CHAPTER IV
CHARACTERISING PARENTS WITH NEW SCIENTIFIC TOOLS

4.1 Introduction

Rice is a one of the world's most important cereal crops, being the main source of calories and income for at least a third of the world's population (Duff, 1991; Hossain and Narciso, 2004; Khush, 1997; Renault and Facon, 2004). Over the past few decades, many rice-producing countries have gained access to export markets, bringing in significant income (Agustiniano-Osornio et al., 2005; Dodsworth, 1997). This is due to both economic development and the capacity and will to adopt the technologies and germplasm delivered by the Green Revolution.

The rice of the Green Revolution was of the non-waxy type, meaning it contains amylose. In Lao PDR, waxy rices, which do not contain amylose, are the staple food crop (Bounphanousay et al., 2000). As a result, Lao PDR did not benefit from the Green Revolution. Lao's resistance to adopting the new high-yielding varieties has preserved the diversity of their waxy rice (Delforge, 2001). However, continuing to grow the low-yielding traditional varieties significantly compromised food security until the late 1990s, at which time, breeding programs tailored to waxy rices led to the release of improved varieties such as Thadokkham1 (TDK 1) and Thasano1 (TSN 1) (Inthapanya et al., 2006). Despite higher yield and acceptable quality, these cultivars do not have the same prized quality as the two most popular traditional varieties, Kai Noy Leuang (KNL) and Hom Nang Nouane (HNN). As Lao PDR negotiates its entry into the World Trade Organisation, export opportunities are emerging for its prized traditional waxy varieties of rice. If the quality of the two traditional varieties can be combined with the agronomic adaptation of high-yielding improved varieties, such as TDK 1 or TSN 1, this could have considerable positive economic and social impact on poor Lao rice farmers.

In order to combine quality with yield, it is necessary to understand important quality traits and be able to select for them in a breeding program. A previous study

indicates that TSN 1, which is responsive to fertilizers and has better milling quality than TDK 1 (Boualaphanh et al., In Press), is a good high-yielding background. However, identifying quality traits is more complex. Current tools of quality evaluation have evolved around the more commonly consumed non-waxy types of rice, and particularly centre around the effects of amylose content (Fitzgerald et al., 2009) these tools might be inappropriate for waxy types.

Over the past few years, genotyping technology has advanced rapidly, to the point where identification of important loci is significantly easier than previously (Hyten et al., 2009; Lee et al., 2009; Newton-Cheh and Hirschhorn, 2005; Raghavan et al., 2007; Shirasawa et al., 2006; Thomson et al., In preparation; Wright et al., 2010) when an appropriate mapping population and accurate phenotyping tools are available (Tarpley and Roessner, 2007). Aside from sensory panels, there is no phenotyping tool that adequately describes the taste of rice (Champagne et al., 2010). In the last decade, technologies for the detection and identification of small molecules in plant tissue, has undergone rapid development (Hall, 2006) to the point where these 'metabolomic' tools can be utilised to reveal biochemically important metabolites that affect the taste and nutritional value of the foods we eat (Ahmed et al., 2008). A suite of plant metabolomic technologies have been developed, targeting different groups of metabolites, and bioinformatic and biostatistic approaches have been developed in concert to enable the robust combination of datasets for previously unheard of comprehensive biochemical analyses.

The objective of the present study is to take a metabolomics approach to (i) identify differences in primary metabolites, volatile compounds and minerals between the grains of TSN 1, KNL and HNN grown at four different N fertilizer regimes; (ii) determine the effect of N on the metabolic signature of each variety; and (iii) to determine whether differences in grain metabolites correlate with genetic differences between the varieties and with known differences in taste and flavour that Lao rice-consumers prize. The over-arching objective is to offer a new generation of selection tools to rice improvement programs, which will also have broader application for all crop improvement programs.

4.2 Materials and methods

4.2.1 Plant materials

Three varieties of rice, *Oryza sativa* L., were used in this study. TSN 1 is in the *indica* germplasm class and is an improved variety with valuable traits of agronomic adaptation, HNN is a traditional land race and is also *indica*, and KNL is a traditional landrace and in the *tropical japonica* germplasm class. Both have valuable traits of quality. Seeds of each were sown in the wet season of 2006 in the Lao PDR. One month later, seedlings were transplanted in a split plot randomised design of sub-plots within four main plots at the Rice and Cash Crop Research Centre, Vientiane, Lao PDR. Each main plot contained three sub-plots of each variety. Each sub-plot was 2 m x 5 m, and 10 rows of 25 plants were transplanted at a spacing of 20 cm between plants and between rows. The main plots received 30 kg ha^{-1} of P and K at transplanting. Different levels of N fertilization were applied to each main plot: 0, 30, 60 or 90 kg N ha^{-1}. For the two high N treatments, 30 kg of N was applied at transplanting and the remainder was applied in two equal top-dressings at 25 and 45 days after transplanting, and for the low N treatment, N was applied in two equal top-dressings at 25 days and 45 days after transplanting. Grain was harvested at maturity and sun-dried. Paddy from each sample was dehulled (Satake Rice Machine, Tokyo, Japan), brown grain was polished (Grainman 60-230-60-2AT, Grain Machinery Mfg. Corp., Miami, FL), and the milled rice was then dispatched to collaborators for profiling of metabolites and elements.

4.2.2 SNP genotyping

DNA was extracted from leaves of the three varieties exactly as previously described (Cuevas et al., 2010). The DNA extracts was quantified to 50 ng/ul prior to genotyping with Nanodrop 1000 (Wilmington, DE, USA). SNP genotyping was carried out using the Illumina BeadXpress (San Diego, CA, USA) at 1536 loci (Zhao et al., 2010). SNP calls were analysed using Alchemy software (Wright et al., 2010). SNP maps were generated and genetic distances were calculated using the software GGT 2.0: Graphical Genotyping (van Berloo, 2008). Euclidean distance was used as a similarity coefficient to estimate genetic distances between each variety.

4.2.3 Metabolite profiling

Milled rice was ground in liquid nitrogen and stored at -80°C. Subsamples of flour were sent to five locations for untargeted metabolite profiling. Primary metabolites were measured on polar extracts by ^{1}H-NMR in one set of extraction and operating conditions at INRA France, and another set of conditions at Rothamsted Research Centre, Harpenden, UK. Primary metabolites were also determined by GC-EI-TOF-MS at Max Planck, Golm, Germany. Volatile compounds were measured at Plant Research International, The Netherlands, and minerals were measured at University of Copenhagen, Denmark.

Primary metabolites by ^{1}H-NMR without elimination of starch residues

Replicate aliquots of rice flour (100 mg) were weighed into 1.5-mL Eppendorf tubes. D_2O–CD_3OD (1 mL, 80:20) containing 0.05% w/v TSP-d4 (sodium salt of deuterated trimethylsilylpropionic acid) was added to each sample. The contents of the tube were mixed thoroughly and heated at 50°C in a water bath for 10 min. The samples were then centrifuged for 5 min; 800 µL of the supernatant was transferred to a clean Eppendorf tube and heated at 90°C in a water bath for 2 min. The high-temperature (90°C) step was incorporated to ensure that enzyme activity had stopped. The samples were cooled at 4°C for 45 min prior to re-centrifugation for 5 min (still at 4°C); 700 µL of the supernatant was transferred to a 5-mm NMR tube.

^{1}H-NMR spectra were acquired under automation at 300 K on an Avance Spectrometer (Bruker BioSpin, Coventry, UK) operating at 600.0528 MHz and equipped with a 5 mm selective inverse probe. Spectra were collected using a water suppression pulse sequence with a pulse and a relaxation delay of 5 s. Each spectrum was acquired using 128 scans of 64K data points with a spectral width of 7309.99 Hz. Spectra were automatically Fourier transformed using an exponential window with a line broadening value of 0.5 Hz. Phasing and baseline correction were carried out within the instrument software. ^{1}H chemical shifts were referenced to -TSP-d4 at d0.00. ^{1}H-NMR spectra were automatically reduced, using Amix (Analysis of MIXtures software, *Bruker BioSpin* GmbH, Rheinstetten, Germany), to ASCII files containing integrated regions or 'buckets' of equal width (0.01 ppm). Spectral intensities were scaled to the -TSP-d4 region (δ0.05 to –0.05). The ASCII file was imported into Excel for the addition of sampling/treatment details. The regions for

unsuppressed water (d4.865–4.775), MeOH-d4 (d3.335–3.285), and TSP-d4 (d0.05 to –0.05) were removed prior to importing the data set into SIMCA-P 11.0 (Umetrics, Umea, Sweden) for multivariate analysis.

Primary metabolites by ^{1}H-NMR after elimination of starch residues

The ground rice samples were extracted as previously described (Moing et al., 2004), with slight modifications. Polar metabolites were extracted from 50 mg of lyophilised powder successively with 2 mL of ethanol/water mixtures: 80/20, 50/50 (v/v) and pure water (4 mL) for 15 min at 80°C. The supernatants were combined, dried under vacuum and lyophilised. Two technological replicates were prepared for each biological sample. The lyophilised extracts were mixed with 500 µL of 100 mM potassium phosphate buffer pH 6.0, 1 mM Ethylene diamine tetraacetic acid disodium salt (EDTA), in D_2O, titrated with KOD solution to pH 6 when necessary. In order to precipitate residual starch, nine volume of absolute ethanol (4.5 mL) were added to rice extracts (500µL), which were then mixed vigorously for 30 s. Samples were stored at 4°C for 24 h and then centrifuged at 30000 rpm for 20 min. The supernatant was then collected and dried under vacuum. Dried extracts were solubilised with 500 µL of D_2O and 5 µL of TSP-d4 solution were added. The mixture was centrifuged at 10000 rpm for 5 min at room temperature. The supernatant was then transferred into a 5 mm NMR tube for acquisition.

^{1}H-NMR spectra were recorded on a Bruker™ Avance Spectrometer (Bruker BioSpin, Wissembourg, France) at 500.162 MHz and 300 K using a 5 mm Broad Band Inverse probe (BBI). Spectra were acquired using a classical mono-dimensional liquid pulse sequence set with a 90° pulse, 32K data points, a spectral width of 6000 Hz, 20 s relaxation delay and 64 scans. The first tube was manually shimmed and analysed while the next ones were analysed using an automated 24-tube sampler (automatic shimming for 10 min). The acquired spectra were line broadened (0.3 Hz), and manually phased using TOPSPIN v1.3 software (Bruker BioSpin, Wissembourg, France) and manually baseline corrected. They were then aligned with the TSP-d4 signal at δ 0.00 ppm. Before statistical analysis, data reduction of the entire spectra (10 to 0.5 ppm with exclusion of 4.9 to 4.5 ppm leading to 456 spectra regions of 0.02 ppm width called buckets) was done, followed by normalization against total intensity. The normalized buckets were mean centered and Pareto scaled before PCA

analysis with Amix software v. 3.5 (*Bruker BioSpin* GmbH, Rheinstetten, Germany). Univariate analyses for the variety effect was performed on each bucket using Kruskal-Wallis test (FDR<0.01) with Multi Experiment Viewer software version 4.6. Resonance identification was performed by comparison with spectra of reference compounds acquired under the same conditions (own local database) and standard spiking.

Primary metabolite profiling with derivatised Gas Chromatography (GC) at MPIMP

Sample extraction, derivatisation, injection and instrumentation were exactly as previously described (Allwood et al., 2009). PCA was determined after mean-centering the data, and Pareto scaling. Discriminating compounds were determined using Kruskal Wallis analysis. Only those compounds with significant co-efficient of variation were considered discriminating.

Headspace analysis of volatile compounds by GC-MS

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-1 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 s-1. 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 s/n of 3 were randomized between 2.4 and 3 times the calculated noise value. Mass signals that were present in ≤ 6 samples were discarded. Signal redundancy per metabolite was removed by means of clustering and mass spectra were reconstructed (Tikunov et al., 2005). Metabolites were identified by matching the mass spectra of obtained metabolites to authentic reference standards and the NIST08, Wiley, and Wageningen natural compounds

spectral library and by comparison with retention indices of the literature (calculated using a series of alkanes and fitted with a third order polynomial function.

The quantitative composition of the volatile blends was evaluated by principal components analysis (PCA) and partial least squares-discriminant analysis (PLS-DA) using the software program SIMCA-P 12.0.1. (Umetrics AB, Umea°, Sweden) (Eriksson et al., 2006). The data were log-transformed and then variables were mean-centred, and Pareto scaled. PLS-DA was applied to find volatiles that discriminate between the three varieties. The number of significant latent variables was determined by cross-validation and the resulting model was validated by permutation testing. Volatile metabolites with significant positive PLS coefficients ($p<0.05$) were considered as most discriminatory.

Mineral elements at University of Copenhagen (UCPH)

The ground rice samples were digested in 100 ml closed vessels in a microwave oven (Multiwave 3000, Anton Paar, Graz, Austria) for 50 min at 210°C with a maximum pressure of 40 bars. The digestion medium consisted of 5 ml 65% HNO_3 and 5 ml 15% H_2O_2. After digestion, the samples were diluted with 3.5% HNO_3 (Hansen et al., 2009). Multi-elemental analysis was performed using ICP-MS (Agilent 7500ce, Agilent Technologies, UK) tuned in standard mode. The plasma power was operated at 1450 ± 50 W and the argon carrier and make-up gases were set at 0.82 and 0.17 l min^{-1}, respectively. Sample uptake was maintained at approximately 0.1 ml min^{-1} by a self-aspirating PFA micro-flow nebulizer. Elimination of spectral interferences was obtained by the use of an octopole ion guide with the cell gasses helium or (Laursen et al., 2009). Seven replicates of certified reference material NIST 8436 (durum wheat, particle size <200 μm; National Institute of Standards and Technology, Gaithersburgh, MD, USA) were included to validate digestion efficiency and analytical accuracy. Only data deviating less than ±10% from the certified mean reference values were retained.

4.2.4 Statistical correlations

For each set of metabolite profiling data, Euclidean distances were computed between the means of each of the clusters in the PCA score plots using *R* statistical software. The varieties in this study are expected to be homozygous at each locus, enabling the genetic distance between the genotype to be expressed as the Euclidean

distance on the basis of the 1536 SNP loci. GGT 2.0 software (van Berloo, 2008) was used to determine genetic distance. Using the Euclidean distances, Procrustes analysis was carried out to determine the correspondence between the group means on the first two principal components of each set of metabolite data and genetic variability.

4.3 Results

4.3.1 Genetic differences between varieties

Three varieties were studied: improved TSN 1 and traditional HNN varieties from the *indica* germplasm class, and traditional KNL from the *tropical japonica* germplasm class. Genome-wide genotyping of single nucleotide polymorphisms (SNPs) at 1536 loci (Figure 4.1) shows that genetic similarity between the traditional and improved *indica* is about 80%, whereas KNL differs from both the *indica* varieties by about 50%. However, it can be seen in Figure 4.1 that the differences between KNL and HNN are not at the same loci as the differences between KNL and TSN 1. The Euclidean distance between each genotype reflects the genetic similarity and is shown in Table 4.3.

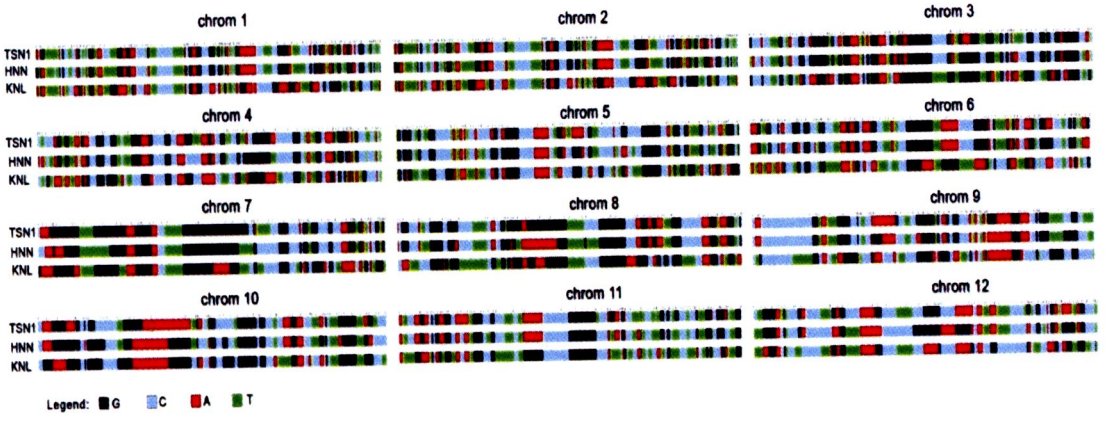

Figure 4.1 Genetic variability between TSN 1, KNL and HNN rice varieties at 1536 SNP loci across all 12 chromosomes. At these loci, the genetic similarity between TSN 1 and KNL is 53.2% and between TSN 1 and HNN is 80.7% and between KNL and HNN is 53.6%.

4.3.2 Discriminating between the quality of each variety

The physical quality of the grains from each variety differs. KNL grains are long and bold, whereas those of TSN 1 and HNN are long and slender. A set of standard quality evaluation data, usually obtained to indicate cooking and sensory quality, was previously determined for in triplicate for each of the three varieties grown in each N treatment (Boualaphanh et al., In Press). The set includes tests of gelatinisation temperature; data derived from viscosity curves, in this case breakdown, setback and retrogradation (Fitzgerald et al., 2003); and hardness and stickiness of the cooked grains. Principal Components Analysis (PCA) was performed on this dataset of six quality evaluation variables after mean-centring and Pareto scaling. PC1 and PC2 together explain 98% of the variability, but examination of the score plot (Figure 4.2A) indicates that the explained variability is unrelated to genotype or to N treatment.

4.3.3 Discriminating between the grain metabolome of each variety

An array of metabolite profiling technologies was applied to the analysis of flour samples from the three varieties grown under different nitrogen fertilizer regimes in order to obtain an overview of the compositional similarities and differences between grain samples from each variety. PCA was used to analyse the profiling data from each analytical technology, and univariate analysis was used to identify compounds that explained the clusters on the PCA score plot. Data was analysed to determine if compounds of taste and flavour could be found that could provide insights into the sensory differences between the varieties. Figures 4.2B-F show PCAs of profiling data using ICP-MS of minerals (Figure 4.2B), ^{1}H-NMR of polar metabolites under two different operating conditions (Figure 4.2C-D), GC-TOF-MS of polar metabolites (Figure 4.2E), and GC-MS of volatile compounds (Figure 4.2F) of the grains of the three varieties and four N treatments.

First, 20 mineral elements were quantified in the polished grains of each variety from each N treatment and PCA of the full set of data was performed. Observation of the scores plot (Figure 4.2B) shows that sample clustering was based on genotype but the minerals in the grains did not lead to clustering on the basis of N treatment. PC1 explains 29% of the variation and PC2 explains 22%. Observation of the loadings plot (data not shown) showed that the minerals discriminating KNL were

the micronutrient minerals Fe, Cu and Zn and that TSN 1 samples were characterized by higher concentrations of several macro minerals.

Major polar metabolites were estimated using ^{1}H-NMR profiling in two different conditions where the major metabolites could be identified. First, rapid ^{1}H-NMR profiling of polar extracts was performed. PCA was done using a matrix containing scaled and integrated data from 950 spectral regions of 0.01 ppm (Figure 4.2C). The first principal component (PC1) of the score plot (Figure 4.2C) explained 81% of total variability and clearly separated HNN from the other two genotypes. The second principal component (PC2), explained 6% of total variability, and clearly separated samples of TSN 1 from KNL (Figure 4.2C). The loadings plot (summarized in Table 4.1) (shows that separation of HNN along PC1 is due to elevated sucrose, asparagine, threonine, lactate, and reduced glycerol and two compounds unable to be identified. Along the PC2 axis, KNL was characterized by elevated sucrose and glycine, and decreased glucose, maltose, asparagine, threonine and choline contents. Figure 4.2C also shows that the major polar metabolites separated the N treatments, especially the lowest and highest, along the PC1 axis, but the samples of different N fertilizer regimes all fell within each genotype cluster.

The data obtained from rapid ^{1}H-NMR profiling were confirmed using ^{1}H-NMR signatures of polar extracts after elimination of residual starch. PCA was performed on these signatures using a matrix containing the data from 456 spectral regions of 0.02 ppm width. PC1 of the score plots (Figure 2D), explaining 63% of the total variability, clearly separated HNN and TSN 1 from KNL. Observation of the loadings plot (in Table 4.1) suggested that the major metabolites contributing to this separation along PC1 were sucrose and raffinose on the negative side and mainly lipids, betaine and choline on the positive side. PC2, explaining 18% of the total variability, separated TSN 1, on the negative side, from the other two varieties (Figure 4.2D). The major metabolites contributing to this separation along PC2 were choline on the negative side and lipids on the positive side. Major and minor differences between varieties were then checked using Kruskal-Wallis test (False Discovery Rate (FDR) <0.01) on each spectra region of 0.02 ppm width. One hundred and eighty-seven spectra regions out of the 456 spectra regions showed a significant effect of genotype (data not shown). HNN grains did not seem to be affected by N regime

while TSN 1 and KNL showed a clear separation between the lowest and highest N levels in the PC1xPC2 plane but the samples of the different N regimes all fell within each variety cluster (Figure 4.2D). However, no common effect of N regime was observed for the three varieties. Therefore the effect of N on the polar metabolites determined by ^{1}H-NMR after elimination of starch residue was less clear (Figure 4.2D) than with rapid ^{1}H-NMR profiling without elimination of starch residue (Figure 4.2C).

Since several overlapping resonances were observed in the ^{1}H-NMR spectra, and GC-MS is more sensitive than ^{1}H-NMR, the data in Figures 4.2C and D were complemented with GC-EI-TOF-MS of polar extracts. PCA was performed using 94 variables, and samples of each variety clustered together (Figure 4.2E). PC1 accounted for 50% of the variation, and along that axis, samples of KNL were separated from those of TSN 1. Along the PC1 axis, HNN and TSN 1 were not clearly separated, but the PC2 axis, accounting for 25% of the variation, separated the HNN samples (Figure 4.2E). Univariate analyses show that the main compounds driving the separation along PC1 were cysteine, pyroglutamic acid, ribonic acid, glycerol, threonine, putrescine, tyrosine and trehalose (Table 4.1). The major metabolites that separated HNN along the PC2 axis are threitol, arabinonic acid, proline, azelaic acid, glycerol, fumaric acid, and allantoin (Table 4.1). Many of the other compounds leading to the separation along the PC axes are sugar alcohols and hexose sugars (Table 4.1).

Volatile compounds also affect taste and flavour and these were determined for each variety using GC-MS. PCA analysis shows volatile compounds separated the three varieties (Figure 4.2F). PC1 explained 34% of the variability, and KNL and HNN were clearly separated along the PC1 axis. PC2 explained 19% of the variability and clearly separated the samples of TSN 1 from the other two (Figure 4.2F). Discriminating compounds, explaining the separation, were those with significant PLS coefficients. The three varieties were separated by 54 compounds, though not all could be identified by the spectral libraries. The major discriminating compound was 2AP, which separated TSN 1 from the other two, and the elevated concentration of 2AP in KNL separated those samples from HNN. Table 4.1 shows the compounds of significant correlation coefficients for each of the varieties. HNN grains are

characterised by ketones, KNL grains by 2AP and several alcohols, and TSN 1 grains, by three volatiles compounds that could be identified (Table 4.2) and twelve that could not.

Volatile compounds also affect taste and flavour and these were determined for each variety using GC-MS. PCA analysis shows volatile compounds separated the three varieties (Figure 4.2F). PC1 explained 34% of total variability, and KNL and HNN were clearly separated along PC1 axis. PC2 explained 19% of total variability and clearly separated the samples of TSN 1 from the other two varieties (Figure 4.2F). Four significant principal components were extracted, cumulatively explaining 77% of variation. Discriminating compounds, explaining the separation, were those with significant PLS coefficients. The three varieties were separated by 54 compounds, though not all could be identified by the spectral libraries. The major discriminating compound was 2AP, which separated TSN 1 from the other two varieties, and the elevated concentration of 2AP in KNL separated those samples from HNN. Table 4.2 shows the compounds of significant PLS coefficients for each of the varieties. HNN grains are characterised by ketones, KNL grains by 2AP and several alcohols, and TSN 1 grains, by only three volatiles compounds that could be identified (Table 4.1) and twelve that could not. Several of the discriminating volatile compounds unique to each variety have quite low odour thresholds and unique flavours (Table 4.2).

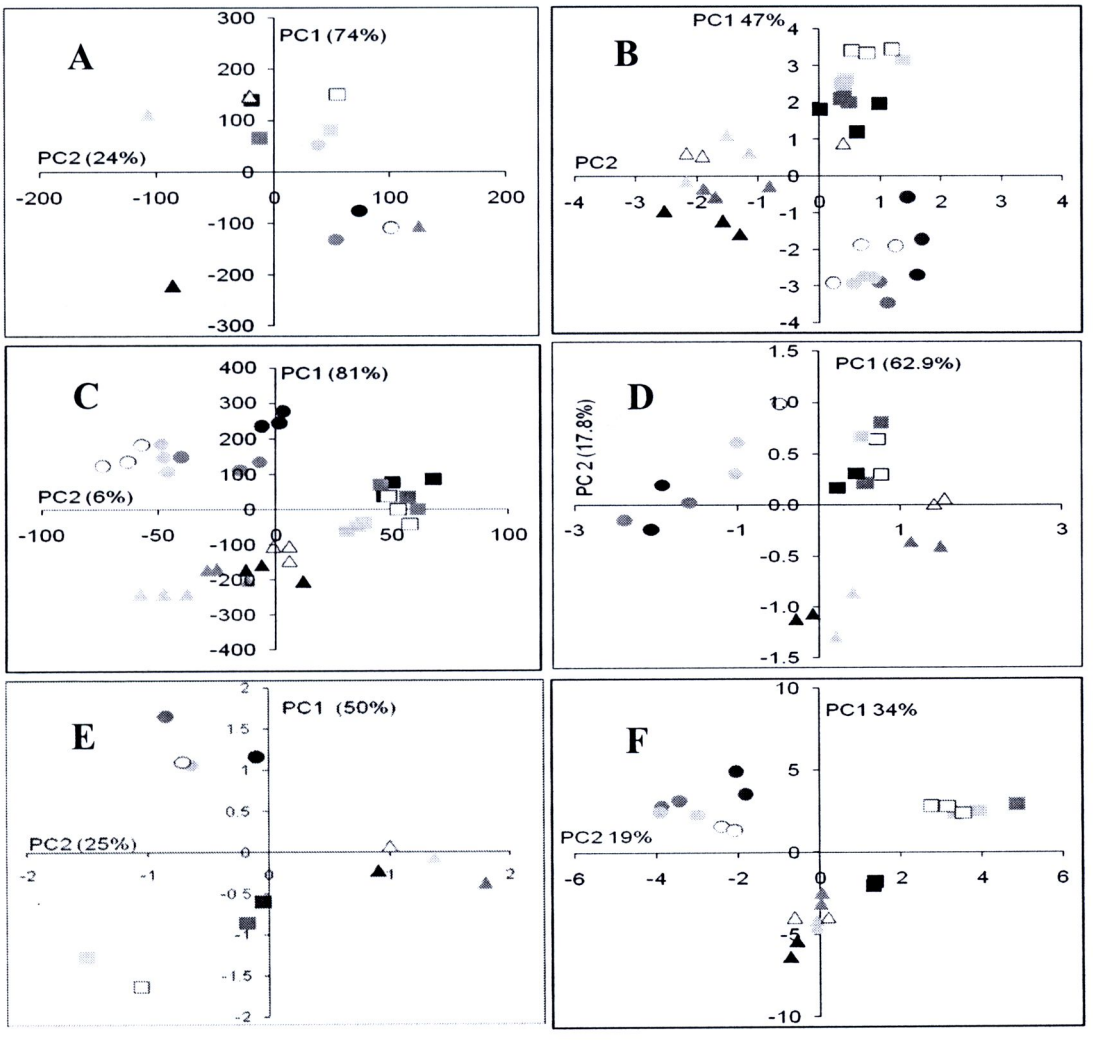

Figure 4.2 PCA score plots showing discrimination between three rice varieties cultivated under four different N regimes on the basis of (A) routine grain quality traits, (B) mineral micronutrients determined by ICP-MS, (C) polar metabolite profiling with 1H-NMR rapid fingerprints or (D) 1H-NMR fingerprints after elimination of starch residue, (E) polar metabolites determined by GC-TOF-MS, (F) volatiles determined by SPME GC-MS. Genotypes: HNN, squares; KNL, circles; TSN 1, triangles. N regime: 0 kg N ha^{-1}, unfilled; 30 kg N ha^{-1}, light grey; 60 kg N ha^{-1}, dark grey; 90 kg N ha^{-1}, black

Table 4.1 Polar compounds discriminating HNN, KNL or TSN 1 detected by rapid ^{1}H-NMR[a], ^{1}H-NMR after elimination of starch residue[b] or derivatised GC-TOF-MS[c] of extracts of polished grains +/++ indicate the tendencies between varieties basedon Kruskal-Wallis test.

Compounds	HNN	KNL	TSN 1
Lipid compounds			
Lipids[b]	++	+	++
Sugars			
Glucose[b]	+	+	++
Sucrose[b]	+	++	+
Sucrose+raffinose[b]	+	++	+
Fructose[c]			+
Sugar alcohols			
Tagatose[c]	++		+
Erythritol[c]			+
Mannitol[c]			+
Threitol[c]			+
Galactitol[c]			+
Xylitol[c]		++	
Organic acids			
Malate[b]	++	+	++
Fumarate[b,c]	++	+	++
Succinate+unknown[b,c]	++	+	++
Azelaic Acid[c]	++		

Table 4.1 Polar compounds discriminating HNN, KNL or TSN 1 detected by rapid ^{1}H-NMR[a], ^{1}H-NMR after elimination of starch residue[b] or derivatised GC-TOF-MS[c] of extracts of polished grains. +/++ indicate the tendencies between varieties basedon Kruskal-Wallis test (Cont.).

Compounds	HNN	KNL	TSN 1
Amino acids and amino compounds			
Alanine+lipids[b]	++	+	++
Asparagine+aspartate[b]	++	+	++
Betaine+unknown[b]	+	+	++
Choline[b]	+	+	++
GABA[b]	+	+	++
GABA+valine+unknown[b]	++	+	++
Putrescine[c]	++	++	-
Valine + lipids[b]	++	+	++

Table 4.2 Volatile compounds that discriminated each variety, their odour thresholds and flavour. The odour/flavour of each compound and its threshold in water is also shown. Stars indicate compounds that are used commercially at 5 ppm or less to add flavour or fragrance to manufactured foods or perfumes. +/++ indicate the tendencies between varieties based on variable importance of prediction (VIP calculated from PLS-DA data).

Compounds	Odour	Threshold	HNN	KNL	TSN 1
Ketones					
2-acetyl-1-pyrroline	Popcorn	0.1	+	++	
2-Heptanone	Fruity	140	++		
2-Octanone	Herbal	50	++		
2-Hexanone			++		
2,3 heptadione	cheese*		++		
3-Octen-2-one	berry*		++		
2,3-Octanedione	dill*		++		
3,5-Octadiene-2-one	Herby		++		
Aldehydes					
Hexanal	Grassy	5	++		
Pentanal	floral	12	++		
2-heptenal	Fatty	13	++		
2-octenal	Herby	3	++		
2-hexenal	Fruity	17	++		
2-butyl 2-octenal	green tea	2	++		
Benzaldehyde	Almond	350	++		
Alcohols					
1-octen-3-ol	Mushroom	1	++		
1-pentanol	Plastic	4000	++		
1-heptanol	Citrus*			++	
1-hexanol	grassy*	2500		++	
Ethanol				++	

Table 4.2 Volatile compounds that discriminated each variety, their odour thresholds and flavour. The odour/flavour of each compound and its threshold in water is also shown. Stars indicate compounds that are used commercially at 5 ppm or less to add flavour or fragrance to manufactured foods or perfumes. +/++ indicate the tendencies between varieties based on variable importance of prediction (VIP calculated from PLS-DA data) (Cont.).

Compounds	Odour	Threshold	HNN	KNL	TSN 1
1-octanol	Citrus	110		++	
2,3-butanediol	creamy*				++
2-ethyl-1-hexanol	Rose				++
Hydrocarbons					
Pentadecane	waxy*		++		
Undecane	herbal*			++	
Longicyclene	Floral				++
dl-Limonene	Citrus	10		++	
1-ethyl-3-methyl benzene				++	
Furans					
2-propylfuran	Fruity	6000	++		
2-butylfuran	Wine	10000	++		
2-pentylfuran	Beany	2000	++		
Carboxylic acids					
nonanoic acid	Rancid	3000		++	
Esters					
ethyl benzoate	Cherry	60		++	

4.3.4 Relating the metabolome of the grain with the genome

For each metabolite profiling platform, Euclidean distances between varieties were calculated using the first two principal components (Table 4.3). Based on 1536 SNP loci of the three varieties, the genetic distance between HNN and TSN 1 was

smallest, as expected since they are both *indica*, and was the similar between KNL and HNN and between KNL and TSN 1 (Table 4.3). The residual sum of squares from Procrustes analysis of the association between Euclidean distances based on the genome and metabolome indicates an almost perfect association between genotype and each metabolite profile (Table 4.3).

Table 4.3 Procrustes analysis of correspondence between Euclidean distances between varieties from each metabolite platform, based on principal components 1 and 2, and Euclidean genetic distance between each variety. Procrustes rotation is calculated as the residual sum of squares scaled so that the total sum of squares is 1.

	Primary polar metaboliltes			Volatiles	Minerals	Genes
	GC -TOF-MS	H^1NMR1	H^1NMR2	GC-MS	ICP-MS	SNPs
Pairs	Euclidean distance					
TSN 1 – HNN	2.675	249.008	1.288	8.057	4.128	12.1
HNN – KNL	2.866	216.273	2.558	7.353	5.961	18.1
TSN 1 – KNL	2.822	426.875	2.936	9.139	4.206	18.4
	Procrustes rotation					
	0.041	0.199	0.024	0.061	0.063	

4.4 Discussion

All rice improvement programs select on the basis of quality, but the present study shows that classic quality evaluation data, obtained previously (Boualaphanh et al., in press), does not discriminate between the genotypes (Figure 4.1F). This finding indicates either that (i) the cooking and sensory properties of the three are similar, or (ii) that the tools of quality evaluation, developed for non-waxy rice, cannot discriminate between waxy types. Bounphanousay (2007) investigated the sensory and cooking properties of KNL and HNN using a large panel of Lao farmers, and

found significant differences between the two varieties, indicating a need to understand the differences, in order to develop ways to screen for them in a quality evaluation program.

In the Lao PDR, as in many rice-consuming countries, the taste and flavour of the rice is considered equally, and possibly more important than the texture (Schiller et al., 2006). Tools to measure taste and flavour are not routinely used in any rice quality program, but the science of metabolomic profiling is developing rapidly and offers new opportunities to begin to collect and test such data. Metabolites are small compounds from biochemical processes in organisms, and differ widely in chemical nature in terms of their solubility, polarity and volatility (Hall, 2006). Many such small molecules are known to correlate with or determine key agronomic and organoleptic traits in plant products. In the present study, volatile and polar primary metabolites were targeted since these compounds are the ones most likely to be detected in the mouths and noses of rice consumers.

PCA demonstrated that the varieties could be separated by most metabolite profiling platforms (Figure 4.2, indicating that each variety has a unique metabolite signature. The relative variation in detected compounds between each variety differs. Many of the compounds in the second half of the citric acid cycle could be identified by GC-TOF-MS in both TSN 1 and HNN, but not in KNL (Table 4.1), and thus appear to discriminate KNL (Figure 4.2E). This pattern might also reflect the genetic relationships between the three varieties (Figure 4.1) – with greater genetic separation between KNL and the two *indica* varieties. On the basis of volatile compounds (Table 4.2), 2-acetyl-1-pyrroline (2AP) was the most discriminating volatile compound, closely followed by 2,3-heptadione, which was detected in HNN. The latter compound is an important food flavour and is used commercially to impart strong buttery and fruity notes to dairy foods and to fragrances. That 2AP comes out on top after these untargeted metabolomics analyses is not surprising and confirms much previous evidence from targeted analyses (Buttery et al., 1983a; Fitzgerald et al., 2008; Itani et al., 2004; Wilkie et al., 2004). However, there were 54 volatile compounds that together discriminated between the varieties (Table 4.2), and a number had relatively low odour thresholds, indicating that aroma of rice is a complex trait. The results from all platforms indicate that compounds and minerals that

discriminated between varieties, forming the unique metabolomic signatures, are compounds with nutritional value or are compounds that are likely to contribute to taste, flavour and aroma.

4.4.1 Nutritionally valuable compounds

There are many metabolites that occur in the milled grains which have attributed nutritional benefits (Table 4.1). Examples are GABA, found in TSN 1, which is involved in a host of cortical functions in mammals (Sanacora et al., 1999) and is involved in signalling and osmotic regulation in plants (Bouché and Fromm, 2004). Choline was detected in KNL and TSN 1 and is involved in brain health and liver function (Zeisel et al., 1991). Glucosamine was found in HNN and is known to contribute to the formation of cartilage and is used to alleviate and prevent pain due to arthritis and bone stress (Hughes and Carr, 2002). Azelaic acid, found in HNN, is an organic compound that contributes to the health of hair and skin (Nguyen and Bui, 1995).

Improving the amount of the micronutrients Fe and Zn in polished rice is the subject of a large international effort under Harvestplus to address hidden hunger. In the present study, the higher concentration of these in KNL discriminated between it and the other two varieties (Table 4.3). Consistent with this, KNL was found to have the highest concentration of Fe, and almost the highest Zn, in another study that measured a number of minerals in polished grains of 56 Lao varieties (Bounphanousay, 2007). The concentration of Fe and Zn in KNL approaches the ambitious targets set by Harvestplus. Consequently, bearing this important trait means KNL would be a useful starting point for further varietal improvement where organoleptic properties are to be combined with enhanced nutritional properties.

Waxy rices are also known as sweet rices. Table 4.1 indicates why this might be so. All three varieties contain various compounds with high sweetness indexes. KNL contains higher levels of sucrose while the other sweet flavours in the three varieties are hexose sugars such as tagatose and psicose, and sugar alcohols such as xylitol and erythritol (Table 4.1). The hexose sugars and sugar alcohols are of similar sweetness to sucrose, but they provide only a fraction of the energy because they are all absorbed in the large intestine. These sweeteners do not induce a glycaemic response to consuming the rice, which could explain some of the variability found in

the glycaemic index of different rices (Wheeler and Pi-Sunyer, 2008) and contribute to the management of diseases like diabetes (Association, 2004).

It is generally assumed that polished rice provides mostly calories, and that the bran layer contains all the compounds of nutritional benefit (Butsat and Siriamornpun, 2010; Heinemann et al., 2008; Sakamoto et al., 2007; Sharif and Butt, 2006; Shen et al., 2009; Yokoyama, 2004; Yu et al., 2007). This is the first demonstration of the range of compounds that occur in polished rice grains, many of which have a role in human physiology and the potential to contribute to human health.

4.4.2 Compounds of taste and flavour

Consumers have difficulty describing the taste of rice. This is why testing for taste is generally not part of standard rice quality programs (Champagne et al., 2010; Fitzgerald et al., 2009). However, in a recent study of the taste and flavour of the polished grains from the major varieties grown throughout South and Southeast Asia, a characteristic flavour profile emerged for those varieties commonly grown and consumed in Southeast Asia consisting of sweet, floral, grassy and dairy notes (Champagne et al., 2010).

KNL and HNN contain polar and volatile compounds that give sweet, floral, fruity and grassy flavours (Tables 4.1 and 4.2). They both also contain unpleasant compounds (off-flavours) such as putrescine, which was first found in putrefying flesh (Olle, 1986), but which is considered to be a precursor in the pathway of aroma in rice (Bradbury et al., 2008). Putrescine was not found in TSN 1 (Table 4.1), consistent with a possible role in 2AP synthesis in the two fragrant rices (Bradbury et al., 2008). It is unlikely that a variety would persist in any country if it smelt or tasted of rotten flesh after cooking, which suggests either that the putrescine could be lost during preparation of the rice, or its effect could be negated by all the floral and fruity volatile compounds that are found in these two varieties (Table 4.1), or the odour threshold of 100 ppm (Christiansen et al., 2008) is too high for detection.

The three varieties differed significantly in their volatile profiles (Table 4.2). HNN and KNL both contained 2AP, as would be expected considering its known role as the key determinant of the fragrance trait (Buttery et al., 1983b). Aside from 2AP, HNN contained a number of ketones. Many of these are commercially available as pure compounds and are used to impart fruity, nutty, floral and butter/dairy aromas

and flavours in foodstuffs. KNL contained grassy compounds that are commercially used to impart fresh grassy fresh. In contrast, the non-fragrant TSN 1 contained few volatile compounds (Table 4.2), corresponding to minimal aroma. Only a few of the volatile compounds detected have an odour threshold that would lead to detection in humans (Table 4.2), based on the detection limit of the GCMS. The strongest of these is 2AP, but some of the fruity and nutty compounds found in HNN also have low odour thresholds (Table 4.2). Moreover, many of the volatile compounds are commercially available and those noted are commonly used at levels below 10 ppm, indicating that a possible role in the taste and flavour of rice. This suggests that HNN might have a more complex flavour profile, whereas KNL has a clean popcorn flavour, with 2AP being dominant and having a sufficiently low odour threshold for human detection. This finding is consistent with the previous work showing that KNL and HNN can be differentiated by consumers in taste trials (Bounphanousay, 2007).

Further analyses are required to correlate the current findings to the situation with cooked rice. The presence of amino acids and reducing sugars among the primary metabolites in the raw grains of each variety leads to the possibility that Maillard reaction products may also be produced during cooking (De Kimpe and Keppens, 1996) leading to other contributions to flavour. Cysteine was found in KNL (Table 4.1) and it is an efficient precursor of α-acetyl N heterocycles, which have a very low odour threshold and contribute to the roasted aroma of cooked rice (Kerler et al., 1997).

While three volatile compounds that were characteristic for TSN 1 could be identified (Table 4.2), another 12 compounds remained unidentified. In the other two varieties, there were also similar numbers of discriminating volatiles that could not yet be identified. Therefore more research is needed to determine fully the volatile cocktail, and the odour threshold of each component in order to build up a more complete picture of flavour and fragrance of rice.

All the metabolite platforms separated the varieties. By comparing the triangle derived from the Euclidean distance between genotypes, with those derived from each metabolite platform (Figure 4.2). Procrustes analysis shows that each metabolite triangle fits very well onto the genotype triangle, indicating that metabolite profiles associate with the genotypes. While this study is of only three genotypes, the results

strongly indicate that metabolomic profiling is an excellent phenotyping tool for association mapping. Such associations could lead to the discovery of new alleles for sensory quality. This could lead to an expanded palette of both DNA and chemical markers for such compounds would greatly enhance the capacity of rice improvement programs to select actively for delicious and nutritious rice.