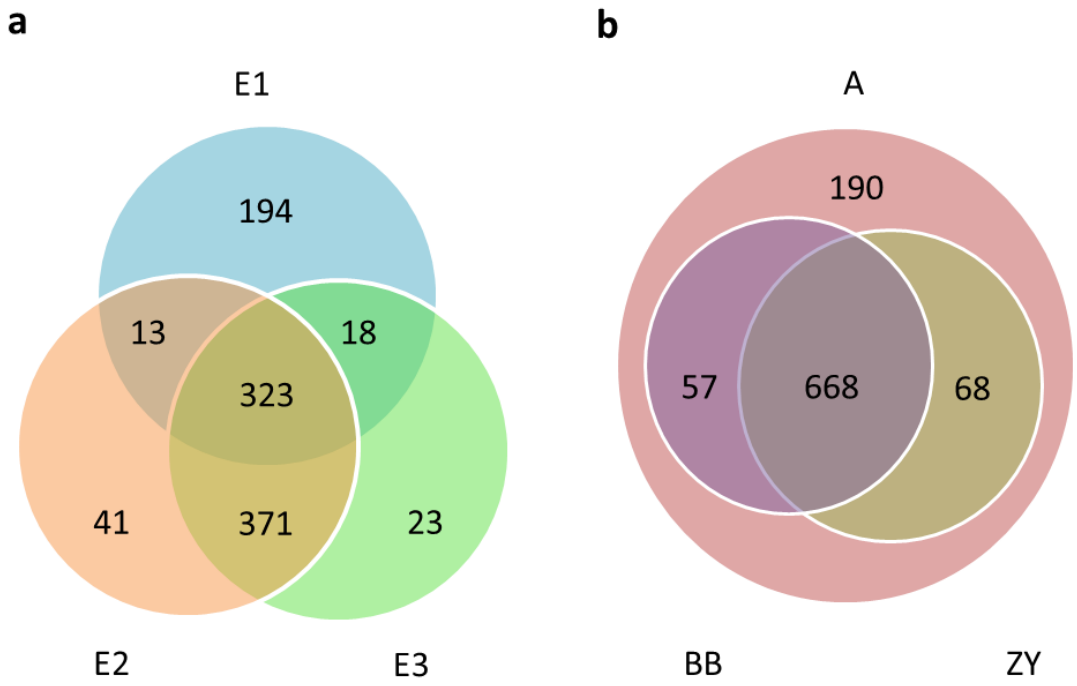
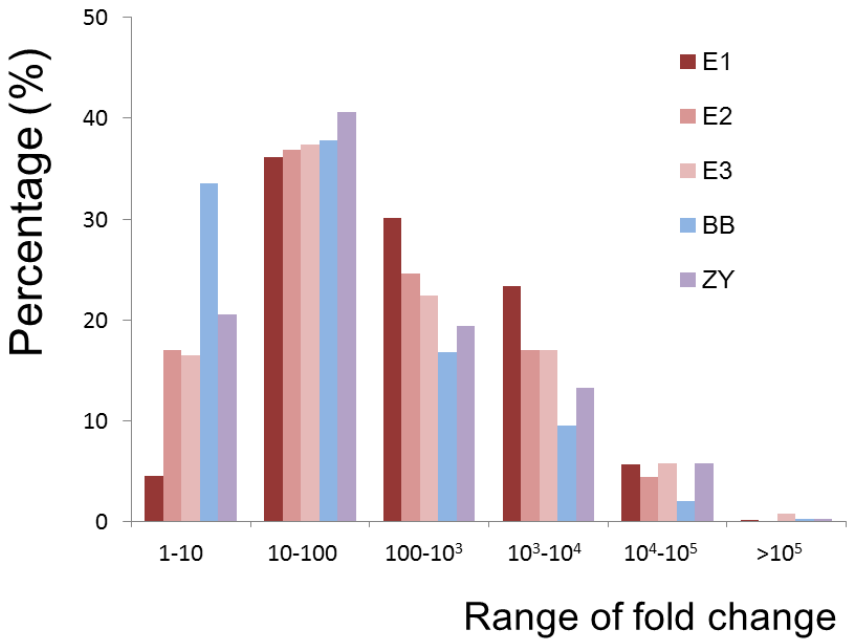
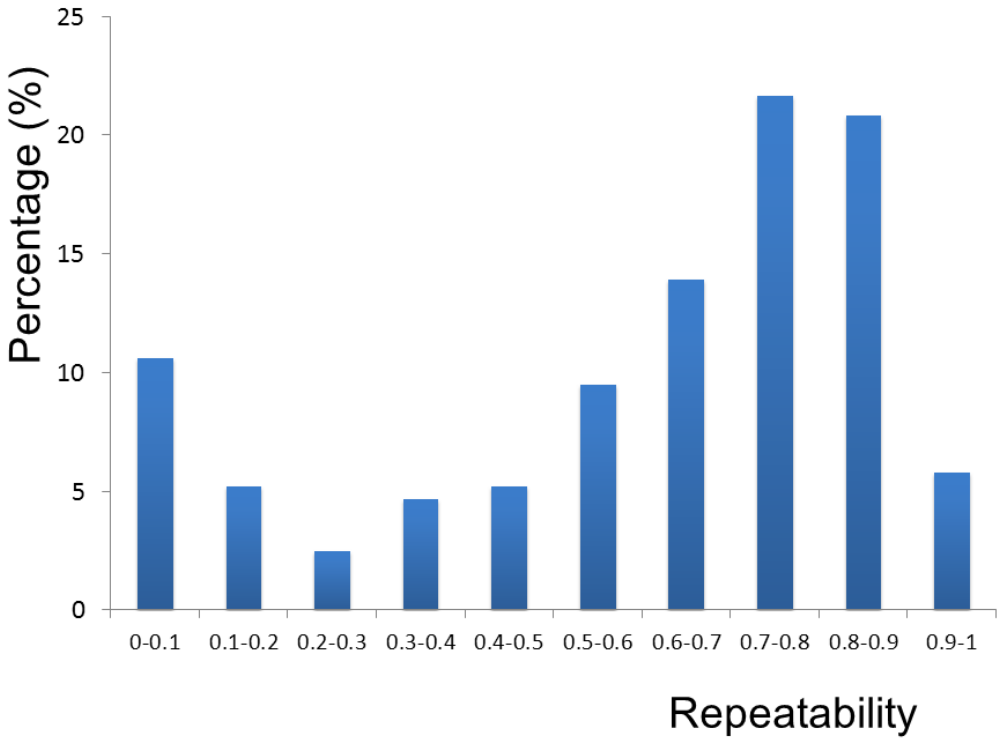
Supplementary Information



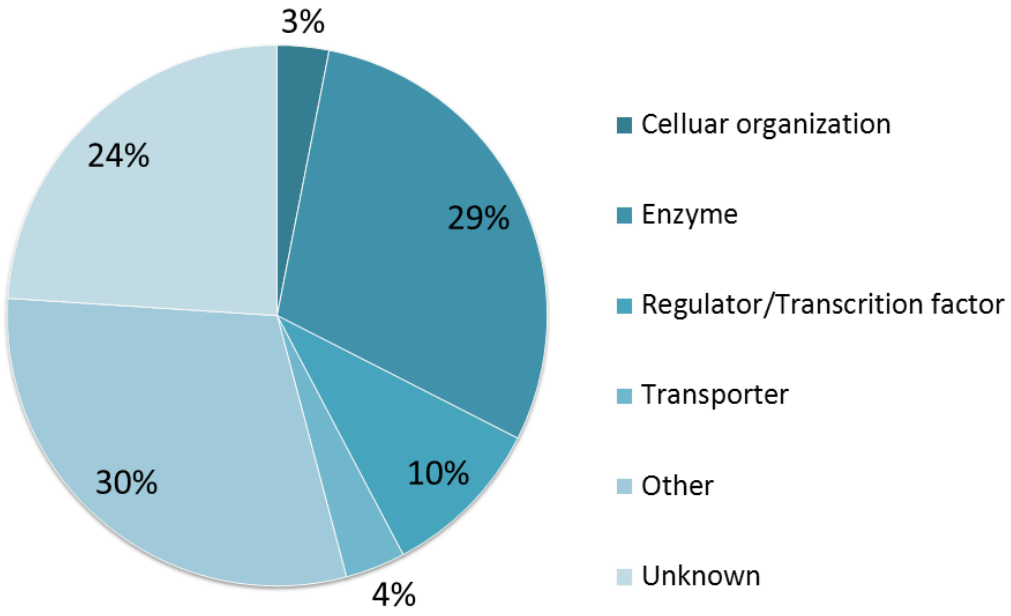
Supplementary Figure 1 Number of detected metabolite features. Venn diagram of the overlap between a) metabolite features detected in different experiments on the association panel and b) the overall metabolite features detected in the association panel and two RIL populations. A, the association panel; BB, B73/By804 RIL population; ZY, Zong3/Yu87-1 RIL population; E1, experiment 1 conducted on the association panel; E2, experiment 2 conducted on the association panel; E3, experiment 3 conducted on the association panel.



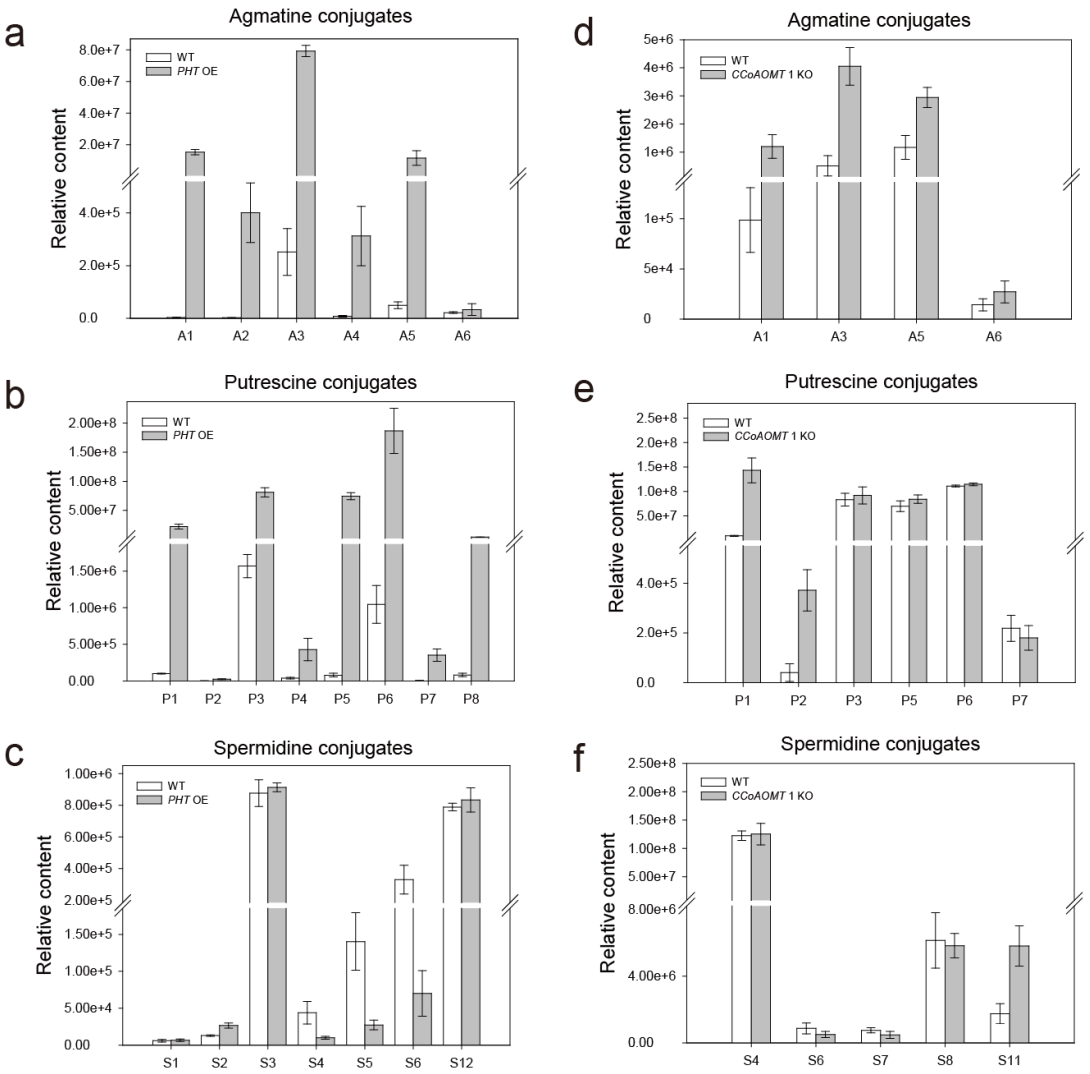
Supplementary Figure 2 Distribution of fold change of metabolite level measured in three experiments of the association panel and two linkage populations. E1, experiment 1 conducted on the association panel; E2, experiment 2 conducted on the association panel; E3, experiment 3 conducted on the association panel; BB, B73/By804 RIL population; ZY, Zong3/Yu87-1 RIL population.



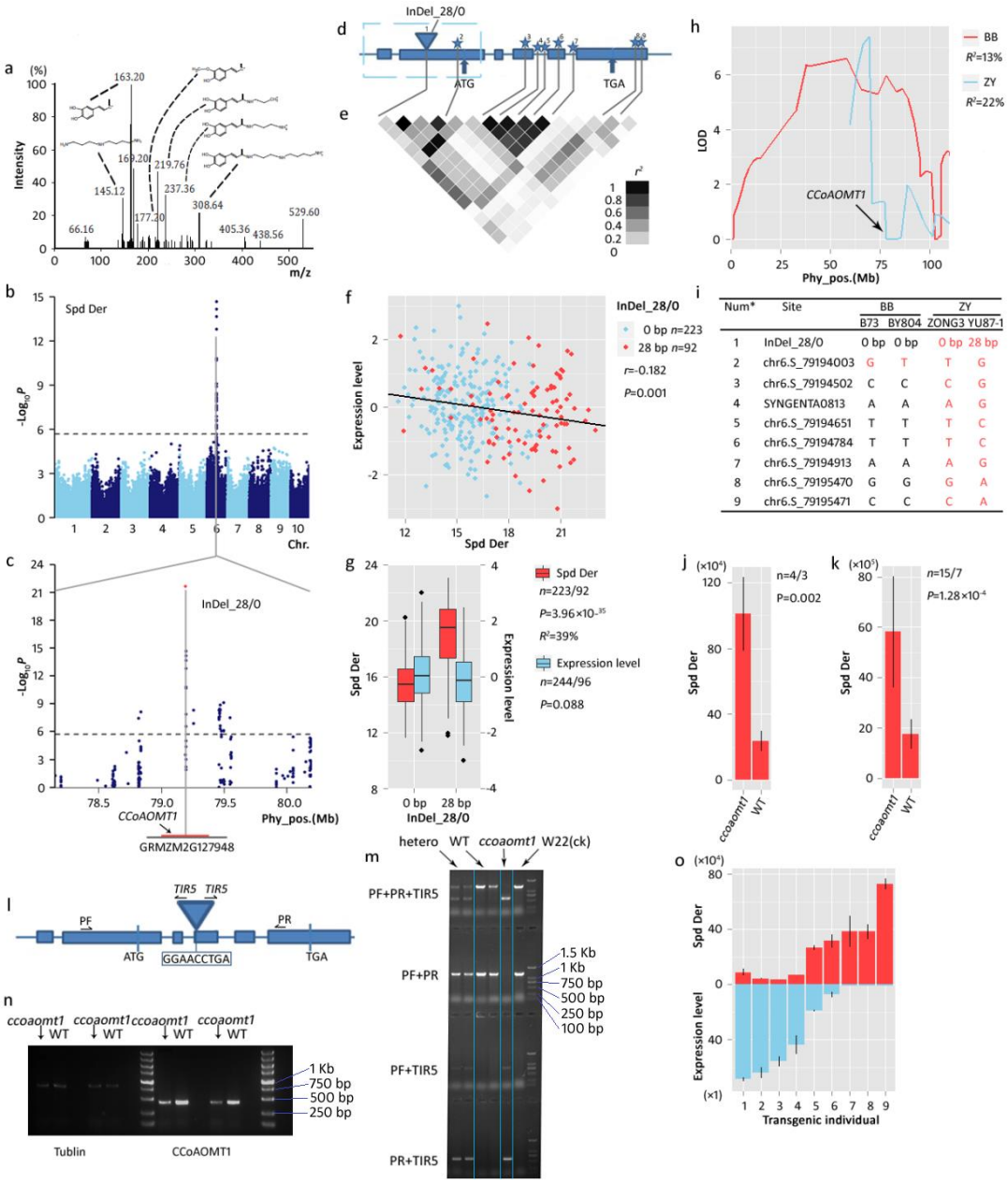
Supplementary Figure 3 Distribution of repeatability of metabolic traits. The vertical bar shows the percentage (%) of metabolic traits that were detected in the association panel across two to three experiments.



Supplementary Figure 4 Classification of identified genes based on GWAS. Candidate genes within the significant loci identified by GWAS across three environments were classified according to their functional annotation.

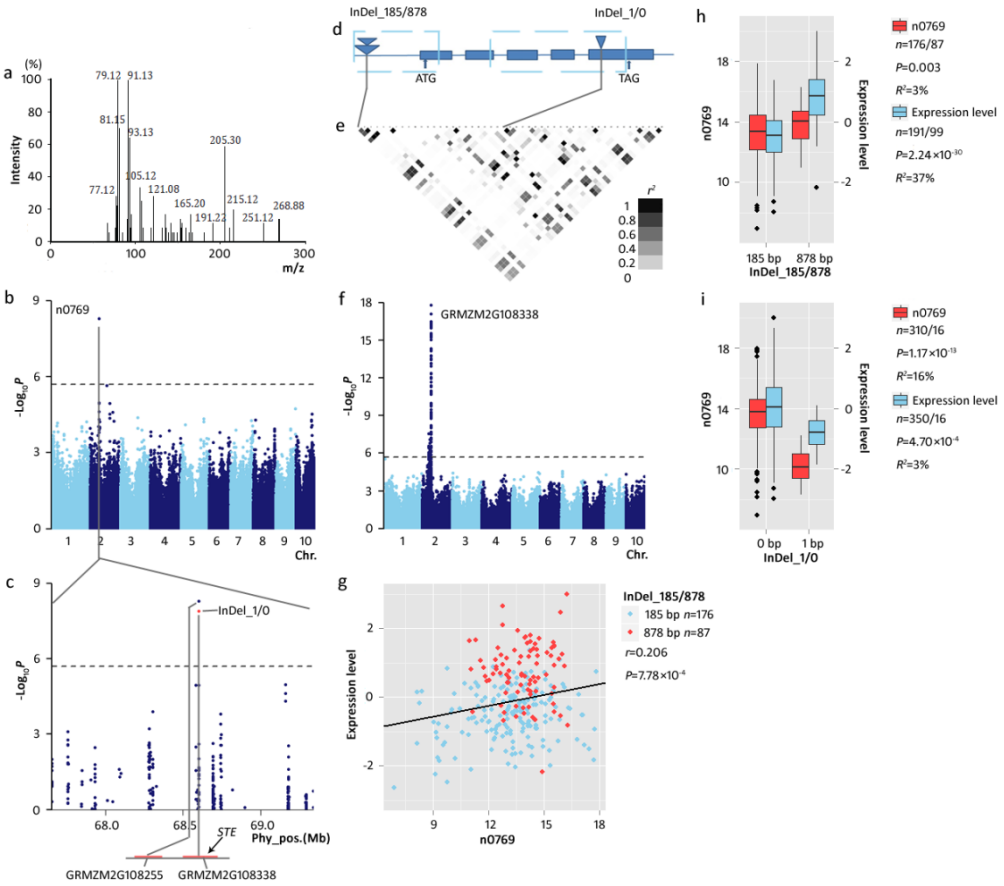


Supplementary Figure 5 Contents of phenolamides in *PHT* overexpression and *CCoAOMT1* knockout lines. a-c show the changes in rice leaf between wild-type (N=10) and *PHT* overexpression lines (N=4). d-f show the changes in maize kernels between wild-type (N=15) and *CCoAOMT1* knockout lines (N=14). Error bars represent the standard error. Abbreviations: A1, *N*'-Coumaroylagmatine; A2, *N*'-Caffeoylagmatine; A3, *N*'-Feruloylagmatine; A4, *N*'-Sinapoylagmatine; A5, *N*'-Feruloyl-*N*''-methoxyagmatine; P1, *N*'-Coumaroylputrescine; P2, *N*'-Caffeoylputrescine; P3, *N*'-Feruloylputrescine; P4, *N*'-Sinapoylputrescine; P5, *N*',*N*''-Di-coumaroyl-putrescine; P6, *N*',*N*''-Di-Feruloyl-putrescine; P7, *N*'-Feruloyl, *N*''-sinapoyl-putrescine; P8, Unknown; S1, *N*'-Coumaroylspermidine; S2, *N*'-Caffeoylspermidine; S3, *N*'-Feruloylspermidine; S4, *N*',*N*''-Di-coumaroyl-spermidine; S5, *N*'-Coumaroyl,*N*''-feruloyl-spermidine; S6, *N*',*N*''-Di-feruloyl-spermidine; S7, *N*-(Coumaroyl-O-hexoside)-spermidine; S8, *N*'-(Caffeoyl-O-hexoside)-spermidine; S11, *N*'-Caffeoyl-*N*''-feruloyl-spermidine derivative; S12, *N*', *N*''-Di-Feruloyl-*N*'''-sinapoylspermidine; S13, *N*', *N*''-Di-coumaroyl-*N*'''-feruloyl-spermidine.



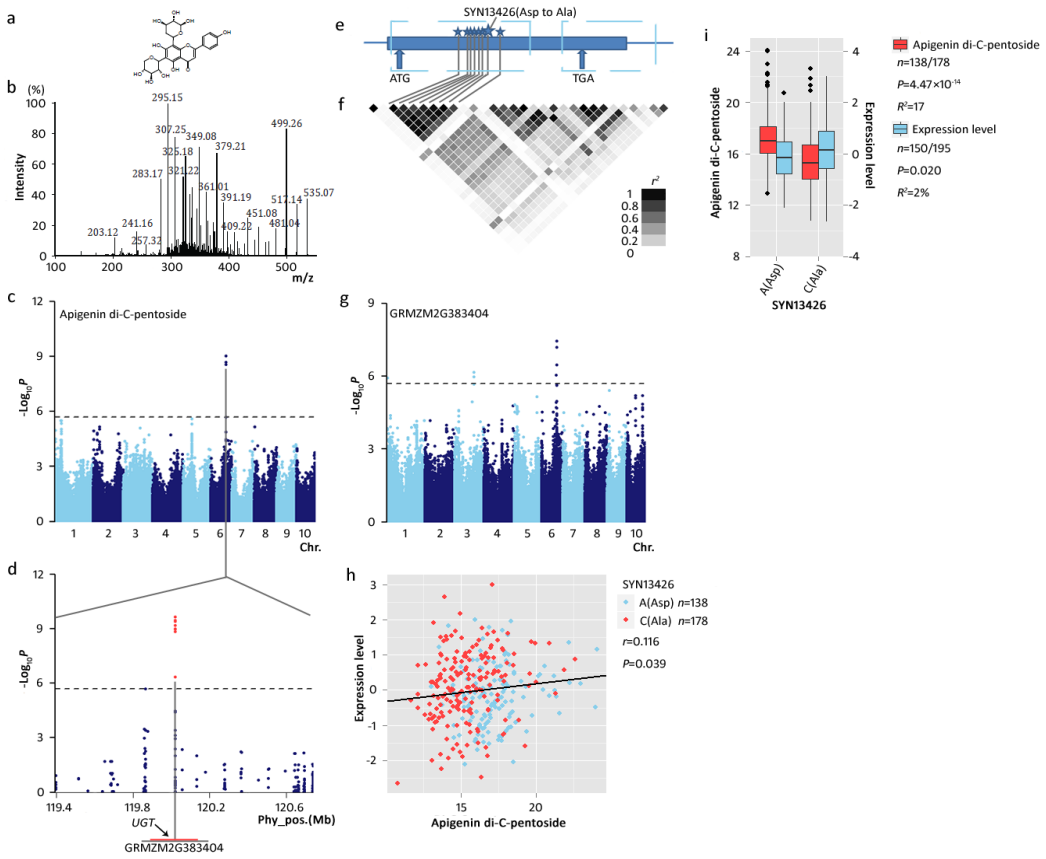
Supplementary Figure 6 Re-sequencing and functional validation of *CCoAOMT1*. (a) LC/MS fragmentation of compound Spd Der, which is a spermidine derivative containing spermidinecaffeoyl and feruloyl. Possible structures of the major fragments are shown. (b) Manhattan plot showing GWAS result of content of Spd Der. (c) Regional association plot for locus *CCoAOMT1*. A new significant site InDel_28/0 with an insertion/deletion polymorphism was identified by re-sequencing *CCoAOMT1* (GRMZM2G127948), shown in red. (d) The gene model of GRMZM2G127948. Filled blue boxes represent exons and UTRs. Dashed box marks the region re-sequenced in this study. The inverted triangle shows site InDel_28/0 while stars represent other significant sites. (e) Pair-wise r^2 value (a measure of LD) among all polymorphisms identified in *CCoAOMT1*. High LD levels between InDel_28/0 and all other significant sites are indicated. (f) Plots of correlation analysis between the content of Spd Der and the normalized expression level of *CCoAOMT1*. Maize inbred lines with different genotypes at the locus InDel_28/0 were shown in red and sky blue, respectively. The *r* is a Pearson correlation

coefficient. (g) Box plot of the content of Spd Der (shown in red) and expression level of *CCoAOMT1* (shown in sky blue); plotted as a function of genotypes at the site InDel_28/0 (0bp/28bp). The vertical lines of the box plot mark the range from 5^{th} to 95^{th} percentile of the total data. (h) QTL intervals located on chromosome 6 identified for the content of Spd Der. *CCoAOMT1* is within both QTL intervals from the linkage mapping based on RIL populations BB (B73/By804, red) and ZY (Zong3/Yu87-1, sky blue). (i) Allele type of the four parental lines at the significant sites. For each RIL, alleles different between the parental lines are shown in red. The numbers (Num) correspond to the superscripts in Fig. d. (j, k) Bar plots of the content of Spd Der in mature kernels of the *CCoAOMT1* knockout maize S3 (N=4) (j) and S4 (N=15) (k) lines, respectively. The number of replicates for wild type (WT) are 3 (j) and 7 (k), respectively. The vertical lines of the bar plot represent standard error. (l) The scaffold of mu1041577 in GRMZM2G127948. The inverted triangle represents the insertion of mu1041577, whose sequence is given. The 4 arrowheads indicate the location of the amplicon between primer set TIR5, PF and PR, whose sequences are presented in Supplementary Table 7. (m) Amplicon pattern of gene *CCoAOMT1* in heterozygous lines, WT, knockout lines and the negative CK (W22, the *Mu* acceptor). (n) Expression level of *CCoAOMT1* in the knockout lines and WT. (o) Bar plot for Spd Der content and OsCCoAOMT1 expression level in rice T1 transformants. Nine transgenic individuals are obtained by overexpressing OsCCoAOMT1 (LOC_Os06g06980, homologous to GRMZM2G127948, with 89% identity at the amino acid level) from rice cultivar Nipponbare driven by the 35S promoter in rice cultivar Zhonghua 11. The y axis represents the content of Spd Der (up) in rice grains (N=2) and the expression level in the leaves (N=3) relative to the lines with the lowest expression level of LOC_Os06g06980 (down), respectively. The vertical lines of the bar plot represent standard error.

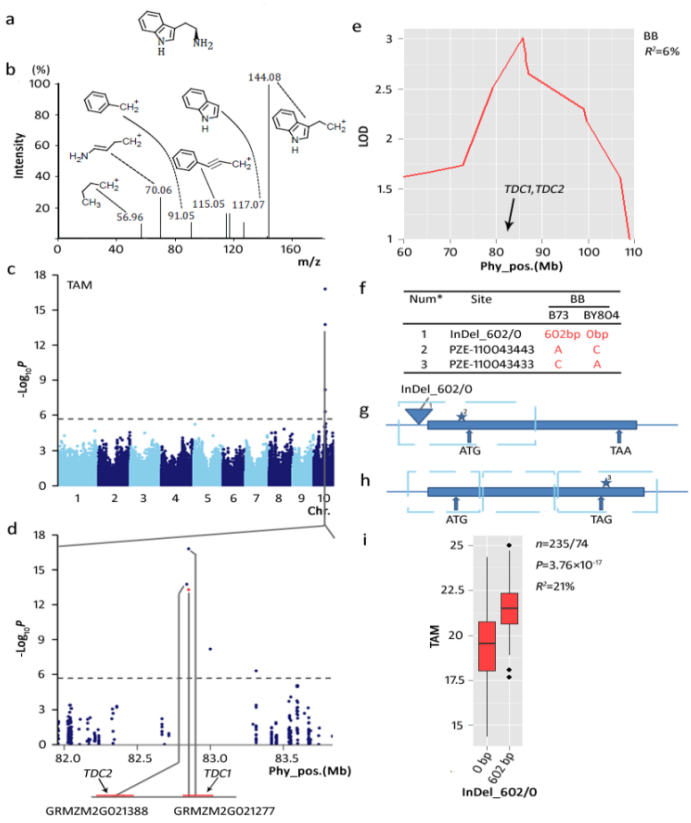


Supplementary Figure 7 Validation of *STE* via re-sequencing. (a) The structure of component n0769 cannot be identified in this study, however, the long fatty acid chain can be found in the LC/MS fragmentation. (b)Manhattan plot displaying the GWAS result of the content of n0769. P value is calculated by MLM, N= 339. (c) Regional association plot for locus *STE*. Two genes, GRMZM2G108338 and GRMZM2G108255, are located next to each other at this locus. A significant polymorphism, 1bp insertion/deletion (the red point) at the site InDel_1/0 in the coding region was identified via re-sequencing. (d) Gene model of *STE*. The filled blue boxes represent exons and UTRs. The dashed boxes mark the region re-sequenced in this study, and the inverted triangles represent the InDels. (e) Representation of the pair-wise r^2 value (a measure of LD) among all the sites in *STE*. (f) Manhattan plot showing the association between expression level of *STE* and genome-wide SNPs. Significant expression differences mapped to SNPs within *STE*, indicating a *cis* transcriptional regulation of this gene. (g) Plot of correlation analysis between the content of n0769 and the normalized expression level of gene *STE*. Maize inbred lines with different genotypes at the site InDel_185/878 were shown in sky blue and red, respectively. The *r* value is based on a Pearson correlation coefficient. (h) Box plot for n0769 content (red) and expression of *STE* (sky blue); plotted as a function of genotypes at the site InDel_185/878. The vertical lines of the box plot mark the range from 5th to 95th percentile of the total data. Both *P* values are based on ANOVA. (i) Box plot for n0769 content (red) and expression of *STE* (sky blue); plotted as a function of genotypes at the site InDel_1/0. The vertical lines mark the range from 5th to 95th percentile of the total data. Both *P* values are calculated based on ANOVA.

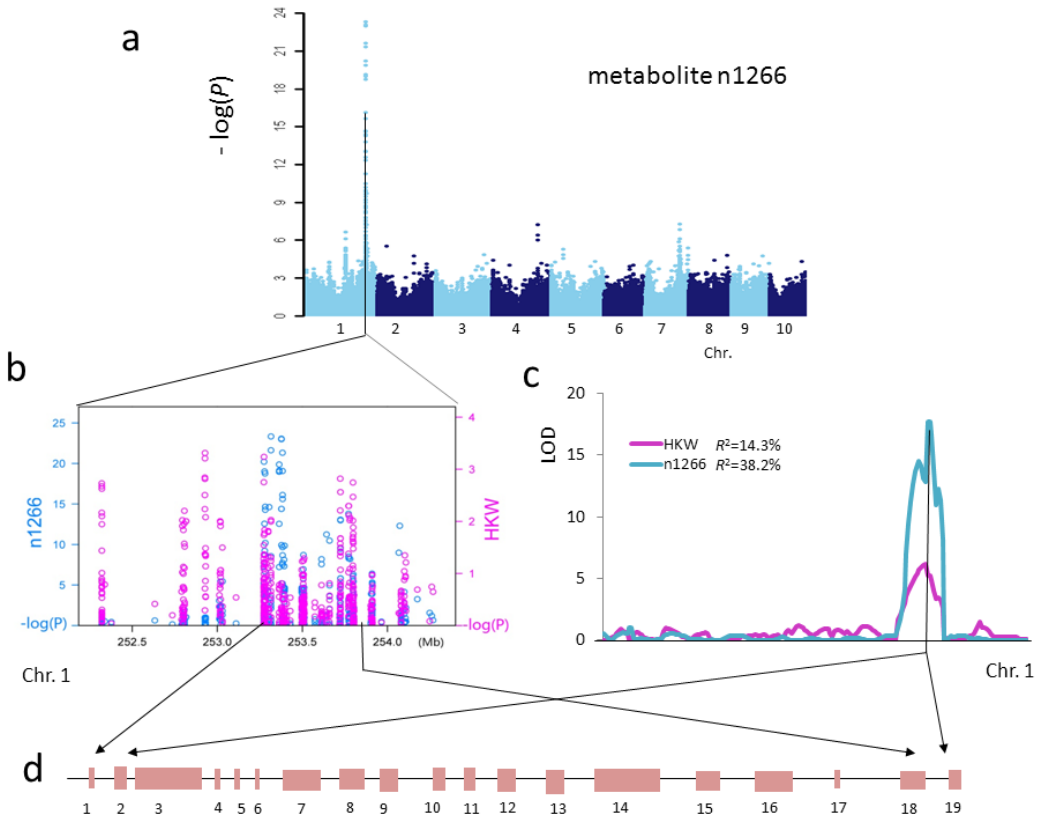
Supplementary Figure 8 Amino acid sequence alignments for *UGT*, *Bz*1 and Os06g18010. *UGT*, a maize gene annotated as UDP glycosyltransferase; *Bz*1, a maize gene modifies purple aleurone and plant color to pale or reddish brown, encodes UDP-glucoronosyl and UDP-glucosyl transferase; Os06g18010 is a rice gene functionally annotated as anthocyanin-3-*O*-glycosyltransferase.



Supplementary Figure 9 Validation of *UGT* via re-sequencing. (a) Structure of Apigenin di-*C*-pentoside (Apidi-*C*-pen), identified in flavonoid pathway. (b) The LC/MS fragmentation of Apidi-C-pen. Possible structures of the major fragments are shown. (c) Manhattan plot displaying the association result of the content of Apidi-*C*-pen. P value is calculated by MLM, N=339. (d) Regional association plot for locus *UGT* (UDP glycosyltransferase). Only polymorphisms in the UGT gene significant above the threshold $P<1.8 \times 10^{-6}$ are shown. Polymorphisms identified by using RNA-seq and theMaizeSNP50 BeadChip were validated by re-sequencing and additional SNPs were identified in the coding region. Red points represent the significant sites identified or validated by re-sequencing. (e) Gene model of *UGT* (GRMZM2G383404). The filled blue boxes represent exons and UTRs. The dashed boxes mark the region re-sequenced in this study, and the stars represent the significant sites identified by re-sequencing. The most likely functional site is chr6.S_120019452, indicated by a large star. The nucleotide change at this site ("A" to "C") causes amino acid substitution (Asp to Ala). (f) Pair-wise r^2 value (a measure of LD) among all identified polymorphisms in *UGT*. Eight significant sites are in one LD block. (g) Manhattan plot showing the association between expression level of *UGT* and genome-wide SNPs. P value is calculated by MLM, N= 368. Significant signals mapped to SNPs located at 155kb upstream of *UGT*, indicating a *cis* transcriptional regulation of this gene. (h) Plot of correlation analysis between the content of Api di-*C*-pen and the normalized expression level of gene *UGT*. Maize inbred lines with different genotypes at the locus chr6.S_120019452 were shown in red and sky blue, respectively. (i) Box plot for the content of Api di-*C*-pen (shown in red) and expression level of *UGT* (shown in sky blue); plotted as a function of genotypes at the site chr6.S_120019452 (AA/GG). The vertical lines of the box plot mark the range from 5th to 95th percentile of the total data. Both *P* values are calculated based on ANOVA, N=216 and 345, respectively.



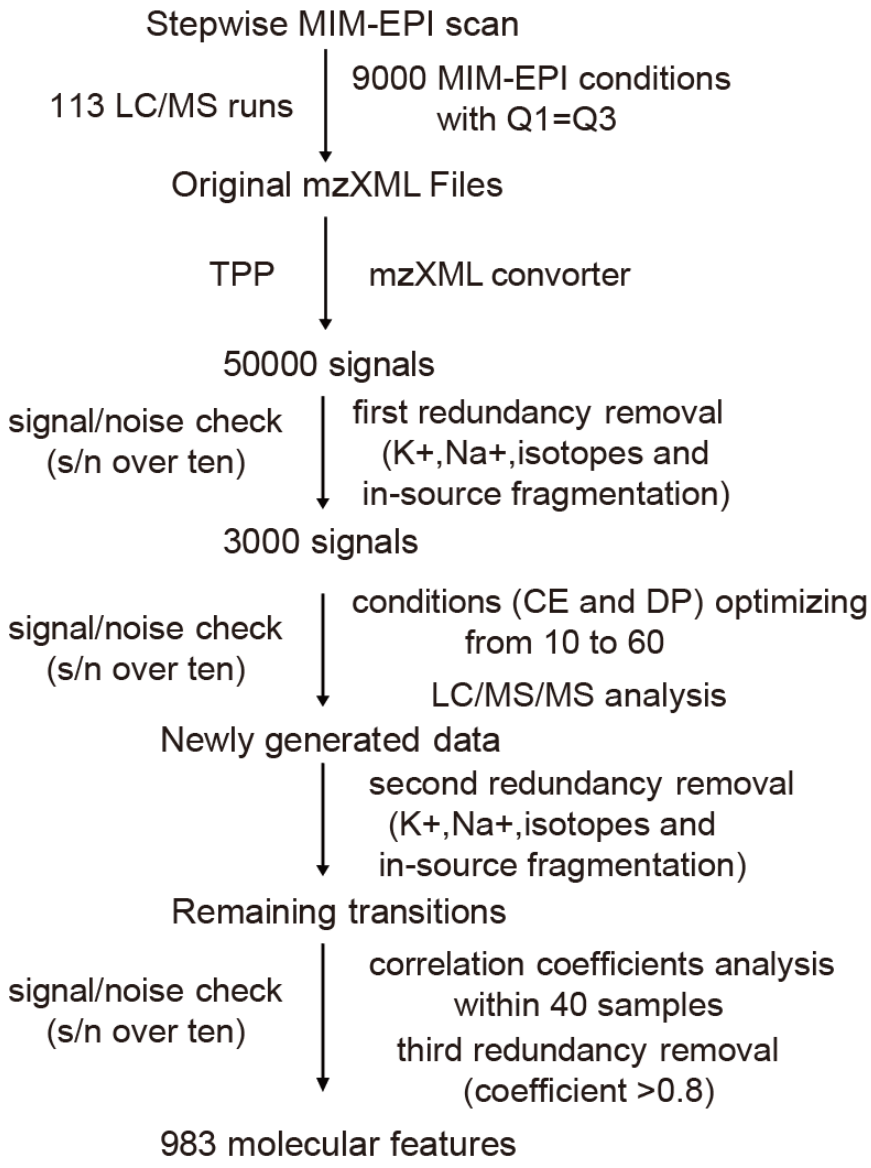
Supplementary Figure 10 Re-sequencing of *TDC1* via re-sequencing. (a) Chemical structure of 3-(2-Aminoethyl) indole Hydrochloride (TAM), a component of the tryptamine pathway. (b)LC/MS fragmentation of TAM. Possible structures of the major fragments are shown. (c) Manhattan plot displaying the GWAS result for the content of TAM. P value is calculated by MLM, N=339. (d) Regional association plot for locus *TDC* (Tryptophan decarboxylase). Two homologous genes *TDC1* (GRMZM2G021277) and *TDC2* (GRMZM2G021388) are located close to each other at this locus. A new insertion-deletion polymorphism, InDel_602/0 (the red point), was identified in the promoter region of *TDC1* by re-sequencing. (e) *TDC1*and *TDC2* are located at the QTL interval for the content of TAM identified by linkage mapping in the RIL population BB (B73/By804). (f) Significant polymorphisms between the two parental lines of this population. Different alleles from B73 and By804 are marked in red, implicating possible functional sites of the locus. The numbers (Num) correspond to the superscripts of the significant sites shown in (g) and (h), respectively. (g) Gene model of *TDC1*. The filled blue box represents exon and UTRs. The dashed box marks the region re-sequenced in this study. The star represents the significant SNP while the inverted triangle represents the InDel_602/0. (h) Gene model of *TDC2*. The filled blue box represents exon and UTRs. The dashed box marks the region re-sequenced in this study, and the star represents the significant SNP. (i) Box plot for the content of TAM; plotted as a function of genotypes at the site InDel_602/0. The vertical lines of the box plot mark the range from 5th to 95th percentile of the total data.



e

No.	Gene	Annotation
1	GRMZM2G472643	Protein kinase domain/Protein tyrosine kinase/Phosphotransferase enzyme family
2	GRMZM2G472651	thylakoid assembly4; mttA/Hcf106 family
3	GRMZM2G066234	Single-strand binding protein family
4	GRMZM2G366373	Aux/IAA-transcription factor 3
5	GRMZM2G066067	UDP-glucuronosyl/UDP-glucosyltransferase
6	GRMZM2G121047	Unknown
7	GRMZM2G141332	Ribonuclease
8	GRMZM2G141379	Zinc finger, C3HC4 type (RING finger)
9	GRMZM2G445139	G10 protein
10	GRMZM2G070061	Prefoldin, subunit 4
11	GRMZM2G070211	WRKY DNA -binding domain
12	GRMZM2G070252	Zeta toxin domain
13	GRMZM2G112596	ATPase-like, ParA/MinD
14	GRMZM2G112672	Calcium-binding EF-hand
15	GRMZM2G021055	Multi antimicrobial extrusion protein
16	GRMZM2G342105	Protein of unknown function (DUF3633)
17	AC211910.4_FG005	Unknown
18	GRMZM2G043191	Endonuclease/exonuclease/phosphatase
19	GRMZM2G043127	Unknown

Supplementary Figure 11 Dissection of metabolite and candidate genes associated with 100-kernel weight (HKW). (a). Manhattan plot displaying the GWAS result for the content of metabolite n1266. P value is calculated by MLM, N= 339. (b). Regional association plots between the candidate locus HKWM1 and two traits (n1266, blue; and HKW, red). *P* value was calculated by using mixed linear model controlling Q and K for n1266, while for HKW, mixed linear model controlling Q was used. (c). HKWM1 is located at the QTL interval for both the content of metabolite n1266 and HKW identified by linkage mapping in the RIL population ZY (Zong3/Yu87-1). (d). Nineteen genes were found within the ~500Kb region of locus HKWM1. (e). Annotation of the 19 genes found within the ~500Kb region of locus HKWM1.



Supplementary Figure 12 Workflow of metabolite profiling and related data analysis in this study. 184 metabolites and 799 metabolite features were obtained after conditions optimizing and three times redundancy removal.

Supplementary Table 1 Summary of candidate gene association results of *PHT*, *CCoAOMT1*, *STE*, *UGT*, *TDC1*, *TDC2* based on re-sequencing.

Gene/Trait	Marker[a]	Chr.	Position (bp)	*P* (Trait)[b]	N[c]	*P* (eQTL)[d]	Region	Codon change	Amino acid replacement
PHT/ n0381-1	chr1.S_140322113**	1	140322113	2.6×10^{-13}	235	0.66	promoter		
	chr1.S_140322090**	1	140322090	1.77×10^{-12}	231	0.84	promoter		
	chr1.S_140322089**	1	140322089	9.25×10^{-13}	235	0.73	promoter		
	chr1.S_140322068**	1	140322068	3.35×10^{-13}	237	0.63	promoter		
	chr1.S_140322063**	1	140322063	6.93×10^{-13}	236	0.44	promoter		
	chr1.S_140322060**	1	140322060	7.58×10^{-12}	237	0.43	promoter		
	InDel_1_140322034**	1	140322034	3.10×10^{-13}	238	0.6	promoter		
	chr1.S_140322028**	1	140322028	1.26×10^{-10}	237	0.06	promoter		
	chr1.S_140322025**	1	140322025	8.84×10^{-13}	237	0.6	promoter		
	chr1.S_140322024**	1	140322024	3.10×10^{-13}	238	0.6	promoter		
	chr1.S_140322015**	1	140322015	6.77×10^{-13}	237	0.77	promoter		
	chr1.S_140322011**	1	140322011	1.46×10^{-12}	238	0.43	promoter		
	chr1.S_140322010**	1	140322010	1.67×10^{-10}	237	0.04	promoter		
	chr1.S_140322009**	1	140322009	3.26×10^{-10}	238	0.04	promoter		
	chr1.S_140322005**	1	140322005	2.45×10^{-10}	239	0.03	promoter		
	chr1.S_140322004**	1	140322004	7.13×10^{-14}	237	0.63	promoter		
	InDel_1_140322003**	1	140322003	3.18×10^{-13}	239	0.52	promoter		
	chr1.S_140321996**	1	140321996	5.62×10^{-10}	249	0.06	promoter		
	InDel_1_140321992**	1	140321992	2.62×10^{-13}	239	0.58	promoter		
	chr1.S_140321991**	1	140321991	1.72×10^{-12}	250	0.47	promoter		
	chr1.S_140321978**	1	140321978	1.18×10^{-8}	248	0.36	promoter		
	chr1.S_140321962**	1	140321962	3.54×10^{-14}	250	0.67	promoter		
	InDel_1_140321926**	1	140321926	2.27×10^{-13}	251	3.42×10^{-5}	promoter		
	chr1.S_140321924**	1	140321924	1.04×10^{-12}	230	0.08	promoter		
	InDel_1_140321888**	1	140321888	1.70×10^{-14}	243	0.72	promoter		
	chr1.S_140321875**	1	140321875	4.31×10^{-9}	247	0.03	promoter		
	InDel_1_140321750**	1	140321750	3.67×10^{-14}	251	0.59	5'UTR		
	PZE-101116556*	1	140321705	8.57×10^{-14}	248	0.69	5'UTR		
	InDel_1_140321622**	1	140321622	4.34×10^{-14}	252	0.57	exon	GCG-0	missing a amino acid
	chr1.S_140321617**	1	140321617	4.66×10^{-13}	252	0.59	exon	GGC-GCC	Gly-Ala
	chr1.S_140321605**	1	140321605	1.26×10^{-11}	250	0.26	exon	CTG-CAG	Leu-Gln
	chr1.S_140321550*	1	140321550	1.05×10^{-13}	252	0.1	exon	GTC-GTA	Val-Val
	chr1.S_140321540*	1	140321540	5.10×10^{-12}	252	0.39	exon	TAC-AAC	Tyr-Asn
	PZE-101116550	1	140321299	2.63×10^{-14}	320	0.26	Intron		
CCoAOMT1/ n1544-1	InDel_28/0**	6	79193717	1.99×10^{-22}	315	0.06	5'UTR		
	chr6.S_79194003	6	79194003	1.23×10^{-8}	339	0.59	5'UTR		
	chr6.S_79194502	6	79194502	1.51×10^{-13}	339	0.01	exon	GTC-GTG	Val-Val
	SYNGENTA0813	6	79194582	6.25×10^{-15}	322	0.02	intron		
	chr6.S_79194651	6	79194651	1.53×10^{-11}	339	0.01	intron		
	chr6.S_79194784	6	79194784	2.01×10^{-15}	339	0.01	exon	GGT-GGC	Gly-Gly
	chr6.S_79194913	6	79194913	2.05×10^{-14}	339	1.70×10^{-3}	intron		
	chr6.S_79195470	6	79195470	1.91×10^{-11}	339	0.1	3'UTR		
	chr6.S_79195471	6	79195471	2.70×10^{-7}	339	0.02	3'UTR		
STE/ n0769	InDel_185/878**	2	68602648	0.13	263	3.65×10^{-13}	promoter		
	InDel_1/0*	2	68601038	3.80×10^{-8}	326	0.02	exon		
UGT/ n1201	chr6.S_120019452*	6	120019452	6.43×10^{-10}	315	0.08	exon	AGC-AAC	Ser-Asn
	chr6.S_120019487*	6	120019487	1.01×10^{-9}	315	0.07	exon	GTA-TTA	Val-leu
	chr6.S_120019493*	6	120019493	3.92×10^{-10}	315	0.08	exon	ATC-CTC	Ile-Leu
	chr6.S_120019528*	6	120019528	3.36×10^{-10}	316	0.05	exon	GGC-GGG	Gly-Gly
	chr6.S_120019564*	6	120019564	2.23×10^{-10}	316	0.1	exon	CAG-CAA	Gln-Gln

	chr6.S_120019581*	6	120019581	1.43×10^{-9}	316	0.07	exon	GTG-GCG	Val-Ala
	SYN13426*	6	120019623	1.08×10^{-9}	316	0.08	exon	GCC-GAC	Ala-Asp
	SYN13424*	6	120019772	4.28×10^{-7}	316	0.1	exon	AAC-CAC	Asn-His
TDC1/ n0671	InDel_602/0**	10	82851072	4.80×10^{-14}	309	-	promoter		
	PZE-110043443	10	82850420	1.52×10^{-17}	311	-	5'UTR		
TDC2/ n0671	PZE-110043433	10	82839422	1.64×10^{-14}	321	-	3'UTR		

[a]Significant polymorphisms within the candidate genes identified by re-sequencing

[b]P values calculated by association analysis between genomic variants obtained after re-sequencing and the metabolic traits using MLM

[c]Number of lines in association analysis between genomic variants obtained after re-sequencing and the metabolic traits using MLM

[d]P values calculated by association analysis between genomic variants and the expression amount of the gene using MLM.

"*" represents genomic variants or markers remodeled after re-sequencing

"**"represents newly found genomic variants or markers after re-sequencing

Supplementary Table 2 List of primers used in this study.

Category	Gene	Primer name	Primer sequence
Sequencing	*PHT*	ACT3F/ACT5R	5'-CGGGTCAAGGCTACAGCGCAA-3'/5'-GGTACCTCGCGACCACAGCG-3'
		ACTCOD1F/ACTCOD2R	5'-GTGACGGTGGACATAACGCG-3'/5'-GTTCACGGCGACCCTTATCT-3'
		ACT3U1F/ACT3U4R	5'-GCTGGCTGGGGTTCAGATT-3'/5'-CATGGTTGTGATGCTTGGTGT-3'
	CCoAOMT1	CCOMT3F/CCOMT3R	5'-CACGACCCGAACAACCGGACG-3'/5'-TGGGCAGGCAGGATCGGAGT-3'
	STE	STE12F/STE5R	5'-GGACGGTGGAAACGTGGGGAAT-3'/5'-TAGTACGCCGGCAGGACCAG-3'
		STECOD1F/STECOD2R	5'-CGGGCCAGGGTGTCGATCTGT-3'/5'-CGAGCGTGTCCGTGGTGGA-3'
	UGT	AGTCOD2F/AGTCOD3R	5'-GTTGCTCCCAAGTGACTCCAA-3'/5'-CACCCAAAGGAACCGACAGT-3'
		AGT3U3F/AGT3U3R	5'-GAGGTCATGTCCGACGAAGC-3'/5'-GAGCCAGGCACCAAGGAGTC-3'
	TDC1	Tyr-3-6F/Tyr-3-2R	5'-GGTTTAGGGCACAAGCATATT-3'/5'-GAAGAAGGCAGAAGTTAGGGTT-3'
		Tyr-1-1F/Tyr-11-5R	5'-CCTTTGCCAGATTGCCAGTT-3'/5'-GCATGACCTCGAAGAGCCT-3'
	TDC2	Tyr-8-2F/Tyr-8-1R	5'-CCTGCTGATTTTGCTATGGT-3'/5'-GGACCTCCTTTAGGGCAAT-3'
		Tyr-10-2F/Tyr-10-2R	5'-CCCTTCGACATTGCCCTAA-3'/5'-ATGTCCTTGAGGTCGGTGA-3'
		Tyr-14-2F/Tyr-14-2R	5'-CACCGTCACCGACCTCAA-3'/5'-CCTTTCTTACCTTTCTGTGCC-3'
PCR assay	*CCoAOMT1*	313CC1F/313CC1R	5'-CACTGGCACGGCGCACCTAT-3'/5'-GGCCAACGTGAAGGTGAAGAA-3'
	STE	STE11F/STE1R	5'-GCAGTAGCGACGGTGTGCGA-3'/5'-ACCCCGTCCTGCTCATGCCC-3'
	TDC1	Tyr-3-6F/Tyr-3-2R	5'-GGTTTAGGGCACAAGCATATT-3'/5'-GAAGAAGGCAGAAGTTAGGGTT-3'
Uniform-MU for DNA	*CCoAOMT1*	PF/PR	5'-CACTGGCACGGCGCACCTAT-3'/5'-GCAGCCGCTCGTGGTAGTTGA-3'
	General	TIR5	5'-GCTCTTCGTCYATAATGRCAATT-3'
Transgenic	*PHT*		5’-attB1- CACGCATATCATTGAGGAAAAAAC-3’/5’-attB2-CTTGCAGTATTACAAAGACGTAC-3’
	CCoAOMT1		5’-attB1- GGAGGGACTCGTTCGTTCA -3’/5’-attB2-TCACTTGACGCGGCGGCA-3’
qRT-PCR	*PHT*		5'-TCGGTAACATGGTGCTGTG-3'/5'-TCGACGAAGGACTGGATGTA-3'
	actin(for rice)		5'-tggcatctctcagcacattcc-3'/5'-tgcacaatggatgggtcaga-3'
Uniform-MU for RNA	*CCoAOMT1*	P-F/P-R	5'-AGCACCCATGGAACCTGATGA-3'/5'-GGAGCCGTTCCACAGCGT-3'

Supplementary Table 3 List of metabolites significantly associated with 100-kernel weight (HKW) at significance level $P \leq 0.05$ by using general stepwise regression.

Experiment	Metabolite	Estimate	Standard Error	t Value	Pr > |t|	N
E2	n0020	1.35	0.36	3.78	0.0002	339
E2	n0136	-1.01	0.20	-4.94	<.0001	339
E2	n0186	0.56	0.21	2.72	0.0069	339
E2	n0673	1.78	0.40	4.44	<.0001	339
E2	n0712	-1.03	0.41	-2.52	0.0123	339
E2	n0743	-0.73	0.26	-2.83	0.005	339
E2	n0790	1.08	0.21	5.23	<.0001	339
E2	n0867	-1.58	0.36	-4.43	<.0001	339
E2	n0956	1.09	0.32	3.46	0.0006	339
E2	n0957	0.48	0.19	2.46	0.0147	339
E2	n0980	-0.45	0.12	-3.87	0.0001	339
E2	n1014	-1.28	0.35	-3.64	0.0003	339
E2	n1043	-2.37	0.15	-15.75	<.0001	339
E2	n1051	-0.29	0.14	-2.11	0.0357	339
E2	n1096	-0.97	0.40	-2.42	0.0163	339
E2	n1166	-1.41	0.37	-3.84	0.0001	339
E2	n1177	-0.61	0.19	-3.25	0.0013	339
E2	n1230	1.54	0.34	4.48	<.0001	339
E2	n1243	0.35	0.15	2.34	0.0198	339
E2	n1266	-0.37	0.12	-2.99	0.003	339
E2	n1329	-0.86	0.26	-3.29	0.0011	339
E2	n1356	-0.77	0.17	-4.43	<.0001	339
E2	n1558	2.45	0.57	4.27	<.0001	339
E2	n1605	-2.28	0.75	-3.04	0.0026	339
E2	n1618	1.73	0.65	2.66	0.0083	339
E2	n0048	1.16	0.38	3.05	0.0025	339
E3	n0486	-0.92	0.19	-4.88	<.0001	335
E3	n1165	-1.88	0.39	-4.8	<.0001	335
E3	n1421	5.27	1.23	4.28	<.0001	335
E3	n0941	-1.42	0.41	-3.48	0.0006	335
E3	n0923	-1.69	0.41	-4.09	<.0001	335
E3	n1027	0.39	0.10	3.88	0.0001	335
E3	n0956	0.86	0.27	3.24	0.0014	335
E3	n1544-1	-0.26	0.08	-3.19	0.0016	335
E3	n1393	-0.69	0.18	-3.88	0.0001	335
E3	n0111	0.38	0.13	2.85	0.0047	335
E3	n1330	-0.60	0.25	-2.39	0.0177	335
E3	n0838	-0.84	0.38	-2.21	0.0283	335
E3	n1104	-0.79	0.30	-2.63	0.0091	335
E3	n1618	-3.00	1.09	-2.76	0.0062	335

E3	n1058	1.08	0.28	3.86	0.0001	335
E3	n0906	-0.54	0.22	-2.52	0.0123	335
E3	n0534	0.39	0.15	2.61	0.0095	335

Supplementary Table 4 List of 100-kernel weight (HKW) QTLs and their co-localized QTLs for metabolic traits.

Trait	Chromosome	Confidence Interval（Mb）[a]	LOD[b]	R^2（%）[c]	Linkage Population[d]
HKW	6	15.2-86.3-86.3	3.78	8.9	BB
HKW	7	21.5-31.8-31.8	2.94	6.5	BB
HKW	1	233.9-249.4-258.4	6.15	14.3	ZY
HKW	5	181.4-181.4-181.4	2.58	6.9	ZY
HKW	7	4.9-9.2-9.2	3.10	8.1	ZY
HKW	8	112.7-118.7-118.7	3.23	8.3	ZY
HKW	10	144.1-144.7-146.1	2.72	5.5	ZY
n0725	6	27.5-86.3-95.6	5.84	13.0	BB
n0908	6	1.3-86.3-95.6	9.22	16.9	BB
n1173	6	1.3-86.3-95.6	6.35	11.3	BB
n1405	6	27.5-86.3-102.8	6.10	12.9	BB
n1578	6	65.6-83.2-86.3	3.08	6.2	BB
n0016	7	6.5-6.5-31.8	4.66	10.4	BB
n0215	7	6.5-6.5-106.8	4.39	10.3	BB
n1503	7	6.5-6.5-20.2	3.15	7.9	BB
n0036	1	233.9-253-253	3.91	12.0	ZY
n0037	1	233.9-253-258.4	10.23	24.5	ZY
n0349	1	233.9-253-258.4	5.79	13.1	ZY
n0472	1	253-253-258.4	3.06	8.9	ZY
n1266	1	233.9-253-258.4	17.70	38.2	ZY
n0087	5	173.4-180.7-181.4	5.86	12.2	ZY
n0298	5	173.4-180.2-181.4	6.13	12.4	ZY
n0386	5	175.6-180.2-181.4	3.17	7.3	ZY
n0610	5	175.6-181.4-181.4	3.47	8.7	ZY
n0640	5	180.2-180.7-181.4	3.15	6.8	ZY
n0841	5	175.6-180.2-181.4	4.57	10.6	ZY
n0151	7	10.8-10.8-10.8	3.05	8.7	ZY
n0202	7	3.2-3.2-9.2	4.77	9.7	ZY
n0203	7	3.2-3.2-9.2	4.84	9.9	ZY
n0209	7	9.2-10.8-10.8	3.91	11.5	ZY
n0216	7	3.2-3.2-9.2	5.80	11.6	ZY
n0291	7	9.2-9.2-10.8	3.50	9.3	ZY
n0429-1	7	9.2-10.8-10.8	4.28	9.3	ZY
n0447	7	9.2-10.8-10.8	3.73	8.3	ZY
n0610	7	4.9-9.2-9.2	5.08	12.8	ZY
n0718	7	3.2-3.2-3.2	3.37	6.4	ZY
n0905-1	7	9.2-10.8-10.8	3.64	9.9	ZY
n0913	7	4.9-6.5-9.2	3.20	7.5	ZY
n1398	7	9.2-10.8-10.8	3.59	6.8	ZY
n1444	7	3.2-3.2-3.2	3.60	7.6	ZY

n1484	7	9.2-9.2-10.8	3.76	10.3	ZY
n2035	7	3.2-3.2-4.9	5.17	11.2	ZY
n0313	8	112.7-112.7-112.7	5.47	12.1	ZY
n0973	8	95.7-110.8-118.7	3.69	8.4	ZY
n0105	10	148.3-148.3-148.3	3.12	8.1	ZY
n0732	10	147-148.3-148.3	3.98	9.8	ZY
n0735	10	147-148.3-148.3	4.20	10.6	ZY
n1104	10	139.1-139.1-139.1	3.00	8.1	ZY
n1258	10	146.1-147-148.3	3.44	8.5	ZY
n1423	10	139.1-139.1-139.1	3.13	7.2	ZY

[a]The number in the middle represents physical position of the peak marker referring to version 5b.60 of the maize reference sequence, flanked by the left and right markers of the confidential interval of the QTL, respectively.

[b]LOD value of the corresponding QTL.

[c]The phenotypic variation explained by the corresponding QTL.

[d]BB and ZY correspond to linkage population B73/By804 and Zong3/Yu87-1 RIL populations, respectively.

Supplementary Note 1: Detailed procedures of LC-MS/MS based metabolite profiling

HPLC conditions

The sample extracts were analyzed using an LC-ESI-MS/MS system (HPLC, Shim-pack UFLC Shimadzu CBM20A system, http://www.shimadzu.com.cn/; MS, Applied Biosystems 4000 Q TRAP, http://www.appliedbiosystems.com.cn/). The analytical conditions were as follows. HPLC: column, shim-pack VP-ODS C18 (pore size 5.0μm, length 2×150mm); solvent system, water (0.04% acetic acid): acetonitrile (0.04% acetic acid); gradient program, 100:0 V/V at 0 min, 5:95 V/V at 20.0 min, 5:95 V/V at 22.0 min, 95:5 V/V at 22.1 min, 95:5 V/V at 28.0 min; flow rate, 0.25ml/min; temperature, 40°C. Injection volume, 5μl. The effluent was alternatively connected to an ESI-triple quadrupole-linear ion trap (Q TRAP)-MS or ESI-QqTOF-MS.

ESI-QqTOF-MS/MS

Full time-of-flight (TOF) scans of HPLC effluents were acquired in the mass range of m/z 100-1,000 using a TOF Agilent 6520 Accurate-Mass time-of-flight (Q-TOF) mass spectrometry equipped with an Dual ESI electrospray ion source in positive ion mode (http://www.chem.agilent.com/en-US/products-services/Instruments-Systems/Mass-Spectrometry/Pages/default.aspx) using the following instrument settings: nebulizer gas, nitrogen, 1.6 bar; dry gas, nitrogen, 6 L/min, 190 °C; capillary voltage, 4,000; end plate offset, 500 V; funnel 1 RF, 200 V; funnel 2 RF, 200 V; in-source collision-induced dissociation (CID) energy, 0 V; hexapole RF, 100 V; quadrupole ion energy, 5 eV; collision gas, argon; collision energy, 10 eV; collision RF 200/400 V (timing 50/50); transfer time, 70 μs; prepulse storage, 5 μs; pulser frequency, 10 kHz; spectra rate, 3 Hz. Internal mass calibration of each analysis was performed by infusion of 20 μL 10 mmol/L lithium formate in isopropanol/ water (1:1, v/v) at the LC method time point of 18 min using a diverter valve.

ESI-Q TRAP-MS/MS

Linear ion trap (LIT) and triple quadrupole (QQQ) scans were acquired on a triple quadrupole-linear ion trap mass spectrometer (Q TRAP), ABI 4000 Q TRAP LC/MS/MS System, equipped with an ESI-TurboIon-Spray interface, operating in a positive ion mode and controlled by Analyst 1.5 software (AB Sciex). The ESI source operation parameters were as follows: ion source, turbo Spray; source temperature 500 °C; ion spray voltage (IS) 5500 V; ion source gas I (GSI), gas II (GSII), curtain gas (CUR) were set at 55 psi, 60 psi, and 25.0psi, respectively; the collision gas (CAD) was high. Instrument tuning and mass calibration were performed with 10 and 100 μmol/L polypropylene glycol solutions in QQQ and LIT mode, respectively.

LIT experiments

The multiple ion monitoring (MIM) scan which served as a survey scan to trigger information-dependent acquisition (IDA) of enhanced product ions (EPI) scan model, MIM-IDA-EPI[1], was carried out to screen metabolites. MIM scan used minimal collision energy (CE) (5eV) in Q2 so that metabolite ions isolated in Q1 passed through Q2 with minimal fragmentation. Q3 was monitored at the same metabolite ions as Q1 instead of fragment ions. EPI scans were typically acquired with DP of 40.0 V, and CE of 30.0 V. The mass range was from m/z 100 to 1,000. A modified MIM-EPI strategy, called stepwise scan MIM-EPI was adopted, in

which Q1 (Q3) was set from 100.1 to 1000.0 Da in positive scan mode, and the mass step is 0.1 Da, such as from 100.1/100.1, 100.2/100.2, 100.3/100.3, to 1000.0/1000.0. Each MIM transition was performed with a 5-ms Dwell time and a 5-ms pause time. We monitored each MIM-IDA-EPI experiment with 80 MIM transitions, product ions of each metabolite ion were scanned from 50.0 Da to 1,000.0 Da in Q3, and the total cycle time for one scan was approximately 1.8 s. All together 113 LC/MS runs were screened.

QQQ experiments

QQQ scans were acquired as multiple reaction monitor (MRM) experiments with collision gas (nitrogen) set to 5psi. Selection of Q3 masses, DP, and CE for individual MRM transitions was done the basis of targeted EPI experiments with further DP and CE optimization. Q1 and Q3 resolutions were set to 'unit'. In our study, 4000 Q Trap was used for the metabolite detection on samples of the association panel in E1 with MRM experiments. 5500 Q was used in E2 and E3 with 'scheduled MRM'[2]experiments because of highly sensitivity and accuracy. Scheduled MRM parameters were set as the following: MRM detection window, 60 s; target scan time, 1 s. Data were processed using Analyst 1.5 software; peak areas were integrated using IntelliQuan algorithm.

Creation of nearly non-redundant MS2T Library

The data matrix was generated from the metabolic profile data. To this end, mzXML files were generated from the original data and the MS2T data matrix was obtained using in-house software written in Perl. We obtained more than 50000 signals with MS/MS spectra from about 9000 MIM-EPI transitions. It should be noted that the original MS2T libraries contain a large amount of data derived from artifacts and also there is redundancy due to the iterative acquisition of MS/MS spectra of the same metabolite.

To produce a matrix containing fewer biased and redundant data, in-house software written in *Perl*[3] was used to remove the redundant signals (first isotopes, in-source fragmentation, K^+ and Na^+) and real peaks with signal/noise over ten were checked manually. A data matrix with about 3000 signals was obtained in this step.

To produce the maximal signal for each precursor-product ion (Q1-Q3) transition, different parameter combinations of collision energy and declustering potential, each from 10 to 60 eV at 10 eV intervals were tested under MRM mode of the LC/MS. The optimized parameters were then used to run LC/MS/MS analysis under MIM-IDA-EPI mode and data matrix was generated. The newly generated data matrix was run through the redundancy removal software (second, isotopes, in-source fragmentation, $K^+$and Na^+).

To further reduce the possible redundancy of the signal, the fact that different transitions from the same metabolite will have high correlation coefficient among different samples was used. In our study, forty different varieties were analyzed for each transition under MIM mode and correlation coefficients between transition pairs were calculated. Transitions that have the same retention time (within 0.1 minute window) and have high correlation coefficient (over 0.9 for metabolite eluted early than 2.5minute, or over 0.8 for metabolite eluted later than 2.5 minute) were regarded as possibly redundant. This was done by using the in-house-developed software (third, coefficient>0.8). The workflow of metabolite profiling and related data analysis is

shown in supplementary Fig. 12. Finally, 184 metabolites and 799 metabolite features were obtained. The mass spectra data of the metabolites detected in this paper are presented in the Supplementary data 6.

Supplementary References

1. Yao, M., Ma, L., Humphreys, W.G., and Zhu, M. Rapid screening and characterization of drug metabolites using a multiple ion monitoring-dependent MS/MS acquisition method on a hybrid triple quadrupole-linear ion trap mass spectrometer. Journal of mass spectrometry: JMS **43,** 1364-1375 (2008).

2. Dresen, S., Ferreiros, N., Gnann, H., Zimmermann, R., and Weinmann, W. 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 (2010).

3. Chen, W. et al. A novel integrated method for large-scale detection, identification and quantification of widely-targeted metabolites: application in study of rice metabolomics. *Molecular Plant*, 6, 1769-1780 (2013)..