Supporting Information

Gong et al. 10.1073/pnas.1319681110

SI Materials and Methods

Plant Materials. The mapping population consisted of 210 recombinant inbred lines (RILs) derived from a cross between ZS97 and MH63, the parents of Shanyou 63, the most widely cultivated hybrid in China. Seventy-one introgression lines (ILs) generated from the same parents as the RILs were used for validating the metabolic quantitative trait locus (mQTL) results. The rice plants examined under field conditions were grown in normal rice-growing seasons in the Experimental Station of Huazhong Agricultural University (Wuhan, China). All seeds were planted in a seed bed in mid-May, and transplanted to the field in mid-June. The planting density was 16.5 cm between plants in a row, and the rows were 26 cm apart. Field management, including irrigation, fertilizer application, and pest control, followed essentially the normal agricultural practice. Leaves of the plants were harvested for genomic DNA extraction.

Sample Preparation. The seeds of 210 RILs and the parents were first soaked in water for 48 h in a chamber (Conviron S10H; Conviron, www.conviron.com) set at 25 °C and 85% relative humidity in the dark, and then transferred to another chamber (Conviron PVG36) for pregermination (35 °C, 85% relative humidity, dark). Germination of the seeds was checked every 2 h, and germinated seeds were transferred to a growth chamber (Conviron S10H) and incubated for 72 h (25 °C, 85% relative humidity, dark). Seeds from 15 seedlings per line were bulk-harvested and frozen in liquid nitrogen for metabolite extraction. One biological replication in 2009 and one in 2010 for each RIL and three for each parent in both years were sampled, one replication at a time.

The flag leaves were harvested at heading date in 2009 using liquid nitrogen from three different plants per line grown in the field for metabolite extraction. Two biological replications for each RIL and three for each parent were sampled.

Metabolite Profiling. The freeze-dried samples were analyzed using a liquid chromatography (LC)–electrospray ionization (ESI)–MS/MS system (HPLC, Shim-pack UFLC SHIMADZU CBM20A system; MS, Applied Biosystems 4000 Q TRAP) and an Agilent 6520 accurate-mass time-of-flight mass spectrometry equipped with a dual ESI electrospray ion source in positive-ion mode. A stepwise multiple ion monitoring-enhanced product ions was used to construct the MS2T library as previously described (1). Quantification of metabolites was carried out using a scheduled multiple reaction monitoring (MRM) method (2). A total of 684 transitions in flag leaf and 318 transitions in germinating seed were monitored, respectively, with positive polarity. The scheduled MRM algorithm was used with an MRM detection window of 80 s and the target scan time of 1.5 s in Analyst 1.5 software. The extracts were absorbed and filtrated (CNWBOND Carbon-GCB SPE Cartridge, 250 mg, 3 mL; ANPEL).

Statistical Analysis. Metabolite (m-trait) data were $\log_2$-transformed for statistical analysis to improve normality. The m-trait data of the RIL population are the mean of biological replicates for the LC-MS/MS as shown below: $P_{m,l} = 1/2(P_{m,l,1} + P_{m,l,2})$, where $P_{m,l}$ represents the m-trait data for metabolite m (m = 1, 2, 3, ..., 684 in flag leaf and m = 1, 2, 3, ..., 318 in germinating seed) in RIL line l (l = 1, 2, 3, ..., 210), and $P_{m,l,1}$ and $P_{m,l,2}$ are the normalized metabolite levels determined in the two replicates, respectively. The values of genetic coefficient of variation (3) were independently calculated for each metabolite (using the mean of the biological replicates of the untransformed m-trait data) as below: σ/μ, σ and μ represent the SD and the mean of each metabolite in the population, respectively. Broad-sense heritability (H^2) (4) was calculated using the following equation by treating RILs as a random effect and the biological replication as the environmental effect: $H^2 = \mathrm{Var}_{(G)}/(\mathrm{Var}_{(G)} + \mathrm{Var}_{(E)})$, where $\mathrm{Var}_{(G)}$ and $\mathrm{Var}_{(E)}$ represent variance derived from genetic and environmental effects, respectively. Pairwise Pearson correlation between metabolites detected was estimated by R (www.r-project.org). Metabolite networks were constructed based on the correlation matrices and demonstrated by the program Cytoscape (5).

QTL Mapping. Bin maps were constructed for the 210 RILs based on individual SNPs and adjacent bins with the same genotype were lumped, resulting in a map consisting of 1,619 recombinant bins without missing data (6, 7). Composite interval mapping (CIM) (8) was performed for each metabolite using the R/qtl function cim (9) with a 10-cM scan window and covariates of five markers. The walking speed was set to zero because the bins were clearly defined, which was different from the nature of traditional molecular markers. The likelihood ratio statistic was computed for each bin. The LOD threshold was set to 3.0 for each metabolite. A 1.5 LOD-drop support interval was used for each QTL as described by Wang et al. (10). The QTL additive effect and variation explained by each QTL were determined using the linear QTL model involving all of the detected QTLs using the R function lm (www.r-project.org).

Detection of mQTL Hot Spots. Distribution of mQTLs along the genome was investigated by dividing the whole genome into 1-cM partitions, and the number of mQTLs in each segment was counted. A permutation test was used to assess the statistical significance of deviation of the observed mQTL distribution per centimorgan from the expectation based on chance events, assuming a uniform distribution throughout the genome. In the permutation, each mQTL was randomly assigned to a 1-cM interval in the map, and the resulting number of mQTLs in each interval was counted. The results of 1,000 permutations showed that, with $P < 0.01$, the cutoff number of mQTLs per centimorgan by chance alone would be eight in flag leaf and six in germinating seed, respectively, and a larger number would be regarded as a mQTL hot spot.

Analysis of Two-Locus Interactions. To identify epistatic interactions between the mQTL hotspots, bins in hotspots region were searched pairwise for interactions using two-way ANOVA (11). The bins most closely associated with each significant mQTL hotspot were used for epistasis analysis (12) against the known metabolites in flag leaf and germinating seed, respectively. The calculation was based on unweighted cell means, and the sums of squares were multiplied by the harmonic means of the cell sizes to form the test criteria. Those that showed significant interactions at $P \leq 0.01$ were subjected to permutation tests, in which the positions of the phenotype scores in the dataset were randomized to perform the two-way ANOVA again. This process was repeated 1,000 times. If no more than 1% of the random F values was larger than the F from the real data, it was regarded to be significant at $P \leq 0.01$.

Constructs and Transformation. The overexpression vector (pJC034) for rice was constructed from the gateway overexpression vector pH2GW7, with the 35S promoter of pH2GW7 replaced by maize ubiquitin promoter. The *OsMaT-2*, *OsMaT-3*, and *Os11g26950* overexpression constructs were made by directionally inserting the full cDNA sequence first into the entry vector pDONR207 and then into the destination vector pJC034 using the Gateway recombination reaction (Invitrogen) (Table S3). The constructs were independently introduced into the *Agrobacterium* strain EHA105, and transformation was done as described previously (13). For each constructs, three independent T1 progeny of overexpression plants that showed the expression level of transgene significantly ($P < 0.001$) correlated with the targeted metabolite were selected for further analysis.

Expression Analyses. We isolated total RNA from rice using an RNA extraction kit (TRIzol reagent; Invitrogen) according to the manufacturer's instructions. The first-strand cDNA was synthesized using 3 μg of RNA and 200 U of M-MLV reverse transcriptase (Invitrogen) according to the manufacturer's protocol. Real-time PCR was performed on an optical 96-well plate in an ABI Stepone plus PCR system (Applied Biosystems) by using SYBR Premix reagent F-415 (Thermo Scientific). *Actin1* was used as an endogenous control (Table S3). The expression measurements were obtained using the relative quantification method (14).

Full Names of Abbreviations of Metabolites. The full names of abbreviations of metabolites are as follows: tri *O*-malhex, tricin *O*-malonylhexoside; tri *O*-hex-*O*-hex, tricin *O*-hexosyl-*O*-hexoside; sin *O*-hex, sinapoyl *O*-hexoside; pyr *O*-hex, pyridoxine *O*-hexoside; tri *O*-hex der, tricin *O*-hexoside derivatives; api 7-*O*-rut, apigenin 7-*O*-rutinoside; suc, sucrose; 3PGA, 3-phosphoglycerate; glu, L-glutamate; asp, L-aspartate; thr, L-threonine; pyr, pyridoxine; pyr *O*-h, pyridoxine *O*-hexoside; phe, L-phenylalanine; ser, L-serine; LPCs, lysophosphatidylcholines; fer, ferulic acid; fer *O*-h, ferulic acid *O*-hexoside; sin, sinapic acid; sin *O*-h, sinapoyl *O*-hexoside; nar, naringenin; kae 3-*O*-h, kaempferol 3-*O*-hexoside; *O*-mque *O*-h, *O*-methylquercetin *O*-hexoside; *C*-h-nar *O*-couh, *C*-hexosyl-naringenin *O*-*p*-coumaroylhexoside; sel, selgin; sel *O*-h, selgin *O*-hexoside; tri, tricin; tri *O*-h, tricin *O*-hexoside; tri 4′-*O*-(R)e *O*-h, tricin 4′-*O*-(syringyl alcohol)ether *O*-hexoside or tricin 4′-*O*-(β-guaiacylglyceryl) ether *O*-hexoside; tri *O*-r, tricin *O*-rutinoside; tri *O*-h-*O*-h, tricin *O*-hexosyl-*O*-hexoside; tri *O*-malh der, tricin *O*-malonylhexoside derivatives; api, apigenin; api *O*-h, apigenin *O*-hexoside; api *O*-r, apigenin *O*-rutinoside; api *C*-h, apigenin *C*-hexoside; *C*-h-api *O*-h-*O*-h, *C*-hexosyl-apigenin *O*-hexosyl-*O*-hexoside; *C*-h-api *O*-(cou/caf)h, *C*-hexosyl-apigenin *O*-(*p*-coumaroyl/caffeoyl)hexoside; api *C*-p, apigenin *C*-pentoside; *C*-p-api *O*-r, *C*-pentosyl-apigenin *O*-rutinoside; *C*-p-api *O*-(cou/caf/fer)h, *C*-pentosyl-apigenin *O*-(*p*-coumaroyl/caffeoyl/feruloyl) hexoside; lut, luteolin; lut *O*-h, luteolin *O*-hexoside; *C*-p-lut *O*-h, *C*-pentosyl-luteolin *O*-hexoside; lut *C*-h luteolin *C*-hexoside; *C*-h-lut *O*-p, *C*-hexosyl-luteolin *O*-pentoside; *C*-h-lut *O*-couh, *C*-hexosyl-luteolin *O*-*p*-coumaroylhexoside; chr, chrysoeriol; chr *O*-h, chrysoeriol *O*-hexoside; chr *O*-malh, chrysoeriol *O*-malonylhexoside; chr *O*-r, chrysoeriol *O*-rutinoside; chr *C*-h, chrysoeriol *C*-hexoside; *C*-h-chr *O*-(cou/fer)h, *C*-hexosyl-chrysoeriol *O*-(*p*-coumaroyl/feruloyl)hexoside; *C*-p-chr *O*-ferh, *C*-pentosyl-chrysoeriol *O*-feruloylhexoside.

GenBank Accession Numbers for Phylogenetic Analysis. GenBank accession numbers are in parentheses. The following gene sequences were used for the phylogenetic analysis of *OsMaT-2* and *OsMaT-3* (Fig. 3*A*): Ss5MaT (AF405707), Pf5MAT (AF405204), NtMaT1 (AB176525), Vh3MaT1 (AY500350), Lp3MaT (AY500352), and the rice gene amino acids sequences from Rice Genome Annotation Project (http://rice.plantbiology.msu.edu/cgi-bin/gbrowse/rice/). The following gene sequences were used for the phylogenetic analysis for glucosyltransferases (Fig. S4*A*): UGT84A1 (z97339), UGT84A2 (ab019232), UGT78D1 (ac009917), UGT73C6 (ac006282), UGT79B1 (ab018115), At3RhaT (NM_102790), At3GlcT (NM_121711), At3AraT (NM_121709), Vv3GlcT (AF000371), Ph3GlcT (AB027454), Pf3GlcT (AB002818), Hv3GlcT (X15694), Zm3GlcT (X13501), At5GlcT (NM_117485), Pf5GlcT (AB013596), Ph5GlcT (AB027455), Vh5GlcT (AB013598), At7RhaT (NM_100480), At7GlcT (NM_129234), DbB5GlcT (Y18871), NtIS5a (AF346431), Gt3′GlcT (AB076697), CmF7G2″RhaT (AY048882), BpA3G2″GlcAT (AB190262), IpA3G2″GlcT (AB192315), and PhA3G2″RhaT (Z25802).

1. Chen W, et al. (2013) A novel integrated method for large-scale detection, identification and quantification of widely-targeted metabolites: Application in study of rice metabolomics. *Mol Plant*, 10.1093/mp/sst080.
2. Dresen S, Ferreirós N, Gnann H, Zimmermann R, Weinmann W (2010) Detection and identification of 700 drugs by multi-target screening with a 3200 Q TRAP LC-MS/MS system and library searching. *Anal Bioanal Chem* 396(7):2425–2434.
3. Chan EK, Rowe HC, Hansen BG, Kliebenstein DJ (2010) The complex genetic architecture of the metabolome. *PLoS Genet* 6(11):e1001198.
4. Visscher PM, Hill WG, Wray NR (2008) Heritability in the genomics era—concepts and misconceptions. *Nat Rev Genet* 9(4):255–266.
5. Smoot ME, Ono K, Ruscheinski J, Wang PL, Ideker T (2011) Cytoscape 2.8: New features for data integration and network visualization. *Bioinformatics* 27(3):431–432.
6. Yu H, et al. (2011) Gains in QTL detection using an ultra-high density SNP map based on population sequencing relative to traditional RFLP/SSR markers. *PLoS One* 6(3): e17595.
7. Xie W, et al. (2010) Parent-independent genotyping for constructing an ultrahigh-density linkage map based on population sequencing. *Proc Natl Acad Sci USA* 107(23): 10578–10583.
8. Zeng ZB (1993) Theoretical basis for separation of multiple linked gene effects in mapping quantitative trait loci. *Proc Natl Acad Sci USA* 90(23):10972–10976.
9. Broman KW, Wu H, Sen S, Churchill GA (2003) R/qtl: QTL mapping in experimental crosses. *Bioinformatics* 19(7):889–890.
10. Wang J, et al. (2010) A global analysis of QTLs for expression variations in rice shoots at the early seedling stage. *Plant J* 63(6):1063–1074.
11. Zhou G, et al. (2012) Genetic composition of yield heterosis in an elite rice hybrid. *Proc Natl Acad Sci USA* 109(39):15847–15852.
12. Rowe HC, Hansen BG, Halkier BA, Kliebenstein DJ (2008) Biochemical networks and epistasis shape the *Arabidopsis thaliana* metabolome. *Plant Cell* 20(5):1199–1216.
13. Hiei Y, Ohta S, Komari T, Kumashiro T (1994) Efficient transformation of rice (*Oryza sativa* L.) mediated by *Agrobacterium* and sequence analysis of the boundaries of the T-DNA. *Plant J* 6(2):271–282.
14. Livak KJ, Schmittgen TD (2001) Analysis of relative gene expression data using real-time quantitative PCR and the $2^{(-\Delta\Delta C(T))}$ method. *Methods* 25(4):402–408.

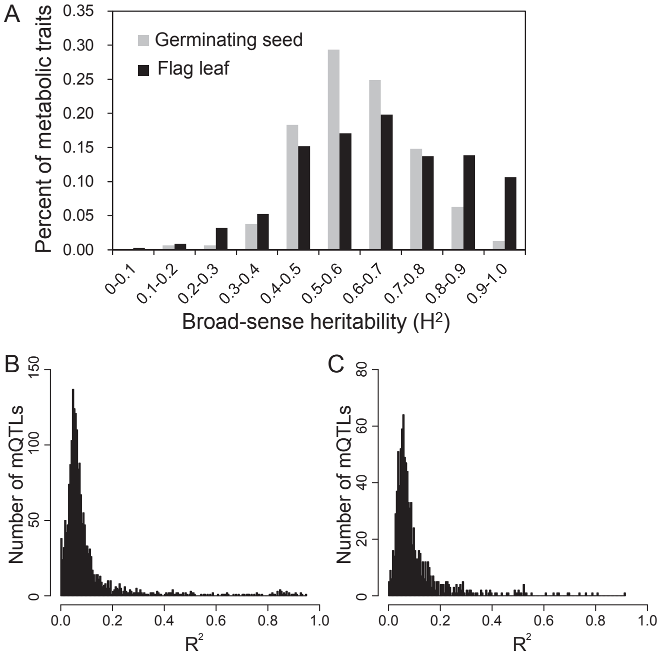

Fig. S1. Distribution of broad-sense heritability and R^2 values. (*A*) Distribution of levels of broad-sense heritability of metabolic traits. Broad-sense heritability (H^2) was estimated by considering variations between the two biological replicates as phenotypic variance derived from environmental factors in flag leaf (black) and germinating seed (gray), respectively. (*B*) The histogram of R^2 values for 1,884 mQTLs in flag leaf. (*C*) The histogram of R^2 values for 937 mQTLs in germinating seed.

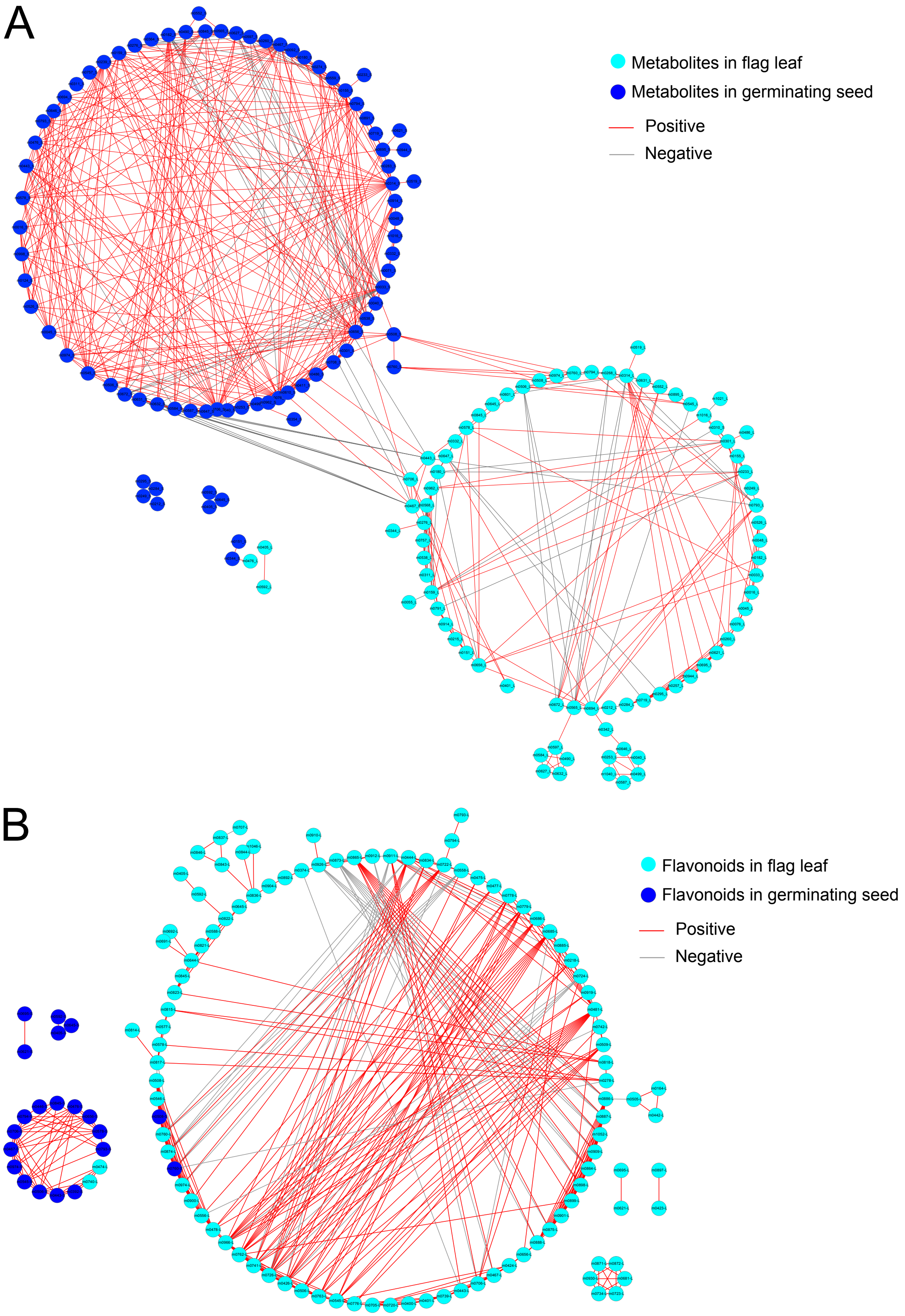

Fig. S2. Network visualization of metabolites analyzed in flag leaf and germinating seed. (*A*) Network visualization of codetected metabolites analyzed in flag leaf and germinating seed. Metabolites are represented as nodes, and their correlation coefficient values as edges. The absolute values of Pearson correlation coefficient values above the threshold ($r^2 = 0.4$) are shown. (*B*) Network visualization of flavonoids analyzed in flag leaf and germinating seed. Flavonoids are represented as nodes, and their correlation coefficient values as edges. The absolute values of Pearson correlation coefficient values above the threshold ($r^2 = 0.5$) are shown.

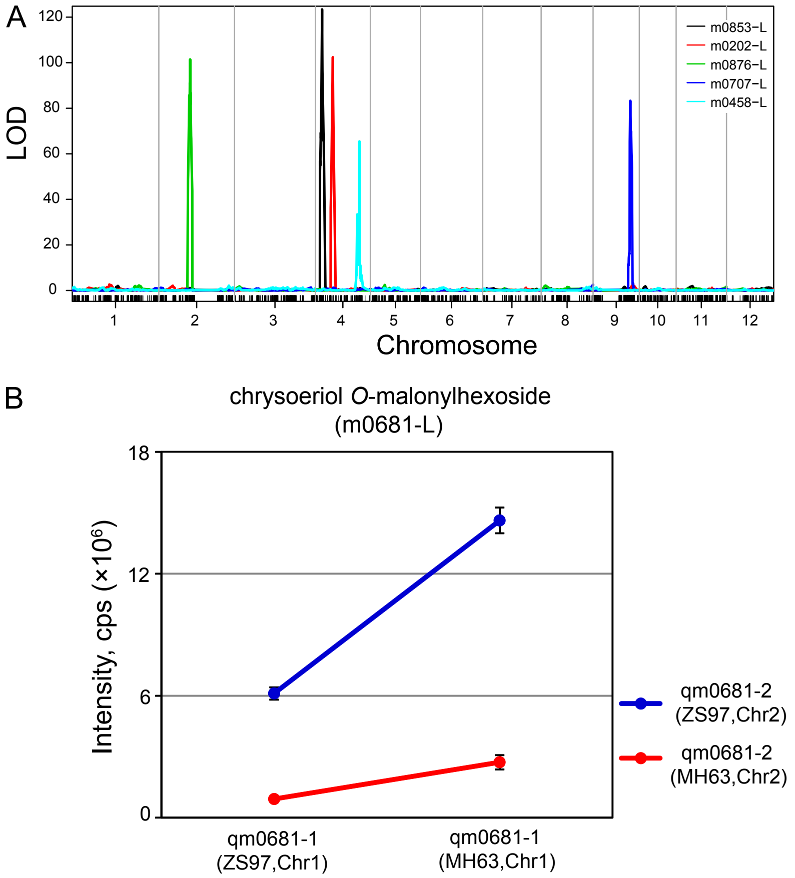

Fig. S3. Examples of metabolic quantitative trait loci (mQTLs) controlling content of metabolites. (*A*) QTL mapping results of five metabolites with one major mQTL, with each of them explaining more than 70% variation of the content. (*B*) Significant epistatic interaction of two major mQTLs controlling m0681-L (chrysoeriol *O*-malonylhexoside) accumulation.

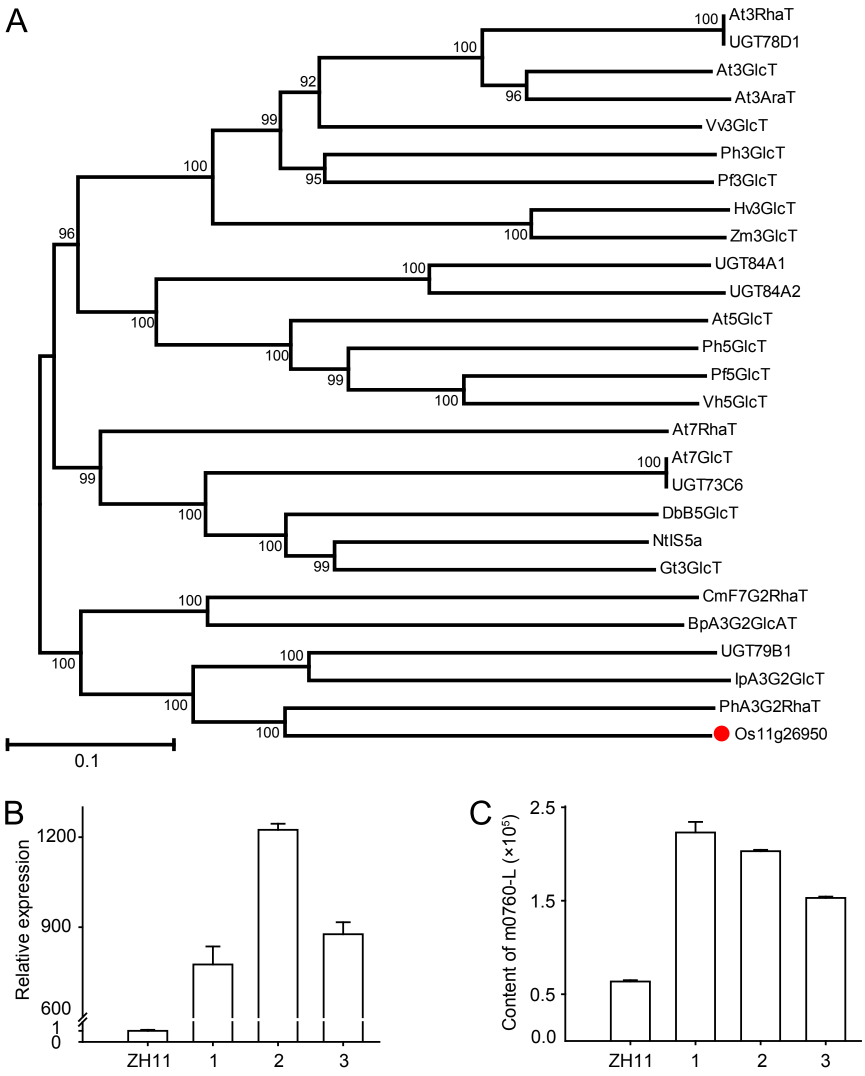

Fig. S4. Functional identification of *Os11g26950*. (*A*) Phylogenetic analysis glucosyltransferase gene in rice with glucosyltransferase genes in other species. The neighbor-joining tree was constructed using aligned full-length amino acid sequences. Bootstrap values from 1,000 replicates were indicated at each node. (Bar: 0.1-aa substitutions per site.) GenBank accession numbers are given in *SI Materials and Methods*. Bar plot for the mRNA level of *Os11g26950* (*B*) and the content of m0760-L (*C*) in rice transgenic individuals (T1), respectively. ZH11 indicates the transgenic background variety. All data are given as mean ± SEM (n = 3).

Table S1. Statistics of metabolic quantitative trait loci (mQTLs) on the chromosomes

Chr	Chromosome length, cM	Total mQTLs of flag leaf	Density of flag leaf, mQTLs/cM	Exp* of flag leaf	SR[†] of flag leaf	Total mQTLs of seed	Density of seed, mQTLs/cM	Exp* of seed	SR[†] of seed
1	200.6	326	1.6	233	6.1	111	0.6	116	−0.4
2	175.4	152	0.9	203	−3.6	61	0.3	101	−4.0
3	187.5	113	0.6	217	−7.1	53	0.3	108	−5.3
4	127.2	151	1.2	147	0.3	38	0.3	73	−4.1
5	116.0	60	0.5	134	−6.4	215	1.9	67	18.1
6	144.4	300	2.1	167	10.3	200	1.4	83	12.8
7	135.4	231	1.7	157	5.9	72	0.5	78	−0.7
8	120.4	100	0.8	140	−3.3	35	0.3	69	−4.1
9	107.2	108	1.0	124	−1.5	42	0.4	62	−2.5
10	85.3	135	1.6	99	3.6	32	0.4	49	−2.4
11	116.7	123	1.1	135	−1.1	50	0.4	67	−2.1
12	109.3	85	0.8	127	−3.7	28	0.3	63	−4.4
Total	1,625.5	1,884	1.2	1,884		937	0.6	937	

*Expected number of mQTLs based on chromosome size. $\chi^2 = 322.92$ ($P < 2.2e^{-16}$) in flag leaf and $\chi^2 = 605.38$ ($P < 2.2e^{-16}$) in seed for the test of goodness-of-fit between the observed and expected numbers of mQTLs on the 12 chromosomes.

[†] SR: standardized residue [=(observed − expected)/$\sqrt{}$expected], which follows a normal distribution asymptotically. Thus, an absolute SR value larger than 2.33 indicates statistical significance at $P < 0.01$. A positive value indicates that the observed number is greater than expected.

Table S2. Flavonoid pathway correlated coexpressed genes with *Os05g48010*

Correlation*	MSU_locus	Annotation
0.86	LOC_Os02g26810	Transcinnamate 4-monooxygenase, putative, expressed
0.76	LOC_Os08g34790	4-Coumarate-CoA ligase 2, putative, expressed
0.75	LOC_Os10g41020	Flavonol synthase/flavanone 3-hydroxylase, putative, expressed
0.75	LOC_Os02g41670	Phenylalanine ammonia-lyase, putative, expressed

*Data from Collection of Rice Expression Profiles (http://crep.ncpgr.cn/crep-cgi/home.pl).

Table S3. Primers used in this study

Primer name	Sequence	Purpose
OsMaT-2OXF	5′-attB1-ATGGCGCCCGCGACACAA-3′	Vector construction
OsMaT-2OXR	5′-attB2-CTACGCCGGGGAGTGGCC-3′	Vector construction
OsMaT-3OXF	5′-attB1-AGACCATGGCGCCGCCAC-3′	Vector construction
OsMaT-3OXR	5′-attB2-CACGCTAGTTGCATTGGGAAGA-3′	Vector construction
Os11g26950-OXF	5′-attB1-CCGTTCACTGCCCTCGAT-3′	Vector construction
Os11g26950-OXR	5′-attB2-GCGTGACGTTCCGTTTTCAG-3′	Vector construction
oja703	5′-CCTTCATACGCTATTTATTTGCTTG-3′	Positive test
OsMaT-2F	5′-AGGTGGACGTCGTGTCCGTG-3′	Expression analysis
OsMaT-2R	5′-GAACCTCTCCATCCGCTCCG-3′	Expression analysis
OsMaT-3F	5′-ACGCTCATCCGCGACGTA-3′	Expression analysis
OsMaT-3R	5′-GGCGCCTTGAACAGATCTTT-3′	Expression analysis
Os11g26950-F	5′-AATGGCGGGAGTTCTTGATG-3′	Expression analysis
Os11g26950-R	5′-TCAGCCCTTGAGCCTTCT-3′	Expression analysis
Actin1F	5′-TGGCATCTCTCAGCACATTCC-3′	Expression analysis
Actin1R	5′-TGCACAATGGATGGGTCAGA-3′	Expression analysis

Dataset S1. Metabolite reporting checklist and recommendations for liquid chromatography–mass spectrometry (LC-MS)

Dataset S1

Dataset S2. Widely targeted metabolites and metabolic quantitative trait loci (mQTLs) results in flag leaf and germinating seed

Dataset S2

Metabolite indicates the number of metabolite. Q1, precursor ions. Q3, product ions. Time, retention time. MH/ZS, the ratio of intensity between Minghui 63 and Zhenshan 97. Chr indicates chromosomal location of the mQTL for the metabolite. LOD, the LOD score of the mQTL. Var, the amount of intensity variation of the metabolite explained by the mQTL. Add, additive effect, the positive value indicates that the allele from Minghui 63 increases phenotypic value. Inf.cM, the genetic position of the inferior support interval bound in centimorgans on each chromosome. Peak.cM, the peak of genetic position. Sup.cM, the superior support interval bound. Inf.Mb, the physical position of the inferior support interval bound in Mb on each chromosome. Peak.Mb, the peak of physical position. Sup.Mb, the superior support interval bound.

Dataset S3. Results of analysis of two-locus interactions

Dataset S3

Dataset S4. Genotype and profiles of 64 metabolites of the 71 introgression lines (ILs) and the parental lines

Dataset S4

A, Zhenshan 97 genotype; B, Minghui 63 genotype; H, heterozygous genotypes.

Dataset S5. The candidate gene list for metabolic quantitative trait loci (mQTLs) results

Dataset S5