

The Jackson Laboratory

The Mouseion at the JAXlibrary

Faculty Research 2019

Faculty Research

12-2019

Comparative analysis of metabolome of rice seeds at three developmental stages using a recombinant inbred line population.

Kang Li

Dehong Wang

Liang Gong

Yuanyuan Lyu

Hao Guo

See next page for additional authors

Follow this and additional works at: <https://mouseion.jax.org/stfb2019>



Part of the [Life Sciences Commons](#), and the [Medicine and Health Sciences Commons](#)

Authors

Kang Li, Dehong Wang, Liang Gong, Yuanyuan Lyu, Hao Guo, Wei Chen, Cheng Jin, Xianqing Liu, Chuanying Fang, and Jie Luo

Comparative analysis of metabolome of rice seeds at three developmental stages using a recombinant inbred line population

Kang Li¹, Dehong Wang¹, Liang Gong², Yuanyuan Lyu¹, Hao Guo¹, Wei Chen³, Cheng Jin¹, Xianqing Liu⁴, Chuanying Fang^{4,*}  and Jie Luo^{1,4,*}

¹National Key Laboratory of Crop Genetic Improvement and National Center of Plant Research, Huazhong Agricultural University, Wuhan 430070, China,

²The Jackson Laboratory for Genomic Medicine, Farmington, CT, USA,

³College of Plant Science and Technology, Huazhong Agricultural University, Wuhan 430070, China, and

⁴Hainan Key Laboratory for Sustainable Utilization of Tropical Bioresource, College of Tropical Crops, Hainan University, Haikou 570288, China

Received 14 May 2019; revised 13 July 2019; accepted 23 July 2019; published online 29 July 2019.

*For correspondence (e-mails chuanyingfang@126.com; jie.luo@mail.hzau.edu.cn).

SUMMARY

Plants are considered an important food and nutrition source for humans. Despite advances in plant seed metabolomics, knowledge about the genetic and molecular bases of rice seed metabolomes at different developmental stages is still limited. Here, using Zhenshan 97 (ZS97) and Minghui 63 (MH63), we performed a widely targeted metabolic profiling in seeds during grain filling, mature seeds and germinating seeds. The diversity between MH63 and ZS97 was characterized in terms of the content of metabolites and the metabolic shifting across developmental stages. Taking advantage of the ultra-high-density genetic map of a population of 210 recombinant inbred lines (RILs) derived from a cross between ZS97 and MH63, we identified 4681 putative metabolic quantitative trait loci (mQTLs) in seeds across the three stages. Further analysis of the mQTLs for the codetected metabolites across the three stages revealed that the genetic regulation of metabolite accumulation was closely related to developmental stage. Using *in silico* analyses, we characterized 35 candidate genes responsible for 30 structurally identified or annotated compounds, among which *LOC_Os07g04970* and *LOC_Os06g03990* were identified to be responsible for feruloylserotonin and L-asparagine content variation across populations, respectively. Metabolite–agronomic trait association and colocation between mQTLs and phenotypic quantitative trait loci (pQTLs) revealed the complexity of the metabolite–agronomic trait relationship and the corresponding genetic basis.

Keywords: rice seeds, metabolome, seeds during grain filling, mature seeds, germinating seeds, metabolic quantitative trait loci.

INTRODUCTION

As readouts of the physiological or biochemical status of an organism, metabolites are essential for plant growth and plant–environment interactions, as well as for human health (Keurentjes, 2009; Saito and Matsuda, 2010; De Luca *et al.*, 2012; Wurtzel and Kutchan, 2016). Benefiting from the extreme diversity of metabolites, plants have become ideal models for dissecting the mechanism of metabolite biosynthesis and its regulation (Keurentjes *et al.*, 2006; Morohashi *et al.*, 2012; Luo, 2015; Fang *et al.*, 2019a,b; Fang and Luo, 2019). In plants, which are sessile in nature, the number of metabolites is estimated to be between 100 000 and 1 million (Dixon and Strack, 2003; Afendi

et al., 2012). Many metabolites display differential shifts during development. For instance, the contents of the majority of C-glycosylated and O-glycosylated flavonoids significantly increase in seedlings during the first 10 days after germination, which decrease slightly during later stages (Dong *et al.*, 2014). Furthermore, the accumulation of anthocyanins and most of the indole-derived glucosinolates in leaves largely increases continuously throughout leaf senescence (Watanabe *et al.*, 2013). Development-dependent accumulation is also observed concerning the content of primary metabolites (Mounet *et al.*, 2007; Hu *et al.*, 2016; Silva *et al.*, 2017). For instance, Hu *et al.* (2016)

found that more than half of the detected amino acids and their derivatives accumulated at a significantly decreased level in rice grains at 28 days after flowering (DAF) compared with 14 DAF. However, knowledge about the diversity of plant metabolites in the same tissue at different stages of development is still largely scarce.

The diversity of plant metabolism across species and within natural accessions of a single species has been well documented (Borevitz *et al.*, 2007; Biais *et al.*, 2010). As one of the most essential crop species, rice (*Oryza sativa* L.) not only provides one-half of the world's population more than 20% of its caloric intake but also serves as a nutrition source (Fitzgerald *et al.*, 2009; Hu *et al.*, 2014; Zhang *et al.*, 2016; Chen *et al.*, 2018). *Indica* and *japonica* are two subspecies of Asian cultivated rice (Ouyang and Zhang, 2013; Zhang *et al.*, 2016). A significant difference between *indica* and *japonica* occurs not only in the genome structure and gene content (Ouyang and Zhang, 2013; Zhang *et al.*, 2016) but also in the metabolite accumulation pattern (Chen *et al.*, 2014; Dong *et al.*, 2014; Hu *et al.*, 2014; Dong *et al.*, 2015; Fang *et al.*, 2016; Peng *et al.*, 2016). For instance, in our previous study in rice grains, a majority of the metabolites whose levels were substantially higher in the *indica* accessions were found to be C-glycosylated and malonylated flavonoids, whereas the *japonica* cultivars exhibited preferential accumulation of most amino acids, nucleic acids, and their derivatives (Chen *et al.*, 2016). Furthermore, Hu *et al.* (2014) identified a significant difference in the abundance of metabolites and in the metabolite–metabolite association networks between mature seeds of *indica* and *japonica*. Genomic studies revealed that *indica* rice is genetically much more diverse than is *japonica* rice (Huang *et al.*, 2010), indicating metabolic diversity across *indica* accessions. A population of RILs was derived from Zhenshan 97 (ZS97, *indica* I) and Minghui 63 (MH63, *indica* II), parents of the most widely cultivated elite hybrid in China (Zhang *et al.*, 2016). Previously, we performed a metabolic quantitative trait loci (mQTLs) analysis in rice grain and flag leaves (Gong *et al.*, 2013). In addition, using a data matrix of metabolites, including 683 in flag leaves and 317 in germinating seeds, we characterized the metabolic diversity between MH63 and ZS97 (Gong *et al.*, 2013).

Exploring the genetic basis underlying the metabolic diversity in plants with linkage and/or association mapping is of high interest (Matsuda *et al.*, 2012; Hu *et al.*, 2014; Kusano *et al.*, 2015; Hu *et al.*, 2016). Many efforts have been made to identify the genetic determinants of the metabolic variation in rice grains among different varieties. mQTL analysis is a powerful strategy for the dissection of the genetic regulation of metabolites. For example, approximately 800 mQTLs for more than 700 metabolite-related traits were identified with backcrossed inbred lines of Sasanishiki (*japonica*) and Hatabaki (*indica*) (Matsuda

et al., 2012). To decode the genetic regulation of the biosynthesis of the nonprotein amino acid, β -tyrosine, RILs were derived from Nipponbare and IR64, which produce and do not produce β -tyrosine, respectively (Yan *et al.*, 2015). Subsequent genetic mapping identified the causal gene that encodes a tyrosine aminomutase, whose function in affecting β -tyrosine accumulation was further confirmed *in vitro* and *in vivo* (Yan *et al.*, 2015). Our previous work also identified the complex genetic regulation underlying the metabolic diversity in flag leaves and/or germinating seeds of MH63 and ZS97. More than 1800 mQTLs for 683 metabolites and more than 800 mQTLs for 317 metabolites were identified in flag leaves and germinating seeds, respectively. In total, 509 mQTLs for 100 codetected metabolites were identified in both tissues, of which 463 mQTLs were distinct; moreover, 23 mQTLs for 19 metabolites were detected simultaneously in both tissues (Gong *et al.*, 2013). Although encouraging discoveries have been obtained for metabolic variation and corresponding genetic bases, knowledge of the diversity of plant metabolites in the same tissue at different developmental stages and of its genetic basis are limited.

In this study, we characterized the metabolic shift in rice seeds at different developmental stages, including seeds during grain filling, mature seeds and germinating seeds. Furthermore, we aimed to decipher the difference and the genetic determinants in grain metabolic dynamics across developmental stages between MH63 and ZS97. Therefore, a population of 210 RILs derived from a cross between ZS97 and MH63 was utilized for metabolic quantitative trait loci (mQTLs) analysis. By combining the results of mQTL analyses and *in silico* analyses, we characterized 35 candidate genes for 30 known compounds, among which the genes controlling the biosynthesis of feruloylserotonin and the content L-asparagine were further validated by transgenic assays. In addition, a metabolite–agronomic trait network was established using the correlation between metabolites in the seeds at three stages and agronomic traits, and mQTL analyses and pQTL analyses were combined to reveal the genetic basis underlying the complex traits consisting of metabolites and agronomics, which provide profound insights into crop improvements.

RESULTS

Metabolic profiling analyses of rice grains at different developmental stages

To obtain a global view of the metabolic variation in rice grains of *indica* cultivars at different developmental stages, metabolic profiling analyses were performed with germinating seeds and mature seeds from 96 and 108 accessions (Table S1), respectively. When the chromatographic and fragmental behaviors were compared directly with those of the commercial standards or were decoded with described

previously strategies (Chen *et al.*, 2013), 167 metabolites in germinating seeds and 296 in mature seeds were structurally identified or annotated (Table S2). To visualize the accumulation pattern of these metabolites in the seeds of *indica* cultivars, hierarchical clustering analyses were conducted. As shown in Figures S1 and S2, significant variation in the content of metabolites in germinating and mature seeds was observed within the *indica* accessions. As the parents of the most widely cultivated hybrid rice in China, MH63 and ZS97 were separated into two distinct clusters according to metabolome data of the germinating and mature seeds.

To further identify the diversity in metabolite content and metabolic shifting between the seeds of MH63 and ZS97 at different developmental stages, widely targeted metabolic profiling was performed with seeds during grain filling, mature seeds and germinating seeds. An MS² spectral tag (MS2T) library was constructed as previously described (Matsuda *et al.*, 2009; Chen *et al.*, 2013). Although 317 metabolites were detected in germinating seeds in our previous study (Gong *et al.*, 2013), we profiled 836 metabolites in germinating seeds by optimizing the method. Combined with 855 compounds in seeds during grain filling and 810 compounds in mature seeds, a data matrix was obtained that included a total of 2501 metabolites, with 372 metabolites detected in common (Figure 1a). Among these compounds, 233 metabolites were structurally identified or annotated and classified as amino acids, anthocyanins, fatty acids, flavonoids, polyamines or vitamins (Chen *et al.*, 2016). Scheduled multiple reaction monitoring (MRM), a high-throughput strategy (Chen *et al.*, 2013) was performed to quantify the metabolites. Principle component analysis (PCA) revealed that 42.4% and 28.9% of the variability were explained by components 1 and 2, respectively (Figure 1b). As shown in the PCA score plots, the separation among rice grains at different stages from the same cultivar was much wider than that among those of different cultivars at the same stage, indicating that the metabolic variation across developmental stages was much more significant than that between two different genotypes. The developmental stage-dependent accumulation pattern was further supported visually by a heatmap based on the metabolome data of MH63 and ZS97 at the three stages. As shown in Figure 1(c) and Table S5, in contrast to those at other stages, rice seeds at each stage accumulate different compounds at high levels. For example, the content of a series of flavonoids and amino acids in germinating seeds is significantly higher than that in mature seeds or seeds during grain filling. Moreover, compounds in the same class displayed different accumulation patterns. For instance, compared with the seeds at the other stages, mature seeds displayed a higher content of adenosine, although most nucleic acid derivatives tend to accumulate at a high level in seeds during grain filling and

in germinating seeds. Although the accumulation pattern of the codetected metabolites across the three stages in MH63 resembled that in ZS97, several compounds displayed different metabolic shifts (Figure 1c). For instance, the content of a tricin *O*-hexoside derivative (m0800) decreased to varying degrees in the mature seeds and germinating seeds compared with the seeds during grain filling of the MH63 cultivar. Although the content of m0800 was significantly lower in the seeds of ZS97 at each stage than in the seeds of MH63, the accumulation pattern of the same compound across the three stages in ZS97 was opposite that in MH63 (Figure S3), suggesting a complex regulatory network of the corresponding metabolite.

To explore the genetic determinants underlying the metabolic variation between MH63 and ZS97, a population of 210 RILs were used for further study. Samples of rice seeds were collected at three different developmental stages, including seeds during grain filling, mature seeds and germinating seeds. The metabolite content varied substantially among the RILs, with average coefficients of variation (CVs) of 60.91%, 71.44% and 60.71% in seeds during grain filling, mature seeds and germinating seeds, respectively (Figure 2a and Table S3). Nearly half of the metabolites have relatively high CVs (CV > 50%) at the three stages, especially for secondary metabolites, including anthocyanins, flavonoids, and terpenes (Table S4).

For the purpose of visualizing the metabolic variation across the RILs, hierarchical clustering analysis was conducted with the metabolome data from the aforementioned stages (Figure 2b). Dramatic metabolic variation was observed in the RILs both within and across stages. Several metabolites were found to accumulate preferentially in the seeds at certain stages. For instance, amino acids such as L-arginine, L-asparagine, L-histidine, L-serine and L-threonine accumulated at relatively high levels specifically in germinating seeds (Figure 2c). Moreover, we also observed a dynamic shifting in metabolites across seeds during grain filling, mature seeds and germinating seeds, indicating a continuous metabolic progress across the three stages. For example, increasing cytidine contents in the three stages was observed, while adenosine displayed the opposite trend (Figure 2d).

mQTL identification for rice seed metabolic variation

Benefiting from an ultra-high-density map consisting of 1619 bins generated by population sequencing (Yu *et al.*, 2011), mQTL mapping yielded 1600, 1506 and 1575 mQTLs with logarithm of odds (LOD) scores >3.0 in seeds during grain filling, mature seeds and germinating seeds, respectively; in addition, 86.0% (735/855), 79.1% (641/810) and 82.7% (691/836) of the detected metabolites had at least one mQTL, respectively. The number of mQTLs for each metabolite varied from one to seven, with 93, 109 and 119 metabolites having more than four mQTLs in seeds during

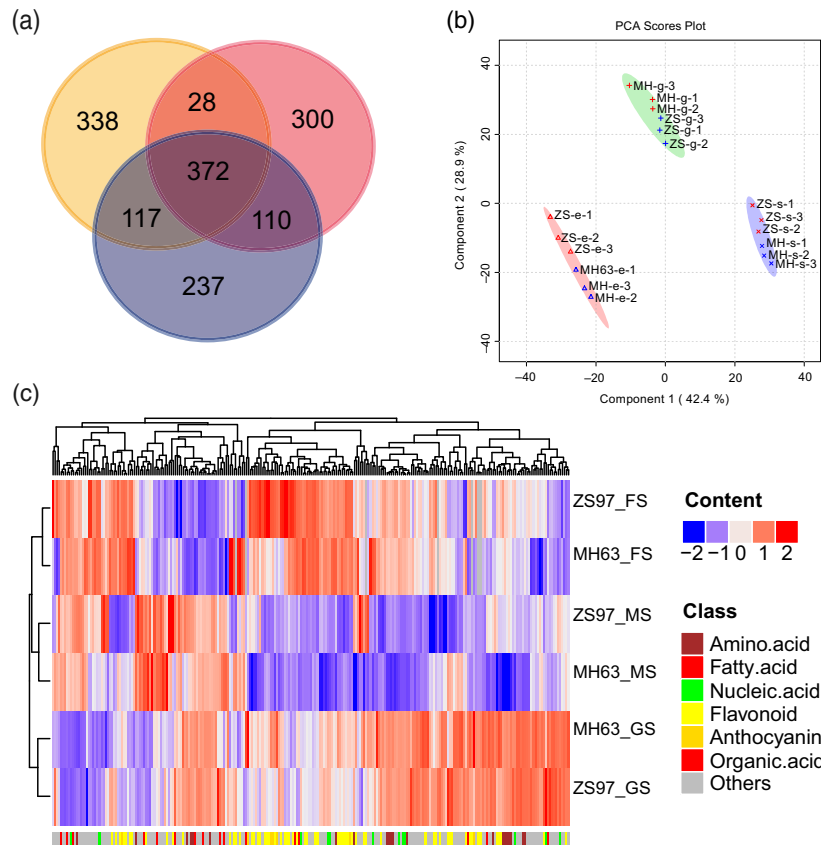


Figure 1. Metabolic variation in rice grains between MH63 and ZS97 across different developmental stages. (a) Venn diagram of the number of metabolites detected in seeds during grain filling (orange), in mature seeds (red) and in germinating seeds (blue). (b) Principle component analysis (PCA) of the metabolite profiling of seeds during grain filling (triangle), of mature seeds (asterisk) and of germinating seeds (plus). Three biological replications were used for metabolite profiling. The letters 'e', 's' and 'g' represent seeds during grain filling, mature seeds and germinating seeds, respectively. (c) Heatmap based on the metabolome data of MH63 and ZS97 rice seeds at three stages. Three biological replications were used for metabolite profiling. The metabolite profiles were analyzed for seeds during grain filling, for mature seeds and for germinating seeds. The content value of each metabolite was normalized, and hierarchical clustering was performed. The red color indicates a high abundance of a metabolite, whereas the blue color represents a low relative abundance of a metabolite. Each rice variety is visualized in a single row, and each metabolite is represented by a single column. The bottom annotation with different colors represents a different class to which the corresponding metabolite belongs. FS, seeds during grain filling; MS, mature seeds; GS, germinating seeds.

grain filling, mature seeds and germinating seeds, respectively (Table S6). Among the 1506 mQTLs detected in mature seeds, 1415, 70 and 21 loci are responsible for <20%, 20–50% and more than half of the variation in the corresponding compounds, respectively. In addition, 1460 and 1488 loci accounted for <20% of the variation in metabolites in germinating seeds and in seeds during grain filling, respectively (Figure 3a). Nevertheless, 26 and 39 mQTLs had effects greater than 50% in germinating seeds and seeds during grain filling, respectively. In total, 112, 91 and 115 mQTLs with effects greater than 20% were obtained in the seeds during grain filling, mature seeds and germinating seeds, respectively (Figure 3a and Table S6).

We found that the mQTLs displayed a significant deviation from random distribution across the whole genome in the seeds during grain filling ($\chi^2 = 292.11$, $P < 2.2e-16$), mature seeds ($\chi^2 = 508.30$, $P < 2.2e-16$) and germinating

seeds ($\chi^2 = 912.66$, $P < 2.2e-16$) of both cultivars (Table S8). The intervals enriching mQTLs may contain major genes for accumulation of a large number of compounds. Totally, 41, 42, and 35 potential mQTL 'hot spots' were characterized in seeds during grain filling, mature seeds, and germinating seeds, respectively. The mQTL hot spots in the mature seeds and germinating seeds were located mainly on chromosomes 5 and 6 and on all chromosomes except chromosome 3 in seeds during grain filling (Figure 3b).

When the mQTLs of individual metabolites were compared, we detected 1,804 distinct loci among the total 2068 loci detected for the 372 codetected metabolites in the seeds at three different stages (Table S7), suggesting that the majority of metabolites across the three stages may be under different genetic control. Further mQTL analysis revealed different types of genetic control of metabolism. Although 372 metabolites were detected in the seeds at all

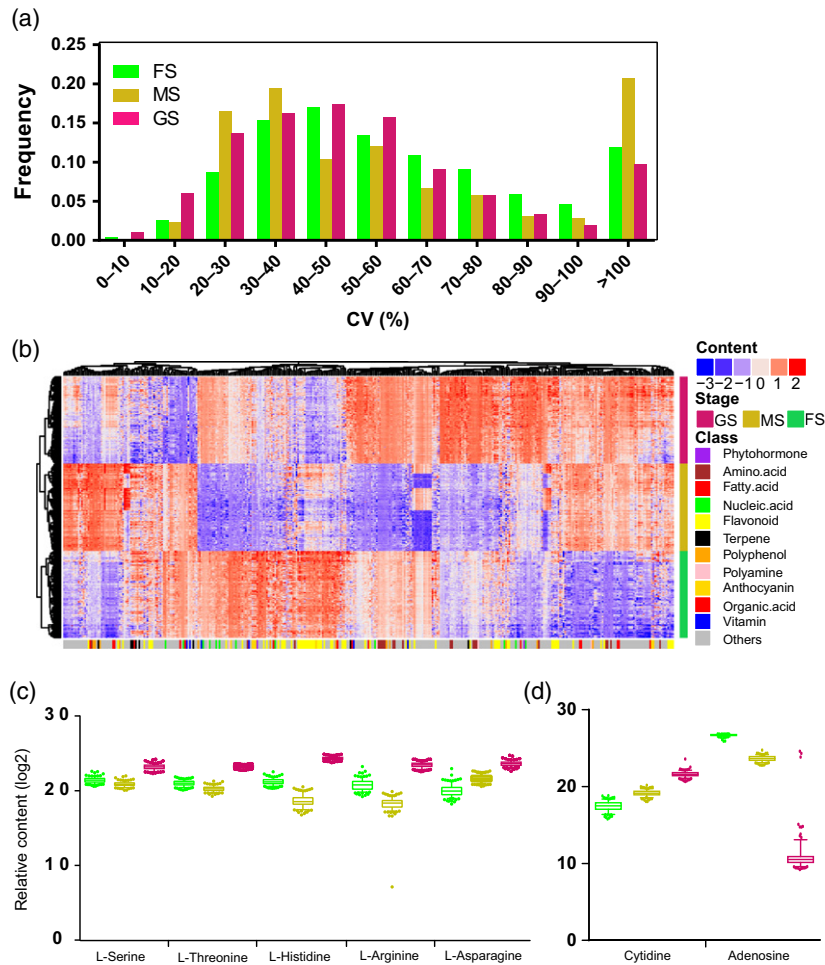


Figure 2. Metabolic variation in rice grains at different developmental stages across 210 RILs. (a) Distribution of the coefficient of variation (CV) across three stages. MS, mature seeds; GS, germinating seeds; FS, seeds during grain filling. (b) Heatmap based on metabolome data of rice seeds at three stages across 210 RILs. Two biological replications were used for metabolite profiling. The metabolite profiles were analyzed for seeds during grain filling, for mature seeds and for germinating seeds. The content value of each metabolite was normalized, and hierarchical clustering was performed. The red color indicates a high abundance of a metabolite, whereas the blue color represents a low relative abundance of a metabolite. Each rice variety is visualized in a single row, and each metabolite is represented by a single column. The bottom annotation with different colors represents a different class to which the corresponding metabolite belongs. The developmental stage is represented by the color of the bar on the right side. MS, mature seeds; GS, germinating seeds; FS, seeds during grain filling. (c) Box plot of the relative content of amino acids that accumulated at relatively high levels specifically in germinating seeds. (d) Different dynamic shifting in metabolites across three stages. The average relative contents of cytidine and adenosine in 210 RILs are given. In (c) and (d), the middle line of the box plots indicates the median, the box indicates the range of the 10th to 90th percentiles of the total data, and the outer dots are outliers. The purple, yellow and green colors represent germinating seeds, mature seeds and seeds during grain filling, respectively.

three stages, 70 mQTLs for 46 metabolites were detected at a specific stage, including 27, 24 and 19 mQTLs for 16, 14 and 16 metabolites in seeds during grain filling, mature seeds and germinating seeds, respectively. Although mQTLs for 102 codetected metabolites were detected in the seeds at two distinct stages, only approximately 18% of those individual metabolites shared overlapping mQTL segments at two stages, mostly (14 metabolites) in mature seeds and germinating seeds. Moreover, mQTLs for 218 codetected metabolites were identified in rice grains at all three stages (Table S7). Further data mining revealed that 60 mQTLs for 48 metabolites shared overlapping segments

across three stages, such as a major mQTL on chromosome 11 for the content of m2148. Overlapping mQTLs for individual compounds across different stages suggested that the genetic regulation of metabolite biosynthesis is conserved at various developmental stages. Although more than half of these metabolites have one or more mQTL(s) sharing overlapping segments in at least two stages, 96 metabolites have distinct mQTLs across three tissues. For example, the most significant mQTL across the three stages for luteolin 6-*C*-glucoside (m0476) was mapped to a 3.6 Mb region on chromosome 8 with LOD score of 5.84 in mature seeds. Although mQTLs overlapping with this major mQTL

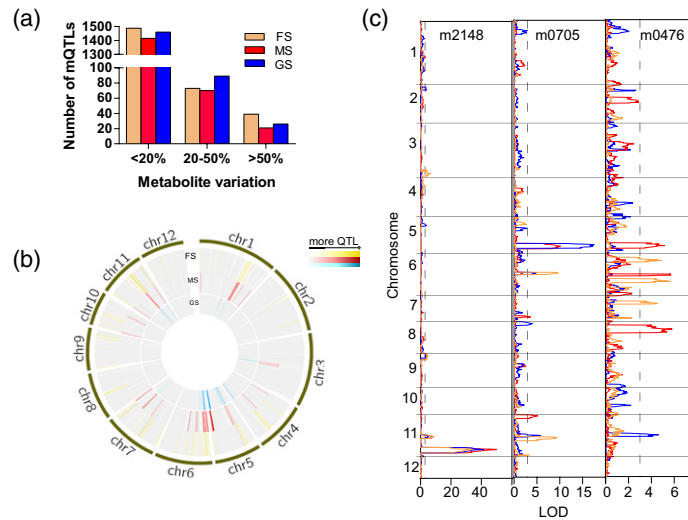


Figure 3. The statistical results of mQTLs at three stages. (a) Distribution of mQTLs accounting for metabolic variation across three stages. FS, seeds during grain filling; MS, mature seeds; GS, germinating seeds. (b) mQTLs and hotspots mapped onto each chromosome in seeds during grain filling (yellow), in mature seeds (red) and in germinating seeds (blue). (c) The development-dependent manner of metabolite biosynthesis control. mQTLs for m2148 at all three stages shared overlapping segments on chromosome 11 (left). Major mQTLs for m0705 in mature seeds and germinating seeds shared a 0.77 Mb segment on chromosome 5 (middle). mQTLs for m0476 did not overlap with each other across the three stages. The orange, red and blue colors represent seeds during grain filling, mature seeds and germinating seeds, respectively. LOD, logarithm of odds.

were not found at the other two stages, mQTLs for the same metabolite were mapped to 1 and 1.8 Mb regions on chromosome 11 and chromosome 6 in germinating seeds and seeds during grain filling, with LOD scores of 4.63 and 5.67, respectively (Figure 3c and Table S7). This finding suggested a development-dependent manner of metabolite biosynthesis control.

***In silico* analysis of candidate genes of mQTLs**

To identify candidate genes underlying the metabolic variation in the seeds of MH63 and ZS97, *in silico* analyses with genomic, transcriptomic and metabolomic data were carried out. To determine the common genetic regulators in the seeds at different stages for each individual metabolite, we focused on those mQTLs with overlapping segments across different stages. For instance, the feruloylserotonin content in MH63 was at least twice as high as that in ZS97 across all three stages (Figure S4b). mQTLs for feruloylserotonin across the three stages shared a 0.34 Mb segment on chromosome 7 and had relatively high LOD values (LOD values = 15.2, 17.1 and 23.9 in seeds during grain filling, mature seeds and germinating seeds, respectively; Figure S4a and Table S7). Based on our prior knowledge about metabolomics and metabolic pathways, we characterized plausible genes, including genes encoding glycosyltransferase, methyltransferase, acetyltransferase and so on (Peng *et al.*, 2016, 2017). A putative acetyltransferase gene, *LOC_Os07g04970*, was identified; the product of this gene may catalyze acylation with two different substrates: feruloyl-CoA and serotonin moieties. The expression of

LOC_Os07g04970 in seed tissues during grain filling, endosperm, was higher than that in germinating seeds in both MH63 and ZS97 (Figure S6a), which is consistent with the metabolic shift of feruloylserotonin (Figure S4b). Genome sequence analysis revealed several variants in the coding region of *LOC_Os07g04970* between MH63 and ZS97, resulting in the alternation of three amino acid residues (Figure S4b). However, the alternated amino acids are not in the conserved regions of the protein (Figure S5). Moreover, a 36-bp deletion was found in the promoter region in ZS97. In the radicle at 48 h after emergence under darkness, compared with that in MH63, the expression level of *LOC_Os07g04970* in ZS97 increased by approximately 14-fold (Figure S4c). Hence, *LOC_Os07g04970* was a candidate gene for the control of the biosynthesis of feruloylserotonin, whose genetic variants in the promoter and coding region might cooperatively affect its content variation.

As mentioned above, nearly 100 individual metabolites, including L-asparagine, have distinct mQTLs across three tissues. The most significant mQTL across the three stages for L-asparagine was mapped to a 0.79 Mb interval on chromosome 6, which displayed the highest LOD score of 12.6 in germinating seeds. Although mQTLs overlapping with this major mQTL were not found at the other two stages, mQTLs for the same metabolite were mapped to 0.86 and 1.41 Mb regions on chromosome 6 and chromosome 9 in mature seeds and seeds during grain filling with LOD scores of 8.3 and 4.3, respectively (Figure 4a). Although a plausible enzymatic gene underlying major QTLs in mature seeds and seeds during grain filling was not found, data

mining led to the identification of *LOC_Os06g03990* as a candidate genetic determinant affecting the content of L-asparagine in germinating seeds. As an aminotransferase, *LOC_Os06g03990* may participate in the transfer of an amino group from L-asparagine to an alpha-keto acid. The expression of *LOC_Os06g03990* in seed tissues during grain filling was higher than that in germinating seeds (Figure S6b), while the germinating seeds displayed higher content of L-asparagine (Figure 2c), suggesting a repressor role of *LOC_Os06g03990* in L-asparagine accumulation. Sequence alignment revealed the alternation of four amino acid residuals between MH63 and ZS97, including one in the aminotransferase class I and II domain conserved region (Figure 4b and Figure S7). There was no significant variation in the expression of *LOC_Os06g03990* across the whole life cycle between MH63 and ZS97 (Figure S6b). Hence, different contents of L-asparagine between MH63 and ZS97 might be the result of protein sequence variation of *LOC_Os06g03990*.

Global *in silico* analyses of the genetic determinants of all detected metabolites were carried out with the above-mentioned approach. In total, we characterized 35 candidate genes responsible for 30 structurally identified or annotated compounds (Table 1).

Validation of candidate genes with introgression and transgenic lines

To validate the candidate genes for metabolic variation, the contents of compounds in different haplotypes were analyzed in RILs. For example, we analyzed the feruloylserotonin content in 106 and 97 lines from the RILs carrying *LOC_Os07g04970* of the MH63 and ZS97 haplotypes, respectively. Feruloylserotonin accumulated in MH63 haplotype lines at an average level that was 2.5-fold that in the ZS97 haplotype lines across the aforementioned stages (Figure S4b). The feruloylserotonin content and genotype association analysis suggested that *LOC_Os07g04970* was responsible for the feruloylserotonin content variation across artificial populations. Candidate genes underlying variation in the content of primary metabolites were also validated with RILs. For instance, a 20% decrease in L-asparagine was observed in lines carrying *LOC_Os06g03990* of the MH63 haplotype compared with lines with the ZS97 haplotype (Figure 4b), which preliminarily confirmed the role of *LOC_Os06g03990* in the genetic determination of the metabolic variation in L-asparagine.

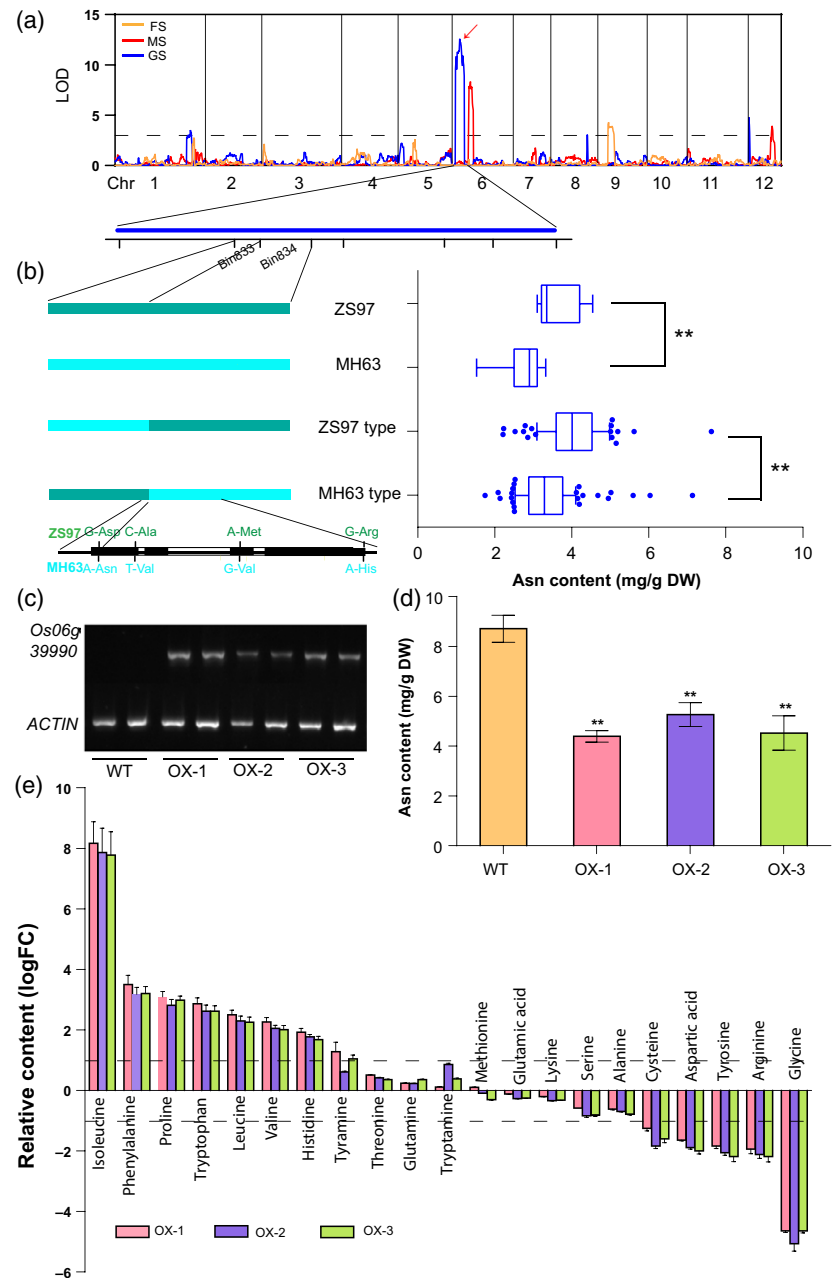
Although the experimental validation of all candidate genes' role in controlling metabolite accumulation is beyond the scope of a single study, we tested the *in vivo* functions of one candidate gene in the accumulation of the corresponding compounds. To confirm the *in vivo* function of *LOC_Os06g03990*, a candidate gene for the content of L-asparagine, overexpression lines were generated by introducing this gene driven by a constitutive promoter into the

Zhonghua 11 (ZH11) background. Expression analysis validated the overexpression of this gene in T3 lines (Figure 4c). Considering that *LOC_Os06g03990* was responsible for L-asparagine content variation only in germinating seeds, metabolic analysis was performed with rice grains at 72 h after germination. The content of L-asparagine in the transgenic lines was found to be less than half of the content in wild-type plants (Figure 4d), validating the repressor role of *LOC_Os06g03990* in L-asparagine accumulation. Moreover, overexpressing *LOC_Os06g03990* also resulted in significantly increased contents of Ile, Phe, Pro, Trp, Lue, Val, His, and tyramine. Moreover, a series of amino acids, such as Gly, Arg, tyrosine, Asp and Cys, displayed declined accumulation levels in the transgenic lines (Figure 4e). This result indicated that *LOC_Os06g03990* exerts global effects on amino acid biosynthesis.

Metabolite–metabolite network and relationships between metabolites and agronomic traits

To evaluate the coregulation of groups of metabolites affected by genetic variation, a series of correlation analyses were conducted using the metabolite profiles of the RIL population at the three aforementioned stages. Significant pairwise correlations ($|r| \geq 0.5$ and $P < 0.01$) between m-traits identified from each tissue were calculated. There were 5544, 3481, and 5286 significant correlations between metabolites detected in seeds during grain filling, in mature seeds, and in germinating seeds, respectively. To visualize the most significant correlation between the majority of structurally identified or annotated metabolites, an association network was constructed using correlation data with $r \geq 0.7$ or $r \leq -0.5$ ($P < 0.01$), encompassing metabolites from amino acids, nucleic acids, fatty acids, flavonoids, anthocyanins, polyamines, polyphenols, and terpenes (Table S9). Compounds of the same chemical class or those involved in the same biochemical pathway tended to display tight correlations with each other. Different stages exhibited distinct correlation networks in general. For instance, despite the high correlation value of triclin 4'-O-(syringyl alcohol) ether O-hexoside (m0821) and triclin 4'-O-(syringyl alcohol) ether O-hexoside (m0823) in seeds during grain filling, their correlation values were lower than 0.5 at the other two stages. Although the correlation values of the same metabolite pairs at different stages varied substantially, conserved correlations across stages were also found. For example, 2-amino-1,3,4,5-eicosanetetrol (m0344) and 4-hydroxysphinganine (m1494) displayed a strong correlation across the three aforementioned stages, with $r > 0.95$ (Figure 5a–c and Table S9). Moreover, we also identified some conserved correlations and subnetworks across different stages. For instance, strong conserved correlations across stages were observed between each pair of tricalysiamide B (m0284), kolaviv acid (m0295), and momilactone A (m1481) and

Figure 4. Validation of the role *LOC_Os06g39990* played in asparagine and amino acid accumulation. (a) QTLs mapped for L-asparagine (Asn) across three stages. The vertical gray lines indicate the separation of chromosomes. The horizontal dotted line indicates the threshold logarithm of odds (LOD) value. FS, seeds during grain filling; MS, mature seeds; GS, germinating seeds. (b) Sequence variation of *LOC_Os06g39990* and metabolic variation in Asn in germinating seeds across the RIL population. The turquoise color indicates that the fragment comes from ZS97, while the fragment from MH63 is cyan. Two biological replications were used for metabolite profiling. The average content of each metabolite was calculated to construct the box plot. The middle line of the box plots indicates the median, the box indicates the range of the 10th to 90th percentiles of the total data, and the outer dots are outliers. DW, dry weight. (c) The expression level of *LOC_Os06g39990* in its T3 overexpression lines (OX-1, OX-2 and OX-3) and wild-type (WT) plants. RNA samples were collected from the second upper leaf of 1-month-old plants. (d) The content of Asn in germinating seeds of transgenic (OX-1, OX-2 and OX-3) and WT plants. DW, dry weight. (e) Overexpressing *LOC_Os06g39990* resulted in altered accumulation levels of most amino acids in the germinating seeds of transgenic (OX-1, OX-2 and OX-3) and WT plants. FC, fold change. In (b) and (d), the asterisks indicate the levels of statistical significance as determined by Student's *t*-test: ** $P < 0.01$. In (d) and (e), three biological replications were used for metabolite profiling. The average content of each metabolite was calculated to construct the plots.



each pair of phytocassane E (m1492), gibberellin A53 (m1635), m0363, m1428, m1474, and m1480 (Figure 5a–c and Table S9).

Variation in metabolite levels constitutes one of the major causes of variations in trait manifestation (Prakash, 2011; Saito, 2013; Chen *et al.*, 2016). It is highly likely that many of the mQTLs would be the causes of phenotypic change. First, we assessed the correlations between the metabolite levels from each stage and seven agronomic traits, including heading date (HD), grain length (GL), grain

width (GW), kilo-grain weight (KGW), grain number per panicle (GN), tillers per plant (TP) and yield per plant (YD). A metabolite-agronomic network was built in which correlations were significant ($|r| > 0.2$, $P < 0.01$). In total, 233 significant correlations between 147 compounds and seven agronomic traits were identified, ranging from -0.63 to 0.68 (Table S10). For each agronomic trait, the number of metabolites that participated in the model ranged from 1 (GN) to 130 (HD), with an average of 29.3. Moreover, for each metabolite, the number of tightly correlated

Table 1 Candidate genes for metabolites detected in this study

ID	Stage ^a	LOD ^b	Interval (Mb)	NAG ^c	Candidate gene	Annotation
m0032	GS	8.3	24.43–25.33	170	<i>Os08g39300</i>	Aminotransferase
m0032	MS	9.6	27.78–28.80	235	<i>Os05g48010</i>	MYB
m0052	GS	13	1.74–2.22	101	<i>Os07g04560</i>	No apical meristem protein
m0054	GS	27.7	19.82–20.02	29	<i>Os12g32850</i>	Cytochrome P450
m0054	MS	50.7	1.74–2.11	830	<i>Os07g04410</i>	Cytochrome P450 90C1
m0076	MS	5.8	6.48–6.76	49	<i>Os06g12320</i>	TAAC
m0155	FS	34.3	21.37–21.60	43	<i>Os09g37200</i>	Transferase family protein
m0159	MS	6.9	24.42–25.16	117	<i>Os03g24339</i>	PWWP
m0164	FS	12.5	5.63–6.90	195	<i>Os09g12150</i>	OsFBX310
m0164	FS	15.5	18.81–18.82	6	<i>Os11g32650</i>	Chalcone synthase
m0214	MS	72.8	20.06–20.21	35	<i>Os09g34214</i>	UDP
m0418	FS	29.8	37.33–37.95	146	<i>Os01g65260</i>	AT
m0478	MS	56.8	31.03–31.05	1	<i>Os01g53460</i>	O-glucosyltransferase
m0506	GS	7.3	5.17–5.27	182	<i>Os01g10440</i>	Xylosyltransferase
m0863	GS	37.3	7.37–10.02	119	<i>Os09g16090</i>	UDP
m0898	FS	68.7	9.56–10.32	144	<i>Os10g18510</i>	UDP
m1015	MS	16.8	2.30–2.74	107	<i>Os06g05910</i>	MDCP
m1073	MS	5.1	20.75–21.80	214	<i>Os10g40200</i>	Aminotransferase
m1085	MS	20.2	0.00–0.49	86	<i>Os04g01590</i>	Arginase
m1085	MS	12.6	27.74–27.95	45	<i>Os05g48450</i>	Aminotransferase
m1085	MS	5.7	2.42–4.02	355	<i>Os06g05980</i>	Transporter family protein
m1089	MS	5.2	23.58–24.08	87	<i>Os12g39080</i>	Amino acid permease
m1090	GS	12.6	1.51–2.03	173	<i>Os06g03990</i>	Aminotransferase
m1231	MS	48.4	7.92–10.80	498	<i>Os12g16230</i>	Exostosin
m1288	GS	110.8	24.80–24.98	25	<i>Os11g42290</i>	Transferase family protein
m1377	MS	62.8	25.79–25.91	23	<i>Os11g42370</i>	Transferase family protein
m1515	FS	9.7	0.00–0.92	213	<i>Os01g01520</i>	Transferase
m1644	FS	29.9	23.36–23.89	98	<i>Os06g39470</i>	Transferase
m1647	GS	8.41	32.72–33.62	178	<i>Os01g56810</i>	CDP
m1650	GS	23.9	1.83–2.17	77	<i>Os07g04970</i>	Transferase family protein
m2085	FS	40.6	9.19–10.32	208	<i>Os10g18430</i>	ACT
m2129	FS	151.4	5.24–5.36	23	<i>Os06g10350</i>	MYB
m2181	MS	84.4	22.45–23.39	166	<i>Os12g37510</i>	UDP
m2254	MS	68.2	24.06–24.66	130	<i>Os05g41645</i>	Chalcone synthase
m2270	FS	8.7	18.31–18.82	114	<i>Os09g30980</i>	UDP

The names of metabolites corresponding to ID are given in Table S5.

Annotation abbreviations: TAAC, transmembrane amino acid transporter; PWWP, PWWP domain containing protein; UDP, UDP-glucuronosyl/UDP-glucosyl transferase; MDCP, methyltransferase domain containing protein; CDP, cytokinin dehydrogenase precursor; ACT, agmatine coumaroyltransferase.

^aGS, germinating seeds; MS, mature seeds; FS, grain-filling seeds.

^bLOD, logarithm of odds.

^cNAG, Number of annotated genes within QTL interval.

agronomic traits ranged from 1 to 3, with an average of 1.39. These results may suggest the complexity of the relationships between different agronomic traits and these metabolites. Most of the metabolite–agronomic trait correlations differed across the three stages, indicating that developmental status exerts effects on the correlations between metabolite–agronomic traits. For instance, 37, 92 and 25 compounds from seeds during grain filling, mature seeds and germinating seeds were found to be significantly correlated with HD. However, in some cases, the same metabolites detected at different stages were in the same model for a single agronomic trait. For instance, m0246, m1481 and m1600 detected at all three stages

appeared in the model for HD, suggesting tighter coregulation of HD and these compounds. In addition, some compounds were found to tightly correlated with multiple agronomic traits, for instance, m1931 participated in the model for four agronomic traits: HD, GL, GW and KGW (Figure 5d and Table S10). In the attempt to decipher the genetic and molecular bases for the variation in complex traits and correlated metabolites, we analyzed the colocalization between the mQTLs and phenotypic QTLs (pQTLs) for tightly correlated compounds and agronomic traits. For example, three QTLs for HD distributed on chromosomes 6, 7 and 11 were shared with mQTLs for at least one metabolite across the three stages, especially with mQTLs

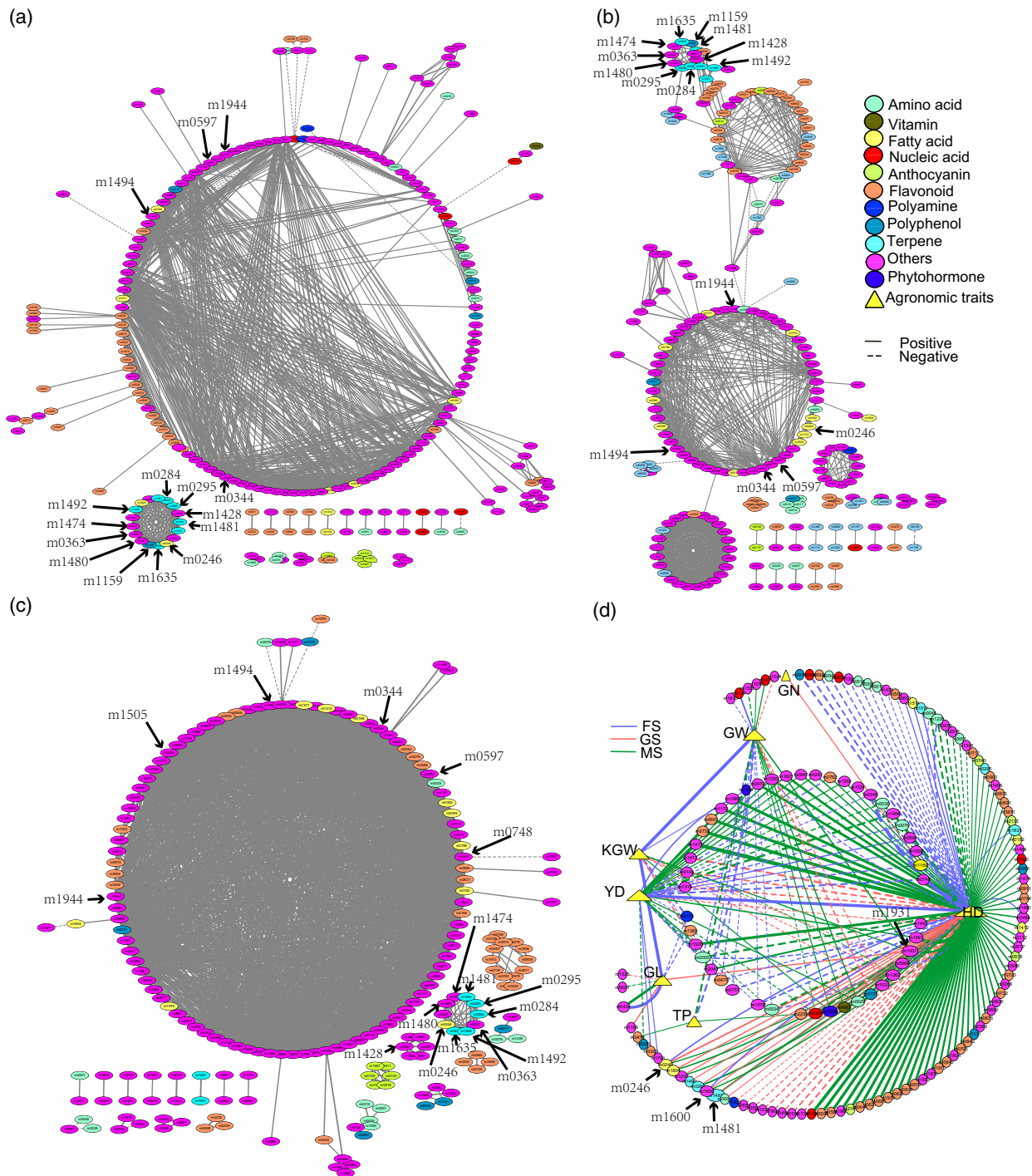


Figure 5. Metabolite–metabolite and metabolite–agronomic trait association network. (a–c) Network visualization of metabolites analyzed in seeds during grain filling (a), in mature seeds (b) and in germinating seeds (c). Only significant correlations are depicted. A significance level of $P < 0.01$ and an r value ≥ 0.7 or ≤ -0.5 were considered significant. (d) Metabolite–agronomic trait association network. This illustration represents the union of the metabolite–agronomic trait association network with the metabolic relevance networks obtained from each stage. Metabolites are represented as circular nodes, while triangular nodes represent agronomic traits. The size of each node is proportional to the number of correlated nodes. The color of each circular node indicates the class to which the corresponding compound belongs. The Pearson product-moment correlation was employed to compute all pairwise correlations between metabolites or metabolites and agronomic traits across the entire set of RILs. The relations are represented as edges. Positive correlations are denoted as solid lines, while dashed lines indicate negative correlations. The width of each edge depends linearly on the absolute value of the corresponding correlation. Computations of the correlations were conducted in the R environment, and Cytoscape was used to generate the graphics of the networks.

detected from mature seeds (Table 2). Further analysis revealed that *Ghd7.1* is located in the QTL region for HD on chromosome 7, which has been well documented as an important determinant of rice HD (Koo *et al.*, 2013; Liu *et al.*, 2013; Yan *et al.*, 2013; Gao *et al.*, 2014). This result suggested a potential role for *Ghd7.1* in coregulating HD and metabolite accumulation, including that of *trans*-zeatin *N*-glucoside.

DISCUSSION

As plant metabolites provide indispensable resources for human nutrition, energy and medicine (Butelli *et al.*, 2008; Chen *et al.*, 2016), dissecting the mechanism of metabolite biosynthesis in plants draws extreme interest (Saito and Matsuda, 2010; Cardoso *et al.*, 2014; Quadrana *et al.*, 2014; Zhao *et al.*, 2016; Fernie and Tohge, 2017; Perchat *et al.*, 2018; Tian *et al.*, 2018). In recent years, the rapid development of analysis approaches for metabolomes and multi-omics techniques have greatly improved our knowledge of the naturally occurring metabolic variation in plants and its underlying genetic determinants in several species (Keurentjes *et al.*, 2008; Shang *et al.*, 2014; Sadre *et al.*, 2016; Tohge *et al.*, 2016; Wen *et al.*, 2016; Fernie and Tohge, 2017; Rai *et al.*, 2017; Westhues *et al.*, 2017; Xiao *et al.*, 2017; Zhu *et al.*, 2018). As one of the most essential crop species, rice (*Oryza sativa* L.) not only feeds approximately half of the human population worldwide but also serves as a nutrition source. Hence, metabolic variation in rice grains between *japonica* and *indica* subspecies and its genetic basis have been well documented (Hu *et al.*, 2014; Chen *et al.*, 2016). However, the accumulation patterns of different metabolites within *indica* subspecies and their genetic basis have rarely been reported. In addition, the genetic and molecular bases of the plant metabolome of seeds is still limited to one developmental stage. To determine the metabolic variation in rice grains at different stages and disclose its genetic determinants, metabolic profiling in seeds during grain filling, mature seeds and germinating seeds was constructed with a population of 210 RILs derived from a cross between ZS97 and MH63 (Zhang *et al.*, 2016).

In our previous study, 317 metabolites were detected in germinating seeds of the same population. Distinct metabolic patterns and their genetic basis in different tissues were documented with data from leaves and germinating seeds (Gong *et al.*, 2013). In this study, by optimizing the method, we profiled more than 2500 metabolites in seeds at different developmental stages, of which 372 metabolites were detected in seeds at three stages, and 338, 300, and 237 were detected only in seeds during grain filling, mature seeds, and germinating seeds, respectively. (Figure 1a and Table S3). Varying metabolite accumulation patterns were found among samples at different stages of the same line and among samples of different lines at the

same stage (Figure 2b). The stage-specific accumulation of metabolites reflects the close association between biochemical synthesis pathways and developmental stage. The specific accumulation of amino acids, such as L-arginine, L-asparagine, L-histidine, L-serine and L-threonine, and the high content of the majority of other amino acids in germinating seeds might be indicative of the active degradation of stored protein. In addition, dynamic shifts in metabolites across the three stages were detected.

Although Hu *et al.* (2014) characterized diverse metabolic shifting in rice grains at different DAF, their work mainly focused on the metabolite profiles concerning the same biological process at various time points. Herein, we identified the metabolic variation in different biological processes in rice grains. Developmental stage and genotype consistently affecting metabolite variation was clarified in both our study and the previously mentioned work, despite differences between developmental stages and cultivars used. Moreover, metabolite-metabolite correlation analysis can be used to evaluate the coregulation of groups of metabolites. Although both various samples were used and different compounds were detected, a strong correlation between L-leucine and L-isoleucine was identified both in this study and the previously mentioned work, suggesting the conserved coregulation of those metabolites. In addition, conserved correlation across three different stages were found in our study. For instance, 2-amino-1,3,4,5-eicosanetetrol (m0344) and 4-hydroxy-sphinganine (m1494) displayed a strong correlation across the three aforementioned stages (Figure 5a–c and Table S9). Further analysis revealed that mQTLs on chromosome 5 for 2-amino-1,3,4,5-eicosanetetrol and 4-hydroxysphinganine shared the same segment, suggesting that there might be a gene in this region regulates the accumulation of those two compounds (Table S7). However, the previous study focused mainly on the relationship among identified compounds, leading to overlooking the correlations between compounds and uncharacterized metabolites. Nonbiased correlation network analysis in this work also uncovered strong conserved (m0597–m1944) and stage-specific (m0748–m1505 in germinating seeds) correlations between unknown metabolites, which might be helpful for studies on the characterization of compounds and their genetic regulation. Although the metabolite–metabolite correlation network displayed large variability across the three stages, metabolites of the same class or those involved in the same biochemical pathway tended to correlate with each other, which may help to elucidate new metabolite synthesis pathways.

A previous study reported that an mQTL analysis of rice grains resulted in the identification of 802 mQTLs for approximately 60% of the compounds detected (Matsuda *et al.*, 2012). Herein, taking advantage of the ultra-high-density genetic map generated by population sequencing

Table 2 Co-localization between the mQTLs and pQTLs for heading date

Trait ^a	Cor ^b	Chr	LOD ^c	Left Marker	Right Marker	Compound
HD		6	7.74	Bin855	Bin856	
m1894_MS	-0.53	6	3.76	Bin843	Bin860	Unknown
m1293_MS	0.51	6	4.54	Bin845	Bin857	2'-Deoxyadenosine monohydrate
m1691_GS	0.64	6	4.25	Bin847	Bin858	Unknown
m1477_MS	0.68	6	3.58	Bin847	Bin859	Unknown
m1689_MS	0.55	6	4.44	Bin850	Bin861	Unknown
m0612_MS	0.52	6	9.08	Bin855	Bin860	Unknown
m0076_MS	-0.53	6	3.25	Bin846	Bin856	L-tryptophan
m0036_MS	-0.46	6	3.38	Bin843	Bin861	L-histidine
m0601_MS	0.56	6	6.12	Bin850	Bin860	Phellodensin F
m1920_MS	0.54	6	5.33	Bin851	Bin859	Unknown
HD		7	5.76	Bin1057	Bin1058	
m0346_FS	-0.63	7	3.77	Bin1056	Bin1058	Unknown
m0739_MS	0.60	7	3.32	Bin1056	Bin1058	C-hexosyl-apigenin O-hexoside
m0368_MS	-0.44	7	5.96	Bin1055	Bin1058	Trans-zeatin N-glucoside
HD		11	4.31	Bin1411	Bin1412	Unknown
m0291_MS	0.36	11	5.05	Bin1412	Bin1418	Protocatechuic acid O-hexoside

^aHD, heading date; GS, germinating seeds; MS, mature seeds; FS, grain-filling seeds.

^bThe correlation between metabolite and phenotype.

^cLOD, logarithm of odds.

technology, we identified 1600, 1506 and 1575 metabolic quantitative trait loci (mQTLs) for approximately 80% of metabolites detected in seeds during grain filling, mature seeds and germinating seeds, respectively (Table S3). Although 372 metabolites were detected in the seeds at all three stages and 366 metabolites had corresponding mQTLs (Figure S8 and Table S7), only 48 metabolites had 60 mQTLs that shared overlapping segments across all three stages, and approximately 1/4 of the metabolites had distinct mQTLs across the three stages. Moreover, 70 mQTLs for 46 metabolites were identified at a specific stage (Table S7), which indicated a development-dependent manner of metabolite biosynthesis control. Further analysis revealed that the 46 metabolites were distributed in various classes, such as fatty acids, polyphenols, amino acids and vitamin, etc. A previous study identified significant mQTLs for asparagine and a putative asparagine synthase as a candidate gene on chromosome 3 via *japonica-indica*-derived IILs (Matsuda *et al.*, 2012). However, the validation of the candidate gene was lacking. In this study, seven putative mQTLs for L-asparagine were identified, which were distributed across five chromosomes (Table S7). A gene encoding an aminotransferase, *LOC_Os06g03990*, was considered as a genetic determinant for the metabolic variation of L-asparagine in RILs. Sequence alignment identified four different amino acid residuals between MH63 and ZS97, including an alternation from valine (V) to methionine (M) in the predicted conserved aminotransferase class I and II domain, which may affect the activity of *LOC_Os06g03990* (Figure 4b and Figure S7). Moreover, *LOC_Os06g03990* was validated as a repressor of L-asparagine content by a

transgenic approach. This finding revealed the genetic basis of metabolite variation among different varieties.

In seeds during grain filling, compared with that in germinating seeds, the expression of *LOC_Os06g03990* increased and the content of L-asparagine decreased (Figure 2c and Figure S6b). In addition, we also found that the expression of *LOC_Os07g04970* in seed tissues during grain filling was higher than in germinating seeds and that the content of ferulylserotonin (Figures S6a and S4b). These results suggested that the altered expression level of the key genes across different stages may be responsible for the metabolic shift.

Moreover, metabolite–agronomic trait association and colocalization between mQTLs and pQTLs revealed the complexity of the metabolite–agronomic trait relationship and the corresponding genetic basis. Through further data mining, more than 30 candidate genes modulating the accumulation of metabolites that are of potential physiological and nutritional importance were identified.

This study significantly improved our knowledge of the genetic and biochemical bases of rice seed metabolome variation at different stages and provides profound insights into rice breeding strategies that increase yields while maintaining high nutritional levels.

EXPERIMENTAL PROCEDURES

Plant materials and growth conditions

The RILs population used for linkage mapping are of 210 lines derived from a cross between ZS97 and MH63 (Zhang *et al.*, 2016). The rice plants used in this study were grown during the

normal rice growing seasons in the field at the experimental farm of Huazhong Agricultural University (Wuhan, China, E 109°51', N 18°25'). All the seeds were germinated for 3 days at 37°C on filter paper soaked in distilled water and then planted in seedbeds in mid-May; the seedlings were subsequently transplanted into the field in mid-June. The field management followed normal agricultural practices.

Sample preparation

(i) Germinating seeds. The seeds of RILs, MH63 and ZS97 were first soaked in water for 2 days under 25°C and 85% relative humidity in the dark. Subsequently, the pregermination (35°C temperature, 85% relative humidity, darkness) was conducted, followed by the incubation for 72 h (25°C temperature, 85% relative humidity, darkness). We harvested the seeds from 15 seedlings per line to extract the metabolites. (ii) Seeds during grain filling. The seeds during grain filling were prepared 10 DAF in 2009 using liquid nitrogen for metabolite extraction. Each sample was from three different plants per line grown in the field. (iii) Mature seeds. Mature seeds were harvested at the mature stage in 2009. Each sample contained 10 seeds for metabolite extraction.

Metabolite profiling

We powdered freeze-dried samples using a mix mill (MM 400 Retsch) with a zirconia bead for 1 min at 30 Hz. Then, 1.0 ml of 70% aqueous methanol per 100 mg of powder was used for metabolite extraction (overnight at 4°C), containing 0.1 mg L⁻¹ lidocaine as internal standard (Chen *et al.*, 2013). We performed an MRM method to quantify m-traits (Dresen *et al.*, 2010; Matsuda *et al.*, 2012; Chen *et al.*, 2013). The relative signal intensities of the metabolites were normalized by first dividing them by the intensities of the internal standard (0.1 mg L⁻¹ lidocaine) and then subjecting them to log₂ transformation to improve the normality further. In total, 810, 836 and 855 transitions in mature seeds, germinating seeds, and seeds during grain filling were monitored, respectively, with positive polarity. We used the scheduled MRM algorithm in Analyst 1.5 software, setting the MRM detection window to 80 sec and the target scan time to 1.5 sec.

Statistical analysis

The metabolite data of the RIL population comprise the means of three technical replications from the LC-MS/MS of one biological replicates. For each individual metabolite, the content was given as the average of the normalized metabolite levels in three replications. Totally, 810, 836 and 855 m-traits were obtained in the mature seeds, the germinating seeds and, the seeds during grain filling, respectively. We calculated the values of the genetic CV for each compound as previously described (Chan *et al.*, 2013). Pairwise Pearson correlations between metabolites detected were estimated by R (www.r-project.org). Subsequently, Gaussian graphical modeling (GGM) was constructed based on pairwise Pearson correlation coefficients. Metabolite networks and metabolite-agronomic trait networks were constructed based on the correlation matrices and realized by the program Cytoscape (3.7.0).

QTL mapping and detection of mQTL hot spots

We conducted a QTL mapping with the RIL population as described in previous works (Weibo *et al.*, 2010; Yu *et al.*, 2011; Gong *et al.*, 2013). Bin maps were composed of 1619 recombinant bins without missing data. We set the LOD threshold to 3.0, with a 1.5-LOD-drop support interval. The whole genome was divided into 1 cM partitions. A permutation test was performed as

previously described (Gong *et al.*, 2013). According to the results of 1000 permutations, the cutoff number of mQTLs per cM by chance alone was estimated to be seven, seven, and eight in the mature seeds, germinating seeds, and the seeds during grain filling, respectively. A larger number of mQTLs in 1 cM indicates the existence of an mQTL hot spot.

Plasmid construction and rice transformation

Gateway recombination reactions (Invitrogen, Waltham, MA, USA) were performed to generate the overexpression construct of *Os06g03990*. Firstly, the full cDNA was amplified from cDNA samples of ZH11, encoding the same protein sequence with that from MH63. Then, the full cDNA was inserted into the donor vector pDONR207, producing the entry clone, which was subjected to an LR reaction with the destination vector pJC034 (Dong *et al.*, 2015). The overexpression construct of *Os06g03990* was subsequently introduced into *Agrobacterium* strain AH105 and then transferred into *japonica* ZH11 as described previously (Hiei *et al.*, 1994). *LOC_Os06g03990* cDNA was amplified by PCR with a set of primers named OJT31 and OJT32 (Table S11), with leaf-derived cDNA used as a template.

RT-PCR

Total RNA was extracted with TRIzol reagent (Invitrogen, Waltham, MA, USA), followed by treatment with DNase I (Thermo Scientific, Waltham, MA, USA). Subsequently, 3 µg of RNA was used for synthesis of the first-strand with M-MLV reverse transcriptase (ZOMANBIO, Beijing, China). We conducted RT-PCR to detect the expression of targeted genes using primers listed in Table S11.

AUTHOR CONTRIBUTIONS

CF and JL conceived and designed this study. KL, DW, LG, YL, HG, CJ and XL performed experiments. WC carried out the metabolite analyses. KL, JL and CF analyzed the data. KL, JL and CF wrote the manuscript. All of the authors discussed the results and commented on the manuscript.

ACKNOWLEDGMENTS

This work was supported by the National Science Fund for Distinguished Young Scholars (31625021), the State Key Program of National Natural Science Foundation of China (31530052), the National Natural Science Foundation of China (31800250, and 31960063), and the Hainan University Startup Fund (KYQD (ZR)1866 to JL, and KYQD(ZR)1824 to CF).

Data availability

All the data used in this study are available from the corresponding author upon request.

CONFLICT OF INTERESTS

The authors declare that they have no conflict of competing interest.

SUPPORTING INFORMATION

Additional Supporting Information may be found in the online version of this article.

Figure S1. Heat map of metabolite accumulations in germinating seeds of *indica* cultivars.

Figure S2. Heat map of metabolite accumulations in mature seeds of *indica* cultivars.

Figure S3. Relative content of m0800 in MH63 and ZS97 seeds at three developmental stages.

Figure S4. Validation of the role *LOC_Os07g04970* played in feruloylserotonin accumulation.

Figure S5. Variations of amino acid residues of *LOC_Os07g04970* between MH63 and ZS97.

Figure S6. Expression profiles of *LOC_07g04970* and *LOC_Os06g03990* in MH63 and ZS97.

Figure S7. Variations of amino acid residues of *LOC_Os06g03990* between MH63 and ZS97.

Figure S8. Venn diagram for the number of metabolites with mQTLs codetected in all three stages.

Table S1. Natural accessions used in metabolic profiling.

Table S2. Metabolites detected in the screening across natural accessions.

Table S3. Metabolites detected in the screening across RILs derived from a cross between MH63 and ZS97.

Table S4. Statistical results of the CVs of various classes metabolites in the seeds across three stages of RILs.

Table S5. Metabolites variation in rice grains between MH63 and ZS97.

Table S6. Statistical results of mQTLs for all of the detected metabolites at three stages.

Table S7. Statistical results of mQTLs for codetected metabolites at three stages.

Table S8. Statistics of metabolic quantitative trait loci (mQTLs) on the chromosomes.

Table S9. Metabolite-metabolite correlationship.

Table S10. Agronomic trait-Metabolite correlationship.

Table S11. Primers used in this study.

REFERENCES

- Afendi, F.M., Okada, T., Yamazaki, M. *et al.* (2012) KNApSACk family databases: integrated metabolite-plant species databases for multifaceted plant research. *Plant Cell Physiol.* **53**, e1.
- Biais, B., Beauvoit, B., William Allwood, J., Deborde, C., Maucourt, M., Goodacre, R., Rolin, D. and Moing, A. (2010) Metabolic acclimation to hypoxia revealed by metabolite gradients in melon fruit. *J. Plant Physiol.* **167**, 242–245.
- Borevitz, J.O., Hazen, S.P., Michael, T.P. *et al.* (2007) Genome-wide patterns of single-feature polymorphism in *Arabidopsis thaliana*. *Proc. Natl Acad. Sci. USA*, **104**, 12057–12062.
- Butelli, E., Titta, L., Giorgio, M. *et al.* (2008) Enrichment of tomato fruit with health-promoting anthocyanins by expression of select transcription factors. *Nat. Biotechnol.* **26**, 1301–1308.
- Cardoso, C., Zhang, Y., Jamil, M. *et al.* (2014) Natural variation of rice strigolactone biosynthesis is associated with the deletion of two MAX1 orthologs. *Proc. Natl Acad. Sci. USA*, **111**, 2379–2384.
- Chan, E.K.F., Rowe, H.C., Hansen, B.G. and Kliebenstein, D.J. (2013) The complex genetic architecture of the metabolome. *PLoS Genet.* **6**, e1001198.
- Chen, W., Gong, L., Guo, Z., Wang, W., Zhang, H., Liu, X., Yu, S., Xiong, L. and Luo, J. (2013) A novel integrated method for large-scale detection, identification, and quantification of widely targeted metabolites: application in the study of rice metabolomics. *Mol. Plant*, **6**, 1769–1780.
- Chen, W., Gao, Y., Xie, W. *et al.* (2014) Genome-wide association analyses provide genetic and biochemical insights into natural variation in rice metabolism. *Nat. Genet.* **46**, 714–721.
- Chen, W., Wang, W., Peng, M. *et al.* (2016) Comparative and parallel genome-wide association studies for metabolic and agronomic traits in cereals. *Nat. Commun.* **7**, 12767.
- Chen, J., Wang, J., Chen, W. *et al.* (2018) Metabolome analysis of multi-connected biparental chromosome segment substitution line populations. *Plant Physiol.* **178**, 612–625.
- De Luca, V., Salim, V., Atsumi, S.M. and Yu, F. (2012) Mining the biodiversity of plants: a revolution in the making. *Science*, **336**, 1658–1661.
- Dixon, R.A. and Strack, D. (2003) Phytochemistry meets genome analysis, and beyond. *Phytochemistry*, **62**, 815–816.
- Dong, X., Chen, W., Wang, W., Zhang, H., Liu, X. and Luo, J. (2014) Comprehensive profiling and natural variation of flavonoids in rice. *J. Integr. Plant Biol.* **56**, 876–886.
- Dong, X., Gao, Y., Chen, W., Wang, W., Gong, L., Liu, X. and Luo, J. (2015) Spatiotemporal distribution of phenolamides and the genetics of natural variation of hydroxycinnamoyl spermidine in rice. *Mol. Plant*, **8**, 111–121.
- Dresen, S., Ferreiros, N., Gnann, H., Zimmermann, R. and Weinmann, W. (2010) Detection and identification of 700 drugs by multi-target screening with a 3200 Q TRAP LC-MS/MS system and library searching. *Anal. Bioanal. Chem.* **396**, 2425–2434.
- Fang, C. and Luo, J. (2019) Metabolic GWAS-based dissection of genetic bases underlying the diversity of plant metabolism. *Plant J.* **97**, 91–100.
- Fang, C., Zhang, H., Wan, J. *et al.* (2016) Control of leaf senescence by an MeOH-Jasmonates cascade that is epigenetically regulated by *OsSRT1* in rice. *Mol. Plant*, **9**, 1366–1378.
- Fang, C., Fernie, A.R. and Luo, J. (2019a) Exploring the diversity of plant metabolism. *Trends Plant Sci.* **24**, 83–98.
- Fang, C., Li, K., Wu, Y. *et al.* (2019b) *OsTSD2*-mediated cell wall modification affects ion homeostasis and salt tolerance. *Plant Cell Environ.* **42**, 1503–1512.
- Fernie, A.R. and Tohge, T. (2017) The genetics of plant metabolism. *Annu. Rev. Genet.* **51**, 287–310.
- Fitzgerald, M.A., McCouch, S.R. and Hall, R.D. (2009) Not just a grain of rice: the quest for quality. *Trends Plant Sci.* **14**, 133–139.
- Gao, H., Jin, M., Zheng, X.M. *et al.* (2014) Days to heading 7, a major quantitative locus determining photoperiod sensitivity and regional adaptation in rice. *Proc. Natl Acad. Sci. USA*, **111**, 16337–16342.
- Gong, L., Chen, W., Gao, Y., Liu, X., Zhang, H., Xu, C., Yu, S., Zhang, Q. and Luo, J. (2013) Genetic analysis of the metabolome exemplified using a rice population. *Proc. Natl Acad. Sci. USA*, **110**, 20320–20325.
- Hiei, Y., Ohta, S., Komari, T. and Kumashiro, T. (1994) Efficient transformation of rice (*Oryza sativa* L.) mediated by *Agrobacterium* and sequence analysis of the boundaries of the T-DNA. *Plant J.* **6**, 271–282.
- Hu, C., Shi, J., Quan, S. *et al.* (2014) Metabolic variation between *japonica* and *indica* rice cultivars as revealed by non-targeted metabolomics. *Sci. Rep.* **4**, 5067.
- Hu, C., Tohge, T., Chan, S.A. *et al.* (2016) Identification of conserved and diverse metabolic shifts during rice grain development. *Sci. Rep.* **6**, 20942.
- Huang, X., Wei, X., Sang, T. *et al.* (2010) Genome-wide association studies of 14 agronomic traits in rice landraces. *Nat. Genet.* **42**, 961–967.
- Keurentjes, J.J. (2009) Genetical metabolomics: closing in on phenotypes. *Curr. Opin. Plant Biol.* **12**, 223–230.
- Keurentjes, J.J., Fu, J., de Vos, C.H., Lommen, A., Hall, R.D., Bino, R.J., van der Plas, L.H., Jansen, R.C., Vreugdenhil, D. and Koornneef, M. (2006) The genetics of plant metabolism. *Nat. Genet.* **38**, 842–849.
- Keurentjes, J.J., Sulpice, R., Gibon, Y., Steinhauser, M.C., Fu, J., Koornneef, M., Stitt, M. and Vreugdenhil, D. (2008) Integrative analyses of genetic variation in enzyme activities of primary carbohydrate metabolism reveal distinct modes of regulation in *Arabidopsis thaliana*. *Genome Biol.* **9**, R129.
- Koo, B.H., Yoo, S.C., Park, J.W., Kwon, C.T., Lee, B.D., An, G., Zhang, Z., Li, J., Li, Z. and Paek, N.C. (2013) Natural variation in *OsPRR37* regulates heading date and contributes to rice cultivation at a wide range of latitudes. *Mol. Plant*, **6**, 1877–1888.
- Kusano, M., Yang, Z.G., Okazaki, Y., Nakabayashi, R., Fukushima, A. and Saito, K. (2015) Using metabolomic approaches to explore chemical diversity in rice. *Mol. Plant*, **8**, 58–67.
- Liu, T., Liu, H., Zhang, H. and Xing, Y. (2013) Validation and characterization of Ghd7.1, a major quantitative trait locus with pleiotropic effects on spikelets per panicle, plant height, and heading date in rice (*Oryza sativa* L.). *J. Integr. Plant Biol.* **55**, 917–927.
- Luo, J. (2015) Metabolite-based genome-wide association studies in plants. *Curr. Opin. Plant Biol.* **24**, 31–38.
- Matsuda, F., Yonekura-Sakakibara, K., Niida, R., Kuromori, T., Shinozaki, K. and Saito, K. (2009) MS/MS spectral tag-based annotation of non-targeted profile of plant secondary metabolites. *Plant J.* **57**, 555–577.

- Matsuda, F., Okazaki, Y., Oikawa, A., Kusano, M., Nakabayashi, R., Kikuchi, J., Yonemaru, J., Ebana, K., Yano, M. and Saito, K. (2012) Dissection of genotype-phenotype associations in rice grains using metabolome quantitative trait loci analysis. *Plant J.* **70**, 624–636.
- Morohashi, K., Casas, M.I., Ferreyra, M.L.F. et al. (2012) A genome-wide regulatory framework identifies maize pericarp color1 controlled genes. *Plant Cell*, **24**, 2745–2764.
- Mounet, F., Lemaire-Chamley, M., Maucourt, M., Cabasson, C., Giraudel, J.L., Deborde, C., Lessire, R., Gallusci, P., Bertrand, A. and Gaudillère, M. (2007) Quantitative metabolic profiles of tomato flesh and seeds during fruit development: complementary analysis with ANN and PCA. *Metabolomics*, **3**, 273–288.
- Ouyang, Y. and Zhang, Q. (2013) Understanding reproductive isolation based on the rice model. *Annu. Rev. Plant Biol.* **64**, 111–135.
- Peng, M., Gao, Y., Chen, W. et al. (2016) Evolutionarily distinct BAHD N-acyltransferases are responsible for natural variation of aromatic amine conjugates in rice. *Plant Cell*, **28**, 1533–1550.
- Peng, M., Shahzad, R., Gul, A. et al. (2017) Differentially evolved glucosyltransferases determine natural variation of rice flavone accumulation and UV-tolerance. *Nat. Commun.* **8**, 12.
- Perchat, N., Saaidi, P.L., Darii, E. et al. (2018) Elucidation of the trigonelline degradation pathway reveals previously undescribed enzymes and metabolites. *Proc. Natl Acad. Sci. USA*, **115**, E4358–E4367.
- Prakash, S. (2011) Human metabolic individuality in biomedical and pharmaceutical research. *Circ. Cardiovasc. Genet.* **4**, 714–715.
- Quadrana, L., Almeida, J., Asis, R. et al. (2014) Natural occurring epialleles determine vitamin E accumulation in tomato fruits. *Nat. Commun.* **5**, 3027.
- Rai, A., Saito, K. and Yamazaki, M. (2017) Integrated omics analysis of specialized metabolism in medicinal plants. *Plant J.* **90**, 764–787.
- Sadre, R., Magallanes-Lundback, M., Pradhan, S., Salim, V., Mesberg, A., Jones, A.D. and DellaPenna, D. (2016) Metabolite diversity in alkaloid biosynthesis: a multilane (diastereomer) highway for camptothecin synthesis in *camptotheca acuminata*. *Plant Cell*, **28**, 1926–1944.
- Saito, K. (2013) Phytochemical genomics—a new trend. *Curr. Opin. Plant Biol.* **16**, 373–380.
- Saito, K. and Matsuda, F. (2010) Metabolomics for functional genomics, systems biology, and biotechnology. *Annu. Rev. Plant Biol.* **61**, 463–489.
- Shang, Y., Ma, Y., Zhou, Y. et al. (2014) Biosynthesis, regulation, and domestication of bitterness in cucumber. *Science*, **346**, 1084–1088.
- Silva, A.T., Ligterink, W. and Hilhorst, H.W.M. (2017) Metabolite profiling and associated gene expression reveal two metabolic shifts during the seed-to-seedling transition in *Arabidopsis thaliana*. *Plant Mol. Biol.* **95**, 481–496.
- Tian, X., Ruan, J.X., Huang, J.Q. et al. (2018) Characterization of gossypol biosynthetic pathway. *Proc. Natl Acad. Sci. USA*, **115**, E5410–E5418.
- Tohge, T., Wendenburg, R., Ishihara, H. et al. (2016) Characterization of a recently evolved flavonol-phenylacyltransferase gene provides signatures of natural light selection in *Brassicaceae*. *Nat. Commun.* **7**, 12399.
- Watanabe, M., Balazadeh, S., Tohge, T., Erban, A., Giavalisco, P., Kopka, J., Mueller-Roeber, B., Fernie, A.R. and Hoefgen, R. (2013) Comprehensive dissection of spatiotemporal metabolic shifts in primary, secondary, and lipid metabolism during developmental senescence in *Arabidopsis*. *Plant Physiol.* **162**, 1290–1310.
- Weibo, X., Qi, F., Huihui, Y., Xuehui, H., Qiang, Z., Yongzhong, X., Sibin, Y., Bin, H. and Qifa, Z. (2010) Parent-independent genotyping for constructing an ultrahigh-density linkage map based on population sequencing. *Proc. Natl Acad. Sci. USA*, **107**, 10578–10583.
- Wen, W., Liu, H., Zhou, Y. et al. (2016) Combining quantitative genetics approaches with regulatory network analysis to dissect the complex metabolism of the maize kernel. *Plant Physiol.* **170**, 136–146.
- Westhues, M., Schrag, T.A., Heuer, C. et al. (2017) Omics-based hybrid prediction in maize. *Theor. Appl. Genet.* **130**, 1927–1939.
- Wurtzel, E.T. and Kutchan, T.M. (2016) Plant metabolism, the diverse chemistry set of the future. *Science*, **353**, 1232–1236.
- Xiao, Y., Liu, H., Wu, L., Warburton, M. and Yan, J. (2017) Genome-wide association studies in maize: praise and stargaze. *Mol. Plant*, **10**, 359–374.
- Yan, W., Liu, H., Zhou, X. et al. (2013) Natural variation in *Ghd7.1* plays an important role in grain yield and adaptation in rice. *Cell Res.* **23**, 969–971.
- Yan, J., Aboshi, T., Teraishi, M. et al. (2015) The tyrosine aminomutase TAM1 is required for beta-tyrosine biosynthesis in rice. *Plant Cell*, **27**, 1265–1278.
- Yu, H., Xie, W., Wang, J., Xing, Y., Xu, C., Li, X., Xiao, J. and Zhang, Q. (2011) Gains in QTL detection using an ultra-high density SNP map based on population sequencing relative to traditional RFLP/SSR markers. *PLoS ONE*, **6**, e17595.
- Zhang, J., Chen, L.L., Xing, F. et al. (2016) Extensive sequence divergence between the reference genomes of two elite indica rice varieties Zhen-shan 97 and Minghui 63. *Proc. Natl Acad. Sci. USA*, **113**, E5163–E5171.
- Zhao, Q., Zhang, Y., Wang, G., Hill, L., Weng, J.K., Chen, X.Y., Xue, H. and Martin, C. (2016) A specialized flavone biosynthetic pathway has evolved in the medicinal plant, *Scutellaria baicalensis*. *Sci. Adv.* **2**, e1501780.
- Zhu, G., Wang, S., Huang, Z. et al. (2018) Rewiring of the fruit metabolome in tomato breeding. *Cell*, **172**, 249–261.