Theor Appl Genet (2011) 122:533–541
DOI 10.1007/s00122-010-1467-6

ORIGINAL PAPER

A SNP in *GmBADH2* gene associates with fragrance in vegetable soybean variety "Kaori" and SNAP marker development for the fragrance

Ruangchai Juwattanasomran · Prakit Somta · Sompong Chankaew · Takehiko Shimizu · Sugunya Wongpornchai · Akito Kaga · Peerasak Srinives

Received: 25 March 2010 / Accepted: 11 October 2010 / Published online: 3 November 2010

Abstract Fragrance in soybean is due to the presence of 2-acetyl-1-pyrroline (2AP). *BADH2* gene coding for betaine aldehyde dehydrogenase has been identified as the candidate gene responsible for fragrance in rice (*Oryza sativa* L.). In this study, using the RIL population derived from fragrant soybean cultivar "Kaori" and non-fragrant soybean cultivar "Chiang Mai 60" (CM60), STS markers designed from *BADH2* homolog were found associating with 2AP production. Genetic mapping demonstrated that QTL position of fragrance and 2AP production coincides with the position of *GmBADH2* (*Glycine max* betaine aldehyde dehydrogenase 2). Sequence comparison of *GmBADH2* between Kaori and non-fragrant soybeans revealed non-synonymous single-nucleotide polymorphism (SNP) in exon 10. Nucleotide substitution of G to A in the exon results in an amino acid change of glycine (GGC; G) to aspartic acid (GAC; D) in Kaori. The amino acid substitution changes the conserved EGCRL**G**PIVS motif of *GmBADH2*, which is essential for functional activity of GmBADH2 protein, to EGCRL**D**PIVS motif, suggesting that the SNP in *GmBADH2* is responsible for the fragrance in Kaori. Five single nucleotide-amplified polymorphism (SNAP) markers which are PCR-based allele specific SNP markers were developed for fragrance based on the SNP in *GmBADH2*. Two markers specific to A allele produced a band in only Kaori, while three markers specific to G alleles produced a band in only CM60. The simple PCR-based allele specific SNAP markers developed in the present study are useful in marker-assisted breeding of fragrant soybean.

Communicated by H. T. Nguyen.

R. Juwattanasomran and P. Somta, have contributed equally to this research.

Electronic supplementary material The online version of this article (doi:10.1007/s00122-010-1467-6) contains supplementary material, which is available to authorized users.

R. Juwattanasomran · P. Somta · S. Chankaew · P. Srinives (✉)
Department of Agronomy, Faculty of Agriculture at Kamphaeng Saen, Kasetsart University, Nakhon Pathom 73140, Thailand
e-mail: agrpss@yahoo.com

T. Shimizu
Institute of Society for Techno-Innovation of Agriculture, Forestry and Fisheries, 446-1 Ippaizuka, Kamiyokoba, Tsukuba, Ibaraki 305-0854, Japan

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

A. Kaga
Soybean Genome Research Team, Division of Genome and Biodiversity Research, National Institute of Agrobiological Sciences, 2-1-2 Kannondai, Tsukuba, Ibaraki 305-8602, Japan

Introduction

Vegetable soybean (also known as edamame or green soybean) is a type of soybean whose young pods at R6 stage are harvested, cooked, and consumed as snack or vegetable (Shanmugasundaram et al. 1991). It has been mainly grown and consumed in East and Southeast Asian countries, viz. China, Indonesia, Japan, Korea, Taiwan, Thailand, and Vietnam. Vegetable soybean is currently gaining popularity in USA and Europe because of its high nutritive values, including proteins, amino acids, vitamins, minerals, and isoflavones. The three largest producers of vegetable soybean are China, Japan, and Taiwan, where

about 1,700,000, 71,000, and 57,000 tons are annually produced, respectively. Japan is the largest importer of about 70,000 tons per annum. Masuda (1991) stated that the export quality of vegetable soybean involved pod and green bean appearance, taste, flavor, texture, and nutritional value.

Vegetable soybean varieties that have unique characteristics such as highly sweet seeds, extra large pods and seeds, and distinct seed fragrance command higher price and demand than those normal varieties. Therefore, soybean breeders always give high priority to these characters in improving vegetable soybean varieties. Fushimi and Masuda (2001) reported that vegetable soybean cultivar "Dadachamame" and "Chakaori" have pleasant fragrance. They found that the fragrance of Dadachamame cultivar is due to 2-acetyl-1-pyrroline (2AP), the same volatile compound that Buttery et al. (1983) found associating with fragrance in rice. The concentration of 2-AP in Dadachamame is highest at the vegetable soybean stage and gradually decreased afterwards (Fushimi and Masuda 2001). Genetic study using organoleptic test by researchers at AVRDC-The World Vegetable Center (formerly Asian Vegetable Research and Development Center; AVRDC) found that the fragrance is controlled by a single recessive gene (AVRDC 2003).

In rice, genetic studies revealed that the fragrance is conditioned by a single recessive gene (Sood and Siddiq 1978), or two recessive genes (Pinson 1994). The major gene controlling fragrance is located on chromosome 8 of rice (Ahn et al. 1992; Lorieux et al. 1996; Jin et al. 2003; Wanchana 2005). Bradbury et al. (2005) and Wanchana (2005) found that an 8-bp deletion in exon 7 of a betaine aldehyde dehydrogenase gene (*OsBADH2*) is responsible for the fragrance in rice. The deletion causes a premature stop codon, and thus generates non-functional *OsBADH2* and fragrance. Shi et al. (2008) reported that a seven base-pair deletion in exon 2 also causes the null function of *OsBADH2* resulting in fragrance in rice. Silencing *OsBADH2* gene in non-fragrant rice varieties using transformation results in elevated 2AP biosysthesis and thus fragrance in those varieties (Vanavichit et al. 2008; Niu et al. 2008). Similarly, transferring functional *OsBADH2* gene into fragrant rice resulted in non-fragrant rice (Chen et al. 2008). Amarawathi et al. (2008) detected two additional QTLs controlling fragrance in rice and suggested that *OsBADH1*, a homolog of *BADH2*, located on rice chromosome 4, is one of candidate genes attributable to fragrance in rice. Although the exact biosynthesis pathway of 2AP is still not known, Bradbury et al. (2008) proposed that functional OsBADH2 in non-fragrant rice changes 4-aminobutanal into 4-aminobutyrate (GABA) resulting in no synthesis of 2AP, whereas non-functional OsBADH2 in fragrant rice fails to do so, and consequently 4-aminobutanal is changed into precursor and 2AP, respectively. Since 2AP is also responsible for the fragrance in soybean, it is possible that the null function of BADH2 protein may be responsible for the fragrance.

In this study, we reported the gene mapping and genetic basis of fragrance in vegetable soybean variety "Kaori". The objective of this study was to (1) locate major QTL(s) associated with the fragrance and (2) sequence and characterize the candidate genes, *BADH1* and *BADH2*, for the fragrance.

Materials and methods

Plant materials and DNA isolation

A RIL (recombinant inbred line) mapping population was developed by single-seed descent method from 82 F_2 plants derived from a cross between "Kaori" and "Chiang Mai 60" (hereafter called CM60). Kaori is a landrace fragrant vegetable soybean variety from Japan, while CM60 is a non-fragrant grain soybean variety from Thailand. Kaori was crossed as female parent to CM60. The F_8 RILs together with their parents were grown during November 2007 to March 2008 at Kasetsart University, Kamphaeng Saen, Nakhon Pathom, Thailand, for DNA extraction and fragrance evaluation by organoleptic test. The F_9 RILs were grown during July to October 2008 at Maejo University, Chiang Mai, Thailand for fragrance evaluation by 2AP analysis.

DNA of Kaori, CM60 and the RILs were extracted from young leaves using a CTAB method described by Godhi et al. (1994) with a minor modification. DNA concentration was quantified in 1.5% agarose gel electrophoresis.

Evaluation for fragrance

Seeds at R6 stage of Kaori, CM60 and 82 RILs were evaluated for fragrance by two methods: 2-AP analysis and organoleptic test. Seeds harvested from 5 to 10 plants were used to represent each RIL. The concentration of 2-AP was assessed by automated headspace gas chromatography (HS-GC) following the method of Sriseadka et al. (2006) with modification in extraction and calibration procedure in order to accommodate the soybean seed samples. An exact concentration of 2,4-dimethylpyridine (DMP), used as internal standard, was spiked in each sample of soybean seed weighed exactly 1.00 g, and the concentration of 2AP was expressed in terms of peak area ratio of 2-AP/DMP. For organoleptic test, a panel of five volunteers evaluated fragrance by chewing the freshly cooked seeds upon blanching in boiling water for 5 min. The fragrance was scored as 1 (no fragrance), or 2 (mildly to moderately fragrant), or 3 (highly fragrant).

Identification of *BADH* gene and STS and SSR marker development

We hypothesized that the fragrance in "Kaori" vegetable soybean is also caused by mutation in *BADH* gene(s). To test our hypothesis, two different cDNA sequences coding peroxisomal betaine aldehyde dehydrogenase from soybean cultivar "Bansei Shirodaizu", GenBank accession no. AB333793 and AB333794, reported by Arai et al. (2008) were obtained from GenBank database (http://www.ncbi.nlm.nih.gov) and then blasted against soybean genome sequence assembly (Glyma 1.01, http://www.phytozome.net) to identify *BADH* genes. Once identified, oligonucleotide primers were designed to amplify all the predicted exons in the *BADH* genes as sequence-tagged-site (STS) markers. Simple sequence repeats (SSRs) were also searched from the predicted genes and the DNA sequences surrounding the genes using SSRIT program (http://www.gramene.org/microsat/ssr.html). Primer3 program (http://frodo.wi.mit.edu) was used to design all the primers (Table S1 and Table S2).

DNA marker analysis and QTL mapping

To initially identify markers associated with fragrance, the bulked-segregant analysis (BSA) technique (Michelmore et al. 1991) was conducted. Two bulks of DNA, each from ten RIL lines of high 2AP (fragrance bulk; FB) and low 2AP (non-fragrance bulk; NFB) were formed from RIL lines based on the results from 2AP analysis. Forty-six markers (31 STSs and 15 SSRs) developed above and 454 SSR markers distributed on 20 linkage groups of soybean (http://www.soybase.org) were used for screening polymorphism between the parents. PCR amplification, electrophoresis, and DNA band visualization were performed as described by Somta et al. (2008) except that annealing temperate for the amplification was 47–55°C depending on the primers used. Markers showing polymorphism between the parents were subsequently used in BSA.

Upon identification of markers showing polymorphism between FB and NFB bulks, markers on the same linkage group that reveal polymorphism between CM60 and Kaori were used to genotype all 82 RILs in the population to locate QTL related to the fragrance on the linkage group. The linkage map was constructed using JoinMap 3.0 (Van Ooijen and Voorrips 2001) at the LOD value of ≥ 3 based on Kosambi's mapping function. QTL for fragrance was located by composite interval mapping (CIM) with QTL Cartographer 2.5 (Wang et al. 2007). Genome-wide significant threshold at $P = 0.01$ was computed by 10,000 runs of a permutation test using the same software.

DNA sequencing and sequence analysis

To compare the difference in coding sequences of *BADH* genes in CM60 and Kaori, five and two primer pairs were designed (Table S3) to amplify genomic sequence including exon coding sequences from GenBank accession no. AB333793 (*BADH1*) and AB333794 (*BADH2*). The start and end positions covering *BADH1* and *BADH2* were Gm06:16161000..16169000 and Gm05:1232000..1238100 (Glyma 1.0 http://www.phytozome.net/soybean.php), respectively. Five primer pairs targeted to amplify five overlapping PCR fragments approximately covering 8 kb region of *BADH1*. The two primer pairs each targeted single genomic PCR fragment of approximately 6.1 kb region containing *BADH2*. The PCR reactions were as follows: The PCR cycling profile was 95°C for 2 min followed by 35 cycles of 94°C for 30 s and 60°C for 3 min. PCR products were run on 1% agarose gel to confirm whether single PCR fragment is amplified. Purified PCR product from ethanol precipitation was used as template DNA in a sequencing reaction. Both strands of each PCR products were sequenced with the Big Dye Terminator v3.1 kit (Applied Biosystems, USA) using internal primers designed to every 300–400 bp based on Glyma1.0. The sequence data were collected using ABI 3130 capillary sequencer. To obtain consensus sequences, multiple chromatogram files were processed for base-calling, and then individual sequences were assembled using Sequencher ver. 4.7 (Gene Codes Corporation) with default parameters.

SNAP marker development for fragrance

To develop single-nucleotide polymorphism (SNP) markers specific to fragrance, the segment of the *BADH* gene harboring the SNP site was used to develop single nucleotide-amplified polymorphism (SNAP) markers (Drenkard et al. 2000). The markers were designed such that the 3′-terminal nucleotide of a primer should be complementary to an allele of a SNP and that the primer should contain an artificial mismatch within 4 bp of the site of SNP. Web-based program SNAPER (http://ausubellab.mgh.harvard.edu) was used to design five SNAP primer pairs (Table 1). PCR condition was the same as described by Somta et al. (2008). PCR cycling profile was 94°C for 2 min followed by 28 cycles of 94°C for 30 s, 55°C for 30 s, 72°C for 1 min, ending with final extension at 72°C for 10 min. After amplification, the PCR products were electrophoresed on a 1.5% agarose gel and stained with ethidium bromide to reveal presence and absence of bands (alleles).

Table 1 Characteristics of SNAP primer pairs specific to the G/A single nucleotide polymorphism of the *GmBADH2* gene

Primer name	Primer sequence	PCR product size (bp)
GmBADH2-G1	F: GGAAGAAGGTTGCAGACCAGG	125
	R: AAAGCATACCTGCCCTTTACTTTAGAA	
GmBADH2-G2	F: TGGAAGAAGGTTGCAGACAAGG	126
	R: AAAGCATACCTGCCCTTTACTTTAGAA	
GmBADH2-G3	F: CCTTGGAAGAAGGTTGCAGACTAAG	129
	R: AAAGCATACCTGCCCTTTACTTTAGAA	
GmBADH2-A1	F: TGGAAGAAGGTTGCAGACGAGA	126
	R: AAAGCATACCTGCCCTTTACTTTAGAA	
GmBADH2-A2	F: TGGAAGAAGGTTGCAGACCAGA	126
	R: AAAGCATACCTGCCCTTTACTTTAGAA	

Results

Fragrance variation in the RIL population and parents

The fragrance evaluation by HS-GC analysis of 2AP revealed that the peak area ratio of 2AP/DMP in the RILs varied from 0 to 0.9231 with the mean of 0.1083. Kaori was highly fragrant with 2AP concentration of 0.6990 while CM60 was non-fragrant with 2AP concentration of 0.0536. The fragrance score evaluated by organoleptic test ranged from 1 to 3 with the mean of 1.99. The scores for Kaori and CM60 were 3 (highly fragrance) and 1 (no fragrance), respectively. Frequency distribution of 2AP concentration and fragrance score of the RIL population is shown in Fig. 1. Relationship between fragrance score and 2AP concentration is shown in Fig. S1 with a significant Pearson correlation (r) of 0.39 ($P < 0.001$, $df = 76$). The fragrance score showed bimodal distribution, while 2AP concentration showed continuous distribution with a bias towards non-fragrant parent, CM60.

Locating *GmBADH* genes and markers associated with the fragrance

By *in silico* search, we identified two annotated genes predicted as *BADH*, viz. Glyma05g01770 and Glyma06g19820. Glyma05g01770 was on chromosome 5 at the position Gm05:1233067..1237025 (−strand) with full genomic sequence length of 3,959 bp. The predicted protein of the Glyma05g01770 was 488 amino acids in length being translated from 16 exons with the total coding sequence (CDS) of 1,467 bp, whereas Glyma06g19820 was on chromosome 6 at the position Gm06:16162014..16167943 (+strand) with full genomic sequence length of 5,930 bp. The predicted protein of the Glyma06g19820 was 503 amino acids in length being translated from 15 exons with the total CDS length of 1,512 bp. Chromosomes 5 and 6 are equivalent to soybean linkage groups A1 and C2, respectively. Homology search for protein sequences of Glyma05g01770 and Glyma06g19820 using blastp against rice non-redundant protein sequence database revealed that

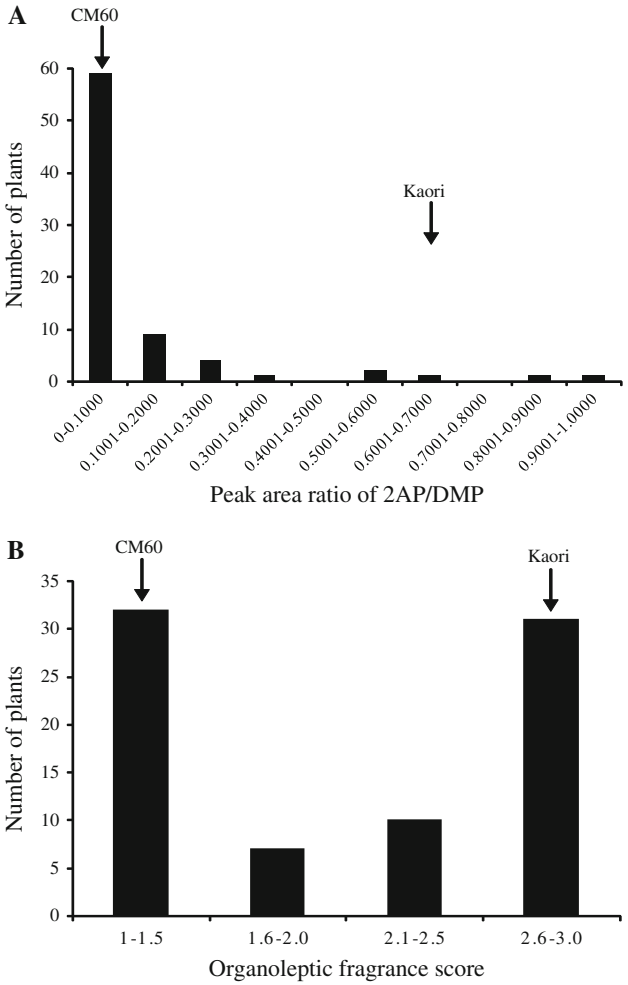

Fig. 1 Frequency distribution of 2AP concentration (peak area ratio of 2AP/DMP per gram of seeds) analyzed by HS-GC (**a**), and of fragrance score evaluated by organoleptic test (**b**) in the RIL population

protein sequence of Glyma05g01770 is similar to BADH2 protein, while that of Glyma06g19820 is similar to BADH1 protein (data not shown). We named Glyma05g01770 and Glyma06g19820 as *GmBADH2* and *GmBADH1*, respectively.

Seventeen STS primer pairs were designed to target all predicted exons of *GmBADH1* and *GmBADH2* (Table S1), and 29 pairs were designed to amplify SSRs in and around *GmBADHs* (Table S2). Two hundred and thirteen out of 500 markers screened showed polymorphism between Kaori and CM60. Only markers BADH2_CDS5-6 which amplified region covering exon 5 and 6, and BADH2_CDS6 which amplified exon 6 showed polymorphism between FB and NFB bulks (Fig. 2), suggesting that *GmBADH2* associated with the fragrance in Kaori and that indel in exon 6 of *GmBADH2* may lead to fragrance in Kaori.

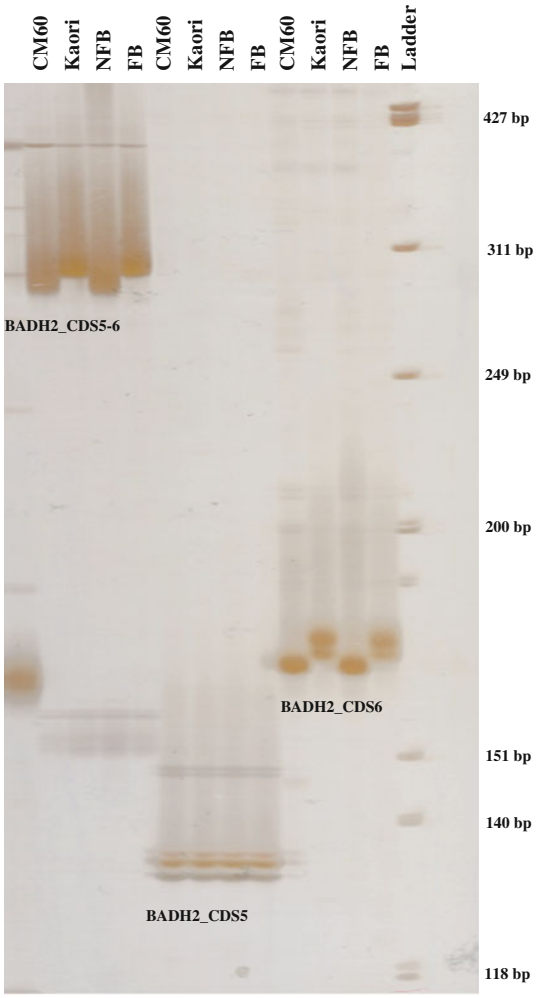

Fig. 2 DNA band polymorphism between CM60 and Kaori, and FB and NFB bulks detected by markers BADH2_CDS5-6, BADH2_CDS5, and BADH2_CDS6. FB and NFB were high-2AP bulk and low-2AP bulk, respectively

To confirm the association between *GmBADH2* gene and fragrance, QTL analysis was conducted using DNA markers on the linkage group LG A1 which *GmBADH2* harbored on. QTL analysis was performed by CIM using two data sets of fragrance evaluation, viz. 2AP concentration and organoleptic score. The significant LOD thresholds for the two traits calculated by permutation tests were 1.7 and 4.8, respectively. CIM consistently revealed that a major QTL for fragrance, designed as *qFgr* for fragrance score and *q2AP* for 2AP, was located on LG A1 with LOD score of 3.1 and 9.9, respectively (Fig. 3). The peak of QTL was highest at the location of markers BADH2-CDS6. *qFgr* explained 16.7% of the 2-AP variation and 42.9% of the fragrance score variation.

Sequencing of *GmBADH1* and *GmBADH2* genes

GmBADH1 and *GmBADH2* genes were sequenced to clarify mutation(s) that caused fragrance in both CM60 and Kaori. The sequencing was repeated several times to obtain precise and reliable data. The sequences of CM60 and Kaori were aligned against the reference sequence (Williams 82) and the sequences from two Japanese cultivars, "Nourin No. 2" (GenBank Acc. no. AK245037 for *GmBADH1*and AK244648 for *GmBADH2*) and "Bansei Shirodaizu" (AB333794 and AB333793), respectively. We found that *GmBADH1* comprises 15 exons with open reading frame of 1,512 nucleotides (Fig. S2) encoding 503

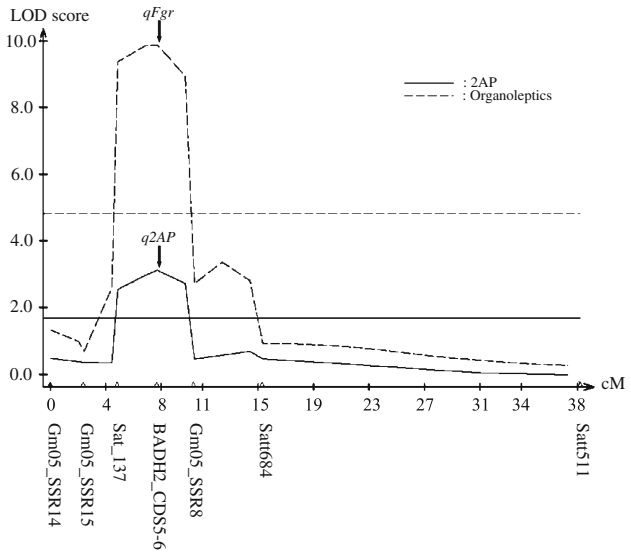

Fig. 3 Fragrance QTL (*qFgr*) position and LOD score plot for composite interval mapping on linkage group A1 in the RIL population of CM60 × Kaori. Position of the marker BADH2_CDS6 which developed from *GmBADH2* gene coincides with the maximum LOD score for fragrance QTL. *Solid* and *dashed lines* parallel to the linkage map represented LOD threshold of 1.7 and 4.8 for 2AP concentration and organoleptic fragrance score, respectively

amino acids (Fig. S3). The coding sequence (CDS) of *GmBADH2* gene which is in exon 12 of Williams 82 (at 2,841 bp of genomic sequence of Glyma05g01770) was 1 bp different from the two Japanese cultivars and CM60 and Kaori (data not shown). Consequently, the next intron position of Williams 82 and predicted amino acid were affected by this difference. Thus, we designed to compare *GmBADH2* sequences of CM60 and Kaori with those of Bansei Shirodaizu and Nourin No. 2. The comparison revealed that *GmBADH2* comprises 15 exons with coding sequence of 1,512 nucleotides (Fig. S4) encoding 503 amino acids (Fig. 4), similar to *GmBADH1*.

In *GmBADH1*, we found only one synonymous mutation caused by a SNP in CM60 that changes thymine (T) to cytosine (C) at the position 69 of the exon 1 (Fig. S2). In case of *GmBADH2*, we found four SNPs, three synonymous and one nonsynonymous mutations in Kaori. The synonymous mutations occurred at position 1,055, 1,154, and 1,160 of the complete coding sequence which is equivalent to nucleotide position 6, 105, and 111 of exon 6, respectively (Fig. 5). The single-base substitutions at these positions change T to C, A to T, and C to T in Kaori. The nonsynonymous mutation was at position 2,214 of the complete coding sequence which is equivalent to nucleotide position 92 of the exon 10 (Fig. 5). The single-point mutation of G to A in Kaori results in an amino acid change of glycine (GGC; G) to aspartic acid (GAC; D) at the position 334 of GmBADH2 protein (Fig. 4).

Development of SNAP markers for *GmBADH2*

To facilitate marker-assisted breeding of fragrant soybean, the SNP between Kaori and CM60 was used to develop

```
Kaori             MSIPIPHRQLFIDGDWKVPVLKNRIPIINPSTQHIIGDIPAATKEDVDLAVAAAKAALSR  60
CM60              MSIPIPHRQLFIDGDWKVPVLKNRIPIINPSTQHIIGDIPAATKEDVDLAVAAAKAALSR  60
Bansei Shirodaizu MSIPIPHRQLFIDGDWKVPVLKNRIPIINPSTQHIIGDIPAATKEDVDLAVAAAKAALSR  60
Nourin No. 2      MSIPIPHRQLFIDGDWKVPVLKNRIPIINPSTQHIIGDIPAATKEDVDLAVAAAKAALSR  60

Kaori             NKGADWASASGSVRARYLRAIAAKITEKKPELAKLEAIDCGKPLDEAAWDIDDVAGCFEF  120
CM60              NKGADWASASGSVRARYLRAIAAKITEKKPELAKLEAIDCGKPLDEAAWDIDDVAGCFEF  120
Bansei Shirodaizu NKGADWASASGSVRARYLRAIAAKITEKKPELAKLEAIDCGKPLDEAAWDIDDVAGCFEF  120
Nourin No. 2      NKGADWASASGSVRARYLRAIAAKITEKKPELAKLEAIDCGKPLDEAAWDIDDVAGCFEF  120

Kaori             YADLAEKLDAQQKAHVSLPMDTFKSYVLKEPIGVVALITPWNYPLLMATWKVAPALAAGC  180
CM60              YADLAEKLDAQQKAHVSLPMDTFKSYVLKEPIGVVALITPWNYPLLMATWKVAPALAAGC  180
Bansei Shirodaizu YADLAEKLDAQQKAHVSLPMDTFKSYVLKEPIGVVALITPWNYPLLMATWKVAPALAAGC  180
Nourin No. 2      YADLAEKLDAQQKAHVSLPMDTFKSYVLKEPIGVVALITPWNYPLLMATWKVAPALAAGC  180

Kaori             AAILKPSELASVTCLELAEICKEVGLPPGVLNILTGLGPEAGAPLAAHPDVDKIAFTGSS  240
CM60              AAILKPSELASVTCLELAEICKEVGLPPGVLNILTGLGPEAGAPLAAHPDVDKIAFTGSS  240
Bansei Shirodaizu AAILKPSELASVTCLELAEICKEVGLPPGVLNILTGLGPEAGAPLAAHPDVDKIAFTGSS  240
Nourin No. 2      AAILKPSELASVTCLELAEICKEVGLPPGVLNILTGLGPEAGAPLAAHPDVDKIAFTGSS  240

Kaori             ATGSKIMTAAAQLIKPVSLELGGKSPIIVFEDVDLDKAAEWTIFGCFWTNGQICSATSRL  300
CM60              ATGSKIMTAAAQLIKPVSLELGGKSPIIVFEDVDLDKAAEWTIFGCFWTNGQICSATSRL  300
Bansei Shirodaizu ATGSKIMTAAAQLIKPVSLELGGKSPIIVFEDVDLDKAAEWTIFGCFWTNGQICSATSRL  300
Nourin No. 2      ATGSKIMTAAAQLIKPVSLELGGKSPIIVFEDVDLDKAAEWTIFGCFWTNGQICSATSRL  300
                                                          *
Kaori             IVHESIATEFLNRIVKWVKNIKISDPLEEGCRLDPIVSEGQYEKILKFISNAKSEGATIL  360
CM60              IVHESIATEFLNRIVKWVKNIKISDPLEEGCRLGPIVSEGQYEKILKFISNAKSEGATIL  360
Bansei Shirodaizu IVHESIATEFLNRIVKWVKNIKISDPLEEGCRLGPIVSEGQYEKILKFISNAKSEGATIL  360
Nourin No. 2      IVHESIATEFLNRIVKWVKNIKISDPLEEGCRLGPIVSEGQYEKILKFISNAKSEGATIL  360

Kaori             TGGSRPEHLKKGFFVEPTVITDVTTSMQIWREEVFGPVLCVKTFSTEEEAIDLANDTVYG  420
CM60              TGGSRPEHLKKGFFVEPTVITDVTTSMQIWREEVFGPVLCVKTFSTEEEAIDLANDTVYG  420
Bansei Shirodaizu TGGSRPEHLKKGFFVEPTVITDVTTSMQIWREEVFGPVLCVKTFSTEEEAIDLANDTVYG  420
Nourin No. 2      TGGSRPEHLKKGFFVEPTVITDVTTSMQIWREEVFGPVLCVKTFSTEEEAIDLANDTVYG  420

Kaori             LGSAVISNDLERCERITKAFKAGIVWINCSQPCFTQAPWGGIKRSGFGRELGEWGLDNYL  480
CM60              LGSAVISNDLERCERITKAFKAGIVWINCSQPCFTQAPWGGIKRSGFGRELGEWGLDNYL  480
Bansei Shirodaizu LGSAVISNDLERCERITKAFKAGIVWINCSQPCFTQAPWGGIKRSGFGRELGEWGLDNYL  480
Nourin No. 2      LGSAVISNDLERCERITKAFKAGIVWINCSQPCFTQAPWGGIKRSGFGRELGEWGLDNYL  480

Kaori             SVKQVTQYISDEPWGWYQSPSRL  503
CM60              SVKQVTQYISDEPWGWYQSPSRL  503
Bansei Shirodaizu SVKQVTQYISDEPWGWYQSPSRL  503
Nourin No. 2      SVKQVTQYISDEPWGWYQSPSRL  503
```

Fig. 4 Alignment of the amino acid sequences of betaine aldehyde dehydrogenase 2 (BADH2) protein encoded by *GmBADH2* gene from non-fragrant soybean CM60, Bansei Shirodaizu, and Nourin No. 2, and fragrant soybean Kaori. *Underlined* are sequences that are believed to be essential for functional activity of the gene. Kaori possesses EGCRL**D**PIVS motif instead of EGCRL**G**PIVS

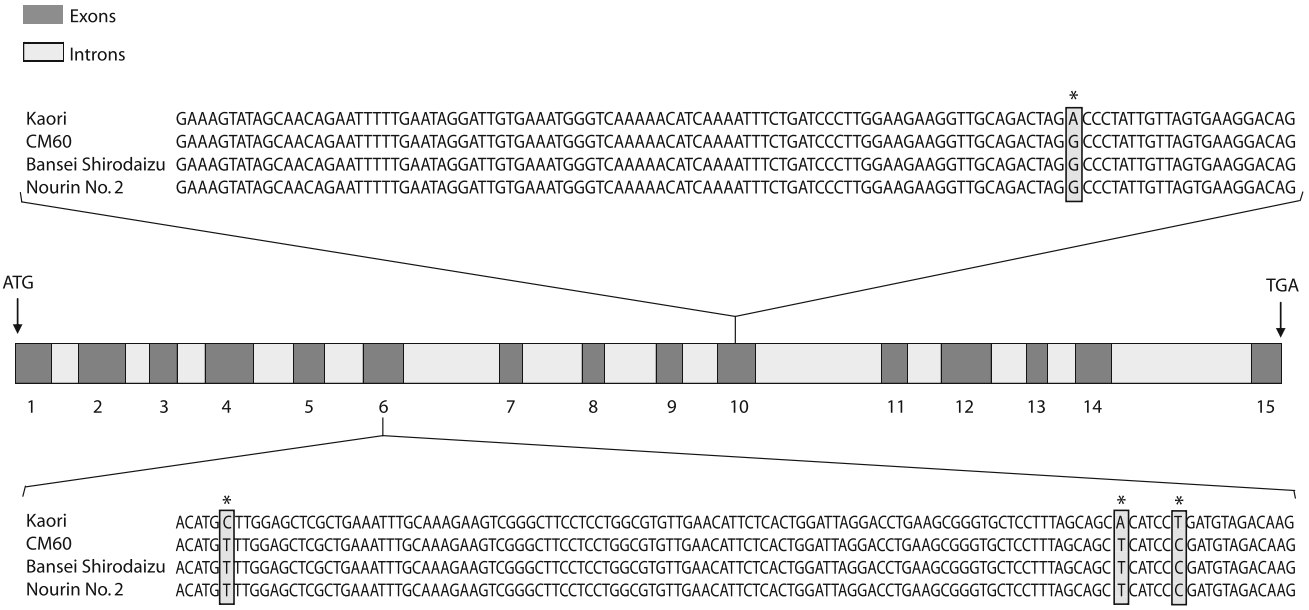

Fig. 5 Structure of *GmBADH2* gene showing start codon (ATG), 15 exons (*black boxes*), 14 introns (*gray boxes*) and stop codon (ACT). The nucleotide sequences of exon 6 and 10 are shown for non-fragrant soybean varieties: CM60, Bansei Shrirodaizu and Nourin No. 2, and fragrant soybean variety: Kaori. Kaori possesses three and one single nucleotide polymorphisms (SNPs) in exon 6 and 10, respectively, as indicated by *asterisks*

PCR-based allele specific SNP markers. In total, five primer pairs were designed for the G/A polymorphism of *GmBADH2* between Kaori and CM60 (Table 1). Three of them were specific to the G allele while two were specific to the A allele. Both primer pairs specific to the A allele generated a band only in Kaori, while all three pairs specific to the G allele generated a DNA band only in CM60 (Fig. 6). When a polymorphic SNAP marker was analyzed in the RIL population, we found that the marker co-segregated perfectly with the markers BADH2_CDS5-6 and BADH2_CDS6 (data not shown).

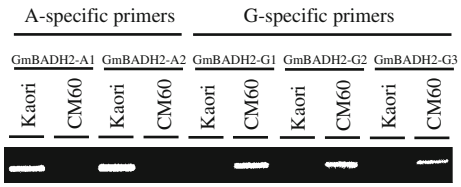

Fig. 6 Patterns of DNA bands amplified by single nucleotide-amplified polymorphism (SNAP) primer pairs in CM60 and Kaori. Primer pairs specific to the A and G alleles of *GmBADH2* were used

Discussion

Fragrance in food crops as a result from 2AP synthesis is intensively studied in rice, a model food crop. The molecular genetic basis of fragrance in rice is due to an 8-bp deletion in exon 7 (Bradbury et al. 2005; Wanchana 2005) and a 7-bp deletion in exon 2 (Shi et al. 2008) of the *fgr* locus which is equivalent to betain aldehyde dehydrogenase (*OsBADH2*) on chromosome 8 (Bradbury et al. 2005). The deletions lead to the introduction of premature stop codon resulting in a truncated OsBADH2 protein and the accumulation of 2AP. In this study, we used gene mapping and candidate gene approaches to revealed genetic nature of fragrance in vegetable soybean. QTL mapping showed that the *GmBADH2* associates with the fragrance. Sequencing of the *GmBADH2* showed three synonymous mutations in exon 6 and a nonsynonymous mutation in exon 10. Thus, the nonsynonymous mutation in exon 10 possibly results in fragrance in Kaori. The band size difference detected by markers BADH2_CDS5-6 and BADH2_CDS6 may arise from the secondary structure of DNA that three SNPs exist in exon 6.

The genetic studies in rice have shown that the fragrance is due to the loss in function of *OsBADH2* (Bradbury et al. 2005; Niu et al. 2008; Shi et al. 2008; Vanavichit et al. 2008). The 8-bp deletion was first identified as the functional mutation generating recessive allele (Bradbury et al. 2005; Wanchana 2005). Chen et al. (2008) showed that only intact 503 aa protein encoded from full-length transcript of *OsBADH2* could inhibit 2AP synthesis. Although we did not find any mutation in *GmBADH2* of Kaori which would cause premature transcript termination, the nonsynonymous mutation in the exon 10 of *GmBADH2* is potentially responsible for the enhanced 2AP (fragrance) in Kaori. Bradbury et al. (2005) showed that functional

OsBADHs protein contains two peptide sequences—VSLELGGKSP and EGCRLGSVVS—and a cysteine residue (28 amino acid residues away from VSLELGGKSP) in exons 8, 10, and 9, respectively, are highly conserved in aldehyde dehydrogenases. These sequences are believed to be essential for functional activity of *OsBADHs* (Bradbury et al. 2005). In the *OsBADH2*, exons 8, 10, and 9 also contain coding regions for those elements, respectively (Bradbury et al. 2005). Interestingly, the exon 10 of *GmBADH2* of the non-fragrant soybeans contains the conserved motif EGCRLGPIVS similar to the motif for the essential functional activity of *OsBADHs*. However, in the fragrant soybean Kaori, the transition of G to A causes a change of the EGCRL**G**PIVS motif to EGCRL**D**PIVS motif (Fig. 4), suggesting that the GmBADH2 in Koari may lose its functional activity because of this mutation. Recently, eight additional non-functional alleles of the *OsBADH2* associating with fragrance in several rice accessions have been identified (Kovach et al. 2009). Four of them are frameshift-inducing indels and one is SNP creating premature stop codon. The other three potentially functional mutations include two SNPs in coding sequence and a 3-bp insertion. Despite the two SNPs result in amino acid substitutions in *OsBADH2* (*badh2.9* and *badh2.10* alleles in their paper) causing no truncation of protein, these coding mutations associated with the fragrance (Kovach et al. 2009). Mutation at the same gene which generates fragrance in rice and soybean suggested that both plant species have similar biochemical pathways for 2AP synthesis. More research is required to clarify the effect of the amino acid substitution in Kaori. It is worth noting that very recently Arikit et al. (2010) and Juwattanasomran et al. (2010) demonstrated that a 2-bp deletion in exon 10 of *GmBADH2* associated with fragrance in soybean. Arikit et al. (2010) also showed that silencing *GmBADH2* gene (equivalent to *GmAMADH2* in the original paper) by means of RNA interference (RNAi) technique results in 2AP biosynthesis in non-fragrant soybean varieties. This supports our finding that mutation in *GmBADH2* associated with fragrance in Kaori.

Although the major gene *GmBADH2* causally involved in fragrance in soybean has been located and identified, the effect of this gene is not highly potent because the phenotypic variance explained for the 2AP concentration and fragrance score was only 16.7 and 42.9%, respectively. Examination of the traits and alleles of the *GmBADH2* gene as genotyped by a SNAP marker of the RILs revealed that all the RILs with high fragrance (fragrance score 3) had *GmBADH2* allele from Kaori and 71.4% of the RILs with mild to moderate fragrance had the Kaori allele (data not shown). However, 15.6% of the RILs with no fragrance also had the Kaori allele. Similarly, all RILs having 2AP concentration comparable to Kaori or higher (0.5170–0.9231) possessed the Kaori allele, while 41.1% of the RILs having 2AP concentration comparable to CM60 or lower (0–0.8211) also possessed the Kaori allele. This suggested that there is possibly another gene(s) conditioning fragrance in soybean. In rice, besides the *OsBADH2* gene on chromosome 8 that strongly affects strength of fragrance, Lorieux et al. (1996) reported two minor QTLs each on chromosome 4 and chromosome 12 also contribute to fragrance in rice, while Amarawathi et al. (2008) identified two minor QTLs each on chromosome 3 and chromosome 4 influence level of the fragrance. Searching rice genome database revealed that *OsBADH1* was likely a candidate gene for QTL for fragrance on rice chromosome 4 (Amarawathi et al. 2008). Therefore, mutation that results in null function of *OsBADH1* or reduction in expression levels of *OsBADH1* may also cause an increase in 2AP accumulation. It is worth noting that Kovach et al. (2009) found two fragrant rice accessions without any mutation in the coding or promoter regions that may alter *OsBADH2* or its expression. In our study, sequencing of the *GmBADH1* revealed no polymorphism, suggesting that this gene has no association with the fragrance in Kaori. In rice, Chen et al. (2008) suggested that the non-functional OsBADH2 protein results in 4-aminobutyraldehyde (AB-ald) accumulation leading to the formation of 2AP in fragrant rice, whereas Bradbury et al. (2008) suggested that a non-functional OsBADH2 protein causing accumulation and spontaneous cyclisation of γ-aminobutyraldehyde (GABald) into Δ^1-pyrroline is a likely cause of 2AP accumulation in rice. Therefore, gene controlling AB-ald or GABald may also involve in 2AP formation in soybean. Without the role of OsBADH, Huang et al. (2008) showed that gene for Δ^1-pyrroline-5-carboxylic acid synthetase (P5CS) may involve in accumulation of 2AP in fragrance rice. P5CS may react directly with methylglyoxal (MG) and lead to 2AP formation in the fragrant rice (Huang et al. 2008). In soybean, on the other hand, P5CS may degrade to 1-pyrroline and interact with MG to produce 2AP (Wu et al. 2009). Therefore, genes regulating these chemicals may also affect fragrance level in soybean.

Breeding for fragrance in soybean and rice by phenotypic selection is difficult. Although fragrance evaluation can be achieved by organoleptic test or GC–MS analysis of 2AP, organoleptic test is subjected to human error, while GC–MS technique is expensive and requires expertise to analyze. Recessive inheritance of the trait also requires repeated evaluation using self-pollination generations during backcross breeding. In this study, a SNP in *GmBADH2* associated with fragrance (2AP) production enabled us to develop a PCR-based SNAP marker that detects polymorphism within the *GmBADH2*. SNAP has several advantages over other marker systems for SNP detection because it is simple and highly reproducible at low cost. The SNAP markers

developed in this study derived from G/A change in *GmBADH2* gene, which causes the trait polymorphism. These markers can facilitate marker-assisted selection for fragrant soybean genotypes in a large-scale breeding program using Kaori as the donor of fragrance gene.

Acknowledgments This research was supported by the Thailand Research Fund (TRF). R. Juwattanasomran is thankful to the Royal Golden Jubilee Ph.D. (RGJ-Ph.D.) program of the TRF for awarding the scholarship. We are thankful to the Center for Agricultural Biotechnology, Kasetsart University for molecular lab facilities and to the Center of Excellence for Innovation in Chemistry (PERCH-CIC), the Commission on Higher Education, Ministry of Education for GC-NPD instrument used for 2AP analysis. We also thank Warunee Sommanas and Worapa Seehalak for their technical supports.

References

Ahn SN, Bollich CN, Tanksley SD (1992) RFLP tagging of a gene for aroma in rice. Theor Appl Genet 84:825–828

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

Arai Y, Hayashi M, Nishimura M (2008) Proteomic analysis of highly purified peroxisomes from etiolated soybean cotyledons. Plant Cell Physiol 49:526–539

Arikit S, Yoshihashi T, Wanchana S, Uyen TT, Huong NTT, Wongpornchai S, Vanavichit (2010) Deficiency in the amino aldehyde dehydrogenase encoded by *GmAMADH*2, the homologue of rice *Os2AP*, enhances 2-acetyl-1-pyrroline biosynthesis in soybeans (*Glycine max* L.). Plant Biotechnol J. doi: 10.1111/j.1467-7652.2010.00533.x

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

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

Bradbury LTM, Gillies SA, Brushett DJ, Waters DLE, Henry RJ (2008) Inactivation of an aminoaldehyde dehydrogenase is responsible for fragrance in rice. Plant Mol Biol 68:443–9449

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

Drenkard E, Richter BG, Rozen S, Stutius LM, Angell NA, Mindrinos M, Cho RJ, Oefner PJ, Davis RW, Ausubel FM (2000) A simple procedure for the analysis of single nucleotide polymorphisms facilitates map-based cloning in Arabidopsis. Plant Physiol 124:1483–1492

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

Godhi MA, Ye G-N, 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

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

Jin QS, Waters D, Cordeiro GM, Henry RJ, Reinke RF (2003) A single nucleotide polymorphism (SNP) marker linked to the fragrance gene in rice (*Oryza sativa* L.). Plant Sci 165:359–364

Juwattanasomran R, Somta P, Kaga A, Chankaew S, Shimizu T, Sorajjapinun W, Srinives P (2010) Identification of a new fragrance allele in soybean and development of its functional markers. Mol Breed. doi:10.1007/s11032-010-9523-0

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

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

Masuda R (1991) Quality requirement and improvement of vegetable soybean. In: Shanmugasundaram S (ed) Vegetable soybean: research needs for production and quality improvement. Asian Vegetable Research and Development Center, Taiwan, pp 92–102

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

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

Pinson SRM (1994) Inheritance of aroma in six rice cultivars. Crop Sci 34:1151–1157

Shanmugasundaram S, Cheng ST, Huang MT, Yan MR (1991) Quality requirement and improvement of vegetable soybean. In: Shanmugasundaram S (ed) Vegetable soybean: research needs for production and quality improvement. Asian Vegetable Research and Development Center, Taiwan, pp 30–42

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

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 BC, 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

Van Ooijen JW, Voorrips RE (2001) JoinMap® 3.0: software for the calculation of genetic linkage maps. Plant Research International, Wageningen

Vanavichit A, Tragoonrung S, Toojinda T, Wanchana S, Kamolsukyunyong W (2008) Transgenic rice plants with reduced expression of *Os2AP* and elevated levels of 2-acetyl-1-pyrroline. US Patent 731918

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

Wang S, Basten CJ, Zeng ZB (2007) Windows QTL Cartographer 2.5. Department of Statistics. North Carolina State University, Raleigh

Wu ML, Chou KL, Wu CR, Chen JK, Huang TC (2009) Characterization and the possible formation mechanism of 2-acetyl-1-pyrroline in aromatic vegetable soybean. J Food Sci 74:192–197