RESEARCH ARTICLE

Whole genome sequencing of *Oryza sativa* L. cv. Seeragasamba identifies a new fragrance allele in rice

Ganigara Bindusree, Purushothaman Natarajan, Sukesh Kalva, Parani Madasamy*

Genomics Laboratory, Department of Genetic Engineering, SRM University, Kattankulathur, India

* parani.m@ktr.srmuniv.ac.in

Abstract

Fragrance of rice is an important trait that confers a large economic benefit to the farmers who cultivate aromatic rice varieties. Several aromatic rice varieties have limited geographic distribution, and are endowed with variety-specific unique fragrances. *BADH2* was identified as a fragrance gene in 2005, and it is essential to identify the fragrance alleles from diverse geographical locations and genetic backgrounds. Seeragasamba is a short-grain aromatic rice variety of the indica type, which is cultivated in a limited area in India. Whole genome sequencing of this variety identified a new *badh2* allele (*badh2-p*) with an 8 bp insertion in the promoter region of the *BADH2* gene. When the whole genome sequences of 76 aromatic varieties in the 3000 rice genome project were analyzed, the *badh2-p* allele was present in 13 varieties (approximately 17%) of both indica and japonica types. In addition, the *badh2-p* allele was present in 17 varieties that already had the loss-of-function allele, *badh2-E7*. Taken together, the frequency of *badh2-p* allele (approximately 40%) was found to be greater than that of the *badh2-E7* allele (approximately 34%) among the aromatic rice varieties. Therefore, it is suggested to include *badh2-p* as a predominant allele when screening for fragrance alleles in aromatic rice varieties.

OPEN ACCESS

Citation: Bindusree G, Natarajan P, Kalva S, Madasamy P (2017) Whole genome sequencing of *Oryza sativa* L. cv. Seeragasamba identifies a new fragrance allele in rice. PLoS ONE 12(11): e0188920. <https://doi.org/10.1371/journal.pone.0188920>

Editor: Swarup Kumar Parida, National Institute for Plant Genome Research, INDIA

Received: July 24, 2017

Accepted: November 15, 2017

Published: November 30, 2017

Copyright: © 2017 Bindusree et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: Data from this study was submitted to NCBI under BioProject ID PRJNA324355 and BioSample ID SAMN05200854. Raw reads of this project were deposited in compressed FASTAQ format at SRA database of NCBI with the accession number SRP076132 (<http://www.ncbi.nlm.nih.gov/sra/SRP076132>).

Funding: This study was financially supported by SRM-DBT Partnership Platform for Contemporary Research Services and Skill Development in Advanced Life Sciences Technologies (Order No.

Introduction

Rice is a major staple food feeding hundreds of millions of people in Asia [1]. The unique fragrance of aromatic varieties makes rice appealing to people in other parts of the world who do not consume it as a staple food. Even in traditional rice-eating countries, highly-priced rice dishes are prepared using aromatic rice varieties. Therefore, the fragrance of aromatic rice varieties is an economically important trait that fetches premium price in domestic as well as international markets. In fact, many rice-growing countries earn huge foreign exchange by exporting aromatic rice varieties to developed countries [2]. Several long-grain and short-grain aromatic rice varieties with considerable variations in fragrance are available in the market catering to consumer preferences. Some of the aromatic rice varieties are traditional cultivars grown in very limited areas, and are known for their unique fragrance [3, 4, 5].

Beyond its importance as the world's premier food crop, rice is also an excellent model plant for crop genomics [6]. Earlier works on rice genomics analyzed genome-wide genetic

BT/PR12987/INF/22/205/2015), SRM University, Chennai, Tamil Nadu, India.

Competing interests: The authors have declared that no competing interests exist.

variations to understand gene functions related to agronomic traits. Genetic studies have shown that the fragrance of aromatic rice is controlled by the *fgr* recessive gene [7, 8], which was later identified to be a gene coding for betaine aldehyde dehydrogenase 2, *BADH2* [9]. The first loss-of-function allele in *BADH2* (*badh2.1* or *badh2-E7*) was identified as an 8-bp deletion in the seventh exon [9]. Subsequently, eighteen *badh2* alleles associated with the rice fragrance were reported [10, 11]. While a majority was loss-of-function mutations due to InDels or SNPs in coding regions, non-synonymous mutations in coding regions, mutations in intron-exon junction, promoter, and 5'UTR were also reported. Such variations may account for the spectrum of unique fragrances observed in aromatic rice varieties. Therefore, it is important to characterize the *badh2* alleles from diverse aromatic rice varieties to generate a panel of fragrance alleles for breeders to choose the desired one.

Seeragasamba (also called Jeeraga Samba) is a short-grain aromatic rice variety that is cultivated in select regions of Tamil Nadu, India [12]. Sakthivel et al. [13] screened the *badh2-E7* allele in 47 aromatic rice varieties from India including Seeragasamba. Though *badh2-E7* allele was present 42 varieties, it is not clear if Seeragasamba had this allele or not. Another study reported that Seeragasamba did not have the *badh2-E2*, *badh2-E7* or *badh2-E8* allele for which screening was undertaken [14]. Therefore, we analyzed the entire *BADH2* gene and its promoter from the whole genome sequence of Seeragasamba; a new *badh2* allele was identified. Evidence for this new allele in other aromatic rice varieties was obtained based on the whole genome sequences of 76 aromatic rice varieties from the 3000 rice genome project.

Materials and methods

Whole-genome sequencing of *Oryza sativa* L. cv. Seeragasamba

Seeds of Seeragasamba were obtained from a farmer's field in Tamil Nadu, India. Seedlings were grown in pots under greenhouse conditions. Genomic DNA was isolated from young leaves using the cetyl trimethyl ammonium bromide (CTAB) method [15]. Quantity and quality of the genomic DNA were assessed using spectrophotometer, fluorimeter, and agarose gel electrophoresis. Preparation of paired-end library and sequencing were carried according to the manufacturer's protocol (Illumina Inc., USA). Paired-end reads were extracted in FASTQ format for further downstream analysis.

Genome mapping and variant calling in Seeragasamba

Quality of the raw paired-end reads was assessed using the FastQC tool [16]. Adapter sequences were trimmed off using the cutadapt tool [17]. The paired-end reads were further filtered by retaining the bases with a minimum Phred quality score of 30 using sickle master (<https://github.com/najoshi/sickle>). The quality filtered reads were mapped to the latest unified build release Os-Nipponbare-Reference-IRGSP-1.0 [18] using Burrows-Wheeler alignment (BWA) software [19]. The aligned reads in the SAM file were sorted using the SortSam function of the Picard tool v1.118 (<https://broadinstitute.github.io/picard/>). The sorted SAM file was converted to BAM file using SAMtools v0.1.19 [20] for variant calling. Variant calling was carried out using mpileup application of SAMtools [20] setting at default parameters. The variants were further filtered based on the following criteria: (1) base quality ≥ 30 , (2) number of reads per base between 5 and 75, (3) variant quality ≥ 90 , (4) mapping quality ≥ 60 , and (5) distance of adjacent variant ≥ 5 . The filtered variants were extracted in Variant Call File (VCF) format and annotated using the SnpEff V3.6 tool [21] and rice7 gene model database (http://sourceforge.net/projects/snpeff/files/databases/v3_6/snpEff_v3_6_rice7.zip). The total numbers of variants were segregated as SNPs and InDels. Variants in the genes, and other genomic regions were annotated as genic and intergenic variants, respectively. According to the location

of the genic SNPs, they were further classified as CDs, UTRs (5'UTRs and 3'UTRs), introns and regulatory sequences. The SNPs found in the coding regions were categorized as non-synonymous (causes change in amino acid), synonymous (causes no change in amino acid), stop loss (removes the existing stop codon), stop gain (introduces a stop codon), start gain (introduces a start codon), and start loss (removes an existing start codon). The SNPs were differentiated as transition (C/T and G/A) and transversion (C/G, T/A, A/C and G/T) SNPs.

Analysis of variants in the aromatic rice varieties of 3000 rice genome project

Whole genome data for 3024 rice varieties were made available from the 3000 rice genome project [22], which included 76 aromatic rice varieties. A searchable Rice SNP-Seek Database containing about 20 million SNPs obtained after aligning these genomes with the Nipponbare reference was created [23] and further updated to include InDels [24]. Using this database, we have analyzed the variants present in the *BADH2* gene (including -1326 bp upstream region) in all the 76 aromatic rice varieties in the database. VCF containing the SNPs from 39 aromatic rice varieties, which contained either *badh2-E7* or *badh2-p* (identified from the current study) or both alleles were extracted from the Rice SNP-Seek Database, and merged with the SNPs from Seeragasamba. The merged VCF was converted into Genomic Data Structure (gds) format using gdsfmt package of R software (<https://www.r-project.org/>). Bi-allelic SNPs were extracted from the gds file. SNPRelate program from R package was used to generate dendrogram using identity-by-state (IBS) distance matrix.

Results and discussion

The major compound responsible for the fragrance of aromatic rice is 2-acetyl-1-pyrroline (2AP), which is present in all aerial parts of the plant [25, 26]. Normally, the product of *BADH2* gene, betaine aldehyde dehydrogenase 2, converts γ -aminobutyraldehyde (GABald) to γ -aminobutyric acid [27]. GABald is diverted to the production and accumulation of 2AP when *BADH2* is partly or fully non-functional. This gives aromatic rice varieties their characteristic fragrance (Fig 1). *BADH2* in chromosome 8 is a homologue of *BADH1* in chromosome 4, and its size is 6154 bp with 15 exons and 14 introns.

We performed whole genome sequencing of the Seeragasamba rice variety via Illumina sequencing by synthesis method to study the mutations in *BADH2* gene and its promoter. Data from this study was submitted to NCBI under BioProject ID PRJNA324355 and BioSample ID SAMN05200854. Raw reads of this project were deposited in compressed FASTAQ format at SRA database of NCBI with the accession number SRP076132 (<http://www.ncbi.nlm.nih.gov/sra/SRP076132>).

Whole genome sequencing of the genomic DNA from Seeragasamba produced 42.6×10^6 raw reads with an average read length of 101 bp. The raw reads were quality filtered, and the resulting 38.6×10^6 high-quality reads totaling 4.2×10^9 bp were used for mapping to Nipponbare reference genome. About 30.8×10^6 high-quality reads (79.8%) were uniquely mapped, which covered 86.5% of the reference genome. Chromosome-wide coverage varied between 82.8 and 91.6% in chromosomes 1 and 12, respectively. The initial variant identification yielded 3,166,688 SNPs and 265,109 InDels. Quality filtering of these variants using the five parameters as described in the materials and methods yielded 671,708 and 60,705 SNPs and InDels, respectively. All quality-filtered variants were annotated; detailed classification of the annotated variants is given in Table 1. Detailed analysis of the annotated variants in the genomic region spanning 20,378,646 bp to 20,385,975 bp in chromosome 8 was carried out to identify *badh2* allele responsible for the fragrance of Seeragasamba.

Fig 1. Biosynthetic pathway of 2-acetyl-1-pyrroline (2AP) in rice. Functional *BADH2* converts γ -aminobutyraldehyde (GABald) to γ -aminobutyric acid (GABA). When *BADH2* is partly or fully non-functional, GABald is diverted to the production of Δ^1 pyrroline and 2AP (responsible for fragrance).

<https://doi.org/10.1371/journal.pone.0188920.g001>

An 8 bp deletion in exon 7 (*badh2.1* or *badh2-E7* allele) resulting in a shift of the reading frame and premature termination of translation is the most predominant loss-of-function mutation in *BADH2* gene [9]. However, InDels in exon 1, 2, 4, 5, 8, 12, 13, and 14 were also reported [10, 11, 28, 29, 30, 31]. Two SNPs resulting in non-sense mutations and premature termination of translation were reported in the exon 10 [10, 13]. A mutation in the splice donor site at exon 1–intron 1 was reported in six Japanese aromatic landraces [32]. Reports on mutations in the non-coding regions of *BADH2* are limited. Nankai 138, a Japanese aromatic rice variety, did not have any mutations in the coding region. It was reported to have an 8 bp insertion in the promoter region upstream of the start codon between -1314 and -1315 position and a 3-bp deletion in the 5'UTR from -81 to -83 positions in *BADH2* (*badh2-p-5'UTR* allele) [33].

Table 1. Annotation of SNPs and InDels in Seeragasamba.

Location/Type of the variant	SNPs	InDels
Intergenic	445,900	8,259
Genic	225,808	52,446
Intron and regulatory sequences	138,005	34,653
UTRs	27,195	11,184
Coding regions (CDs)	60,608	6,609
Synonymous SNPs	25,668	-
Non-Synonymous SNPs	34,940	-
Stop loss	243	-
Stop gain	908	-
Start gain	1055	-
Start loss	95	-
Transition SNPs (Ts)	457,091	-
Transversion SNPs (Tv)	214,617	-
Ts/Tv ratio	2.1298	-

<https://doi.org/10.1371/journal.pone.0188920.t001>

Seeragasamba also contained the same 8-bp insertion in the promoter region without any mutation in the coding region. It did not have the 3-bp deletion in the 5'UTR. This allele was named as *badh2-p*. Whole genome sequences of 76 aromatic varieties from Bangladesh, Bhutan, India, Iran, Japan, Liberia, Madagascar, Myanmar, Nepal, Pakistan, Philippines, Taiwan, and Thailand were available in the 3000 rice genome project [22]. A detailed variant analysis of *BADH2* gene in these aromatic varieties was carried out, and the presence of *badh2* alleles was documented (Table 2). Twenty-six varieties (approximately 34%) had the *badh2-E7* allele,

Table 2. List of *badh2* alleles identified from the whole genome sequences of the aromatic rice varieties in the 3000 rice genome project.

S.No	Name of the variety	Variety Group	Origin	Allele 1	Allele 2
1	P 335::IRGC 69001–1	Aus/boro	Liberia	<i>badh2-E7</i> [#]	NIL
2	BASMATI 385	Intermediate type	Pakistan	<i>badh2-E7</i>	NIL
3	MADHUWA KARIA::IRGC 16138–2	Basmati/sadri	Nepal	<i>badh2-E7</i>	NIL
4	BARA PASHAWARI 390::IRGC 27779–1	Basmati/sadri	Pakistan	<i>badh2-E7</i>	NIL
5	BADSHABHOG 4–60::IRGC 37793–2	Basmati/sadri	Bangladesh	<i>badh2-E7</i>	NIL
6	MILAGROSA, ZAWA BANDAY	Indica	Philippines	<i>badh2-E7</i>	NIL
7	DOMSIAH	Japonica	Iran	<i>badh2-E7</i>	NIL
8	POHHERLIMASION::IRGC 62025–1	Basmati/sadri	Nepal	<i>badh2-E7</i>	NIL
9	TAROM MOLAI	Intermediate type	Iran	<i>badh2-E7</i>	NIL
10	JC 157	Aus/boro	India	<i>badh2-E7</i>	<i>badh2-p</i> *
11	JC 1	Basmati/sadri	India	<i>badh2-E7</i>	<i>badh2-p</i>
12	BASMATI 1::IRGC 27798–1	Basmati/sadri	Pakistan	<i>badh2-E7</i>	<i>badh2-p</i>
13	HIRA NAKHI::IRGC 70840–1	Indica	India	<i>badh2-E7</i>	<i>badh2-p</i>
14	KASHA::IRGC 83865–1	Basmati/sadri	Bangladesh	<i>badh2-E7</i>	<i>badh2-p</i>
15	BASMATI SURKH 161::IRGC 27797–1	Basmati/sadri	Pakistan	<i>badh2-E7</i>	<i>badh2-p</i>
16	BEGUNBICHI 33::IRGC 29260–1	Basmati/sadri	Bangladesh	<i>badh2-E7</i>	<i>badh2-p</i>
17	PANKAIT 31::IRGC 29377–1	Basmati/sadri	Bangladesh	<i>badh2-E7</i>	<i>badh2-p</i>
18	BAJAL::IRGC 45024–2	Basmati/sadri	India	<i>badh2-E7</i>	<i>badh2-p</i>
19	KEYA NUNIA::IRGC 46117–2	Basmati/sadri	India	<i>badh2-E7</i>	<i>badh2-p</i>
20	SADRI RICE 1	Basmati/sadri	Iran	<i>badh2-E7</i>	<i>badh2-p</i>
21	RATO BASMATI::IRGC 59205–1	Indica	Nepal	<i>badh2-E7</i>	<i>badh2-p</i>
22	CR 44–1::IRGC 45397–1	Basmati/sadri	India	<i>badh2-E7</i>	<i>badh2-p</i>
23	IET 14720::IRTP 23179-G1	Tropical japonica	Na	<i>badh2-E7</i>	<i>badh2-p</i>
24	NS 1576::IRGC 68951–1	Indica	Madagascar	<i>badh2-E7</i>	<i>badh2-p</i>
25	KARNAL LOCAL	Basmati/sadri	India	<i>badh2-E7</i>	<i>badh2-p</i>
26	SETO JHINUWA::IRGC 62043–1	Indica	Nepal	<i>badh2-E7</i>	<i>badh2-p</i>
27	CODE NO 31225::IRGC 46865–2	Intermediate type	India	<i>badh2-p</i>	NIL
28	ARC 10497	Basmati/sadri	India	<i>badh2-p</i>	NIL
29	IRAT 118::IRGC 55785–1	Temperate japonica	Madagascar	<i>badh2-p</i>	NIL
30	HALIDA::IRGC 59005–1	Indica	Nepal	<i>badh2-p</i>	NIL
31	TAIPEI 167::IRGC 65371–1	Indica	Taiwan	<i>badh2-p</i>	NIL
32	HANSRAJ::IRGC 70836–1	Basmati/sadri	India	<i>badh2-p</i>	NIL
33	TK RED 35–799::IRGC 38535–1	Basmati/sadri	Bangladesh	<i>badh2-p</i>	NIL
34	GUDURA::IRGC 61955–1	Basmati/sadri	Nepal	<i>badh2-p</i>	NIL
35	ARC 13502::IRGC 41126–2	Basmati/sadri	India	<i>badh2-p</i>	NIL
36	JALADHI 1::IRGC 45858–2	Intermediate type	India	<i>badh2-p</i>	NIL
37	T 3::IRGC 46760–2	Basmati/sadri	India	<i>badh2-p</i>	NIL
38	ARC 14398::IRGC 41544–2	Basmati/sadri	India	<i>badh2-p</i>	NIL
39	ARC 7296::IRGC 20585–1	Basmati/sadri	India	<i>badh2-p</i>	NIL

[#] Bradbury et al. 2005 *badh2-E7* allele

*new allele identified from the current study

<https://doi.org/10.1371/journal.pone.0188920.t002>

Fig 2. The dendrogram showing the relationship between Seeragasamba and 39 aromatic rice varieties based on whole genome SNPs. The SNP data for the 39 aromatic rice varieties were extracted from the Rice SNP-Seek Database. Name of the aromatic rice variety followed by its type in bracket, and its geographical origin are given for each entry.

<https://doi.org/10.1371/journal.pone.0188920.g002>

which was reported to be the most predominant fragrance allele in aromatic rice varieties. No other reported fragrance alleles with mutations in the coding region or non-coding region were found among the 76 aromatic varieties.

The new allele identified here, *badh2-p*, was present in 13 varieties (approximately 17%). Interestingly, 17 varieties with the *badh2-E7* allele also had *badh2-p* allele bringing the total number varieties with *badh2-p* allele to 30 (approximately 40%). This is higher than the frequency of *badh2-E7* allele in aromatic rice varieties. The presence of the *badh2-p* allele is not significant in the varieties with the *badh2-E7* allele, which represents a loss-of-function mutation. Among the varieties that did not have any apparent loss-of-function mutation (InDels and non-sense SNPs in coding regions), the frequency of *badh2-p* allele was approximately 30% (13 out of 50 varieties). The varieties with *badh2-p* allele included both indica and japonica types. Similarity analysis of 40 aromatic rice varieties using genome-wide SNPs, showed grouping of Seeragasamba with six other aromatic rice varieties from India, Nepal and Taiwan, all of which contained the 8 bp insertions in the promoter region (Fig 2 and Table 2). Previous studies might have underestimated the frequency of the *badh2-p* allele in aromatic rice varieties because of targeted screening for *badh2-E7* and other alleles rather than sequencing the whole gene or genome. This study strongly indicates that *badh2-p* should be included in the screening for fragrance alleles in aromatic rice varieties. Indels near the *cis* elements can greatly influence the expression of the cognate genes. Genes with differential expression harbor more indels in the promoter, especially between 500–1500 bp upstream regions, than those that are not differentially expressed [34]. However, experimental promoter analysis is needed to establish effect of *badh2-p* allele on the expression of BADH2 gene.

Acknowledgments

This study was financially supported by SRM-DBT Partnership Platform for Contemporary Research Services and Skill Development in Advanced Life Sciences Technologies (Order No. BT/PR12987/INF/22/205/2015), SRM University, Chennai, Tamil Nadu, India. The HPCC facility (Genome Server) at SRM University was used for the genome analysis.

Author Contributions

Conceptualization: Parani Madasamy.

Data curation: Ganigara Bindusree, Purushothaman Natarajan, Sukesh Kalva, Parani Madasamy.

Formal analysis: Ganigara Bindusree, Purushothaman Natarajan, Sukesh Kalva, Parani Madasamy.

Funding acquisition: Parani Madasamy.

Investigation: Ganigara Bindusree, Purushothaman Natarajan, Parani Madasamy.

Methodology: Ganigara Bindusree, Purushothaman Natarajan, Sukesh Kalva, Parani Madasamy.

Project administration: Parani Madasamy.

Resources: Purushothaman Natarajan, Parani Madasamy.

Software: Purushothaman Natarajan, Sukesh Kalva, Parani Madasamy.

Supervision: Parani Madasamy.

Validation: Ganigara Bindusree, Purushothaman Natarajan, Parani Madasamy.

Visualization: Ganigara Bindusree, Purushothaman Natarajan, Sukesh Kalva, Parani Madasamy.

Writing – original draft: Ganigara Bindusree, Parani Madasamy.

Writing – review & editing: Ganigara Bindusree, Purushothaman Natarajan, Sukesh Kalva, Parani Madasamy.

References

1. Wakte K, Zanan R, Hinge V, Khandagale K, Nadaf A, Henry R. Thirty-three years of 2-acetyl-1-pyrroline, a principal basmati aroma compound in scented rice (*Oryza sativa* L.): a status review. *Journal of the science of food and agriculture*. 2017; 97(2):384–95. Epub 2016/07/05. <https://doi.org/10.1002/jsfa.7875> PMID: 27376959.
2. Giraud G. The world market of fragrant rice, main issues and perspectives. *International Food and Agribusiness Management Review*. 2013; 16(2):1–20.
3. Roy S, Banerjee A, Mawkhlieng B, Misra A, Pattanayak A, Harish G, et al. Genetic diversity and population structure in aromatic and quality rice (*Oryza sativa* L.) landraces from North-Eastern India. *PloS one*. 2015; 10(6):e0129607. <https://doi.org/10.1371/journal.pone.0129607> PMID: 26067999
4. Wettewa W, Kottearahchi N. Sequence analysis of the mutation in the 7th exon of badh2 gene in traditional aromatic rice varieties in Sri Lanka. *Journal of Agricultural Sciences*. 2014; 9(1).
5. Myint KM, Courtois B, Risterucci A-M, Frouin J, Soe K, Thet KM, et al. Specific patterns of genetic diversity among aromatic rice varieties in Myanmar. *Rice*. 2012; 5(1):20. <https://doi.org/10.1186/1939-8433-5-20> PMID: 27234242
6. Tyagi AK, Khurana JP, Khurana P, Raghuvanshi S, Gaur A, Kapur A, et al. Structural and functional analysis of rice genome. *Journal of genetics*. 2004; 83(1):79–99. PMID: 15240912
7. Ahn S, Bollich C, Tanksley S. RFLP tagging of a gene for aroma in rice. *Theoretical and Applied Genetics*. 1992; 84(7–8):825–8. <https://doi.org/10.1007/BF00227391> PMID: 24201481
8. Lorieux M, Petrov M, Huang N, Guiderdoni E, Ghesquière A. Aroma in rice: genetic analysis of a quantitative trait. *Theoretical and Applied Genetics*. 1996; 93(7):1145–51. <https://doi.org/10.1007/BF00230138> PMID: 24162494
9. Bradbury LM, Fitzgerald TL, Henry RJ, Jin Q, Waters DL. The gene for fragrance in rice. *Plant biotechnology journal*. 2005; 3(3):363–70. Epub 2006/11/30. <https://doi.org/10.1111/j.1467-7652.2005.00131.x> PMID: 17129318.
10. Kovach MJ, Calingacion MN, Fitzgerald MA, McCouch SR. The origin and evolution of fragrance in rice (*Oryza sativa* L.). *Proceedings of the National Academy of Sciences of the United States of America*. 2009; 106(34):14444–9. Epub 2009/08/27. <https://doi.org/10.1073/pnas.0904077106> PMID: 19706531.
11. He Q, Park Y-J. Discovery of a novel fragrant allele and development of functional markers for fragrance in rice. *Molecular Breeding*. 2015; 35(11):217.
12. Mohan MM, Balakrishnan A, Renganayaki P. Research Note A high yielding seeragasamba rice culture VG 09006 and its medicinal properties. *Electronic Journal of Plant Breeding*. 2013; 4(2):1148–54.
13. Sakthivel K, Rani NS, Pandey MK, Sivaranjani A, Neeraja C, Balachandran S, et al. Development of a simple functional marker for fragrance in rice and its validation in Indian Basmati and non-Basmati fragrant rice varieties. *Molecular breeding*. 2009; 24(2):185–90.
14. Rai VP, Singh AK, Jaiswal HK, Singh SP, Singh RP, Waza SA. Evaluation of molecular markers linked to fragrance and genetic diversity in Indian aromatic rice. *Turkish Journal of Botany*. 2015; 39(2):209–17.
15. Murray MG, Thompson WF. Rapid isolation of high molecular weight plant DNA. *Nucleic acids research*. 1980; 8(19):4321–5. Epub 1980/10/10. PMID: 7433111.
16. Andrews S. FastQC: A quality control tool for high throughput sequence data. Reference Source. 2010.
17. Martin M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet journal*. 2011; 17(1):pp. 10–2.
18. Kawahara Y, de la Bastide M, Hamilton JP, Kanamori H, McCombie WR, Ouyang S, et al. Improvement of the *Oryza sativa* Nipponbare reference genome using next generation sequence and optical map data. *Rice (New York, NY)*. 2013; 6(1):4. Epub 2013/11/28. <https://doi.org/10.1186/1939-8433-6-4> PMID: 24280374.
19. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics (Oxford, England)*. 2009; 25(14):1754–60. Epub 2009/05/20. <https://doi.org/10.1093/bioinformatics/btp324> PMID: 19451168.

20. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* (Oxford, England). 2009; 25(16):2078–9. Epub 2009/06/10. <https://doi.org/10.1093/bioinformatics/btp352> PMID: 19505943.
21. Cingolani P, Platts A, Wang le L, Coon M, Nguyen T, Wang L, et al. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w1118; iso-2; iso-3. *Fly*. 2012; 6(2):80–92. Epub 2012/06/26. <https://doi.org/10.4161/fly.19695> PMID: 22728672.
22. The 3,000 rice genomes project. *GigaScience*. 2014; 3:7. Epub 2014/05/30. <https://doi.org/10.1186/2047-217X-3-7> PMID: 24872877.
23. Alexandrov N, Tai S, Wang W, Mansueto L, Palis K, Fuentes RR, et al. SNP-Seek database of SNPs derived from 3000 rice genomes. *Nucleic acids research*. 2015; 43(Database issue):D1023–7. Epub 2014/11/29. <https://doi.org/10.1093/nar/gku1039> PMID: 25429973.
24. Mansueto L, Fuentes RR, Borja FN, Detras J, Abriol-Santos JM, Chebotarov D, et al. Rice SNP-seek database update: new SNPs, indels, and queries. *Nucleic acids research*. 2017; 45(D1):D1075–D81. Epub 2016/12/03. <https://doi.org/10.1093/nar/gkw1135> PMID: 27899667.
25. Buttery R, Juliano B, Ling L. Identification of rice aroma compound 2-acetyl-1-pyrroline in pandan leaves. *SOC CHEMICAL INDUSTRY 14 BELGRAVE SQUARE, LONDON, ENGLAND SW1X 8PS*; 1983. p. 478–.
26. Yoshihashi T, Huong NT, Inatomi H. Precursors of 2-acetyl-1-pyrroline, a potent flavor compound of an aromatic rice variety. *Journal of agricultural and food chemistry*. 2002; 50(7):2001–4. Epub 2002/03/21. PMID: 11902947.
27. Bradbury LM, Gillies SA, Brushett DJ, Waters DL, Henry RJ. Inactivation of an aminoaldehyde dehydrogenase is responsible for fragrance in rice. *Plant molecular biology*. 2008; 68(4–5):439–49. Epub 2008/08/16. <https://doi.org/10.1007/s11103-008-9381-x> PMID: 18704694.
28. Amarawathi Y, Singh R, Singh AK, Singh VP, Mohapatra T, Sharma TR, et al. Mapping of quantitative trait loci for basmati quality traits in rice (*Oryza sativa* L.). *Molecular Breeding*. 2008; 21(1):49–65.
29. Shao G, Tang A, Tang S, Luo J, Jiao G, Wu J, et al. A new deletion mutation of fragrant gene and the development of three molecular markers for fragrance in rice. *Plant breeding*. 2011; 130(2):172–6.
30. Shao G, Tang S, Chen M, Wei X, He J, Luo J, et al. Haplotype variation at *Badh2*, the gene determining fragrance in rice. *Genomics*. 2013; 101(2):157–62. Epub 2012/12/12. <https://doi.org/10.1016/j.ygeno.2012.11.010> PMID: 23220350.
31. Shi W, Yang Y, Chen S, Xu M. Discovery of a new fragrance allele and the development of functional markers for the breeding of fragrant rice varieties. *Molecular Breeding*. 2008; 22(2):185–92.
32. Ootsuka K, Takahashi I, Tanaka K, Itani T, Tabuchi H, Yoshihashi T, et al. Genetic polymorphisms in Japanese fragrant landraces and novel fragrant allele domesticated in northern Japan. *Breeding science*. 2014; 64(2):115–24. Epub 2014/07/06. <https://doi.org/10.1270/jsbbs.64.115> PMID: 24987297.
33. Shi Y, Zhao G, Xu X, Li J. Discovery of a new fragrance allele and development of functional markers for identifying diverse fragrant genotypes in rice. *Molecular breeding*. 2014; 33(3):701–8.
34. Zhang HY, He H, Chen LB, Li L, Liang MZ, Wang XF, et al. A genome-wide transcription analysis reveals a close correlation of promoter INDEL polymorphism and heterotic gene expression in rice hybrids. *Molecular plant*. 2008; 1(5):720–31. Epub 2009/10/15. <https://doi.org/10.1093/mp/ssn022> PMID: 19825576.