

Multi-platform metabolomics analyses of a broad collection of fragrant and non-fragrant rice varieties reveals the high complexity of grain quality characteristics

R. Mumm^{1,3} · J. A. Hageman⁴ · M. N. Calingacion^{2,5,21} · R. C. H. de Vos^{1,3,6} ·
H. H. Jonker^{1,3} · A. Erban⁷ · J. Kopka⁷ · T. H. Hansen⁸ · K. H. Laursen⁸ ·
J. K. Schjoerring⁸ · J. L. Ward⁹ · M. H. Beale⁹ · S. Jongee¹⁰ · A. Rauf¹¹ ·
F. Habibi¹² · S. D. Indrasari¹³ · S. Sakhan¹⁴ · A. Ramli¹⁵ · M. Romero¹⁶ ·
R. F. Reinke^{17,22} · K. Ohtsubo¹⁸ · C. Boualaphanh^{19,20} · M. A. Fitzgerald^{2,21} ·
R. D. Hall^{1,3,5,6}

Received: 30 June 2015 / Accepted: 18 September 2015 / Published online: 23 January 2016
© The Author(s) 2016. This article is published with open access at Springerlink.com

Abstract The quality of rice in terms not only of its nutritional value but also in terms of its aroma and flavour is becoming increasingly important in modern rice breeding where global targets are focused on both yield stability and grain quality. In the present paper we have exploited advanced, multi-platform metabolomics approaches to determine the biochemical differences in 31 rice varieties

from a diverse range of genetic backgrounds and origin. All were grown under the specific local conditions for which they have been bred and all aspects of varietal identification and sample purity have been guaranteed by local experts from each country. Metabolomics analyses using 6 platforms have revealed the extent of biochemical differences (and similarities) between the chosen rice genotypes. Comparison of fragrant rice varieties showed a difference in the metabolic profiles of jasmine and basmati varieties. However with no consistent separation of the germplasm class. Storage of grains had a significant effect on the metabolome of both basmati and jasmine rice varieties but changes were different for the two rice types. This shows how metabolic changes may help prove a causal relationship with developing good quality in basmati rice or incurring quality loss in jasmine rice in aged grains. Such

This article is dedicated to Supanee Jongee, who contributed greatly to breeding for quality rice in Thailand and who sadly passed away before this paper was published.

R. Mumm, J. A. Hageman, M. A. Fitzgerald, and R. D. Hall have contributed equally to this work.

Electronic supplementary material The online version of this article (doi:10.1007/s11306-015-0925-1) contains supplementary material, which is available to authorized users.

✉ R. Mumm
roland.mumm@wur.nl;
http://www.wur.nl

¹ Plant Research International, Wageningen University and Research Centre, Droevendaalsesteeg 1, Wageningen, The Netherlands

² Grain Quality, and Nutrition Centre, International Rice Research Institute, DAPO 7777, Metro Manila, Philippines

³ Centre for BioSystems Genomics, P.O. Box 98, 6700 AB Wageningen, The Netherlands

⁴ Biometris-Applied Statistics, Wageningen University and Research Centre, Droevendaalsesteeg 1, Wageningen, The Netherlands

⁵ Laboratory of Plant Physiology, Wageningen University and Research Centre, Droevendaalsesteeg 1, Wageningen, The Netherlands

⁶ Netherlands Metabolomics Centre, Einsteinweg 55, 2333 CC Leiden, The Netherlands

⁷ Max-Planck-Institute of Molecular Plant Physiology (MPIMP), Am Mühlenberg 1, 14476 Potsdam-Golm, Germany

⁸ Plant and Soil Science Section, Department of Plant and Environmental Sciences, Faculty of Science, University of Copenhagen (UC), Thorvaldsensvej 40, 1871 Frederiksberg C Copenhagen, Denmark

⁹ The National Centre for Plant and Microbial Metabolomics, Rothamsted Research, West Common, Harpenden, Herts AL52JQ, UK

¹⁰ Ubonratchathani Rice Research Centre, Ubon Ratchathani, Thailand

¹¹ Rice Programme, National Agricultural Research Centre, Islamabad, Pakistan

metabolomics approaches are leading to hypotheses on the potential links between grain quality attributes, biochemical composition and genotype in the context of breeding for improvement. With this knowledge we shall establish a stronger, evidence-based foundation upon which to build targeted strategies to support breeders in their quest for improved rice varieties.

Keywords Rice · Multi-platform analysis · Jasmine · Basmati · Grain quality · PLS-DA

1 Introduction

Rice is not only the most important global food crop, it is also a major export commodity and source of income for many developed and developing countries (Khush and Juliano 1985). In particular, fragrant rice types such as basmati, produced in India and Pakistan, and jasmine which is widely cultivated in S.E. Asia, are specifically favoured not only locally, but also outside their production regions where they find major and increasing markets, particularly in Europe, N. America and Australia. For example, the volume of Thai rice imported into the EU increased more than three fold in the period 2000–2006 (Supakornchuwong and Suwannaporn 2012) and in Australia, imports from Thailand are required to meet a very large shortfall in fragrant rice each year. The global market for basmati export was worth \$US1 billion in 2007 (Archak et al. 2007) and has since continued to rise. These two types of fragrant rice command the highest prices on the international market, being up to three times more expensive than high quality, non-fragrant rice (www.oryza.com).

The basmati rices of the highest quality are single line selections from traditional varieties of the Punjab region,

shared by India and Pakistan, and are in the *aromatic* germplasm class (McNally et al. 2009). The high quality jasmine varieties, such as Khao Dawk Mali (KDML) 105 from Thailand, are also single line selections, but from traditional *indica* varieties and were domesticated in the Greater Mekong Sub region (GMS) (McNally et al. 2009). Aroma originated in the *aromatic* germplasm class (Kovach et al. 2009) and introgression into *indica* varieties occurred probably as a result of hybridisation during domestication. Traditional varieties are generally low yielding, and many countries, including Thailand, Pakistan and India, invest heavily in programs attempting to combine the quality of these aromatic rices with superior agronomic adaptation (Fitzgerald et al. 2009). Fragrant rice is also found in the other germplasm classes, but many of these are improved varieties which tend to be consumed predominantly in the countries where they have been developed and cultivated.

A major limitation for breeding programs aiming to increase the yield potential of fragrant rice is a lack of understanding of the particular fragrance that is characteristic of basmati and jasmine rices. All fragrant rice varieties contain the fragrant compound 2-acetyl-1-pyrroline (2AP) (Buttery et al. 1982), and the genetics driving 2AP accumulation are now well known (Bradbury et al. 2005; Kovach et al. 2009). Currently, 2AP is the only component that rice breeders can select for, either phenotypically or genetically (Bryant and McClung 2011). However, it has already been recognised that the importance of 2AP to rice fragrance has been over-emphasised relative to the many other (unknown) contributory compounds (Limpawattana et al. 2008). More than 100 other volatile compounds have been found in rice grains (Buttery et al. 1983, 1988), each of which has the potential to influence the sensory experience. Even compounds present at very low levels can still impart significant fragrant/

¹² Grain Quality Division, Rice Research Institute of Iran (RRII), Km 5 Tehran Rd, 41996-13475 Rasht, Islamic Republic of Iran

¹³ Indonesian Center for Rice Research (ICRR) BB Padi, Jl. Raya 9, Sukamandi, Subang 41256, Jawa Barat, Indonesia

¹⁴ Cambodian Agricultural Research and Development Institute, CARDI Rd, Phnom Penh, Cambodia

¹⁵ Pusat Penyelidikan Padi dan Tanaman Industri, MARDI, Seberang Perai Beg Berkunci 203 Pejabat Pos Kepala Batas, 13200 Seberang Perai Pulau, Penang, Malaysia

¹⁶ Rice Chemistry and Food Science Division, Philippine Rice Research Institute, Maligaya, Science City of Muñoz 3119, Nueva Ecija, Philippines

¹⁷ Graham Centre for Agricultural Innovation, Agricultural Institute (An Alliance Between NSW Department of Primary Industries and Charles Sturt University), Wagga Wagga, NSW, Australia

¹⁸ Department of Applied Biological Chemistry, Faculty of Agriculture, Niigata University, Niigata, Japan

¹⁹ Department of Plant Science and Agricultural Resources, Faculty of Agriculture, Khon Kaen University, Khon Kaen 40002, Thailand

²⁰ Rice and Cash Crops Research Centre, National Agriculture and Forestry Research Institute, PDR, Vientiane, Lao

²¹ School of Agriculture and Food Science, University of Queensland, St Lucia, QLD 4072, Australia

²² Present Address: Plant Breeding, Genetics and Biotechnology Division, International Rice Research Institute, DAPO 7777, Metro Manila, Philippines

flavour notes when they have low odour thresholds, as is the case for 2AP (Champagne 2008). Rice fragrance and flavour are highly complex sensory attributes, involving a wide range of chemical compound classes. Our previous work has shown that the grains from aromatic rices from Southeast Asia, South Asia and Central Asia differ significantly both in taste (Champagne et al. 2010), and in the metabolomic composition of the grains (Calingacion et al. 2014).

Metabolomics is an analytical technology focused on profiling myriad of small metabolites present in biological samples (Hall 2006; Saito et al. 2006). As an untargeted approach, it has the potential to play a key role in defining better what is meant by grain quality, and the technology can help identify those components which either positively or negatively influence sensory characteristics without any great requirement for prior knowledge (Stewart et al. 2011). A range of approaches has been developed, generally centred around combinations of separation (e.g. gas chromatography; liquid chromatography; capillary electrophoresis) and detection (e.g. mass spectrometry; nuclear magnetic resonance) techniques (Hall and Hardy 2012; Hall 2011). Following on from some of the more traditional approaches, metabolomics is now also becoming established as a valuable technology to evaluate rice quality (Calingacion et al. 2012; Fitzgerald et al. 2009).

Uniquely, in the study described here, we have deliberately sourced a wide range of fragrant and non-fragrant rice varieties specifically from their countries of origin through the International Network for Quality Rice (INQR). These materials have been produced by INQR members under the cultivation conditions for which each specific variety has been bred. We have therefore been able to avoid the potential limitations of growing all varieties at a single location which would have entailed that most would have been cultivated under unadapted conditions. Furthermore, in contrast to many previous reports, usually based on store-bought materials, we can guarantee that pure samples of confirmed varietal identification have been used in the present study. We have also had full knowledge and control of the post-harvest handling of the grains (storage, transport and polishing) which can also greatly affect final sensory quality. Within the EU consortium META-PHOR (www.meta-phor.eu) we have had access to six contrasting but complementary analytical platforms capable of profiling a broad range of small molecules and mineral micronutrients (Hall 2007). These platforms have already demonstrated their value by providing new insights into both the primary and secondary metabolite profiles of rice breeding materials (Calingacion et al. 2012) and have the potential to identify biochemical markers appropriate for use in breeding programmes for different types of fragrant rice (Calingacion et al. 2014).

The objectives of this study were to investigate: (i) differences in the metabolomic signature of fragrant rice from each germplasm class, (ii) the effect of storage on both jasmine and basmati rices, and (iii) differences between improved and traditional varieties of both basmati and jasmine rices. The results clearly indicate both the complexity and the dynamism of rice metabolomic profiles and indeed, emphasise how much we still do not know about which metabolites are involved in c.q. causal of, key quality traits.

2 Materials and methods

2.1 Plant materials

Thirty one rice (*Oryza sativa* L.) varieties were produced in their local environments and resulted in a total of 44 samples (Table S1). Samples included fragrant varieties from the *aromatic*, *sadri*, *indica*, *tropical japonica*, and *temperate japonica* germplasm classes (Garris et al. 2005; McNally et al. 2009), as well as non-fragrant varieties from each germplasm class. Basmati varieties from the Punjab, and basmati-like samples from Central Asia, are from the *aromatic* and *sadri* classes respectively. The others were jasmine style rices from the *indica* and the *japonica* classes. The set included traditional and improved varieties (Table S1). The basmati set included samples that were freshly harvested and also that had been stored for up to 1 year in containers at room temperature, as typically occurs in commercial practice to achieve the quality that basmati consumers prefer. The *sadri* rice from Iran is usually consumed after 3 months storage, and our set included freshly harvested grains as well as grains stored for 3 months at ambient temperature. Fragrant *indica* rice is usually consumed soon after harvest, but samples stored at ambient temperature for 6 months and 1 year were also prepared to compare the aging of the *indica* jasmine with that of the basmati type grains.

Replicate grain samples from each variety were harvested at maturity and sun-dried to an average of 14 % moisture. Freshly harvested or stored paddy from each sample was dehulled (Satake Rice Machine, Tokyo, Japan), and brown grain was polished (Grainman 60-230-60-2AT, Grain Machinery Mfg. Corp., Miami, FL, USA) at IRRRI (Philippines). Polished rice samples was couriered from each rice producing country to Plant Research International, Wageningen where they were immediately placed at 4 °C before grinding.

2.2 Metabolomic profiling

Approximately 50 g polished rice of each variety was ground to a fine powder in liquid nitrogen to generate a stock sample which was thereafter stored at -80 °C.

Aliquots of these powders were then couriered on dry ice to each metabolomic profiling partner, selected for their expertise with a particular metabolomics platform. Volatile compounds were measured using GC-qMS of the headspace sampled by solid-phase micro extraction (SPME) at PRI (Wageningen UR, NL). Primary polar metabolites were profiled by GC-TOF-MS at MPIMP-Golm (DE), both semi-polar and polar compounds by $^1\text{H-NMR}$ and direct infusion ESI MS at Rothamsted Research (UK) and, finally multi-elemental analysis was performed using ICP-MS at UC (DK).

2.2.1 Headspace analysis of volatile compounds by SPME-GC-MS

Headspace volatiles were collected by solid phase micro-extraction (SPME) using a 65 μm polydimethylsiloxane-divinylbenzene fibre (Supelco, Bellefonte, USA) as described previously (Verhoeven et al. 2012). The volatile compounds were thermally desorbed at 250 °C by inserting the fibre for 1 min into the GC injection port (GC 8000, Fisons Instruments, Cheshire, UK). The released compounds were transferred onto the analytical column (HP-5, 30 m \times 0.25 mm ID, 1.05 μm film thickness) in splitless mode. The temperature program ran from 45 °C (2-min hold) and rose 5 °C min^{-1} to 250 °C (5-min hold). The column eluent was ionised by electron impact ionisation at 70 eV (MD800 electron impact MS, Fisons Instruments, Cheshire, UK). 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 using MetAlign software (Lommen 2009) to extract and align the mass signals ($s/n \geq 3$). Mass signals below an s/n of 3 were randomized and given a value between 2.4 and 3 times the calculated noise value. Mass signals present in ≤ 6 samples were discarded. Signal redundancy per metabolite was removed by means of clustering and mass spectra were reconstructed using MSClust (Tikunov et al. 2012). Metabolites were identified by matching the mass spectra of detected metabolites to authentic reference standards and the NIST08, Wiley, and Wageningen Natural Compounds spectral library and by comparison with retention indices in the literature (Strehmel et al. 2008).

2.2.2 $^1\text{H-NMR}$ fingerprinting analysis

The methods previously optimised for $^1\text{H-NMR}$ profiling of wheat flour (Baker et al. 2006) and $^1\text{H-NMR-ESI-MS}$ profiling of Arabidopsis (Ward et al. 2010) were followed. To replicate aliquots of rice flour (100 mg) in 1.5 mL Eppendorf tubes was added $\text{D}_2\text{O-CD}_3\text{OD}$ (1 mL, 4:1) containing 0.05 % w/v TSP- d_4 (sodium salt of deuterated trimethylsilylpropionic acid). The contents of the tube were

mixed thoroughly and heated at 50 °C in a water bath for 10 min. After centrifugation for 5 min, 800 μL of the supernatant was transferred to a clean Eppendorf tube. To ensure cessation of hydrolytic enzyme activity, the samples were heated at 90 °C in a water bath for 2 min, cooled at 4 °C for 45 min and re-centrifuged for 5 min (still at 4 °C); 700 μL of the supernatant was transferred to a 5-mm NMR tube and a further 50 μL to a glass auto-sampler vial containing 950 μL of $\text{H}_2\text{O:methanol}$ (4:1) for DI-MS.

$^1\text{H-NMR}$ 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 the zgpr water suppression pulse sequence with pre-saturation and a 90° pulse and a relaxation delay of 5 s. Each spectrum was acquired using 128 scans of 64 K 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. ^1H chemical shifts were referenced to -TSP- d_4 at $\delta 0.00$. $^1\text{H-NMR}$ 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- d_4 region ($\delta 0.05$ to -0.05). The ASCII file was imported into Excel for the addition of sampling/treatment details. HCA clustering was carried out in Spotfire (Tibco) using extracted characteristic spectral regions, identified via comparisons to a library of authentic standards, which were reduced to unit variance prior to analysis.

2.2.3 Direct infusion ESI-MS fingerprinting analysis (DI-MS)

For direct infusion ESI-MS, 100 μL of each sample was infused into the spectrometer (Esquire 3000, Bruker Daltonics... Bruker Daltonics, Coventry, UK) by flow injection using an Agilent 1100 series HPLC system with degasser, quaternary pump and auto-sampler. The flow rate was 100 $\mu\text{L min}^{-1}$. Mass spectra were recorded from 1.7 to 4.2 min after the sample had entered the flow. Spectra were recorded in both positive and negative electrospray ionization modes on the same sample via an alternating sequence of the two ionisation modes, each time recording an average of 25 scans. The spectra were recorded over an m/z range of 50–1000 using the “smart tuning” function (Bruker Daltonics) with a target mass of m/z 300, a trap-drive and stability of 100 % and a scan speed of 13,000 $m/z \text{ min}^{-1}$. The spectra were recorded using Ion Charge Control with a maximum accumulation time of 40 ms for 20,000 (negative mode) or 50,000 (positive mode) ions.

The nebuliser pressure was 20 psi and the dry gas flow rate was 6 L min⁻¹ at 350 °C.

Using Data Analysis 3.2 (Bruker Daltonics) the accumulated positive and negative ion spectra for each sample were separately combined and exported as ASCII files of m/z-intensity pairs. For spectral alignment AMIX software was used to reduce the ASCII files to a single data table for each ionisation mode, containing integrated regions or buckets of equal width (m/z 1). Individual spectral intensities were scaled to total spectrum intensity and exported as a CSV file for annotation in Excel and subsequent PCA analysis in SIMCA-P.

2.2.4 Gas chromatography–time of flight mass spectrometry (GC–TOF–MS): derivatized metabolite fractions

Sample extraction, chemical derivatization, injection, instrumentation, chromatogram and data pre-processing were exactly as previously described (Allwood et al. 2009). Retention indices (RI) were generated for each individual chromatogram based on a spiked mixture of alkanes. Data were aligned according to RI and full mass unit into a numerical data matrix using the TagFinder software (Luedemann et al. 2008). Clusters of at least three corresponding mass fragments were selected for relative metabolite quantification. Metabolites were identified by TagFinder supported matching to reference spectra and RI from the Golm Metabolome Database (Luedemann et al. 2008). The matching process was manually supervised with RI deviations <1 % (Strehmel et al. 2008).

2.3 Multi-elemental analysis

Whole grain milled samples were sent to the University of Copenhagen for multi-elemental analysis. The grain samples were surface decontaminated by 3 consecutive rinses in Milli-Q-water (Millipore, Bedford, Massachusetts, USA) and samples were subsequently freeze-dried for 48 h (Christ Alpha 2-4, Martin Christ GmbH, Osterode, Germany). Whole rice grain samples were then 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 bar. The digestion medium consisted of 250 mg dry grain sample, 5 mL 65 % ultrapure HNO₃ (J.T. Baker Instra-Analysed Reagent) and 5 mL 15 % H₂O₂ (30 % Extra-Pure, Riedel de Häen, Selze, Germany). After digestion, the samples were diluted to 3.5 % v/v HNO₃ with ultrapure Milli-Q water (Laursen et al. 2009). Multi-elemental analysis was performed using ICP-MS (Agilent 7500ce, Agilent Technologies, Manchester, UK) tuned in standard mode. The plasma power was operated at 1500 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 0.6 mL min⁻¹ by a PFA micro-flow nebulizer. Elimination of spectral interferences was achieved using an octopole ion guide with the cell gases helium or hydrogen as described previously (Laursen et al. 2009). Using this instrumental set-up, 13 elements were analysed: Na, Mg, Al, P, S, K, Ca, Mn, Fe, Ni, Cu, Zn and Mo. Seven replicates of a 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 element concentrations deviating less than ±10 % from the certified mean reference values were accepted. The limit of detection (LOD) was determined as three times the standard deviation of a minimum of 7 blanks. Only data above LOD were included for chemometric analysis.

2.4 Statistical analysis

Raw data obtained by the different platforms were auto-scaled (van den Berg et al. 2006). To obtain an overview of the whole data set, principal component analysis (PCA) was performed using the software package SIMCA-P (version 13.0. Umetrics, Umeå, Sweden). The number of significant PC's was determined using cross validation (Eriksson et al. 2006). Hierarchical cluster analysis, using Euclidean distance and Ward's linkage method, was applied to depict the similarities of the samples as measured by the different analytical platforms.

To identify the metabolites that are important for describing the differences between the groups, a double cross validated partial least squares discriminant analysis (PLS-DA) was used (Smit et al. 2007; Westerhuis 2008). PLS-DA is a classification technique that creates linear combinations (latent variables) of the original variables to predict the class labels, indicated by zeroes and ones.

A two nested leave-one-out cross validation or double cross validation was used to obtain an unbiased estimate of the prediction error. In this way, class membership of a sample was predicted without its use in the model building process. In the outer cross validation, one sample was set aside, and the model built using the remaining samples. To determine the optimal number of latent variables for this PLS-DA model, the remaining samples were subjected to an inner leave-one-out cross validation. By iteratively repeating this procedure and leaving out all samples one at a time, we obtained a class prediction and model per sample. The number of misclassifications from this procedure was compared to the results from a permutation test (Smit et al. 2007) to check if the results could have been obtained by chance alone.

To find metabolites that are predictive for the different groups, we focused on those that had high absolute values for the regression coefficients in the PLS-DA models. Since the double cross validation procedure yielded as many models as the number of samples that we had, we summarized the many models using rank products (Smit et al. 2007). The absolute regression coefficients from each model were ranked (with rank 1 for the highest absolute value). To summarize the importance of each metabolite, the rank numbers from all models were multiplied together, giving a rank product. Metabolites with a low rank product are typically the most important metabolites, since that metabolite was often at the top of the list of sorted regression coefficients. When rank products are (very) high, the metabolite in question was typically in the lower regions of the sorted list, indicating that often other metabolites were more important.

3 Results

Thirty one fragrant and non-fragrant rice varieties from 10 different countries were used in this study. Of these varieties, 23 were fragrant and 8 were not (Table S1). All samples have been analysed by six generally complementary metabolomics platforms: SPME–GC–MS of headspace extracts (518 reconstructed peaks), $^1\text{H-NMR}$ of 25 % MeOH extracts (95 frequency bins), DI-MS of 25 % MeOH extracts (634 mass signals in negative ionization mode; 367 mass signals in positive ionization mode), GC–TOF–MS of derivatized polar extracts (82 compounds; mainly primary metabolites), and ICP-MS for micronutrients (13 elements). The raw data from these six analytical platforms were concatenated, resulting in a dataset consisting of 1709 variables, and were subjected to different multivariate statistical techniques including principal component analysis (PCA), hierarchical clustering analysis (HCA), and partial least squares-discriminant analysis (PLS-DA).

Figure 1a represents a PCA of all 31 varieties, combining the data from all six platforms. The first PC, explaining 24.8 % of the total variation, mainly separates the fresh basmati varieties from the majority of the other varieties. Interestingly, IR6 (1073) which is a freshly harvested and non-fragrant variety from the *tropical japonica* germplasm class grown in Pakistan, resembles the non-stored *aromatic* basmati rices from Pakistan with regard to its metabolite profile. In the first two principal components, there are no other notable groupings. The PCA loadings show that the metabolites from the DI-MS and $^1\text{H-NMR}$ platforms are at the extremes of PC1, whereas metabolites from the other platforms are more randomly distributed (Fig. 1b). Metabolites that were analysed by NMR are highly correlated with and are highly abundant in the fresh

basmati samples (Fig. 1a, b). Since the $^1\text{H-NMR}$ extraction protocol specifically targets polar metabolites such as carbohydrates, amino acids and organic acids the data would suggest that fresh basmati samples are richer in this class of metabolites compared to other varieties. For micronutrients, no systematic differences were found between rice samples across the 13 elements analysed.

Figure 2 represents an HCA of all the samples of freshly harvested samples of both jasmine and basmati rice, combined for all platforms. It is clear that the first cluster break separates the jasmine samples from the basmati samples, with two unexpected exceptions: Kai Noy Leuang (KNL) from Lao PDR clusters with the basmati samples, and Domsiah from Iran, clusters with the jasmine samples. There is no consistent pattern of germplasm class in the clustering. Even though major differences in the metabolic profiles of the different varieties can be recognized, the picture is still too complex to answer specific questions. Consequently, to reduce the level of complexity and the number of variables in the data set, a number of logical inter-group comparisons were performed, using the data from individual platforms, to examine in detail the presence of discriminatory compounds between specific sub-groups. We focused on (i) the metabolome of stored basmati varieties from the Punjab and freshly harvested jasmine varieties from Thailand according to the manner in which they are normally consumed; (ii) the metabolic differences between freshly harvested and stored basmati varieties; and (iii) the metabolomic differences between freshly harvested basmati and jasmine varieties. These analyses were also used to reveal which metabolomics platforms are most valuable for discriminatory analyses with the view to delivering selection tools to breeders.

3.1 As rice is marketed: fresh Thai jasmine versus stored basmati varieties

Figure 3 represents dendrograms from an HCA depicting the relationship between metabolite profiles of traditional and improved samples of basmati and jasmine rice, as they would normally be sold and consumed: freshly harvested for jasmine, and stored locally at ambient temperature for about 12 months, for basmati. The volatile and derivatized polar metabolites both showed the expected separation between basmati and jasmine rices (Fig. 3a, b), but the other platforms (Fig. 3c–e) misclassified some of the samples. Statistically, PTT 1 was misclassified by PLS-DA by all platforms except for GC–TOF–MS. However, even though the volatile profile of PTT 1 was significantly different from the Thai varieties, using HCA, PTT 1 still clustered within the sub-group of jasmine varieties (Fig. 3a). Six of the ten most discriminating volatile compounds are more abundant in the stored basmati

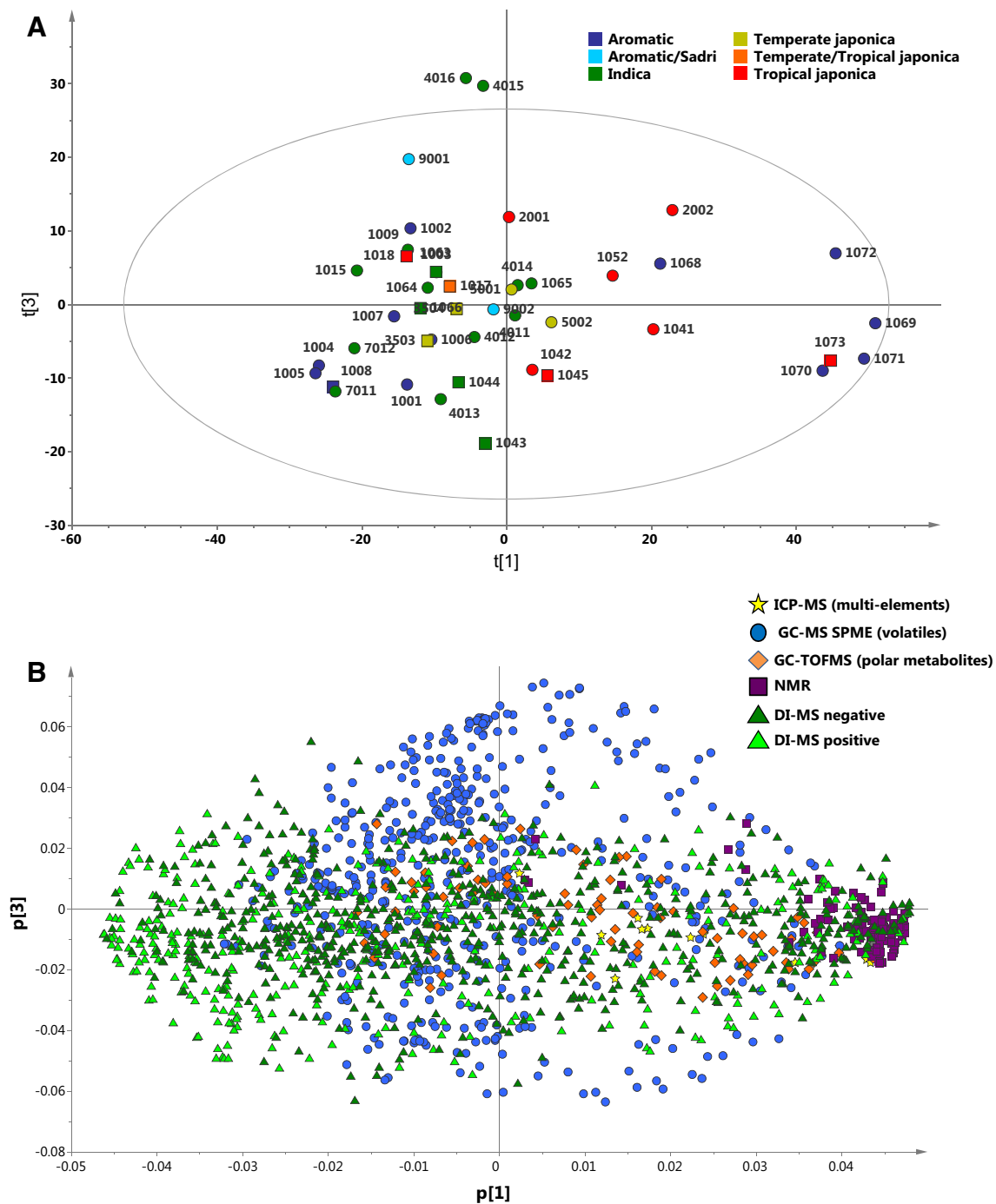


Fig. 1 Principal component analysis (PCA) of the metabolome as detected by six different analytical platforms of 23 fragrant and 8 non-fragrant rice varieties. **a** Scores plot showing the first PC and third PC explaining 24.8 and 6.2 % of the variation respectively. *Numbers* next to the *symbols* represent the metaphor code of the rice varieties as

given in Table S1. Different colors indicate the different germplasm classes. *Circles* indicate fragrant varieties while *boxes* represent non-fragrant varieties. **b** Loadings plot of the first PC versus the third PC. The five different analytical platforms are indicated by the *different colored symbols* (Color figure online)

samples than in the jasmine varieties, but none of these could be unambiguously annotated (Table 1). The most discriminating polar (GC-TOF-MS) compounds included the organic acids tetradecanoic acid, nonanoic acid and fumaric acid, which were more abundant in basmati

varieties, while erythronic acid and its lactone derivative, raffinose, and 2-amino-malonic acid were more abundant in jasmine varieties. For metabolites detectable by NMR, PTT 1 appears to be more similar to the basmati rices according to HCA and was consequently misclassified by

Fig. 2 Hierarchical cluster analysis (HCA) of the metabolome as detected by five different metabolomics platforms of 20 fragrant rice varieties that were analyzed soon after harvest (fresh). Samples are colored according to their germplasm class. The boxed label indicates samples which were incorrectly classified by double cross validated PLS-DA (Color figure online)

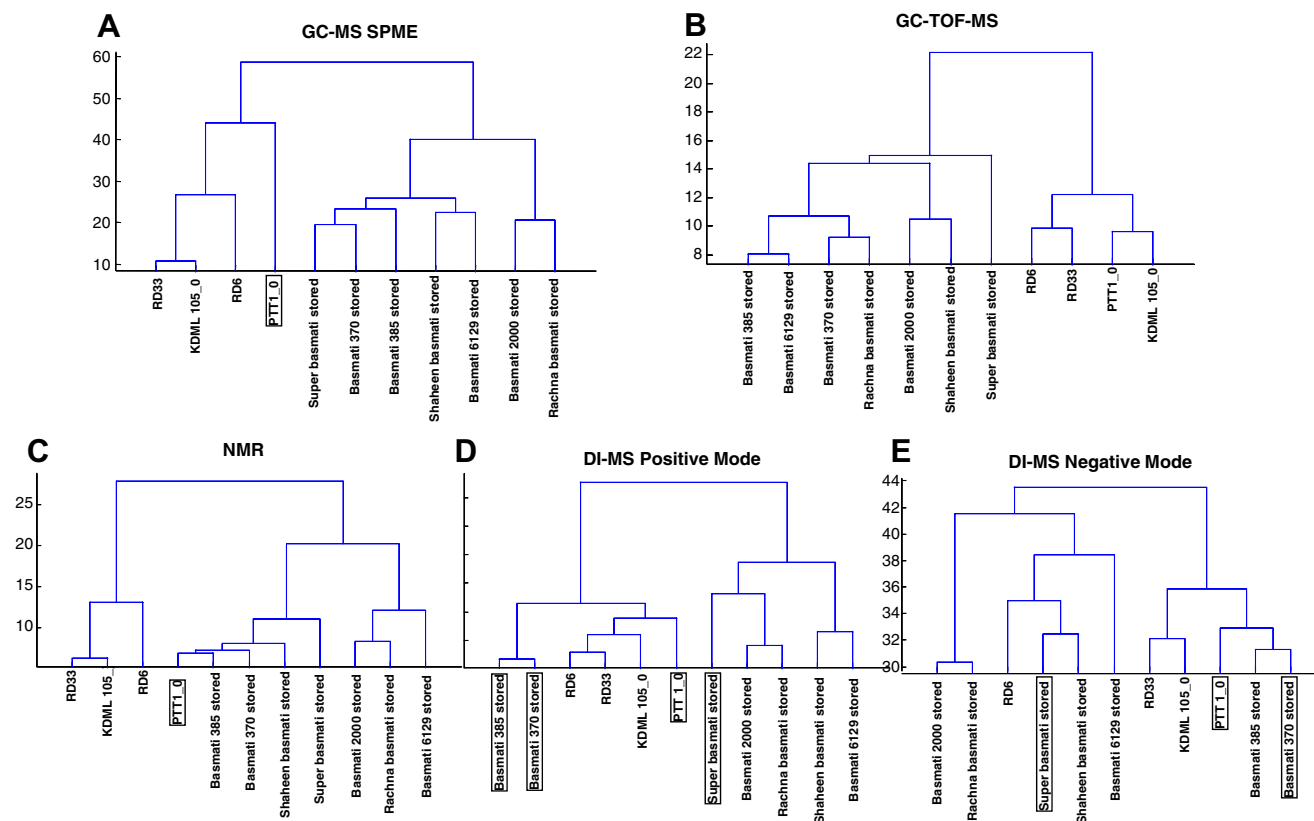
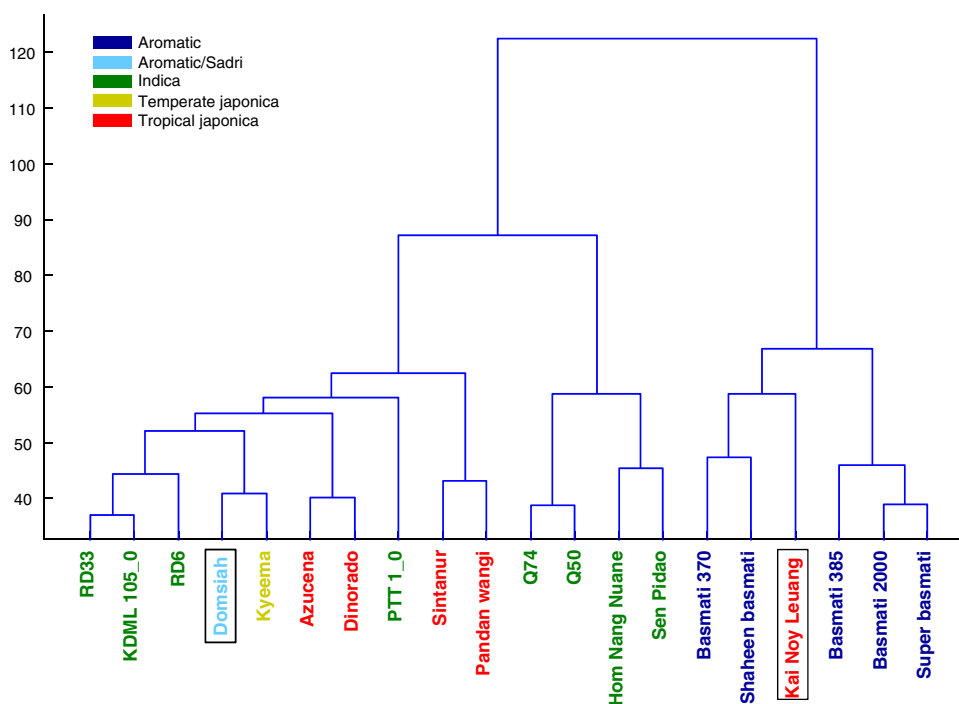


Fig. 3 Hierarchical cluster analysis (HCA) of the metabolic profiles of four fresh Thai jasmine varieties and seven stored basmati varieties. Each analytical platform is represented by a separate

dendrogram. The *boxed label* indicates samples which were incorrectly classified by double cross-validated PLS-DA

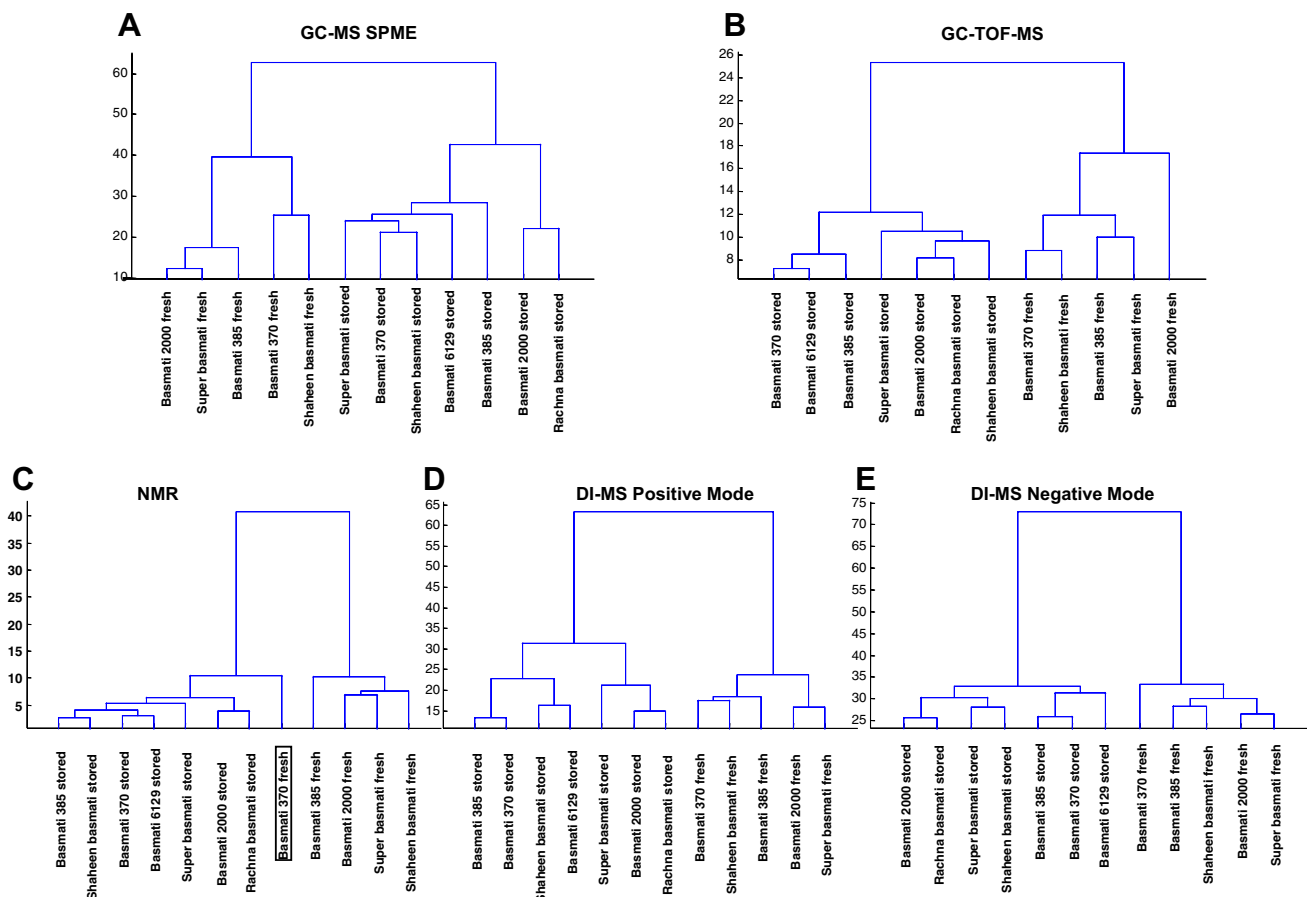


Fig. 4 Hierarchical cluster analysis (HCA) of the metabolic profiles of fresh and stored basmati varieties. Each metabolomics platform is represented by a separate dendrogram. The *boxed label* indicates the sample was incorrectly classified by double cross-validated PLS-DA

PLS-DA. Both positive and negative DI-MS techniques were unable to show the expected classification. Only when these two DI-MS modes were used to look at semi-polar metabolites were the improved basmati varieties observed to have a different profile compared to the traditional varieties (Basmati 370 and 385) (Fig. 3d, e).

3.2 Fresh versus stored basmati varieties

Freshly harvested basmati could be discriminated from stored basmati by every analytical platform using HCA (Fig. 4). With the exception of the $^1\text{H-NMR}$ profile of Basmati 370, all samples were also correctly classified by PLS-DA (Fig. 4). The metabolites most responsible for the separations are shown in Table 2. Of the 10 most discriminatory volatile compounds, three were more abundant in fresh basmati while seven had higher values in stored basmati varieties (Table 2). The differences between stored and fresh basmati varieties with regard to polar metabolites, as analysed by GC-TOF-MS and $^1\text{H-NMR}$, were mainly based on metabolites more abundant in the fresh basmati samples. The differences detected by DI-MS positive mode were less pronounced than

for the other platforms. Very few of the 50 most discriminating metabolites can yet be annotated (Table 2).

3.3 Fresh Thai jasmine versus fresh basmati varieties

To determine whether the metabolites in basmati and jasmine differ immediately after harvesting, fresh samples of basmati and jasmine rices were compared (Fig. 5). The overall picture from all platforms clearly indicated that grains of fresh basmati and jasmine rices differ in their suite of metabolites. The metabolic profile of the improved jasmine variety PTT 1 was significantly different from the other jasmines with regard to the volatile patterns (as determined by PLS-DA), but in spite of this, PTT 1 always still fell in the jasmine sub-cluster (Fig. 5). The profiles of Basmati 370 as analysed by $^1\text{H-NMR}$ and DI-MS negative mode were not classified as basmati varieties by PLS-DA. Unexpectedly, when analysed by $^1\text{H-NMR}$, the improved variety Basmati 2000 clustered using HCA within the basmati samples (Fig. 5c) but was misclassified when using PLS-DA. The metabolites from all platforms that

Table 1 List of metabolites detected by the analytical platforms that significantly contribute to the metabolic differences of fresh and stored basmati varieties as determined by partial least squares-discriminant analysis (PLS-DA) (Color table online)

Metabolite ID	RankProduct from CV PLS-DA	Fresh Thai Jasmynes		Stored Basmati	
		Mean	SD	Mean	SD
GCMS volatiles					
GCMS 4371 (hydrocarbon)	692224	71500	56400	18400	6070
GCMS 6908	82620000	921	1420	10900	2830
GCMS 6429	152064000	2110	2240	14200	1590
GCMS 5939 (hydrocarbon)	4694215680	306000	51700	16800	13000
GCMS 7510	13801881600	82	15	1530	453
GCMS 7610	31465324800	1090	1030	3810	725
GCMS 7716	45396120000	179	219	3900	702
GCMS 5295 (hydrocarbon)	3.98649E+11	28400	12400	10000	9140
GCMS 7336	1.09369E+12	139	163	2800	456
GCMS 4164	1.67556E+12	12900	2070	617	429
GCTOF polar					
A147011-101	2	18.2	4.6	2.5	0.8
A213001-101	270000	16.3	1.7	9.9	1.1
Tetradecanoic acid	4478976	5.6	0.5	9.5	1.6
Erythronic acid-1,4-lactone	36960000	7.7	4.3	0.3	0.3
Malonic acid, 2-amino-	796068000	5.3	2.4	1.3	0.2
Raffinose	1646023680	5.0	1.8	0.5	0.2
Erythronic acid	2194698240	11.3	6.7	1.4	1.1
Nonanoic acid	6372753408	2.4	1.2	5.4	0.9
Fumaric acid	57231360000	1.0	0.3	4.9	1.6
A211001-101	71751680000	16.1	2.5	10.1	1.2
NMR					
Acetate	3	58.1	17.1	199.0	52.6
Glycerol	3840	437.0	99.9	582.0	172.0
7.975-7.955 (unknown aromatic)	1399680	0.9	0.1	0.6	0.1
Nicotinic acid	68812800	5.0	2.0	11.9	6.3
2.425-2.395	483840000	16.7	2.7	17.1	3.7
Phenylalanine	2037420000	3.8	0.2	2.6	0.6
Raffinose	9484231680	90.6	9.4	60.0	7.1
Fumarate	28686873600	9.9	1.2	9.3	1.8
9.295-9.195 (unknown carbohydrate)	1.05641E+11	1.5	0.1	1.0	0.4
7.875-7.845 (unknown aromatic)	2.64856E+11	1.1	0.0	1.2	0.3
DI-MS Positive mode					
MS+ 449	240	0.00115	0.00009	0.00186	0.00011
MS+ 529	2592	0.00309	0.00016	0.00225	0.00013
MS+ 543 (Raffinose K+ adduct)	1209600	0.00224	0.00014	0.00152	0.00008
MS+ 699	65028096	0.00069	0.00006	0.00047	0.00004
MS+ 544	93837744000	0.00122	0.00013	0.00079	0.00008
MS+ 528	2.27753E+11	0.00564	0.00098	0.00280	0.00079
MS+ 530	5.05228E+11	0.00127	0.00011	0.00114	0.00010
MS+ 299	1.08992E+12	0.00047	0.00003	0.00065	0.00010
MS+ 471	3.97715E+12	0.00062	0.00005	0.00087	0.00008
MS+ 527 (Raffinose Na+ adduct)	6.13735E+12	0.01010	0.00206	0.00474	0.00156
DI-MS Negative mode					
MS- 503 (Raffinose)	4608	0.00391	0.00032	0.00207	0.00049
MS- 826	26611200	0.00128	0.00014	0.00087	0.00010
MS- 529	28901376	0.00129	0.00017	0.00083	0.00011
MS- 656	65610000	0.00109	0.00006	0.00078	0.00009
MS- 653	847143360	0.00548	0.00079	0.00260	0.00084
MS- 654	2670796800	0.00355	0.00071	0.00150	0.00051
MS- 825	3102624000	0.00151	0.00014	0.00087	0.00020
MS- 463	18065203200	0.00102	0.00010	0.00153	0.00018
MS- 227	49397040000	0.00259	0.00038	0.00178	0.00013
MS- 575	2.37702E+11	0.00093	0.00006	0.00146	0.00021

Metabolites were ranked in descendant order of the rank products that was determined by double cross validation. Mean and standard deviation (sd) of the metabolites of the different compared varieties is given. For clarity, colour codes indicate the higher (red) and lower (blue) values for each

Table 2 List of metabolites detected by the analytical platforms that significantly contribute to the metabolic differences of fresh traditional Thai jasmine varieties and stored basmati varieties as determined by PLS-DA (Color table online)

Metabolite ID	RankProduct from CV PLS-DA	Fresh Basmati		Stored Basmati	
		Mean	SD	Mean	SD
GCMSvolatiles					
GCMS 8019	5832	5.52e+01	1.99e+01	1.44e+03	239
GCMS 4487	9408	1.39e+03	2.53e+02	1.24e+02	24
GCMS 6429	41817600	2.10e+03	2.04e+03	1.42e+04	1590
GCMS 2738	7838208000	1.40e+04	3.63e+03	9.46e+02	725
Benzyl alcohol	17831923200	4.78e+05	1.26e+05	4.25e+04	40400
GCMS 7716	208653096960	8.53e+02	5.00e+02	3.90e+03	702
GCMS 7678	391184640000	3.88e+02	4.76e+02	3.98e+03	804
GCMS 6665	3450669465600	1.13e+02	3.70e+01	5.79e+03	1690
GCMS7007	4237695221760	1.48e+02	8.46e+01	4.78e+03	1190
GCMS7801	7320202606080	3.37e+02	4.74e+02	3.31e+03	711
GCTOF polar					
A147011-101	72	1.47e+01	3.03e+00	2.50e+00	0.8
2 amino-Malonic	380160	3.82e+00	8.19e-01	1.32e+00	0.2
A131003-101	3849120	5.61e-01	2.01e-01	6.61e+00	1.8
Citric acid	478800000	1.80e+01	6.35e+00	1.80e+00	1.0
Tetradecanoic acid	4978713600	4.68e+00	6.38e-01	9.46e+00	1.6
A283005-101	8544463200	5.12e+00	1.99e+00	4.37e-01	0.2
A166004-101	149099258880	1.92e+01	4.17e+00	7.43e+00	3.8
Adipo-2,6-lactam	156513337344	2.53e+00	4.35e-01	1.19e+00	0.3
Valine	181683532800	1.55e+01	4.80e+00	3.68e+00	1.8
Raffinose	386311680000	2.09e+01	9.32e+00	4.91e-01	0.2
NMR					
Glutamate/glutamine	56	1.22e+02	2.14e+01	5.01e+01	5.1
Glutamate	1175040	1.22e+02	2.45e+01	3.91e+01	11.1
3.685-3.655	47900160	2.28e+03	5.46e+02	5.69e+02	180.0
Glucose	541900800	5.15e+03	1.27e+03	1.21e+03	398.0
Glucose	2280960000	1.67e+03	3.67e+02	5.28e+02	121.0
Glucose	9001692000	1.15e+03	2.90e+02	2.66e+02	85.0
5.185-5.165	100111334400	2.62e+01	2.83e+00	1.71e+01	1.4
Sucrose	120374231040	9.65e+02	2.66e+02	1.72e+02	71.3
Fructose	654151680000	8.31e+02	1.28e+02	4.15e+02	63.1
Sucrose	1102248000000	1.10e+03	3.13e+02	1.86e+02	79.2
DI-MS Positive mode					
MS+515	64260	1.48e-03	1.34e-04	1.30e-03	0.00018
MS+516	1350000	9.93e-04	1.24e-04	8.78e-04	0.00009
MS+336	8631360	1.00e-03	1.56e-04	7.81e-04	0.00019
MS+350	18800640	1.04e-03	1.57e-04	7.28e-04	0.00008
MS+386	195955200	9.00e-04	7.84e-05	1.10e-03	0.00008
MS+531	24690355200	2.19e-03	3.27e-04	1.84e-03	0.00023
MS+219	1614820147200	5.61e-04	6.68e-05	1.28e-03	0.00010
MS+348	3808715904000	7.82e-04	1.38e-04	7.78e-04	0.00012
MS+702	4822364160000	5.10e-03	1.95e-04	5.80e-03	0.00060
MS+318	11581503897600	7.15e-04	8.45e-05	5.42e-04	0.00008
DI-MS Negative mode					
MS-324	6	4.77e-04	6.95e-05	9.37e-04	0.00003
MS-378	9216	2.55e-03	3.28e-04	5.97e-04	0.00015
MS-343	7741440	3.81e-03	4.19e-04	1.07e-03	0.00034
MS-532	116640000	1.07e-03	8.82e-05	2.22e-03	0.00020
MS-379	381024000	0.00297	3.89e-04	1.07e-03	0.00015
MS-529	1770209280	2.32e-03	3.18e-04	8.26e-04	0.00011
MS-679	2222640000	4.03e-03	3.21e-04	1.86e-03	0.00032
MS-132	174596083200	1.03e-02	6.57e-04	1.35e-02	0.00042
MS-704	776301926400	7.87e-04	1.30e-04	1.70e-03	0.00017
MS-491	2201696640000	5.17e-02	7.93e-03	1.26e-02	0.00571

Metabolites were ranked in descendant order of the rank products that was determined by double cross validation. Mean and standard deviation of the metabolites of the different compared varieties is given. For clarity, colour codes indicate the higher (red) and lower (blue) values for each

contributed the most to the separation between fresh jasmine and basmati varieties are shown in Table 3.

3.4 Effects of storage on metabolome of Thai jasmine rices

While clear changes in the metabolome occur during storage in Basmati rices (Fig. 4) little is known about changes to jasmine rices upon storage. We therefore analysed two important jasmine rices from Thailand: KDML 105 and PTT 1 that were either fresh, stored at ambient temperature for 6 months or for 12 months. In addition, two other fresh fragrant jasmine varieties RD 6 and RD 33 were included as extra reference materials. When combining all metabolic platforms, a clear aging effect could be seen (Fig. 6a). Unsurprisingly, samples that had been stored for 12 months were particularly characterised by an increase/appearance of many specific volatile compounds (Fig. 6b). Interestingly, in the un-stored samples, PTT 1 is separated from the other jasmine rices along PC2, and after 6 months storage both KDML 105 and PTT 1 have moved along PC1, but notably with an opposite separation along PC2. After 12 months of storage, both samples have moved further along PC1, and there is no longer any separation along PC2 (Fig. 6a).

When comparing the profiles of jasmine and basmati rices in response to storage it appears that aging in jasmine rice apparently results in different metabolome changes than for basmati (Fig. 6c). Thai jasmine varieties stored for 12 months are characterised by an increase of particularly volatile compounds (Fig. 6d) as indicated by the strong positive loadings of PC2 (corresponding to the position of the jasmine samples stored for 12 months in Fig. 6c). In contrast, basmati samples are also characterised by an increase of semi-polar metabolites as analysed by DI-MS.

4 Discussion

Basmati and jasmine rices are arguably the most popular rices in the world. Consumers pair these rices with appropriate cuisines, and can easily tell the difference between both types, on the basis of aroma, taste, flavour and texture (Champagne et al. 2010). However, breeding programs aiming to develop new varieties of each type are hampered by a lack of knowledge about ways to discriminate between the two rice types. Both types are fragrant, both express 2AP (Fitzgerald et al. 2008) as a major fragrance compound (Buttery et al. 1982), and both carry exactly the same genetic mutation that leads to the expression of 2AP (Kovach et al. 2009). Although factors such as grain appearance and cooked rice texture differ

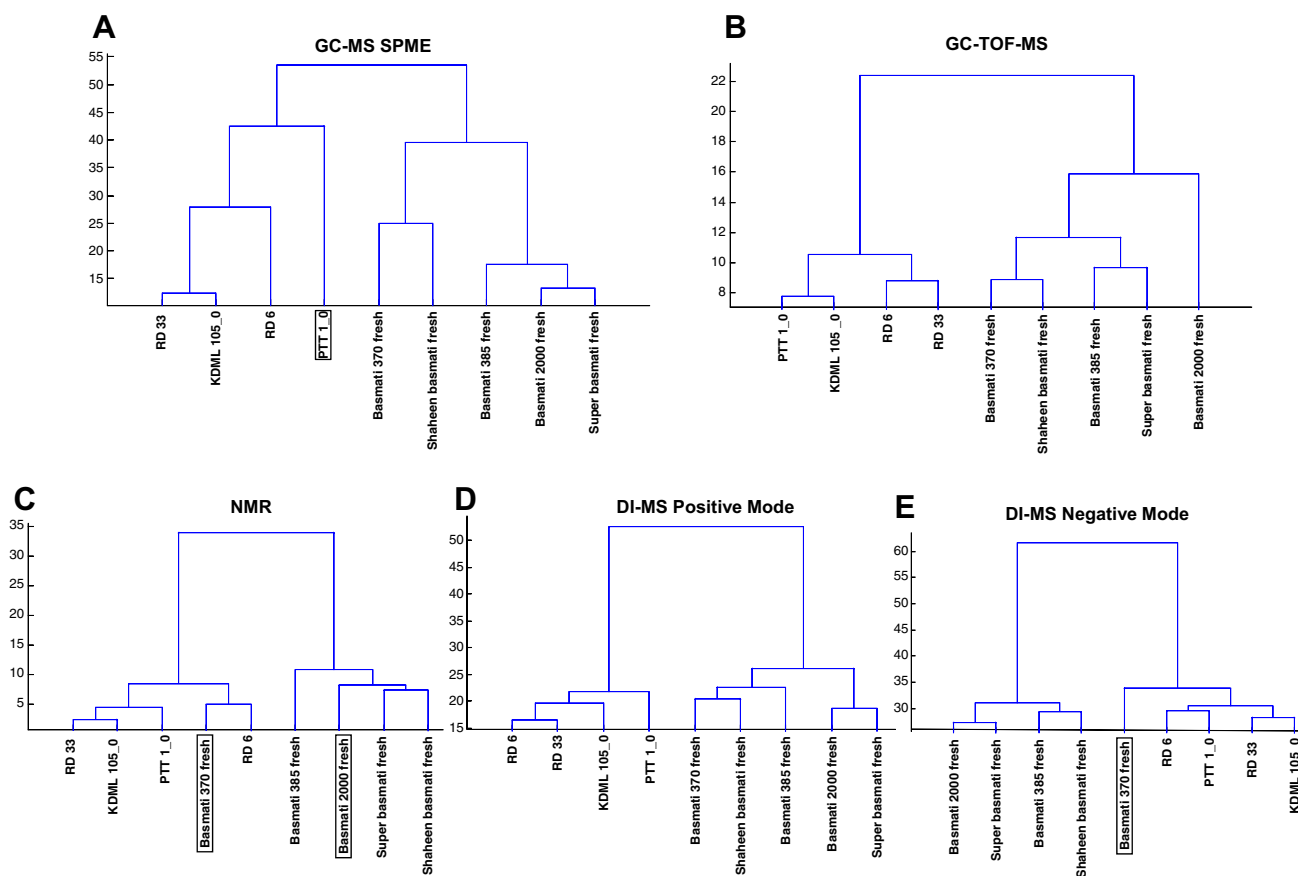


Fig. 5 Hierarchical cluster analysis (HCA) of the metabolic profiles of fresh Thai jasmine varieties and fresh basmati varieties. Each metabolomics platform is represented by a separate dendrogram. The

boxed label indicates samples that were incorrectly classified by double cross-validated PLS-DA

between jasmine and basmati rices selection for 2AP either by phenotype and/or genotype cannot discriminate the two classes. The analysis of volatile compounds in the head-space of aromatic rice has intensively been investigated over the years (Bryant and McClung 2011). However, all previous studies that included both jasmine and basmati varieties either used samples imported from Asia and grown in the country of study, or used commercial samples purchased from a local (super)market. Limitations to using these types of samples are that the pre- and post-harvest environments can have a significant effect on the volatile profile of jasmine and basmati rice (Yoshihashi et al. 2004)—as has also clearly been demonstrated here (Figs. 4, 6). Furthermore, basmati rice (Bhattacharjee et al. 2002), and commercial rice samples are often mixtures of several varieties, of uncertain origin (Sivaranjani et al. 2010; Steele et al. 2008).

The particular value of the collection of rice varieties used in the present study is that all were confirmed to be 100 % pure; grown specifically for this project, each in their adapted environment and conditions; and delivered with full harvest, postharvest and storage histories.

Secondly, the samples have been analysed by six metabolomics platforms, each measuring mostly different groups of compounds. Results in Fig. 2 clearly demonstrate that the jasmine rices of the GMS and from other countries, all share common metabolic elements, which separates them at the first HCA cluster break, from the basmati samples. Deeper analysis of these data would provide the best opportunity yet to understand differences between basmati and jasmine rices to determine potential use for breeders, consumers and food chemists alike.

The results presented in Fig. 1a include all samples and data from all platforms. The most notable discrimination occurs on PC1 where the fresh basmati samples are shifted to the extreme of PC1. This is possibly explained by Fig. 1b, which shows that the polar and semi-polar compounds detected by NMR and DI-MS are generally polarized at the extremes of PC1. These compounds are mostly involved in primary metabolism of carbon and nitrogen (Moing et al. 2004). That the compounds from these three platforms separate fresh basmati samples from the others suggests the presence of specific metabolites or biochemical processes unique to fresh basmati and which appear to

Table 3 List of metabolites detected by the analytical platforms that significantly contribute to the metabolic differences of fresh traditional Thai jasmine varieties and freshly harvested basmati varieties as determined by PLS-DA (Color table online)

Metabolite ID	RankProduct from CV PLS-DA	Fresh Basmati		Fresh Thai Jasmynes	
		Mean	SD	Mean	SD
GCMS volatiles					
GCMS 6061 (hydrocarbon)	9072	6100	2340	16800	2700
GCMS 5592 (hydrocarbon)	19152	3990	4040	46600	3770
GCMS 5495	1866240	278	190	3620	419
GCMS 7003 (hydrocarbon)	174073536	1330	1050	4060	743
GCMS 5507 (hydrocarbon)	182626920	3700	2680	35900	5530
GCMS 5441	259096320	1840	1200	34700	2280
GCMS 6925 (hydrocarbon)	418057200	14300	6320	38500	2850
GCMS 6980 (hydrocarbon)	1311552000	14300	6320	38500	2850
GCMS 4894 (silane artefact)	9295104000	3970	110	5790	258
2-butoxy ethanol	21003840000	13300	2480	2610	2860
GCTOF polar					
A131003-101	6000	0.6	0.2	7.3	0.5
Threitol	14364	7.1	2.0	11.6	2.0
A166004-101	37440	19.2	4.2	9.1	1.1
Psicose	10027008	14.7	2.4	21.1	2.8
Citric acid	72990720	18.0	6.4	1.0	0.4
A198001-101	164934000	20.2	4.0	12.5	3.9
Valine	1121734656	15.5	4.8	2.9	1.4
A213001-101	2412305280	12.7	3.0	16.3	1.7
A125004-101	2889457920	6.8	3.8	1.2	0.9
Erythronic acid-1,4-lactone	5157636120	3.0	2.4	7.7	4.3
NMR					
Glutamate	432	122.0	24.5	43.9	4.5
Glutamate / Glutamine	175104	122.0	21.4	59.6	5.4
3.685-3.655	864000	2280.0	546.0	740.0	113.0
3.775-3.725	7464960	1780.0	420.0	609.0	83.7
Glucose	48000000	1150.0	290.0	364.0	42.6
Glucose	104976000	5150.0	1270.0	1680.0	257.0
Glucose	161021952	1880.0	486.0	589.0	102.0
Glucose	507091200	1670.0	367.0	670.0	95.3
5.185-5.165	842063040	26.2	2.8	18.5	1.1
Sucrose	5544045000	965.0	266.0	280.0	48.2
DI-MS Positive mode					
MS+ 529	465696	0.00262	0.00021	0.00309	0.00016
MS+ 515	1983744	0.00148	0.00013	0.00114	0.00007
MS+ 350	887112000	0.00104	0.00016	0.00072	0.00014
MS+ 413	4311180720	0.00106	0.00006	0.00154	0.00006
MS+ 318	4324725000	0.00072	0.00008	0.00049	0.00010
MS+ 325	9044421120	0.00066	0.00009	0.00074	0.00005
MS+ 699	20451816000	0.00057	0.00005	0.00069	0.00006
MS+ 250	1.6118E+11	0.00022	0.00002	0.00031	0.00002
MS+ 291	1.6511E+11	0.00053	0.00007	0.00070	0.00007
MS+ 726	4.18164E+11	0.00046	0.00003	0.00063	0.00006
DI-MS Negative mode					
MS- 401	912	0.00104	0.00008	0.00054	0.00006
MS- 428	336600	0.00021	0.00002	0.00046	0.00006
MS- 378	6633900	0.00255	0.00033	0.00097	0.00018
MS- 343	7464960	0.00381	0.00042	0.00160	0.00035
MS- 132	151165440	0.01030	0.00066	0.01340	0.00053
MS- 514	271656000	0.00461	0.00025	0.00309	0.00034
MS- 491	1087257600	0.05170	0.00793	0.02000	0.00309
MS- 361	6038323200	0.00111	0.00047	0.00295	0.00022
MS- 902	8137584000	0.00028	0.00003	0.00048	0.00005
MS- 682	8423509248	0.00141	0.00048	0.00340	0.00028

Metabolites were ranked in descendant order of the rank products that was determined by double cross validation. Mean and standard deviation of the metabolites of the different of the metabolites is given. For clarity, colour codes indicate the higher (red) and lower (blue) values for each

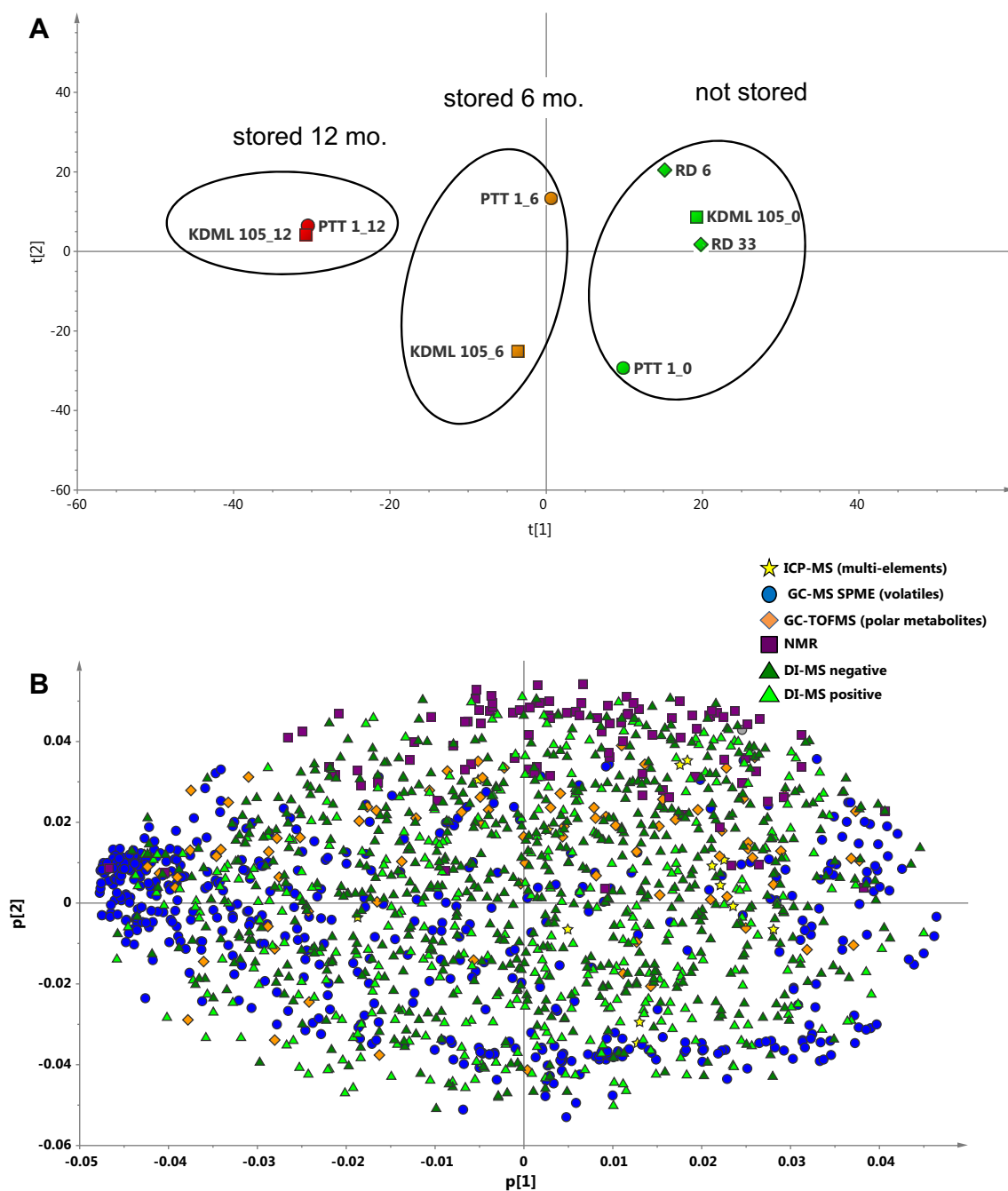


Fig. 6 PCA of the metabolome as detected by five different metabolomics platforms of five basmati varieties from the Punjab region and Thai jasmine varieties that were fresh or stored for up to a maximum of 12 months. **a** Score plot of fresh and stored Thai jasmine varieties showing PC1 versus PC2 explaining 24.9 and 18.6 % of the variation respectively. Grain samples were analyzed soon after harvest (*green*), 6 months (*orange*) and 12 months (*red*) after storage

at room temperature. **b** Corresponding loading plot of the first PC versus the second PC. **c** Score plot of basmati and Thai jasmine varieties showing PC1 versus PC2 explaining 34.4 and 10.6 % of the variation respectively. **d** Loading plot of the first PC versus the second PC. The five different analytical platforms are indicated by the *different colors* (Color figure online)

be no longer present after storage. Such changes should be borne in mind when assessing progeny for selection in any breeding programme. Consistent with this, Tables 2 and 3 reveal that, of the 30 most discriminating compounds from NMR and the DI-MS platforms, 14 are common in the

comparisons between fresh and stored basmati (Table 2) and fresh basmati versus fresh jasmine (Table 3), signaling the importance of these compounds in discriminating fresh basmati samples from all the others. Basmati grains differ from all other types in that the grains only expand

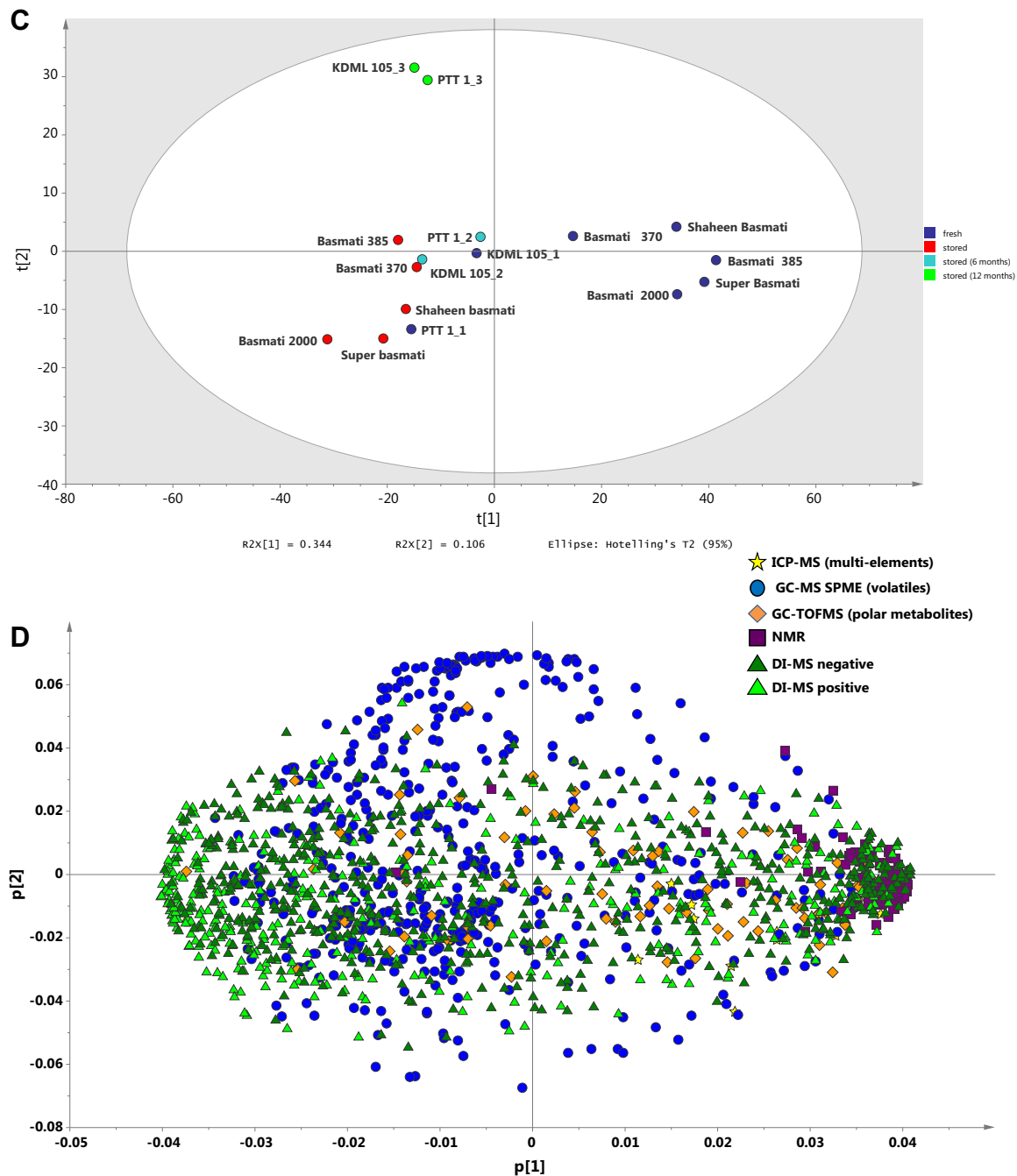


Fig. 6 continued

longitudinally when cooked (Bhattacharjee et al. 2002), which is attributed to the unique packing of starch granules and alignment of cells within the grains (Sekhar and Reddy 1982; Singh et al. 2000). It is therefore possible that the discriminating biochemical metabolites could be those participating in the pathways during grain-filling that lead to the unique internal arrangement of basmati grains. It is well known that next to the carbohydrate polymers, starch grains are highly complex and can comprise varying

amounts of other components including lipids, sugars, amino acids, phosphorylated residues and mineral elements (Tester et al. 2004).

Jasmine rices are consumed as close to harvest as possible, while basmati consumers prefer to store their rice for about a year before consumption, an action which changes the flavour and aroma of the rice significantly (Bhattacharjee et al. 2002). Consumers can readily tell the difference between a jasmine and a basmati rice (Champagne

et al. 2010), and Fig. 2 shows that the metabolomics analyses separate jasmine varieties from different parts of the world from the basmati varieties. In the present study, volatile and primary polar compounds in the grains of fresh jasmine and stored basmati rice are clearly different (Fig. 3a–c). However the compounds detected by DI-MS do not classify the types quite so well (Fig. 2d, e). None of the ten most discriminating volatile compounds, of which six were higher in basmati and four higher in jasmine, could be identified (Table 2). Seven of the ten most discriminating primary polar compounds were identified, but no clear biochemical relationship can be extrapolated. Nevertheless, the volatile and primary polar compounds are the most discriminating between these two types of rice, and are potentially valuable sources of markers, despite not yet being annotated.

Storage of basmati rice for 1 year leads to essential changes to the texture and cooking properties of the grains (Martin and Fitzgerald 2002), and also, the development of the typical basmati aroma and flavour (Bhattacharjee et al. 2002). All five platforms separated the fresh and stored basmati varieties (Fig. 4) indicating that the chemical changes occurring on storage are extensive. Again, only a few metabolites could be identified (Table 3), but we can speculate that lipid oxidation is likely to occur, leading to a different suite of odorous and malodorous compounds in the stored grains, such as fatty acids, aldehydes and furans (Zhou et al. 2003).

Even though we have demonstrated significant chemical differences between jasmine and basmati rices in the forms that they are generally used (i.e. fresh jasmine and stored/aged basmati) (Fig. 3; Table 2), as this stands, this is not useful to rice improvement programs, where selections are made on freshly harvested material. However, here we also demonstrate that freshly-harvested basmati and jasmine rice are clearly metabolically different (Fig. 5). All platforms could discriminate the types, providing potential opportunities for selection tools for breeding programs targeting both markets. While many of the discriminatory compounds have yet to be annotated, they can already be reliably detected by the platforms used, which is important for use in a breeding context. We have previously shown, in both rice and *Arabidopsis*, that metabolomic information from the platforms used can be exploited for association mapping, irrespective of any annotation (Calingacion et al. 2012; Fu et al. 2009; Keurentjes et al. 2006). Therefore, as long as compounds can be reliably detected, and heritability is confirmed, it should be possible in the future to identify genes that associate with these important discriminating, but as yet unidentified metabolites.

Stored basmati is not considered ‘rancid’ by basmati consumers whereas stored jasmine rice is (Laohakunjit and Kerdchoechuen 2007). Stored jasmine rice is often referred

to as being ‘musty’ and ‘old’ and is unacceptable to consumers. On all platforms we see clearly that in jasmine rices metabolite profiles change as storage time increases (Fig. 6). Importantly, the metabolic shifts do not involve the same compounds that discriminate between fresh and stored basmati rice. This indicates that aging proceeds differently in the two types of rice, in agreement with consumer data. Interestingly, it is clear that RD 33, RD 6 and KDML 105 differ from PTT 1 when freshly harvested, but after 12 months of storage at room temperature, there appears to be little difference between the metabolomes of KDML and PTT 1 (Fig. 6). KDML 105 is marketed in the high quality Hom Mali class but although PTT 1 is not (Pitiphunpong et al. 2011), it is often found as an adulterant in rice marketed as Hom Mali rice (Attaviroj et al. 2011). The data presented here can (i) be used to support policy development in Thailand in the areas of rice marketing, storage, and stockpiling, and (ii) provide potential tools for detecting adulteration of the Hom Mali class in fresh samples of rice.

The jasmine variety PTT 1, which has been improved through breeding for a number of agronomic characteristics, does not cluster with the other Thai jasmine rices (Figs. 2, 6), and indeed, was often misclassified in the other comparisons. This variety is a semi-dwarf jasmine rice that has undergone several breeding cycles (www.irri.org). RD 6 was a product of mutation breeding using KDML 105 (Somrith 1998), and RD 33 is the product of a back-crossing program using a variety carrying blast resistance as the donor parent and KDML105 as the recurrent parent (Chamarek et al. 2008). Therefore RD 33 and RD 6 are both derivatives of KDML 105 while PTT 1 is not. Recently, the genetic structure of Thai rices was determined, and PTT 1, while falling close to the other Thai rices, was one of two genetically isolated accessions (Chakhonkaen et al. 2012). This is consistent with the biochemical compositional differences between PTT 1 and the other jasmine rices in many of the datasets as observed here.

As with many other crops in the past, breeding for rice quality has often taken second place to crop yield and disease resistance. This was primarily because of a pre-occupation with productivity at the start of the green revolution, and subsequent to that, limited resources and the lack of a sufficient number of appropriate tools for quality evaluation as a means to assist the rice breeder to select also for quality attributes (Calingacion et al. 2014). Large-scale sensory analyses are limited because of the high costs and the often small amount of material that is produced during cultivar development. Systematic robust, non-subjective, predictive models for quality evaluation are therefore needed to allow us to visualise and exploit key traits and subsequently, help support direct breeding

strategies (Limpawattana et al. 2008). For this we need a broad-based analytical approach to complement already-existing genome-based approaches (e.g. Zhao et al. 2010). Metabolomics platforms have the potential to bridge this gap in our analytical potential, and in combination with other methodologies such as Genotyping by Sequencing (GBS), we have developed a powerful approach not only to understand the biochemical differences but also the genetics driving these. Their further use in helping us to continue to build a more detailed picture of the complexity of rice aroma and rice taste will be a very welcome addition to the plant breeders arsenal of tools required for breeding new high-yielding and high-quality rice varieties and targeting these accurately to specific local and international export markets. A logical next step would be to link the biochemistry determined by genetics as described here (which plays a major role in consumer choice in a retail situation) to the biochemistry resulting from cooking in order to understand better how the natural biochemical profile is converted into the processed one which plays the second major role by determining consumer preference at kitchen level.

5 Conclusions

Here, we demonstrate what rice consumers of course already know, that grains from jasmine and basmati rice have significantly different metabolomes which is consistent with recognized differences in sensory properties. We can now specify these differential properties and the most significant differences were found in the groups of volatile (SPME–GC–MS) and primary polar compounds (GC–TOF–MS) suggesting that these platforms have greatest potential value in future application. Importantly, we have also shown that the differences in biochemical profile between the two types is not simply a function of ageing. With all platforms, fresh Thai jasmine rice and fresh basmati rice were clearly found to differ, thus providing the potential to develop selection tools for breeding programs. Furthermore, we have shown differences between major Thai rice varieties that can lead to policy development in important areas of rice marketing. The next steps are to focus on annotating the discriminating compounds and to develop a conduit of delivery to rice breeding programs to enable robust selection of traits that define sensory and cooking quality.

Acknowledgments This work was supported by the EU Framework Programme VI Project META-PHOR (Grant No. FOOD–CT–2006–036220). RDH, JH, RCHdV and RM also acknowledge funding support from the Netherlands Metabolomics Centre (NMC) and the Centre for Biosystems Genomics (CBSG) which are part of the

Netherlands Genomics Initiative (NGI). The INQR members acknowledge support from their institutes for rice trials and sample preparation.

Compliance with ethical standards

Conflict of interests The authors have declared that no competing interests exist.

Ethical Approval This article does not contain any data related to human or animal studies.

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

References

- Allwood, J. W., Erban, A., de Koning, S., Dunn, W., Luedemann, A., Lommen, A., et al. (2009). Inter-laboratory reproducibility of fast gas chromatography–electron impact–time of flight mass spectrometry (GC–EI–TOF/MS) based plant metabolomics. *Metabolomics*, 5, 479–496. doi:10.1007/s11306-009-0169-z.
- Archak, S., Lakshminarayanareddy, V., & Nagaraju, J. (2007). High-throughput multiplex microsatellite marker assay for detection and quantification of adulteration in basmati rice (*Oryza sativa*). *Electrophoresis*, 28, 2396–2405. doi:10.1002/elps.200600646.
- Attaviroj, N., Kasemsumran, S., & Noomhorm, A. (2011). Rapid variety identification of pure rough rice by fourier-transform near-infrared spectroscopy. *Cereal Chemistry Journal*, 88, 490–496. doi:10.1094/cchem-03-11-0025.
- Baker, J., Hawkins, N., Ward, J., Lovegrove, A., Napier, J., Shewry, P. R., et al. (2006). A metabolomic study of substantial equivalence of field grown GM wheat. *Plant Biotechnology Journal*, 4, 381–392.
- Bhattacharjee, P., Singhal, R., & Kulkarni, P. (2002). Basmati rice: A review. *International Journal of Food Science & Technology*, 37, 1–12.
- Bradbury, L. M. T., Henry, R. J., Jin, Q. S., Reinke, R. F., & Waters, D. L. E. (2005). A perfect marker for fragrance genotyping in rice. *Molecular Breeding*, 16, 279–283.
- Bryant, R. J., & McClung, A. M. (2011). Volatile profiles of aromatic and non-aromatic rice cultivars using SPME/GC-MS. *Food Chemistry*, 124, 501–513.
- Buttery, R. G., Ling, L. C., & Juliano, B. O. (1983). Additional aroma components of rice. *Abstracts of Papers of the American Chemical Society* 186, 164-AGFD.
- Buttery, R. G., Ling, L. C., & Juliano, B. O. (1982). 2-Acetyl-1-pyrroline: An important aroma component of cooked rice. *Chemistry and Industry*, 23, 958–959.
- Buttery, R. G., Turnbaugh, J. G., & Ling, L. C. (1988). Contribution of volatiles to rice aroma. *Journal of Agricultural and Food Chemistry*, 36, 1006–1009.
- Calingacion, M., Boualaphanh, C., Daygon, D., Anacleto, R., Biais, B., Deborde, C., et al. (2012). A metabolomics approach to identify new traits of rice quality in traditional and improved varieties of Laos. *Metabolomics*, 8, 771–783. doi:10.1007/s11306-011-0374-4.
- Calingacion, M., Laborte, A., Nelson, A., Resurreccion, A., Concepcion, J. C., Daygon, V. D., et al. (2014). Diversity of global rice

- markets and the science required for consumer-targeted rice breeding. *PLoS One*, 9, e85106. doi:10.1371/journal.pone.0085106.
- Chakhonkaen, S., Pitnjam, K., Saisuk, W., Ukoskit, K., & Muangprom, A. (2012). Genetic structure of Thai rice and rice accession obtained from the International Rice Research Institute. *Rice*, 5, 19.
- Chamarek, V., Konboon, Y., Wongboon, W., Jongee, B., Romyen, P., & Lertna, S. (2008). *Drought-prone rainfed lowlands Ubun Rachathani and Nakorn Ratchasima, Thailand. Consortium for Unfavourable Rice* (pp. 458–460). Laguna: IRRI.
- Champagne, E. T. (2008). Rice aroma and flavor: A literature review. *Cereal Chemistry*, 85, 445–454. doi:10.1094/CCHEM-85-4-0445.
- Champagne, E. T., Bett-Garber, K. L., Fitzgerald, M. A., Grimm, C., Lea, J., Ohtsubo, K., et al. (2010). Important sensory properties differentiating premium rice varieties. *Rice*, 3, 270–281.
- Eriksson, L., Johansson, E., Kettaneh-Wold, N., Wikström, C., & Trygg, J. (2006). *Multi- and megavariate data analysis part I: Basic principles and applications* (Vol. 1). Umea: Umetrics AB.
- Fitzgerald, M. A., McCouch, S. R., & Hall, R. D. (2009). More than just a grain of rice, the global quest for quality. *Trends in Plant Science*, 14, 133–139. doi:10.1016/j.tplants.2008.12.004.
- Fitzgerald, M. A., Sackville-Hamilton, N. R., Calingacion, M. N., Verhoeven, H. A., & Butardo, V., Jr. (2008). Is there a second gene for fragrance in rice? *Plant Biotechnology Journal*, 6, 416–423.
- Fu, J., Keurentjes, J. J. B., Bouwmeester, H., America, T., Verstappen, F. W. A., Ward, J. L., et al. (2009). System-wide molecular evidence for phenotypic buffering in Arabidopsis. *Nature Genetics*, 41, 166–167.
- Garris, A. J., Tai, T. H., Coburn, J., Kresovich, S., & McCouch, S. R. (2005). Genetic structure and diversity in *Oryza sativa* L. *Genetics*, 169, 1631–1638.
- Hall, R. D. (2006). Plant metabolomics: From holistic hope, to hype, to hot topic. *New Phytologist*, 169, 453–468.
- Hall, R. D. (2007). Food metabolomics: META-PHOR A new European research initiative. *Agro Food Industry Hi-Techniques*, 18, 14–16.
- Hall, R. D. (2011). Plant Metabolomics in a nutshell: Potential and future challenges. In R. D. Hall (Ed.), *Biology of plant metabolomics annual plant reviews* (Vol. 43, pp. 1–24). Oxford: Wiley Blackwell.
- Hall, R., & Hardy, N. (2012). Practical applications of metabolomics in plant biology. In N. W. Hardy & R. D. Hall (Eds.), *Plant metabolomics. Methods in Molecular Biology* (Vol. 860, pp. 1–10). Totowa: Humana Press.
- Keurentjes, J. J. B., Fu, J., de Vos, C. H. R., Lommen, A., Hall, R. D., Bino, R. J., et al. (2006). The genetics of plant metabolism. *Nature Genetics*, 38, 842–849.
- Khush, G. S., & Juliano, B. O. (1985). Breeding for high-yielding rices of excellent cooking and eating qualities. Paper presented at the International Rice Research Conference, Los Banos, Laguna, June 1–5, 1985.
- Kovach, M. J., Calingacion, M., Fitzgerald, M. A., & McCouch, S. R. (2009). The origin and evolution of fragrance in rice (*Oryza sativa* L.). *Proceedings of the National Academy of Science*, 106, 14444–14449.
- Laohakunjit, N., & Kerdchoechuen, O. (2007). Aroma enrichment and the change during storage of non-aromatic milled rice coated with extracted natural flavor. *Food Chemistry*, 101, 339–344.
- Laursen, K., Hesselhøj Hansen, T., Persson, D., Schjoerring, J., & Husted, S. (2009). Multi-elemental fingerprinting of plant tissue by semi-quantitative ICP-MS and chemometrics. *Journal of Analytical Atomic Spectroscopy*, 24, 1198–1207.
- Limpawattana, M., Yang, D. S., Kays, S. J., & Shewfelt, R. L. (2008). Relating sensory descriptors to volatile components in flavor of specialty rice types. *Journal of Food Science*, 73, S456–S461. doi:10.1111/j.1750-3841.2008.00952.x.
- Lommen, A. (2009). MetAlign: Interface-driven, versatile metabolomics tool for hyphenated full-scan mass spectrometry data preprocessing. *Analytical Chemistry*, 81, 3079–3086.
- Luedemann, A., Strassburg, K., Erban, A., & Kopka, J. (2008). TagFinder for the quantitative analysis of gas chromatography - mass spectrometry (GC-MS) based metabolite profiling experiments. *Bioinformatics*, 24, 732–737.
- Martin, M., & Fitzgerald, M. A. (2002). Proteins in rice grains influence cooking properties. *Journal of Cereal Science*, 36, 285–294.
- McNally, K. L., Childs, K. L., Bohnert, R., Davidson, R. M., Zhao, K., Ulat, V. J., et al. (2009). Genomewide SNP variation reveals relationships among landraces and modern varieties of rice. *Proceedings of the National Academy of Sciences*, 106, 12273–12278. doi:10.1073/pnas.0900992106.
- Moing, A., Maucourt, M., Renaud, C., Gaudillere, M., Brouquisse, R., Lebouteiller, B., et al. (2004). Quantitative metabolic profiling by 1-dimensional ¹H-NMR analyses: Application to plant genetics and functional genomics. *Functional Plant Biology*, 31, 889–902. doi:10.1071/FP04066.
- Pitiphunpong, S., Champangern, S., & Suwannaporn, P. (2011). The jasmine rice (KDML 105 variety) adulteration detection using physico-chemical properties. *Chiang Mai Journal of Science*, 38, 105–115.
- Saito, K., Dixon, R. A., & Willmitzer, L. (2006). *Plant metabolomics*. Heidelberg: Springer.
- Sekhar, B. P. S., & Reddy, G. M. (1982). Amino acid profiles in some scented rice varieties. *Theoretical and Applied Genetics*, 62, 35–37. doi:10.1007/bf00276278.
- Singh, V., Singh, U., & Khush, G. (2000). Basmati rice of India. *Aromatic Rices*, 135–154.
- Sivaranjani, A. K. P., Pandey, M. K., Sudharshan, I., Kumar, G. R., Madhav, M. S., Sundaram, R. M., et al. (2010). Assessment of genetic diversity among basmati and non-basmati aromatic rices of India using SSR markers. *Current Science*, 99, 221–226.
- Smit, S., van Breemen, M. J., Hoefsloot, H. C. J., Smilde, A. K., Aerts, J. M. F. G., & de Koster, C. G. (2007). Assessing the statistical validity of proteomics based biomarkers. *Analytica Chimica Acta*, 592, 210–217. doi:10.1016/j.aca.2007.04.043.
- Somrith, B. (1998). *Rice production in Thailand*. Cairo: International Rice Commission.
- Steele, K. A., Ogden, R., McEwing, R., Briggs, H., & Gorham, J. (2008). InDel markers distinguish basmatis from other fragrant rice varieties. *Field Crops Research*, 105, 81–87.
- Stewart, D., Shepherd, L. V., Hall, R. D., & Fraser, P. D. (2011). Crops and tasty, nutritious food—How can metabolomics help? In R. D. Hall (Ed.), *Biology of plant metabolomics. Annual plant reviews* (Vol. 43, pp. 181–217). Oxford: Wiley Blackwell.
- Strehmel, N., Hummel, J., Erban, A., Strassburg, K., & Kopka, J. (2008). Retention index thresholds for compound matching in GC-MS metabolite profiling. *Journal of Chromatography*, 871, 182–190.
- Supakornchuwong, C., & Suwannaporn, P. (2012). Attitudes toward rice compared with potatoes and pasta among British, French, Dutch and Belgian consumers. *Journal of Sensory Studies*, 27, 71–77. doi:10.1111/j.1745-459X.2011.00369.x.
- Tester, R. F., Karkalas, J., & Qi, X. (2004). Starch—Composition, fine structure and architecture. *Journal of Cereal Science*, 39, 151–165.
- Tikunov, Y. M., Laptinok, S., Hall, R. D., Bovy, A., & De Vos, R. C. H. (2012). MSClust: A tool for unsupervised mass spectra

- extraction of chromatography-mass spectrometry ion-wise aligned data. *Metabolomics*, 8, 714–718.
- van den Berg, R., Hoefsloot, H., Westerhuis, J., Smilde, A., & van der Werf, M. (2006). Centering, scaling, and transformations: Improving the biological information content of metabolomics data. *BMC Genomics*, 7, 142. doi:[10.1186/1471-2164-7-142](https://doi.org/10.1186/1471-2164-7-142).
- Verhoeven, H. A., Jonker, H. H., de Vos, R. C. H., & Hall, R. D. (2012). Solid-phase micro-extraction (SPME) GC-MS analysis of natural volatile components in melon and rice. In N. G. Hardy & R. D. Hall (Eds.), *Plant metabolomics methods* (pp. 85–100). Ithaca, NY: Humana Press.
- Ward, J., Forcat, S., Beckmann, M., Bennett, M., Miller, S., Baker, J., et al. (2010). The metabolic transition during disease following infection of *Arabidopsis thaliana* by *Pseudomonas syringae* pv. tomato. *The Plant Journal*, 63, 443–457.
- Westerhuis, J. A. (2008). Assessment of PLS-DA cross validation. *Metabolomics*, 4, 81–89.
- Yoshihashi, T., Nguyen, T. T. H., & Kabaki, N. (2004). Area dependency of 2-acetyl-1-pyrroline content in an aromatic rice variety, Khao Dawk Mali 105. *Japan Agricultural Research Quarterly*, 38, 105–109.
- Zhao, K., Wright, M., Kimball, J., Eizenga, G., McClung, A., Kovach, M., et al. (2010). Genomic diversity and introgression in *O. sativa* reveal the impact of domestication and breeding on the rice genome. *PLoS One*, 5, e10780. doi:[10.1371/journal.pone.0010780](https://doi.org/10.1371/journal.pone.0010780).
- Zhou, Z., Blanchard, C., Helliwell, S., & Robards, K. (2003). Fatty acid composition of three rice varieties following storage. *Journal of Cereal Science*, 37, 327–335.