Profiling and Association Mapping of Grain Metabolites in a Subset of the Core Collection of Chinese Rice Germplasm (*Oryza sativa* L.)

Qiaojun Lou,[†,§] Chenfei Ma,[‡,§] Weiwei Wen,[†] Jia Zhou,[‡] Liang Chen,[†] Fangjun Feng,[†] Xiaoyan Xu,[†] Xin Lu,[‡] Lijun Luo,[†] Hanwei Mei,*,[†] and Guowang Xu*,[‡]

[†]Shanghai Agrobiological Gene Center, Shanghai 201106, China

[‡]CAS Key Laboratory of Separation Science for Analytical Chemistry, Dalian Institute of Chemical Physics, Chinese Academy of Sciences, Dalian 116023, China

S Supporting Information

ABSTRACT: In this study, metabolic profiles of a set of 48 rice germplasms from the Chinese core collection were obtained by gas chromatography and time-of-flight mass spectrometry (GC-TOF-MS). Forty-one metabolites were identified and relatively quantified according to the internal standard (IS). Wide ranges of variations for all metabolites were observed among rice accessions. The maximum/minimum ratios varied from 4.73 to 211.36. The metabolites were categorized into seven groups based on their chemical characteristics. Clustering analysis and a correlation network showed that most of the metabolites had variations among rice accessions in the same direction. Using 218 molecular markers, association mapping was conducted to identify the chromosomal loci influencing the concentrations of identified metabolites. Twenty markers were identified associating with the concentrations of 29 metabolites [-lg(P) > 3]. Allelic effects were investigated in detail in two markers (RM315 and RM541) as examples.

KEYWORDS: Metabolites, *Oryza sativa* L., GC-MS, association mapping

■ INTRODUCTION

Rice is one of the world's most important food crops, providing the caloric needs for at least one-third of the world population. As a model cereal crop, rice has several huge collections of germplasms around the world.[1] The genetic diversity stored in this large number of accessions is undoubtedly a great treasure for human being. However, it is still a great challenge to characterize the genetic variations in the germplasm collections of rice and other major crops because of the huge number of accessions and the abundant characteristics to be evaluated. As a potential solution, core collections and minicore collections of germplasms have been developed for several crops in China.[2−4] A set of germplasms containing a few thousand or several hundred accessions, but bearing a high proportion of genetic diversity, can be applied to fine genotyping and phenotyping experiments.

Metabolites are the end products of cellular regulatory processes, and metabolomics could link genotypes and phenotypes together.[5] Cereal chemistry has played an important role in studies of rice cooking quality and the determination of nutrient composition. For example, the cooked rice of *japonica* varieties is usually softer and stickier than *indica* varieties due to a lower amylose content and higher gel consistency. The protein content is considered in routine tests, and the concentrations and compositions of amino acids are frequently reported.[6] However, the identification of the global metabolic components in rice grains has been seldom reported in large sets of germplasms. After high-throughput methodologies of analytical chemistry were developed, metabolite and element profiling in large populations became possible. Many analytical methods like gas chromatography—mass spectrometry (GC-MS), high-performance liquid chromatography mass spectrometry (HPLC-MS), nuclear magnetic resonance (NMR), and inductively coupled plasma atomic emission spectroscopy (ICP-AES) have been adopted for such investigations in crop species.[7−9] As a useful metabolic tool, GC-MS coupled with multivariate analysis was widely used in nontargeted tests of metabolites in different rice germplasms.[10,11] However, no study has reported an investigation of the metabolic diversity in a large germplasm set and linked metabolite contents with DNA markers by association mapping.

Association mapping, especially linkage-disequilibrium (LD) mapping, has been applied in a number of plant species.[4,12,13] Compared with traditional quantitative trait loci (QTL) linkage analysis, association mapping has the following advantages: It provides better genome coverage of marker polymorphisms than any biparental population, it has a higher mapping resolution[14] and the possibility to detect multiple allelic effects, and it does not require the development of segregating populations. Several research groups have successfully conducted whole-genome association analyses of multiple agronomic traits using microsatellite (SSR) markers,[15,16] which became a popular and indispensable method in agriculture genetic studies.

In this study, the metabolic profiles of 48 rice varieties from a subset of a core collection of Chinese rice germplasms were obtained by gas chromatography and time-of-flight mass spectrometry (GC-TOF-MS). Variations in their metabolite characteristics were investigated by assessing their respective ranges among the germplasms as well as their categorical distributions and correlations. Dominant correlations among nutrients are useful in crop breeding, especially in multiple nutrient components improvement at selection stage. Markers that associated with metabolite concentrations were detected by association analysis using a mixed linear model (MLM), and the allelic effects of two associated markers (RM315 and RM541) were presented as examples. Association mapping and allelic effect analysis are helpful for right parental germplasms screening and marker-assisted selection (MAS) in breeding for specific nutrients.

■ MATERIALS AND METHODS

Plant Materials. A total of 48 accessions of rice varieties (*Oryza sativa* L.) were selected from a core collection of Chinese rice germplasms.[2] Rice accessions in this study were landraces or modern varieties from 19 provinces or municipalities in China, including Anhui, Fujian, Guangdong, Guangxi, Guizhou, Hebei, Heilongjiang, Henan, Hunan, Jiangsu, Jiangxi, Liaoning, Ningxia, Shanghai, Shanxi, Sichuan, Tianjin, Xizang, and Yunnan (Table S1 in the Supporting Information). According to the information from the germplasm database, concerning morphological traits and surveying by molecular markers,[16] those 48 accessions were divided into two subgroups, including 25 *indica* varieties (I01—I25) and 23 *japonica* varieties (J01—J23).

Grain samples used in this experiment were collected from the plants grown in the summer season of 2007 in Qingpu District, Shanghai, China. The field management followed normal agricultural practices. Each genotype was transplanted in five rows, with 20 cm between plants and 30 cm between rows. For each variety, about 30 rice plants were planted concurrently, and the seeds of these plants were collected together. After harvest, the grains were stored in refrigeration house of Shanghai Agrobiological Gene Center at low temperature from 5 to15 °C and 45% relative humidity (RH) until they were sampled in December 2007.

Metabolite Profiling. *Chemicals.* All solvents used in chromatograph with HPLC grade, and a total of 14 chemical standards (see Table 1) for qualitative analysis were purchased from Sigma-Aldrich (Beijing, China). Ultrapure water used in this experiment was produced by Milli-Q filtration (Millipore, MA).

Sample Preparation. The sample pretreatment in this experiment was similar to a previously described method.[17] Brown rice grains were ground into powder and sieved by 60 mesh sieves. Then, the powder was frozen dried by lyophilizer (Labconco, United States) for further extraction. Three hundred milligrams of powder, 3 mL of methanol—water solution (4:1, V/V), and 100 μL of capric acid [0.30 mg/mL, internal standard (IS)] were added into a 10 mL glass tube. After vortex shaking for 1 min, the mixture was immersed for 30 min, sonicated for 60 min, and centrifuged at 12000g for 10 min to obtain the supernatant. Two milliliters of the supernatant solution was freeze dried for 12 h. Then, the dried extracts were mixed with 90 μL of *N,O*-bis (trimethylsilyl) trifluoroacetamide having 1% of trimethylchlorosilane and 80 μL of pyridine . At last, the mixture was silanized at 75 °C for 45 min to increase samples' volatility. In method validation, sample pretreatments were repeated by five times in the same conditions, and the reproducibility was acceptable. More details of this procedure can be found in ref 17. For each rice variety, two samples were repeatedly analyzed as technical replicates, and each sample was duplicated and analyzed by GC-MS. The average peak areas of each component in all repeated assay for each variety were utilized for statistic analysis.

Metabolites Identification. A LECO Pegasus 4D TOF-MS (LECO Corp., MI) equipped with an Agilent 6890N GC was used to identify and quantify the metabolites. The conditions of GC-MS for analysis were according to our previous report[17] with little modification. A peak table was obtained from the LECO ChromaTOFTM workstation (version 3.25). In accordance with a previously described approach,[18] the signal-to-noise ratio of 100 was considered as a suitable threshold; about 200 peaks fulfilled this criterion and were subjected to further assessment. Peak alignment was carried out with homemade software, following the same strategies described in our previous work.[19] The chromatogram having the most peaks among all samples was chosen as reference, and the match window of retention time was 6 s. According to the "80% rule",[20] peaks with zero areas in more than 20% of the samples were removed from the total peak table. Finally, 76 compounds remained for further investigation, and 41 metabolites were identified by comparing their mass spectra to the data in NIST library (National Institute of Standards and Technology, United States). Among the 41 metabolites, 21 metabolites were confirmed through standards identification. The relative quantification of metabolites was determined by the IS and represented by mg per 100 g of dried powder of brown rice grains.

Metabolites Statistic Analysis. The statistical software package of S-PLUS 6.1 for windows was used in this experiment. Pair-wise correlations were first estimated among metabolites. Agglomerative hierarchical clusterings were constructed for 41 metabolites. The Pajek program (v1.23) was suitable for large network analysis and was used to develop two networks based on the array of correlation coefficients among metabolites, according to a previously described method (Batagelj and Mrvar, http://pajek.imfm.si/doku.php?id=pajek, 21). One network showed the metabolites with very high correlation coefficients in the positive direction ($r \geq 0.70$), while the other showed some metabolites having weak negative correlations ($r \leq -0.10$).

Association Mapping. A total of 218 markers [SSR + insertion-deletion markers (Indel)] were used in population genotyping (Table S2 in the Supporting Information). The software SPAGeDi[22,23] was used to estimate the kinship coefficients (K values). According to the report of Wen et al.,[16] 52 unlinked or loosely linked marker loci (i.e., four on chromosome 7 together with 48 on other chromosomes; mostly with physical distances larger than 1 Mb) were used to analyze the population structure. Associations between 41 metabolites and 218 markers were calculated using a MLM function based on the (Q + K) method in TASSEL2.0.[24] Significant marker-trait associations were indicated by $-\log(P) \geq 3$. Pair-wise mean comparisons were conducted by S-Plus with Window V6.1 to test the differences among the accession groups defined by alleles of associated markers, based on the Fisher LSD method.

■ RESULTS

Variation in Rice Grain Metabolites. Among 41 metabolites detected in this study, the mean concentrations had a range from 1.11 to 60.95 mg/100 g (Table 1). Of these 41 metabolites, 23 metabolites had nonzero values in all rice accessions, while 18 metabolites had zero values in some accessions. Excluding the zero values, there were still wide variations of metabolite concentrations among rice accessions. The ratios of maximum to minimum values varied from 4.73 in ethanedioic acid (M01) to 211.36 in raffinose (M38). The 41 metabolites were briefly categorized into seven groups based on their chemical characteristics, including 15 amino acids, 8 sugars, 6 organic acids, 6 fatty acids, 2 sterols, 2 polyols, and 2 others.

Table 1. Summary of the Variations of 41 Metabolites Detected in Brown Rice of 48 germplasm Accessions

code	names	categories[a]	N[b]	mean ± SD[c]	max	min[d]	max/min ratio	corr. pairs[e]
M01	ethanedioic acid	OA	44	2.24 ± 1.05	4.69	0.99	4.73	12
M02	carbodiimide	AA	48	4.17 ± 1.98	11.20	1.33	8.42	33
M03	1,2-ethanediol	polyol	45	1.11 ± 0.49	2.74	0.40	6.85	15
M04	lactate	OA	46	1.85 ± 1.44	7.38	0.31	24.17	30
M05[f]	L-alanine	AA	48	4.99 ± 3.00	20.84	1.45	14.40	36
M06[f]	glycine	AA	46	1.13 ± 0.82	3.75	0.24	15.50	16
M07[f]	L-norvaline	AA	48	2.60 ± 1.57	9.81	0.85	11.48	28
M08[f]	L-norleucine	AA	46	1.73 ± 1.04	5.09	0.50	10.11	17
M09[f]	phosphoric acid	others	47	7.97 ± 5.38	30.91	0.76	40.66	4
M10	glycerol	polyol	47	19.6 ± 10.33	48.39	8.31	5.82	28
M11	nicotinic acid	OA	47	1.46 ± 0.78	3.83	0.46	8.37	30
M12[f]	L-proline	AA	48	5.31 ± 3.93	21.76	0.91	24.04	33
M13[f]	serine	AA	48	5.30 ± 2.46	13.60	1.89	7.21	33
M14[f]	L-threonine	AA	48	2.44 ± 1.27	5.79	0.82	7.04	28
M15	mercaptosuccinic acid	OA	37	1.84 ± 1.16	6.51	0.49	13.36	33
M16	pyroglutamic acid	AA	48	3.41 ± 1.85	12.54	1.50	8.35	34
M17[f]	L-aspartic acid	AA	48	9.00 ± 3.56	20.98	4.02	5.22	31
M18[f]	4-aminobutanoic acid	AA	45	2.37 ± 1.58	7.52	0.59	12.76	30
M19	citrulline	AA	48	2.73 ± 1.41	6.42	0.41	15.80	33
M20[f]	glutamic acid	AA	48	11.98 ± 5.84	41.63	4.98	8.35	31
M21[f]	L-asparagine	AA	46	12.48 ± 6.90	42.17	4.73	8.92	17
M22	xylitol	sugar	48	5.89 ± 4.56	29.66	1.37	21.67	22
M23	glycerol-3-phosphate	others	48	14.06 ± 6.93	37.64	6.05	6.22	33
M24	gluconic acid lactone	sugar	48	10.56 ± 5.19	33.39	3.28	10.17	33
M25[f]	citric acid	OA	48	9.04 ± 5.15	29.33	3.09	9.48	31
M26	tetradecanoic acid	FA	48	7.58 ± 4.12	18.79	2.56	7.35	30
M27[f]	D-glucose	sugar	48	27.61 ± 20.42	107.01	8.12	13.18	28
M28	L-tyrosine	AA	44	1.49 ± 0.95	4.34	0.32	13.52	3
M29[f]	mannitol	sugar	48	5.12 ± 5.50	31.00	0.58	53.66	13
M30[f]	glucitol	sugar	48	2.37 ± 2.53	11.18	0.30	37.81	30
M31[f]	hexadecanoic acid	FA	48	48.72 ± 19.25	109.80	22.61	4.86	32
M32[f]	linoleic acid	FA	48	60.95 ± 27.15	159.03	12.23	13.00	18
M33[f]	oleic acid	FA	47	40.98 ± 18.91	108.63	5.77	18.84	17
M34	octadecanoic acid	FA	48	14.89 ± 5.98	33.42	6.99	4.78	27
M35	ethyl tartrate	OA	48	7.08 ± 3.56	24.84	1.76	14.09	31
M36	galacturonic acid	sugar	35	1.92 ± 1.16	5.17	0.38	13.44	13
M37	monopalmitin	FA	34	5.41 ± 2.37	10.85	1.92	5.65	28
M38[f]	raffinose	sugar	48	31.24 ± 30.52	123.46	0.58	211.36	1
M39[f]	trehalose	sugar	47	27.15 ± 12.8	72.92	4.59	15.87	21
M40	stigmasterol	sterol	41	1.92 ± 1.06	4.96	0.29	17.25	7
M41	cholesterol	sterol	46	5.47 ± 2.84	16.14	0.43	37.61	24

[a] OA, organic acid; AA, amino acid; and FA, fatty acid. [b] The number of accessions for whose metabolites observations was above the assay limits. [c] Metabolite concentrations were presented as relative quantification data (mg/100 g). [d] The minimum values among the observations besides the zero values. [e] The number of significant positive correlation pairs ($P \leq 0.01$) between this metabolite and each of other 40 metabolites. [f] Those metabolites validated by standard chemicals.

Clustering and Correlations among Metabolites. On the basis of the variations of metabolites among 48 rice accessions, hierarchical clustering was constructed (Figure 1) to expatiate the similarity between the metabolites in grains. It was observed that a large proportion of the metabolites were clustered into three branches with small dissimilarities: The first group had seven amino acids, four organic acids, two sugars, one sterol, and one polyol; the second group had five amino acids, one fatty acid, and one sterol; a small separate branch had two sugars; and the third group had three amino acids, two organic acids, two fatty acids, one sugar, and two others. Taking 34 of the above-mentioned metabolites together as a dominant branch, only seven remaining metabolites formed three more branches. In the order of ascending dissimilarity from the major group, there were one group including glycerol, trehalose, and D-glucose, one group with raffinose alone, and another group including hexadeconoic acid, oleic acid, and linoleic acid. Although dry grains were used as samples in this experiment, which are "inactive" organism in
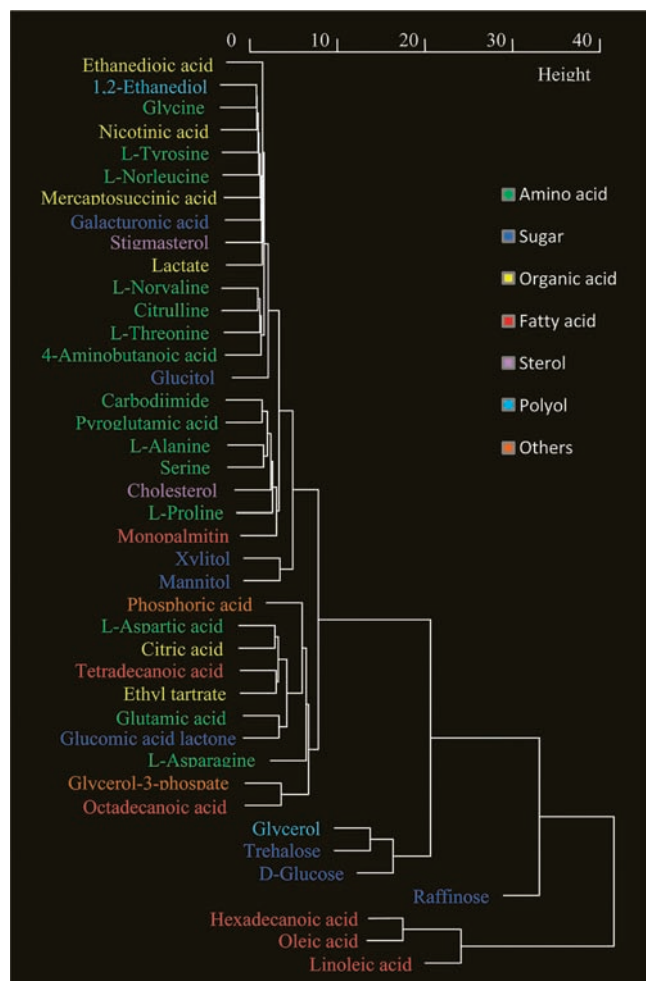
Figure 1. Hierarchical clustering of 41 metabolites based on their content in brown grains of 48 rice accessions. The distance between the clusters represents the similarities between the items that they contain. The color scheme of the metabolites follows the symbols on the top right corner, indicating their different categories; for example, serine with a green color belongs to amino acid.

Figure 2. Correlations among metabolites detected in brown rice samples of 48 accessions. (A) Positive correlation networks among 23 metabolites with $r \geq 0.70$. (B) Negative correlation networks among 13 metabolites with $r \leq -0.10$. Circles filled with different colors represent different categories of metabolites, and the color scheme follows that of Figure 1. Sizes of circles were determined by the number of correlation pairs.

the plant, the clustering structures still implied some variations among the metabolites parallel to metabolic pathways, for example, the variations among hexadeconoic acid, oleic acid, and linoleic acid. These three fatty acids connect closely in metabolic reaction in living organisms.

Similar to the results obtained from the clustering analysis, positive correlations were widely observed among most metabolites. Among 40 pairwise relationships from one metabolite to all other metabolites, 19 metabolites had 30 or more significant positive correlative pairs; 9 metabolites had 20–30 positive correlative pairs; 9 metabolites had 10–20 pairs; and only 4 metabolites had less than 10 significant positive correlative pairs (Table 1).

Positive correlations with the highest coefficients ($r \geq 0.70$) were presented as a network frame in Figure 2A. It was found that 10 amino acids had close correlations with each other and with metabolites in other categories, especially serine (M13), L-alanine (M05), glutamic acid (M20), L-proline (M12), etc. Five other metabolites, glycerol-3-phosphate (M23), gluconic acid lactone (M24), ethyl tartrate (M35), citric acid (M25), and tetradecanoic acid (M26), showed high correlations with each
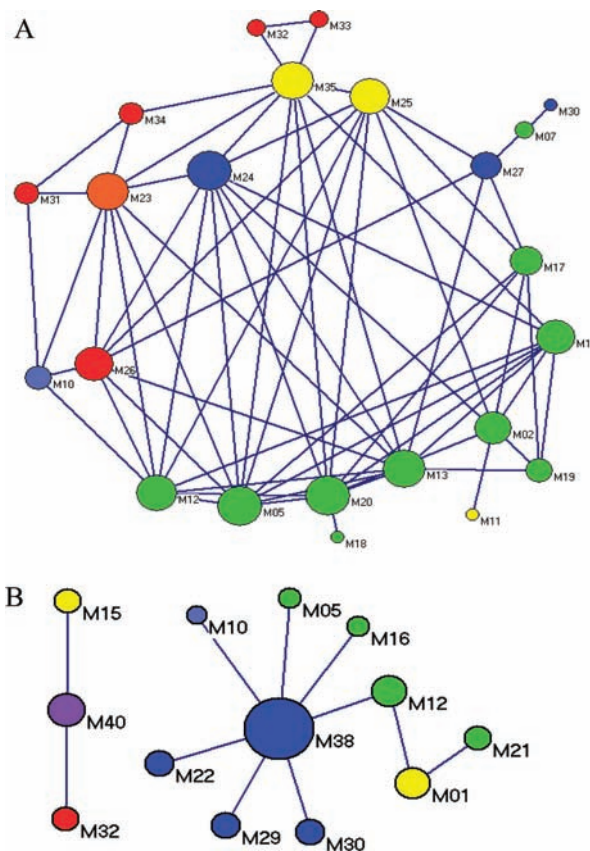
other and with many amino acids. The high correlations among glycerol-3-phosphate (M23), tetradecanoic acid (M26), glycerol (M10), hexadecanoic acid (M31), octadecanoic acid (M34), ethyl tartrate (M35), and gluconic acid lactone (M24) were probably coincident with their relations in sugar and fatty acid pathways.

Two groups of metabolites had pairwise negative correlations ($r \leq -0.10$, Figure 2B). First, stigmasterol (M40) had negative correlations with linoleic acid (M32) and mercaptosuccinic acid (M15). Second, it was observed that raffinose (M38) had negative correlations with seven other metabolites including four sugar alcohols. The coefficients varied between -0.11 and -0.22 for the above-mentioned negative correlations. In general, the negative relationships among rice grain metabolites seemed much weaker than the positive ones. However, the negative correlations involving raffinose were interesting as raffinose was regarded as one of the antinutritional oligosaccharides, so it would be useful in crop's nutrition modification.

Association between Metabolite Concentration and Molecular Markers. On the basis of MLM analysis, significant associations were declared at $-\log(P) \geq 3$ as shown in Figure 3. The number of associated molecular markers for each metabolite varied from one to three in most cases, but there were four or five markers for several metabolites such as L-alanine (M05), D-glucose (M27), and mannitol (M29). On the other hand, one
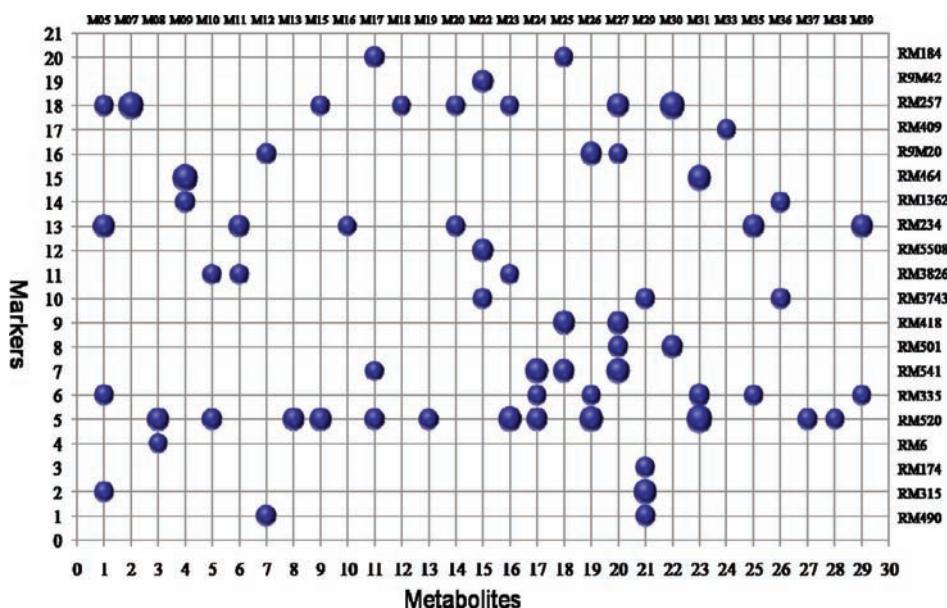
Figure 3. Marker loci associated with the concentrations of metabolites based on the mixed linear model [MLM, $-\log(P) \geq 3$]. Totally, 29 metabolites (x-axis) and 20 markers (y-axis) were indentified having significant associations between each other. The markers from RM490 to RM184 were ordered by their chromosome locations from the top of chr. 1 to bottom of chr. 12, and the metabolites (x-axis) were ordered by their retention time. Sizes of the blue bubbles represent the values of $-\log(P)$ approximately.

marker could be involved in associations with only one metabolite, or with several metabolites, for example, RM520 on chromosome 3 associated with 12 metabolites and RM257 on chromosome 9 associated with eight metabolites.

Taking two markers as examples, the effects of multiple alleles were interpreted in detail (Table 2). RM315 was associated with the concentrations of L-alanine (M05) with $-\log(P) = 3.36$. Eight varieties (alleles 0 and 4) had significantly higher means than those of other genotypes (alleles 1, 2, and 3). This marker was also associated with mannitol (M29) and had three groups of genotypes according to different allelic effects at high test significance ($P \leq 0.01$); that is, three varieties with allele 0 had a mean value of 2.36 mg/100 g, five varieties with allele 4 had a mean value of 1.29 mg/100 g, and 40 genotypes with alleles 1, 2, and 3 had means near 0.5 mg/100 g.

RM541 was associated with the concentrations of four metabolites with $-\log(P)$ values varied from 3.01 to 4.79. Six groups with alleles 0–5 had 5, 8, 3, 4, 25, and 3 varieties, respectively. For L-aspartic acid (M17), varieties with allele 5 had the highest concentrations that were highly different from the group means with alleles 0, 4, 2, and 1. The mean value of the group with allele 3 was also significantly higher than those of groups with alleles 4 and 1. The different alleles' means of gluconic acid lactone (M24) showed similar patterns to L-aspartic acid, but four varieties within the group with allele 3 had the highest mean value. The means of groups with different RM541 alleles had almost the same trends in citric acid (M25) and D-glucose (M27) as in L-aspartic acid. In other words, rice germplasm accessions having alleles 3 and 5 of RM541 had higher concentrations of those four metabolites in grains than the other accessions (Table 2).

■ DISCUSSION

Metabolites Related to Grain Quality and Nutrients. Dried cereal grains are organisms at relatively inactive stage in the life cycle of the plant. Grains can be used as seed stocks and serve as storage of nutrients when used as food or feed. Variations in the chemical compositions of grains are important to grain consumers and to both seed and food/feed industries for the purposes of quality control.

Taking nutrient values as an objective of rice breeding programs, the key question is which metabolites should be increased or decreased in concentration and to what extent? Programs like Harvest-Plus have been making great efforts to develop crop germplasms with increased nutritional content and/or availability of vitamin A, Fe, Zn, etc. (http://www.harvestplus.org). High lysine content is another popular target in cereal breeding programs.[25] However, the answers to above questions become complicated and uncertain when considering factors such as rice plant growth, grain yield, performance under storage, and consumption habits (e.g., favored taste, brown rice vs milled rice, etc.). For example, low phytic acid varieties (from mutants or transgenic breeding) can provide much more free phosphorus as compared to other varieties and can increase the absorption of Ca, Fe, and Zn at the same time.[26] However, varieties with low phytic acid usually show difficult growth or development, lower grain yield, and rapid loss of seed germination ability.[27]

Highly positive correlations were found in dominant pairwise relationships among the metabolites in this study. This kind of information is valuable for breeders to select parental germplasms and to enable the integrated improvement of multiple nutrient factors (like several amino acids) at the same time (Figure 3). A single positive correlation and seven pairs of negative correlations ($r \leq -0.10$) between raffinose (M38) and other metabolites seemed to be encouraging for breeders, as reducing raffinose content would increase or at least would not decrease the concentrations of other metabolites such as alanine and proline.

Associated Molecular Markers Valuable for Nutrient Breeding Programs. To enable effective selection of multiple nutrients in rice grains, a highly efficient platform is necessary to

Table 2. Pair-Wise Multiple Comparisons of Average Metabolite Contents in Rice Germplasms Grouped by the Alleles of RM315 and RM541

markers	metabolites	−log(P)	alleles[a]	N[b]	means[c]	P < 0.05	P < 0.01
RM315	L-alanine (M05)	3.36	0	3	1.31	a	A
			4	5	1.01	a	A
			3	17	0.54	b	B
			2	17	0.54	b	B
			1	6	0.45	b	B
	mannitol (M29)	4.74	0	3	2.36	a	A
			4	5	1.29	b	B
			1	6	0.48	c	C
			3	17	0.45	c	C
			2	17	0.40	c	C
RM541	L-aspartic acid (M17)	3.01	5	3	1.88	a	A
			3	4	1.59	ab	AB
			0	5	1.09	bc	B
			4	25	1.02	c	B
			2	3	1.00	bc	B
			1	8	1.00	c	B
	gluconic acid lactone (M24)	4.78	3	4	2.56	a	A
			5	3	1.76	b	AB
			0	5	1.53	bc	B
			2	3	1.44	bc	B
			4	25	1.10	c	B
			1	8	1.06	bc	B
	citric acid (M25)	3.99	5	3	2.01	a	A
			3	4	1.92	a	A
			0	5	1.26	ab	AB
			4	25	1.02	b	B
			2	3	0.95	b	AB
			1	8	0.74	b	B
	D-glucose (M27)	4.79	5	3	8.27	a	A
			3	4	6.36	a	AB
			0	5	3.45	b	BC
			2	3	3.05	b	BC
			1	8	2.72	b	C
			4	25	2.69	b	C

[a] Different allele types of the marker, arranged in descending order of the means. [b] N represents the number of rice variety with a fixed type allele; for example, 3 at first line represents that there are three rice varieties in allele type 0 at the loci of RM315. [c] Means followed by different letters were significantly different by the LSD test at the level $P \leq 0.05$ (in lowercase) and $P \leq 0.01$ (in uppercase).

measure the concentrations of all metabolites or a group of targeted metabolites in a large quantity of breeding materials; however, these procedures are quite time-consuming and costly. As in the cases of many other complex traits, MAS can play an important role in breeding programs targeted for specific grain nutrients.

Like traditional mapping, association mapping is also an efficient way to find powerful molecular markers for crop improvement by MAS.[14,28] Several associations between metabolites and markers found in this experiment coincided with the data presented on Rice Annotation Project (RAP, http://www.dna.affrc.go.jp/database/). For example, RM257 was found associated with L-norvaline (amino acid, -lg(p) = 5.48) and glucitol (sugar, -lg(p) = 5.67) and 4-aminobutanoic acid (amino acid, -lg(p) = 3.00) in this experiment. Then through searching in the RAP database, three genes (Os09t0432600, Os09t0432600, and Os09t0433900) that may significantly related to these three

metabolites were found near the RM257 (±50Kb of RM257 on chromosome 9). Os09t0432600 was similar to an amino acid transporter, Os09t0432900 was supposed to be a glycosyl transferase, and Os09t0433900 was similar to the alanine aminotransferase (EC 2.6.1.2), which correlated with 4-aminobutanoic acid in the pathway of arginine and proline metabolism (KEEG, http://www.genome.jp/kegg/). So the association result identified in this experiment was powerful and could be used in crop improvement through MAS, even there were some shortage in this mapping population.

In this study, a lot of associated molecular markers were identified for 29 metabolites in the population of core rice germplasms. In each associated marker locus, the favorable allele could be found by mean comparison among different alleles. Then, the germplasm having adequate phenotypic performance and possessing the favored allele in the associated loci could be selected

9262

dx.doi.org/10.1021/jf201602g |J. Agric. Food Chem. 2011, 59, 9257–9264

as the donor parent to make crosses with modern varieties as the recurrent parent in marker-assisted backcross breeding procedures. Instead of frequent measurements of metabolites, materials could be selected for backcrossing or selfing in early generations by tracing the targeted alleles of associated markers.

We also found that several metabolites can be coinfluenced by one marker locus (Figure 3), and favored alleles can be shared by different metabolites (Table 2). In this case, the concurrent improvements for several nutrients (and/or antinutrients) could be expected to occur in the breeding population derived from a single cross with the parental line and marker allele with multiple contributions.

It should be noted that this experiment was only a preliminary study in this area. More detailed case analysis of candidate alleles and donor parents for marker-assisted selection must be conducted before the above-mentioned strategy can be actually effective in breeding practice.

■ ASSOCIATED CONTENT

ⓢ Supporting Information. Rice accessions used in this study (Table S1) and a list of 218 molecular markers used in association mapping (Table S2). This material is available free of charge via the Internet at http://pubs.acs.org.

■ AUTHOR INFORMATION

Corresponding Author
*Tel: +86 021 62202938 or 0411-84379530. Fax: +86 021 62204010 or 0411-84379559. E-mail: hmei@sagc.org.cn (H.M.) or xugw@dicp.ac.cn (G.X.).

Author Contributions
§Q.L. and C.M. contributed equally to this research.

■ ACKNOWLEDGMENT

■ ABBREVIATIONS

GC-TOF-MS, gas chromatography and time-of-flight mass spectrometry; HPLC-MS, high-performance liquid chromatography mass spectrometry; ICP-AES, inductively coupled plasma atomic emission spectroscopy; indel, insertion—deletion markers; IS, internal standard; LD, linkage-disequilibrium mapping; MAS, marker-assisted selection; MLM, mixed linear model; NIST library, USA National Institute of Standards and Technology; NMR, nuclear magnetic resonance; QTL, quantitative trait loci; RH, relative humidity; SSR, microsatellite marker

■ REFERENCES

(1) Yan, W.; Rutger, J. N.; Bockelman, H. E.; Tai, T. USDA rice core collection and its genetic assessment. In *Proceedings of International Symposium on Genetic-Based Plant Germplasm Research*; Beijing, China, 2005; pp 25—28.

(2) Zhang, H; Zhang, D; Wang, M; Sun, J; Qi, Y; Li, J; Wei, X; Han, L; Qiu, Z; Tang, S; Li, Z. A core collection and mini core collection of *Oryza sativa* L. in China. *Theor. Appl. Genet.* **2010**, *122*, 49–61.

(3) Hao, C. Y.; Dong, Y. C.; Wang, L. F.; Zhang, H. N.; Ge, H. M.; Jia, J. Z.; Zhang, X. Y. Genetic diversity and construction of core collection in Chinese wheat genetic resources. *Chin. Sci. Bull.* **2008**, *53*, 1518–1526.

(4) Wang, R. H.; Yu, Y. T.; Zhao, J. R.; Shi, Y. S.; Song, Y. C.; Wang, T. Y.; Li, Y. Population structure and linkage disequilibrium of a mini core set of maize inbred lines in China. *Theor. Appl. Genet.* **2008**, *117*, 1141–1153.

(5) Fiehn, O. Metabolomics-the link between genotypes and phenotypes. *Plant Mol. Biol.* **2002**, *48*, 155–171.

(6) Hagop, E. G.; Younis, S. A.; Shahatha, H. A. Proteins and amino acids of some local varieties of rice seeds (*Oryza sativa* L.). *Plant Foods Hum. Nutr.* **1990**, *40*, 309–315.

(7) Harrigan, G. G.; Stork, L. G.; Riogdan, S. G.; Reynold, T. L.; Ridley, W. P.; Masucci, J. D.; MacIsaac, S; Halls, S. C.; Orth, R; Smith, R. G.; Wen, L; Brown, W. E.; Welsch, M; Riley, R; McFarland, D; Pandravada, A; Glenn, K. C. Impact of genetics and environment on nutritional and metabolite components of maize grain. *J. Agric. Food Chem.* **2007**, *55*, 6177–6185.

(8) Mochida, K.; Furuta, T.; Ogihara, Y.; Shinozaki, K.; Kikuchi, J. In Metabolic Phenotyping of Genetically Diverged Species in Gramineae. Appels, R., Eastwood, R., Lagudah, E., Langridge, P., Mackay, M., McIntyre, L., Sharp, P., Eds.; *The 11th International Wheat Genetics Symposium Proceedings*. Sydney University Press: Sydney, 2008.

(9) Zeng, Y. W.; Zhang, H. L.; Wang, L. X.; Pu, X. Y.; Du, J; Yang, S. M.; Liu, J. F. Genotypic variation in element concentrations in brown rice from Yunnan landraces in China. *Environ. Geochem. Health* **2010**, *32*, 165–177.

(10) Kusano, M.; Fukushima, A.; Kobayashi, M.; Hayashi, N.; Jonsson, P.; Moritz, T.; Ebana, K.; Saito, K. Application of a metabolomic method combining one-dimensional and two-dimensional gas chromatography-time-of-flight/mass spectrometry to metabolic phenotyping of natural variants in rice. *J. Chromatogr., B: Anal. Technol. Biomed. Life Sci.* **2007**, *855*, 71–79.

(11) Shu, X. L.; Frank, T; Shu, Q. Y.; Engel, K. H. Metabolite profiling of germinating rice seeds. *J. Agric. Food Chem.* **2008**, *56*, 11612–11620.

(12) Kraakman, A. T.; Niks, R. E.; Van den Berg, P. M.; Stam, P; Van Eeuwijk, F. A. Linkage disequilibrium mapping of yield and yield stability in modern spring barley cultivars. *Genetics* **2004**, *168*, 435–446.

(13) Aranzana, M. J.; Kim, S.; Zhao, K.; Bakker, E.; Horton, M.; Jakob, K.; Lister, C.; Molitor, J.; Shindo, C.; Tang, C.; Toomajian, C.; Traw, B.; Zheng, H.; Bergelson, J.; Dean, C.; Marjoram, P.; Nordborg, M. Genome-wide association mapping in Arabidopsis identifies previously known flowering time and pathogen resistance genes. *PLoS Genet.* **2005**, *1*, 531–539.

(14) Flint-Garcia, S. A.; Thuillet, A. C.; Yu, J.; Pressoir, G.; Romero, S. M.; Mitchell, S. E.; Doebley, J.; Kresovich, S.; Goodman, M. M.; Buckler, E. S. Maize association population: A high-resolution platform for quantitative trait locus dissection. *Plant J.* **2005**, *44*, 1054–1064.

(15) Zhang, N.; Xu, Y.; Akash, M.; McCouch, S.; Oard, J. H. Identification of candidate markers associated with agronomic traits in rice using discriminant analysis. *Theor. Appl. Genet.* **2005**, *110*, 721–729.

(16) Wen, W. W.; Mei, H. W.; Feng, F. J.; Yu, S. B.; Huang, Z. C.; Wu, J. H.; Chen, L; Xu, X. Y.; Luo, L. J. Population structure and association mapping on chromosome 7 using a diverse panel of Chinese germplasm in rice. *Theor. Appl. Genet.* **2009**, *119*, 459–470.

(17) Zhou, J.; Ma, C. F.; Xu, H. L.; Yuan, K. L.; Lu, X.; Zhu, Z.; Wu, Y. N.; Xu, G. W. Metabolic profiling of transgenic rice with *cryIAc* and *sck* genes: An evaluation of unintended effects at metabolic level by using GC-FID and GC-MS. *J. Chromatogr. B* **2009**, *877*, 725–732.

(18) Begley, P.; McIntyre, S. F.; Dunn, W. B.; Broadhurst, D. I; Halsall, A.; Tseng, A.; Knowles, J.; HUSERMET Consortium; Goodacre, R.; Kell, D. B. Development and performance of a gas

chromatography-time-of-flight mass spectrometry analysis for large-scale non targeted metabolomic studies of human serum. *Anal. Chem.* **2009**, *81*, 7038–7046.

(19) Ma, C. F.; Wang, H. H.; Lu, X; Xu, G. W.; Liu, B. Y. Metabolic fingerprinting investigation of *Artemisia annua* L. in different stages of development by gas chromatography and gas chromatography-mass spectrometry. *J. Chromatogr. A* **2008**, *1186*, 412–419.

(20) Bijlsma, S.; Bobeldijk, L.; Verheij, E. R.; Ramaker, R.; Kochhar, S.; Macdonald, I. A.; Van Ommen, B.; Smilde, A. K. Large-scale human metabolomics studies: A strategy for data (pre-) processing and validation. *Anal. Chem.* **2006**, *78*, 567–574.

(21) Schauer, N.; Semel, Y.; Roessner, U.; Gur, A.; Balbo, I.; Carrari, F.; Pleban, T.; Perez-Melis, A.; Bruedigam, C.; Kopka, J.; Willmitzer, L.; Zamir, D.; Fernie, A. R. Comprehensive metabolic profiling and phenotyping of interspecific introgression lines for tomato improvement. *Nat. Biotechnol.* **2006**, *24*, 447–454.

(22) Hardy, O. J.; Vekemans, X. SPAGeDi: A versatile computer program to analyze spatial genetic structure at the individual or population levels. *Mol. Ecol. Notes* **2002**, *2*, 618–620.

(23) Loiselle, B. A.; Sork, V. L.; Nason, J.; Graham, C. Spatial genetic structure of a tropical understorey shrub: Psychotria officinalis (*Rubiaceae*). *Am. J. Bot* **1995**, *82*, 1420–1425.

(24) Yu, J.; Holland, J. B.; McMullen, M. D.; Buckler, E. S. Genetic design and statistical power of nested association mapping in maize. *Genetics* **2008**, *178*, 539–551.

(25) Shewry, P. R. Improving the protein content and composition of cereal grain. *J. Cereal Sci.* **2007**, *46*, 239–250.

(26) Raboy, V. Seeds for a better future: 'Low phytate' grains help to overcome malnutrition and reduce pollution. *Trends Plant Sci.* **2001**, *6*, 458–462.

(27) Bregitzer, P.; Raboy, V. Effects of Four Independent Low-Phytate Mutations on Barley Agronomic Performance. *Crop Sci.* **2006**, *46*, 1318–1322.

(28) Yan, J. B.; Kandianis, C. B.; Harjes, C. E.; Bai, L; Kim, E. H.; Yang, X. H.; Skinner, D. J.; Fu, Z. Y.; Mitchell, S.; Li, Q.; Fernandez, M. G. S.; Zaharieva, M.; Babu, R.; Fu, Y.; Palacios, N.; Li, J. S.; DellaPenna, D.; Brutnell, T.; Buckler, E. S.; Warburton, M. L.; Rocheford, T. Rare genetic variation at Zea mays *crtRB1* increases β-carotene in maize grain. *Nat. Genet.* **2010**, *42*, 322–327.