

# CHAPTER V

## POPULATION DEVELOPMENT USING NEW GENERATION TOOLS

### 5.1 Introduction

Introgression lines (ILs) are valuable genetic stocks that have been widely used for genetic studies, but it takes at least six generations of backcrossing to obtain near ILs. Over the last few years, molecular markers have proven powerful tools for the development of ILs and near isogenic lines both for genes of interest and obtaining background purity (Jairin et al., 2009; Keurentjes et al., 2007; Li et al., 2005; Neeraja et al., 2007). However, using a statistically significant number of markers on many progeny becomes costly. Now, genome-wide single nucleotide polymorphism (SNP) technology is being developed that allows whole populations to be screened economically, and to enable selection of specific progeny based on genome-wide background purity (McNally et al., 2006; McNally et al., 2009). This technology has recently become available for populations derived from *indica* and *tropical japonica* varieties of rice (Thomson et al., In preparation; Wright et al., 2010), which provides new opportunities to develop valuable genetic stocks to search for genes of interest more quickly and cheaply than has been previously possible.

In many countries of Asia, traditional landraces of rice continue to be grown, despite active breeding programs developing and releasing higher yielding varieties. These landraces persist because the eating quality of the grain has not been able to be captured in higher yielding varieties, and because very few of the genes leading to particular traits of qualities are known (Fitzgerald et al., 2009). Even when a new variety, such as Apo (PSBRc9), offers superior yield under drought stress (Venuprasad et al., 2008), widespread adoption does not automatically follow because a variety must be accepted by consumers before farmers will adopt and grow it. Apo has been released in many Asian countries, including the Lao PDR, for cultivation in drought-prone areas. Even though it has given significantly higher yield in all areas tested, consumers have generally rejected it on the basis of its eating quality.

High-yielding varieties with acceptable eating quality are available in Lao, but two traditional varieties that are highly aromatic and prized for their taste, continue to be grown in the major rice-growing areas. These are Kai Noy Leung (KNL) and Hom Nang Nouane (HNN) (Inthapanya et al., 2006b). The Lao rice improvement program has previously focused on conventional crossing and testing, but has not succeeded in capturing the aromatic quality of either KNL or HNN into high-yielding backgrounds. HNN carries the most common mutation leading to aroma (Bounphanousay et al., 2008). This is a mutation in the betaine aldehyde dehydrogenase (*BADH2*) gene (Bradbury et al., 2005a). The genetic basis of aroma in KNL is unknown (Fitzgerald et al., 2008b). The objective of this study is to develop two sets of ILs, as close to isogenic lines as possible, with a fragrance pathway in each set, using KNL as the donor parent of the fragrance pathway for one set, and HNN as the donor of the fragrance pathway for the other set. For both, the recurrent parent chosen was the high-yielding improved variety, Thasano 1 (TSN 1). Genome-wide SNP genotyping at 384 loci (Thomson et al., In preparation; Zhao et al., 2010) will be undertaken at BC<sub>4</sub>, offering the possibility to select particular progeny carrying the gene of interest in a background that is primarily from the recurrent parent. Such genetic stocks are extremely valuable for research projects, and by coupling trait-based relevant phenotyping with new generation genotyping, these stocks can now be obtained economically and in a short time-frame.

## 5.2 Materials and methods

### 5.2.1 Plant materials

Two populations of introgression lines (ILs) (*Oryza sativa*. L) were developed. The donor parent in the first was HNN, which carries the functional nucleotide polymorphism (FNP) on exon 7 of *BADH2*, and KNL was the donor in the second. For both populations, the recurrent parent was TSN 1, a non-aromatic, high-yielding variety. The first crosses were made in the rainy season of 2006 at the Rice and Cash Crop Research Centre in Vientiane, Lao PDR, and all subsequent crosses were made there also.

At each generation, progeny of the HNN population were genotyped for the FNP at the *BADH2* locus exactly as previously described (Fitzgerald et al., 2008).

F<sub>1</sub> progeny that were heterozygous at that locus were selected and crossed back to TSN 1. This procedure was repeated for four generations. The BC<sub>4</sub>F<sub>1</sub> progeny were allowed to self, to give BC<sub>4</sub>F<sub>2</sub>. The set of 437 BC<sub>4</sub>F<sub>2</sub> ILs was then grown in the Lao PDR in the wet season of 2009. At early tillering, leaf material was harvested from each line for DNA extraction, and grain was harvested at maturity.

For the set with KNL as the donor parent, fragrant progeny were selected based on a sensory test for fragrance. This test was carried out by cooking leaf material (2 g) in KOH (1.7%, 10 ml) for 10 min. The KOH volatilises 2-acetyl-1-pyrroline (2AP) and it is able to be detected by smelling (Juliano, 1985). Those found to be fragrant were back-crossed to TSN 1 in the same way as for the HNN set. The BC<sub>4</sub>F<sub>1</sub> progeny were allowed to self, to give BC<sub>4</sub>F<sub>2</sub> in the wet season of 2009. Leaf material was harvested from each line at early tillering for DNA extraction and to determine the presence of 2AP, and grain was harvested at maturity.

### **5.2.2 Measurement of 2-acetyl-1-pyrroline (2AP)**

The major compound of fragrance, 2AP, was measured in the BC<sub>4</sub>F<sub>2</sub> leaf samples of the KNL set by coupled gas chromatography-mass spectrometry (Agilent 6890N gas chromatograph equipped with a 5975N mass selective detector). 2AP was extracted by heating 1 g of fresh leaves in distilled water for 30 mins at 85°C. The leaves were discarded and dichloromethane and saturated sodium chloride were added to the aqueous extract. The mixture was shaken for 1 min. The organic layer, containing the 2AP, was filtered over anhydrous sodium sulfate and concentrated over a stream of nitrogen. The operating conditions of the GC-MS and detection of 2AP were exactly as previously described (Kovach et al., 2009).

### **5.2.3 SNP genotyping to select pure NILs**

A subset of each BC<sub>4</sub>F<sub>2</sub> population was selected for SNP genotyping based on fragrance genotype and phenotype, and on similarity of plant type to TSN 1. A set of 46 of the BC<sub>4</sub>F<sub>2</sub> lines with the fragrance allele derived from HNN and 45 that did not carry the fragrance gene were selected. For the KNL set, SNP genotyping was carried out for the 34 BC<sub>4</sub>F<sub>2</sub> lines for which 2AP was detected in leaves by GCMS, and 34 without 2AP in leaves. DNA was extracted from leaf tissue of all the samples selected for genotyping exactly as previously described (Fulton et al., 1995). DNA in the extracts were quantified to 50 ng/ul using Nanodrop 1000 (Wilmington, DE, USA).

Genotyping was carried out at 384 Single Nucleotide Polymorphism (SNP) loci using an Illumina BeadXpress GoldenGate Genotyping Assay (San Diego, CA, USA) (Thomson et al., In preparation). SNP calls were analysed using Alchemy software (Wright et al., 2010). The proportion of donor introgression in each line was determined using GGT 2.0: Graphical GenoTyping (van Berloo, 2008).

#### **5.2.4 Detecting volatile compounds in the selected introgression lines**

Following the SNP genotyping, lines from each population were selected based on fragrance genotype, phenotype and minimal background introgression from the donor parent. Five were selected from the HNN set and six from the KNL set. These samples were planted in the Lao PDR for two more generations, to stabilise heterozygosity at background loci and to increase seed quantities. At BC<sub>4</sub>F<sub>4</sub>, grains were harvested at maturity from the eleven selected lines, and sent to the International Rice Research Institute for processing. Paddy was dehulled (THU35A 250V 50Hz Test Husker, Satake) and polished (Grainman 60-230-60-2AT, Grain Machinery Mfg. Corp.). Polished grain of each of the eleven samples and parents was sent to Plant Research International Wageningen, The Netherlands for profiling of volatile compounds.

Headspace volatiles were collected by solid phase micro extraction (SPME) using a 65-mm polydimethylsiloxane-divinylbenzene fiber (Supelco, Bellefonte, USA) as described in detail (Verhoeven et al., in press). The volatile compounds were thermally desorbed at 250°C by inserting the fiber for 1 min into the GC injection port (GC 8000, Fisons Instruments). The released compounds were transferred on the analytical column (HP-5, 30 m x 0.25 mm ID, 1.05 µm – film thickness) in splitless mode. The temperature program started at 45°C (2-min hold) and rose 5°C min<sup>-1</sup> to 250°C (5-min hold). The column effluent was ionised by electron impact (EI) ionisation at 70 eV (MD800 electron impact MS, Fisons Instruments). Mass scanning was done from 35 to 400 m/z with a scan time of 2.8 scans sec<sup>-1</sup>. GC-MS raw data were processed by using MetAlign software (Lommen, 2009) to extract and align the mass signals (s/n ≥ 3). Mass signals that were below a s/n of 3 were awarded randomized values between 2.4 and 3 times the calculated local noise value. Only mass signals that were present in at least 6 samples were retained for analysis, all others were discarded. Signal redundancy per metabolite was removed by means of

clustering and mass spectra were reconstructed (Tikunov et al., 2005). The quantitative composition of the volatile blends was evaluated by principal components analysis (PCA) using the software program SIMCA-P 12.0.1. (Umetrics AB, Umea°, Sweden) (Eriksson et al., 2006). Data were log-transformed and then variables were mean-centred, and pareto scaled.

## 5.3 Results

### 5.3.1 Population development

Table 5.1 shows, for both populations at each generation, the number of plants crossed, the number of grains planted, and the number of seeds harvested. For the KNL set, 34 BC<sub>4</sub>F<sub>2</sub> lines were identified by the sensory test and GC analysis to be fragrant. For the HNN set, 77 BC<sub>4</sub>F<sub>2</sub> lines were fragrant and carried the fragrance gene, and a several non-fragrant lines were also selected.

**Table 5.1** Number of plants for each population obtained in each generation and number selected for backcrossing.

Generation		Donor Parent	
		HNN	KNL
F <sub>1</sub>	Planted	20	15
	Crossed	5	4
BC <sub>1</sub> F <sub>1</sub>	Planted	19	11
	Crossed	8	6
BC <sub>2</sub> F <sub>1</sub>	Planted	202	317
	Crossed	71	21
BC <sub>3</sub> F <sub>1</sub>	Planted	107	127
	Crossed	34	13
BC <sub>4</sub> F <sub>1</sub>	Planted	163	161
	Selected plants	35	27
BC <sub>4</sub> F <sub>2</sub>	Planted	437	250
	Selected plants	127	34

Germination was poor for all  $F_1$  seeds, and at  $BC_3F_1$  and  $BC_4F_2$ , many lines were lost due to an infestation of the brown planthopper.

**Table 5.2** Sensory evaluation of aroma in each  $BC_4F_2$  line derived from KNL and TSN 1

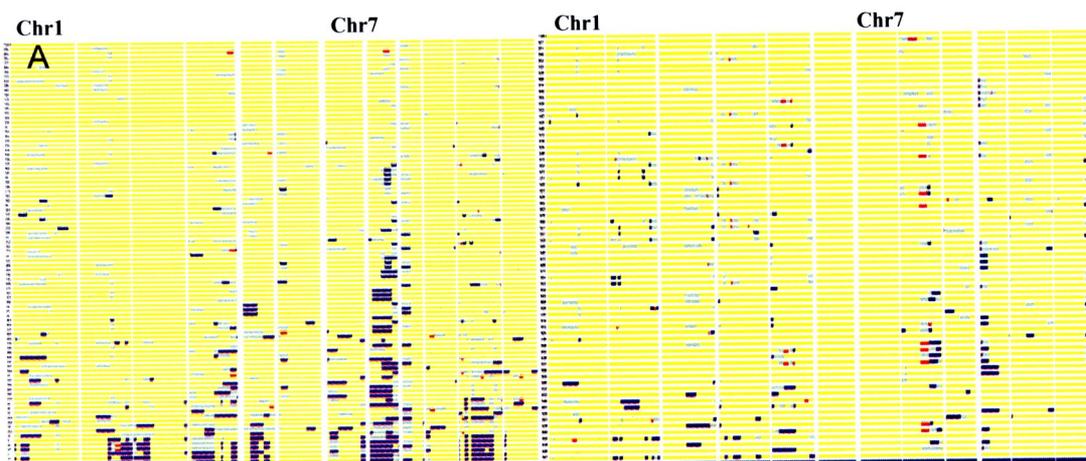
Sensory assessment of fragrance in each line	
Strong	523, 529, 557, 558, 625, 626, 629, 631, 637, 639, 664, 668
Intermediate	501, 606
Slight	505, 530, 539, 560, 577, 588, 609, 610, 619, 622, 632, 633, 634, 635, 636, 638, 640, 643, 665, 666

Table 5.2 shows the amount of fragrance in each  $BC_4F_2$  line with KNL as the donor parent detected by sensory evaluation. The data show that the range is from strong to slight, with all of the selected lines showing fragrance

### 5.3.2 SNP profiling

Figure 5.1a shows the SNP profile of selected  $BC_4F_2$  HNN line, ordered first by the presence of the fragrance gene on chromosome 8, and secondly by the percent of genetic similarity to HNN. Figure 5.1b shows a similar map for the KNL set, which are sorted first by the fragrance phenotype and secondly by the proportion of KNL present. It can be seen for both populations that the lines differ in the proportion of donor parent, and they are heterozygotic at several loci. Also notable is that there are more loci that differ between TSN 1 and KNL.





**Figure 5.1** SNP maps of the BC<sub>4</sub>F<sub>2</sub> lines from the HNN population (A) and the KNL population (B). Of the 384 loci assayed, HNN differed from TSN 1 at 137 where as KNL differed from TSN 1 at 220 loci. Yellow represents TSN 1, blue represents introgression from the donor parent, and red shows heterozygotic loci.

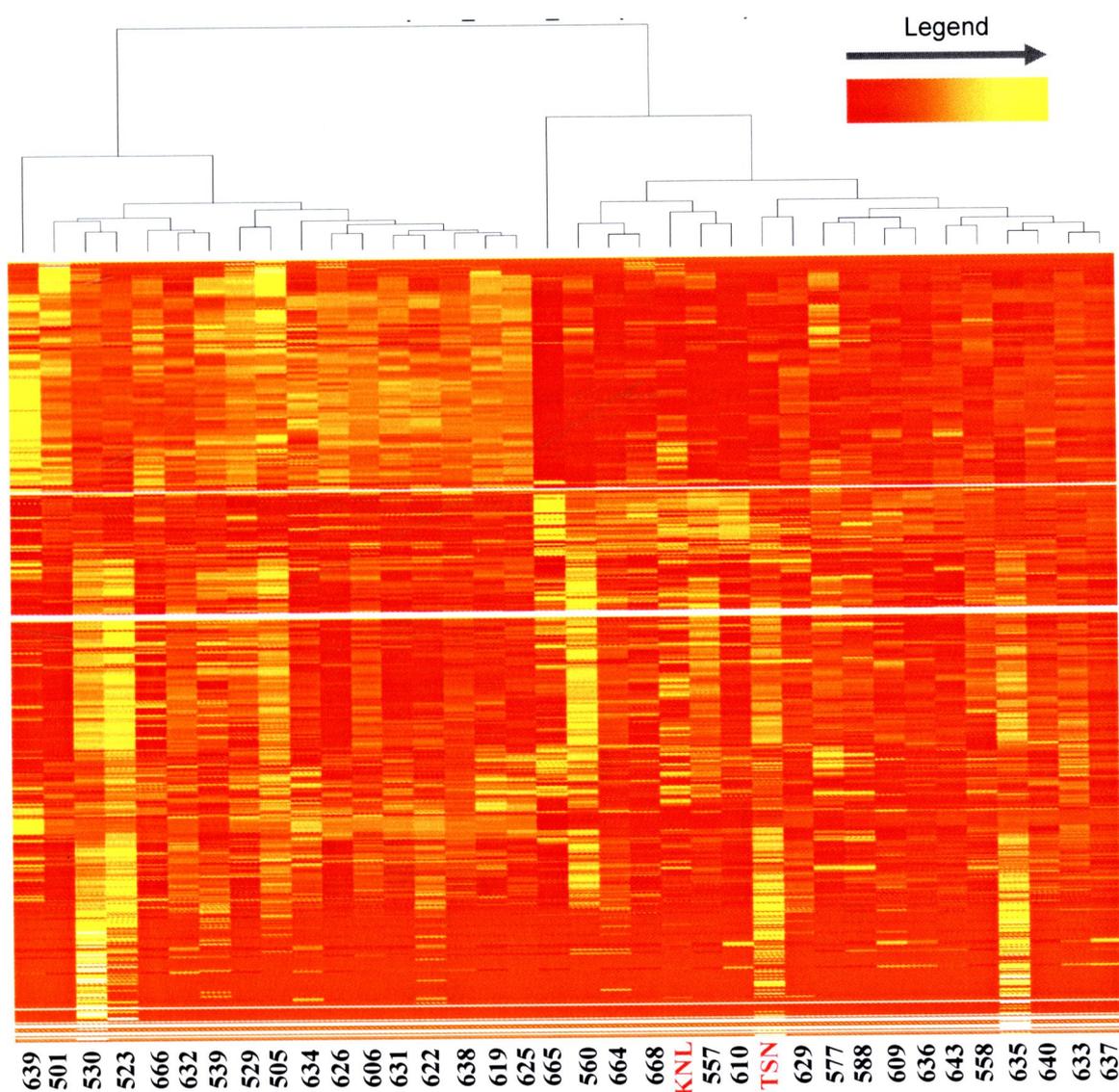
Using SNP profiling at 1536 loci (Chapter IV), KNL and TSN 1 differ at 57% of loci, and TSN 1 and HNN differ at 35% of the 1536 loci. This same difference is found for the SNP genotyping at 384 loci, which was used for the BC<sub>4</sub>F<sub>2</sub> populations. The panels in Figure 5.1 indicate that the set derived from KNL as the donor parent carries less introgression than when HNN was the donor parent. Table 5.3 shows the percent similarity of each BC<sub>4</sub>F<sub>2</sub> line to the donor parent based on the SNP genotyping at 384 loci. The similarity of the lines with KNL as the donor parent ranges from 0.1 to 7.5%, whereas those with HNN as the donor range from 0.1 – 34% (Table 5.3), close the maximum given that HNN and TSN 1 differ by only about 30% at the 1536 loci. Only four of the lines N with <1% carry the fragrance allele from HNN. Moreover, the higher the amount of donor introgression, the more likely it seems that the fragrance gene is present (Table 5.3).

**Table 5.3** Percent Similarity of each BC<sub>4</sub>F<sub>2</sub> line to the donor parent. Those derived from HNN and carry the HNN fragrance allele are in italics.

Similarity to donor parent (%)	Line number	
	KNL	HNN
≤ 0.1	523, 529, 545, 554, 571, 680	<i>280, 221, 223, 116, 142, 266, 269, 133, 146, 228, 234, 236, 241, 288, 233, 435</i>
≤ 0.5	447, 464, 507, 512, 531, 616, 532, 638, 640, 649, 665	
≤ 1	461, 504, 540, 553, 577, 592, 622, 626, 635, 679, 670	<i>108, 126, 438, 85, 98, 145, 148, 271, 326, 218, 303, 315, 370, 378</i>
1 - 1.5	530, 560, 590, 611, 617, 658, 666, 668	<i>54, 405, 261, 263, 268, 362, 397</i>
1.5 – 2	557, 562, 563, 664, 603, 609, 619	
2.0 - 3.0	448, 505, 506, 606, 634, 637	<i>374, 144, 89, 91, 93, 226, 246, 431</i>
3.0 - 4.0	444, 546, 558, 625, 631, 639, 652, 593	<i>389, 102, 240, 249, 248, 92, 96, 235, 439</i>
4.0 - 5.0	633, 684, 548, 629	<i>219, 404, 436</i>
5.0 - 7.5	610, 636, 588, 593, 587	<i>180, 253, 281, 327, 349, 129, 283, 300, 307, 380, 392, 256</i>
7.0 – 10		<i>44, 52, 429, 430, 30, 416, 367, 205, 176</i>
10.0 - 20.0		<i>204, 400, 254, 63, 29, 40,</i>
20.0 - 35.0		<i>1, 20, 17, 6, 15, 14</i>

### 5.3.3 Metabolite profiling

Since a fragrance locus was unknown in KNL, profiling of polar compounds in leaves was carried out. A significant amount of variability was found in the number of polar compounds in each line. Figure 5.2 shows a heat map of all the compounds detected, and shows how each line clusters in relation to the parents.

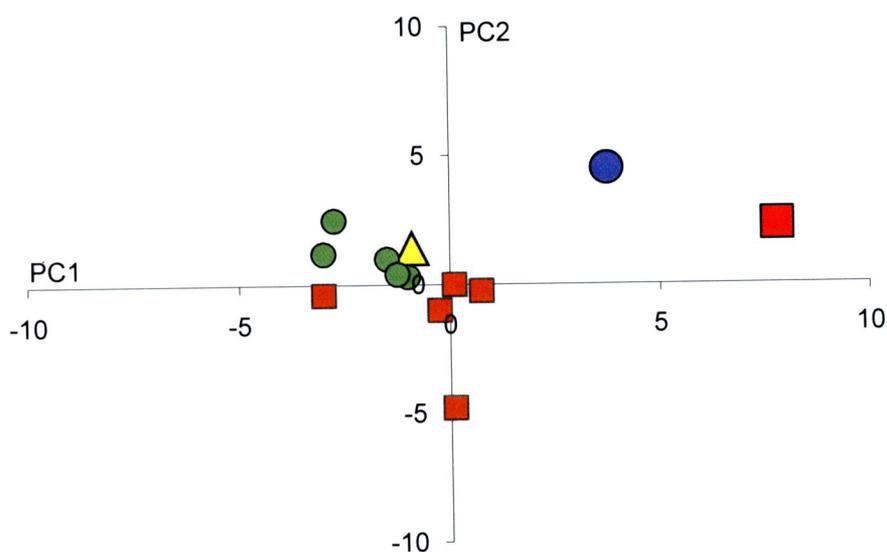


**Figure 5.2** Heat map showing groupings by polar primary compounds detected in  $BC_4F_2$  leaves of the 34 selected lines derived from KNL and TSN 1.

Figure 5.3 shows two primary clusters. Under the second, KNL falls into one sub-cluster and TSN 1 into another. The other primary cluster is clearly discriminated by the compounds in the top half of the heat map. In the KNL cluster, five of the  $BC_4F_2$  lines are present. Four of the lines share less than 1% genetic similarity with KNL, and the fifth shares 5.1% similarity with KNL (Table 5.3). By sensory testing, the three that cluster the closest to KNL show the same strength of fragrance as KNL, while two at each end of the cluster show slight fragrance (Table 5.2). The TSN 1

subcluster contains 11 lines, ranging from 0.5 – 6.6% similarity with KNL (Table 5.3). No relationship was found between the distance of clustering and the proportion of donor parent. The first cluster of lines consists of more than half of those found to have only a slight fragrance by sensory.

Using all the data sets available, the purest lines with aroma were selected from each of the populations. Figure 5.3 shows a principal components analysis of the volatile compounds in the selected lines and the parents, and clearly shows that the selected lines from each population cluster close to the recurrent parent and not to either of the donor parents. Generally the lines derived from KNL form a cluster slightly separated from those derived from HNN (Figure 5.3).



**Figure 5.3** PCA of volatile compounds in parents and selected  $BC_4F_2$  lines showing that all lines cluster near TSN 1. KNL blue, HNN red, TSN 1 yellow. HNN progenies are orange and KNL progenies are green.

#### 5.4 Discussion

In developing the populations, fragrant progeny at each generation derived from HNN and TSN 1 was selected using the fragrance marker (Bradbury et al., 2005b). However, since the fragrance genes of KNL are unknown, progeny at each generation had to be selected by detecting 2AP using a sensory technique (Juliano, 1985). The volume of samples prohibited the detection of 2AP using GC (Bergman et

al., 2000) for reasons of cost and throughput, so sensory testing was employed. Sensory testing is subjective, and therefore is less likely to be accurate compared to an objective quantification of 2AP. This possibly explains why fewer lines derived from KNL were progressed through back-crossing generations compared with the HNN set. In the interests of speed, the 690 BC<sub>4</sub>F<sub>2</sub> lines were sown in April 2009 in Laos, soon after harvesting the BC<sub>4</sub>F<sub>1</sub> lines - a month before the main crop is sown in Laos. In the WS of 2009 in Laos, an infestation of brown plant hopper affected rice production. The populations, being a month earlier than the main crop, were the only plants at tillering and therefore more susceptible to the plant hopper, so about 50 promising BC<sub>4</sub>F<sub>2</sub> lines (based on fragrance and plant type), were lost to the insects. It was observed that the fragrant lines were in this population damaged more by the hoppers in this trial than the non-fragrant lines.

Of the 437 harvested lines derived from HNN, 46 that carried the fragrance gene were chosen based on resemblance of the plant type to TSN 1. Another 45, also with the TSN 1 plant type, that did not carry the fragrance gene, were selected. These lines were genotyped for at 384 loci using the *indica/indica* chip, because both are in the *indica* germplasm class. For the KNL set, all 34 fragrant lines that survived the plant hopper infestation were selected and 40 non-fragrant lines, with plant-type like TSN 1, were selected for the SNP genotyping at 384 loci using the *indica/japonica* chip, because KNL is in the *tropical japonica* germplasm class. Figure 5.1 and Table 5.3 show that the introgression lines derived from KNL carry fewer introgressions from the donor parent (KNL) than the lines derived from HNN.

With a marker available to detect the gene of interest, in this case fragrance at the *BADH2* locus, the SNP panel (Figure 5.1a) provides an ideal tool for selecting the purest introgression lines that carry the gene of interest. For the HNN set, 5 introgression lines were selected for multiplication and further work based on genetic purity. These are line 116, 142, 266, 269 and 280. Using the 137 polymorphic SNP markers and the fragrance gene, all these lines carry 0.1% HNN introgression. Selection of pure, fragrant lines from the KNL set was more difficult.

Figure 5.1b shows that many of the fragrant lines carry an introgression from KNL on chromosome 8 and 10. The region on chromosome 8 is close to the *BADH2* gene, suggesting that another allele of the *BADH2* gene might underlie 2AP synthesis

in KNL. However, four of the lines that exhibited fragrance levels equal to KNL did not have this introgression on chromosome 8.

The nitrogen precursors of the fragrant compound 2AP are proline, ornithine and arginine (Yoshihashi et al., 2002), which are further metabolised to putrescine and  $\gamma$ -amino butyraldehyde (Bradbury et al., 2008). The carbon precursor of 2AP has not been identified, but labelling with  $^{13}\text{C}$  shows that the carbon moiety is not derived from proline (Yoshihashi et al., 2002). However, many physiological amino acids and compounds are involved in the two pathways of synthesis of putrescine (Liao et al., 2008) with any one of them potentially providing the carbon moiety. These amino acids are all polar compounds and primary metabolites that are able to be detected by techniques such as non-targeted metabolomic profiling with NMR.

The primary polar compounds in the leaves of the 34 putatively fragrant BC<sub>4</sub>F<sub>2</sub> lines, and the parents are shown in Figure 5.2. Comparing the SNP map (Figure 5.1b) with the heat map of primary metabolites (Figure 5.2) produces a number of interesting associations. Genetic purity of the lines does not correlate with the metabolites. The second primary cluster in Figure 5.2 contains both parents and within that primary cluster is a secondary cluster of KNL and 5 lines and another secondary cluster of TSN 1 and 12 lines. Within the secondary clusters of the second primary cluster, there is no association with genetic purity. Four of the lines within the KNL secondary cluster in Figure 5.2 are strongly fragrant (Table 5.2) and genetically very pure (Table 5.3). This indicates that the compounds leading to fragrance in KNL could be present in these 4 lines in a background that is mainly derived from TSN 1. In the TSN cluster in Figure 5.2, there are twelve lines ranging in genetic purity (Table 5.3), and 3 of these have fragrance equal to KNL (Table 5.2). The first primary cluster contains 17 lines and four of them as strongly fragrant as KNL (Table 5.2). Consumers are able to differentiate the flavour and fragrance of KNL from that of HNN (Bounphanousay et al., 2008), suggesting that KNL has a different cocktail of compounds that constitute its taste. The first primary cluster is possibly comprised of lines that carry some of those taste compounds that are not related to 2AP and are possibly secondary flavour components. The TSN 1 secondary cluster in the second primary cluster perhaps contains lines that just carry the pathway of 2AP synthesis,

and not the pathways of some of the other compounds that define the special flavour and fragrance of KNL.

Selection of the most biologically valuable lines from the KNL set is based on the three datasets - fragrance, genetic purity and metabolite content. The lines selected for multiplication are those that are strongly fragrant from each primary and secondary cluster, and the most genetically pure. The lines, in order of genetic purity from 0.9% - 4.7% KNL, are 626, 557, 609, 664, 558, 631 and 629.

## 5.5 Conclusion

Generally, near introgression lines are reached after at least 6 generations of back-crossing. The use of the 21<sup>st</sup> century tools of metabolomic and SNP profiling enables the selection of relatively genetically pure introgression lines after only 4 generations of back-crossing. By studying a trait common to all germplasm classes, genetic purity can be reached earlier by selecting parents from different germplasm classes, which limits introgression, and leads to greater purity at earlier generations. The selected lines, the first of their type in Laos, will provide a valuable resource for identifying variability at important loci governing farmer and consumer adoption. The selected BC<sub>4</sub>F<sub>2</sub> lines from each of the HNN and the KNL set will be multiplied and used to understand the genetics of fragrance in KNL, and complete the understanding of the genetic pathway of fragrance involving the *BADH2* gene.