Theor Appl Genet (2013) 126:2897–2906
DOI 10.1007/s00122-013-2180-z

ORIGINAL PAPER

# Gene discovery and functional marker development for fragrance in sorghum (*Sorghum bicolor* (L.) Moench)

**Chutintorn Yundaeng · Prakit Somta · Sithichoke Tangphatsornruang · Sugunya Wongpornchai · Peerasak Srinives**

Received: 26 March 2013 / Accepted: 12 August 2013 / Published online: 22 August 2013


**Abstract**

***Key message*** **Sequence analysis and genetic mapping revealed that a 1,444 bp deletion causes a premature stop codon in *SbBADH2* of sorghum IS19912. The non-function of *SbBADH2* is responsible for fragrance in sorghum IS19912.**

*Abstract* 2-acetyl-1-pyrroline (2AP) is a potent volatile compound causing fragrance in several plants and foods. Seeds of some varieties of rice, sorghum and soybean possess fragrance. The genes responsible for fragrance in rice and soybean are orthologs that correspond to *betaine aldehyde dehydrogenase 2* (*BADH2*). Genotypes harboring fragrance in rice and soybean contain a premature stop codon in *BADH2* which impairs the synthesis of full length functional BADH2 protein leading to the accumulation of 2AP. In this study, we reported an association between the *BADH2* gene and fragrance in sorghum. An $F_2$ population of 187 plants developed from a cross between KU630 (non-fragrant) and IS19912 (fragrant) was used. Leaves of $F_2$ and $F_3$ progenies were evaluated for fragrance by organoleptic test, while seeds of $F_2$ plants were analyzed for 2AP. The tests consistently showed that the fragrance is controlled by a single recessive gene. Gene expression analysis of *SbBADH1* and *SbBADH2* in leaves of KU630 and IS19912 at various stages revealed that *SbBADH1* and *SbBADH2* were expressed in both accessions. Sequence comparison between KU630 and IS19912 revealed a continuous 1,444 bp deletion encompassing exon 12 to 15 of *SbBADH2* in IS19912 which introduces a frameshift mutation and thus causes a premature stop codon. An indel marker was developed to detect polymorphism in *SbBADH2*. Bulk segregant and QTL analyses confirmed the association between *SbBADH2* and fragrance.

Communicated by E. S. Guiderdoni.

**Electronic supplementary material** The online version of this article (doi:10.1007/s00122-013-2180-z) contains supplementary material, which is available to authorized users.

C. Yundaeng
Program in Plant Breeding, Faculty of Agriculture at Kamphaeng Saen, Kasetsart University, Kamphaeng Saen, Nakhon Pathom 73140, Thailand

P. Somta (✉) · P. Srinives
Department of Agronomy, Faculty of Agriculture at Kamphaeng Saen, Kasetsart University, Kamphaeng Saen, Nakhon Pathom 73140, Thailand
e-mail: pksomta@gmail.com

S. Tangphatsornruang
National Center for Genetic Engineering and Biotechnology, 113 Phaholyothin Road, Khlong Nueng, Khlong Luang, Pathumthani 12120, Thailand

S. Wongpornchai
Center of Excellence for Innovation in Chemistry, Department of Chemistry, Faculty of Science, Chaing Mai University, Chaing Mai 50200, Thailand

## Introduction

Sorghum (*Sorghum bicolor* (L.) Moench) is among the most important cereal crops of the world. It ranks after maize, wheat, rice, and barley with an annual production of about 65.5 million tons from the planted area of 45 million ha (FAO 2010). This crop is widely grown in dry areas of tropical and sub-tropical regions of Asia, Africa, and America because of its relatively high drought tolerance. Sorghum seeds are utilized for both human food and animal feed, while leaves and stalk are used as forage. Several kinds of foods such as porridge, cookies, cakes, snack and

malted beverages can be made from sorghum grain. Sweet sorghum cultivars which have juicy stem with high sugar content (mainly sucrose) are popularly processed for syrup, sugar and alcohol. In addition, sorghum lignocelluloses biomass can be valorized for second generation biofuel which is utilized for energy purpose. Therefore, sorghum has potential to be a source of biofuel.

Fragrance is an important characteristic of food products. It can increase consumer's acceptance and command higher price of foods. Natural pleasant fragrance/aroma has been recognized in some cultivars of three major crops, viz. rice (*Oryza sativa* L.) (Kadam and Patankar 1938), sorghum (Kottur 1919) and soybean (*Glycine max* (L.) Merr.) (Fushimi and Masuda 2001). For sorghum, Kottur (1919) reported that a landrace cultivar "Ambemohor" from India has a flavor-like fragrant rice cultivar "Ambemohor", while Ayyangar (1938) reported that sorghum cultivar "Kinungapembo" from Tanzania (Tanganyiga in the original paper) possessed fragrant seeds. The fragrance in sorghum is the same as fragrance in rice (Prasada and Murty 1979). Fragrant rice cultivars such as "Basmati" and "Khao Dok Mali 105", and fragrant vegetable soybean cultivars such as "Chamame" and "Kaori" are internationally accepted and command higher market price than normal cultivars. Breeding for fragrant cultivars is a major goal in rice and vegetable soybean breeding programs. The fragrance in rice and soybean is largely due to the volatile chemical compound 2-acetyl-1-pyrroline (2AP) (Buttery et al. 1983; Fushimi and Masuda 2001). The 2AP also associates with flavor in a variety of cooked foods such as popcorn, corn tortillas, fried mungbean, taro and bread crust, to name a few (see Adams and De Kimpe 2006 for a review). Fragrance in rice and soybean can be determined by three methods, viz. (1) chewing cooked/boil seed (2) smelling KOH-treat leaf or seed, and (3) quantification of 2AP content in leaf or seed by gas chromatography techniques. Itani et al. (2004) showed a strong correlation between sensory test and 2AP concentration in rice. Juwattanasomran et al. (2011) reported that the correlation coefficient between chewing cooked seed and 2AP analysis in soybean is 0.39 and highly significant.

Fragrance in rice, soybean and sorghum is controlled by a single recessive gene (AVRDC 2003; Berner and Hoff 1986; Ayyangar 1938; Murty et al. 1982; Sood and Siddiq 1978). Molecular genetics studies in rice revealed that the *OsBADH2* gene encoding betaine aldehyde dehydrogenase 2 protein is largely responsible for the fragrance (Bradbury et al. 2005; Wanchana 2005). The fragrant rice possesses non-functional *OsBADH2*. Up to date, nine fragrance alleles resulting from insertion, deletion, and nucleotide substitution of the *OsBADH2* gene have been identified in rice (Kovach et al. 2009). In vegetable soybean, the non-functional *GmBADH2* gene is also responsible for the fragrance (Juwattanasomran et al. 2011). Two fragrance alleles of *GmBADH2* have been reported in soybean (Juwattanasomran et al. 2011, 2012).

Although the genetic control of fragrance in sorghum has long been reported (Ayyangar 1938), the molecular basis of the trait has not yet been identified. Because the gene *BADH2* is responsible for fragrance in rice (Bradbury et al. 2005) and soybean (Juwattanasomran et al. 2011), we hypothesized that this gene is also responsible for the fragrance in sorghum. Information on the causal gene of fragrance in sorghum will be useful for breeding new fragrant cultivar(s). In this study, we reported the gene mapping and genetic basis of fragrance in sorghum accession "IS19912". The objectives of the present study were (1) to determine whether the *SbBADH2* gene is responsible for fragrance, and (2) to develop a functional marker for fragrance in sorghum.

## Materials and methods

### Plant materials

An $F_2$ population of 187 individuals derived from a cross between IS19912 (as female parent) and KU630 (as male parent) was used as the mapping population. IS19912 (Basmati) is a landrace fragrant sorghum from Madhya Pradesh, India (Prasada and Murty 1979), while KU630 is a non-fragrant sorghum from Thailand. The fragrance in IS19912 is the same as the fragrance in rice (Prasada and Murty 1979). The $F_2$ plants and 10 plants for each of the parental accessions were grown in an experimental field at Kasetsart University, Kamphaeng Saen Campus, Nakhon Pathom, Thailand during March to June 2011 for DNA extraction and fragrance evaluation. Subsequently, $F_{2:3}$ population and the parents were sown in a randomized complete block design (RCBD) with two replicates during September to December 2011 for fragrance evaluation. In each replicate, each family had 10 $F_3$ plants.

DNA of the parental and $F_2$ plants was extracted from young leaves as per Lodhi et al. (1994) with minor modification. DNA concentration was estimated in 1.2 % agarose gel electrophoresis by comparing with a known concentration λ DNA, and adjusted to 10 ng/μl.

### Evaluation for fragrance

Two methods, viz. 2AP quantification and organoleptic test were employed for evaluation of the fragrance. For 2AP quantification, seeds of parental and $F_2$ plants were individually harvested and used for 2AP analysis. The 2AP was quantified using automated headspace gas chromatography (HS-GC) with nitrogen-phosphorus detection (NPD).

The HS-GC–NPD instrumentation and conditions as well as sample preparation and calibration procedure followed those described by Sriseadka et al. (2006) with modification in extraction and calibration steps in order to accommodate the sorghum seed samples. One microliter of an exact concentration of 2,6-dimethylpyridine (DMP), used as internal standard, was spiked in each sample of sorghum seed weighed exactly 1.00 g prior to HS-GC–NPD analysis. The detector responses in terms of peak area ratios between 2AP and 2,6-DMP and the amount of standard 2AP were linear and reproducible with a correlation coefficient (*r*) of 0.9997.

For organoleptic test, young leaves of 50-day-old parental and $F_2$ plants were excised and used for the test. The test was performed following the KOH method of Sood and Siddiq (1978) in which 3.0 g of the leaves were cut in pieces of about 0.5 × 0.5 cM, put in a 25 ml tube and immersed in a 15 ml of 1.7 % KOH (w/v) at 40 °C for 10 min. Subsequently, the tube was opened and smelled directly by a panel of five volunteers. The fragrance was scored as 1 (no fragrance), 2 (mildly to moderately fragrance), or 3 (highly fragrance). Apart from the $F_2$ population, the organoleptic test was also conducted in the $F_{2:3}$ population. In each replicate, young leaves from five plants (equal weight from each plant) of each family were harvested and pooled. The leaves were then subjected to the fragrance test using the same procedures as for the $F_2$ population. Although only five plants per family were sampled which may lead to bias selection towards some genotypes (homozygous dominant and heterozygous) and thus evaluation accuracy, two replicates of evaluation can reduce such bias.

### Correlation and Chi square ($\chi^2$) analyses

Correlation coefficient of 2AP concentration and fragrance score in the $F_2$ population was calculated using the *R*-Program 2.10.0 (R Development Core Team 2010).

Based on the organoleptic tests for the fragrance, $F_2$ plants and $F_{2:3}$ families with the score of 2 or 3 were classified as fragrant, and those with the score of 1 were classified as no fragrant. Chi square ($\chi^2$) analysis was carried out using the *R*-Program 2.10.0 to test for the goodness-of-fit of a 3:1 or 1:3 (no fragrance: fragrance) ratio for the fragrance segregation in $F_2$ or $F_{2:3}$ populations, respectively.

### Expression analysis of the *SbBADH1* and *SbBADH2* genes

Mutation(s) in coding sequence of the *BADH2* gene causes fragrance in rice (Bradbury et al. 2005; Wanchana 2005) and soybean (Juwattanasomran et al. 2011, 2012). Thus, the expression of *BADH1* and *BADH2* genes was initially considered responsible for fragrance in sorghum. KU630 and IS19912 were grown in plots in a greenhouse under natural condition during May to June 2012. At 15, 30, and 45 days after planting (DAP), 0.5 g of young fresh leaves from three plants of each accession was separately harvested for extraction of total RNA using Concert™ plant RNA reagent (Invitrogen, USA). Each RNA sample was from a single different plant. The RNA was then treated with DNA-free™ DNaseI (Ambion, USA) and converted to cDNA using RevertAid H Minus First Strand cDNA Synthesis Kit (Fermentas, USA) according to manufacturer's instruction. cDNA concentration was quantified using ND-1000 Spectrophotometer (NanoDrop Technologies, Inc., DE, USA).

Two cDNA sequences for *BADH* genes of sorghum (Wood et al. 1996) were downloaded from GenBank (accessions AB333793 and AB333794). The sequences were subjected to BLAST search and alignment against the whole genome sequence of sorghum (Paterson et al. 2009) at http://www.phytozome.net/sorghum (JGI Sbi1.4). Annotated gene sequences identified as *SbBADH1* (Sb06g019200) and *SbBADH2* (Sb07g020650) were used in designing primers for real-time PCR analysis. Primers were designed for *SbBADH1*, *SbBADH2,* and for the reference *ACTIN* (Sb09g000730) using Primer3 (Rozen and Skaletsky 2000) (Table S1). Real-time PCR analysis was performed in a CFX96™ real-time PCR detection system (Bio-Rad, USA). Three analyses (three biological replicates) were conducted for each growing stage. Reaction mixtures contained water, 1 × Master mix of FastStart SYBR Green Master (Rox) (Roche Molecular Biochemicals), 300 nM of forward primer, 300 nM of reverse primer, and 50 ng cDNA. Thermocycle conditions included initial denaturation at 95 °C for 3 min, followed by 60 cycles at 95 °C for 10 s, 57 °C for 15 s and 72 °C for 30 s. After 60 cycles, a melting curve was generated by slowly increasing (0.5 °C per 1 s) the temperature from 65 to 95 °C, while the fluorescence was measured. Fluorescent data were acquired during each extension phase. Expression of *SbBADH1* and *SbBADH2* was quantified with *ACTIN* normalization using $2\text{-}\Delta\Delta C_T$ method (Livak and Schmittgen 2001).

### Sequencing of the *SbBADH2* gene and development of allele-specific marker

A genomic region of 7.304 kb (Sb07: 53503330.0.53510634; http://www.phytozome.net/sorghum) covering the *SbBADH2* gene was amplified using 10 primer pairs (Table S2). A PCR was performed in a total volume of 25 µl containing water, 20 ng genomic DNA, 1 × of *Taq* buffer, 2 mM of $MgCl_2$, 0.2 mM dNTPs, 0.1 U of *Taq* DNA polymerase (Fermentas, USA), and 0.5 µM each of forward and reward primers. Amplification was performed in a GeneAmp PCR 9700 System thermocycler (Applied Biosystems, USA) programmed as follows: 94 °C for 3 min, 35 cycles of 94 °C for 30 s, 55 °C for 30 s, 72 °C for 2 min, and 72 °C

for 5 min. After amplification, PCR products were electrophoresed on agarose gel and purified using E-gel® CloneWell Agarose Gels (Invitrogen, USA). The purified PCR products were sequenced by Macrogen Co. Ltd. (South Korea) using BigDye® Terminator v3.1 Cycle Sequencing Kit on an ABI 3730xl DNA Analyzer (Applied Biosystems, USA). The individual sequences were assembled using software Sequencher 5.0 (Gene Code Corporation, USA). The *SbBADH2* sequences from IS19912, KU630 and the reference sequence (Sb07g020650; http://www.phytozome.net/sorghum) were aligned using ClustalW (http://www.ebi.ac.uk/clustalw) to identify polymorphic sites between the two accessions. Predicted protein sequences encoded from these genes were also aligned using the same software.

Upon identification of a large indel polymorphism between the two accessions, a PCR marker specific to the mutant allele was developed using Primer3 (Rozen and Skaletsky 2000). The marker was tested in the parents by PCR amplification. The amplification was the same as that described for SSR analysis (see below). The PCR products were run on 1 % agarose gel electrophoresis, stained with ethidium bromide and visualized under ultraviolet light. The markers were then analyzed in the $F_1$ plant and $F_2$ individual plants.

### Bulk segregant analysis and QTL mapping

Since the fragrance in sorghum is controlled by a single gene (Ayyangar 1938; Murty et al. 1982; see also the Results in this study). Bulk segregant analysis (BSA) (Michelmore et al. 1991) and QTL analysis were used to locate the gene controlling fragrance. Two DNA bulks, fragrance bulk (FB) and non-fragrance bulk (NFB), were formed. FB and NFB were each derived from 10 $F_2$ plants showing high 2AP and fragrance score of 3, and none or low 2AP and fragrance score of 1, respectively.

Two hundred and ninety-eight simple sequence repeat (SSR) markers for sorghum developed from the whole genome sequence of sorghum (Paterson et al. 2009) reported by Yonemaru et al. (2009) were screened for polymorphism between IS19912 and KU630. These markers were chosen to evenly cover the ten chromosomes of the sorghum genome. In addition, eight SSR markers were developed to detect polymorphism within the *SbBADH2* gene or nearby genomic regions (Table S3). PCR amplification, electrophoresis and DNA band visualization were performed as described by Somta et al. (2008) with an exception that a PCR contained 20 ng of genomic DNA. Polymorphic markers were used to detect polymorphism between FB and NFB. The marker(s) showing polymorphism between FB and NFB were used to analyze the $F_2$ population to locate the QTL controlling the fragrance.

A linkage map was constructed using MAPMAKER (Lander et al. 1987) with the minimum LOD score of 3 and the maximum recombination frequency ($r$) of 4.0. Map distance was calculated using Kosambi (1944) mapping function. Location of the QTL for the fragrance was determined using composite interval mapping (CIM) implemented in QTL cartographer 2.5 (Wang et al. 2010). Significant LOD threshold for the QTL at $P = 0.01$ was determined by a 10,000 runs of permutation test.

### DNA sequence submission

Sequences of *SbBADH2* of KU630 and IS19912 were submitted to GenBank with the accession numbers KC687089 and KC6897090, respectively.

## Results

### Fragrance variation in the parents, $F_2$ and $F_{2:3}$ populations

The fragrance evaluation by HS-GC analysis of 2AP in seeds showed that the peak area ratio of 2AP/2,6-DMP in the $F_2$ population varied from 0 to 1.2338 with the mean of 0.1101. IS19912 was highly fragrant with 2AP concentration of 0.3019, while KU630 was non-fragrant with 2AP concentration of zero. Fragrance scores evaluated by organoleptic test of leaves in the $F_2$ population ranged from 1 to 3 with the mean of 1.74. The scores for IS19912 and KU630 were three (highly fragrant) and one (non-fragrant), respectively. Results of organoleptic test in the $F_{2:3}$ population correspond to those in the $F_2$ population. The fragrance scores varied between 1 and 3 with the mean of 1.89. IS19912 had a score of 3, while KU630 had a score of 1. Frequency distribution of 2AP concentration in the $F_2$ population and fragrance score of the $F_2$ and $F_{2:3}$ populations are shown in Fig. 1. 2AP concentration and fragrance score in the $F_2$ population showed a skewed distribution toward KU630 (Fig. 1a ,b). Fragrance score in the $F_{2:3}$ population showed more or less bi-modal distribution (Fig. 1b).

Correlation coefficient ($r$) between fragrance score and 2AP concentration in the $F_2$ population was 0.52. The correlation was significant at $P < 0.0001$ at 185 degree of freedom.

### Inheritance of the fragrance

Inheritance of the fragrance was determined in the $F_2$ and $F_{2:3}$ populations based on organoleptic tests. In the $F_2$ population, 50 plants were fragrant and 137 plants were non-fragrant, whereas in the $F_{2:3}$ population in which five bulked plants represented each family, 144 families were fragrant and 43 families were non-fragrant. In the $F_2$ population, the segregation of non-fragrant:fragrant fitted well

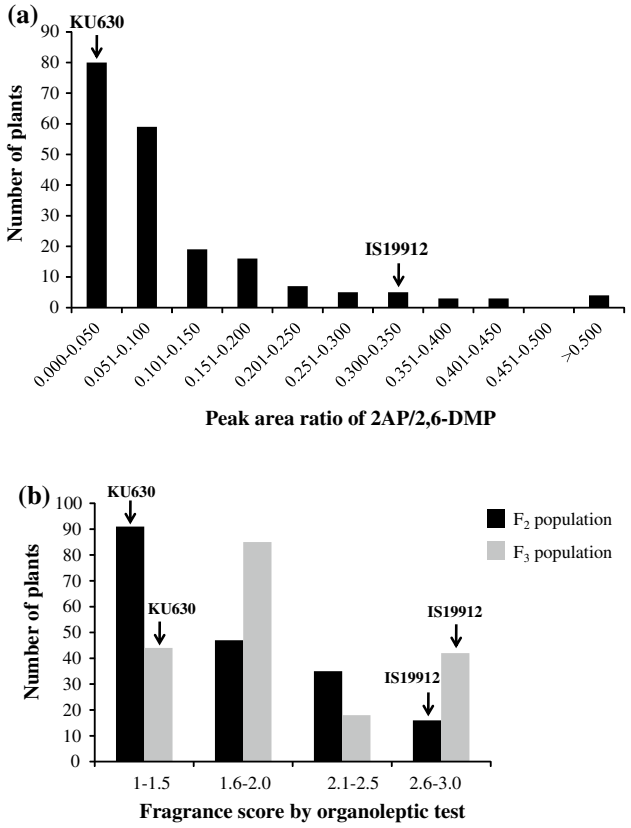


**Fig. 1** Frequency distribution of 2AP concentration (peak area ratio of 2AP/2,6-DMP per gram of seeds) in the $F_2$ population analyzed by HS-GC-NPD (**a**), and of fragrance score in the $F_2$ and $F_{2:3}$ populations evaluated by organoleptic test (**b**)

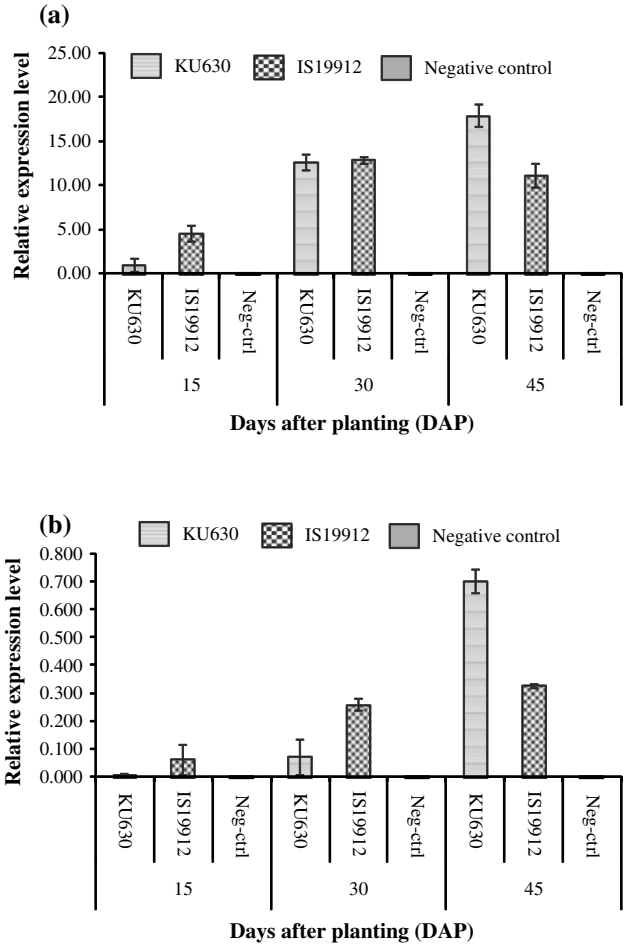


**Fig. 2** Expression of *SbBADH1* (**a**) and *SbBADH2* (**b**) in fully expanded leaves of KU630 and IS19912 at 15, 30 and 45 days after planting (DAP)

with a 3:1 ratio ($\chi^2 = 0.30$, $P = 0.58$). This suggested that the fragrance in IS19912 is conditioned by a single recessive gene. For the $F_{2:3}$ population, the segregation ratio for non-fragrant:fragrant was 1:3 ($\chi^2 = 0.40$, $P = 0.53$). This confirmed that the fragrance in IS19912 is a monogenic recessive trait.

## Expression of *SbBADH1* and *SbBADH2*

BLAST search against sorghum reference genome sequence identified three annotated loci as *BADH* genes. They were Sb06g019210, Sb06g019200, and Sb07g020650. Sb06g019210 located on chromosome 6 at the position chromosome_6: 48776457–48778890 (plus strand) with full genomic sequence length of 2,434 bp and total coding sequence (CDS) of 585 bp being translated from five exons. The predicted protein of this locus was 194 amino acids in length. Sb06g019210 located on chromosome 6 is the same as Sb06g019200 at the position chromosome_6: 48773986–48776105 (plus strand) with full genomic sequence length of 2,120 bp, the predicted protein was 258 amino acids in length being translated from nine exons with the CDS of 777 bp. Sb07g020650 was on chromosome 7 at the position chromosome_7: 53503552–53509157 (plus strand) with full genomic sequence length of 5,606 bp and the predicted protein of 505 amino acids in length. The protein is translated from 15 exons with the total CDS of 1,518 bp. BLAST search using CDS of the three loci against soybean and rice showed that Sb06g019210 and Sb06g019200 are parts of the *BADH1* (*SbBADH1*) gene, and that Sb07g020650 is the *BADH2* (*SbBADH2*) gene (data not shown).

Expression of *SbBADH1* and *SbBADH2* in IS19912 and KU630 was determined in leaves at 15, 30, and 45 DAP using real-time PCR analysis. *SbBADH1* and *SbBADH2* were expressed in both sorghum accessions. For the *SbBADH1*, both sorghum varieties showed constitutive expression of this gene at all three stages (Fig. 2a). KU630 showed the highest expression level of *SbBADH1* at 45 DAP, followed by 30 and 15 DAP, respectively. IS19912 had the greatest expression level at 30 DAP, followed by

45 and 15 DAP, respectively. In case of *SbBADH2*, KU630 and IS19912 showed similar expression patterns (Fig. 2b). They showed the higher expression at 45 DAP, followed by 30, and 15 DAP, respectively. However, $\chi^2$ tests revealed that expression levels of both *SbBADH1* and *SbBADH2* at the three stages in KU630 and IS19912 were not different ($\chi^2 = 3.75$ and $P = 0.15$ for *SbBADH1* and $\chi^2 = 0.27$, $P = 0.87$ for *SbBADH2*).

## Sequence of *SbBADH2* and allele-specific marker

A genomic region covering *SbBADH2* was amplified by 10 primer pairs (Table S1) and sequenced in both IS19912 and KU630 to identify possible mutation(s) that may cause dysfunction of the gene. The sequence length of *SbBADH2* in KU630 was 6,447 bp, while that of IS19912 was only 5,003 bp (Fig. 3a). When the sequences of IS19912 and KU630 were aligned against the reference genome sequence, eight mutations, seven base substitutions and one indel were identified among the three sequences. Two of the seven base substitutions were in exons 1 and 3, while the other 5 were in introns (2 in intron 2, and each one in introns 8, 9 and 11) (Fig. S1). However, only the mutation in exon 1 causes the amino acid change (from tryptophan in reference sequence to glycine in IS19912 and KU630) (Fig. S2). For the indel, a continuous 1,444 bp deletion was found in IS19912 at the position chromosome_7: 53508468-53509912. The deletion encompasses exons 12–15 and 3′ UTR of the *SbBADH2* in IS19912 which causes frameshift reading and introduces a premature stop codon (Fig. 3b). This mutant allele was named as *Sbbadh2-1*. CDS of *SbBADH2* of KU630 and the reference sequence were the same in length (data not shown). They were 1,518 nt long with 15 exons encoding 505 amino acids. In contrast, CDS of IS19912 was only 1,261 nt long encoding 420 amino acids (Fig. S2).

The indel marker SbBADH2-EX12-15 (forward primer: 5′-GGTTTGGCTGGTGCTGTAAT-3′; reverse primer: 5′- AACTTCCACCGAGAATGACG-3′) was developed to detect size polymorphism due to the 1,444 bp deletion in IS19912. It was targeted to amplify 343 bp PCR product in IS19912 and 1,787 bp PCR product in KU630. The marker showed clear size polymorphism between the two sorghum varieties and among their $F_2$ progenies (Fig. 4).

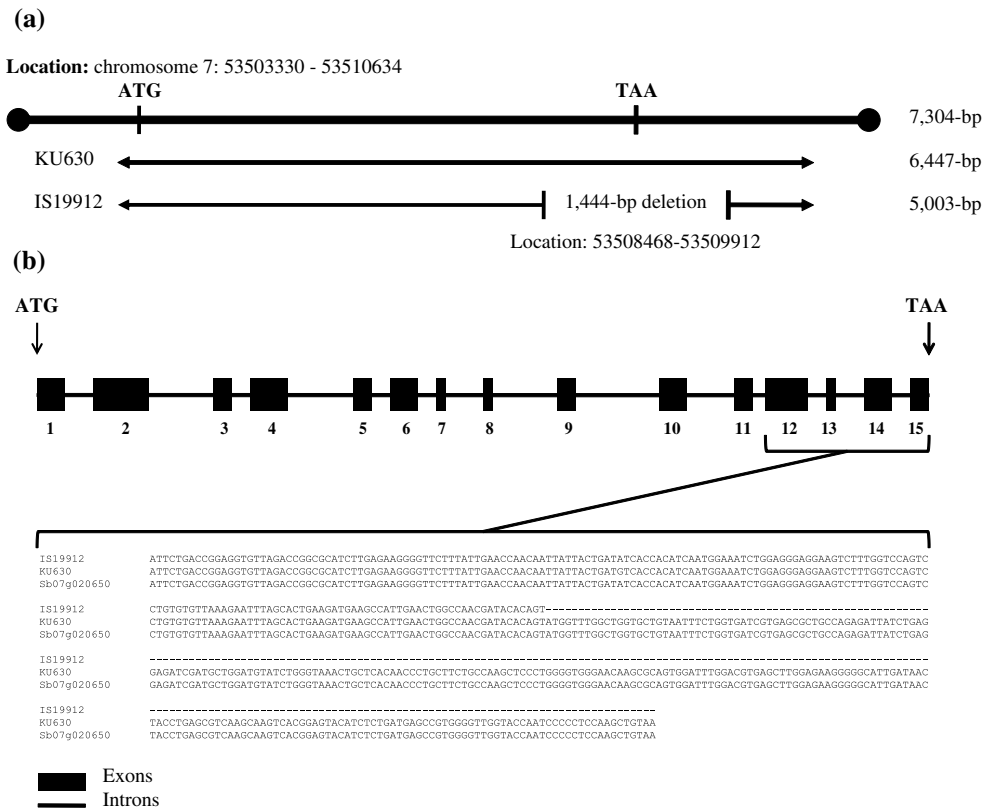


**Fig. 3** A scheme depicting the location of 1,444 bp deletion on chromosome 7 which introduces a premature stop codon in *SbBADH2* gene of sorghum IS19912 (**a**). The deletion encompasses exons 12–15 of *SbBADH2* (**b**). Start codon (ATG), 15 exons (*black boxes*), 14 introns (*solid line*) and stop codon (TAA) are illustrated. Alignment of coding sequence of exons 12–15 of *SbBADH2* for fragrant sorghum IS19912, non-fragrant sorghum KU630, and the reference sequence is shown

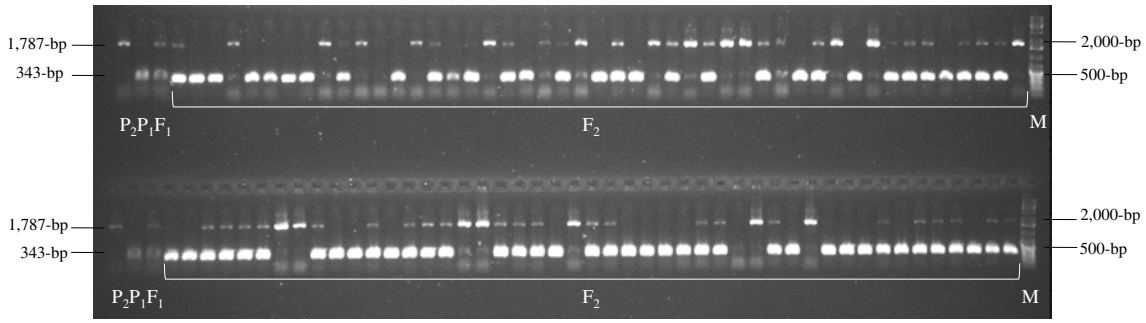


**Fig. 4** Segregation of the allele *Sbbadh2-1* in 94 $F_2$ progenies from the cross between IS19912 ($P_1$) and KU630 ($P_2$) as detected by the indel marker SbBADH2-EX12-15

## Bulk segregant analysis and QTL mapping

BSA and QTL mapping were conducted to determine the association between fragrance and *SbBADH2*. Marker screening revealed that 93 out of 298 SSRs surveyed gave polymorphism between the parents. When these 93 polymorphic markers and the marker SbBADH2-EX12-15 were used in BSA, nine markers including SbBADH2-EX12-15 showed polymorphism between FB and NFB. These markers and five polymorphic markers from chromosome 6 were used to constructed a linkage map. The linkage map was 182 cM in total length covering chromosomes 6 (Fig. 5a) and 7 (Fig. 5b). Marker orders of the map were congruent with the positions of the markers on the sorghum reference genome.

LOD threshold determined by permutation test of 2AP concentration, and fragrance score in the $F_2$ and $F_{2:3}$ populations were 3.6, 7.8, and 3.4, respectively. CIM identified a single major QTL for 2AP concentration, *q2AP*, in the $F_2$ population with a LOD score of 7.9 (Fig. 5b). The *q2AP* was on chromosome 7 and was 4.9 cM away from the marker SbBADH2-EX12-15. The *q2AP* explained 58.49 % of the total variation in 2AP concentration with additive and dominant effects of −0.36 and −0.42, respectively. CIM detected a QTL for fragrance score, *qFgr*, in $F_2$ and $F_{2:3}$ populations with the LOD score of 28.1 and 9.6,

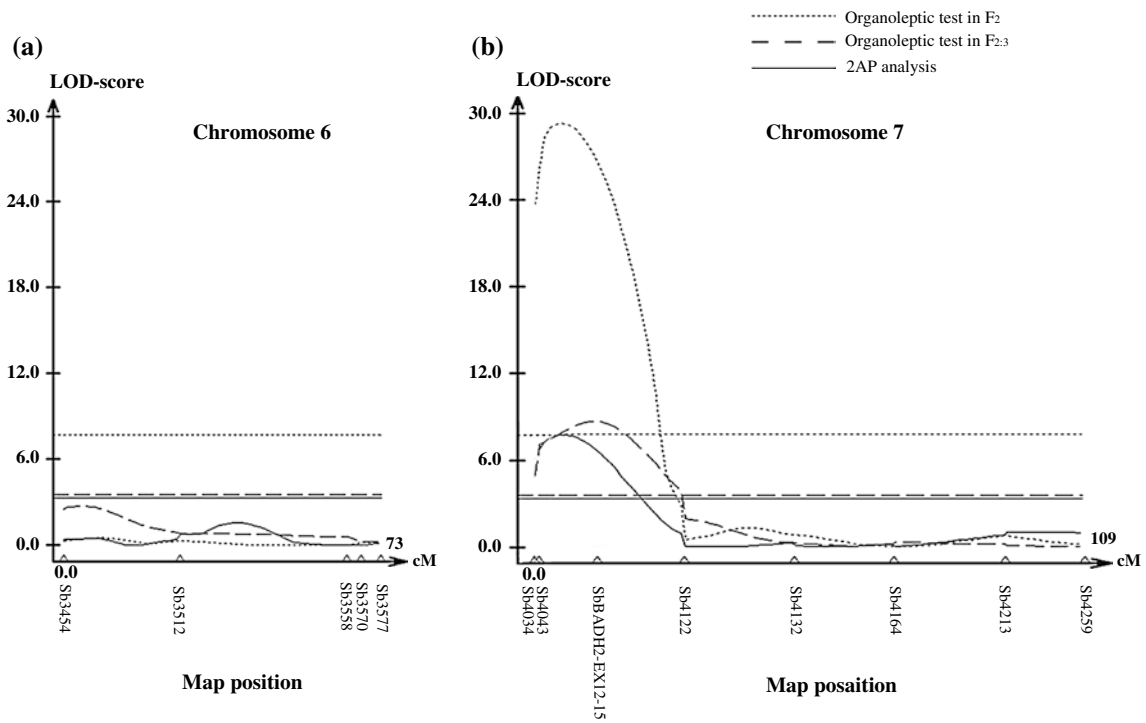


**Fig. 5** LOD-score plots for QTLs controlling fragrance score (*qFgr*) and for 2AP concentration (*q2AP*) detected on chromosomes 6 (**a**) and 7 (**b**) by composite interval mapping in the $F_2$ or $F_{2:3}$ population of IS19912 × KU630

respectively (Fig. 5b). The *qFgr* was located on chromosome 7. It is located at 5.7 and 1.2 cM from the marker SbBADH2-EX12-15 in $F_2$ and $F_{2:3}$ populations and accounted for 12.08 and 29.62 % of the fragrance score, respectively. *qFgr* possessed additive and dominant effects of −0.34 and −0.69 in the $F_2$ population, and of −0.65 and 0.14 in the $F_{2:3}$ population, respectively. At all QTLs detected, alleles from IS19912 increased 2AP concentration or fragrance score.

## Discussion

Fragrance in consumed parts of crops such as rice grains and soybean seeds is persuasive to consumers and is value-added in agricultural products. This trait is governed by a single recessive gene (Sood and Siddiq 1978; Berner and Hoff 1986; AVRDC 2003; Juwattanasomran et al. 2012). In sorghum, previous studies by Ayyangar (1938) and Murty et al. (1982), respectively, showed that the fragrance in "Kinungapembo" and "IS19907" (KEP-472) is a monogenic recessive trait. In our study, we used IS19912 as source of fragrance and found that segregation ratio of fragrance score from organoleptic tests in the $F_2$ plants (3 no fragrance: 1 fragrance) and the $F_{2:3}$ families (1 no fragrance: 3 fragrance) confirmed the result of Ayyangar (1938) and Murty et al. (1982) that fragrance in sorghum is controlled by a single recessive gene. However, it is still necessary to find out if the same gene controls fragrance in different fragrant sorghum cultivars.

Fragrance in rice and soybean is mainly due to the presence of 2AP chemical (Buttery et al. 1983; Fushimi and Masuda 2001). We showed that grains of the fragrant sorghum (IS19912) contain high concentration of 2AP (0.3019), while those of non-fragrant sorghum (KU630) are lacking of 2AP (0.0000). Thus, the fragrance in sorghum appears to be due to an accumulation of 2AP. Highly significant correlations between 2AP concentration and fragrance score in the $F_2$ and $F_{2:3}$ populations confirm that 2AP is the major chemical contributing to fragrance in sorghum. In rice and soybean, genome mapping and molecular studies revealed that the loss in function of *OsBADH2* on chromosome 8 and *GmBADH2* on chromosome 6, respectively, is the mechanism controlling fragrance (2AP). Mutation(s) in *OsBADH2* that causes defective function of the BADH2 protein results in fragrance. A 7 bp deletion in exon 2 (Shi et al. 2008), an 8 bp deletion in exon 7 (Bradbury et al. 2005; Wanchana 2005), and 803 bp deletion including exons 4 and 5 (Shao et al. 2011) of *OsBADH2*, all causing premature stop codon, are responsible for fragrance in rice. Similarly, mutation(s) in the *GmBADH2* gene causing disrupted function of BADH2 gives rise to fragrance. In soybean, a 2 bp deletion in exon 10 of the *GmBADH2* gene on chromosome 6 introduces a premature stop codon that is responsible for fragrance in soybean (Juwattanasomran et al. 2011). In this study, our results confirm the key role of *BADH2* in fragrance (2AP) formation in plants. A 1,444 bp deletion encompassing exon 12–15 of *SbBADH2* of the fragrant sorghum IS19912 brings in a premature stop codon (Figs. 4 and S1) and thus causing truncated gene product (Fig. 3a, b), and resulting in fragrance production. The association between 1,444 bp deletion in *SbBADH2* and fragrance of IS19912 was confirmed by genotyping the parental lines and the $F_2$ population with the marker SbBADH2-EX12-15 (Fig. 4). The QTL analysis in the $F_2$ population also located the marker SbBADH2-EX12-15 and its flanking markers as a major locus responsible for fragrance in sorghum (Fig. 5). Our results further highlight the association of *BADH2* and domestication for fragrance in crop plants. Although the pathway(s) leading to 2AP synthesis remains unclear, two pathways have been proposed; one is BADH2 dependent while the other is not. In the first pathway, functional BADH2 inhibits the biosynthesis of 2AP by converting γ-aminobutyraldehyde (GABald) to γ -aminobutyric acid (GABA), whereas non-functional BADH2 leads to an accumulation of GABald which is then acetylated to form 2AP (Bradbury et al. 2008). In the second pathway, $\Delta^1$-pyrroline-5-carboxylate synthetase (P5CS) catalyzes the formation of $\Delta^1$-pyrroline-5-carboxylic acid (P5C) which reacts directly with methylglyoxal (MG) to form 2AP (Huang et al. 2008). However, up to the present, only the gene encoding BADH2 protein has been proven to be directly involved in 2AP production. Chen et al. (2012), Niu et al. (2008) and Vanavichit et al. (2005) demonstrated that when *OsBADH2* in non-fragrant rice varieties were knocked-down, those varieties had significantly elevated 2AP content in grain, while wild type had no 2AP. In contrast, Chen et al. (2008) showed that when a functional *OsBADH2* was expressed in fragrant rice cultivars, such cultivars became non-fragrant with no or very low 2AP content. Arikit et al. (2010) showed that silencing the *GmBADH2* gene results in 2AP biosynthesis in non-fragrant soybean.

*SbBADH2* was discovered as a major locus controlling fragrance in sorghum. This gene explained as high as 58.49 % of the total variation in 2AP concentration, but explained 12.08 and 29.62 % of the total fragrance score variation in the $F_2$ and $F_{2:3}$ populations, respectively. Although the low phenotypic variance explained for fragrance scores in both $F_2$ and $F_{2:3}$ populations may stem from the inaccuracy of the sensory test, this also suggests that there may be other genes conditioning the fragrance. Similar results have been reported in rice and soybean. Amarawathi et al. (2008) reported that the major locus (*OsBADH2*) for fragrance in rice explained only 18.9 % of the total variation of sensory fragrance score.

Juwattanasomran et al. (2011) reported that *GmBADH2* accounted for as low as 16.7 % of the total variation of 2AP concentration. Besides *OsBADH2* which is the major locus conferring fragrance in rice, three minor loci, one each on chromosomes 3, 4 and 12, were also contributing to the fragrance (Lorieux et al. 1996; Amarawathi et al. 2008). Amarawathi et al. (2008) further showed that *OsBADH1* is the candidate gene for the fragrance QTL on the chromosome 4. Later, Singh et al. (2010) demonstrated that a haplotype of *OsBADH1* associates with fragrance in rice, although they found that *OsBADH2* is the major gene controlling the fragrance. In our study, *SbBADH1* showed no association with fragrance in sorghum IS19912, because no QTL was found on the chromosome 6 where *SbBADH1* locating on (based on the reference sequence *SbBADH1* locates between SSR markers SB3558 and SB3570) (Fig. 5). However, this result is specific for the cross between IS19912 and KU630. There is still a possibility that *SbBADH1* may be involved in fragrance in other fragrant sorghum genotypes or other crosses. Since, 2AP can be synthesized from two distinct pathways with different precursors and intermediate chemicals, genetic factor(s) regulating such chemicals (GABald, P5CS, P5C, and MG) may also affect fragrance level in sorghum. Additional studies are necessary to determine whether any of the genes encoding these chemicals associate with the fragrance. It is worth noting that 11 mutant alleles of *OsBADH2* associated with fragrance in rice have been identified (Kovach et al. 2009; Shao et al. 2011) but several fragrance rice accessions possess no mutation in *OsBADH2* (Kovach et al. 2009).

*SbBADH2*, *OsBADH2*, and *GmBADH2* are orthologs (Arikit et al. 2010). Although domestication histories of sorghum, rice and soybean are different, mutation(s) in *BADH2* was associated with fragrance in all these crops (this study; Bradbury et al. 2005; Kovach et al. 2009; Shi et al. 2008; Shao et al. 2011; Arikit et al. 2010; Juwattanasomran et al. 2011, 2012). Functional loss or reduction of BADH2 in these crops appears not to affect survival, and may positively affect performance under drought condition in case of rice (Yoshihashi et al. 2004).

Breeding for fragrance in crops can be difficult, albeit fragrance is principally controlled by a single gene. Fragrance evaluation can be performed by sensory tests such as sniffing and chewing, or by 2AP quantification using advanced chromatography analyses such as gas chromatography–mass spectrometry (GC–MS) or HS-GC. Although sensory test is rapid and inexpensive, it is subjected to human error and thus may give inconsistent and unreliable results (Juwattanasomran et al. 2011). While 2AP analysis is accurate and reliable, the techniques are expensive and require expertise to analyze, hence make it impractical for a large-scale breeding program. In addition, recessive nature of the fragrance requires repeated evaluation after self-pollination generations during backcross breeding. An indel marker SbBADH2-EX12-15 can be used as a functional marker for genotyping the fragrance allele *Sbbadh2-1* to overcome the above-mentioned difficulties. The marker is co-dominant and detects amplicon size polymorphism of *SbBADH2* between fragrance and non-fragrance sorghums (Fig. 5). This marker is also economic and simple because it is PCR-based and detectable by agarose gel electrophoresis with high reproducibility. The marker can facilitate a marker-assisted selection for fragrant sorghum genotypes with the *Sbbadh2-1* allele in a large breeding program using IS19912 or other fragrant sorghums that possess this fragrance allele.

**Acknowledgments** This research was supported by the Royal Golden Jubilee Ph.D. Scholarship co-funded by the Thailand Research Fund (TRF) and Kasetsart University to P. Somta and C. Yundaeng.

**Conflict of interest** The authors declare that they have no conflict of interest.

**Ethical standards** All the experiments performed in this study comply with the current laws of Thailand.

## References

Adams A, De Kimpe N (2006) Chemistry of 2-acetyle-1-pyrroline, 6-acetyl-1,2,3,4-tetrahydropyridine, 2-acetyl-2thiazoline, and 5-acetyl-2,3-dihydro-4*H*-thiazine: extraordinary maillard flavor compounds. Chem Rev 106:2299–2319

Amarawathi Y, Singh R, Singh AK, Singh VP, Mohapatra T, Sharma TR, Singh NK (2008) Mapping of quantitative trait loci for basmati quality traits in rice (Oryza sativa L.). Mol Breed 21:49–65

Arikit S, Yoshihashi T, Wanchana S, Uyen TT, Huong NTT, Wongpornchai S, Vanavichit (2010) Deficiency in the amino aldehyde dehydrogenase encoded by *GmAMADH2*, the homologue of rice *Os2AP*, enhances 2-acetyl-1-pyrroline biosynthesis in soybeans (*Glycine max* L.). Plant Biotechnol J 9:75–87

AVRDC (2003) AVRDC progress report 2002. AVRDC-The World Vegetable Center, Shanhua

Ayyangar GNR (1938) Studies in sorghum. J Madras Univ 11:131–143

Berner DK, Hoff BJ (1986) Inheritance of scent in American long grain rice. Crop Sci 26:876–878

Bradbury LMT, Fitzgerald TL, Henry RJ, Jin QS, Waters DLE (2005) The gene for fragrance in rice. Plant Biotechnol J 3:363–370

Bradbury LMT, Gillies SA, Brushett DJ, Waters DLE, Henry RJ (2008) Inactivation of an aminoaldehyde dehydrogenase is responsible for the fragrance in rice. Plant Mol Biol 68:443–449

Buttery RG, Ling LC, Juliano BO, Turnbaugh JG (1983) Cooked rice aroma and 2-acetyl-1-pyrroline. J Agric Food Chem 31:823–826

Chen S, Yang Y, Shi W, Ji Q, He F, Zhang Z, Cheng Z, Liu X, Xu M (2008) *Badh2*, encoding betaine aldehyde dehydrogenase, inhibits the biosynthesis of 2-acetyl-1-pyrroline, a major component in rice fragrance. Plant Cell 20:1850–1861

Chen M, Wei X, Shao G, Tang S, Luo J, Hu P (2012) Fragrance of rice grain achieved via artificial microRNA-induced down-regulation of *OsBADH2*. Plant Breed 131:584–590

Food and Agriculture Organization (2010) FAOSTAT. http://faostat.fao.org. Accessed 7 Feb 2013

Fushimi T, Masuda R (2001) 2-acetyl-1-pyrroline concentration of the vegetable soybean. In: Lumpkin T, Shanmugasundaram S (eds) Proceeding of the 2nd international vegetable soybean conference. Washington State University, Pullman, p 39

Huang TC, Teng CH, Chang JL, Chuang HS, Ho CT, Wu ML (2008) Biosynthetic mechanism of 2-acetyl-1-pyrroline and its relationship with $\Delta^1$-pyrroline-5-carboxylic acid and methylglyoxal in aromatic rice (*Oryza sativa* L.) callus. J Agric Food Chem 56:7399–7404

Itani T, Takeda A, Tamaki M, Hayata Y, Fushimi T, Hashizume K (2004) Chemical and sensory evaluation of aromatic rice cultivar. In: Fischer T, Turner N, Angus J, McIntyre L, Robertson M, Borell A, Lloyd D (eds) New Directions for a Diverse Planet, Proceedings for the 4th International Crop Science Congress, Brisbane, Australia. www.cropscience.org.au/icsc2004/poster/5/1/1/811_itani.htm. Accessed 6 23 June 2013

Juwattanasomran R, Somta P, Chankaew S, Shimizu T, Wongpornchai S, Kaga A, Srinives P (2011) A SNP in *GmBADH2* gene associates with fragrance in vegetable soybean variety "Kaori" and SNAP marker development for the fragrance. Theor Appl Genet 122:533–541

Juwattanasomran R, Somta P, Kaga A, Chankaew S, Shimizu T, Sorajjapinun W, Srinives P (2012) Identification of a new fragrance allele in soybean and development of its functional marker. Mol Breeding 29:13–21

Kadam BS, Patankar VK (1938) Inheritance of aroma in rice. Chronic Bot 4:32

Kosambi DD (1944) The estimation of map distances from recombination values. Ann Eugen 12:172–175

Kottur GL (1919) Classification and description of the jowars of the Bombay and Karnataka. Bulletin Number 92, Department of Agriculture, Bombay

Kovach MJ, Calingacion MN, Fitzgerald, McCough SR (2009) The origin and evolution of fragrance in rice (*Oryza sativa* L.). Proc Natl Acad Sci USA 106:14444–14449

Lander ES, Green P, Abrahamson J, Barlow A, Daly MJ, Lincoln SE, Newberg LA (1987) MAPMAKER: an interactive computer package for constructing primary genetic linkage maps of experimental and natural populations. Genomics 1:174–181

Livak KJ, Schmittgen TD (2001) Analysis of relative gene expression data using real-time quantitative PCR and the 2-$\Delta\Delta C_T$ method. Methods 25:402–408

Lodhi MA, Ye GN, Weeden NF, Reisch BI (1994) A simple and efficient method for DNA extraction from grapevine cultivars and *Vitis* species. Plant Mol Biol Rep 12:6–13

Lorieux M, Petrov M, Huang N, Guiderdoni E, Ghesquiere A (1996) Aroma in rice: genetic analysis of a quantitative trait. Theor Appl Genet 93:1145–1151

Michelmore RW, Paran I, Kesseli RV (1991) Identification of markers linked to disease-resistance genes by bulked segregant analysis: a rapid method to detect markers in specific genomic regions by using segregating populations. Proc Natl Acad Sci USA 88:9828–9832

Murty DS, Nicodemus KD, House LR (1982) Inheritance of basmati and dimpled seed in sorghum. Crop Sci 22:1080–1082

Niu X, Tang W, Huang W, Ren G, Wang Q, Luo D, Xiao Y, Yang S, Wang F, Lu BL, Gao F, Lu T, Liu Y (2008) RNAi-directed downregulation of *OsBADH2* results in aroma (2-acetyl-1-pyrroline) production in rice (*Oryza sativa* L.). BMC Plant Biol 8:100. doi:10.1186/1471-2229-8-100

Paterson AH, Bowers JE, Bruggmann R, Dubchak I, Grimwood J, Gundlach H, Haberer G, Hellsten U, Mitros T, Poliakov A, Schmutz J, Spannagl M, Tang H, Wang X, Wicker T, Bharti AK, Chapman J, Feltus FA, Gowik U, Grigoriev IV, Lyons E, Maher CA, Martis M, Narechania A, Otillar RP, Penning BW, Salamov AA, Wang Y, Zhang L, Carpita NC, Freeling M, Gingle AR, Hash CT, Keller B, Klein P, Kresovich S, McCann MC, Ray Ming, Peterson DG, Rahman M, Ware D, Westhoff P, Mayer KFX, Messing J, Rokhsar DS (2009) The *Sorghum bicolor* genome and diversification of grasses. Nature 457:551–556

Prasada KE, Murty DS (1979) A *basmati* (scented) sorghum from Madhya Pradesh. Curr Sci 48:824–826

R Development Core Team (2010) R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna. ISBN 3-90005107-0

Rozen S, Skaletsky HJ (2000) Primer3 on the WWW for general users and for biologist programmers. In: Krawetz S, Misener S (eds) Bioinformatics methods and protocols: methods in molecular biology. Humana Press, Totowa, pp 365–386

Shao GN, Tang A, Tang SQ, Luo J, Jiao GA, Wu JL, Hu PS (2011) A new deletion mutation of fragrant gene and the development of three molecular markers for fragrance in rice. Plant Breed 130:172–176

Shi WW, Yang Y, Chen SH, Xu ML (2008) Discovery of a new fragrance allele and the development of functional markers for the breeding of fragrant rice varieties. Mol Breeding 22:185–192

Singh A, Singh PK, Singh R, Pandit A, Mahato AK, Gupta DK, Tyagi K, Singh AK, Singh NK, Sharma TR (2010) SNP haplotypes of the BADH1 gene and their association with aroma in rice (*Oryza sativa* L.). Mol Breed 26:325–328

Somta P, Musch W, Kongsamai B, Chanprame S, Nakasathien S, Toojinda T, Sorajjapinun W, Seehaluk W, Tragoonrung S, Srinives P (2008) New microsatellite markers isolated from mungbean (*Vigna radiata* (L.) Wilczek). Mol Ecol Resour 8:1155–1157

Sood BG, Siddiq EA (1978) A rapid technique for scent determination in rice. Indian J Genet Plant Breed 38:268–271

Sriseadka T, Wongpornchai S, Kitsawatpaiboon P (2006) Rapid method for quantitative analysis of the aroma impact compound, 2-acetyl-1-pyrroline, in fragrance rice using automated headspace gas chromatography. J Agric Food Chem 54:8183–8189

Vanavichit A, Tragoonrung S, Toojinda T, Wanchana S, Kamolsukyunyong W (2005) Transgenic rice plants with reduced expression of Os2AP and elevated levels of 2-acetyl-1-pyrroline. United States Patent, Patent No. US 7,319,181 B2

Wanchana S (2005) Identification of genes controlling grain aroma and amylose content for positional cloning and marker-assisted selection program in rice (*Oryza sativ*a L.). Kasetsart University, Thailand

Wang S, Basten CJ, Zeng ZB (2010) Windows QTL Cartographer 2.5. Department of Statistics. North Carolina State University, Raleigh

Wood AJ, Saneoka H, Rhodes D, Joly RJ, Goldsbrough PB (1996) Betaine aldehyde dehydrogenase in *Sorghum*. Plant Physiol 110:1301–1308

Yonemaru J, Ando J, Mizubayashi T, Kasuga S, Matsumoto T, Yano M (2009) Development of genome-wide simple sequence repeat markers using whole-genome shotgun sequences of sorghum (*Sorghum bicolor* (L.) Moench). DNA Res 16:187–193

Yoshihashi T, Nguyen TTH, Kabaki N (2004) Area dependency of 2-acetyl-1-pyrroline content in an aromatic rice variety, Khao Dawk Mali 105. Jarq-Japan Agric Res Q 38:105–109