

# Allelic Variation and Haplotype Diversity of the *badh2* Gene in Thai Aromatic Landrace Rice Varieties

Sudarat Sinchai, Hathairat Urairong and Kittaporn Rumjuankiat\*

College of Agricultural Innovation and Food Technology, Rangsit University, Pathum Thani 12000, Thailand

(\*Corresponding author's e-mail: [kittaporn.r@rsu.ac.th](mailto:kittaporn.r@rsu.ac.th))

Received: 3 August 2025, Revised: 11 September 2025, Accepted: 1 October 2025, Published: 10 December 2025

## Abstract

Thai aromatic landrace rice varieties are vital reservoirs of genetic diversity for fragrance traits, offering valuable potential for rice breeding. This study characterized grain morphology, aroma, and the *badh2* gene variation in aromatic and non-aromatic rice varieties. Most aromatic varieties featured slender grains, typically with white pericarps. Sensory evaluation revealed varying aroma intensities across the panel. *Badh2* gene sequences were analyzed for 14 allelic variations across introns and exons. Six exonic variations defined 4 haplotypes. Haplotype 1, including KDML105 and 17 others, shared an 8-bp deletion in exon 7. Haplotype 2, found only in Hawm Hua Bawm, had the same alleles as Haplotype 1 but a unique SNP (G > A) in exon 13, a novel allele associated with aroma expression. Haplotype 3 consisted of 8 varieties, lacking the 8-bp deletion in exon 7 and carrying the new allele in exon 13. Haplotype 4 was found in Hawm Gra Dang Ngah 59. All the alleles resembled non-aromatic rice, suggesting that aroma may result from functional mutations elsewhere in the *badh2* gene or be influenced by different genes. Predicted functional impacts of exonic variants support the role of *badh2* gene diversity in shaping aroma. These findings enhance our understanding of aroma genetics in rice and provide useful markers for breeding fragrant cultivars.

**Keywords:** Aromatic rice, Genetic diversity, *badh2* gene, Haplotype, Landrace rice, KDML105, Allelic variation

## Introduction

The growing demand for fragrant rice now comprises 15% - 18% of the global rice trade, [1], with most consumers willing to pay premium prices, due to its superior quality, tender texture, unique flavor, and distinctive aroma when cooked, Thai Hom Mali Rice is one of the most popular and valuable aromatic rice varieties in the domestic and international markets, produced from the paddy of fragrant non-glutinous rice varieties, which are sensitive to photoperiod and cultivated as main crops in Thailand. The Department of Agriculture, under the Ministry of Agriculture and Cooperatives, has certified these varieties as Khao Dawk Mali 105 (KDML105) and RD15. Their grains are naturally aromatic, depending on age, and retain a soft texture when cooked [2].

The unique aroma of fragrant rice varieties emanates from a complex mixture of over 300 volatile compounds, with 2-acetyl-1-pyrroline (2-AP) the

principal volatile substance [3,4]. Rice aroma can be detected by grain-chewing and the 1.7% KOH methods [5-8], with gas chromatography mass spectrometry selected ion monitoring (GC-MS-SIM) used for quantitative determination [6,9-11], and specific PCR amplification [12] for the classification of aromatic and non-aromatic rice varieties.

The aroma in rice grains is governed by a single recessive *fgr* gene on the eighth chromosome of the rice genome [13], encoded by the betaine aldehyde dehydrogenase 2 (*badh2*) gene. The complete *badh2* gene has 6,154 bp on the long arm of chromosome 8, encodes 503 amino acids, and comprises 15 exons and 14 introns [12,14].

The biosynthesis and accumulation of 2-AP are linked to the BADH2 enzyme, which converts gamma-amino butyraldehyde (AB-ald) into gamma-aminobutyric acid (GABA), resulting in a non-fragrant rice variety. The mutation of the *badh2* gene causes the

BADH2 enzyme to become non-functional; it cannot convert AB-ald into GABA, leading to higher levels of 2-AP production and, consequently, fragrant rice [15,16]. The 2-AP is present in all parts of the aromatic rice plant except for the roots and is found in very low concentrations in non-aromatic rice varieties [11]. Mutation types in the *badh2* gene include single-nucleotide polymorphisms (SNPs), deletions, and insertions (InDel) that disrupt normal enzyme function. Nonsense mutations also introduce premature stop codons, resulting in truncated protein synthesis. Deletions or insertions alter the reading frame and cause frameshift mutations, which usually result in non-functional proteins.

Previous research on the nucleotide mutation of the *badh2* gene across various aromatic rice cultivars reported 21 possible allele types including a 3-base pair (bp) deletion in the 5' untranslated region, an 8-bp insertion in the promoter site (-1,314) [17], a miniature interspersed transposable element (MITE) insertion in the promoter region [18], a 2-bp deletion in exon 1, a 7-bp deletion in exon 2 [14], an 8-bp deletion with 3 SNPs in exon 7 [12], a 7-bp insertion in exon 8 [19], an 806-bp deletion between exons 4 and 5 [20], a 1-bp (T) deletion, SNP (G/T), and 1-bp (T) insertion in exon 10 [14], a 3-bp deletion in exon 12 [21], a 1-bp (T) deletion, 3-bp (TAT) insertion and SNP (C/T) in exon 13 [14], and a 1-bp (G) insertion (premature stop codon), SNP G/T and SNP C/T in exon 14 [18]. Among the functional alleles, an 8-bp deletion in exon 7 induces a premature stop codon, resulting in the ultimate loss of BADH2 enzyme function. This functional mutation has been recorded in most aromatic rice varieties such as Basmati from India and Pakistan and KDML105 from Thailand [22]. Different aromatic varieties carry unique mutations in the *badh2* gene that contribute to functional or nonsynonymous allelic mutations [14]. Phitaktansakul *et al.* [1] conducted whole-genome resequencing and analyzed the *badh2* gene variation in 475 rice accessions from the Korean World Rice Collection. They identified 33 polymorphic sites in the *badh2* gene coding region including 26 SNPs and InDels, 5 alleles in the 5'UTR, and 2 alleles in the 3'UTR, with 38 haplotypes identified based on 26 SNPs. Chan-in *et al.* [23] investigated the allelic variation of the *badh2* gene, which covers introns 4 to 8 in 22 Thai fragrant rice

samples. They identified 4 haplotypes. Among these, only Haplotype 1, which contains an 8-bp deletion along with 3 SNPs, was associated with strong aromatic varieties including Niaw Dam Luem Pra, Kaow Hawm, Phayaa Luem Dang, Buer Ner Moo4, Buer Ner Moo-CMU, KDML105, and PTT1. The other 3 haplotypes did not contain an 8-bp deletion.

Single amino acid substitutions (SAASs) are typically caused by SNPs in the gene coding regions [24-26]. Experimental methods can accurately evaluate the impact of SAASs on protein function; however, these approaches are time-consuming, resource-intensive, and difficult to manipulate [26]. Studies to develop methods for predicting the effects of SAASs on protein function in plants are limited because of the lack of plant SAAS data from molecular experiments [27]. The Plant Protein Variation Effect Detector (PPVED) has demonstrated high predictive accuracy for plant SAASs. PPVED facilitates the identification and characterization of genetic variants that account for the observed phenotypic differences in plants by addressing the challenges in functional genomics and systems biology. Owing to diverse *badh2* alleles being associated with different aroma profiles, they are valuable targets for marker-assisted selection (MAS) in global aromatic rice breeding and conservation. Thai landrace aromatic rice possesses valuable genetic diversity, rich in genetic resources for discovering new aromatic flavor types for developing new aromatic rice in breeding programs. However, there are limited aromatic genes (*badh2*) in the landrace varieties that have been characterized. In this study, we attempted to analyze the shape and pericarp color of aromatic rice grains, their sensory test for aroma, and evaluate the haplotype diversity of aromatic landrace rice samples by examining allelic variation in the *badh2* gene. The impacts of variants detected on the function of the *badh2* gene were also predicted.

## Materials and methods

### Plant materials

Seed samples of 25 fragrant rice varieties were collected from morphological characterization experimental plots of 10 traditional plus 2 improved varieties (KDML105 and RD15) from Sakon Nakhon Rice Research Center in the northeast, 11 landraces

from Phatthalung Rice Research Center in the south, and 2 landrace varieties from farmers' fields in the north. Four improved varieties including aromatic (Pathum Thani 1, RD43) and non-aromatic (RD79, RD87), were purchased from private seed enterprises (PSEs). The 29 varieties used in this study are listed in **Table 1**.

### Grain morphology and sensory test

The morphological characteristics of brown rice grains including length, width, and thickness were measured using a Vernier caliper (Draper Tools, Ltd., Chandler's Ford, United Kingdom) with a precision of  $\pm 0.05$  mm. Grain shape was recorded as the length-to-width ratio and classified into 4 categories according to the International Rice Research Institute (IRRI, 2002) as slender (L/W ratio  $> 3.0$ ), medium (L/W ratio between 2.1 and 3.0), bold (L/W ratio between 1.1 and 2.0), and round (L/W ratio  $< 1.0$ ). Pericarp color as white, red, brown, and purple was also determined [15].

An aromatic sensory test was performed using a modified method based on Amarawathi *et al.* [19]. Ten milled rice grains were placed in a 15 mL test tube containing 8 mL of 1.7% KOH solution, and the tube was incubated at 30 °C in a water bath with the lid closed for 10 min. After incubation, the aroma was assessed by a panel of 6 trained individuals. Panelists were trained with smell reference samples and practice until they could anchor their perception. Present samples in a random order to avoid bias, and a short break was taken between samples to rinse the noses. The average grade was scored as strong aroma (odor is very strong, easily noticeable; score 3), moderate aroma (odor is not strong but noticeable; score 2), slight aroma (odor is faint, requires concentration to detect; score 1), and no aroma (no detectable odor; score 0).

### Sequencing of the *badh2* gene

The collected seed samples were germinated for 15 days. Genomic DNA was extracted from the seedlings using the CTAB method described by Doyle and Doyle [28]. The complete coding nucleotide sequence of the *badh2* gene was downloaded from GenBank (Nanjing11, Acc. no EU770319.1) and used to design the primers. The full length of the *badh2*

gene sequence, including start to stop codons, is roughly 6,000 bp. Due to the limitations of the sequencing technique, 6 primer pairs were designed to amplify overlapping DNA fragments, with the initial sequence on the same strand as follows:

AroB1 F: 5'-ATCGCTTTCCACCTCAACGC-3'

R: 5'-AGGGATCCATCCATCAACAGG-3'

AroB2 F: 5'-GTGCTTCGTTAGTTGGCAGG-3'

R: 5'-ACGGAGAGAGTAGCTGCTAGG-3'

AroB3 F: 5'-TTTTTCTCTCATCCTGCGCT-3'

R: 5'-TCCACAGAAATTTGGAAACAAACC-3'

AroB4 F: 5'-CCTCCTGTAATCATGTATACCCCA-3'

R: 5'-TCCACTCAACAGCTGTACAAGA-3'

AroB5 F: 5'-CGTTTTGTCTGGTTATGGACTCT-3'

R: 5'-GGTAGCACCTTGCTTTTGG-3'

AroB6 F: 5'-GTTTTGACCCCATGGCACTT-3'

R: 5'-ATGGGCGTGCATGCGTA-3'

Fifty microliters of PCR reactions contained 5  $\mu$ L of 10X PCR buffer, 2.5  $\mu$ L of 2.5 mM MgCl<sub>2</sub>, 1  $\mu$ L of 0.2 mM dNTPs, 0.5  $\mu$ L of each primer (10 ng/ $\mu$ L), 0.5  $\mu$ L of 5U Taq DNA polymerase (Thermo Scientific, Massachusetts, USA), 35  $\mu$ L of ddH<sub>2</sub>O, and 5  $\mu$ L of 50 ng/ $\mu$ L genomic DNA. The amplification profile consisted of 2 min at 94 °C, followed by 45 s at 94 °C, 45 s at 60 °C, and 1 min at 72 °C for 25 cycles, with a final extension at 72 °C for 5 min. The amplified products were visualized using 1.5% agarose gel electrophoresis. The purified PCR products with confirmed quality and quantity were subjected to nucleotide sequencing using an automated sequencer operated by Solutions for Genetic Technologies (SolGent, Daejeon, South Korea). The primers used for sequencing were the same as those used in the PCR reactions, and sequencing was performed in both directions using forward and reverse primers.

### Allelic variation and haplotype analysis of the *badh2* gene

BioEdit software version 7.4.2 was used to combine 6 DNA sequence fragments, each originating from a different rice variety, into contiguous sequences of the *badh2* gene. For multiple alignment, the full-length contig of the *badh2* gene of each variety was imported into MEGA7 software to compare all sequences simultaneously. Polymorphic or variant

alleles in the alignment sequence were analyzed in the form of insertions/deletions (InDels), single-nucleotide polymorphisms (SNPs), and variable numbers of nucleotide repeats. The impacts of polymorphic alleles were used to characterize the haplotype diversity of the aromatic rice samples.

### Prediction of protein function

The *badh2* gene nucleotide sequence of each variety was translated into an amino acid sequence using MEGA XI software (<https://www.megasoftware.net/Hawme>) to identify the SAASs in the coding region caused by single-nucleotide variants. The amino acid positions that exhibited changes and impacted normal protein function in plants, leading to physiological or morphological changes, were further analyzed to predict their potential effects on protein function using the Plant Protein Variation Effect Detector (PPVED) program, accessible at <http://www.ppved.org.cn>. If a non-synonymous substitution, as a change in nucleotide sequence (mutation), resulted in a different amino acid protein sequence and impacted the gene function, then the PPVED program predicted this as functional. Conversely, a synonymous substitution that did not alter the gene function of the *badh2* gene, a predicted class, was evaluated as neutral.

## Results and discussion

### Grain morphology and sensory test

The grain shape, pericarp color, and aroma score of the 29 Thai rice varieties are listed in **Table 1**. The

brown rice samples were analyzed and classified as detailed below.

Slender grains consisted of 13 traditional aromatic varieties (Khao Dawk Mali 105, Hawm Khi Kwai, Hawm Phu Khiew, Hawm Tha Wee, Pathum Thep, Daw Hawm, Dawk Pha Yom, Hawm Luang Ram, Hawm Bang Kaew, Hawm Na Khao, Baow Hawm, Hawm Jan, and Hawm Hua Bawn) with 5 improved rice varieties (Pathum Thani 1, RD15, RD43, RD79, and RD87). Only RD15, a photosensitive, improved variety derived from mutated KDML105 with gamma radiation of 15 Krad, is mentioned. Its grain shape, cooking quality, softness, and aroma are very similar to KDML105, except that it matures 15 days earlier.

Medium grains included 11 Thai varieties (Khi Tom Hawm, Hawm Nang Nuan, Hawm Thung, Hawm Pamah, Khao Hawm, Buer Soo, Buer Ner Moo, Hawm Tai, Hawm Gra Dang Ngah 59, Khao Hawm Tai, and Luang Hawm). No bold or round-grain rice varieties were collected in this study.

Most of the rice samples had a white-colored pericarp, with the others showing purple (Hawm Phu Khiew, Khi Tom Hawm, and Buer Soo), red (Hawm Luang Ram, Hawm Gra Dang Ngah 59), and brown (Buer Ner Moo). Pigmented rice pericarps have become increasingly popular due to their health benefits, with high levels of phenolics, flavonoids, anthocyanins, high antioxidant activity, anti-inflammatory potential, antimicrobial properties, and potential as a functional food [29].

**Table 1** Grain morphological characteristics and aromatic scores of 24 traditional fragrant, 3 improved fragrant, and 2 improved non-fragrant rice varieties.

Rice variety	Source	Grain size of brown rice (mm)			GL/GW ratio	Grain shape	Pericarp color	Aroma (0 - 3)*
		Length	Width	Thickness				
1. Hawm Khi Kwai	NE	7.69	2.15	1.83	3.58	Slender	Red	2
2. Hawm Phu Keiw	NE	7.14	2.18	1.72	3.28	Slender	Purple	2
3. Pathum Thep	NE	7.77	2.15	1.75	3.61	Slender	White	2
4. Khao Hawm	NE	7.29	2.80	2.09	2.60	Medium	White	3
5. Hawm Tha Wee	NE	7.69	2.17	1.82	3.54	Slender	White	2
6. Hawm Nahng Nuan	NE	7.14	2.41	1.86	2.96	Medium	White	2
7. Hawm Pamah	NE	7.42	2.55	1.93	2.91	Medium	White	2
8. Daw Hawm	NE	7.48	2.46	1.88	3.04	Slender	White	2

Rice variety	Source	Grain size of brown rice (mm)			GL/GW ratio	Grain shape	Pericarp color	Aroma (0 - 3)*
		Length	Width	Thickness				
9. Hawm Tung	NE	7.37	2.50	1.99	2.95	Medium	White	1
10. Khi Tom Hawm	NE	7.50	2.61	2.07	2.87	Medium	Purple	1
11. Khao Dawk Mali 105	NE	7.56	2.18	1.79	3.47	Slender	White	3
12. RD15	NE	7.60	2.16	1.79	3.52	Slender	White	2
13. Hawm Tai	S	5.96	2.25	1.72	2.65	Medium	White	1
14. Hawm Jan	S	8.02	2.20	1.84	3.65	Slender	White	2
15. Lueang Hawm	S	7.91	2.75	2.02	2.88	Medium	White	2
16. Dawk Pha Yom	S	7.30	2.20	1.80	3.32	Slender	White	2
17. Hawm Luang Ram	S	7.70	2.08	1.83	3.70	Slender	Red	2
18. Hawm Hua Bawn	S	8.15	2.08	1.72	3.92	Slender	Red	3
19. Hawm Bang Keaw	S	7.37	2.11	1.74	3.49	Slender	White	1
20. Hawm Na Khao	S	8.09	2.11	1.82	3.83	Slender	White	1
21. Khao Hawm Tai	S	7.60	2.60	1.80	2.92	Medium	White	2
22. Baow Hawm	S	6.97	2.17	1.69	3.21	Slender	White	1
23. Hawm Gra Dang Ngah 59	S	6.21	2.27	1.70	2.74	Medium	Red	2
24. Buer Soo	N	6.41	2.79	1.76	2.30	Medium	Purple	1
25. Buer Ner Moo	N	7.40	2.98	2.09	2.48	Medium	Brown	2
26. Pathum Thani 1	PSE	7.46	2.11	1.72	3.54	Slender	White	2
27. RD43	PSE	7.84	2.15	1.89	3.65	Slender	White	2
28. RD79	PSE	7.48	2.17	1.80	3.45	Slender	White	0
29. RD87	PSE	7.11	2.25	1.76	3.16	Slender	White	0

Note: GL = Grain Length; GW = Grain Width; GT = Grain Thickness; PSE = Private Seed Enterprises; N = North; S = South; NE = Northeast; \* = Aroma score

Rice grain aroma sensory testing revealed that 7 rice varieties possessed a slight aroma (Khi Tom Hawm, Hawm Thung, Buer Soo, Hawm Bang Kaew, Hawm Tai, Hawm Na Khao, and Baow Hawm), with 17 varieties having a moderate aroma (Pathum Thani 1, RD15, Hawm Khi Kwai, Hawm Phu Khiew, Hawm Tha Wee, Hawm Nang Nuan, Pathum Thep, Hawm Pamah, Daw Hawm, Bue Ner Moo, Dawk Pha Yom, Khao Hawm Tai, Hawm Luang Ram, Hawm Kra Dang Ngah 59, Hawm Jan, RD43, and Luang Hawm), 3 varieties having a strong aroma ( Khao Dawk Mali 105, Hawm Hua Bawn, and Khao Hawm) and 2 varieties with no aroma.

The aroma of rice grains collected from different locations varies due to diverse genetic backgrounds, environmental conditions such as soil fertility, cultural

practices, and post-harvest management [30]. Photosensitive aromatic landrace rice varieties were used in this study. These cultivars should only be planted during the wet season, and each variety has its own flowering and harvesting times depending on the day length. Freshly harvested rice has a stronger aroma than rice stored for a long period because the aromatic compound fragrances diminish over time [8].

#### Nucleotide sequence and allelic variation analysis of the *badh2* gene

The 6 nucleotide sequence fragments of each rice variety, derived from the designed primers, were assembled into a single contig of the *badh2* gene from the start codon at position 1 to the stop codon at position 5886. The *badh2* gene of the 29 Thai rice



Genetic group	Rice variety	Nucleotide position of the <i>badh2</i> gene														
		Intron 2		Intron 4			Exon 7		Intron 8			Exon 10	Intron 10		Exon 13	
		829	2123	2398	2876	2878	2880 - 2887	2888	3217	3378	3364	4369	4421	4470	5236	
	Lueang Hawm	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
4*	Hawm Tha Wee	.	(AT)6	.	.	.	.	.	.	.	.	.	.	.	.	.
	Hawm Nahng Nuan	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
	Hawm Pamah	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
	Dawk Pha Yom	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
	Hawm Luang Ram	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
	RD43	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
5*	Hawm Hua Bawn	.	(AT)12	-	.	.	.	.	.	.	.	.	.	.	.	A
6*	Hawm Bang Keaw	.	.	-	A	A	GATTATGG	C	.	.	.	A	G	T	.	.
	Hawm Na Khao	.	.	-	.	.	.	.	.	.	.	.	.	.	.	.
7*	Khi Tom Hawm	.	(AT)6	-	.	.	.	.	.	.	.	.	.	.	.	.
	Hawm Tung	.	.	-	.	.	.	.	.	.	.	.	.	.	.	.
	Daw Hawm	.	.	-	.	.	.	.	.	.	.	.	.	.	.	.
	Khao Hawm Tai	.	.	-	.	.	.	.	.	.	.	.	.	.	.	.
	Hawm Tai	.	.	-	.	.	.	.	.	.	.	.	.	.	.	.
	Baow Hawm	.	.	-	.	.	.	.	.	.	.	.	.	.	.	.
8*	Hawm Gra Dang Ngah 59	TT del	.	-	.	.	.	.	.	.	T	G	C	G	G	G
9**	RD79	TT	.	-	.	.	.	.	.	.	.	.	.	.	.	.
	RD87	.	.	-	.	.	.	.	.	.	.	.	.	.	.	.

Note: \* = Aromatic rice; \*\* = non-aromatic rice

### Haplotype diversity of aromatic rice

Based on the nucleotide sequence variation within the exon regions of the *badh2* gene, this study identified 6 variable sites that allowed the classification of 27 Thai aromatic and 2 non-aromatic rice varieties into 4 haplotypes. Haplotypes 1 and 2 shared the same nucleotide substitutions and an 8-bp deletion, except in exon 13 of Haplotype 2, where nucleotide G was replaced with A. Haplotypes 3 and 4 did not have the 8-bp deletion, as detailed in **Table 3**.

**Haplotype 1** was characterized by specific genotypic variations including exon 7, an 8-bp deletion at positions 2880 - 2887 (E7Del8), SNPs at positions 2876 (E7/2876-T), 2878 (E7/2878-T), and 2888

(E7/2888-T), exon 10 at position 4369 (E10/4369-G), and exon 13 at position 5236 (E13/5236-G). This haplotype was the most common, and found in 17 aromatic rice varieties (62.96% of the studied aromatic rice samples) including KDML105, Pathum Thani 1, RD15, RD43, Buer Soo, Leung Hawm, Hawm Khi Kwai, Hawm Phu Khiew, Hawm Tha Wee, Hawm Nang Nuan, Pathum Thap, Hawm Pamah, Khao Hawm, Buer Ner Moo, Dawk Pha Yom, Hawm Luang Ram, and Hawm Jan.

**Haplotype 2** shared the most variable alleles with Haplotype 1 but differed at exon 13, where the SNP at position 5236 showed allele A instead of G (E13/5236-A). This haplotype was found in only 1 variety, Hawm

Hua Bawn from Krabi Province, which exhibits a unique aroma reminiscent of taro when cooked [33]. The most popular variety, KDML105 in Haplotype 1, has an aroma similar to popcorn or pandan leaves (*Pandanus amaryllifolius*) [34].

**Haplotype 3** showed variation at exon 7, with no 8-bp deletion and SNPs at positions 2876 (E7/2876-A), 2878 (E7/2878-A), 2888 (E7/2888-C), exon 10 (E10/4369-A), and exon 13 (E13/5236-A). Haplotype 3 was found in 8 varieties: Khi Tom Hawm, Hawm Thung, Daw Hawm, Hawm Tai, Hawm Bang Kaew, Khao Hawm Tai, Hawm Na Khao, and Baow Hawm. Similar findings were reported by Sakthivel *et al.* [16]; Chan-in *et al.* [23]. Many aromatic rice cultivars do not carry the functional mutation in exon 7, such as traditional non-Basmati aromatic cultivars, Jow Pluak Dam (a breeding line from the Department of Agriculture, Thailand), and 3 landraces from Loei Province: Pla Sew, Sew Kliang Hawm, and Hawm Sa-ngium.

**Haplotype 4** comprised a single fragrant rice called Hawm Gra Dang Ngah 59, an aromatic landrace from Narathiwat Province, notable for its ylang-ylang fragrance [35], together with the non-aromatic varieties RD79 and RD85. Key features included A alleles at exon 7 positions 2876-A and 2878-A, with no 8-bp deletion, and alleles E10/4369-G and E13/5236-G. All alleles in this haplotype were similar to non-aromatic rice, suggesting that the aroma in Hawm Gra Dang Ngah 59 may result from mutations in the promoter region, 5' or 3' UTR of the *badh2* gene, or be influenced by other genes. This classification aligned with Fitzgerald *et al.* [31], who studied 313 aromatic landrace rice varieties from 17 countries. Most (279 varieties) had aromatic alleles at exon 7, except for 15 varieties that did not, and 19 were mixtures, possibly affected by mutations outside exon 7.

An additional functional mutation was identified in aromatic rice varieties that did not possess the 8-bp deletion in exon 7. The *badh2* gene was sequenced in 26 aromatic rice varieties from diverse geographical regions, revealing eight nonsynonymous polymorphisms. Four of these were frameshift mutations: A 7-bp deletion in exon 2 (found in Chinese cultivar Hsings Keng-Nuo), a 2-bp deletion in exon 1 (in cultivars from Madagascar), a T-insertion in exon 10 (in cultivar Pare Bane Pulut from Indonesia), a G > T SNP in exon 10 (of Indian rice cultivar

Vashunparag), and a T-deletion in exon 10 (in a cultivar from Malaysia). A premature stop codon mutation (G-insertion in exon 14) was also found in 6 Indian cultivars, which all probably produced truncated BADH2 proteins [14]. A separate 7-bp deletion in exon 2 was identified as a frameshift mutation in 12 Chinese aromatic varieties, which also lacked MITE in the promoter region [32]. Several nonsynonymous alleles in the *badh2* gene have been used to design a functional marker to distinguish between scented and non-scented rice [17,32,36].

### Prediction of protein function

Single amino acid substitutions (SAASs) are usually caused by SNPs or other variants in the gene coding region [24-26]. Some SAASs impact normal plant protein functions, leading to distinct physiological or morphological changes [25,37,38]. Nucleotide variations resulting from SNPs or InDels alter codons and change amino acids and gene functions.

In **Table 3**, an SNP at position 2876 (exon 7), where A was substituted by T, resulted in a codon change from AAA to ATA, causing an amino acid substitution from Lysine (K) to Isoleucine (I); this is a missense mutation. Protein function prediction software (PPVED) indicated this change as functional. Similarly, at position 2878 (exon 7), the SNP A to T caused a codon change from AAG to TAG, resulting in a substitution from Lysine (K) to a stop codon or a nonsense mutation. This premature stop codon was also predicted as a functional change. At position 2888 (exon 7), the SNP C to T changed the codon from GCT to GTT, replacing Alanine (A) with Valine (V); however, the PPVED indicated this as a neutral or synonymous substitution, a nucleotide mutation that altered the amino acid sequence of a protein but did not change the gene function. For position 4369 (exon 10), an SNP from G to A altered the codon from GGG to AGG, changing Glycine (G) to Arginine (R), and this was also predicted to be neutral. At position 5236 (exon 13), an SNP from G to A caused a codon shift from GGT to GAT, leading to a change from Glycine (G) to Aspartic acid (D). This missense alteration was predicted as functional. Position 5236, located in exon 13, was identified as a novel variant in aromatic rice. PPVED software has a predictive accuracy of 87%

higher than that of comparable software. PPVED facilitated the identification and characterization of genetic variants that explained the observed phenotype

variations in plants, contributing solutions for challenges in functional genomics and systems biology [39].

**Table 3** Haplotype patterns of Thai aromatic rice landrace varieties classified based on 6 allelic variations in exons of the *badh2* gene.

Haplotype	Coding region/Nucleotide position of the <i>badh2</i> gene						Number of varieties
	Exon 7			Exon 10	Exon 13		
	2876	2878	2880 - 2887	2888	4369	5236	
1	T	T	-----	T	G	G	17
2	T	T	-----	T	G	A	1
3	A	A	GATTATGG	C	A	A	8
4	A	A	GATTATGG	C	G	G	1
Reference*	A	A	GATTATGG	C	G	G	
Codon	AAA>ATA	AAG>TAG		GCT>GTT	GGG>AGG	GGT>GAT	
Amino acid**	K>I	K>STOP		A>V	G>R	G>D	
Predicted class	Functional	Functional		Neutral	Neutral	Functional	

Note: \* = Reference (Nanjing11\*, acc. NO. EU770319.1); \*\* = A: Alanine, K: Lysine, I: Isoleucine, V: Valine, G: Glycine, R: Arginine, D: Aspartic acid

Nowadays, high-yielding, improved rice varieties are widely accepted and have become popular among Thai farmers. The cultural practice of monoculture involves growing improved high-yielding rice crops over a large area, leading to reduction in the genetic diversity of rice, with genetic erosion of genes or landrace varieties. Genetic erosion is detrimental to rice landrace varieties in Thailand, where over 600 aromatic rice landraces are cultivated across diverse environments from deep-water to upland areas. The Thai jasmine rice type is widely cultivated throughout the country under various names including Hom Mali and Khao Dawk Mali 105 (KDML105) [40]. Some rice landraces are conserved in the Gene Bank under the Rice Department and the Department of Agriculture, but their aromatic genes under the landrace genetic background have not been well evaluated. This study characterized the variations of the complete coding sequences within the *badh2* gene, which codes for a protein, including the start and stop codons in aromatic landrace varieties. However, this research did not explore the regulatory elements, promoters, and 5' and 3' untranslated regions (UTRs) that can also affect gene transcription and translation. These aspects should be addressed in future studies. Classifying haplotype diversity of the *badh2* gene in landrace varieties will lead to the exploration of superior,

favorable, and rare alleles, such as different alleles of aroma genes and alleles of abiotic stress genes that contribute to the necessary traits in rice breeding and conservation.

## Conclusions

This study provides new insights into the phenotypic and genetic diversity of Thai aromatic rice, revealing a wide range of grain morphologies, pericarp colors, and aromatic profiles, by identifying 14 polymorphic alleles in the *badh2* gene, including a novel SNP in exon 13 with potential functional effect on the aroma gene. The classification of varieties into 4 major haplotypes enhances the genetic framework for distinguishing aromatic landraces. Importantly, the validation of the well-known 8-bp deletion and discovery of new variants offer valuable molecular markers for use in marker-assisted selection. These findings have direct implications for rice breeding programs aiming to improve or preserve aroma traits, while also supporting efforts in the conservation of traditional landraces. Future research should investigate the biochemical effects of newly identified variants and assess their impact on aroma expression to pave the way for the development of high-quality aromatic rice cultivars to challenge consumer preferences.

## Acknowledgements

The authors express sincere appreciation to the College of Agricultural Innovation and Food Technology, Rangsit University, Thailand for facilities support, which significantly contributed to the successful completion of this research. We would also like to thank our fellow students for their invaluable assistance, encouragement, and support throughout this study.

## Declaration of generative AI in scientific writing

The authors declare that Generative AI was used to enhance the clarity, grammar, and language flow of the manuscript. Scientific data analysis and interpretation of the results were not generated by AI tools. Following the use of AI, the authors have carefully reviewed and edited the text and accept full responsibility for the final version of the published article.

## Credit author statement

**Kittaporn Rumjuankiat:** Conceptualization; Methodology; Validation. **Hathairat Urairong:** Supervision; Data curation; Writing - Original draft preparation. **Sudarat Sinchai:** Visualization; Investigation.

## References

- [1] R Phitaktansakul, KW Kim, KM Aung, TZ Maung, MH Min and A Somsri. Multi-omics analysis reveals the genetic basis of rice fragrance mediated by betaine aldehyde dehydrogenase 2. *Journal of Advanced Research* 2022; **42**, 303-314.
- [2] National Bureau of Agricultural Commodity and Food Standards. Thai agricultural standard: Thai Hom Mali rice (TAS 4000-2003). Ministry of Agriculture and Cooperatives, Bangkok, Thailand, 2003.
- [3] K Mahattanatawee and RL Rouseff. Comparison of aroma active and sulfur volatiles in 3 fragrant rice cultivars using GC-Olfactometry and GC-PFPD. *Food Chemistry* 2014; **154**, 1-6.
- [4] S Mathure, K Wakte, N Jawali and A Nadaf. *Breeding Science* 2014; **64(1)**, 9-17.
- [5] LMT Bradbury, RJ Henry, Q Jin, RF Reinke and DLE Waters. A perfect marker for fragrance genotyping in rice. *Molecular Breeding* 2005; **16(4)**, 279-283.
- [6] F Liu, W Wang, F Sun, W Liu and D Liang. Detection of 2-acetyl-1-pyrroline in fragrant rice by KOH immersion method. *Journal of Northeast Agricultural University* 2014; **21(4)**, 25-30.
- [7] S Sood and EA Siddiq. A rapid technique for scent determination in rice. *Indian Journal of Genetics and Plant Breeding* 1978; **38(2)**, 268-271.
- [8] CH Yeh, YS Lee, HL Chen and TP Huang. Post-harvest changes in aroma compounds of aromatic rice during storage. *Journal of Food Composition and Analysis* 2024; **125**, 105652.
- [9] NL Hien. Aroma characterization in rice using gas chromatography - mass spectrometry. *Plant Breeding and Biotechnology* 2006; **47**, 351-357.
- [10] R Widjaja, JD Craske and M Wootton. Comparative studies on volatile components of non-fragrant and fragrant rices. *Journal of the Science of Food and Agriculture* 1996; **70(2)**, 151-161.
- [11] T Yoshihashi, NTT Huong and TTC Hoa. Method for quantification of 2-acetyl-1-pyrroline, a potent flavor compound of aromatic rice, using stable isotope dilution and gas chromatography/mass spectrometry. *Journal of Agricultural and Food Chemistry* 2004; **52(20)**, 6047-6051.
- [12] LMT Bradbury, TL Fitzgerald, RJ Henry, Q Jin and DLE Waters. The gene for fragrance in rice. *Plant Biotechnology Journal* 2005; **3(3)**, 363-370.
- [13] SN Ahn, SK Bajaj, SR McCouch, H Dulay, JS Lob Fernando and FG Gamas. Inheritance of aroma in rice: Identification of a gene for fragrance. *Journal of Genetics and Breeding* 1992; **46(2)**, 123-128.
- [14] MJ Kovach, MN Calingacion, MA Fitzgerald and SR McCouch. The origin and evolution of fragrance in rice (*Oryza sativa* L.). *Proceedings of the National Academy of Sciences* 2009; **106(34)**, 14444-14449.
- [15] S Chen, Y Yang, W Shi, Q Ji, F He, Z Zhang, Z Cheng, X Liu and M Xu. *Badh2*, encoding betaine aldehyde dehydrogenase, inhibits the

- biosynthesis of 2-acetyl-1-pyrroline, a major component in rice fragrance. *Plant Cell* 2008; **20(7)**, 1850-1861.
- [16] K Sakthivel, RM Sundaram, N Shobha Rani, SM Balachandran and CN Neeraja. Genetic and molecular basis of fragrance in rice. *Biotechnology Advances* 2009; **27(4)**, 468-473.
- [17] H Shi, DS Tan, WB Tang and QQ Liu. Development of a functional marker for fragrance gene in rice. *Rice Science* 2013; **20(1)**, 21-28.
- [18] F Bourgis, R Guyot, H Gherbi, E Tailliez, I Amabile, J Salse, M Lorieux and A Ghesquière. Characterization of the major fragrance gene from an aromatic japonica rice and analysis of its diversity in Asian cultivated rice. *Theoretical and Applied Genetics* 2008; **117(3)**, 353-368.
- [19] M Amarawathi, R Singh, AK Singh, VP Singh, T Mohapatra, TR Sharma and NK Singh. Mapping of quantitative trait loci for basmati quality traits in rice (*Oryza sativa* L.). *Indian Journal of Genetics and Plant Breeding* 2008; **68(2)**, 120-126.
- [20] G Shao, S Tang, M Luo, M Li, P Hu and W Zhai. A novel deletion mutation of the *BADH2* gene in fragrant rice. *Plant Molecular Biology Reporter* 2011; **29**, 471-476.
- [21] X He, Z Lin, D Zhang and Q Zhang. Identification of 3 bp deletion in Exon 12 of *badh2*. *Molecular Breeding* 2015; **35**, 161.
- [22] R Roy, VK Singh, AK Shukla, P Sood and AK Singh. Allelic diversity of *BADH2* gene in aromatic rice cultivars including KDML105 and Basmati. *Journal of Plant Biochemistry and Biotechnology* 2020; **29(4)**, 742-750.
- [23] P Chan-in, S Jamjod, N Yimyam, B Rerkasem and T Pusadee. Grain quality and allelic variation of the *Badh2* gene in Thai fragrant rice landraces. *Agronomy* 2020; **10(6)**, 779.
- [24] MA Care, CJ Needham and AJ Bulpitt. Deleterious SNP prediction: Methods and resources. *Briefings in Functional Genomics* 2007; **6(5)**, 367-380.
- [25] Z Wang, T Mou, Y Tan and J Zhang. Development of functional markers for *BADH2* gene and fragrance analysis in rice. *Breeding Science* 2016; **66(1)**, 90-96.
- [26] PC Ng and S Henikoff. Predicting the effects of amino acid substitutions on protein function. *Annual Review of Genomics and Human Genetics* 2006; **7**, 61-80.
- [27] AV Kovalev, J Li, W Zhou and J Guo. PPVED: Prediction of the effect of single amino acid substitution on protein function in plants. *Frontiers in Plant Science* 2018; **9**, 1403.
- [28] JJ Doyle and JL Doyle. Isolation of plant DNA from fresh tissue. *Focus* 1990; **12(1)**, 13-15.
- [29] W Punfa, S Wongpornchai and S Siriamornpun. Pigmented rice pericarp: A potential source of phenolic compounds and functional properties. *The Journal of Cereal Science* 2024; **113**, 105009.
- [30] O Kongpun, S Wanchana, T Toojinda and A Vanavichit. Environmental and genetic influence on rice grain aroma variation in Thailand. *Rice Science* 2024; **31(2)**, 157-167.
- [31] MA Fitzgerald, NR Sackville Hamilton, MN Calingacion, HA Verhoeven and VM Butardo. Is there a second fragrance gene in rice? *Plant Biotechnology Journal* 2008; **6(4)**, 416-423.
- [32] W Shi, X Zhao, J Rong and BR Lu. Genetic diversity of fragrant rice (*Oryza sativa* L.) germplasm investigated by microsatellite markers. *Genetic Resources and Crop Evolution* 2008; **55(4)**, 497-506.
- [33] A Songkairat. 2019, Genetic diversity and aromatic characteristics of Thai local rice varieties. Master's Thesis. Kasetsart University, Bangkok Thailand.
- [34] S Laksanalamai and S Ilangantileke. *Pandanus amaryllifolius*: A source of aroma compound in rice. *International Rice Research Notes* 1993; **18(1)**, 15-16.
- [35] S Chaithong, P Sukkarn, C Aenglong, W Woonnoi, W Klaypradit, W Suttithumsatid, N Chinfak, J Seatan, S Tanasawet and W Sukketsiri. Biological activities and phytochemical profile of Hawm Gra Dang Ngah rice: Water and ethanolic extracts. *Foods* 2025; **14(7)**, 1119.
- [36] KD Wakte, PR Zanan, VC Hinge, SB Khandagale, R Nadaf and RA Henry. Genetic diversity analysis and identification of aroma gene in Indian aromatic rice accessions using *BADH2* gene specific markers. *Physiology and*

- Molecular Biology of Plants* 2016; **22(4)**, 551-563.
- [37] Y Li, Y Huang, J Bergelson, M Nordborg and JO Borevitz. Association mapping of local climate-sensitive quantitative trait loci in *Arabidopsis thaliana*. *Proceedings of the National Academy of Sciences* 2012; **107(49)**, 21199-21204.
- [38] Y Xu, R Wang, Y Tong, H Zhao, Q Xie and Y Liu. Mapping of QTLs for grain shape and size using a high-density SNP map in rice. *PLoS One* 2018; **13(5)**, e0196962.
- [39] X Gou, X Feng, H Shi, T Guo, R Xie, Y Liu, Q Wang, H Li, B Yang, L Chen and Y Lu. PPVED: A machine learning tool for predicting the effect of single amino acid substitution on protein function in plants. *Plant Biotechnology Journal* 2022; **20**, 2128-2143.
- [40] A Vanavichit. Exploration and utilization of rice landrace genes for rice improvement in Thailand. *Rice Science* 2018; **25(3)**, 119-124.