



# Grain size and shape reflects variability in metabolite and elemental composition in traditional rice varieties

Jhansi Narmada Reddy Kotamreddy<sup>1</sup> · Monica Barman<sup>1</sup> · Laxuman Sharma<sup>2</sup> · Adinpunya Mitra<sup>1</sup>

Received: 22 July 2019 / Accepted: 23 September 2019 / Published online: 28 September 2019  
© Springer Science+Business Media, LLC, part of Springer Nature 2019

## Abstract

Traditional rice varieties serve as a pool for diverse germplasm. Present study investigates the compositional variability of metabolites and elements in traditional rice varieties with varying grain size and shape. Grain quality characterization of traditional rice varieties including medium, small and bold shaped grain types was performed. Additionally, whether these differences in the size and shape of varieties are reflected in the metabolite and elemental composition of their grains was explored. Two major clusters were obtained from hierarchical clustering of varieties based on the similarities of grain size and shape. The second cluster was further separated into four sub-clusters (clusters II, III, IV and V). A single variety representative of each cluster was selected for further analysis. Metabolite analysis and elemental analysis of grains was carried out by gas chromatography mass spectrometry (GC–MS) and inductively coupled plasma mass spectrometry (ICP–MS), respectively. A total of 47 metabolites belonging to different metabolite groups and a total of 35 minerals were identified from dehulled grains of selected varieties. Distinct differences detected amongst the representative varieties from five clusters with respect to their metabolite and elemental profiles were further confirmed by partial least square discriminant analysis (PLS-DA) of the data obtained from both GC–MS and ICP–MS platforms. This study further highlights the usefulness of metabolite and elemental profiling complemented by suitable statistical tools to explore the prospect of an inherent correlation between grain quality features, biochemical composition and genetic background.

**Keywords** Traditional rice varieties · Metabolites · Elements · Gas chromatography mass spectrometry · Inductively coupled plasma mass spectrometry · Partial least squares discriminant analysis

## Introduction

Traditional rice varieties are those that have come into existence over a series of evolution and adaptation processes in a particular agro-environment and agricultural practices. Preferences of local farmers combined with natural, diverse agro-climates, soil and varied agricultural practices led to huge diversity of germplasm among traditional rice varieties

[1]. Popularization of high yielding varieties, led by green revolution created a setback to the cultivation of traditional rice varieties [2]. Low yields of traditional rice varieties also favored the shift to cultivation of high yielding varieties by farmers. However, each traditional rice variety has a specialty trait that has either agricultural and/or food significance. Some varieties are suited to particular type of soil, others can withstand drought and/or flood, while some others are salinity tolerant [3, 4]. Some traditional varieties are known for their fragrance and/or color of the rice, while some are popular for their health benefits owing to their nutrient content and bioactivities [5]. If specialty traits of traditional rice varieties can be coupled with agronomic competence of high yielding rice varieties, the outcome may create enthusiasm among the growers. This in turn, if matched by willingness of consumers to try out these improved rice varieties, may have a positive effect on the economy of Indian farmers.

**Electronic supplementary material** The online version of this article (<https://doi.org/10.1007/s11694-019-00273-7>) contains supplementary material, which is available to authorized users.

✉ Adinpunya Mitra  
adin@iitkgp.ac.in; adinpunya@gmail.com

<sup>1</sup> Natural Product Biotechnology Group, Agricultural and Food Engineering Department, Indian Institute of Technology Kharagpur, Kharagpur 721 302, India

<sup>2</sup> Department of Horticulture, Sikkim University, Gangtok, Sikkim 737 102, India

In order to develop improved varieties with both good yield and grain quality, it is imperative to identify traditional rice varieties with traits that influence the grain quality, a suitable high yielding background, and to select them for a breeding program. Grain quality evaluation encompasses measurement of physical traits such as length (L) and breadth (B) of grains [6], yield trait such as thousand grain weight (TGW) and cooking quality traits that influence the texture of cooked rice, such as amylose content (AC), gelatinization temperature (GT), which is an indirect indicator of cooking time [6]. Other grain quality features important for millers and consumers include the shape and size of grains, which are influenced by the length and breadth of grains. Grain size and shape affect milling quality of rice and also the appearance of milled rice grain. In rice, these traits are under polygenic control and each of the sub-species of rice (*indica*, *japonica*) carries diverse combinations of alleles for grain size and shape [7]. Majority of the previous studies however aimed at achieving the right combination of alleles resulting in the desired shape and size of grain that have potential implications in developing high yielding varieties with particular grain morphology to satisfy various quality requirements of consumers [7]. As there are growing numbers of health-conscious consumers, nutritional quality adds a new dimension to the visual appeal of grain. It would be interesting to determine if there are any plausible qualitative or quantitative differences in the metabolite content of grains with different shape and size.

Apart from these traits, biochemical composition of grains including metabolites and mineral elements also affect the nutritional quality of rice, which in turn depends on type of rice [8, 9]. For instance, basmati and jasmine type fragrant varieties differ significantly in their metabolome, owing to differences in the composition of volatile and primary polar compounds. Major polar compounds causing the distinction between the two fragrant varieties are organic acids. Organic acids such as myristic acid, fumaric acid and nonanoic acid are present in higher quantities in basmati rice, whereas higher amounts of erythronic acid and 2-amino-malonic acid are present in jasmine rice [8]. Metabolite analysis of 26 *indica* rice varieties revealed varietal differences in their metabolic constituents. Hierarchical clustering based on the metabolite profile resulted in grouping of varieties into five clusters. A distinct difference in the contents of organic acids, fatty acids, amino acids, sugar alcohols, sucrose, campesterol and fucosterol among the varieties belonging to different clusters was observed [10]. Metabolite profiling in combination with multivariate analysis demonstrated distinct differences in the metabolite profile of dehulled colored and non-colored rice [11, 12]. Frank et al. have effectively demonstrated the use of multivariate analysis to separate red, black and non-colored rice based on their metabolite profiles. They observed that colored

rice was shown to contain higher amounts of amino acids, organic acids, fatty acids and fatty acid methyl esters [11]. Another study by Kim et al. revealed a strong correlation between primary metabolites and the individual phenolics contents of grains upon metabolite profiling of colored rice varieties. Their study also demonstrated the usefulness of multivariate data analysis to discriminate between colored and non-colored rice grains [12]. Discrimination of varieties based on their geographical origin was demonstrated by multi-elemental profiling and subsequent multivariate analysis [9, 13–15]. Chung et al. validated the geographical origin of rice by performing multi-elemental analysis and chemometric studies. Their quantification analysis showed that 11 out of 25 elements such as, Copper (Cu), silver (Ag), zinc (Zn), chromium (Cr), calcium (Ca) have contributed to distinguish the rice varieties of different origin [13]. A similar study by Cheajesadagul et al. showed differentiation of Thai jasmine rice from rice samples of different geographical origin using multi-element fingerprinting followed by multivariate analysis. Elements such as, boron (B), molybdenum (Mo), cadmium (Cd), Cu and magnesium (Mg) served as the major variables of such separation of Thai jasmine rice from other samples. Furthermore, same strategy was utilized to separate Thai jasmine rice based on the region of origin within Thailand. Elements such as Mo, rubidium (Rb), Cd and B are responsible for distinguishing Thai jasmine rice of different regions of Thailand [14]. In contrast, application of elemental analysis to classify type of rice [9] has received less attention, even though this information has a direct relevance to nutritional quality of rice. Promchan et al. have used elemental imaging to classify rice varieties based on their origin and type [9]. To our knowledge, the findings reported in this manuscript is the only study that demonstrates the application of multi-elemental analysis to differentiate rice grains based on shape and size.

It is therefore important to look upon the grain quality features of selected traditional rice varieties with special emphasis on their nutritional profiles. Further, in order to elucidate compositional differences of metabolites and elements between the rice varieties, their metabolic and elemental profiles should be subjected to multivariate statistical analysis. This manuscript reports our attempt to characterize 12 traditional rice varieties in terms of grain quality features. Further, clustering of these varieties based on the differences and similarities in their size and shape was done. Partial least squares discriminant analysis (PLS-DA) of the data obtained by gas chromatography mass spectrometry (GC-MS) and inductively coupled plasma mass spectrometry (ICP-MS) platforms from representative varieties of each cluster was done to see if grain morphological differences are reflected in their metabolite and elemental profiles. This study also encourages the utilization of new selection tools for breeding programs aimed at developing improved rice.

## Materials and methods

### Plant materials

Twelve traditional rice varieties (Table S1) were grown in the farm of Agricultural and Food Engineering department of IIT Kharagpur (22.330239°N latitude, 87.323653°E longitude and 29 m altitude) in the year 2015 (July–November). The set included three non-fragrant and nine fragrant traditional rice varieties. A number of these varieties are currently cultivated across the state of West Bengal, India. These offer vast genetic diversity which is important for breeding initiatives.

Seedlings (17–21 days old) were transplanted in experimental field plot (2.0 × 2.5 m) with plant to plant and row to row spacing of 25 cm × 25 cm. Standard agricultural practices were followed. Nitrogen (N), phosphorous (P) and potassium (K) were applied in the ratio 100:50:60 kg/ha. Both P and K were applied before transplantation, while N was applied in three equal split doses. Each variety was replicated three times in each main plot. Grains from each subplot were harvested at maturity and sun-dried until the moisture content reached 12–14%. Moisture was measured using universal moisture testing machine (Indosaw, Osaw Industrial Products, India).

Around 500 g of paddy was dehulled using Satake rubber roll dehusker (Satake engineering Co. Ltd, Japan). Twenty grams of dehulled rice were ground and screened. Ground rice powder was stored at –80 °C until further analysis for determining metabolite and elemental composition of grains.

### Chemical reagents

All the chemicals used in this study were purchased from Sigma-Aldrich, now incorporated in Merck. All the solvents used are MS grade and were purchased from Merck.

### Measurement of grain quality traits

Grain dimensions were measured using Vernier calipers (Mitutoyo, Japan). A total of ten grains per each variety were taken for the measurement of grain dimensions. Classification of grains into different shapes and sizes was done based on the length and L/B ratio of grain [16]. A more detailed description of size and shape based on grain dimensions of the studied varieties is given in Supplementary material (Table S1). A count of 1000 grains was done to measure TGW, a yield associated trait.

Cooking quality parameters like AC and alkali spreading value (ASV) were determined from milled rice. Dehulled rice was polished using Satake polisher (Satake engineering

Co. Ltd, Japan). Broken rice grains were separated from whole grains in a grader. AC and ASV of milled grains were determined following the method of Bhonsle and Krishnan [17]. AC of each variety was calculated against a standard curve of amylose and expressed in mg per gram of fresh weight. ASV, which in turn determines GT, was determined by placing six milled grains in a petri plate (9 mm diameter) equidistant from each other. Then 10 mL of 1.7% of KOH was added to each plate. Plates are covered and incubated at room temperature for 23 h. ASV was calculated using a seven-point scale [16].

### Metabolite extraction

Five representative varieties were selected from each cluster on the basis of their superior agronomic performances in the studied agro-climatic zone (data not shown) for metabolite profiling using GC–MS. Metabolite extraction and derivatization was done following the method of De and Nag [10] with slight modifications. Dehulled grain powder (100 mg) was used for extraction of polar metabolites with 1 mL of 100% methanol (mixture was vortexed thoroughly and incubated at 70 °C for 15 min). Ribitol and norleucine (20 µL of 0.2 mg/mL solution) were added prior to extraction as internal standards. Collected supernatant was distributed into microcentrifuge tube (50 µL). Supernatant was then vacuum dried. Dried methanolic extracts were methoxyaminated (20 µL of 30 mg/mL solution of methoxyamine hydrochloride in pyridine) for 120 min at 37 °C, followed by derivatization with *N*-methyl-*N*-(trimethylsilyl) trifluoroacetamide (TMS) and *N*-tert-butyldimethylsilyl-*N*-methyltrifluoroacetamide with 1% tert-butyldimethylchlorosilane (TBS). During derivatization the samples were incubated at 37 °C for TMS analysis and 65 °C for TBS analysis. Retention times were calibrated by addition of a C8–C20 and C20–C40 *n*-alkane mixture for TMS analysis and C8–C20 *n*-alkane mixture for TBS analysis to each sample prior to GC injection.

### Metabolite analysis by GC–MS

The derivatized metabolites were analyzed on a GC–MS instrument (TRACE™ 1300, Thermo Scientific) by injecting 1 µL of the sample with a split ratio of 10:1. The separation of derivatized compounds was performed on a TB-5 column (30 m × 0.32 mm id, film thickness 0.25 µm) with helium as a carrier gas. For TMS analysis, conditions maintained by GC–MS were followed according to Samanta et al. [18], with slight modifications. The inlet temperature was set at 280 °C and the flow rate of carrier gas through the column was 1.5 mL min<sup>–1</sup>. The column oven temperature of GC was held at 70 °C for 4 min, and then ramped to 300 °C at a rate of 5 °C min<sup>–1</sup> and held

for 10 min. The interface and the ion source temperatures were set at 300 °C and 230 °C, respectively. For TBS analysis, conditions maintained by GC–MS were followed according to Jacobs et al. [19], with slight modifications. The inlet temperature was set at 230 °C and the flow rate of carrier gas through the column was 1.5 mL min<sup>-1</sup>. The column oven temperature of GC was held at 100 °C for 10 min, and then ramped to 106 °C at a rate of 1 °C min<sup>-1</sup> and to 300 °C at the rate of 7 °C min<sup>-1</sup> and finally held for 10 min. The interface and the ion source temperatures were set at 300 °C and 230 °C, respectively. The electron energy for ionization mode was set at 70 eV (ISQ QD, single quadrupole mass spectrometer, Thermo Scientific) for both TMS and TBS analysis and mass spectra were recorded using full scan monitoring mode with a mass scanning range of *m/z* 40–600 a.m.u. Reagent blanks and quality control (QC) samples were included in the runs. The QC samples were a mixture of all the samples (pooled QC), and were injected after every seven to nine samples. Raw files obtained after GC–MS analysis were converted to NetCDF file format and deconvoluted by Automated Mass Spectral Deconvolution and Identification System (AMDIS). Peaks were identified by comparing the mass spectrum of the compound to that from mass spectral library NIST 14 (if at least three specific *m/z* fragments per each compound are matched) and comparing retention index from the literature. The relative abundance of various metabolites was calculated as the ratio of peak area of an individual compound to that of an internal standard and normalized to sample weight before statistical analysis.

### Elemental extraction and profiling by ICP-MS analysis

Dehusked rice grain powder was digested using an acid mixture (a ratio of 9:4 nitric acid to perchloric acid). Digestion was performed following the method of [20] with a few modifications. Briefly, 0.5 g of ground dehulled rice sample was digested with 20 mL of acid. The digestion flask was heated until the digest was clear. The clear digest obtained was later diluted to 50 mL with distilled water. A blank was prepared with same treatment and without any sample.

Acid digested extract was then analyzed by ICP-MS system (Perkin Elmer, Nex ION 300 X) with cross flow nebulizer as essentially described by Mumm et al. [8]. Mineral contents of digested samples are estimated by multi-elemental standard solutions no. 1, 3 and 5 (provided by Perkin Elmer). The concentration of estimated minerals is expressed in ng/g of fresh weight of the sample.

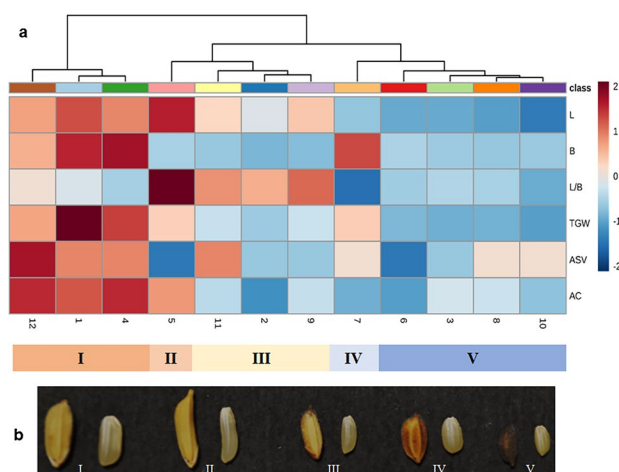
### Statistical analysis

All statistical analyses were performed on means of triplicate measurements. Supplementary data on quantitative traits were statistically analyzed using SPSS software (Version 17.0). Metaboanalyst, an open source software was used for clustering, heat map generation and PLS-DA [21]. After clustering of varieties based on their grain quality features and heat map generation, a variety representative of each cluster was selected for metabolite and elemental analysis. After GC–MS and ICP-MS analysis, data generated was subjected to PLS-DA analysis to evaluate similarities and differences between varieties with varying size and shape. Output from PLS-DA consists of scores plot and loadings plot. Scores plot represent the separation of samples, while loadings plot signifies the variables causing the separation. Variable importance in the projection (VIP) scores were also obtained from PLS-DA. Variables with high VIP scores contribute most to the separation of samples.

## Results and discussion

### Grain quality characterization and clustering

Grain quality is affected by both physical traits such as length, breadth, L/B ratio and TGW and cooking quality traits such as AC and ASV. Therefore, grain quality diversity of selected traditional rice varieties was assessed. Quantitative representation of all these characteristics is done in the form of a table in supplementary data (Tables S2, S3). Visual representation of variation of grain quality characters in different rice varieties is done in the form of a heat map (Fig. 1a). Considerable diversity of studied characteristics was observed among the varieties. Numerous reports supporting this observation are available in literature, depicting the diversity of traditional rice varieties regarding their grain quality features [1, 22–25]. Saha et al. collected the indigenous rice varieties of West Bengal, India and also performed morphological characterization of the collected germplasm. They reported a huge diversity among the studied germplasms in terms of their agronomic features, physical and cooking quality traits of grain including aroma [1]. In another study, Roy and Sharma collected germplasm from various agro-ecological pockets of West Bengal, India, and 16 agro-morphological traits and eight grain quality traits were evaluated from those collected landraces. All the traits evaluated showed genetic diversity among the studied landraces, serving as a source of information for crop breeders [23]. Agro-morphological characterization of 124 landraces of Nagaland, India revealed a considerable genetic variability in plant architecture, physical and quality grain traits [24].



**Fig. 1** **a** Heat map and hierarchical clustering analysis generated from evaluation of grain quality traits of all 12 varieties. 1—Champakushi, 2—Radhunipagol, 3—Gobindobhog, 4—Kalonunia, 5—Dudheswar, 6—Badshabhog, 7—Kanakchur, 8—Gopalbhog, 9—Kataribhog, 10—Tulsibhog, 11—Tulaipanji and 12—Tulsimukul. *L* Length of grain, *B* breadth of grain, *L/B* length-to-breadth ratio, *TGW* 1000-grain weight, *ASV* alkali spreading value; *AC* amylose content. Heat map shows clustering of varieties (Cluster I, II, III, IV and V) based on the similarity of their grain size and shape and these clusters are labeled with different colors. In the heat map, lowest value is indicated by dark blue, the highest value by bright red and values in the middle are indicated by light grey, with a corresponding gradient between extremes. Rows and columns represent individual grain quality traits and average values of each trait in all 12 rice varieties, respectively. **b** Visual phenotypes displaying the differences in size and shape of the varieties selected from each cluster for metabolite and elemental analysis. 1—Champakushi, 2—Dudheswar, 3—Radhunipagol, 4—Kanakchur and 5—Tulsibhog (Color figure online)

To identify varieties with similar grain quality traits, cluster analysis was performed based on the above-mentioned traits, which classified the varieties into two major clusters and second cluster was separated into four sub-clusters, thus making altogether five clusters (Fig. 1a). From the clustering analysis, it is clear that grouping of all twelve varieties into separate clusters is caused by physical traits such as length, TGW and cooking quality traits such as AC. Length of grains also influences L/B ratio, which in turn determines the shape of grain. Grain dimensions are varietal dependent traits. Long, slender grains tend to break more during milling, than short, bold grains. Thus, grain size and shape have a significant effect on type of milling equipment to be used and milling recovery [26]. A comprehensive study by Xie et al. [27] demonstrated that a positive correlation exists among grain morphological traits, grain weight, and recovery of milled rice. Grain dimensions similarly influence cooking quality traits of grain [28]. AC is one of the grain quality traits that influence the texture of cooked rice [6]. Varieties with high amylose content tend to be dry and separate upon cooking, while varieties with low amylose content tend to be glutinous and sticky [29]. In this study,

varieties from cluster I are characterized by medium sized and medium shaped grains. Varieties of this cluster have high TGW and AC. Tulsimukul, Champakushi and Kalonunia constituted the first cluster and these are non-fragrant type rice. Cluster II comprised of a single variety (Dudheswar). It has a long and slender shaped grain. This is a fragrant rice variety with intermediate AC. Radhunipagol, Kataribhog and Tulaipanji are grouped into cluster III. They belong to medium sized and medium shaped grains and are fragrant type rice with low AC. Cluster IV is comprised of Kanakchur variety and is characterized by its distinct fragrance and bold shaped grain. Varieties grouped into cluster V are Gobindobhog, Badshabhog, Gopalbhog and Tulsibhog. They belong to short grain fragrant rice types. Varieties from this cluster have low AC and TGW. However, flavour of cooked rice is as important as its texture. Flavour is imparted to grains by their biochemical constituents [8]. These include metabolites, mineral elements and volatile compounds. Studying the biochemical composition of grains adds another dimension to the grain quality, and existence of any correlation between the metabolite and elemental composition of grains with known differences in size and shape is worth exploring. Accordingly, a variety from each cluster was selected which represents the major features of that particular cluster for metabolite and elemental analysis. They are Champakushi from cluster I, Dudheswar, only variety in cluster II, Radhunipagol in cluster III, Kanakchur, also only variety in cluster IV and Tulsibhog in cluster V. Picture of grains depicting the diversity of selected rice varieties in terms of their size and shape is given in Fig. 1b. In addition, PLS-DA of data obtained from GC-MS and ICP-MS platforms (by metabolite and elemental profiling of selected varieties) was performed.

### Compositional variation of metabolites in rice grains on the basis of their size and shape

Metabolite analysis of the rice grains was done by derivatization of methanolic extracts of dehulled rice powder and subsequent analysis by GC-MS. Derivatization was done by silylation using TMS and TBS. The derivatized products were run in GC and identified by their mass fragmentation spectra and confirmed by retention index calculation. GC-MS chromatograms corresponding to TMS and TBS analysis are given in Figs. S1 and S2 of supplementary material. A total of 47 metabolites were identified from all the five varieties. List of metabolites identified, along with three major *m/z* fragments considered for the identification, and corresponding CAS number are presented in the in Table 1. Major groups of metabolites identified were sugars, sugar alcohols, amino acids, fatty acids, organic acids, phytosterols, phenolics etc. Similar metabolite groups were identified from 26 different Indian indica rice varieties by De and Nag

**Table 1** List of metabolites identified by TMS and TBS derivatization and subsequent analysis by GC–MS

TMS derivatives				TBS derivatives			
Compound name	Group	m/z fragments	CAS No	Compound name	Group	m/z fragments	CAS no.
1,3-Propanediol	Other groups	147,130,115	17,887-80-8	Tiglic acid	Organic acid	157,75,158	–
Oxalic acid	Organic acid	147,73,148	18,294-04-7	Guaiacol	Other groups	166,181,151	–
Vanillic acid	Phenolic compound	73, 267,297	2078-15-1	Lactic acid	Organic acid	147,73,261	–
Glycerol-3-phosphate	Lipid precursor	73,357,299	31,038-11-6	Glycolic acid	Organic acid	73,147,189	67,226-76-0
Azelaic acid	Organic acid	73,75,55	17,906-08-0	L-Alanine	Amino acid	73,158,147	92,751-15-0
Fructose	Sugar	73,103,307	56,196-14-6	Glycine	Amino acid	73,147,218	107,715-88-8
Galactose	Sugar	73,147,205	128,705-64-6	Urea	Other groups	231,147,73	82,475-73-8
Glucose	Sugar	73,147,205	130,405-10-6	L-Valine	Amino acid	73,186,260	107,715-89-9
p-Coumaric acid	Phenolic compound	73,219,293	10,517-30-3	Leucine	Amino acid	73,200,57	107,715-90-2
Mannitol	Sugar alcohol	73,319,205	14,317-07-8	Isoleucine	Amino acid	73,200,274	92,771-658
Sorbitol	Sugar alcohol	73,147,319	13,829,212 <sup>a</sup>	GABA	Amino acid	73,147,75	110,024-92-5
Galactitol	Sugar alcohol	73,217,307	18,919-39-6	Succinic acid	TCA	73,75,147	98,847-52-0
Gluconic acid	Sugar acid	73,333,147	34,290-52-3	Proline	Amino acid	73,184,147	107,715-91-3
Palmitic acid	Fatty acid	313,73,117	55,520-89-3	Glycerol	Polyol	89,73,147	82,112-23-0
Myo-inositol	Sugar	73,217,147	2582-79-8	Phosphoric acid	Other groups	73,57,383	85,197-28-0
Linoleic acid	Fatty acid	75,73,67	56,259-07-5	L-Pyroglutamic acid	Amino acid derivative	73,272,147	107,716-03-0
Oleic Acid	Fatty acid	117,73,339	21,556-26-3	L-Serine	Amino acid	73,362,390	107,715-93-5
<i>trans</i> -9-Octadecenoic acid	Fatty acid	73,75,117	96,851-47-7	Myristic acid	Fatty acid	75,285,43	104,255-79-0
Stearic acid	Fatty acid	73,117,341	18,748-91-9	Malic acid	TCA	73,115,147	99,461-86-6
1-Monopalmitin	Monoglyceride	73,43,371	1188-74-5	L-Aspartic acid	Amino acid	73,75,302	107,715-96-8
Sucrose	Sugar	361,73,362	19,159-25-2	L-Glutamic acid	Amino acid	73,432,75	107,715-97-9
1-Monooleoylglycerol	Monoglyceride	73,397,129	54,284-47-8	L-Asparagine	Amino acid	73,417,147	96,381-41-8
Stigmasterol	Phytosterol	83,129,255	14,030-29-6	Citric acid	TCA	73,147,459	99,477-48-2
$\beta$ -Sitosterol	Phytosterol	129,357,121	2625-46-9				

GABA  $\gamma$ -Aminobutyric acid, TCA tricarboxylic acid cycle metabolites

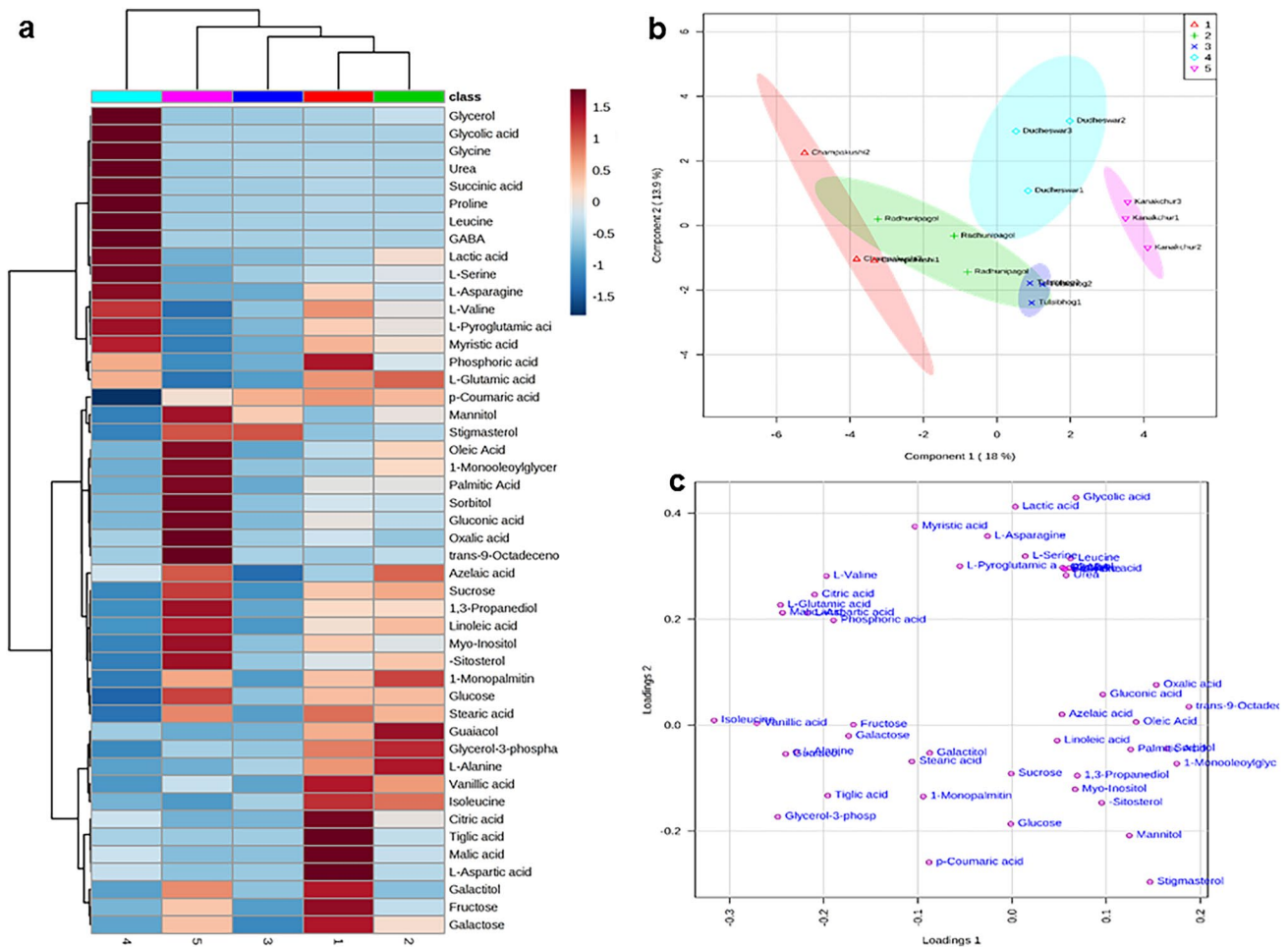
<sup>a</sup>PubChem CID

[10]. List of identified metabolites along with their mean relative abundance values and retention index are given in Table S4 in supplementary material. Relative abundances of identified metabolites vary significantly among the studied varieties and are represented in the form of a heat map (Fig. 2a). Grains of Champakushi are rich in organic acids, monosaccharides and sugar alcohols, while grains of Dudheswar are rich in amino acids content. Grains of Kanakchur are rich in fatty acids and their derivatives, sugar alcohols, organic acids, disaccharides and even some phytosterols. Unique aroma of Kanakchur could be a due to the presence of linoleic and oleic acids in higher amounts. These unsaturated fatty acids are produced during hydrolysis of lipids and their breakdown products constitute major volatile compounds in milled rice [6]. Metabolites identified in this study are vital for growth and development of plants and their

adaptation to different environmental stresses. Besides, most of the metabolites detected by GC–MS analysis have known nutritional benefits, e.g.  $\gamma$ -aminobutyric acid (GABA) and azelaic acid [30].

### Compositional variation of mineral elements in rice grains on the basis of their size and shape

Elemental analysis of the rice grains was also performed. Preparation of samples digested with nitric acid and perchloric acid, followed by ICP-MS analysis resulted in identification and quantification of a total of 35 elements with varying quantities in different varieties. Essential minerals such as Ca, Mg, phosphorous (P), sodium (Na), potassium (K) and trace elements such as Cu, Zn, cobalt (Co), iron (Fe), sulphur (S), manganese (Mn) are identified. Significant differences



**Fig. 2** Variability in the composition of metabolites in selected rice varieties. 1—Champakushi, 2—Radhunipagol, 3—Tulsibhog, 4—Dudheswar, 5—Kanakchur. **a** Heat map showing the variations of relative abundances of identified metabolites. In the heat map, lowest value is indicated by dark blue, the highest value by bright red and values in the middle are indicated by light grey, with a corresponding gradient between extremes. Corresponding PLS-DA showing a clear

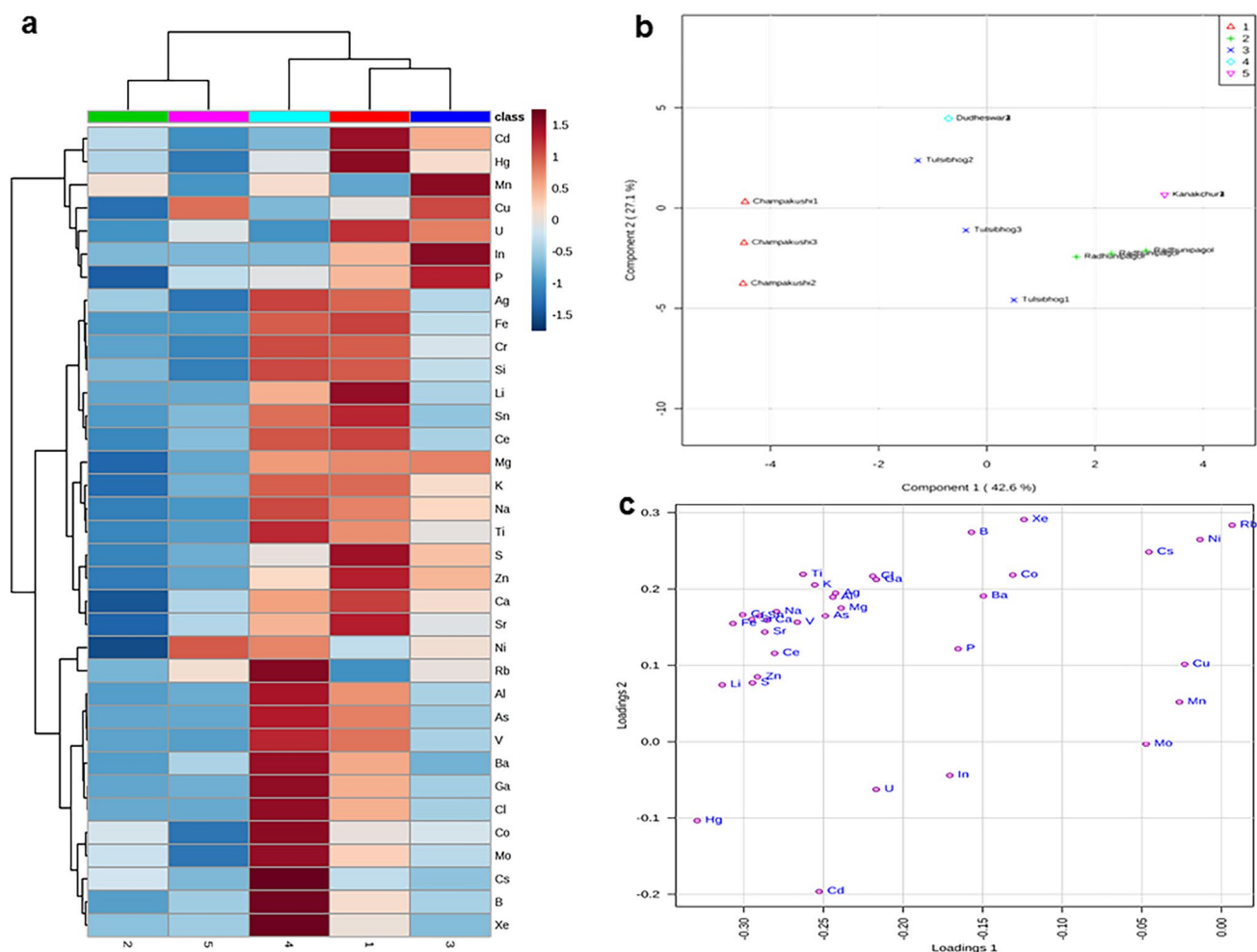
separation of varieties based on grain size and shape. **b** Scores plot generated from GC-MS based metabolite profile of the above-mentioned varieties as collectively shown in PC1 and PC2. Different colored symbols indicate different varieties. **c** Loadings plot of PC1 versus PC2 generated from accumulated metabolites in the five selected varieties. *GABA*  $\gamma$ -amino butyric acid (Color figure online)

in the quantities of some of the macro- and micronutrient elements were observed between all the five selected varieties. Quantitative data is given in Table S5 as supplementary material. Quantitative variation of elemental concentration in the five varieties is represented as heat map in Fig. 3a. Based on the quantitative variation of elemental content in the grains, the studied varieties were clearly divided into two groups. First group constitute varieties Kanakchur and Radhunipagol. This group is characterized by comparatively low mineral content. However, Kanakchur contains relatively high amounts of Cu and Nickel (Ni). Second group is characterized by high mineral content in their grains. All the three varieties in this group viz. Champakushi, Tulsibhog and Dudheswar showed relatively similar quantities of essential elements such as Na, K, Ca and Mg. However,

content of P varied significantly among these three varieties. Highest level of P was found in Tulsibhog amongst the studied varieties. Among the trace elements, Mn and Cu were present in higher amounts in the grains of Tulsibhog. Higher quantities of S and Zn were shown to be present in Champakushi. However, grains of both Champakushi and Dudheswar showed relatively high Fe content.

### Partial least squares-discriminant analysis (PLS-DA)

Impact of germplasm, geographical origin, grain type and color, aromaticity etc. on variability of grain composition in terms of metabolite and mineral contents was well demonstrated by multivariate analysis techniques [11, 14, 31, 32]. PLS-DA is a supervised multivariate analysis technique that



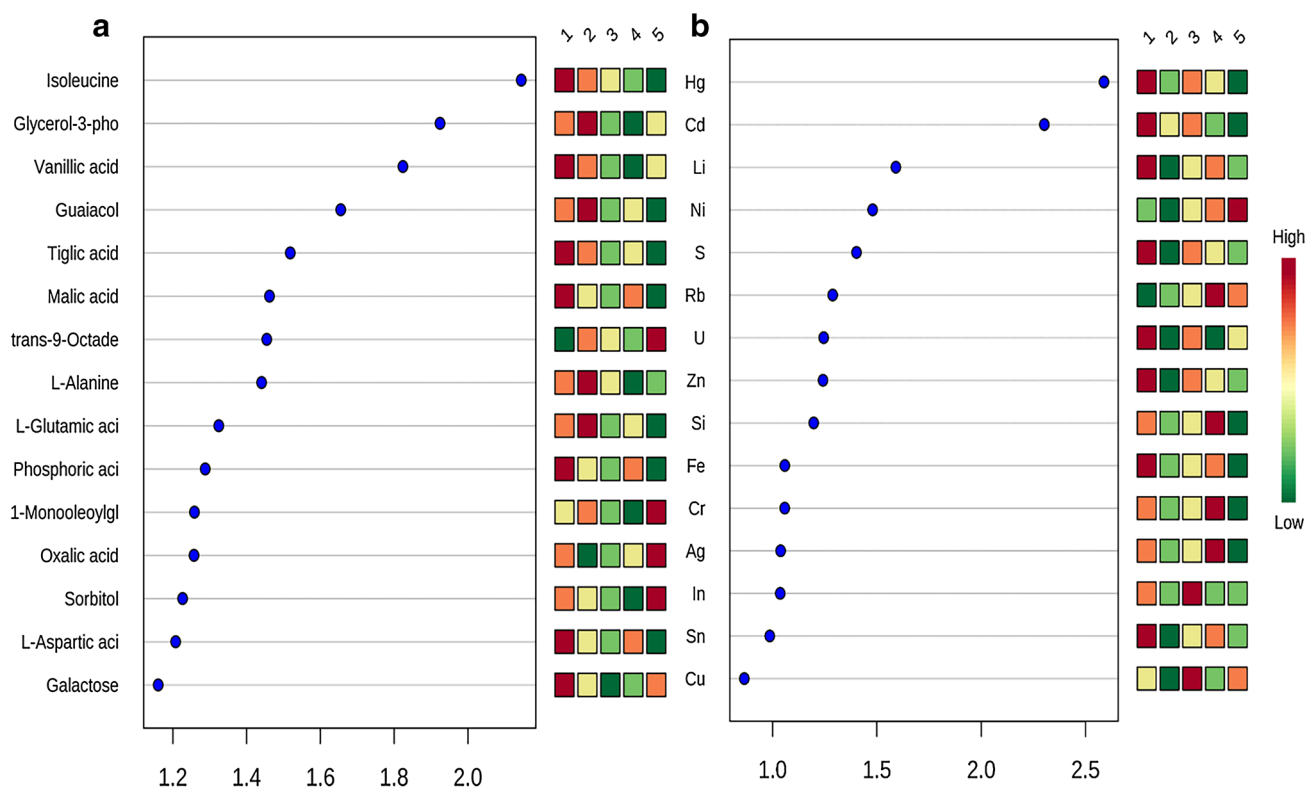
**Fig. 3** Variability in the composition of mineral elements in selected rice varieties. 1—Champakushi, 2—Radhunipagol, 3—Tulsibhog, 4—Dudheswar, 5—Kanakchur. **a** Heat map showing the variations of absolute quantities of identified elements. In the heat map, lowest value is indicated by dark blue, the highest value by bright red and values in the middle are indicated by light grey, with a corresponding gradient between extremes. Corresponding PLS-DA showing a clear

was performed in order to identify whether metabolite and elemental profiling can differentiate varieties with different grain size and shape and also to identify the variables responsible for the distinction between varieties. PLS-DA provides a better interpretation of large and complex datasets through scores plot that are easily interpretable, has lesser dimensions and clearly resolve sample groups [14]. PLS-DA for metabolite and elemental profiles of all five studied varieties was carried out separately. As anticipated, the selected varieties were distinctly separated from each other in both the cases. This indicates that, unique metabolite and elemental profiles exists for morphologically different traditional rice varieties. Scores plot for metabolite profile is shown in Fig. 2b. Together PC1 and PC2 explained 31.9% of the variation among all studied varieties. From the loadings

separation of varieties based on grain size and shape. **b** Scores plot generated from ICP-MS based elemental profile of the above-mentioned varieties explains 69.7% of the variation between the selected varieties as collectively shown in PC1 and PC2. **c** Loadings plot of PC1 versus PC2 generated from accumulated elements in all the selected varieties (Color figure online)

plot (Fig. 2c) and VIP scores (Fig. 4a) the most discriminant metabolites responsible for the variation among varieties could be identified. PC1 explaining 18% of the total variation, primarily separates varieties based on grain size and shape. Champakushi (Long, medium grain), Radhunipagol (medium size and medium-shaped grain), Kanakchur (short, bold grain) are well separated from Dudheswar (long and slender grain) and Tulsibhog (short grain). Loadings along PC1 were caused by organic acids (such as, tiglic acid, azelaic acid, gluconic acid and oxalic acid), fatty acids (such as oleic acid, linoleic acid, and palmitic acid) and amino acids (such as, alanine and isoleucine). Phenolic acids such as, vanillic acid and *p*-coumaric acid, also contribute to the separation along PC1. Vanillic acid was reported to be as one of the major phenolic compounds enabling discrimination





**Fig. 4** VIP (variable importance in projection) score-plots derived from the PLS-DA analysis showing the variables with high discriminating power. Coloured boxes indicate the relative concentrations of

corresponding variable in each rice variety. **a** Metabolites with high VIP scores. **b** Mineral elements with high VIP scores (Color figure online)

of Korean rice based on geographical origin [33]. Dittgen et al. [34] reported that *p*-Coumaric acid is one of the phenolic compounds that could discriminate a black rice variety growing in different locations. PC2 which explains 13.9% of the variation could separate long and slender Dudheswar, from short grained Kanakchur and Tulsibhog. Loadings along PC2 were caused by organic acids (such as glycolic acid, lactic acid), fatty acids (myristic acid), amino acids such as serine, leucine, and amino acid derivative L-pyroglytamic acid. PLS-DA for elemental composition resulted in PC1 and PC2 that could explain 69.7% of variation (Fig. 3b). PC1 explains 42.6% of the variation among varieties. Champakushi, Radhunipagol and Kanakchur are separated from Dudheswar and Tulsibhog along PC1. This separation of varieties based on elemental profile is similar to the separation of varieties achieved by metabolite profile. Interestingly, non-fragrant variety Champakushi can be clearly differentiated from other fragrant varieties along PC1 in PLS-DA analyses for both metabolite profiling and elemental profiling data. From the loadings plot (Fig. 3c) and VIP scores (Fig. 4b), elements causing separation between the varieties could be identified as Li, Ni, S, Zn, Cu, Si, Ag and Fe. Previous studies identified Cu, Ag, Zn, and Cr [14] and K, Ca, Zn, Mg, Mn, Sr, Ni, and Cd [31] as critical for discrimination of

geographical origin of rice. It was clear from the heat map that grains of Dudheswar accumulated higher quantities of minerals in their grains, followed by Champakushi and Tulsibhog. However, it is interesting to note that elements such as Cd, Zn, Fe, mercury (Hg), lithium (Li) and strontium (Sr) that are contributing to the separation of studied varieties as indicated by their high VIP scores are present in relatively higher quantities in Champakushi than in Dudheswar. This study encompassing comparative metabolite and elemental profiling revealed how the selected varieties are distinct from each other. Grain morphology and characteristic biochemical signatures of all five selected varieties are specified in Table 2.

## Conclusion

In the present study, we were able to categorize all the 12 traditional rice varieties into different clusters based on grain quality evaluation. Traits such as grain dimensions, TGW and AC contributed to the discrimination among studied varieties. Five major clusters are obtained and varieties from each cluster differ in their grain size and shape. Out of five clusters, one variety from each cluster was selected

**Table 2** Characteristic features of five selected traditional rice varieties

Variety	Size	Shape	Abundant metabolites	Abundant minerals
Champakushi	Long	Medium	High levels of monosaccharides, High levels of organic acids and phenolic acids, high levels of few amino acids, low levels of fatty acids	Cd, U, Li, In, P, Sn, Ce, Mg, S, Zn, Ca, Sr
Radhunipagol	Medium	Medium	High levels of phenolic acids, high levels of amino acids, moderate levels of monosaccharides and fatty acids	Cd, Mn, Co, Mo, Cs, Ag
Tulsibhog	Short	Medium	High levels of phenolic acids, moderate levels of sugar alcohols and monoglycerides, low levels of other metabolites	Cd, Mn, Cu, U, In, P, Mg, K, S, Zn, Ca, Ni
Dudheswar	Long	Slender	High levels of amino acids and organic acids, low levels of fatty acids, low levels of mono- and disaccharides	Rb, Al, Va, Ba, Ga, Co, Mo, Cs, B, Ti, Na, K, Mg, Ag, Fe, Si
Kanakchur	Short	Bold	High levels of fatty acids and sterols, moderate levels of amino acids and monosaccharides	Cu, Ni, U, Rb, Ba, Ca, Sr, P

as a representative variety and a combination of platforms for metabolite and elemental profiling were used. Around 47 metabolites and 35 minerals could be identified in all five varieties. These five varieties, which have distinct differences in grain size and shape showed a unique metabolite and elemental profile as revealed by PLS-DA analysis. All the identified metabolites and elements are present in varying quantities, so the differentiation between biochemical compositions of grains from selected varieties is typically quantitative. Further, this study highlights the discriminating power of metabolite and elemental composition in rice depending on grain dimensions.

**Acknowledgements** This work was a part of a Mega Project on Sustainable Food Security (File No: 4-25/2013/TS-I) funded by Ministry of Human Resource Development, Government of India. JNRK was a recipient of a doctoral fellowship from the institute.

## References

- P.S. Saha, S. Paul, B. Ghosh, *Int. J. Biores. Stress Manag.* **3**, 368 (2012)
- B. Bowonder, *Technol. Forecast. Soc. Change.* **15**, 297 (1979)
- R.C. Rabara, M.C. Ferrer, C.L. Diaz, M.C.V. Newingham, G.O. Romero, *Agronomy* **4**, 217 (2014)
- A.N.M.R. Bin Rahman, J. Zhang, *Rice* **11**, 10 (2018)
- R. Valarmathi, M. Raveendran, S. Robin, N. Senthil, *J. Plant Biochem. Biotechnol.* **24**, 305 (2015)
- J.C.T. Concepcion, S. Ouk, A. Riedela, M. Calingacion, D. Zhaod, M. Ouk, M.J. Garsonc, M.A. Fitzgerald, *Food Chem.* **240**, 1014 (2018)
- M.A. Fitzgerald, S.R. McCouch, R.D. Hall, *Trends Plant Sci.* **14**, 133 (2009)
- R. Mumm, J.A. Hageman, M.N. Calingacion, R.C.H. Vos, H.H. Jonker, A. Erban, J. Kopka, T.H. Hansen, K.H. Laursen, J.K. Schjoerring, J.L. Ward, M.H. Beale, S. Jongee, A. Rauf, F. Habibi, S.D. Indrasari, S. Sakhan, A. Ramli, M. Romero, R.F. Reinke, K. Ohtsubo, C. Boualaphanh, M.A. Fitzgerald, R.D. Hall, *Metabolomics* **12**, 38 (2016)
- J. Promchan, D. Günther, A. Siripinyanond, J. Shiowatana, J. Cereal Sci. **71**, 198 (2016)
- B. De, G. Nag, *C. R. Biol.* **337**, 283 (2014)
- T. Frank, B. Reichardt, Q. Shu, K.H. Engel, *J. Cereal Sci.* **55**, 112 (2012)
- J.K. Kim, S.Y. Park, S.H. Lim, Y. Yeo, H.S. Cho, S.H. Ha, *J. Cereal Sci.* **57**, 14 (2013)
- I.M. Chung, J.K. Kim, J.K. Lee, S.H. Kim, *J Cereal Sci.* **65**, 252 (2015)
- P. Cheajesadagul, C. Arnaudguilhem, J. Shiowatana, A. Siripinyanond, J. Szpunar, *Food Chem.* **141**, 3504 (2013)
- C. Kukulamude, S. Kongsri, *Food Control* **91**, 357 (2018)
- Grain quality manual, IRRI Rice knowledge bank. [www.knowledgebank.irri.org](http://www.knowledgebank.irri.org)
- S.J. Bhonsle, S. Krishnan, *Indian. J. Agric. Sci.* **2**, 101 (2010)
- T. Samanta, J.N.R. Kotamreddy, B.C. Ghosh, A. Mitra, *Acta Physiol. Plant.* **39**, 11 (2017)
- A. Jacobs, C. Lunde, A. Bacic, M. Tester, U. Roessner, *Metabolomics* **3**, 307 (2007)
- A.O. Oko, B.E. Ubi, A.A. Efiuse, N. Dambaba, *Int. J. Agric. For.* **2**, 16 (2012)
- J. Xia, I. Sinelnikov, B. Han, D.S. Wishart, *Nucleic Acids Res.* **43**, 251 (2015)
- A.K. Sinha, P.K. Mishra, *Int. J. Curr. Res.* **5**, 2764 (2013)
- S.C. Roy, B.D. Sharma, *Physiol. Mol. Biol. Plants.* **20**, 209 (2014)
- S. Roy, R.S. Rathi, A.K. Mishra, B.P. Bhatt, D.C. Bhandari, *Plant Genet. Resour. C* **12**, 58 (2014)
- A.K. Pachauri, A.K. Sarawgi, S. Bhandarkar, G.C. Ojha, *J. Pharmacogn. Phytochem.* **6**, 75 (2017)
- J. Rehal, G.J. Kaur, A.K. Singh, *Int. J. Curr. Microbiol. Appl. Sci.* **6**, 1278 (2017)
- L. Xie, S. Tang, N. Chen, J. Luo, G. Jiao, G. Shao, X. Wei, P. Hu, *Cereal Chem.* **90**, 587 (2013)
- K.S. McKenzie, J.N. Rutger, *Crop Sci.* **23**, 306 (1983)
- T. Mahmood, M.A. Turner, F.L. Stoddard, *Starch* **59**, 357 (2007)
- M.N. Calingacion, C. Boualaphanh, V.D. Daygon, R. Anacleto, R.S. Hamilton, B. Biais, C. Deborde, M. Maucour, A. Moing, R. Mumm, R.C.H. Vos, A. Erban, J. Kopka, T.H. Hansen, *Metabolomics* **8**, 771 (2012)
- I.M. Chung, J.K. Kim, K.J. Lee, S.K. Park, J.H. Lee, N.Y. Son, Y.I. Jin, S.H. Kim, *Food Chem.* **240**, 840 (2018)
- C. Hu, J. Shi, S. Quan, B. Cui, S. Kleessen, Z. Nikoloski, T. Tohge, D. Alexander, L. Guo, H. Lin, J. Wang, X. Cui, J. Rao, Q. Luo, X. Zhao, A.R. Fernie, D. Zhang, *Sci. Rep.* **4**, 5067 (2014)
- S.Y. Park, J.K. Kim, S.Y. Lee, S.D. Oh, S.M. Lee, J.S. Jang, C.I. Yang, Y.J. Won, Y. Yeo, *Plant Omics* **7**, 430 (2014)

34. C.L. Dittgen, J.F. Hoffmann, F.C. Chaves, C.V. Rombaldi, J.M.C. Filho, N.L. Vanier, *Food Chem.* **288**, 297 (2019)

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.