# CHAPTER VI
# FRAGRANCE PATHWAY HOM NANG NOUANE LINES

## 6.1 Introduction

Aromatic rice commands a price premium in all rice markets because the fragrance trait is prized by many rice consumers. The pleasant aroma in rice has been mainly attributed to the presence of the volatile compound 2-acetyl 1-pyrroline (2AP) (Buttery et al., 1982), and also to a host of other odour-active volatiles (Boualaphanh et al., submitted). The genetics of 2AP accumulation has been studied extensively over the past few years and it has been shown to be due to allelic variation at the betaine aldehyde dehydrogenase locus (*BADH2*) located on chromosome 8 (Bradbury et al., 2008; Bradbury et al., 2005a). Ten alleles leading to different amounts of 2AP have been identified, but one, bearing an 8 bp deletion and 3 SNPs on exon 7, is the most common allele found in varieties currently cultivated (Fitzgerald et al., 2008; Kovach et al., 2009). The involvement of *badh2* in fragrance was confirmed by the loss of 2AP after transformation of a fragrant variety with functional *BADH2* (Chen et al., 2008). A recent study shows that *BADH2* participates in the pathway of γ-aminobutryric acid (GABA) synthesis by oxidising 4-aminobutanal, produced by oxidative deamination of putrescine (Fogel et al., 1979), to GABA (Bradbury et al., 2008). Varieties carrying *badh2* cannot carry out the oxidation step, and they accumulate both 4-aminobutanal in equilibrium with its more energetically stable isomer Δ-1-pyrroline (Bradbury et al., 2008; Chen et al., 2008). Other compounds found to be in higher concentrations in aromatic rice are 1-pyrroline-5-carboxylic acid and methylglyoxal (Huang et al., 2008). However, initiation of the pathway of 2AP synthesis, the identity of other compounds involved in the pathway, and the origin of the acetyl group on the 2AP molecule are all unknown factors that could be involved in determining the amount of 2AP that accumulates.

It is likely that 2AP is synthesised within the rice grain. It is not the product of cooking resulting from a Maillard reaction, because it is possible to detect 2AP in uncooked grains of rice (Nadaf et al., 2006; Sriseadka et al., 2006). The amount of

2AP in rice grains can vary considerably. Screening of 279 varieties carrying the common fragrance allele, the functional nucleotide polymorphism (FNP) on exon 7, showed that 2AP accumulation ranged from 0.25 ppm to 3.5 ppm (Fitzgerald et al., 2008b). This variation has been suggested to result from mutations at the *BADH1* locus on chromosome 4, leading to additional 2AP (Singh et al., 2010). However, it is also possible that allelic variability at other as yet unknown loci in the pathway leading to 2AP through *badh2* could be responsible for altering the rate of the complete pathway with different amounts of 2AP as a consequence. The discovery of genes that modify the amount of 2AP would be greatly facilitated by the identification of the compounds that are involved in the pathway of 2AP synthesis.

Detection of compounds involved in biochemical pathways has recently become easier due to advances in metabolomic profiling of compounds and metabolites in plants (Hall, 2006). 2AP is detected as a volatile compound from the grain or from a solvent extraction of the grain (Bergman et al., 2000). The known precursor, Δ-1-pyrroline can be detected by gas chromatography (Bradbury et al., 2008), however the chemical, spatial or temporal attributes of other compounds that contribute to the pathway are unknown. The objective of the present study is to identify compounds that could contribute to the pathway of 2AP synthesis using introgression lines (ILs) developed using next generation genotyping tools and a number of metabolomic profiling platforms. These specifically-developed ILs carry the fragrance pathway, with minimal other introgression from the donor parent, in a non-fragrant background. Volatile compounds and primary and secondary polar metabolites were determined to identify the compounds that are unique to the fragrant ILs compared with the non-fragrant recurrent parent.

## 6.2 Materials and methods

### 6.2.1 Plant materials

A backcross population was constructed using a traditional variety of rice (*Oryza sativa.* L) from the Lao PDR, Hom Nang Nouane (HNN), which carries the FNP on exon 7 of *BADH2*, as the donor parent. The recurrent parent was Thasano 1 (TSN 1), a non-aromatic variety, also from the Lao PDR. The objective of the backcrossing was to transfer just the fragrance pathway from HNN to the TSN 1

background. The first crosses were made in the wet season of 2006 at the Rice and Cash Crop Research Centre in Vientiane, Lao PDR. At each generation, progeny were genotyped for the FNP at the *BADH2* locus exactly as previously described (Fitzgerald et al., 2008b). The $F_1$ progeny that were heterozygous at that locus were selected and crossed back to TSN 1. This procedure was repeated for four generations. The $BC_4F_1$ progeny were allowed to self, to give $BC_4F_2$. The set of 437 $BC_4F_2$ ILs was then grown in the Lao PDR in the rainy season of 2009. At early tillering, leaf material was harvested from each line for DNA extraction, and grain was harvested at maturity.

### 6.2.2 Genotyping

DNA extraction and genotyping for the FNP at the *BADH2* locus was done exactly as previously described (Fitzgerald et al., 2008b).

In order to estimate the degree of genetic difference between the two *indica* parents, DNA was extracted from TSN 1 and HNN, and genotyping was carried out at 1536 single nucleotide polymorphism (SNP) loci using an Illumina BeadExpress platform at Cornell University, Ithaca, NY (Zhao et al., 2010). SNPs were annotated using the Alchemy algorithm (Wright et al., 2010).

Selection of the ILs with the least donor parent DNA was carried out by genotyping at 384 single nucleotide polymorphism (SNP) loci, a subset of the 1536 set specific for indica x indica crosses (Thomson et al., In preparation). For SNP genotyping, a set of 46 lines carrying the FNP and 45 without the FNP were selected from the $BC_4F_2$ lines, based on morphological similarity of the plants to TSN 1. SNPs were genotyped using an Illumina Beadxpress GoldenGate Genotyping Assay (San Diego, CA, USA). The proportion of donor introgression was determined using GGT 2.0: Graphical GenoTyping (http://www.ncbi.nlm.nih.gov/pubmed/1822293). A set of five ILs, with the fragrance FNP and the least introgression from HNN, were selected and were grown for two more generations to minimise heterozygosis and for seed increase. When they reached $BC_4F_4$, seed was planted and at the three leaf stage, leaves were collected for DNA extraction and SNP profiling at the 384 loci to determine the segregation pattern at the loci that were heterozygous at the $F_2$ stage.

### 6.2.3 Measurement of 2AP

Harvested rice grains from the five $BC_4F_4$ lines selected was dehulled (Satake Rice Machine, Tokyo, Japan) and then polished in a Paintshaker fitted with home-made capsules. A sub-sample of the polished grain was ground to coarse flour. 2AP was quantified by gas chromatography (GC) as described previously (Fitzgerald et al., 2008b).

### 6.2.4 Metabolite profiling

Metabolite profiling was carried out on the five $BC_4F_4$ lines and the parents to determine volatile compounds and primary metabolites in the grains. Polished grains of each of the five lines and the parents were sent to metabolomic profiling partners in the META-PHOR project (www.meta-phor.eu). All samples were frozen in liquid nitrogen before grinding to a fine powder which was then maintained in the deep frozen state until the moment of extraction. The frozen powder was divided into replicate aliquots and distributed on dry ice to the Rothamsted for profiling of primary and secondary polar metabolites, Max Plank for high resolution primary metabolites and PRI for volatiles compounds. Extraction, instrument parameters, data collection and analysis were exactly as described in Chapter IV.

## 6.3 Results

### 6.3.1 SNP profiling

Figure 6.1 shows the genetic differences between TSN 1 and HNN at 1536 SNP loci. The two varieties were polymorphic at 35% of the 1536 loci. Of the 437 $BC_4F_2$ lines, 28% were homozygous for the *badh2* allele and 26% were homozygous for the *BADH2* allele. Of those, 46 ILs with *badh2* and 45 ILs with *BADH2* were genotyped at 384 loci. Figure 6.2 shows the SNP map of the 91 genotyped ILs at the 384 loci ordered by the percentage of similarity to the donor parent, HNN. This figure shows that there is a large difference in the amount of donor parent genome in each of the ILs, and a large proportion of heterozygotic loci. Many samples also show significant donor introgression on chromosome 8, near the *BADH2* locus. Table 6.1 shows that 15 of the 91 lines have less than 0.1% donor genome, but only four of those carry the FNP. A large number of lines have more than 10% donor

introgression, with one having 33%. As the proportion of donor introgression increases, the likelihood of *badh2* being included does too (Table 6.1).

The fragrance gene is located on chromosome 8. Figure 6.2B shows the range in donor introgression on chromosome 8 for all those with the badh2 allele. One of the SNP loci is within the flanking region of the BADH2 gene, but many of the lines with the badh 2 allele have donor introgression on more than half of the chromosome.

Five $BC_4F_2$ lines were selected based on the FNP at the fragrance locus and percent similarity to the recurrent parent. Three of these, lines 266, 269 and 280, have > 0.1% donor introgression and two, lines 85 and 148, have 0.8% similarity, but most of the loci with donor introgression are segregating (Figure 6.3a). Two generations later, at $BC_4F_4$, much less heterozygosis is seen and the lines all carry < than 0.1% donor introgression (Figure 6.3b). All five lines show introgression from the donor parent at chromosome 8, and the *badh2* allele (Figure 6.3c).

### 6.3.2 Metabolite profiling

Figures 6.4a-b show principal component analyses (PCA) of the two parents and the five ILs from two profiling platforms: GCMS for volatile compounds and $H^1$ NMR for primary polar metabolites. In all cases, HNN is distinct from the other samples, and ILs cluster with TSN 1 are slightly separated from the other three and TSN 1. Table 6.2 shows those compounds that could be identified that discriminated between TSN 1 and the $BC_4F_4$ ILs. Only 2 volatile compounds were present in the ILs and not TSN 1. The most discriminatory compound was 2 acetyl 1-pyrroline, which was present in all ILs and not TSN 1 and a branched alkane that could not be further identified.

Significant differences were found in the presence or amount of some primary and secondary polar compounds. Decreases in sucrose and amino acid concentration were major factors that caused separation along the PC2 axis (Figure 6.4b). Interestingly, a simple compound carrying an acetyl group was detected in the ILs that caused some of the separation from TSN 1 (Figure 6.4a).
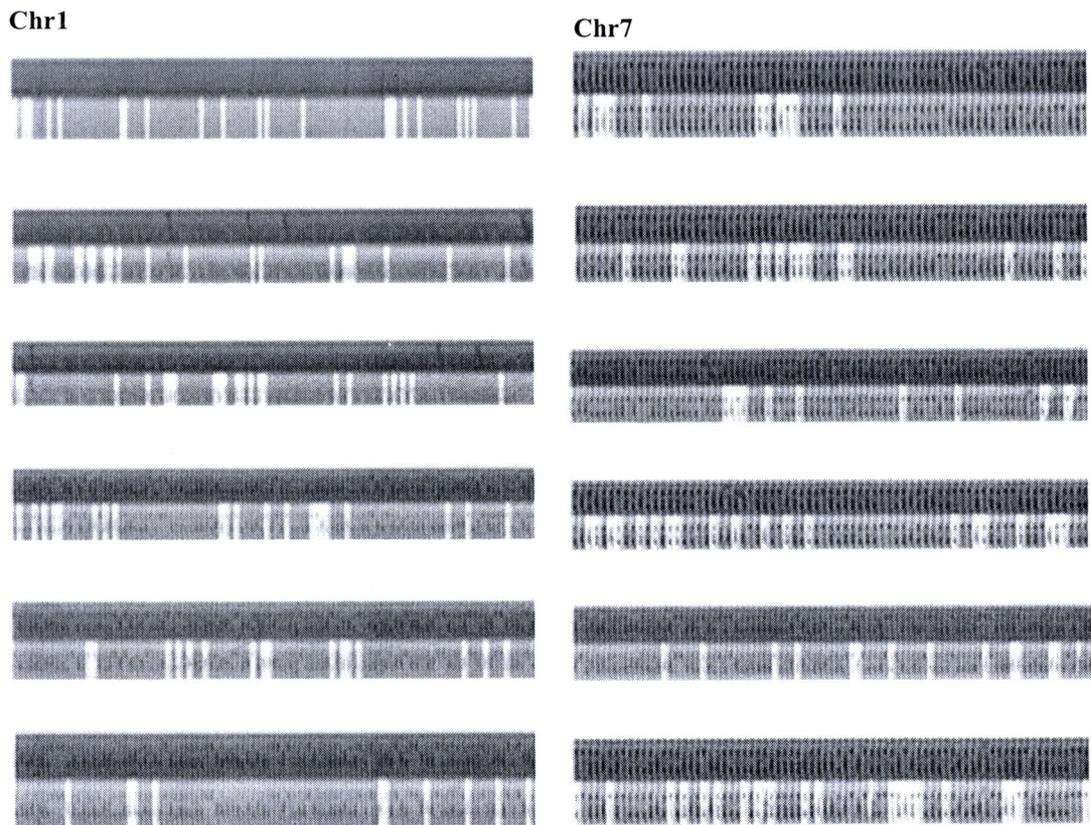
**Chr1**

**Chr7**



**Figure 6.1** Comparison between the genotype of TSN 1 and HNN at each chromosome at1536 loci. HNN is shown at the top of each chromosome, and polymorphic regions between TSN and HNN are light grey. Chromosomes 1 – 6 are in the first block and 7 – 12 in the second. The parents differ at 35% of loci.
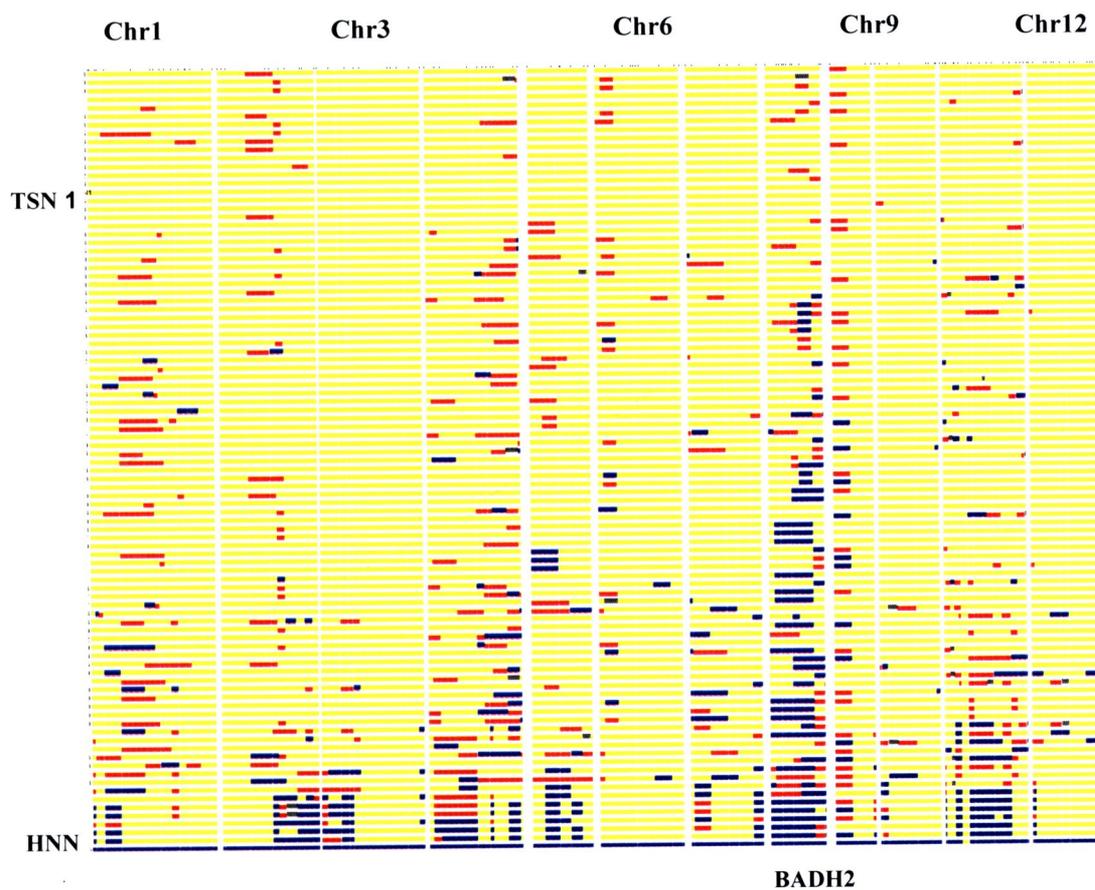
**Figure 6.2** SNP map at 384 loci of the 91 BC$_4$F$_2$ lines derived from HNN and TSN 1. TSN 1 is yellow and the first line, HNN is dark blue and the last line. Each line in the figure shows the genotype of one BC$_4$F$_2$ line. Introgression from HNN can be seen as blue regions and heterozygotic regions are red.
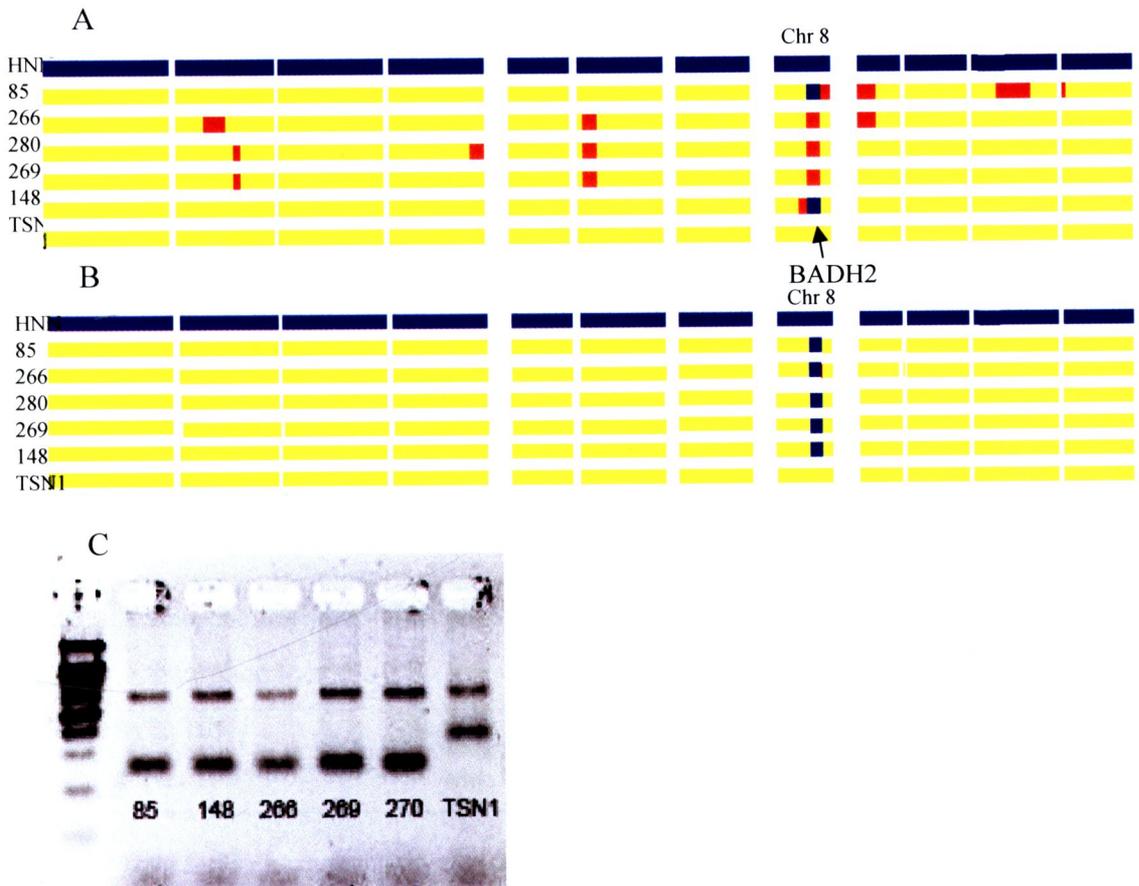
**Figure 6.3** A: SNP map of the parents and selected BC$_4$F$_2$ lines carrying the Fragrance gene but with most of the TSN 1 genome. Regions of HNN Introgression are mostly heterozygotic (red). B: SNP map of the same lines at BC$_4$F$_4$ showing HNN introgression only at a region on chromosome 8 (blue). C: the presence of the *badh2* allele in the five introgression lines and *BADH2* in the recurrent parent, TSN 1.
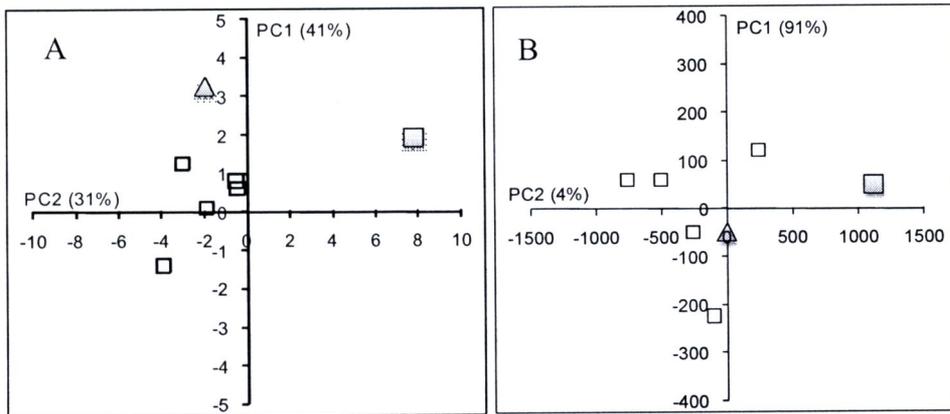
**Figure 6.4** PCA of volatile compounds (A) and polar metabolites (B).

**Table 6.1** Percent similarity of lines to HNN with and without the FNP at the fragrance locus.

| Similarity to HNN (%) | Number of lines with FNP | Number of lines without FNP |
|---|---|---|
| ≤ 0.1 | 4 | 11 |
| 0.1 – 1 | 8 | 7 |
| 1 – 2 | 3 | 4 |
| 2.0 - 3.0 | 1 | 7 |
| 3.0 - 4.0 | 3 | 5 |
| 4.0 - 5.0 | 0 | 3 |
| 5.0 - 7.5 | 9 | 3 |
| 7.0 – 10 | 7 | 2 |
| 10.0 - 20.0 | 4 | 2 |
| 20.0 - 35.0 | 6 | 0 |

**Table 6.2** Volatile, primary polar and secondary metabolites that differ between the recurrent parent and the introgression lines. + indicates that a compound has increased in the ILs, - means the amount of a compound has decreased in the ILs, and ++ means that a compound is unique to the ILs.

| Compounds | |
|---|---|
| **Volatiles** | |
| 2-acetyl-1-pyrroline | ++ |
| branched alkanes | ++ |
| other alkanes | ++ |
| **Primary metabolites** | |
| asparagine/aspartate | + |
| acetyl bearing metabolite | + |
| Valine | + |
| Proline | - |
| Glutamate | - |
| Sucrose | - |
| Choline | - |
| Putrescine | ++ |
| GABA | ++ |
| **Putative Secondary metabolites** | |
| hydroxy glutamate/acetyl cysteine | + |
| carnitine/hydroxyl lysine/ hydroxyl adipate/hydroxyl oxoglutarate | + |
| Tyramine | + |
| Phenylalanine | + |
| leucine/isoleucine/carbamoyl putrescine/acetyl beta alanine/glutamate semi aldehyde | + |
| Citrate | + |
| mannitol/sorbitol/galactitol | - |
| acetylcarnitine/tryptophan/ oxaloglutarate | - |

## 6.4 Discussion

### 6.4.1 Developing high quality genetic stocks

Introgression lines, carrying minimal donor introgression aside from the desired locus, are extremely useful for identifying genes and determining their

function (Myint et al., 2009, Jairin, 2009). However, ILs have historically needed to undergo many generations of backcrossing to reach the desired level of genetic purity. The past decade has borne witness to an exponential development in genetics and genotyping capacity, driven by high throughput technologies allowing genetic selection and faster development of genetic stocks. Figure 6.1 shows that 35% of the loci are polymorphic between the two parents used to develop the ILs, and Figure 6.2 shows that there are large differences in the amount of donor DNA between the lines from the population, ranging from 0.1% to almost the maximum possible at 33%. If this variability is not visible in morphological characteristics, the whole population needs to be progressed through several more backcrossing events to maximise background purity. By carrying out genome-wide SNP genotyping, it has been possible to select relatively rapidly those lines with the desired donor gene and with the lowest other donor introgression after just four generations of backcrossing. Furthermore, it is possible to select the $F_2$ material with the desired introgression, but only heterozygotic loci at other points of introgression. By selecting five such lines in the present study, and progressing them to $BC_4F_4$, segregation occurred at the heterozygotic loci, and most of them are now very pure contain a small genome region bearing the *badh2* FNP (Figure 6.3c), and no other donor introgression at any of the 384 SNP loci.

Many of the $BC_4F_2$ lines showed donor introgression on chromosome 8 (Figure 6.2), and the five selected lines did also (Figure 6.3a). One of the 384 SNPs on chromosome 8 is just ~300kb 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 HNN allele at ~300 kb upstream of the *BADH2* locus is a strong indicator of introgression of the *badh2* allele. However, many lines showed large donor intrgoression around that SNP, indicating a significant amount of genetic drag around the *badh2* allele. With SNP genotyping providing a lens into the genome of the ILs, it was possible to select a set of ILs with minimum donor introgression, the *badh2* allele, and minimal genetic drag around the badh2 allele.

### 6.4.2 Possible pathways to 2AP

Analysis of the volatile compounds from the recurrent parent and the ILs showed 2AP, the major fragrance compound, as the most discriminatory, follow by a branched alkane. It is likely therefore, that the biochemical pathway leading to 2AP does not occur in the volatile phase.

The synthesis of 2AP requires a carbon and a nitrogen precursor. Putrescine is one of the major discriminating compounds between the ILs and HNN (Table 6.2), and it is one of three polyamines that are widespread in living organisms (Takahashi and Kakehi, 2010). Polyamines are involved in a host of regulatory, developmental and stress response processes (Takahashi and Kakehi, 2010) and their concentration in tissues must be strictly regulated (Moschou et al., 2008). Catabolism is one such regulatory process (Kusano et al., 2008), and for putrescine this is achieved through oxidative deamination by a copper containing diamine oxidase (CuOA) (Cona et al., 2006). One of the products of putrescine catabolism is 4-aminobutanal, which immediately cyclises to the more energetically stable Δ-1-pyrroline (Fogel et al., 1979), and can be metabolised further by an aldehyde dehydrogenase to GABA (Sebela et al., 2000). GABA is an important metabolite associated with regulation of cytosolic pH, signalling, protection against stresses and regulation of carbon flow into the citric acid cycle (Bouché and Fromm, 2004). GABA was found in TSN 1 grains at higher concentrations than in the ILs.

Catabolism of polyamines is the major source of Δ-1-pyrroline (Angelini et al., 2010). Interestingly, in rice grains, Δ-1-pyrroline is oxidised to GABA by BADH2 (Bradbury et al., 2008), and in soybeans by a homologue of BADH2 (Arikit et al., 2011). This strongly indicates that BADH2 in rice is the amino aldehyde dehydrogenase that is part of the metabolic pathway of polyamine catabolism. Thus we support the recommendation by Bradbury et al. (2008) to re-annotate the *BADH2* gene to an amino aldehyde dehydrogenase (AADH).

Most oxidative deamination of polyamines occurs in the apoplast (Cona et al., 2006), where two of the reaction products, $NH_3$ and $H_2O_2$, are thought to contribute to cell wall development, mediate defence responses to attacking insects, and signal programmed cell death (Cona et al., 2006; Moschou et al., 2008). However the AADH gene in rice, which oxidises the Δ-1-pyrroline, presumably produced by a

CuAO, carries a signal peptide targeting it to a peroxisome (Bradbury et al., 2008). BADH2 from rice has previously been detected in peroxisomes with immunolabelling (Nakamura et al., 1997a). Peroxisome targeting has also been found for the homologous AADH in soybean (Arikit et al., 2011), barley (Nakamura et al., 1997b), and tobacco (Luo et al., 2001). The location of this part of the pathway in peroxisomes in several plant species suggests either that Δ-1-pyrroline is transported from the apoplast to the peroxisome to be oxidised to GABA, or that there is a CuOA that is expressed in the peroxisome. *In silico* analysis shows that a copper-containing primary amine oxidase (EC 1.4.3.21; Os04g0476100) is present in rice (Kyoto Encyclopedia of Genes and Genomes http://www.genome.jp/kegg/). It has approximately 77% similarity witha N-methylputrescine oxidase (MPO1 and MPO2) reported in tobacco, which oxidatively deaminates N-methylputrescine to 4-methyl-aminobutanal (Katoh et al., 2007), and it carries a C-terminal peroxisomal targeting signal (PTS1). It is therefore possible that the catabolism of putrescine to GABA could occur completely within peroxisomes, which means that 2AP in fragrant rice is probably localized within the peroxisome also.

Perhaps CuAOs are specifically localised based on their biochemical role. Apoplastic CuAO activity is reported to be for cell wall functions mediated by the peroxide formed, whereas CuAO activity leading to GABA requires the other product of the reaction, 4-aminobutanal. Perhaps this reaction occurs in peroxisomes so that the GABA formed can be used within the cell, and the peroxide formed can be neutralised in the peroxisome. In peroxisomes of transgenic tobacco carrying a peroxisomal BADH2 gene (which could also be AADH), both catalase and peroxidase are significantly upregulated (Luo et al., 2001) suggesting neutralisation of peroxide. Thus, it seems that that Δ-1-pyrroline formed from putrescine is likely to be part of a pathway localised in peroxisomes that leads to GABA by the action of AADH. Presumably this process occurred in TSN 1 since GABA was found in high concentration in those grains, but much lower in the grains of the ILs.

Each of the five ILs carries a mutation in the AADH gene, and that mutation leads to a stop codon, rendering the enzyme non-functional (Bradbury et al., 2005a). Presumably, this leads to an accumulation of 4-aminobutanal/ Δ-1-pyrroline. The cyclised form is more energetically stable, but perhaps accumulation of Δ-1-pyrroline

inhibits further cyclisation, leading to a simultaneous accumulation of 4-aminobutanal. It can be predicted that this would inhibit the activity of CuAO, leading to an accumulation of putrescine. The concentration of polyamines in cells is thought to be under tight regulation (Moschou et al., 2008), suggesting that an accumulation of putrescine could have significant effects on the biochemistry of the grain.

Acetylation of the accumulating Δ-1-pyrroline leads to a more energetically stable compound than Δ-1-pyrroline, and it would alter the equilibrium conditions between Δ-1-pyrroline and 4-aminobutanal, which would allow the oxidative deamination of putrescine to proceed. Acetylation of the Δ-1-pyrroline leads to grains with a pleasant popcorn aroma, which is in strong contrast to the aroma the grain would have if putrescine accumulated. Either ancient farmers selected grains that were able to acetylate Δ-1-pyrroline, preventing the accumulation of putrescine, or acetylation is a response to the accumulation of Δ-1-pyrroline, allowing polyamine concentrations to be regulated. Some varieties of soybean carry a mutation in the homologous AADH gene, and these varieties also acetylate Δ-1-pyrroline to form 2AP (Arikit et al., 2011), this cross-species similarity provides some evidence that acetylation could be a mechanism to regulate polyamine concentration in the absence of AADH activity.

A polar compound carrying an acetyl group was detected in the ILs but not in TSN 1 (Table 6.2). This could be a potential source of acetylation of Δ-1-pyrroline. Acetyl CoA is another possibility, and it is formed in the oxidation of higher polyamines (Cona et al., 2006). This compound carries the acetyl group on a protein, so is unlikely be detected by the platforms used in the present study. Acetyl CoA could react with the Δ-1-pyrroline through the agency of an acetyl transferase. While we cannot determine the acetylation mechanism, we propose that 2AP in rice and soybean accumulates in response to the disruption of polyamine catabolism, and that the acetylation process enables the catabolism of putrescine to proceed, so that polyamine concentration is regulated in the cell.

Several studies suggest that in conditions of abiotic stress, plants catabolise putrescine (Quinet et al., 2010) to produce GABA which mediates a stress response (Cona et al., 2006). When soybean, rice and oat plants were exposed to salinity stress, a significant increase in GABA was observed, concomitant with a significant increase

in CuOA and aldehyde dehydrogenase activity, and a decrease in putrescine concentration (Quinet et al., 2010; Xing et al., 2007). Moreover, inhibitors of CuOA led to decreases in GABA, and increases in salinity effects, indicating that GABA from polyamine catabolism plays an important role in response to abiotic stresses (Xing et al., 2007). In fragrant rice, which cannot produce GABA through AADH, acetylation to remove of the Δ-1-pyrroline from the pathway would allow the deamination of putrescine in response to salinity stress, but instead of GABA, 2AP accumulates. Perhaps this explains the often repeated anecdote that fragrant rice is always much more fragrant when grown in conditions of slight salinity stress.

**6.5 Conclusion**

It is proposed that the accumulation of the key fragrance compound 2AP in rice grains occurs as a fortuitous side-effect of polyamine catabolism, and the acetylation step occurs to enable the plant to maintain its ability to regulate polyamine concentration in the grains.