these 271 accessions, for re-sequencing the *GBSS1* gene to identify the allelic variants associated with the AAC. The details of the accessions used for re-sequencing are given in Table 1.

### *PCR amplification, sequencing and sequence analysis of GBSS1 gene*

The genomic sequence of *GBSS1*(Os06g0133000, LOC_Os06g04200) gene was downloaded from the NCBI and four of the six primers *viz*., GB 1, GB 2, GB 4, GB 5 were taken from the previous work of Biselli et al. (2014) while two primes namely GB3 and GB6 were newly designed using the Primer3 Plus3.3.0 (Untergasser et al. 2012) to amplify the overlapping fragments from the *GBSS1* gene that together represent the entire gene. The gene was amplified in a set of six overlapping fragments of size ranging from 908 bp to 1,214 bp, together accounting to a total length of 6kb including 1 kb promoter region. The newly designed primers, GB 3 (forward 5′-3′, GATCTGCTCAAAGCTCTGTG) and (Reverse 5′-3′, AGCGTCCTTGTACTGGTCGT), and GB6 (Forward 5′-3′, AGGGGATGAGATACGGAACG) and (reverse 5′-3′, GCACACCCAGAAGAGTACAA) with expected fragment lengths of 908 and 1077, respectively were used for sequencing.

Genomic DNA isolation was carried out using leaf samples collected from 2 week old seedlings by adopting the CTAB DNA extraction protocol (Murray and Thompson 1980). The PCR reaction was carried out using 15 ng of genomic DNA, 1X PCR assay buffer with 1.5 mM $MgCl_2$, 5 pmol of forward and reverse primers, 0.05 mMdNTPs, and 1U of Taq polymerase. The final reaction volume was made

**Table 1.** A list of aromatic germplasm lines used for sequencing

| S. No. | Genotype | Description | Amylose (%) |
|---|---|---|---|
| 1 | Manipur Black | Aromatic short grain | 4.81 |
| 2 | Bhogali Bora | Aromatic short grain | 6.90 |
| 3 | NJ-72 | RIL (NPT11 × Jaya) | 7.36 |
| 4 | ANP 121 | KDML 105 (Traditional short/ Evolved Exotic) | 15.25 |
| 5 | ANP 462 | IR 62873-224-1-6 (Traditional long/ Evolved Exotic) | 16.29 |
| 6 | ANP 471 | Hasan Serai (Traditional long/ Evolved Exotic) | 16.46 |
| 7 | ANP 133 | Pant Sugandh Dhan-15 (Evolved Basmati) | 17.16 |
| 8 | ANP 227 | Sumati (Aromatic short grain) | 17.62 |
| 9 | BGP-138 | P 1570 (IRBB 60 × PB1121) | 17.77 |
| 10 | Thai rice | Aromatic | 18.03 |
| 11 | BGP-126 | P 1551 (IET 17270 × P1302) | 18.06 |
| 12 | ANP 524 | UPR 3429-2-1-1 (Evolved Basmati) | 18.23 |
| 13 | ANP 454 | Haryana Gaurav (Evolved Basmati) | 19.22 |
| 14 | ANP 217 | IET 21953 (UPR 3506-7-1-1) (Evolved Basmati) | 20.17 |
| 15 | ANP 546 | IET 16327 (Evolved Basmati) | 21.57 |
| 16 | ANP 284 | Basmati Jamuna (Traditional Basmati) | 22.32 |
| 17 | ANP 266 | Basmati Kamon (Traditional Basmati) | 23.30 |
| 18 | ANP 245 | Ranbir Basmati (Traditional Basmati) | 24.38 |
| 19 | ANP 262 | Basmati 6129 (Traditional Basmati) | 25.25 |
| 20 | ANP 173 | Elayachi (Aromatic short grain) | 26.26 |
| 21 | ANP 166 | Chhatri (Aromatic short grain) | 27.22 |
| 22 | ANP 45 | RAU 3076 (Aromatic short grain) | 28.32 |
| 23 | ANP 320 | RRB 2005-1 (Aromatic short grain) | 29.28 |
| 24 | ANP 487 | Hari Shankar (Aromatic short grain) | 30.29 |
| 25 | ANP 486 | KOLIHA (Aromatic short grain) | 31.28 |
| 26 | ANP 235 | GR-104 (Aromatic short grain) | 32.20 |
| 27 | ANP 10 | RAU 3061 (Aromatic short grain) | 33.51 |
| 28 | ANP 61 | IGSR-3-1-5 (Aromatic short grain) | 35.88 |

up to 10 µl using nuclease-free water. PCR amplification on Eppendorf Thermocycler (Eppendorf, Germany) was performed with the following profile: 94°C per 4 min followed by 40 cycles of 94°C per 40 sec, 60°C per 50 sec and elongation at 72°C for 1 min and a final extension at 72°C for 10 min. The PCR products were resolved on 2% agarose gel with ethidium bromide and images were capture using a gel documentation system (BIO-RAD Gel Doc XR+, California, United States). PCR products of appropriate size were eluted from the gel using the QIA-quick® Gel Extraction kit (Qiagen) and purified products were sequenced using Sanger's chemistry-based ABI 3730XL DNA Analyzer Sequencer (ABI, Applied Biosystems Amersham). Genotypes were sequenced in replicates both in forward and reverse directions.

The sequence reads from each variety were assembled into a single contigs using DNA software Sequencher® version 5.4.6 (http://www.genecodes.com). The contigs were then aligned to the reference *GBSS1* sequence from Nipponbare in order to identify the sequence variants from each genotype.

### *Nucleotide Diversity and Haplotype analysis*

The assembled contigs of all the 28 varieties were aligned using an online alignment tool, Multiple alignments using the fast Fourier transform (MAFFT) (Katoh et al. 2019) and phylogenetic analysis was carried out using the software DnaSP v6 (Rozas et al. 2017). For haplotype analysis, the haplotypes were generated in DnaSP and the median-joining haplotype network diagram was visualized in Network software (Leigh and Bryant 2015).

## Results

### *Assessment of AAC in a rice germplasm collection*

The aromatic rice accessions analysed in the study depicted enormous variability for AAC which ranged from 4.81% (Manipur Black Rice) to 35.88% (IGSR -3-1-5). When grouped into different AAC classes, no accessions were found in the waxy (0–2%), 3 in very low (3–9%), 12 in low (10–19%), 78 in intermediate (20–25%) and 178 accessions in high amylose category (>25) were observed in the study. The results reveal predominance of intermediate and high amylose genotypes in the aromatic rice cultivars.

### *Discovery of sequence polymorphisms in GBSS1 gene*

Sequence analysis of the *GBSS1* gene revealed a total of 72 sequence variants in these 28 aromatic accessions. Of these 72 variants, 57 were SNPs while 15 were InDels (8 deletions and 7 Insertions) in the size range of 1 to 23 bp. Of the 57 SNPs identified here, 27 were present in the promoter region, 12 in the UTR, 13 in the introns while only 5 SNPs were observed in the exonic regions. Of all the SNPs, 79% (45) were transition substitutions while 21% (12) were transversions. The 5 exonic SNPs were distributed one each in exons 2 (G/A polymorphism at position 1,570); exon 6 (A/C polymorphism at position 2,385); exon 9 (T/C polymorphism at position 3,103); exon 10 (C/T polymorphism at position 3,377) and exon 11 (C/T polymorphism at 3,780). The SNP A/C in exon 6 is a transversion while all the other exonic SNPs were the results of transition mutations (Fig. 1). The study identified seven insertions in the *GBSS1* gene of the aromatic accessions. The 23 bp insertion located in the exon 2 is the prominent among the InDels and it is specifically observed in the low amylose accessions Manipur Black rice, Bhogali Bora and NJ 72. The details of other InDels including their size, location and sequence polymorphism are given in Table 2.

### *Nucleotide diversity analysis of GBSS1 alleles*

Nucleotide polymorphism analysis of *GBSS1* sequences from the 28 accessions carried out in DnaSP showed average pairwise nucleotide diversity (π) as 0.00374 and the sliding window analysis of π in DnaSP showed significant diversity in the 200-850 bp, 1300-1600 bp, 4200-4300 bp and 4600-4800 bp regions coinciding with the promoter; UTR; exon 9 and intron 9, and intron 10 and exon 11 regions (Fig. 2).

The silent Watterson's nucleotide diversity estimator, Theta per site from S ($\theta$w) and the Theta from Eta over the *GBSS1* alleles were found to be same as 0.00241 while Theta per sequence from S was 14.514. The average number of nucleotide differences, *k* was found to be 22.417. The Tajima's D statistic is positive (2.079) and statistically significant at P=0.05.

### *Haplotype analysis based on the exonic SNPs*

Haplotype analysis carried out based on five high quality SNPs in the protein coding region showed nine haplotypes with haplotype diversity (ouz). Hap3 is the major haplotype encompassing 8 of the 28 accessions while Hap6 is the smallest haplotype with only one genotype Ranbir Basmati in it (Fig. 3). Haplotypes 1, 2 and 8 diverged from Hap3 through three independent mutational events i.e. Hap3 differs from Hap1, Hap2 and Hap8 based on the SNP T/C in exon 9, SNP G/A in exon 2 and the SNP T/C in exon 10 respectively. Further association analysis among the haplotypes revealed that Hap4was significantlyassociated with Hap7 (P=0.043*) and Hap9 (P=0.038*). The Hap5 and Hap7 were also significantly associated with Hap9 and found

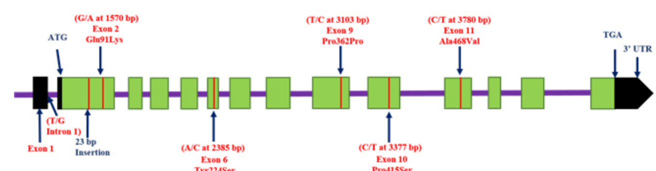



**Fig. 1.** Representation of waxy gene (*GBSS1*) structure and identification of SNPs in *Wx* locus. The boxes denote the exon and lines between boxes indicate the intron

**Table 2.** Waxy gene Insertion and deletion distribution in aromatic rice germplasm

| S. No. | Type | Physical position | Location | Reference | Variant |
|---|---|---|---|---|---|
| 1 | Insertion | 1387 bp-1410 bp | Exon 2 | - | ACGGGT TCCAGG GCCTCAA GCCC |
| 2 | Insertion | 1765 bp | Exon 3 | - | G |
| 3 | Insertion | 3755 bp | Exon 11 | - | G |
| 4 | Insertion | 4679 bp | UTR | - | G |
| 5 | Insertion | 4722 bp | UTR | - | G |
| 6 | Insertion | 4800 bp | UTR | - | G |
| 7 | Insertion | 4828 bp | UTR | - | G |
| 8 | Deletion | 83 bp | UTR | C | - |
| 9 | Deletion | 84 bp | UTR | T | - |
| 10 | Deletion | 966 bp | UTR | A | - |
| 11 | Deletion | 2318 bp | Intron 5 | C | - |
| 12 | Deletion | 3020 bp | Exon 9 | A | - |
| 13 | Deletion | 3543 bp | Intron 10 | A | - |
| 14 | Deletion | 3544 bp | Intron 10 | T | - |
| 15 | Deletion | 3545 bp | Intron 10 | A | - |

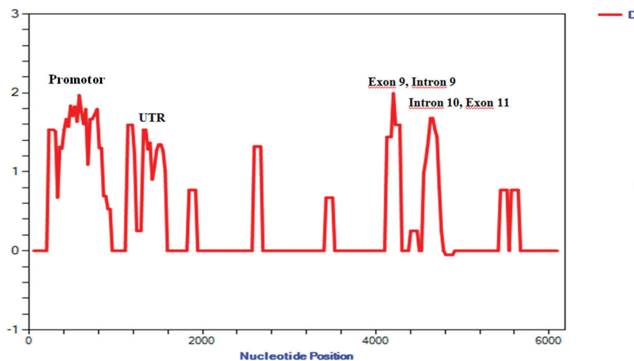
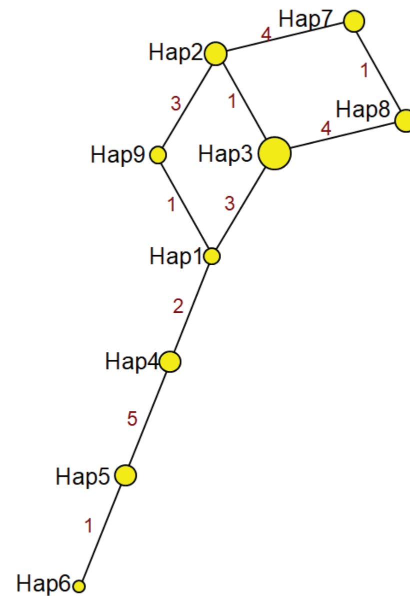



Fig. 2. Nucleotide diversity in 28 aromatic rice accessions

to be statistically significant at P value 0.014* and 0.031* respectively as assessed by the unpaired t-test.

## Discussion

Eating and cooking quality determined by the AAC content, is the major criteria affecting the consumer preference and market value of the rice varieties. Hence, manipulating the starch composition is one of the critical criteria in rice improvement programmes. Aromatic rice is a special type of rice highly valued among the consumers for their pleasant aroma and unique taste. These varieties are often utilized for the preparations of special dishes like biryani, flavoured rice and some sweet dishes like kheer etc. In the present study, we have analysed the variation for AAC content, the major component of rice starch, in the aromatic rice germplasm and observed significant variation in the range of 4.81%




Hap1 = Nipponbare, Nj-72; Hap2 = KDML 105, IR 62873-224-1-6, Haryana Gaurav, IET 21953, IGSR-3-1-5; Hap3 = Hasan Serai, Pant Sugandh Dhan-15, Sumati, UPR 3429-2-1-1, Hari Shankar, Koliha, Manipur black, Bhogali Bora; Hap4 = IET 16327, Basmati Jamuna, Basmati 6129; Hap5 = Basmati Kamon, Elayachi, Chhatri; Hap6 = Ranbir Basmati; Hap7 = RAU 3076, RAU 3061; Hap8 = RRB 2005-1, GR-104, Thai rice; Hap9 = P 1551, P 1570

**Fig. 3.** Haplotypes based on exonic $SNP_S$

(Manipur Black rice) to 35.88% (IGSR-3-1-5). This range is much higher than reported by the previously by Vemireddy et al. (2014) based on 58 aromatic accessions from India (8.8% to 27.9%), Siriamornpun et al. (2016) based on 25 aromatic rice varieties cultivated in Thailand (13.1% to 25.6%) and Mian et al. (2017) based on 32 aromatic rice germplasm from Bangladesh (12.5% to 26.5%). The accessions with higher AAC content, identified in this study, can be utilized as donors in the breeding programmes aiming at manipulating the rice starch composition of aromatic accessions in order to develop low Glycaemic Index (GI) rice varieties as AAC is significantly associated with GI (Chung et al. 2011; Bao et al. 2017; Zhou et al. 2018).

Sequence analysis of 6000 bp(6 kb) region of *GBSS1* gene encompassing 1.1 kb promoter and 13 exons and 12 introns, revealed the presence of significant sequence variants among the aromatic rice germplasm. The frequency of SNPs in the intronic regions was almost 3 folds higher than in exons. Of the 5 exonic SNPs identified in the study, four SNPs resulted in non-synonymous substitution of amino acid while the SNP in exon 9 resulted in synonymous substitution (Pro362Pro). Of the four non-synonymous SNPs, two SNPs *viz.,* A/C in exon 6 (Tyr224Ser) and C/T in exon 10 (Pro415Ser) have been reported previously by Larkin and Park (2003). According to them, all the accessions with intermediate range of AAC carried the C allele in exon 6 while the accessions in low and high AAC category either carried A or C allele at this locus. Chen et al. (2008) also reported the

predominance of C allele in the intermediate AAC class (21 to 22%) after screening a set of 171 accessions originating from 43 countries. Similarly, in the present study we observed that the accessions in the AAC range of 21.56 to 27.21% (IET 16327, Basmati Jamuna, Basmati Kamon, Ranbir Basmati, Basmati 6129, Elayachi, Chhatri) carried the C allele at exon 6 while all other accessions both in low and high AAC range carried the A allele in contrast to either A or C allele reported by Larkin and Park (2003). The C/T SNP in the exon 10 is not found associated with any AAC class as C and T alleles are distributed randomly across the low, intermediate and high AAC classes while Larkin and Park (2003) reported the predominance of T allele in the high AAC class although few accessions in the high AAC class also carried the C allele. The differences in the two studies could be attributed to the different accessions included in the analysis.

Additionally, the two non-synonymous SNPs in exons 2 and 11 are novel and identified for the first time in this study. The G/A SNP in the exon 2 resulted in the substitution of glutamic acid at 91$^{st}$ position to lysine residue. This substitution does not appear to significantly influence the *GBSS1* activity as G/A SNP did not show any significant association with AAC content. The C/T SNP in exon 11 resulted in the substitution of alanine residue at 468$^{th}$ position to valine. Only four of the 28 accessions *viz.,* Basmati Kamon, Ranbir Basmati, Elayachi, Chhatricarried the T allele at this locus while all the remaining accessions showed C allele.

Wanchana et al. (2003) reported the 23 bp duplication (insertion) in exon 2 of the waxy gene, which originated from tropical glutinous rice. This duplication resulted in a frame shift mutation leading to the formation of pre-mature stop codon and production of non-functional GBSS1 protein thereby zero AAC in the glutinous rice cultivars. Later, Biselli et al. (2014) also reported the 23 bp duplication in a glutinous rice cultivar Calmochi 101. However, in this study, we observed the 23 bp exon 2 duplication in the accessions Manipur black rice (4.81%), Bhogali Bora (6.20%) and NJ-72 (7.36%). So, the study reveals the distribution of 23 bp exon 2 duplication in the very low AAC category (3–9%) in addition to the glutinous or waxy rice with 0–2% AAC.

The SNP G/T was discovered near the 5' splicing location of the first intron in aromatic rice germplasm lines. Due to this mutation the improper splicing of pre-mRNA results in 10-fold less *GBSS1* protein, leading in lower AC (Wang et al. 1995). In the present study accessions with very low (3–9%) as well as low (10–19%) AAC were closely associated to the T SNP in the first intron, while genotypes with intermediate and high amylose class showing the presence of G (Ayres et al. 1997; Larkin and Park 1999). These results are in line with earlier research found that waxy cultivars with T in the first intron had lower amylose content (Hirano 1998; Yamanaka et al. 2004). Also, Ayres et al. (1997) and Larkin and Park (1999) reported that cultivars with high or intermediate AAC in rice germplasm from the United States and Europe were connected to the G genotype in the 1st intron. This SNP, T in the 1st intron can therefore be utilized to differentiate between low and intermediate to high amylose content groups (Ayres et al. 1997; Liu et al. 2009). These studies revealed that the T SNP in the 1st intron was crucial for controlling the *GBSS* mRNA processing and steady levels, as well as the following low amylose in these cultivars.

It is generally accepted that the process of domestication decreases the diversity of nucleotides at genes associated to domestication, which regulate particular features chosen during domestication. This is demonstrated by the fact that aromatic rice has significantly less nucleotide diversity ($\pi$ = 0.00374) at the *GBSS1* locus. Low nucleotide diversity is frequently found in genes associated with particular traits (Olesen and Purugganan 2002; Zhang et al. 2009). The present results also suggested that the sequence diversity of the *GBBSI* genes was significantly decreased by domestication. Diversity of the *GBSS1* gene showed that the majority of polymorphic sites were found near the start and end of the promoter region, and the coding region of *GBSS1* gene had much less variation than the non-coding region. Similar kind of studies were also reported previously about domestication and decreased diversity in *Wx* gene by Olesen and Purugganan (2002) and Qiao et al. (2012). In the present study, positive values of Tajima D values (2.079) indicate an excess of intermediate frequency polymorphisms which could be attributable to balancing or overdominant selection (Biswas and Akey 2006).

The haplotype network constructed based on the five exonic SNPs in the *GBSS1* gene demonstrated the genetic relatedness of the accessions used in the study. The traditional non-Basmati short grain aromatic cultivar Bogali Bora clustered with Manipur black rice in Hap3, the largest haplotype while the Basmati types were found dispersed in haplotypes Hap4, Hap5 and Hap6. This indicates divergence of Basmati types from aromatic non-Basmati accessions at the *GBSS1* locus.

Amylose, one of the two major components of rice starch is the primary determinant of an array of eating and cooking quality attributes in rice. The ECQs are particularly important in the aromatic rice specially priced for their unique aroma, taste and mouth feel. Although, previous studies reported allelic variation in *GBSS1* gene of rice, the status of this locus in aromatic rice germplasm has not been much explored. The purpose of the study was to assess the sequence variation and polymorphism in the *GBSS1* gene of the aromatic rice germplasm. The study revealed significant variation for AAC in aromatic rice *vis-a-vis* assessing the sequence diversity in *GBSS1* locus. The novel variants identified in the study would help in differentiating aromatic rice accessions belonging to different categories of AAC and

thus hold value in breeding programs aiming at altering the starch composition in rice grain.

## Authorts' contributions

Conceptualization of research (AKS, HB); Designing of the experiments (AKS, HB, NS); Contribution of experimental materials (GKS, PKB); Execution of field/lab experiments and data collection (DD); Analysis of data and interpretation (DD, HB ); Preparation of the manuscript (DD, HB, AM, RKE, MN, KKV).

## Acknowledgement

The authors are grateful to ICAR-Division of Genetics, Indian Agricultural Research Institute, New Delhi, sanctioned NPTC project and provided funds for this research work.

## References

Ayres N. M., McClung A. M., Larkin P. D., Bligh H. F. J., Jones C. A. and Park W. D. 1997. Microsatellite and single nucleotide polymorphism differentiate apparent amylose classes in an extended pedigree of US rice germplasm. Theor. Appl. Genet., **94**: 773-781.

Bao J S., Zhou X., Xu F F., He Q. and Park Y. 2017. Genome-wide association study of the resistant starch content in rice grains. Starch-Starke, **69**: 1600343.

Bergman C. J., Delgado J. T., McClung A. M. and Fjellstrom R. G. 2001. An improved method for using a microsatellite in the rice waxy gene to determine amylose class. Cereal Chem., **78**: 257-260

Biselli C., Cavalluzzo D., Perrini R., Gianinetti A., Bagnaresi P., Urso S., Orasen G., Desiderio F., Lupotto E. and Cattivelli L. et al. 2014. Improvement of marker-based predictability of apparent amylose content in japonica rice through GBSS1 allele mining. Rice, **1**: 1-12.

Biswas S. and Akey J.M., 2006. Genomic insights into positive selection. Trend. Genet., **22**(8): 437-446.

Bligh H F G., Till R. J. and Jones C. A. 1995. A microsatellite sequence closely linked to the Waxy gene of *Oryza sativa*. Euphytica, **86**: 83-85.

Bollinedi H., Vinod K. K., Bisht K., Chauhan A., Krishnan S. G., Bhowmick P. K., Nagarajan M., Rao D. S., Ellur R. K. and Singh A. K. 2020. Characterising the diversity of grain nutritional and physico-chemical quality in Indian rice landraces by multivariate genetic analyses. Indian J. Genet. Breed., **80**(1): 26-38. doi: 10.31742/IJGPB.80.1.4

Cai X. L., Wang Z.Y., Xing, Y. Y., Zhang J. L. and Hong M. M. 1998. Aberrant splicing of intron 1 leads to the heterogeneous 5'UTR and decreased expression of waxy gene in rice cultivars of intermediate amylose content. Plant J., **14**: 459-465.

Chen M. H., Bergman C., Pinson S. and Fjellstrom R. 2008. Waxy gene haplotypes: Associations with apparent amylose content and the effect by the environment in an international rice germplasm collection. J. Cereal Sci., **47**: 536-545.

Chung H J., Liu Q., Lee L. and Wei D. Z. 2011. Relationship between the structure, physicochemical properties and in vitro digestibility of rice starches with different amylose contents. Food Hydrocolloids, **25**(5): 968-975.

Hirano Y. H., EiguchiM. and Sano Y. 1998. A single base change altered the regulation of Waxy gene at the post-transcriptional level during evolution of rice. Mol. Biol. Evol., **15**: 978-987.

Isshiki M., Morino K., Nakajima M., Okagaki R. J., Wessler S. R., Izawa T. and Shimamoto K. 1998. A naturally occurring functional allele of the rice waxy locus has a GT to TT mutation at the 5 splice site of the Wrst intron. Plant J., **15**: 133-138.

Juliano B. O., Perez C. M., Blakeney A. S., Castillot D., Kongseree N., Laingnelet B., Lapis E. T., Murty V. S., Paule C. M. and Webb B. D. 1981. International cooperative testing on the amylose content of milled rice. Starch/Staerke, **33**: 157.

Katoh K., Rozewicki J. and Yamada K. D. 2019. MAFFT online service: Multiple sequence alignment, interactive sequence choice and visualization. Brief. Bioinformatics, **20(**4): 1160-1166.

Kumar I. and Khush G. S. 1986. Gene dosage effects of amylose content in rice endosperm. Jpn. J. Genet., **61**: 559-568.

Larkin P. D. and Park W. D. 2003. Association of waxy gene single nucleotide polymorphism with starch characteristics in rice. Mol. Breed., **12**: 335-339.

Larkin P.D. and Park W.D. 1999. Transcript accumulation and utilization of alternate and non-consensus splice sites in rice granule-bound starch synthase are temperature-sensitive and controlled by a single-nucleotide polymorphism. Plant Mol. Biol., **40**: 719-727.

Leigh J. W. and Bryant D. 2015. Popart: Full-feature software for haplotype network construction. Method. Ecol. Evol., **6**(9): 1110-1116.

Liu L. L., Ma X.D., Liu S.J., Zhu C. L., Jiang L., Wang Y. H., Shen Y., Ren Y. L., Dong H., Chen L.M. et al. 2009. Identification and characterization of a novel waxy allele from Yunnan rice landrace. Plant Mol. Biol., **71**: 609-626.

Mian M. H., Karim A. J. M. S., Bhuiyan M. J. U., Ahmed A. U. and Rashid M. H. A. 2017. Genetic diversity of aromatic rice germplasm of Bangladesh based on SSR markers and amylose content. J. Food Sci. Technol., **54**(10): 3063-3072.

Mikami I., Aikawa M., Hirano H. Y. and Sano Y. 1999. Altered tissue-specific expression at the Wx gene of the opaque mutants in rice. Euphytica, **105**: 91-99.

Mikami I., Uwatoko N., Ikeda Y., Yamaguchi J., Hirano H. Y., Suzuki Y. and Sano Y. 2008. Allelic diversification at the wx locus in landraces of Asian rice. Theor. Appl. Genet., **116**: 979-989.

Murray M. G. and W. F. Thompson. 1980. Rapid isolation of high molecular weight plant DNA. Nucleic Acids Res., **8**: 4321-4325.

Olsen K.M. and Purugganan M.D. 2002. Molecular evidence on the origin and evolution of glutinous rice. Genetics, **162**(2): 941-950.

Qiao W., Chen Y., Wang R., Wei X., Cao L., Zhang W. and Yang Q. 2012. Nucleotide diversity in waxy gene and validation of single nucleotide polymorphism in relation to amylose content in Chinese microcore rice germplasm. Crop Sci., **52**(4): 1689-1697.

Rozas J., Ferrer-Mata A., Sanchez-DelBarrio J. C., Guirao-Rico S., Librado P., Ramos-Onsins S. E. and Sanchez-Gracia A. 2017. DnaSP 6: DNA Sequence Polymorphism analysis of large data sets. Mol. Biol. Evol., **34**(12): 3299-3302.

Sano Y. 1984. Differential regulation of waxy gene expression in rice endosperm. Theor. Appl. Genet., **68**: 467-473.

Sato H., Suzuki Y., Sakai M. and Imbe T. 2002. Molecular characterization of Wx-mq, a novel mutant gene for low-amylose content in endosperm of rice (*Oryza sativa* L.). Breed. Sci., **52**: 131-135

Singh A. K., Gopalakrishnan S., SinghV. P., PrabhuK. V., Mohapatra

T., Singh N. K., Sharma T. R., Nagarajan M., Vinod K. K., Singh D., Singh U. D., Chander S., Atwal S. S., Seth R., Singh V. K., Ellur R. K., Singh A., Anand D., Khanna A., Yadav S., Goel N., Singh A., Shikari A. B., Singh A. and Marathi B. 2011. Marker assisted selection: a paradigm shift in Basmati breeding. Indian J. Genet. Plant Breed., **71**(2): 120-128.

Siriamornpun S., Tangkhawanit E., Kaewka K. and Songjinda P. 2016. Amylose content and gelatinization properties of Thai aromatic rice varieties. Int. J. Food Properties, **19**(2): 276-286.

Tan Y. F. and Zhang Q. F. 2001. Correlation of simple sequence repeat (SSR) variants inthe leader sequence of the waxy gene with amylose of the grain in rice. Acta Bot. Sin., **43**: 146-150

Thomas R., Wan-Nadiah W. A. and Bhat R. 2013. Physiochemical properties, proximate composition, and cooking qualities of locally grown and imported rice varieties marketed in Penang, Malaysia. Int. Food Res. J., **20**(3): 1345-1351.

Umemoto T. and K. Terashima. 1999. Differences in amylopectin structure between two rice varieties in relation to the effects of temperature. Starch., **51**: 58-62.

Vemireddy L. R., Reddy S. Y. and Reddy B. V. 2014. Genetic variability and correlation studies for amylose content and gelatinization temperature among aromatic rice (*Oryza sativa* L.) germplasm. J. Crop Sci. Biotech., **17**(2): 67-74.

Wanchana S., Toojinda T., Tragoonrung S. and Vanavichit A. 2003. Duplicated coding sequence in the waxy allele of tropical glutinous rice (*Oryza sativa* L.). Plant Sci., **165**: 1193-1199.

Wang Z. Y., Zheng F. Q., Shen G. Z., Gao J. P., Snustad D.P., Li M. G., Zhang J. L. and Hong M. M. 1995. The amylose content in rice endosperm is related to the post-transcriptional regulation of the waxy gene. Plant J., **7**: 613-622.

Yamanaka S., Nakamura I., Watanabe K.N. and Sato Y.I. 2004. Identification of SNPs in the waxy gene among glutinous rice cultivars and their evolutionary significance during the domestication process of rice. Theor. Appl. Genet., **108**: 1200-12.

Yang J., Wang J., Fan F., Zhu J., Chen T., Wang C., Zheng T., Zhang J., Zhong W. and Xu J. 2013. Development of AS-PCR marker based on a key mutation confirmed by resequencing of Wx-mp in Milky Princess and its application in japonica soft rice (*Oryza sativa* L.) breeding. Plant Breed., **132**: 595-603.

Zhang C., Zhu J., Chen S., Fan X., Li Q., Lu Y., Wang M., Yu H., Yi C. and Tang S. 2019. $Wx^{lv}$, the Ancestral Allele of Rice waxy gene. Mol. Plant., **12**: 1157-1166.

Zhang D., Zhang H., Wang M., Sun J., Qi Y., Wang F., Wei X., Han L., Wang X. and Li Z. 2009. Genetic structure and differentiation of *Oryza sativa* L. in China revealed by microsatellites. Theor. Appl. Genet., **119**: 1105-1117.

Zhou H., Xia D., Zhao D., Li Y., Li P., Wu B., Gao G., Zhang Q., Wang G., Xiao J. et al. 2021. The origin of $Wx^{la}$ provides new insights into the improvement of grain quality in rice. J. Integr. Plant Biol., **63**: 878–888.

Zhou X., Ying Y N., Hu B L., Pang Y H. and Bao J. S. 2018. Physicochemical properties and digestibility of endosperm starches in four indica rice mutants. Carbohydr Polym, **195**: 1–8.