CHAPTER VII

IDENTIFICATION OF A NOVEL ALLELE OF *Badh 2* IN LAO TRADITIONAL VARIETY KAI NOY LEUANG

7.1 Introduction

Aroma in rice is a highly prized trait in rice. As previously mentioned, it arises from one of several mutations in the the *betaine aldehyde dehydrogenase 2* (*BADH2*) gene. These mutations lead to a truncated, non-functional BADH2 protein (Bradbury et al., 2008; Bradbury et al., 2005a; Kovach et al., 2009).

Most aromatic varieties carry a 7 bp deletion in the *BADH2* gene (Fitzgerald et al., 2008), but recent research shows that this gene has mutated a number of times, and that these mutations associate with geographical origin (Kovach et al., 2009). The fragrance gene appears to have originated in *japonica*, and been introgressed into *indica* during domestication (Kovach et al., 2009). Interestingly, all those varieties found to contain uncommon alleles of the *BADH2* gene are *japonica*, suggesting that the introgression into *indica* is relatively recent.

One of the varieties used in the present study, KNL, is a *tropical japonica* landrace. It is a highly aromatic rice, and does not carry the common mutation in *BADH2* (Bounphanousay et al., 2009; Fitzgerald et al., 2008b; Kovach et al., 2009). KNL grains contain high levels of 2AP, and this was confirmed by mass spectroscopy (Fitzgerald et al., 2008). Sequencing of the coding and promoter regions of the BADH2 gene of KNL revealed no mutation (Kovach et al., 2009), suggesting that this variety could carry a new gene for aroma.

The objective of this chapter is to use different populations, genome-wide genotyping, and appropriate phenotyping to determine the genetic basis of aroma in KNL.

7.2 Materials and methods

7.2.1 Development of introgression lines for mapping

The introgression lines were developed from a cross between Thasano 1 (TSN 1), a high-yielding but non-aromatic variety, and KNL, an aromatic variety. Backcrossing was performed to transfer the aroma alleles from KNL to TSN 1, with KNL as the donor parent. At each subsequent generation, progeny with the TSN 1 plant type that were fragrant were selected for the next generation of backcrossing. Fragrant lines were selected by sensory evaluation (Juliano, 1985). Those found to be fragrant were back-crossed to TSN 1 until BC_4F_1 and then allowed to self. The BC_4F_2 were phenotyped for 2AP content and used for SNP genotyping. The crosses were made in the wet season of 2006 at the Rice and Cash Crop Research Center (RCCRC) in Vientiane, Lao PDR.

A second population was developed using KNL and another aromatic traditional variety, HNN, that carries the common mutation in BADH2. The cross was made in the rainy season of 2006 at RCCRC, Vientiane, Lao PDR and progeny were grown until F_3. F_4 seeds were sent to IRRI and were advanced to F_5 without selection.

A third population was created between TSN 1 and KNL and grown to F2. This population was genotyped and phenotyped to validate the discovery of the genetics of aroma in KNL.

7.2.2 Quantification of 2-AP by GC-MS

2AP was extracted either from grains or fresh leaves. 2-AP content was quantified in grains exactly as described on (Fitzgerald et al., 2008a). For leaves, 2AP was extracted with dichloromethane using a method described previously (Bergman et al., 2000) with slight modifications. One gram of fresh leaves in distilled water was heated for 30 mins at 85°C. The leaves were discarded and dichloromethane and saturated sodium chloride were added to the filtrate. The mixture was shaken for 1 min. The organic layer was filtered over anhydrous sodium sulfate, concentrated over stream of nitrogen and immediately injected into Agilent 6890N gas chromatography-mass spectrometry. 2AP was extracted from grains of freshly harvested KNL, HNN and progeny of each population exactly as previously described. KNL and HNN were also stored for two years, and 2AP quantified at various times during storage.

2AP from leaves and grain was injected into a HP-5MS column (length, 30 m; internal diameter, 0.25mm; thickness, 0.25um). Samples were injected in splitless mode (2uL) with a starting temperature of 35°C for 1.8min, which was then increased at 5°C/min to 70°C, then at 10°C/min to 100°C, and then at 25°C/min to 270°C. Temperature was then held at 270°C for 2 min. The temperatures of injector, detector, interface, and ion source were 155°C, 300°C, 280°C and 230°C, respectively. Helium was the carrier gas at 1.0mL/min. Mass spectra were acquired over 50-550 amu at 2 scans per second with an ionising electron energy of 70 eV. The peaks were identified using NIST2005.

7.2.3 DNA extraction and genotyping

The BC_4F_2 lines were selected for genome wide genotyping. DNA extraction was done using Qiagen DNEasy Plant Mini Extraction Kit (Qiagen, Valencia, CA, USA) following the manufacturer's protocol. Prior to genotyping, the DNA was quantified to 50ng/ul using a Nanodrop ND-100 Spectrophotometer (Willmington, DE, USA). Genotyping was done with Illumina Beadxpress Golden Gate Genotyping Assay (San Diego, CA, USA) at 384 SNP loci specific to show polymorphisms in populations derived from indica and japonica parents (Thomson et al., In preparation). The raw data from Illumina Beadxpress was analysed with Alchemy (Wright et al., 2010). SNP maps were created using GGT 2.0 (van Berloo, 2008).

7.2.4 Sequencing of *BADH2*

In order to determine whether KNL carries a novel allele of BADH2, or another fragrance gene, previous sequencing gaps from (Kovach et al., 2009) were filled. Primers that amplify regions spanning these gaps (32bp to 1kb) and 1kb upstream and downstream *BADH2* were designed using Primer 3 version 4.0 (Table 1). Purified PCR products were sequenced on ABI Prism 3700/3100 DNA analyzers (Applied Biosystems, Carlsbard, CA, USA) at the Cornell Life Sciences Core Laboratories Center. Sequences were aligned using CodonCode Aligner software with TSN 1 as the reference sequence.

7.3 Results

7.3.1 Characterisation of KNL

GC-MS results show that KNL contained a higher concentration of 2AP than HNN, 0.3 and 0.2ug/g respectively (Figure 7.1A). Furthermore, a significant decrease in 2AP content is observed in HNN upon storage for 3 to 22 months (Figure 7.1B). In KNL, a decrease in the concentration of 2AP is also observed from 3 months to 7 months storage but the decrease was slow between 7 and 22 months. The decline of 2AP concentration during storage in KNL is not statistically significant (5% LSD 0.07) compared to that of HNN (5% LSD 0.07).

7.3.2 Gene mapping

Two hundred and twenty SNPs found to be polymorphic between KNL and TSN 1 were used for mapping fragrance in the BC_4F_2 population. A major QTL was found on chromosome 8, near to *BADH2*. Figure 7.2 shows the introgressed segment from KNL in the fragrant lines. The presence of 2AP in the leaves was determined for each of the lines, and most of those with KNL introgression near BADH2 were aromatic. For this population, no other candidate regions were detected that associated with the presence of 2AP in leaves.

7.3.3 Sequencing of *BADH2*

In order to prove the hypothesis of a novel fragrance allele, KNL, another aromatic variety, KTL, with no known mutation in *BADH2*, and TSN 1 were sequenced. PCR amplification of exons 3 to 5 of *BADH2* showed a large polymorphism between the three varieties detectable even in a 1.2% agarose gel (Figure 7.3). Sequencing results of the region shown in Figure 7.4A confirms an 806bp deletion in KNL and KTL that spans exon 4 to 5, including 86bp of exon 4, the whole of intron 4 (697bp) and 23bp of exon 5. Translation of the KNL sequence using the ExPaSy Translate tool (http://www.expasy.ch/tools/dna.html) shows that the alteration of the coding region causes a reading frame shift resulting in a premature stop codon. The translated protein of varieties carrying this mutation is only 135 amino acids long, 363 shorter than the functional protein (Figure 7.4B). Henceforth, we will refer to the new allele as badh2.11. We were not able to detect any additional alteration in the other coding regions as well as the upstream and downstream region of *BADH2* sequenced in this study.
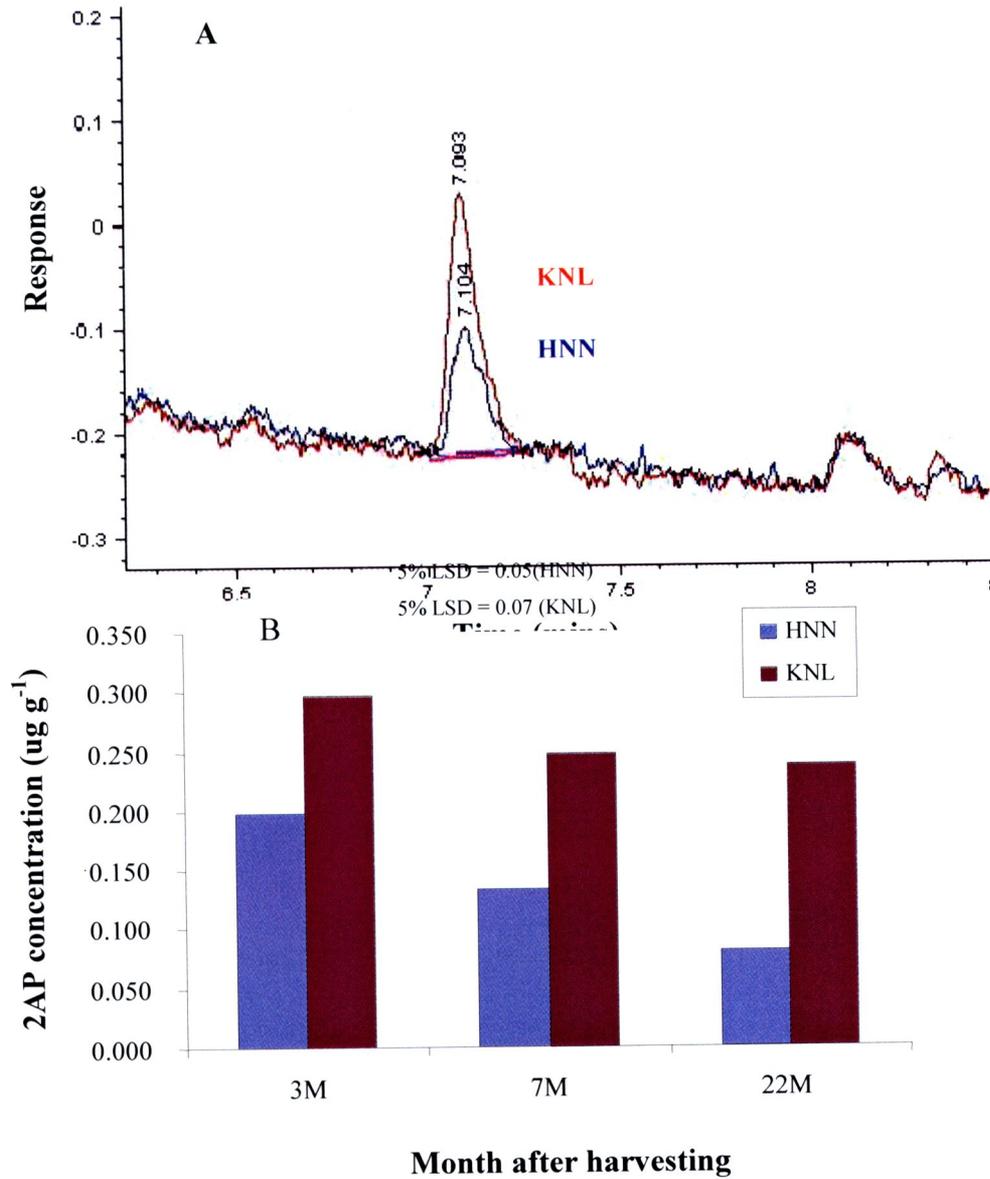
Figure 7.1 A: 2AP content of KNL vs other Laos traditional fragrant variety, HNN. **7.1 B:** 2AP content of KNL and HNN upon storage. Decrease over time is not significant KNL, but is for HNN LSD 5%.
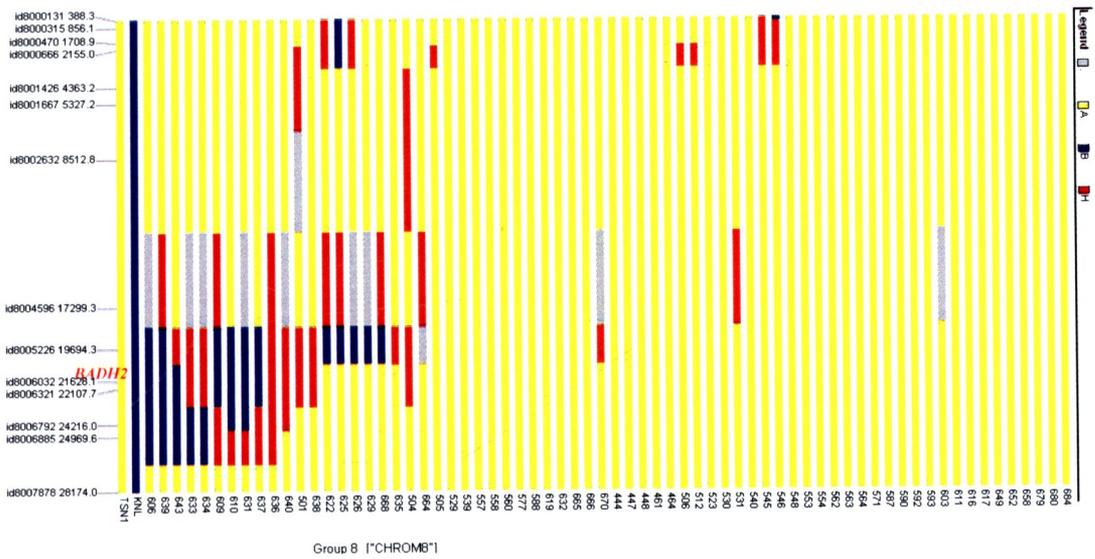
Figure 7.2 KNL introgression in chromosome 8 of fragrant lines. Yellow segments are from TSN 1, blue is from KNL, red are heterozygotes and gray means no data.
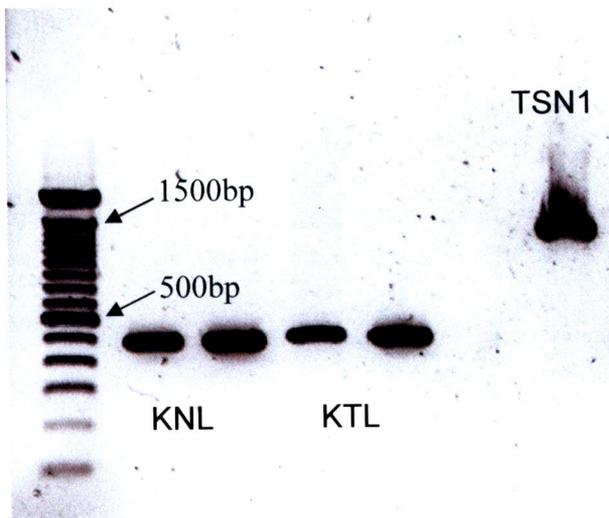

Figure 7.3 PCR product of *BADH2* exons 3 to 5 in KNL, TSN 1 and KTL showing huge difference in the size of the bands.

```
KNL   CAGGACGATGTTGCTGGATGCTTTGAGTACTTTGCAGATCTTGCAGAATCCTTGGACAAA
KTL   CAGGACGATGTTGCTGGATGCTTTGAGTACTTTGCAGATCTTGCAGAATCCTTGGACAAA
TSN1  CAGGACGATGTTGCTGGATGCTTTGAGTACTTTGCAGATCTTGCAGAATCCTTGGACAAA

KNL   AG----------------------------------------------------------
KTL   AG----------------------------------------------------------
TSN1  AGGCAAAATGCACCTGTCTCTCTTCCAATGGAAAACTTTAAATGCTATCTTCGGAAAGAG

KNL   ------------------------------------------------------------
KTL   ------------------------------------------------------------
TSN1  CCTATCGGTGTAGTTGGGTTGATCACACCTTGGTATTTCACATTTTTCTCTCATCCTGCG

KNL   ------------------------------------------------------------
KTL   ------------------------------------------------------------
TSN1  CTTATATTTATTTATGACCCAAGCATGGTACTAAATAGTACTAGTAACATGCATATACTG

KNL   ------------------------------------------------------------
KTL   ------------------------------------------------------------
TSN1  AATGAGTTTACAACTTTACATGATTTTTTTGAACTATGA......TTTGTTGTTTCTTTG

KNL   ------------------------------------------------------------
KTL   ------------------------------------------------------------
TSN1  TTATGTTCTTATCGACAATTACAAATTTGATTCTGAGAATCATGTTCGGGATGTGTATTT

KNL   -------------------------------GCAACATGGAAGGTAGCTCCTGCCCTGGCTGC
KTL   -------------------------------GCAACATGGAAGGTAGCTCCTGCCCTGGCTGC
TSN1  CTACTGCAGGAACTATCCTCTCCTGATGGCAACATGGAAGGTAGCTCCTGCCCTGGCTGC

KNL   TGGCTGTACAGCTGTACTAAAACCATCTGAATTGGCTTCCGTGTAAGTTTAACATGTTAA
KNL   TGGCTGTACAGCTGTACTAAAACCATCTGAATTGGCTTCCGTGTAAGTTTAACATGTTAA
TSN1  TGGCTGTACAGCTGTACTAAAACCATCTGAATTGGCTTCCGTGTAAGTTTAACATGTTAA

Exons
```

Figure 7.4A Sequence alignments of (A) exons 4 to 5 of KNL and KTL showing the new *BADH2* allele with TSN 1 as the reference sequence

```
KNL/1-135     1 MATAIPQRQLFVAGEWRAPALGRRLPVVNPATESPIGEIPAGTAEDVDAAVAAAREALKRNRGRDWARAPG71
TSN1/1-503    1 MATAIPQRQLFVAGEWRAPALGRRLPVVNPATESPIGEIPAGTAEDVDAAVAAAREALKRNRGRDWARAPG71

KNL/1-135    72 AVRAKYLRAIAAKIIERKSELARLETLDCGKPLDEAAWDMDDVAGCFEYFADLAESLDKRQHGR------135
TSN1/1-503   72 AVRAKYLRAIAAKIIERKSELARLETLDCGKPLDEAAWDMDDVAGCFEYFADLAESLDKRQNAPVSLPMEN142

KNL/1-135       -----------------------------------------------------------------
TSN1/1-503  143 FKCYLRKEPIGVVGLITPWNYPLLMATWKVAPALAAGCTAVLKPSELASVTCLELADVCKEVGLPSGVLNI213

KNL/1-135       -----------------------------------------------------------------
TSN1/1-503  214 VTGLGSEAGAPLSSHPGVDKVAFTGSYETGKKIMASAAPMVKPVSLELGGKSPIVVFDDVDVEKAVEWTLF284

KNL/1-135       -----------------------------------------------------------------
TSN1/1-503  285 GCFWTNGQICSATSRLILHKKIAKEFQERMVAWAKNIKVSDPLEEGCRLGPVVSEGQYEKIKQFVSTAKSQ355

KNL/1-135       -----------------------------------------------------------------
TSN1/1-503  356 GATILTGGVRPKHLEKGFYIEPTIITDVDTSMQIWREEVFGPVLCVKEFSTEEEAIELANDTHYGLAGAVL426

KNL/1-135       -----------------------------------------------------------------
TSN1/1-503  427 SGDRERCQRLTEEIDAGIIWVNCSQPCFCQAPWGGNKRSGFGRELGEGGIDNYLSVKQVTEYASDEPWGWY497

KNL/1-135       ------
TSN1/1-503  498 KSPSKL                                                          503
```

Figure 7.4B Translated peptide sequence showing the premature stop codon that renders *BADH2* in KNL nonfunctional.

7.3.4 Development of functional markers

To develop a cost-efficient assay that can differentiate accessions carrying the mutation, a primer pair flanking the deletion was designed (Table 7.1). Primer Badh2.11F and Badh2.11R anneal to sequences external to the area where the mutation occurs. Badh2.11R also pairs with Badh2.11ASPF, a primer specific to the sequence internal to the deletion producing a 344bp segment in non-fragrant varieties (Figure 7.5).
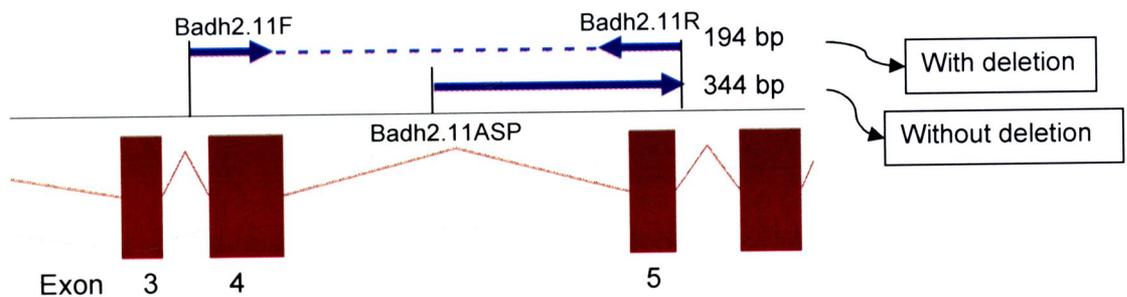

Figure 7.5 Diagram of the relative positions of the designed primers for the novel fragrance allele.

Using these primers in one PCR reaction gives rise to three possible results. First, a single 194bp band in homozygous fragrant accessions; in the second case, a single 344bp band present in homozygous non-fragrant accessions; and the third case, both 194 and 344bp bands are produced indicating heterozygous individuals (Figure 7.6).
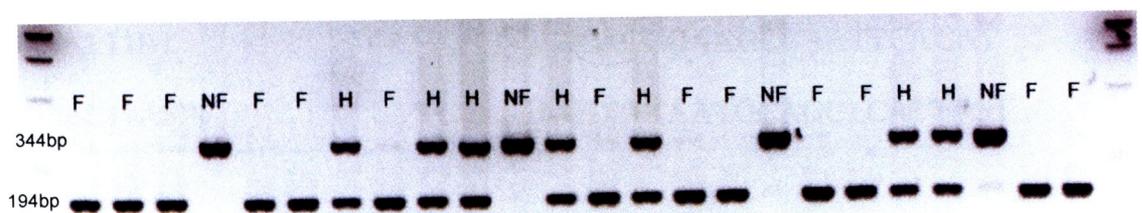

Figure 7.6 PCR banding patterns of the allele-specific markers in segregating F_2 population. F - fragrant; NF - non-fragrant; and H - heterozygous for the mutation.

Table 7.1 Primers used in sequencing *BADH2* gene

Primer Name	Forward Sequence (5' – 3')	Reverse Sequence (5' – 3')	Included region	Product size (bp)
BADH2_Exon1	CGAAGTCCGTACCAACTGC	AGTGTATATGGCTCACGGCC	Downstream	1075
BADH2_Exon3	GATTGTGGGAAGCCTCTTGA	ATTAGGTTCTGAAGCCGGTG	Exon 3 to 5	1296
BADH2_Exon 7	CAGCTGCTCCTATGGTTAAGG	TTCCTAACTGCCTTCCTTGC	Exon 8	702
BADH2_Exon9	GCAAGGAAGGCAGTTAGGAA	AACTTTCATGGTCAGGAGCAA	Exon 9	713
BADH2_Exon10	TTGAGTGGACTCTCTTTGGTTG	GCATACACCGACTAATTCATTCA	Exon 10	685
BADH2_Exon14	CCGGAATTATCTGGGTGAAC	TAGAAGCGTGGCACGTAATC	Exon 15	693
BADH2_Pro1	CCTTATTGCAATGATGGTCA	TATTCAGTTGCTGTGGTGATTT	Promoter	580
BADH2_Pro2	GACGTCGGGAGTTTGACTG	GCAGTTGGTACGGACTTCG	Promoter	908

Table 7.2 Number of individuals showing aromatic (+2AP) and nonaromatic (-2AP) phenotypes in the F_2 population and the average amount of 2AP.

Gene	(+) 2AP	2AP (ug/g)	(-) 2AP
badh2.11/badh2.11	60	0.10	4
Badh2.11/badh2.11	29	0.07	94
Badh2.11/Badh2.11	0	-	51

Table 7.3 Allele specific primers designed to amplify the novel *BADH2* allele.

Primer name	Sequence (5'-3')
Badh2.11F1	TGCCTACACAGCAATCTTTCC
Badh2.11R1	ACACGGAAGCCAATTCAGAT
Badh2.11ASPF	CGTGTTAATGCAGCTCATTTGT

7.3.5 Badh2.11 assay in segregating F_2 population

The designed allele specific primers were used to screen F_2 progenies of the cross between TSN 1 and KNL. Of the 300 individuals, 77 were found to be homozygous for badh2.11, 68 were homozygous non-fragrant, while 153 are heterozygous. The primers were also used to screen the F_5 population derived from

KNL and HNN. One hundred lines carrying the HNN allele (*badh*2.1) and a hundred with the KNL allele (*badh*2.11) were selected and grown. 2AP was quantified in grains of the two hundred lines. Figure 7.7 shows that those lines carrying *badh*2.11 have more 2AP in grains than the hundred lines carrying the *badh*2.1 allele. However, the amount of 2AP in outliers is not different between alleles.
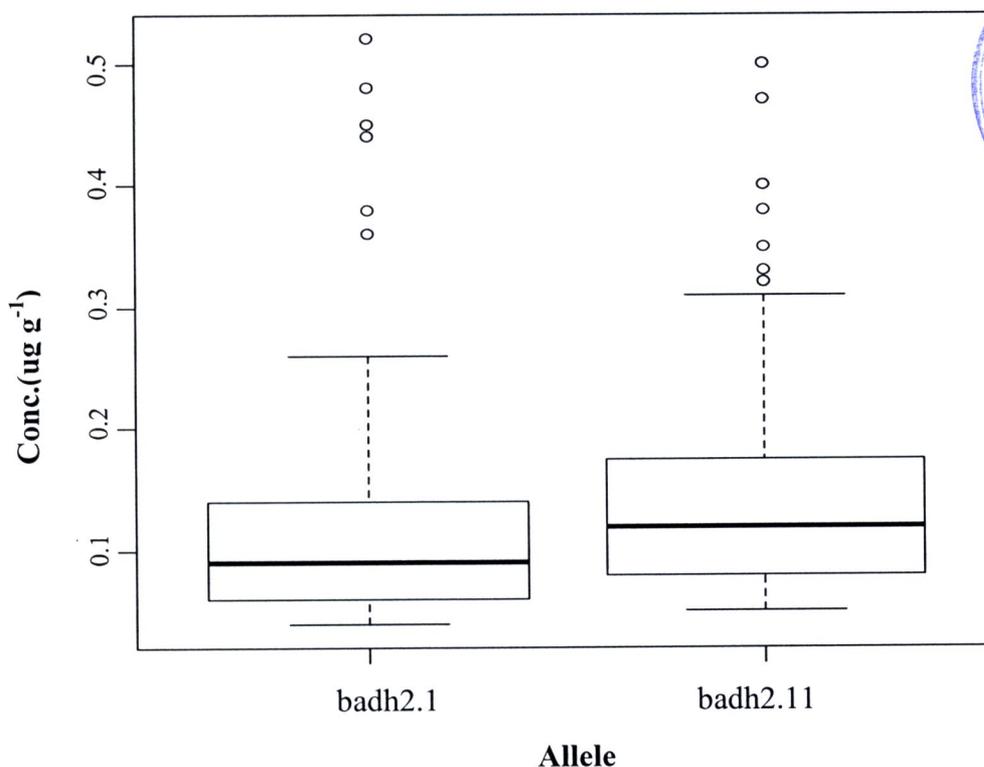

Figure 7.7 Box and whiskers plot of 2AP values in grains of progeny carrying the badh2.1 and 2.11 allele.

7.4 Discussion

KNL is a valuable donor parent of fragrance in rice. Not only that it contains high amount of 2AP compared to other Lao varieties, the 2AP content does not decrease significantly upon storage (Figure 7.1 A and B). Understanding the cause of fragrance in KNL is thus important.

BADH2 is the key enzyme for fragrance in rice. Loss of function of BADH2 favours the production of 2AP via acetylation of Δ1-pyrroline (Bradbury et al., 2008). We report here an 806bp deletion in the *BADH2* gene spanning across exons 4 to 5,

which induces a premature stop codon rendering the BADH2 enzyme nonfunctional. This is by far, the largest alteration of *BADH2* reported. It could also be considered as one of the rarest alleles of *BADH2* since among the 300 and 100 aromatic accessions previously screened (Fitzgerald et al., 2008; Kovach et al., 2009), only 2 varieties have been found, KNL and KTL, that carry this allele.

Association studies done in a segregating F_2 population shows a co-dominant monogenic inheritance consistent with previous reports (Bradbury et al., 2005a; Shi, 2008). In all lines homozygous for the *badh2.11* allele, 2AP could not be detected in four of them (Table 7.2). This may be explained by the effects of environmental and postharvest conditions which could influence the amount of 2AP, or the amount of 2AP was below the threshold limit detectable by gas chromatography. *BADH2* is a recessive gene (Bradbury et al., 2005a), implying that heterozygote individuals should not contain 2AP. However, in our case, phenotyping of the F_2s were done in the F_3 seeds which may have segregated into pure fragrant lines. Thus, 2AP was also detected in some that were heterozygote at F_2, but in a significantly lower amount (Table 7.2).

One of the SNP loci is just 300 kb upstream of the *BADH2* locus (Thomson et al., In preparation). Fragrance arises in *indica* rices from an introgression of a region of *japonica* genome that includes the *badh2* allele and flanking regions of ~650 kb upstream and 330 bp downstream of it (Kovach et al., 2009). Therefore the presence of the KNL allele at ~300 kb upstream of the *BADH2* locus is a strong indicator of introgression of the *badh2* allele. However, a small number of the lines showed the introgression but did not have the allele or 2AP indicating that the linkage between the SNP locus and the BADH2 locus could be broken.

Using a population derived from HNN (badh2.1) and KNL (badh2.11), progeny show that badh2.11, leads to significantly higher 2AP in grains than is generally detected in grains of varieties with badh2.1 (Figure 7.7). The outliers in that figure also suggest that other genes could also affect the amount of 2AP that accumulates. This suggests that the large deletion could lead to a larger phenotype than the smaller deletion. Kovach et al., 2009 show that the different alleles of BADH2 lead to different amounts of 2AP, and the alleles defined by a SNP have the least 2AP. This could suggest that the SNPs and small insertions or deletions

(Kovach et al., 2009) do not fully disable transcription, leading to some enzyme activity and oxidation of 4-aminobutanal.

The quality of 2AP from badh2.11 has been shown to be higher than that of other alleles. It does not change over storage and the amount is greater. Therefore, the discovery of this allele could be very important for rice improvement programs, and the development of indica and japonica germplasm carrying this allele, as well as primers, is a valuable output of this study.