PROFESSOR WEIWEI WEN (Orcid ID : 0000-0002-3413-4766)

DR JIANBING YAN (Orcid ID : 0000-0001-8650-7811)

DR YINGJIE XIAO (Orcid ID : 0000-0001-7276-653X)

Article type : Original Article

Metabolomic analysis reveals differences in evolution between maize and rice

Min Deng[1,2], Xuehai Zhang[3], Jingyun Luo[2], Haijun Liu[2], Weiwei Wen[4], Hongbing Luo[1], Jianbing Yan[2] and Yingjie Xiao[2,*]

[1]College of Agronomy, Hunan Agricultural University, Changsha, Hunan, 410128, China

[2]National Key Laboratory of Crop Genetic Improvement, Huazhong Agricultural University, Wuhan, 430070, China

[3]National Key Laboratory of Wheat and Maize Crops Science, College of Agronomy, Henan Agricultural University, Zhengzhou, 450002, China

[4]Key Laboratory of Horticultural Plant Biology (Ministry of Education), Huazhong Agricultural University, Wuhan, 430070, China

[*]Correspondence to Yingjie Xiao (yxiao25@mail.hzau.edu.cn)

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the Version of Record. Please cite this article as doi: 10.1111/TPJ.14856

Running title: Metabolism provides insights into maize and rice evolution.

Keywords: maize, rice, metabolomics, evolution, adaptation.

SUMMARY

Metabolites are the intermediate and final products of metabolism, which play essential roles in plant growth, evolution and adaptation to changing climates. However, it is unclear how evolution contributes to metabolic variation in plants. Here, we investigated the metabolomics data from leaf and seed tissues in maize and rice. By a principal component analysis (PCA) based on leaf metabolites but not seed metabolites, it was able to be clearly separated for rice *Indica* and *Japonica* accessions, while two maize subgroups, temperate and tropical, showed more visible admixture. Rice and maize seed exhibited significant interspecific differences in metabolic variation, while within rice, leaf and seed displayed similar metabolic variations. Among 10 metabolic categories, flavonoids had higher variation in maize than rice, indicating flavonoids are a key constituent of interspecific metabolic divergence. Interestingly, metabolic regulation was additionally found to be dramatically reshaped from positive to negative correlations, indicative of the differential evolutionary processes in maize and rice. Moreover, perhaps due to this divergence significantly more metabolic interactions were identified in rice than maize. Furthermore, in rice, the leaf was found to harbor much more intense metabolic interactions than the seed. Our result suggests that metabolomes are valuable for tracking evolutionary history, thereby complementing and extending genomic insights concerning which features are responsible for interspecific differentiation in maize and rice.

Introduction

Metabolites serves as the intermediate and ultimate products of biological processes, being critical in plant growth, evolution and adaptation to changing climate conditions (Carreno-Quintero et al., 2013). In addition, metabolites are indispensable for human nutrition, energy and medicines (De Luca et al., 2012; Saito & Matsuda, 2010; Keurentjes, 2009). The metabolome represents the entire set of metabolites from a cell, tissue, organ or organism at a particular developmental or physiological stage (Carreno-Quintero et al., 2013). As such metabolomics aims to identify and quantify metabolites in an organism for investigating their dynamics, compositions and interactions with internal and external environments (Orešič, 2009). Moreover, by contrast to analyses such as genomics, transcriptomics, and proteomics which are largely a function of the genetic blueprint of an organism metabolomics focuses on investigating downstream biological processes that are closer to the end-phenotype (Keurentjes, 2009). Therefore, metabolomics analysis could provide an alternative to understand the evolutionary and breeding processes in plants.

Recently, with fast advancements in metabolomics technologies, metabolomics analysis methods have become an important platform benefiting for functional genomics and systems biology (Raamsdonk et al., 2001; Saito & Matsuda, 2010), complementary to transcriptomic and proteomic analyses (Hollywood et al., 2006; Weckwerth & Morgenthal, 2005). With the development of high-throughput and low-cost genotyping methods, most association and linkage mapping studies have been used to dissect the genetic mechanisms of complex traits in plants. A recent study combining genome-wide association analysis with metabolomics proved this to be a powerful approach for dissecting the genetic and biochemical basis of plant metabolism. Integrated multi-level "omics" analysis is also an effective tool in identifying critical functional genes underlying metabolic pathways (Chen et al., 2014; Wen et al., 2014; Wen et al., 2015). Moreover, studies of the chemical structures and annotations of newly identified metabolites are capable of not only uncovering novel metabolic pathways but also of improving the understanding of previously identified metabolic pathways (Gong et al., 2013; Wen et al., 2014; Chen et al.,

2014). Additionally, identifying the loci encoding known metabolites could facilitate gene annotation (Wen et al., 2014; Peng et al., 2016; May et al., 2008). Metabolites can also be used as biomarkers of for example 100-kernel weight, which would facilitate the cloning of QTL and the identification of the genetic architecture underlying such traits, thereby enhancing crop breeding (Wen et al., 2014; Chen et al., 2016).

Maize (*Zea mays* L.) and rice (*Oryza sativa* L.) are two important crops in the grass family. These crops are not only staple foods for humans and animals, but also serve as highly important genetic materials. Maize was domesticated from teosinte (*Zea mays* ssp. *parviglumis*) approximately 9,000 years ago in the mid- to lowland tropical growth environment of southwestern Mexico. Domestication of maize is thought to have occurred via a single event (Matsuoka et al., 2002; van Heerwaarden et al., 2011). The separation of maize into tropical and temperate lines apparently occurred approximately 3,400-6,700 years ago (Liu et al., 2015). Similarly, rice is thought to have been domesticated from wild rice (*Oryza rufipogon*) ~10,000 years ago (Molina et al., 2011; Gross & Zhao, 2014), and *Indica* rice was cultivated in the Ganges in India ~4000 years ago (Fuller et al., 2010; Fuller, 2011). However, whether *Indica* and *Japonica* have single or multiple origins has been the subject of several recent genetic and archaeological analyses and remains a source of some debate (Fuller & Sato, 2008; Civan et al., 2015; Huang & Han, 2015). The large number of plant metabolites, which is estimated to exceed 200,000 in nature (Fiehn, 2001; Fiehn, 2002), show remarkable biological variability in terms of abundance (Morgenthal et al., 2006). Therefore, the analysis of metabolomes would be a new angle to investigate the origin of *Indica* and *Japonica* rice.

Metabolomic studies have been reported in many species including maize (Riedelsheimer et al., 2012; Riedelsheimer et al., 2013; Rao et al., 2014; Wen et al., 2014; Obata et al., 2015; Wen et al., 2015; Li et al., 2019; Zhou et al., 2019; Xu et al., 2019), rice (Matsuda et al., 2012; Gong et al., 2013; Chen et al., 2014; Matsuda et al., 2015; Peng et al., 2016; Valette et al., 2019), tomato (*Solanum lycopersicum*) (Do et al., 2010; Toubiana et al., 2012; Sauvage et al., 2014; Alseekh et al., 2015; Toubiana et al., 2015; Li et al., 2019), *Arabidopsis thaliana* (Keurentjes et al., 2008; Lisec et al., 2008; Rowe et al., 2008; Chan et al., 2011; Joseph et al., 2013), loblolly pine (Eckert

et al., 2012), wheat (*Triticum aestivum*) (Graham et al., 2009; Warth et al., 2015) and soybean (*Glycine maximum*) (Lin et al., 2014). However, there have been very few reports concerning comparative metabolomics studies between maize and rice (Wang et al., 2014). As such how the metabolome was involved in the differential evolution of maize and rice remains elusive.

Three metabolic datasets were obtained in several previous studies, including the data for mature seed in a maize association panel (Wen et al., 2014), and the metabolic data of leaf (Chen et al., 2014) and seed (Chen et al., 2016) of a rice association panel. In the present study, we intend to explore: (i) the metabolic basis underlying the differential evolution of maize and rice; (ii) the spatio-temporal- metabolomics between rice tissues and (iii) evolutionary changes in the interactive metabolomics networks of maize and rice.

Results

Metabolome-based population differentiations in maize and rice

Three metabolomics datasets of maize and rice were used in the present study. In the maize dataset, a high-throughput liquid chromatography tandem mass spectrometry (LC-MS/MS) analysis had previously detected 748 and 735 metabolites in two environments, respectively (Wen et al., 2014). There were 694 metabolites commonly detected in both environments, and 54 and 41 metabolites which were specifically detected within a single environment. In the rice dataset, both leaf and seed tissues were used in the liquid chromatography-electrospray ionization tandem mass spectrometry (LC-ESI-MS/MS) analysis, for which 840 and 837 metabolites, were detected respectively (Chen et al., 2014 and Chen et al., 2016).

After filtering metabolic data to remove those with high missing value rate and low correlation between different environments (for details see Experimental Procedures), we obtained 576, 645 and 587 distinct metabolites in maize seed, rice seed and rice leaf, respectively, and the best linear unbiased prediction (BLUP) values of each metabolite were used in the following analyses (Supplemental Table 1-3). The details of metabolic distribution within the populations are provided in Figure S1. Then, ten classes of metabolites were identified as described previously

(Chen et al., 2016). Most of the metabolites are unknown in both maize and rice (Figure S2), leaving 103, 252 and 238 metabolites which could be fully annotated in maize seed, rice seed and rice leaf, respectively. Among which, 35% (36), 45% (115), and 52% (125) are flavonoids in maize seed, rice seed, and rice leaf, respectively (Figure 1A).

Based on metabolic variation within population, we performed a principal component analysis (PCA) of the three datasets. We can see that, the rice subspecies, *Indica* and *Japonica* accessions, can apparently be separated by the top principal components when analyzing leaf metabolite data, but not seed metabolite data; however tropical and temperate maize lines where not fully separable from one another (Figure 1B). In order to figure out which specific metabolic category contributes to the population differentiation in maize and rice, the PCA were re-performed using only data from the metabolites with known annotation. In rice seed, flavonoid related metabolites could clearly discriminate *Indica* and *Japonica* accessions. A similar pattern was observed in rice leaf, but more annotation categories (flavonoids, phenolamides, polyphenols and others) appeared to be responsible for the differentiation of *Indica* and *Japonica* rice in this instance (Figure 1B and Figure S3). That said, none of metabolite annotations seemed to have high discriminating power for tropical and temperate maize, with the results for individual compound classes that proved anything a little bit worse than that the result using all metabolites (Figure 1B). These observations suggest that the key metabolites, such as flavonoids and phenolamide, probably played vital roles in the domestications of maize and rice. However, considerable further research will be needed in order to verify if that is indeed the case.

Dramatic metabolic differentiation between maize and rice

Rice and maize differentiated to form independent species tens of millions of years ago (Xu et al., 2008; Wang et al., 2011). Afterwards, the intraspecific evolutionary process includes the post-domestication adaptation from tropical to temperate climes in maize, and the subspecies differentiation between *Indica* and *Japonica* in rice. To investigate the metabolic changes that occurred in parallel to the differentiation between rice and maize, we focused on the metabolic data for rice and maize collected in seed tissue, the key organ in diverse crop domestications.

Within a species, a metabolite that exhibited significantly different mean levels between subpopulations (p<0.01) was regarded to be involved in its intraspecific evolution. It was found that rice subspecies differentiation was accompanied by changes in 43.5% of seed metabolites associated with amino acid derivatives exhibiting significantly different mean values between *Indica* and *Japonica*, significantly higher than the ability of the post-domestication adaptation that resulted in changes in the levels of only 14.3% of metabolites between temperate and tropical in maize (p=6.74×10^{-4}). Intriguingly, about 33.3% of polyphenols-associated metabolites appeared to be changed between *Indica* and *Japonica* rice, significantly higher than maize that observing none between temperate and tropical maize (p<2.20×10^{-16}). Overall, we found that on average 39.8% (0.0% to 71.4%) of metabolites were significantly altered in level between temperate and tropical maize, by contrast, 56.0% (33.3% to 83.3%) of metabolites between *Indica* and *Japonica* (Figure 2A, Supplementary Table 4 and 5). The result suggest considerable evolutionary differences exist between rice and maize and furthermore that several key metabolites may be involved.

The metabolic variation within populations was next estimated using the coefficient of variation (CV) of each metabolite in rice and maize population, respectively. Globally, rice metabolic variation ranged from 0.01 to 4.17, significantly higher than that found in maize (from 0.04 to 3.85, p=0.011) (Figure S4, Supplementary Table 7 and 8). According to a compound class analysis, it was found that maize had significantly higher metabolic variation of flavonoids than rice (p= 1.67×10^{-6}; Figure 2B). It is worth noting, however, that the metabolic variation of amino acid derivatives was also marginally higher in maize than rice (p=0.08) (Figure 2B). This is surprisingly in contrast to the trend that rice preserved a higher proportion of difference in the means of amino acid related metabolites between its subpopulations than maize (Figure 2A). That said, between subpopulations within species, the coefficient of variation for any metabolite seemed not be significantly different between temperate and tropical maize, and between *Indica* and *Japonica* rice (Figure S5).

Spatio-temporal metabolic variation in rice

In order to understand how spatio-temporal factors could impact on metabolic variation, we

compared the metabolic data collected in seed and leaf tissue of a rice population. It was found that nearly all metabolite categories showed higher proportion of differential-metabolites between *Indica* and *Japonica* in rice leaf than seed, particularly for flavonoids (p=1.67×10^{-12}), phenolamides (p=7.72×10^{-4}), and polyphenols (p=4.66×10^{-5}) (Figure 3A, Supplementary Table 5 and 6). There was only one exception that, the vitamin category appeared to contain ~40% differential metabolites between *Indica* and *Japonica* in leaf, which was marginally lower than that (~50%) in seed. Moreover, on average 73.8% (from 40.0% to 90.4%) metabolites showed significant difference between *Indica* and *Japonica* in leaf, whilst 56.0% (from 33.3% to 83.3%) metabolites significant difference between *Indica* and *Japonica* in seed (Figure 3A, Supplementary Table 5 and 6). That said, the leaf tissue displayed metabolic variation of between 0.02 and 3.58, which was similar to that of the seed (between 0.01 and 4.17; p=0.80) (Figure S4, Supplementary Table 8 and 9). The amino acid and nuclei acid derivatives appeared to be more diverse in seeds than that in leaves (Figure 3B). Furthermore, when assessed by metabolite category, there was no significant difference between *Indica* and *Japonica* leaves (Figure S5).

Two-dimensional metabolic variation in maize and rice

To investigate metabolic networks between maize and rice, we computed metabolite-metabolite correlations according to metabolic categories. We filtered out the minor correlations in order to alleviate the noise due to chance, using the 1000 permutation-based thresholds for different datasets (Supplementary Table 10). In maize seed tissue ($|r|\geq$ 0.174, $p\leq$1.92×10^{-3}), there were 653 pairs of positive correlations and 68 negative correlations in the whole population, while 337 and 315 positive correlations, twelve and six negative correlations were detected in temperate ($|r|\geq$ 0.307, $p\leq$1.42×10^{-3}) and tropical ($|r|\geq$ 0.309, $p\leq$3.28×10^{-4}) lines, respectively. In rice seed tissue ($|r|\geq$ 0.126, $p\leq$4.47×10^{-3}), we detected a total of 5,474 positive and 741 negative correlations in the whole population, while 3,603 and 2,716 positive correlations, 371 and 84 negative correlations were detected when the *Indica* ($|r|\geq$ 0.176, $p\leq$3.22×10^{-3}) and *Japonica* ($|r|\geq$ 0.255, $P\leq$1.55×10^{-3}) accessions, respectively, were assessed independently. In rice leaf tissue ($|r|\geq$ 0.132, $p\leq$2.15×10^{-3}), a total of 9,504 positive and 2,123 negative correlations were

identified in all rice accessions, whilst 8,861 and 4,477 positive correlations, 389 and 37 negative correlations in *Indica* ($|r| \geq 0.185$, $p \leq 1.51 \times 10^{-3}$) and *Japonica* ($|r| \geq 0.310$, $p \leq 8.97 \times 10^{-5}$) accessions, respectively. Then, the metabolic networks for maize and rice were built based on the pairwise correlation relationship (Figure S6, Supplementary Table 11-13). Interestingly, upon different metabolic data, the topology of metabolic networks appeared to be dramatically reshaped response to different spatio-temporal developmental stages and evolutionary processes.

It was found that rice metabolites displayed more intense interactions than maize, while within rice, the metabolites in leaf exhibited higher correlations than those in seed (Figure 4 and Figure S7). By comparing inter- and intra-category, it was found that the intra-category metabolic correlations were significantly higher than that inter-category ones in maize seed and rice leaf tissues (Figure S8). However, in the rice seed tissue, the difference between inter- and intra-category was largely reduced, especially for *Japonica* accessions, the intra-category only showed marginal higher metabolic correlations than the inter-category (p=0.02; Figure S8).

Furthermore, it was found that the maize metabolome exhibited significant higher correlations than that of rice in multiple metabolic categories including amino acids, flavonoids, amino acid derivatives, lipids and nucleic acid derivatives (Figure 5A). Intriguingly, in rice, the leaf tissue appeared to be more extreme in that it consistently exhibited higher metabolic correlations than the seed tissue in all categories (Figure 5B). By comparing two subgroups within species, we found that the majority of metabolic categories did not display much difference with regard to the metabolic correlations between temperate and tropical maize lines, except for amino acids that the mean pairwise metabolic correlations appeared to be considerably decreased between tropical and temperate maize ($p=7.70 \times 10^{-3}$; Figure S9A). A striking phenomenon was observed in rice. In the seed tissue, the metabolic correlations from amino acid derivatives, flavonoids and lipids were found to be significantly lower in *Indica* than that in *Japonica* accessions ($p<0.05$; Figure S9B), however, the leaf tissue revealed a reverse metabolic pattern to the seed, in that the *Indica* accessions exhibited significantly stronger metabolic correlations in amino acids and flavonoids than the *Japonica* accessions (Figure S9C).

Discussion

It is of great interest to understand how crop plants evolved to adapt environmental resilience upon domestication (Hu et al., 2014). Currently, the history of crop domestications had been well explored based on genomic variation, plant morphology and even growth habits (Xu et al., 2008, Liu et al., 2015, Purugganan et al., 2009, Meyer et al., 2012, Doebley et al., 2006,). Nevertheless, the knowledge regarding how metabolomics is involved in plant evolutionary process remains ambiguous. In the present study, we provided metabolomics clues to better interpret the diverse evolution between rice and maize.

Metabolic insights on evolution of maize and rice

Maize was domesticated from *Zea mays* ssp. *parviglumis* in the mid- to lowland tropical growth environment of southwestern Mexico based on a single domestication event (Matsuoka et al., 2002; van Heerwaarden et al., 2011). In the current study, cluster analysis of maize based on metabolomics data separated the maize lines into two major subgroups, congruent to temperate and tropical origins. We did however still find a proportion of temperate lines were clustered close to the tropical groups, or *vice versa* (Figure S10A), which could be explained by their admixed origins or recent gene introgressions from tropical germplasms for improving the agronomic and resistance ability of the temperate inbred lines (Yang et al., 2011; Wu et al., 2016). This result suggests that metabolomics data are somewhat valuable to discriminate the majority of the tropical and temperate lines, but as yet not to the extent that be done with the rice subspecies. The relevant metabolites might be associated to plant morphology, yield, and resistance to biotic and abiotic stress, which would probably be the targets selected by breeders and ancient farmers alike. The QTL of some metabolic and agronomic traits were reported to be co-localized in maize, and the use of metabolites as the intermediate traits or biomarkers could provide new insights into the genetic improvement of crops (Wen et al., 2014).

Rice was domesticated from *Oryza rufipogon* approximately 10,000 year ago (Molina et al., 2011; Gross & Zhao, 2014). However, many questions of rice domestication remain elusive, such as its geographic origin, direct wild progenitor, and whether the two subspecies of cultivated rice

arose from a single or multiple domestication event (Huang et al., 2012; Fuller & Sato, 2008; Civan et al., 2015; Huang & Han, 2015). Our cluster analysis revealed that metabolomics could be used to distinguish between rice accessions with different origins (Figure S10B). Moreover, the results of cluster analysis seemed to support the occurrence of a single domestication event, because the two major groups both contain *Indica* and *Japonica* subgroups, which implied the likelihood is very low that rice simultaneously separated into *Indica* and *Japonica* at multiple domestication events. Therefore, metabolomics analysis could provide alternative angles to, perhaps even more efficiently, explore the evolutionary history in crops.

Key metabolites contribute to maize and rice domestications

In this study, many metabolites exhibited significantly different contents and variations between temperate and tropical maize, and between *Japonica* and *Indica* rice (Figure 2, 3, Figure S6 and Supplementary Table 1-3, 7-9), which implied their participations in the domestication processes. In maize, tropical lines showed significantly greater variation in most metabolic categories than temperate lines, indicating that the maize post-domestication adaptation from tropical to temperate climates may involve broad series of metabolic regulations (Matsuoka et al., 2002; van Heerwaarden et al., 2011).

Flavonoids, a class of secondary metabolites containing a C6-C3-C6 carbon framework (Tohge et al., 2013), play vital roles in protecting the plant from harmful effects of UV irradiation in sexual reproduction, as well as serving as anti-pathogenic compounds (Zhao, 2015; Koes et al., 1994; Casas et al., 2016). Phenolamides, widespread across many plant species, play roles in plant defense against biotic stress (microorganisms, bacteria, viruses, fungi, and insects), abiotic stress (dehydration, salt stress, and UV irradiation) and floral induction and development (Dong et al., 2015). In the present study, we found many metabolites differed between subgroups within maize and rice (Supplementary Table 4-9), among which, the flavonoids were the most detected metabolites in all metabolic categories (Figure 1 and Figure S2). The flavonoid related metabolites showed very high proportion of metabolites in difference between subgroups in rice (45%) and maize (36%), which was congruent to previous reports (Chen et al., 2014; Chen et al., 2016). The PCA results revealed that the flavonoids and phenolamide-related metabolites had the high ability

to discriminate subgroups of maize or rice accessions (Figure 1B). Interestingly, the flavonoids were found to exhibit higher metabolic variation in seeds of the maize population than in those of the rice population (p=1.67 $\times$ 10^{-6}, Figure 2B). These results suggest that flavonoids and phenolamides are key metabolites contributing to differential evolution in rice and maize.

Metabolic networks are dynamically reshaped upon plant evolution

Many studies reported that the selection for genomic and transcriptomic regulations play fundamental roles in crop evolutions and breeding improvement (Hufford et al. 2012; Huang et al., 2012; Molina et al., 2011), even more important than the selection on genes or coding regions in maize (Liu et al., 2015). We built metabolic regulatory networks based on pairwise correlations in maize and rice. In the intraspecific layer, the tropical maize exhibited more positive metabolic correlations than the temperate maize, while the *Japonica* rice appeared to have more positive correlations than the *Indica* rice (Figure 4 and Figure S7). This demonstrates that the adaptation could drive the broadscale rewiring of regulatory relationships between metabolites. Intriguingly, this occurred by consistent reshaping of positive intraspecific regulatory relationships to negative ones in both maize and rice. In the interspecific layer, maize appeared to exhibit strong positive correlations between intra-category metabolites. By contrast, negative correlations seemed to be widespread not only between intra- but also inter-category metabolites in rice seed tissue (Figure 4 and Figure S7). The opposing direction of metabolic regulations in the two species needs further research but may be an indicator of the metabolic bases of interspecific evolution between rice and maize. Moreover, by comparing leaf to seed tissue in rice, we found that metabolic regulations were reshaped dramatically, recovering many positive correlations between intra-category metabolites in the rice leaf that were analogous to the patterns observed in the maize seed (Figure 4 and Figure S7). These results illustrate a dynamic landscape of metabolic variation in the interspecific and intraspecific evolution of maize and rice, indicating the neglected role of regulatory networks during artificial crop selection (Studer et al., 2017). Nevertheless, we realize that the conclusions of this study need to be cross-validated using independent population-based metabolomics data; and further efforts should focus on elucidating the mechanism underlying metabolomics evolution in rice and maize using integrated multi-omics methods.

Experimental procedures

Metabolic datasets

The metabolic data used in this study were obtained from a diverse maize association-mapping panel (Wen et al., 2014) and a rice association-mapping panel (Chen et al., 2014; Chen et al., 2016). The maize data, including 789 distinct metabolite features that were detected and quantified, from mature maize kernels of an association panel consisting of 368 diverse inbred lines from temperate, tropical and mixed subgroups (Wen et al., 2014). The chemical structures of 126 metabolites have been identified or annotated (Wen et al., 2014). In previous study, maize association panel was planted in three environments, the samples from two environments (Yunnan and Chongqing) were measured using the same equipment, and the third one (Hainan) used different equipment (Wen et al., 2014). To integrate multiple environmental metabolic data in a BLUP data, here we only used the data from Yunnan (N 24 25', E 102 30', referred to as E1) and Chongqing (E 106 50', N 29 25', referred to as E2) environments aiming to reduce systematic bias. Detailed information about field trials and LC-MS/MS-based metabolite profiling of mature kernels from the maize association panel was reported previously (Wen et al., 2014).

For rice, the data obtained from a diverse worldwide collection of 524 rice accessions, including both landraces and elite varieties, derived from four subgroups: *Indica*, *Japonica*, *Aus,* and *intermediate* rice (Chen et al., 2014). In brief, metabolic profiling analyses in rice were conducted at the five-leaf and mature grain stages, respectively, based on liquid chromatography–electrospray ionization–MS/MS, revealed 840 and 837 distinct metabolites were detected and quantified in the leaf (Chen et al., 2014) and seed (Chen et al., 2016) tissues, respectively. The chemical structures of 276 and 309 metabolites had been identified or annotated in leaf and seed tissues, respectively (Chen et al., 2014; Chen et al., 2016). For leaf metabolomics, two biological sample sets of rice association panel were extracted for leaf tissue in two different places at Huazhong Agricultural University, China (Wuhan, N 18 25', E 109 51'). Detailed information about field trials and metabolite profiling is provided in (Chen et al., 2014). For seed metabolomics, four sets of samples (two years, 2012 and 2013; and two replications per year) of

mature seeds per accession (Wuhan, N 18 25', E 109 51', referred to as 2012R1, 2012R2, 2013R1 and 2013R2) were extracted using an LC–electrospray ionization (ESI)-MS/MS system. Detailed information about field trials and metabolite profiling was provided previously (Chen et al., 2016).

Data analysis

Before analysis, metabolic data in maize and rice were filtered for quality control, using the criteria of <10% missing rate in population and significant correlation between multiple environments (p<0.05). In rice data, the combinations of year and repeat were treated as environments. For each retained metabolite, the best linear unbiased predictor (BLUP) value for each line across environments was used to reduce environmental noise, based on the mixed linear model implemented in the R package 'lme4' (www.r-project.org). Ultimately, in rice and maize data, the BLUP value for each line was used to be metabolic phenotype for following analyses.

The metabolites used in the present study was classified into annotation categories, following the criteria described previously (Chen et al., 2016). In brief, the amino acids category contains maize seed, rice seed and rice leaf amino acid. The amino acid derivatives category related amino acids from maize seed metabolome, amino acid derivatives from rice seed and leaf metabolomes. The lipids category is consisted of lysophosphatide, fatty acids and phosphorylcholine from maize seed metabolome, fatty acid from rice seed and leaf metabolomes. The class of nucleic acid derivatives contains the nucleoside and nucleotide from maize seed metabolome, nucleic acid derivative from rice leaf metabolome, and nucleic acid derivative and nucleic acid from rice seed metabolome. The category of flavonoids includes flavonoid, anthocyanin and phytoalexin from maize seed metabolome, anthocyanin and flavonoid from rice seed and leaf metabolomes. The phenolamides category is consisted of phenolamides from maize seed metabolome, phenolamine from rice leaf metabplome, phenolamine and polyamine from rice seed metabolome. The polyphenols class includes carboxylic acids and organic acid from maize seed metabolome, polyphenol from rice leaf and seed metabolomes. The category of vitamins includes vitamin from maize seed metabolome, vitamine from rice seed and leaf metabolomes and vitamine derivative from rice seed metabolome. The others category is consisted of alkaloid, polypeptide and

hormones from maize seed metabolome, others, terpene, carotene and alkaloid from rice leaf metabolome, others, phytohormone, alkaloid, terpenoid and saccharides from rice seed metabolome, and unknown category.

In maize and rice metabolomics data, principal component analysis (PCA) and the hierarchical cluster analysis were performed using SIMCA-P version 13.0 software with autofit model and scores standard (Wu et al., 2010); To understand intraspecific pattern in maize and rice, the maize association mapping panel can be divided into tropical, temperate and mixed origins, respectively, referred as the post-domestication adaptation (http://www.maizego.org/Resources.html), whilst the rice panel can be majorly divided into *Indica* and *Japonica* origins, respectively, referred as the subspecies differentiation (Chen et al., 2014). The information of population structure in rice and maize populations were detailed in Supplementary Data 1-3. For each metabolite category, the mean of metabolite content between temperate and tropical maize, and between *Indica* and *Japonica* rice was compared using student's t-test in excel 2013 ($p<0.01$). The proportion of significant metabolites for each category was estimated in rice kernel, rice leaf and maize kernel, respectively. The proportion of significant metabolites for a category was compared between rice kernel and maize kernel, the significance of different proportions indicated the metabolite category may be associated to distinct evolutions between rice and maize. The similar analysis was performed to compare the significant proportion for each metabolite category between rice leaf and seed tissues, reflecting the metabolic spatial-temporal difference. The proportion comparison was done by binomial test in the R function 'binom.test' (Dorai-Raj S. and Dorai-Raj M. S. 2009), for exemplifying between rice and maize kernel, by assuming the proportion in the rice kernel as the null proportion and then comparing it with the observed proportion in maize kernel. Moreover, the metabolic variation within population were measured by coefficient of variation (CV) to normalize all metabolites in different tissues and species. For each metabolite category, the difference of metabolite variation between tropical and temperate maize, between *Indica* and *Japonica* rice or between maize and rice, was tested based on student's t-test using the excel 2013 ($p<0.01$).

To explore two-dimensional variation in maize and rice, the Pearson correlation of pairwise inter-metabolites and intra-metabolites were calculated in subpopulations using the R function 'cor.test' (www.r-project.org). The heatmap showing metabolite relationships was draw using the R package 'pheatmap' (Kolde R. and Kolde M. R. 2015). The significant metabolite-metabolite correlations (p<0.001) were retained to construct metabolic network using the software Cytoscape version 3.2.0 (http://www.cytoscape.org/). To summarize the significant metabolic correlations, a 1,000-permutation test was implemented to determine the threshold of correlation coefficient (r) and p value. To do so, we reshuffled the lines randomly and obtained a permutated metabolic data. We then calculated the Pearson correlation coefficient and p value between pairwise metabolites by choosing one metabolite in original metabolic data and another in the permutated data. In each round of permutation, to reduce the impact of outliers, we recorded the 1th percentile small p and 1th percentile large absolute r values following the rules (Chen & Storey, 2006). This permutation process was repeated 1000 rounds, obtain a distribution of recorded absolute r and p values, respectively. The 99th percentile of absolute r distribution was defined to as the r threshold while the 1th percentile as the p threshold under the global 1% type-I error rate. Following this procedure, we obtained the threshold of r and p value for the maize and rice metabolism data, respectively. The permutation-based thresholds for all analyzed datasets were provided as Supplementary Table 10.

Author contributions

Y.X. and J.Y. designed the research. M.D., X.Z., J.L, H.L, W.W., H.L. participated in data analysis. M.D. and Y.X. wrote the manuscript.

Acknowledgements

This research was supported by the National Key Research and Development Program of China (2016YFD0101803), Hunan Agricultural University Talent Startup funding, the Science Fund for Young Scholars in Hunan Agricultural University (19QN15), the Fundamental Research Funds for the Central Universities (2662019QD050), the Natural Science Foundation of Hunan Province (2018JJ3234) and the fund from the Henan Provincial Higher Education Key Research Project (20A210003).

Conflict of interest

The authors declare no conflict of interest.

Figure Legend

Figure 1 Metabolomics composition and utility for inferring population structure. (A) The distribution of known metabolites. The know metabolites were classified into ten annotation categories. **(B) The principal component analysis of maize and rice populations.** Different sets of metabolites were used to evaluate population structure, which can be referred to be the temperate, tropical and mixed subgroups in maize; and the *Japonica*, *Indica*, *Aus* and *Intermediate* subgroups in rice.

Figure 2 Interspecific metabolomics between maize and rice. (A) The frequency of domestication-associated metabolites based on metabolic category. (B) The population variation of metabolites based on metabolic category. The metabolic data in seed tissue for maize and rice were used. The comparison for the domestication-associated metabolite between maize and rice was based on binomial test, while the comparison for population variation between maize and rice was based on the Student's-t test. $p<0.05$ was used as the significance threshold.

Figure 3 Spatial-temporal metabolomics between leaf and seed tissues in rice. (A) The frequency of domestication-related metabolites based on metabolic category. (B) The population variation of metabolites based on metabolic category. The metabolic data in leaf and seed tissues in rice were used. The comparison for the domestication-associated metabolite between leaf and seed was based on binomial test, while the comparison for population variation between leaf and seed was based on the Student's-t test. $p<0.05$ was used as the significance threshold.

Figure 4 The 2-D metabolomics for maize and rice evolutions. The heatmaps exhibited the metabolite-metabolite correlations between maize and rice, between temperate and tropical maize, and between *Japonica* and *Indica* rice.

Figure 5 The intra-category metabolic correlations in maize and rice. (A) Maize seed and rice seed. (B) Rice seed and leaf. The comparison of Pearson correlation coefficients of intra-category metabolites between maize and rice seed, and between rice leaf and seed, was based on the Student's-t test. $p<0.05$ was used as the significance threshold.

Supporting Information

Figure S1. The distribution of normalized metabolic values. (A) Maize seed. (B) Rice seed. (C) Rice leaf.

Figure S2. The distribution of all metabolites from maize and rice. (A) Maize seed. (B) Rice seed. (C) Rice leaf.

Figure S3 The PCA of maize and rice populations based on different sets of metabolites. (A-B) Maize seed. (C-I) Rice seed. (J-M) Rice leaf.

Figure S4. The distribution of coefficient of variation per metabolite in maize seed, rice seed and rice leaf, respectively.

Figure S5 The population variation of metabolites between subgroups based on metabolite category. (A) Temperate and tropical in maize seed. (B) *Japonica* and *Indica* in rice seed. (C) *Japonica* and *Indica* in rice leaf. The comparison between subgroups in maize and rice was based on the Student's-t test. $p<0.05$ was used as the significance threshold.

Figure S6 Regulatory networks based on metabolic correlations. (A) Maize seed. (B) Rice seed. (C) Rice leaf. Positive and negative correlations are represented by blue and green edges, respectively. Each color denotes a compound class as shown in the bottom left legend.

Figure S7 The distributions of intra-category metabolic correlations in maize and rice. The comparison between different metabolic data was based on Student's-t test. $p<0.05$ was used as the significance threshold.

Figure S8 The distributions of inter- and intra-category metabolic correlations within subgroups in maize and rice. The comparison between different metabolic data was based on Student's-t test. $p<0.05$ was used as the significance threshold.

Figure S9 The intra-category metabolic correlations for subgroups in maize and rice. (A) Maize seed. (B) Rice seed. (C) Rice leaf. The comparison of Pearson correlation coefficients of intra-category metabolites between subgroups was based on the Student's-t test. $p<0.05$ was used as the significance threshold.

Figure S10 Cluster analysis of maize and rice metabolomes. (A) Cluster analyses of 307 maize inbred lines with 576 metabolites. (B) Cluster analyses of 524 rice accessions with 645 metabolites. The clades indicate the accessions, colored corresponding to the subgroups in maize and rice.

Supplementary Table 1. The best linear unbiased predictor value for each metabolic trait in maize seed.

Supplementary Table 2. The best linear unbiased predictor value for each metabolic trait in rice seed.

Supplementary Table 3. The best linear unbiased predictor value for each metabolic trait in rice leaf.

Supplementary Table 4. The summary of annotated metabolites in maize seed and the adaptation related metabolites between temperate and tropical.

Supplementary Table 5. The summary of annotated metabolites in rice seed and the subspecies differentiation related metabolites between *Indica* and *Japonica*.

Supplementary Table 6. The summary of annotated metabolites in rice leaf and the subspecies differentiation related metabolites between *Indica* and *Japonica*.

Supplementary Table 7. The information of annotated metabolites and population variation in maize seed.

Supplementary Table 8. The information of annotated metabolites and population variation in rice seed.

Supplementary Table 9. The information of annotated metabolites and population variation in rice leaf.

Supplementary Table 10. The permutation-based thresholds for all analyzed datasets.

Supplementary Table 11. The Pearson correlation coefficient and *p* value between pairwise metabolites in maize seed.

Supplementary Table 12. The Pearson correlation coefficient and *p* value between pairwise metabolites in rice seed.

Supplementary Table 13. The Pearson correlation coefficient and *p* value between pairwise metabolites rice leaf.

References

Alseekh S., Tohge T., Wendenberg R., Scossa F., Omranian N., Li J., Kleessen S., Giavalisco P., Pleban T., & Mueller-Roeber B. (2015). Identification and mode of inheritance of quantitative trait loci for secondary metabolite abundance in tomato. *Plant cell, 27*(3), 485-512.

Carreno-Quintero N., Bouwmeester H.J., & Keurentjes J.J.B. (2013). Genetic analysis of metabolome–phenotype interactions: from model to crop species. *Trends Genet, 29*(1), 41-50.

Casas M.I., Falcone-Ferryra M.L., Jiang N., Mejia Guerra M.K., Rodriguez E.J., Wilson T., Engelmeier J., Casati P., & Grotewold E. (2016). Identification and characterization of maize salmon silks genes involved in insecticidal maysin biosynthesis. *Plant cell*. 28(6), 1297-1309.

Chan E.K.F., Rowe H.C., Corwin J.A., Joseph B., & Kliebenstein D.J. (2011). Combining genome-wide association mapping and transcriptional networks to identify novel genes controlling glucosinolates in *Arabidopsis thaliana*. *Plos Biol, 9*(8), 1713.

Chen L., & Storey J.D. (2006). Relaxed significance criteria for linkage analysis. *Genetics*, 173(4):2371-81.

Chen W., Gao Y.Q., Xie W.B., Gong L., Lu K., Wang W.S., Li Y., Liu X.Q., Zhang H.Y., Dong H.X., Zhang W., Zhang L.J., Yu S.B., Wang G.W., Lian X.M., & Luo J. (2014). Genome-wide association analyses provide genetic and biochemical insights into natural variation in rice metabolism. *Nat Genet, 46*(7), 714-721.

Chen W., Wang W., Peng M., Gong L., Gao Y., Wan J., Wang S., Shi L., Zhou B., Li Z., Peng X., Yang C., Qu L., Liu X., & Luo J. (2016). Comparative and parallel genome-wide association studies for metabolic and agronomic traits in cereals. *Nat commun, 7*(1), 1-10.

Civan P., Craig H., Cox C.J., & Brown T.A. (2015). Three geographically separate domestications of Asian rice. *Nat plants,* 1(11), 1-5.

Doebley, J. F., Gaut, B. S., & Smith, B. D. (2006). The molecular genetics of crop domestication. *Cell*, *127*(7), 1309-1321.

De Luca V., Salim V., Atsumi S.M., & Yu F. (2012). Mining the biodiversity of plants: a revolution in the making. *Science, 336*(6089), 1658-1661.

Do P.T., Prudent M., Sulpice R., Causse M., & Fernie A.R. (2010). The influence of fruit load on the tomato pericarp metabolome in a *Solanum chmielewskii* introgression line population. *Plant Physiol, 154*(3), 1128-1142.

Dong X., Gao Y., Chen W., Wang W., Gong L., Liu X., & Luo J. (2015). Spatiotemporal distribution of phenolamides and the genetics of natural variation of hydroxycinnamoyl spermidine in rice. *Mol Plant, 8*(1), 111-121.

Dorai-Raj, S., and Dorai-Raj, M. S. (2009). Package 'binom'. CRAN. *Google Scholar*.

Eckert A.J., Wegrzyn J.L., Cumbie W.P., Goldfarb B., Huber D.A., Tolstikov V., Fiehn O., & Neale D.B. (2012). Association genetics of the loblolly pine (*Pinus Taeda*, Pinaceae) metabolome. *New Phytol, 193*(4), 890-902.

Fiehn O. (2001). Combining genomics, metabolome analysis, and biochemical modelling to understand metabolic networks. *Comp Funct Genom, 2*(3), 155-168.

Fiehn O. (2002). Metabolomics – the link between genotypes and phenotypes. *Plant Mol Biol, 48*(1), 155-171.

Fuller D.Q. (2011). Finding plant domestication in the Indian subcontinent. *Current Anthropology, 52*(S4), S347-S362.

Fuller D.Q., & Sato Y.-I. (2008). *Japonica* rice carried to, not from, Southeast asia. *Nat Genet, 40*(11), 1264-1265.

Fuller D.Q., Sato Y.-I., Castillo C., Qin L., Weisskopf A.R., Kingwell-Banham E.J., Song J., Ahn S.-M., & Van Etten J. (2010). Consilience of genetics and archaeobotany in the entangled history of rice. *Archaeol Anthrop Sci, 2*(2), 115-131.

Gong L., Chen W., Gao Y., Liu X., Zhang H., Xu C., Yu S., Zhang Q., & Luo J. (2013). Genetic analysis of the metabolome exemplified using a rice population. *Proc. Natl. Acad. Sci. USA, 110*(50), 20320-20325.

Graham S.F., Amigues E., Migaud M., & Browne R.A. (2009). Application of nmr based metabolomics for mapping metabolite variation in European wheat. *Metabolomics, 5*(3), 302-306.

Gross B.L., & Zhao Z. (2014). Archaeological and genetic insights into the origins of domesticated rice. *Proc. Natl. Acad. Sci. USA, 111*(17), 6190-6197.

Hollywood K., Brison D.R., & Goodacre R. (2006). Metabolomics: current technologies and future trends. *Proteomics, 6*(17), 4716-4723.

Hu C., Shi J., Quan S., Cui B., Kleessen S., Nikoloski Z., Tohge T., Alexander D., Guo L., Lin H., Wang J., Cui X., Rao J., Luo Q., Zhao X., Fernie A.R., & Zhang D. (2014). Metabolic variation between Japonica and Indica rice cultivars as revealed by non-targeted metabolomics. *Sci Rep*, 4, 5067

Huang X., & Han B. (2015). Rice domestication occurred through single origin and multiple introgressions. *Nat plants,* 2(1), 1-1.

Huang X., Kurata N., Wei X., Wang Z.X., Wang A., Zhao Q., Zhao Y., Liu K., Lu H., Li W., Guo Y., Lu Y., Zhou C., Fan D., Weng Q., Zhu C., Huang T., Zhang L., Wang Y., Feng L., Furuumi H., Kubo T., Miyabayashi T., Yuan X., Xu Q., Dong G., Zhan Q., Li C., Fujiyama A., Toyoda A., Lu T., Feng Q., Qian Q., Li J., & Han B. (2012). A Map of rice genome variation reveals the origin of cultivated rice. *Nature, 490*(7421), 497-501.

Hufford, M. B., Xu, X., Van Heerwaarden, J., Pyhäjärvi, T., Chia, J. M., Cartwright, R. A., Elshire R. J., Glaubitz J. C., Guill K. E., Kaeppler S. M., Lai J., Morrell P. L., Shannon L. M., Chi Song C., Springer N. M., Swanson-Wagner R. A., Tiffin P., Wang J., Zhang G., Doebley J., McMullen M. D., Ware D., Buckler E. S., Yang S. & Ross-Ibarra J. (2012). Comparative population genomics of maize domestication and improvement. *Nat Genet*, 44(7), 808-811.

Joseph B., Corwin J.A., Li B., Atwell S., & Kliebenstein D.J. (2013). Cytoplasmic genetic variation and extensive cytonuclear interactions influence natural variation in the metabolome. *Elife, 2*, e00776.

Keurentjes J.J., Sulpice R., Gibon Y., Steinhauser M.C., Fu J., Koornneef M., Stitt M., & 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*(8), R129.

Keurentjes J.J.B. (2009). Genetical metabolomics: closing in on phenotypes. *Curr Opin Plant Biol, 12*(2), 223-230.

Kolde, R., and Kolde, M. R. (2015). Package 'pheatmap'. *R Package*, 1(7).

Koes R.E., Quattrocchio F., & Mol J.N.M. (1994). The flavonoid biosynthetic pathway in plants: function and evolution. *BioEssays, 16*(2), 123-132.

Li K., Wen W.W., Alseekh S., Yang X.H., Guo H., Li W.Q., Wang L.X., Pan Q.C., Zhan W., Liu J., Li Y.H., Wu X., Brotman Y., Willmitzer L., Li J.S., Fernie A.R., & Yan J.B. (2019). Large-scale metabolite quantitative trait locus analysis provides new insights for high-quality maize improvement. *Plant J*, 99(2), 216-230.

Li, Y., Lu, Y., Li, L., Chu, Z., Zhang, H., Li, H., Fernie A.R. & Ouyang, B. (2019). Impairment of hormone pathways results in a general disturbance of fruit primary metabolism in tomato. *Food chem*, 274, 170-179.

Lin H., Rao J., Shi J., Hu C., Cheng F., Wilson Z.A., Zhang D., & Quan S. (2014). Seed metabolomic study reveals significant metabolite variations and correlations among different soybean cultivars. *J Integr Plant Biol, 56*(9), 826-836.

Lisec J., Meyer R.C., Steinfath M., Redestig H., Becher M., Witucka-Wall H., Fiehn O., Torjek O., Selbig J., Altmann T., & Willmitzer L. (2008). Identification of metabolic and biomass QTL in *Arabidopsis thaliana* in a parallel analysis of RIL and IL populations. *Plant J, 53*(6), 960-972.

Liu H., Wang X., Warburton Marilynl., Wen W., Jin M., Deng M., Liu J., Tong H., Pan Q., Yang X., & Yan J. (2015). Genomic, transcriptomic, and phenomic variation reveals the complex adaptation of modern maize breeding. *Mol plant, 8*(6), 871-884.

Matsuda F., Nakabayashi R., Yang Z., Okazaki Y., Yonemaru J.-I., Ebana K., Yano M., & Saito K. (2015). Metabolome-genome-wide association study dissects genetic architecture for generating natural variation in rice secondary metabolism. *Plant J, 81*(1), 13-23.

Matsuda F., Okazaki Y., Oikawa A., Kusano M., Nakabayashi R., Kikuchi J., Yonemaru J., Ebana K., Yano M., & Saito K. (2012). Dissection of genotype-phenotype associations in rice grains using metabolome quantitative trait loci analysis. *Plant J, 70*(4), 624-636.

Matsuoka Y., Vigouroux Y., Goodman M.M., Sanchez G.J., Buckler E., & Doebley J. (2002). A single domestication for maize shown by multilocus microsatellite genotyping. *Proc. Natl. Acad. Sci. USA, 99*(9), 6080-6084.

May P., Wienkoop S., Kempa S., Usadel B., Christian N., Rupprecht J., Weiss J., Recuenco-Munoz L., Ebenhöh O., Weckwerth W., & Walther D. (2008). Metabolomics- and proteomics-assisted genome

annotation and analysis of the draft metabolic network of *Chlamydomonas Reinhardtii*. *Genetics, 179*(1), 157-166.

Meyer R. S., DuVal A. E., & Jensen H. R. (2012). Patterns and processes in crop domestication: an historical review and quantitative analysis of 203 global food crops. *New Phytologist*, *196*(1), 29-48.

Molina J., Sikora M., Garud N., Flowers J.M., Rubinstein S., Reynolds A., Huang P., Jackson S., Schaal B.A., Bustamante C.D., Boyko A.R., & Purugganan M.D. (2011). Molecular evidence for a single evolutionary origin of domesticated rice. *Proc. Natl. Acad. Sci. USA, 108*(20), 8351-8356.

Morgenthal K., Weckwerth W., & Steuer R. (2006). Metabolomic networks in plants: transitions from pattern recognition to biological interpretation. *Biosystems, 83*(2–3), 108-117.

Obata T., Witt S., Lisec J., Palacios-Rojas N., Florez-Sarasa I., Araus J.L., Cairns J.E., Yousfi S., & Fernie A.R. (2015). Metabolite profiles of maize leaves in drought, heat and combined stress field trials reveal the relationship between metabolism and grain yield. *Plant Physiol*, 169(4), 2665-2683.

Orešič M. (2009). Metabolomics, a novel tool for studies of nutrition, metabolism and lipid dysfunction. *Nutr, Metab Cardiovas, 19*(11), 816-824.

Peng M., Gao Y., Chen W., Wang W., Shen S., Shi J., Wang C., Zhang Y., Zou L., Wang S., Wan J., Liu X., Gong L., & Luo J. (2016). Evolutionarily distinct BAHD N-acyltransferases are responsible for natural variation of aromatic amine conjugates in rice. *Plant cell, 28*(7), 1533-1550.

Purugganan M. D., & Fuller D. Q. (2009). The nature of selection during plant domestication. *Nature*, *457*(7231), 843.

Raamsdonk L.M., Teusink B., Broadhurst D., Zhang N., Hayes A., Walsh M.C., Berden J.A., Brindle K.M., Kell D.B., Rowland J.J., Westerhoff H.V., Van Dam K., & Oliver S.G. (2001). A functional genomics strategy that uses metabolome data to reveal the phenotype of silent mutations. *Nat Biotech, 19*(1), 45-50.

Rao J., Cheng F., Hu C., Quan S., Lin H., Wang J., Chen G., Zhao X., Alexander D., Guo L., Wang G., Lai J., Zhang D., & Shi J. (2014). Metabolic map of mature maize kernels. *Metabolomics, 10*(5), 775-787.

Riedelsheimer C., Brotman Y., Meret M., Melchinger A.E., & Willmitzer L. (2013). The maize leaf lipidome shows multilevel genetic control and high predictive value for agronomic traits. *Sci Rep, 3*, 2479.

Riedelsheimer C., Lisec J., Czedik-Eysenberg A., Sulpice R., Flis A., Grieder C., Altmann T., Stitt M., Willmitzer L., & Melchinger A.E. (2012). Genome-wide association mapping of leaf metabolic profiles for dissecting complex traits in maize. *Proc. Natl. Acad. Sci. USA, 109*(23), 8872-8877.

Rowe H.C., Hansen B.G., Halkier B.A., & Kliebenstein D.J. (2008). Biochemical networks and epistasis shape the *Arabidopsis thaliana* metabolome. *Plant cell, 20*(5), 1199-1216.

Saito K., & Matsuda F. (2010). Metabolomics for functional genomics, systems biology, and biotechnology. *Annu Rev Plant Biol, 61*, 463-489.

Sauvage C., Segura V., Bauchet G., Stevens R., Do P.T., Nikoloski Z., Fernie A.R., & Causse M. (2014). Genome-wide association in tomato reveals 44 candidate loci for fruit metabolic traits. *Plant Physiol, 165*(3), 1120-1132.

Studer, A. J., Wang, H., & Doebley, J. F. (2017). Selection during maize domestication targeted a gene network controlling plant and inflorescence architecture. *Genetics*, 207(2), 755-765.

Tohge T., Watanabe M., Hoefgen R., & Fernie A.R. (2013). The evolution of phenylpropanoid metabolism in the green lineage. *Crit Rev Biochem Mol, 48*(2), 123-152.

Toubiana D., Batushansky A., Tzfadia O., Scossa F., Khan A., Barak S., Zamir D., Fernie A.R., Nikoloski Z., & Fait A. (2015). Combined correlation-based network and mQTL analyses efficiently identified loci for branched-chain amino acid, serine to threonine, and proline metabolism in tomato seeds. *Plant J, 81*(1), 121-133.

Toubiana D., Semel Y., Tohge T., Beleggia R., Cattivelli L., Rosental L., Nikoloski Z., Zamir D., Fernie A.R., & Fait A. (2012). Metabolic profiling of a mapping population exposes new insights in the regulation of seed metabolism and seed, fruit, and plant relations. *PLoS Genet, 8*(3), e1002612.

Valette, M., Rey, M., Gerin, F., Comte, G., & Wisniewski-Dyé, F. (2020). A common metabolomic signature is observed upon inoculation of rice roots with various rhizobacteria. *J Integr Plant Biol*, 62(2), 228-246.

Van Heerwaarden J., Doebley J., Briggs W.H., Glaubitz J.C., Goodman M.M., De Jesus Sanchez Gonzalez J., & Ross-Ibarra J. (2011). Genetic signals of origin, spread, and introgression in a large sample of maize landraces. *Proc. Natl. Acad. Sci. USA, 108*(3), 1088-1092.

Wang L, Czedik-Eysenberg A, Mertz RA, Si Y, Tohge T, Nunes-Nesi A, Arrivault S, Dedow LK, Bryant DW, Zhou W, Xu J, Weissmann S, Studer A, Li P, Zhang C, Larue T, Shao Y, Ding Z, Sun Q, Patel RV, Turgeon R, Zhu XG, Provart NJ, Mockler TC, Fernie AR, Stitt M, Liu P, Brutnell TP. (2014) Comparative analyses of C_4 and C_3 photosynthesis in developing leaves of maize and rice. *Nat Biotechnol*, 32(11), 1158-1165.

Wang X, Tang H, Paterson A H. (2011). Seventy million years of concerted evolution of a homoeologous chromosome pair, in parallel, in major Poaceae lineages. *Plant Cell*, 23(1): 27-37.

Warth B., Parich A., Bueschl C., Schoefbeck D., Neumann N.K.N., Kluger B., Schuster K., Krska R., Adam G., Lemmens M., & Schuhmacher R. (2015). GC–MS based targeted metabolic profiling identifies changes in the wheat metabolome following deoxynivalenol treatment. *Metabolomics, 11*(3), 722-738.

Weckwerth W., & Morgenthal K. (2005). Metabolomics: from pattern recognition to biological interpretation. *Drug Discov Today, 10*(22), 1551-1558.

Wen W., Li K., Alseekh S., Omranian N., Zhao L., Zhou Y., Xiao Y., Jin M., Yang N., Liu H., Florian A., Li W., Pan Q., Nikoloski Z., Yan J., & Fernie A.R. (2015). Genetic determinants of the network of primary metabolism and their relationships to plant performance in a maize recombinant inbred line population. *Plant cell, 27*(7), 1839-1856.

Wen W.W., Li D., Li X., Gao Y.Q., Li W.Q., Li H.H., Liu J., Liu H.J., Chen W., Luo J., & Yan J.B. (2014). Metabolome-based genome-wide association study of maize kernel leads to novel biochemical insights. *Nat commun,* 5(1), 1-10.

Wu Y.S., San Vicente F., Huang K.J., Dhliwayo T., Costich D.E., Semagn K., Sudha N., Olsen M., Prasanna B.M., Zhang X.C. & Babu R. (2016). Molecular characterization of CIMMYT maize inbred lines with genotyping-by-sequencing SNPs. *Theor Appl Genet*, 129(4), 753-765.

Wu, Zaibin, Dapeng Li, Jie Meng, and Huiwen Wang. (2010) Introduction to SIMCA-P and its application. Handbook of Partial Least Squares. Springer, Berlin, Heidelberg, 757-774.

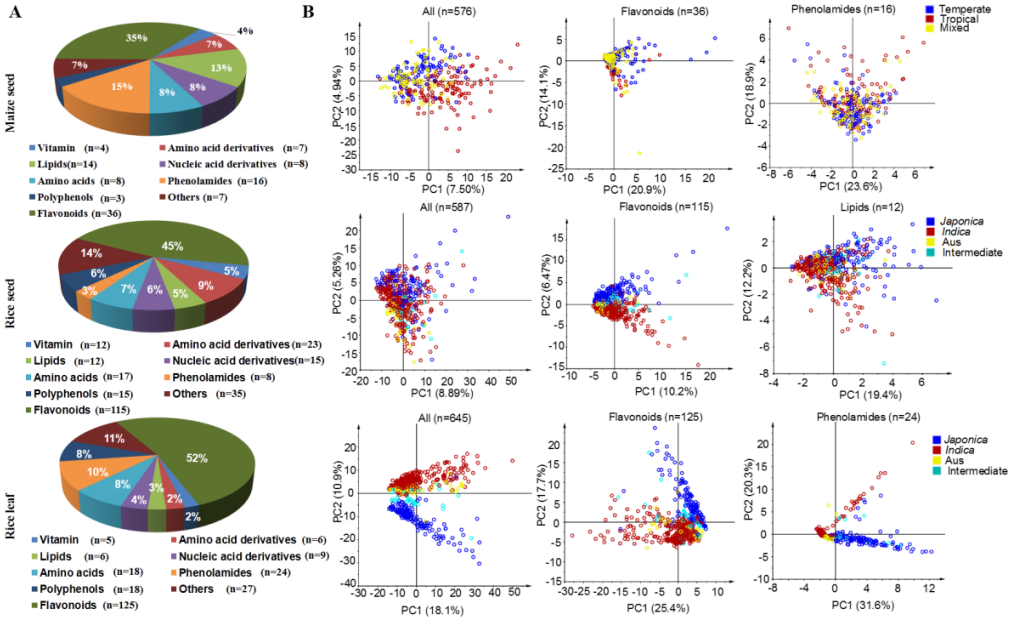
Xu, G., Cao, J., Wang, X., Chen, Q., Jin, W., Li, Z., & Tian, F. (2019). Evolutionary metabolomics identifies substantial metabolic divergence between maize and its wild ancestor, teosinte. *Plant Cell*, 31(9), 1990-2009.

Xu J H, Messing J. (2008). Organization of the prolamin gene family provides insight into the evolution of the maize genome and gene duplications in grass species. *Proc. Natl. Acad. Sci. USA*, 105(38): 14330-14335.

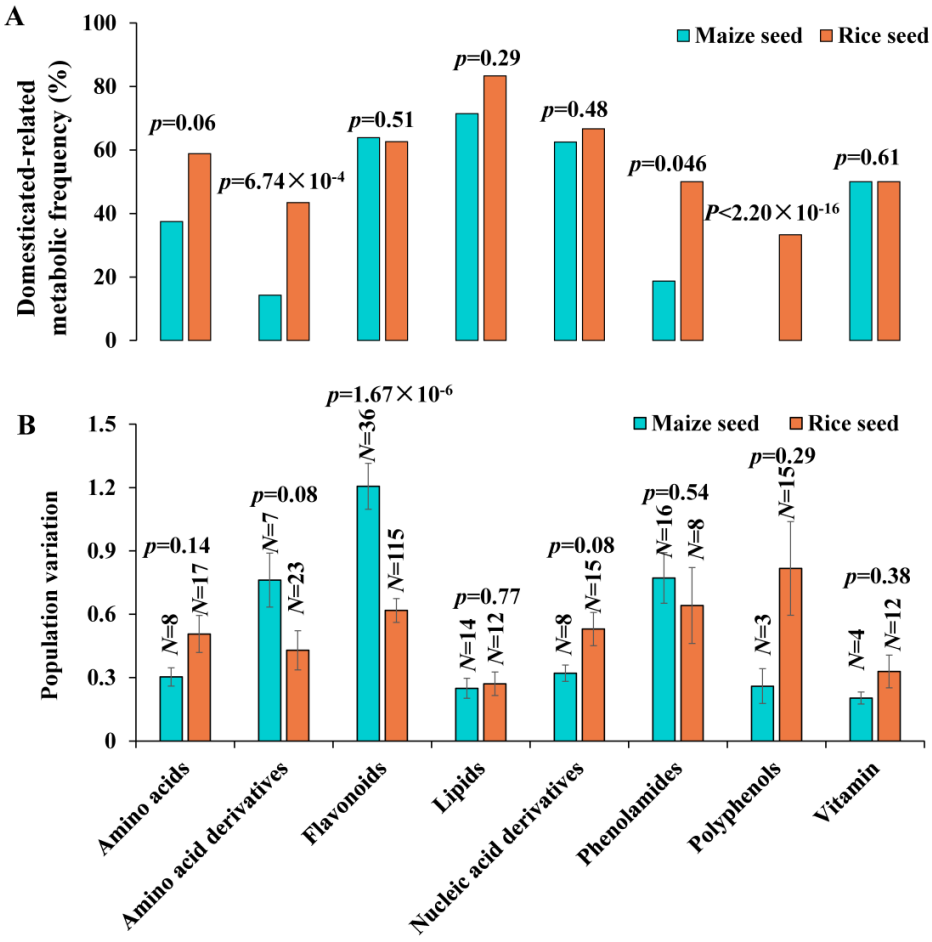
Yang X.H., Gao S.B., Xu S.T., Zhang Z.X., Prasanna B.M., Li L., Li J.S. & Yan J.B. (2011). Characterization of a global germplasm collection and its potential utilization for analysis of complex quantitative traits in maize. *Mol Breeding*, 28(4), 511-526.

Zhao J. (2015). Flavonoid transport mechanisms: how to go, and with whom. *Trends Plant Sci, 20*(9), 576-585.

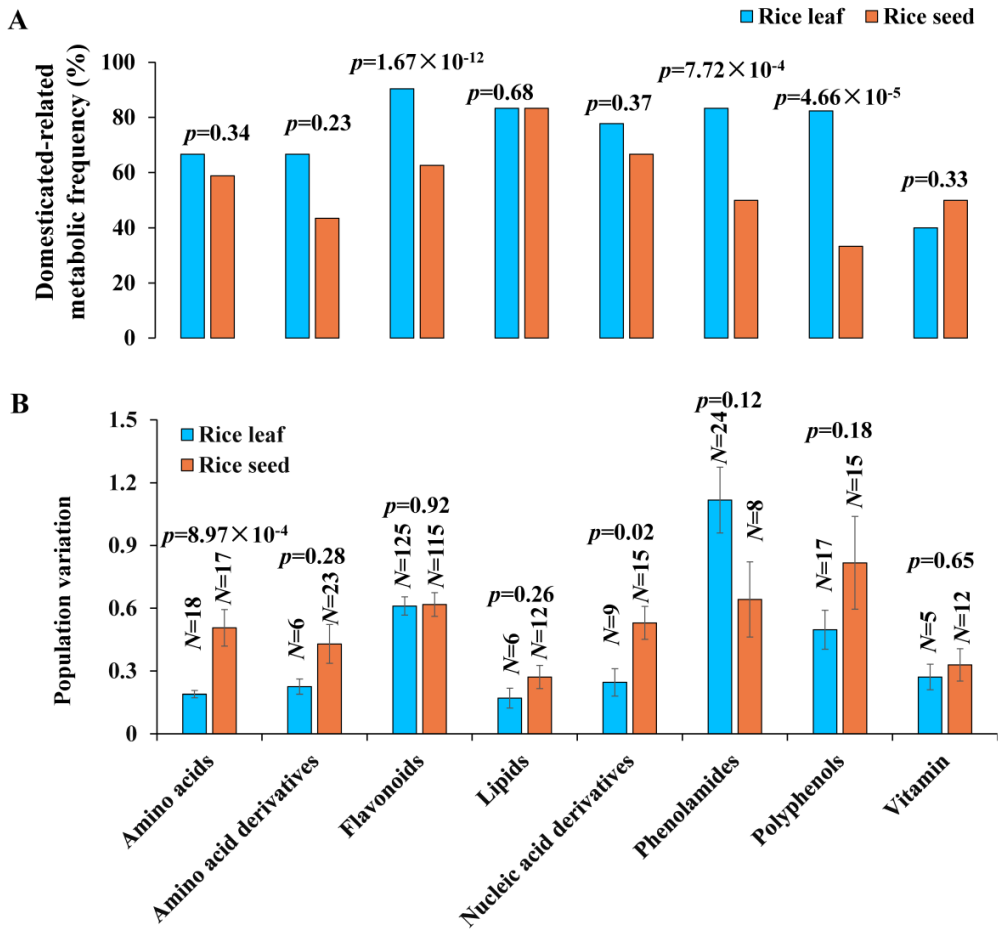
Zhou, S., Kremling, K.A., Bandillo, N., Richter, A., Zhang, Y.K., Ahern, K.R., Artyukhin A.B. Hui J.X., Younkin G.C., Schroeder F.C., Buckler, E.S. & Jander G. (2019). Metabolome-scale genome-wide association studies reveal chemical diversity and genetic control of maize specialized metabolites. *Plant Cell*, 31(5), 937-955.



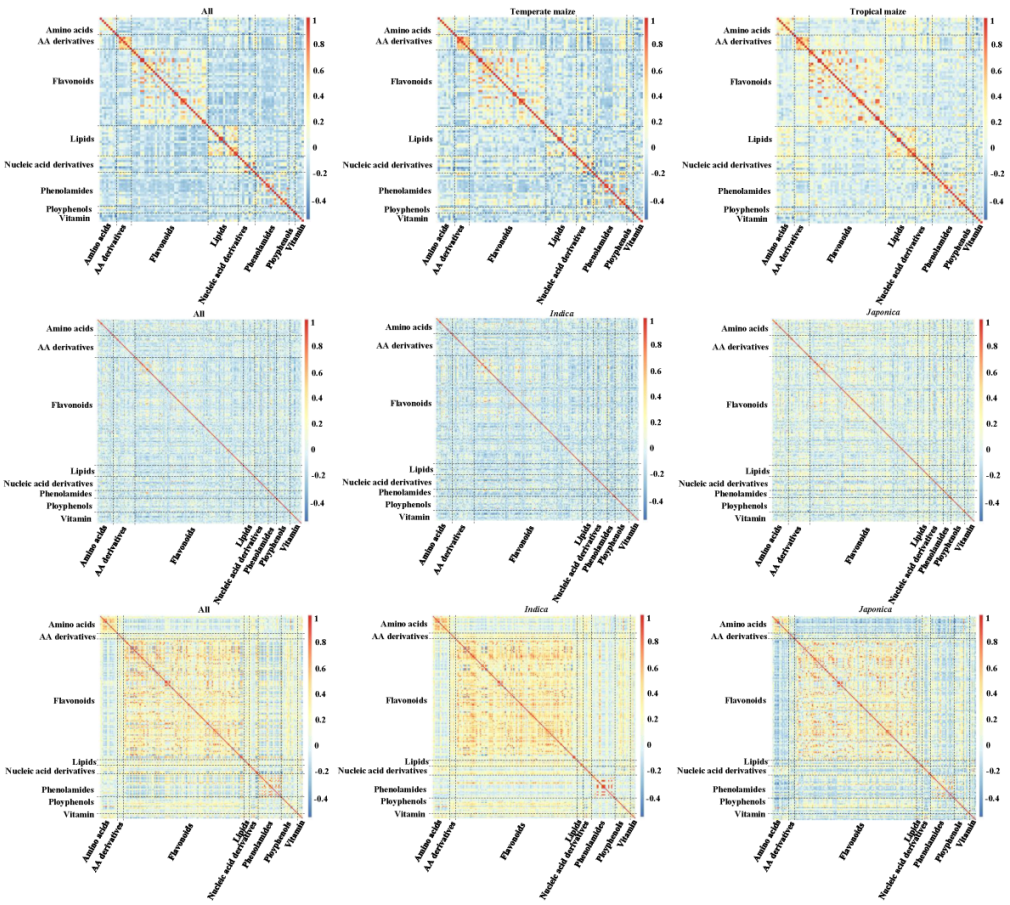
tpj_14856_f1.tif



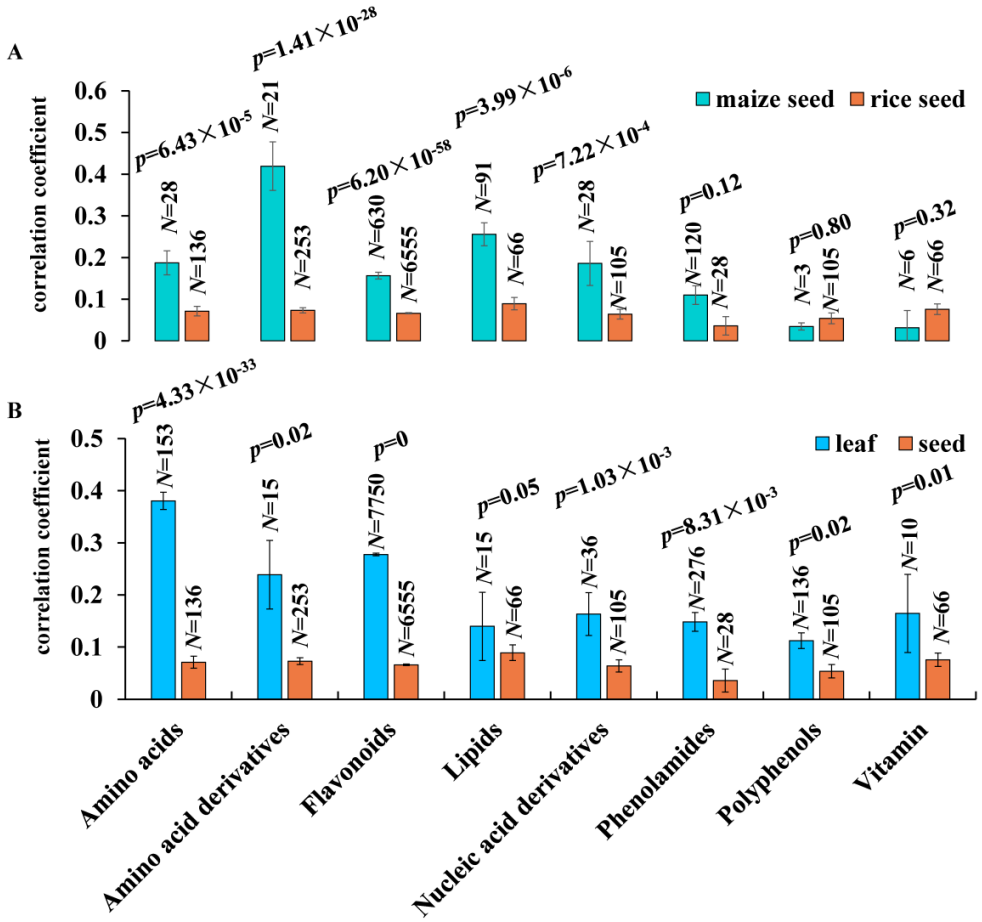
tpj_14856_f2.tif



tpj_14856_f3.tif



tpj_14856_f4.tif



tpj_14856_f5.tif