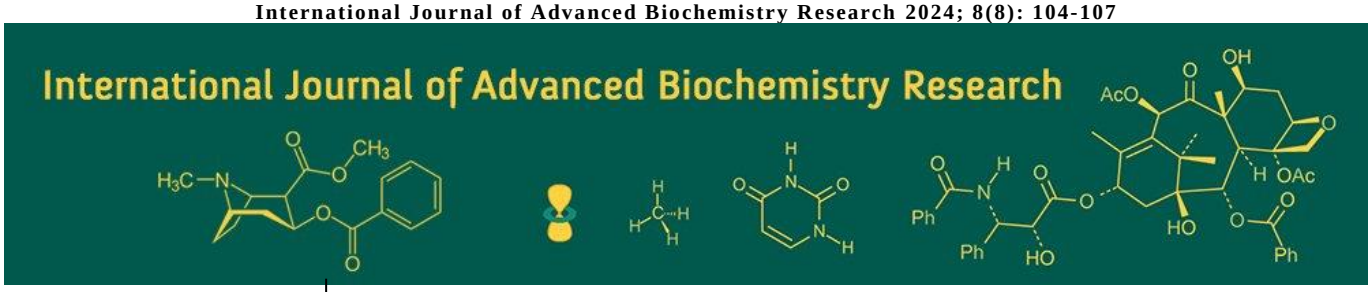

ISSN Print: 2617-4693
ISSN Online: 2617-4707
IJABR 2024; 8(8): 104-107
www.biochemjournal.com
Received: 03-06-2024
Accepted: 06-07-2024

Praveen Kumar R
ICAR-Indian Institute of Rice Research, Hyderabad, Telangana, India

Sathish Kumar P
ICAR-Indian Institute of Rice Research, Hyderabad, Telangana, India

MSRSC Sekhar
ICAR-Indian Institute of Rice Research, Hyderabad, Telangana, India

A Krishna Satya
Department of Biotechnology, Acharya Nagarjuna University, Andhra Pradesh, India

P Sudhakar
Department of Biotechnology, Acharya Nagarjuna University, Andhra Pradesh, India

LV Subba Rao
ICAR-Indian Institute of Rice Research, Hyderabad, Telangana, India

Corresponding Author:
LV Subba Rao
ICAR-Indian Institute of Rice Research, Hyderabad, Telangana, India

Allele mining of Badh1 gene in aromatic rice germplasm

Praveen Kumar R, Sathish Kumar P, MSRSC Sekhar, A Krishna Satya, P Sudhakar and LV Subba Rao

DOI: https://doi.org/10.33545/26174693.2024.v8.i8b.1707

Abstract
In this study, we explored the allelic diversity of the Badh1 gene in rice using a core set of 99+2 allele sequences. Phylogenetic analysis revealed seven distinct clusters, which were compared with haplogroups based on whole gene sequences of synonymous (Syn) and non-synonymous (non-Syn) SNPs. Detailed SNP and indel (insertion-deletion) analyses were conducted on alleles from six different categories of rice, highlighting significant variations in the exonic and intronic regions. These findings provide insights into the genetic diversity and evolutionary patterns of the Badh1 gene in rice. We also discovered the glutamic acid and cysteine active sites, as well as the N- and C-terminals, in the reference protein. All 99+2 genotypes were also found at these places. In certain accessions, the active sites for cysteine, glutamic acid, and C-terminal were not detected. Additionally, those genotypes lack 2-AP.

Keywords: Badh1, core set, SNP and indel

Introduction
The Badh1 gene plays a crucial role in rice aroma and stress tolerance. Understanding its allelic diversity is essential for rice breeding programs aimed at improving these traits. This study aims to analyze the phylogenetic diversity of Badh1 alleles and to identify SNPs and indels within different categories of rice accessions.

A suitable organic solute, glycinebetaine (GB) is produced by many different organisms, including bacteria, animals, and plants, in response to stressors such as salt, drought, and temperature changes. Betaine is an enzyme. E.C. No. 1.2.1.8] aldehyde dehydrogenase (BADH) is engaged in the process of creating GB from its precursor Aldehyde betaine. Numerous flowering plants, such as sorghum, barley, amaranth, spinach, and mangroves, are proven betaine accumulators and tolerate salt and drought stress partly through this mechanism, but other species like tobacco, tomato and rice are considered non-accumulators of GB (Ishitani *et al*. 1993; Rathinasabapathi *et al*. 1993; Shirasawa *et al*. 2006) [2, 9, 10]. The transformation of the Badh gene from bacterial and plant sources into betaine-deficient plant species has led to the accumulation of glycine betaine in their systems, resulting in increased tolerance to salt and drought stress. (Liang *et al*. 1997; Mohanty *et al*. 2002) [4, 6]. In response to salt and drought stress, the synthesis of BADH is significantly upregulated in the leaves of spinach, barley, and sorghum. (Weretilnyk and Hanson 1990; Ishitani *et al*. 1995; Wood *et al*. 1996) [12, 3, 13]. Rice (*Oryza sativa* L.) is generally considered a non-accumulator of GB, although it does express BADH at low levels. (Fitzgerald *et al*. 2008) [1] This interest leads to an exploration of the phylogenetic evolution of the enzyme in rice and a search for variations in the BADH gene sequence within the rice germplasm, along with its association with important traits. Rice possesses two functional genes encoding the BADH enzyme: BADH1, located on chromosome 4, and BADH2, located on chromosome 8. Both genes contain 15 exons and exhibit high sequence homology with their orthologs in other species. BADH1 in rice is induced by salt and water stresses, while BADH2 is expressed constitutively at low levels. The expression of both genes also appears to be regulated by post-translational processing mediated by paired short direct repeats in response to stress. (Niu *et al*. 2007) [7].

Materials and Methods

Plant Materials

A total of 101 rice accessions, comprising 99 aromatic and 2 non-aromatic varieties, were selected for allele mining of the Badh1 gene. These accessions were sourced from various geographical locations across India known for aromatic rice cultivation. The seeds were multiplied and maintained at the Indian Institute of Rice Research, Hyderabad, for field evaluation. Additionally, seeds were multiplied in the Geographical Indicator (GI) region for 2-AP analysis, following the procedure described by Peddamma *et al*. (2018) [8]. The 99 aromatic accessions were categorized into five groups (Supplementary Table 1):

- Aromatic Short Grain (ASG)-Indigenous (41 accessions)
- ASG-Exotic (13 accessions)
- Traditional Basmati Indigenous (15 accessions)
- Evolved Basmati Indigenous (14 accessions)
- Traditional Evolved and Exotic Basmati (16 accessions)
- Non-aromatic (2 accessions)

PCR Amplification and Sequencing

Genomic DNA was extracted from the leaves of 20-day-old seedlings using a modified cetyltrimethylammonium bromide (CTAB) method. The Badh1 gene sequence was retrieved from the Nipponbare reference genome in the NCBI database (www.ncbi.nlm.nih.gov). Ten different forward and reverse primer sets were designed to cover the entire reference Badh1 gene sequence using Primer 3 software (Figure 1 and Table 1). Annealing temperatures were optimized for all primers using gradient PCR. The PCR was conducted in a 45 µL reaction volume containing 40 ng/µL genomic DNA, 1 mM dNTPs, 5 µM primers, 1.5 U high-fidelity Taq DNA polymerase (Fermentas), Taq buffer, and nuclease-free water. The PCR conditions included an initial denaturation at 94 °C for 5 minutes, followed by 35 cycles of denaturation at 94 °C for 30 seconds, annealing at 54-64 °C (depending on the primer) for 1 minute, extension at 72 °C for 1 minute, and a final extension at 72 °C for 10 minutes. Amplicons were resolved on a 3% agarose gel to confirm their sizes, purified using EXOSAP, and sequenced using Sanger's deoxy method. Each amplicon was sequenced twice using specific forward and reverse primers, and contigs were assembled using DNA Baser software.

Sequence Analysis of Badh1 Alleles

The contig sequences from individual genotypes were aligned with the reference Badh1 sequence using MEGA 4.1 software (Tamura *et al*., 2007) [11]. SNPs (synonymous and non-synonymous), indels, Ka/Ks ratio, and Tajima's D values were determined using DnaSP 5.0 software (Librado *et al*., 2009) [5]. The alleles were also processed using the FGENESH gene prediction tool (www.softberry.com) to determine gene structure. Coding regions of alleles were translated, and protein sequence comparisons with reference protein sequences were conducted using BioEdit software. Haplotype diversity was analyzed by including all polymorphic sites in the total contig sequences using TASSEL V5 software. Phylogenetic relationships among all sequenced Badh1 alleles, including the Nipponbare reference allele, were determined using MEGA V7 software with the neighbor-joining method and 1000 bootstrap replications.

Results and Discussion

Results: Seven distinct clusters were identified, as shown in Table 1.

Table 1: Seven distinct clusters were identified, as shown

S. No	MEGA Cluster/Group	No. of Accessions
1	Group 1A	10
2	Group 1B	9
3	Group 2	7
4	Group 3	2
5	Group 4	7
6	Group 5	12
7	Group 6	8
8	Group 7	48

Haplogroup Comparison

- **Data Source:** Whole gene sequences of Syn and non-Syn SNPs.
- **Comparison:** Clusters generated using MEGA were compared with haplogroups.
- **Details:** Given in Table 2.

Table 2: Haplogroup Comparison

S. No	Haplotype	Acc No.	Category	MEGA Cluster	2-AP
1	H1	36	Cat-I	Group 5	0.016
		62	Cat-III	Group 5	0.009
2	H12	11	Cat-I	Group 2	0.004
		17	Cat-I	Group 2	0.008
3	H22	22	Cat-I	Group 4	0.003
		38	Cat-I	Group 4	0.010
		39	Cat-I	Group 4	0.000
		97	Cat-V	Group 4	0.0054
4	H30	30	Cat-I	Group 1B	0.006
		82	Cat-IV	Group 1B	0.006
		84	Cat-IV	Group 1B	0.008
5	H46	49	Cat-II	Group 2	0.011
		83	Cat-IV	Group 2	0.006

Allele Analysis

Category I - ASG-Indigenous

- **Accessions:** 41 alleles.
- **SNP and Indel Analysis:** Detailed in Table 3.

Table 3: SNP and Indel Analysis

S. No	Total No. of SNPs	Syn SNP	Non-Syn SNP	SNPs in Exon	Position of SNPs	Total No. of Indels	Indels	Position of Indels	2-AP
1	45	11.00	32.00	0	45	12	12	0	0
2	7	1.00	6.00	0	7	6	6	0	0.012
3	10	1.50	7.50	0	10	3	2	1 (E1:1)	0.013
4	20	6.00	14.00	4	16 (E3-1, E9-3)	10	9	1 (E8:1)	0.007
5	52	0.00	6.00	1	51 (E3-1)	12	11	1 (E2:1)	0
6	13	1.00	12.00	8	5 (E12-8)	6	3	1 (E1:1, E9:1, E10:1)	0.009
7	6	0.00	6.00	1	5 (E10-1)	4	4	0	0.007
8	7	1.00	6.00	0	7	3	1	2 (E1:1, E9:1)	0.011
9	18	5.00	13.00	2	16 (E3-1, E9-1)	4	2	2 (E1:1, E7:1)	0.006
10	17	3.00	14.00	0	17	3	2	1 (E1:1)	0.015
11	2	0.00	2.00	0	2	4	3	1 (E1:1)	0.004
12	7	1.00	5.00	1	6 (E3-1)	9	8	1 (E11:1)	0.01
13	10	4.00	6.00	2	8 (E10-2)	7	7	0	0.005
14	7	1.00	4.00	1	6 (E10-1)	9	8	1 (E12:1)	0.01
15	3	0.00	3.00	2	1 (E2-2)	6	2	4 (E1:1, E2:2, E12:1)	0.008
16	21	7.00	13.00	3	18 (E3-1, E9-2)	10	7	3 (E1:1, E7:1, E9:1)	0.01
17	8	1.50	6.50	0	8	3	2	1 (E1:1)	0.009
18	2	0.00	2.00	1	1 (E9-1)	3	2	1 (E1:1)	0.009
19	21	6.00	15.00	4	17 (E3-1, E9-3)	6	5	1 (E1:1)	0.01
20	19	6.00	12.00	1	18 (E9-1)	4	2	2 (E3:1, E8:1)	0.005
21	6	0.00	6.00	2	4 (E2-2)	2	1	1 (E1:1)	0.006
22	6	0.00	6.00	0	6	1	1	0	0.01
23	12	3.00	9.00	0	12	2	1	1 (E1:1)	0.014
24	5	0.00	5.00	0	5	3	2	1 (E1:1)	0.013
25	6	0.00	6.00	0	6	6	6	0	0.011
26	10	3.00	7.00	0	10	2	1	1 (E1:1)	0.009
27	7	0.00	6.00	1	6 (E10-1)	1	0	1 (E10:1)	0.008
28	16	2.50	13.50	0	16	1	1	0	0.005
29	6	0.00	6.00	0	6	2	1	1 (E1:1)	0.006
30	14	0.00	10.00	2	12 (E3-1, E9-1)	1	1	0	0.01
31	10	0.00	8.00	1	9 (E3-1)	2	1	1 (E1:1)	0.014
32	12	4.00	8.00	2	10 (E3-1, E9-1)	3	2	1 (E1:1)	0.012
33	4	0.00	4.00	0	4	0	0	0	0.005
34	2	0.00	2.00	0	2	0	0	0	0.01
35	4	0.00	4.00	1	3 (E10-1)	1	0	1 (E10:1)	0.006
36	7	1.00	5.00	2	5 (E10-2)	2	1	1 (E10:1)	0.005
37	4	0.00	4.00	2	2 (E10-2)	2	1	1 (E10:1)	0.005
38	5	0.00	5.00	1	4 (E10-1)	2	1	1 (E10:1)	0.005
39	5	0.00	5.00	2	3 (E10-2)	2	1	1 (E10:1)	0.005
40	3	0.00	3.00	1	2 (E10-1)	0	0	0	0.005
41	5	0.00	5.00	1	4 (E10-1)	1	0	1 (E10:1)	0.005

Category II - ASG-Exotic

- **Accessions:** 3 alleles.
- **SNP and Indel Analysis:** Detailed in Table 4.

Table 4: SNP and Indel Analysis

S. No	Total No. of SNPs	Syn SNP	Non-Syn SNP	SNPs in Exon	Position of SNPs	Total No. of Indels	Indels	Position of Indels	2-AP
1	6	2.00	4.00	0	6	5	5	0	0.009
2	18	3.00	15.00	1	17 (E3-1)	7	6	1 (E1:1)	0.012
3	1	0.00	1.00	0	1	1	1	0	0.005

Category III - Basmati

- **Accessions:** 2 alleles.
- **SNP and Indel Analysis:** Detailed in Table 5.

Table 5: SNP and Indel Analysis

S. No	Total No. of SNPs	Syn SNP	Non-Syn SNP	SNPs in Exon	Position of SNPs	Total No. of Indels	Indels	Position of Indels	2-AP
1	2	1.00	1.00	0	2	2	1	1 (E10:1)	0.011
2	9	2.00	7.00	0	9	0	0	0	0.009

Category IV - Wild

- **Accessions:** 18 alleles.
- **SNP and Indel Analysis:** Detailed in Table 6.

Table 6: SNP and Indel Analysis

S. No	Total No. of SNPs	Syn SNP	Non-Syn SNP	SNPs in Exon	Position of SNPs	Total No. of Indels	Indels	Position of Indels	2-AP
1	3	1.00	2.00	0	3	1	1	0	0.01
2	8	2.00	6.00	0	8	4	3	1 (E1:1)	0.01
3	13	2.50	10.50	3	10 (E10-3)	2	1	1 (E10:1)	0.008
4	2	0.00	2.00	0	2	0	0	0	0.009
5	4	0.00	4.00	1	3 (E10-1)	1	0	1 (E10:1)	0.011

In the above tables, the values for total number of SNPs, Synonymous SNPs (Syn SNP), Non-Synonymous SNPs (Non-Syn SNP), SNPs in Exon, Position of SNPs, Total Number of Indels, Indels, Position of Indels, and 2-AP are indicated for each respective category. Further analyses and interpretations are provided in the subsequent sections.

Discussion

- The results indicated that the majority of the genetic variations were SNPs, with a smaller proportion being Indels. The identification of synonymous and non-synonymous SNPs helps understand the potential impact on protein function.
- The analysis highlighted the significant genetic diversity present in the Aromatic and non-Basmati categories. This diversity could be harnessed for breeding programs aimed at improving specific traits.
- The 2-AP levels were correlated with the presence of specific SNPs and Indels, suggesting potential markers for selecting high 2-AP content lines.

Conclusion

This study provides a comprehensive overview of the genetic variation within the different rice categories, highlighting the potential for using SNP and Indel markers in breeding programs. The high genetic diversity observed, particularly in the Aromatic and non-Basmati categories, presents opportunities for the development of new rice varieties with improved traits, including aroma and yield. Future work should focus on validating these markers in larger populations and different environments to ensure their robustness and utility in breeding programs.

References

1. Fitzgerald TL, Waters DLE, Henry RJ. The effect of salt on betaine aldehyde dehydrogenase transcript levels and 2-acetyl-1-pyrroline concentration in fragrant and non-fragrant rice (*Oryza sativa*). Plant Science. 2008;175(4):539-546. doi:10.1016/j.plantsci.2008.06.005.
2. Ishitani M, Arakawa K, Mizuno K, Kishitani S, Takabe T. Betaine aldehyde dehydrogenase in the Gramineae. Levels in leaves of both betaine accumulating and non-accumulating cereal plants. Plant and Cell Physiology. 1993;34(3):493-495.
3. Ishitani M, Nakamura T, Han SY, Takabe T. Expression of the betaine aldehyde dehydrogenase gene in barley in response to osmotic stress and abscisic acid. Plant Molecular Biology. 1995;27(2):307-315.
4. Liang Z, Ma D, Tang L, Hong Y, Luo A, Zhou J, *et al*. Expression of the spinach betaine aldehyde dehydrogenase (BADH) gene in transgenic tobacco plants. Chinese Journal of Biotechnology. 1997;13:153-159.
5. Librado P, Rozas J. DnaSP v5: A software for comprehensive analysis of DNA polymorphism data. Bioinformatics. 2009;25(11):1451-1452.
6. Mohanty A, Kathuria H, Ferjani A, Sakamoto A, Mohanty P, Murata N, *et al*. Transgenics of an elite indica rice variety Pusa Basmati 1 harbouring the codA gene are highly tolerant to salt stress. Theoretical and Applied Genetics. 2002;106(1):51-57.
7. Niu XL, Zheng WJ, Lu BR, Ren GJ, Huang WZ, Wang SH, *et al.* An unusual post-transcriptional processing in two betaine aldehyde dehydrogenase loci of cereal crops directed by short, direct repeats in response to stress conditions. Plant Physiology. 2007;143(4):1929-1942.
8. Kumar PS, Ragichedu PK, Maddala S, Rao DS, Lella VSR, Konne K, *et al*. Insight of aroma in brown rice through chemical assessment of 2-Acetyl-1-pyrroline (2AP) in aromatic germplasm of India. Cereal Chemistry. 2018. doi:10.1002/cche.10081.
9. Rathinasabapathi B, Gage DA, Mackill DJ, Hanson AD. Cultivated and wild rices do not accumulate glycine betaine due to deficiencies in two biosynthetic steps. Crop Science. 1993;33(3):534-538.
10. Shirasawa K, Takabe T, Kishitani S. Accumulation of glycinebetaine in rice plants that overexpress choline monooxygenase from spinach and evaluation of their tolerance to abiotic stress. Annals of Botany. 2006;98(3):565-571.
11. Tamura K, Dudley J, Nei M, Kumar S. MEGA4: molecular evolutionary genetics analysis (MEGA) software version 4.0. Molecular Biology and Evolution. 2007;24(8):1596-1599.
12. Weretilnyk EA, Hanson AD. Molecular cloning of a plant betaine-aldehyde dehydrogenase, an enzyme implicated in adaptation to salinity and drought. Proceedings of the National Academy of Sciences of the United States of America. 1990;87(7):2745-2749.
13. Wood AJ, Saneoka H, Rhodes D, Joly RJ, Goldsbrough PB. Betaine aldehyde dehydrogenase in sorghum. Plant Physiology. 1996;110(4):1301-1308.