Metabolomics-driven gene mining and genetic improvement of tolerance to salt-induced osmotic stress in maize

Xiaoyan Liang[1]*, Songyu Liu[2]*, Tao Wang[1]*, Fenrong Li[1], Jinkui Cheng[1,3], Jinsheng Lai[1,2,3], Feng Qin[1,3] (iD), Zhen Li[1], Xiangfeng Wang[2] (iD) and Caifu Jiang[1,3] (iD)

[1]State Key Laboratory of Plant Physiology and Biochemistry, College of Biological Sciences, China Agricultural University, Beijing, 100094, China; [2]Laboratory of Agrobiotechnology and National Maize Improvement Center of China, Department of Agronomy and Biotechnology, China Agricultural University, Beijing, 100193, China; [3]Center for Crop Functional Genomics and Molecular Breeding, China Agricultural University, Beijing, 100094, China

Authors for correspondence:
Zhen Li
Email: lizhenchem@cau.edu.cn

Xiangfeng Wang
Email: xwang@cau.edu.cn

Caifu Jiang
Email: cfjiang@cau.edu.cn

Summary

- The farmland of the world's main corn-producing area is increasingly affected by salt stress. Therefore, the breeding of salt-tolerant cultivars is necessary for the long-term sustainability of global corn production.
- Previous studies have shown that natural maize varieties display a large diversity of salt tolerance, yet the genetic variants underlying such diversity remain poorly discovered and applied, especially those mediating the tolerance to salt-induced osmotic stress (SIOS). Here we report a metabolomics-driven understanding and genetic improvement of maize SIOS tolerance.
- Using a LC-MS-based untargeted metabolomics approach, we profiled the metabolomes of 266 maize inbred lines under control and salt conditions, and then identified 37 metabolite biomarkers of SIOS tolerance (METO1-37). Follow-up metabolic GWAS (mGWAS) and genotype-to-phenotype modeling identified 10 candidate genes significantly associating with the SIOS tolerance and METO abundances. Furthermore, we validated that a citrate synthase, a glucosyltransferase and a cytochrome P450 underlie the genotype–METO–SIOS tolerance associations, and showed that their favorable alleles additively improve the SIOS tolerance of elite maize inbred lines.
- Our study provides a novel insight into the natural variation of maize SIOS tolerance, which boosts the genetic improvement of maize salt tolerance, and demonstrates a metabolomics-based approach for mining crop genes associated with this complex agronomic trait.

Introduction

Maize (*Zea mays* ssp. *mays*) is one of the major cereal crops feeding the increasing number of people on the earth. However, although the estimated demand for corn products will double by 2050 (Schnable, 2015), corn production is increasingly challenged by a wide range of environmental stresses (Zuo *et al.*, 2015; Wang *et al.*, 2016; Yang and Guo, 2018; Li *et al.*, 2019; C. Zhang *et al.*, 2019; M. Zhang *et al.*, 2019; Liu *et al.*, 2020). Soil salinity is one of the major abiotic stresses restricting maize production, affecting millions of hectares of corn-producing lands worldwide (Hanks *et al.*, 1987; Fu, 1983; Egamberdieva *et al.*, 2019). To secure the sustainability of global corn production, there is an urgent need to understand the molecular mechanisms of maize salt tolerance, for example, the identification and application of quantitative trait loci (QTLs) associated with the key salt-tolerant physiological processes. Previous studies have demonstrated that high concentrations of soil salts damage crops

by causing ion (mainly sodium (Na^+)) toxicity and osmotic stress (Ren *et al.*, 2005; Munns & Tester, 2008), hence the identification of QTLs associated with Na^+ homeostasis and the tolerance to salt-induced osmotic stress (SIOS) are the key tasks of improving crop salt tolerance. However, whilst decades of effort have identified a number of salt-tolerant QTLs regulating Na^+ homeostasis in maize and other cops (Ren *et al.*, 2005; Munns *et al.*, 2012; Yang *et al.*, 2014; Zhang *et al.*, 2018; Wang *et al.*, 2021), the QTLs underlying the natural variation of SIOS tolerance in major crops (including maize) remain poorly understood, largely as a consequence of the complexity of SIOS response and the difficulty of quantifying SIOS tolerance (Munns & Tester, 2008; Ismail & Horie, 2017). To overcome these difficulties, the development of accurate phenotyping methods and sophisticated gene mining strategies are highly desirable.

A metabolome comprises the end products of gene expression (Ghatak *et al.*, 2018; Wang *et al.*, 2019). In recent years, the development of cost-effective metabolomics profiling strategies and large-scale genotyping-by-sequencing (GBS) of natural populations have enabled metabolome-based genome-wide

*These authors contributed equally to this work.

association study (mGWAS), which is emerging as a promising approach to disclose the genotype–metabolite–phenotype relationship in crops (Jiao *et al.*, 2012; Zhu *et al.*, 2018; Xu *et al.*, 2019). Metabolite biomarkers can provide sensitive chemical fingerprints for various complex traits (e.g. fruit flavor, stress tolerance), making the metabolite biomarker-based GWAS to be a cutting-edge approach for targeted mining of QTLs associated with complex agronomic traits (Weckwerth, 2003; Weckwerth *et al.*, 2004; Nakabayashi & Saito, 2015; Zhu *et al.*, 2018; Zhu *et al.*, 2019; Weckwerth *et al.*, 2020). Given the fact that SIOS tolerance is a complex trait, and that metabolites can act as biomarkers of complex traits, here we employed a metabolite biomarker-based approach to investigate the natural variations of SIOS tolerance in maize (Fig. 1a). We conducted a comprehensive metabolomics profiling of 266 maize inbred lines under control and salinity conditions, then characterized 37 metabolite biomarkers of SIOS tolerance (termed METO), before identifying a set of single nucleotide polymorphisms (SNPs) and candidate genes associated with SIOS tolerance by METO-based mGWAS, which enables the modeling of SIOS tolerance. We also experimentally validated that a citrate synthase, a glucosyltransferase and a cytochrome P450 confer the genotype–METO–SIOS tolerance associations, and demonstrated that the favorable alleles of these genes have practical implications for breeding maize with salt (SIOS) tolerance. These observations provided novel insights into the understanding and genetic improvement of crop salt (SIOS) tolerance, that support a proof-of-concept case for using the metabolite biomarker-based approach to understand and improve complex multigenic traits in crops.

Materials and Methods

Plant growth and sample collection

Pots of 10 cm diameter and 12 cm high were filled with uniformly mixed Pindstrup substrate (www.pindstrup.com) and watered to soil saturation with 100 mM NaCl solution. Eight seeds of the same inbred line were planted in each pot, grown in a glasshouse for 12 d. For each sample, shoot tissues from five seedlings were pooled, snap-frozen in liquid N_2, and stored at −80°C for later metabolomics profiling. All samples were collected in a 3-h time window between 09:00 and 12:00 h.

Extraction of metabolites

The materials were ground into powder in liquid N_2, and partitioned into two sample sets for the extraction of water-soluble and lipid-soluble metabolites. For each extraction, 100 mg of powder was weighed and transferred to a 1.5-ml tube, to which 1 ml of methanol (for lipid-soluble metabolites) or 75% (v/v) methanol (for water-soluble metabolites) were added. Then the tube was placed in an ultrasonic cleaner for 30 min, and then centrifuged at −21 130 *g* and 4°C for 10 min. The lipid-soluble and water-soluble supernatants were combined (1 : 1, v/v), and 0.6 ml of the mixture was dried and reconstituted in 80 μl of

50% (v/v) methanol and filtered through a 0.1-mm membrane, in preparation for LC-high resolution (HR)MS analysis.

Metabolomics profiling and data processing

We used an UPLC-HRMS system (UPLC, Acquity I-Class (Waters Corp., Milford, MA, USA); HRMS, Q-Exactive Focus (Thermo Fisher Scientific, Waltham, MA, USA) equipped with a heated electrospray ionization source to profile the metabolomics. The MS analysis was performed in the positive ion mode. We obtained scans in the mass range of 70–1000 *m/z*, at three scans per second with a resolution of 70 000. For the MS/MS assay, a normalized collision energy of 35 V, an isolation window of 0.8 *m/z*, and a mass resolution of 35 000 were used. Data acquisition from the raw MS output files were achieved using XCALIBUR software (Thermo Fisher Scientific). Progenesis QI (Waters) was used to extract the mass spectral features, which is very efficient in peak picking. Principal component analyses (PCA) were performed using SIMCA-P 13 software (Umetrics). A PLS-DA (partial least squares discriminate analysis) model was used for the calculation of variable importance in projection (VIP) values. Annotated compounds were identified by searching the accurate mass of the molecular ions and the fragment ions against compound databases, including Metabolite and tandem MS Database (METLIN), Plant Metabolic Pathway Databases (PlantCyc) and Arabidopsis Metabolic Pathway Databases (AraCyc), as well as the published literatures (Wen *et al.*, 2014, 2016; Chen *et al.*, 2016; Xu *et al.*, 2019; Zhou *et al.*, 2019). The identified compounds were classified according the criterion of HMDB or Lipidmap databases.

Measurement of leaf water content

In order to measure the leaf water content (WC), plots with 10 cm square length and 12 cm height were filled with uniformly mixed culture substrate. Eight plants were planted in each pot, grown for 7 d, and then watered to soil saturation with 300 mM NaCl (treatment) or water (control). The leaf or shoot tissue were harvested, weighed (FW), dried at 80°C for 24 h to constant weight, weighed again to obtain the DW, and then the water contents (expressed as %FW) were calculated by (FW − DW)/FW×100%.

Characterization of metabolite biomarker of SIOS tolerance (METO)

First, the WC of the 1st and 2nd leaves were transformed to normal distribution using the *Box-cox* function of the PYTHON/SCIPY package (Virtanen *et al.*, 2020), and then the SIOS sensitivity of the inbred lines was categorized into three classes based on leaf water content, using K-means clustering implemented in PYTHON/SCIKIT-LEARN (K-means function, n_clusters = 3, random_state = 3). The inbred lines with high, low and middling WC were designated as SIOS-sensitive, SIOS-tolerant and moderately salt-tolerant classes, respectively. Second, the maximum mutual information coefficient (MIC) between metabolite abundance and SIOS sensitivity was calculated by PYTHON/MINEPY.
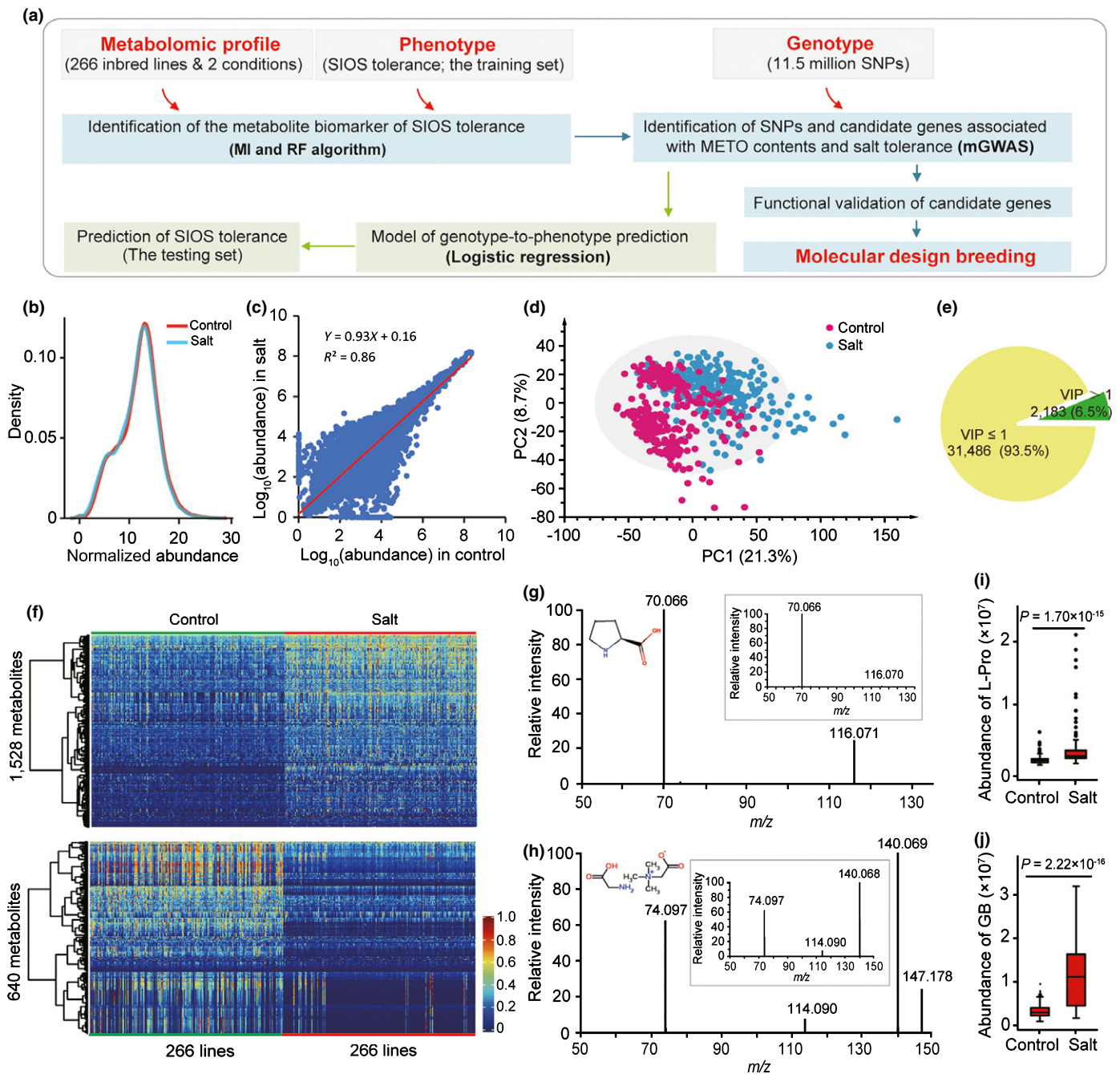
Fig. 1 Study design and metabolomics profiling of 266 maize inbred lines. (a) Schematic overview of this study. SIOS, salinity-induced osmotic stress; MI, mutual information; RF, random forest. (b, c) Distribution (b) and correlation (c) of the normalized abundances of 33 619 metabolites identified in the control and salt-treated samples. (d) Principal components analysis (PCA) with the metabolomes of 266 maize inbred lines grown under control (carmine) and saline (blue) conditions. (e) Pie chart showing the numbers and percentages of metabolites with indicated variable importance in projection (VIP). (f) Hierarchically clustering heatmap of the levels of salt-responsive (up- or downregulated) metabolites in 266 inbred lines. (g, h) MS/MS fragmentation profiles of proline (g) and glycine betaine (h) annotated in our study. The insets are profiles of the reference substance. (i, j) The relative abundances of proline (i) and glycine betaine (j) in the control and salt-treated samples. Statistical significances in (i) and (j) were determined by a two-sided Student's *t*-test.

The script for MIC calculation can be accessed at GitHub (https://github.com/liusy-jz/MIC_calc). Third, the random forest classification model in PYTHON/SCIKIT-LEARN was trained using the abundances of the top 200 MIC metabolites 1000 times to generate a 1000×200 metabolite feature importance matrix. Finally, the 1000×200 feature importance matrix was used to cluster the metabolites using the K-means clustering method of PYTHON/SCIKIT-LEARN, hence identified the metabolite with the highest feature importance (termed metabolite biomarker of SIOS tolerance, METO). PCA analysis of the identified METOs was conducted using the scikit-learn package and visualized by R/SCATTERPLOT3D (Ligges & Maechler, 2003).

mGWAS analysis

The metabolite-(METO-)based genome-wide association study (mGWAS) was performed using the 11 655 404 SNPs identified by genotyping-by-sequencing (GBS) across the 266 inbred lines with metabolomics profile. The abundances of the 37 METOs in control and salt samples were used to perform the mGWAS. A kinship matrix was used to reflect the family relatedness across the population, and the principal components were adjusted for covariates to control the effects of population structure. A mixed-model approach implemented in Gemma was established for the GWAS analysis (Zhou & Stephens, 2012). A P-value cut-off was computed by $0.05/n$ ($n = 11\,655\,404$), and we obtained the $-\log_{10}(P)$ threshold of 8.4 as the final significance value.

Logistic regression (LR) model of SIOS tolerance

We identified 22 peak SNPs from the results of the METO-based mGWAS analysis. With these peak SNPs as the genotype input (randomly selected SNP as negative control), we used the LR model in PYTHON/SCIKIT-LEARN to build the prediction model (penalty = 'l2'; tol = 0.0001; C = 1.0; max_iter = 100). In the training process, the WCs of the training set samples were used, and the area under receiver operating characteristic curve (AUROC) value was used to evaluate the accuracy of the model, and the stability of the model was evaluated through the five-fold cross-validation method (Pedregosa et al., 2011), and the robustness of the model was evaluated using the genotypes and WCs of the testing set samples.

RR-BLUP- and CropGBM-based prediction of SIOS tolerance

In order to predict the SIOS tolerance (leaf WC) through RR-BLUP (Endelman et al., 2011), 100 000 randomly selected SNPs were used. The prediction effectiveness (information gain, IG) for 22 peak SNPs were derived from CropGBM, which can be accessed at https://github.com/YuetongXu/CropGBM. The 10 SNPs with prediction effectiveness greater than zero were retained and then one-by-one added into the prediction model in descending order of prediction effectiveness to predict leaf WC. The prediction accuracy is calculated as the Pearson correlation coefficient between the predicted and observed leaf WC in test population. In the meantime, 100 times of predictions with 22 randomly selected SNPs were used to calculate the Pearson correlation as random prediction baseline. The prediction accuracy of baseline was from the first to third quartile value of 100 predictions accuracy.

Generation of the ZmSOT-overexpressing lines

The coding sequences of maize sulfotransferases (*ZmSOT*)5 (citrate synthase *ZmCS3*), *ZmSOT7* (glucosyltransferase *ZmUGT*), *ZmSOT9* (cytochrome P450 polypeptide *ZmCYP93D1*) and *ZmSOT10* (*ZmCYP709B2*) were cloned to PBCXUN vector using the primers listed in Supporting Information Table S13

(see later). The constructs were transformed into *Agrobacterium* strain EHA105, and then into immature embryo to regenerate T_1 seedlings. The homozygous overexpression lines were obtained by antiherbicide selection of the self-pollinated T_1, T_2 and T_3 generation plants.

Generation of CRISPR-Cas9 lines

CRISPR-Cas9 knockout lines of *ZmCS3*, *ZmCYP93D1* and *ZmCYP709B2* were generated according to previous reports (Xing et al., 2014; Cao et al., 2019). We designed gRNAs using CRISPR-P (http://crispr.hzau.edu.cn/CRISPR2/). pCAMBIA-derived CRISPR/Cas9 binary vector with two gRNA expression cassettes targeting *ZmCS3*, *ZmCYP93D1* or *ZmCYP709B2* was generated using the primers listed in Table S13. The transformation of immature embryo and the regeneration of T_1 plants were as described above. The mutant lines were confirmed by sequencing the genomic fragment covering the gRNA-targeted sites of T_1 and T_2 plants.

RNA extraction and quantitative reverse transcription (qRT)-PCR

In order to analyze the transcript levels of *ZmCS3*, *ZmUGT*, *ZmCYP709B2* and *ZmCYP93D1* in the CRISPR-Cas9 lines, *ZmSOT* overexpressing lines and selected inbred lines, plants were grown under indicated conditions for 2 wk, shoot tissue from five seedlings were pooled, ground into powder in liquid N_2, and then total RNA was extracted using RNA prep pure plant kit (Tiangen Biotech, Beijing, China). Two micrograms of RNA were used to synthesize cDNA using M-MLV Reverse Transcriptase, and then qRT-PCR analyses were conducted using the PowerUp^{TM} SYBR Green Master Mix (Applied Biosystems, Carlsbad, CA, USA). *ZmUbi2* provided a control. The primers were shown in Table S13. A $2^{-\Delta\Delta Ct}$ based calculation was used to calculate the transcript level.

Results

Metabolomics profiling of 266 maize lines under control and salinity conditions

Natural maize varieties confer a rich genetic diversity of salt tolerance, which is due mostly to the functional variation of genes mediating SIOS tolerance and Na$^+$ homeostasis (Saneoka et al., 1995; Luo et al., 2019; C. Zhang et al., 2019; M. Zhang et al., 2019; Sandhu et al., 2020). Our previous studies showed that maize *ZmNC1* (*Na$^+$ Content 1*), *ZmNC2* and *ZmNSA1* (*Na$^+$ Content under Saline-Alkaline Condition*) underlie the natural variation of the shoot Na$^+$ exclusion, a key process protecting shoot tissue from Na$^+$ toxicity (Zhang et al., 2018; C. Zhang et al., 2019; M. Zhang et al., 2019; Cao et al., 2020). In this study, we aimed to use metabolites as biomarkers to investigate the genetic basis underlying the natural variation of SIOS tolerance in maize (Fig. 1a). First, 266 representative inbred lines were selected randomly from a maize population that was genotyped

in our previous studies (Jiao *et al.*, 2012; C. Zhang *et al.*, 2019; M. Zhang *et al.*, 2019) (Table S1), the seedlings were grown for 10 d under control and salinity conditions, and then their metabolome were profiled using a LC-MS-based untargeted metabolomics approach in the positive ion mode (see the Materials and Methods section). A total of 33 619 distinct mass spectral features were detected and relatively quantified (Table S2; Fig. S1). The overall abundance distributions of the identified metabolites were comparable between the control and salt-treated samples (Fig. 1b), and most of the metabolites were detected under both conditions (Fig. 1c). We performed a PCA with all metabolites, and observed that the control and salt samples were significantly differentiated (Fig. 1d). A total of 2183 (6.5%) metabolites showed a VIP > 1.0 (Fig. 1e), among which 1528 and 640 metabolites were up- and downregulated by salt stress ($P < 0.01$), respectively (Fig. 1f). We unraveled 358 mass spectral features showing an *m/z* comparable with the metabolites annotated by previous studies (Wen *et al.*, 2016; Chen *et al.*, 2016; Xu *et al.*, 2019; Zhou *et al.*, 2019), among which 242 were confirmed by the MS/MS fragmentation pattern (Table S3). These annotated metabolites covered a wide range of metabolite classes, including flavonoids, amino acids, lipids, benzoxazinoids, terpenoids and so on (Table S3). To exemplify the quality of our data, we showed that the proline and glycine betaine annotated in this study showed MS/MS fragmentation patterns comparable with that of the reference substances under the same MS setting (Fig. 1g,h), and the abundances of these well-known osmoregulators in the salt-treated samples were significantly higher than those of the control (Fig. 1i,j).

Identification of the metabolite biomarkers of SIOS tolerance

The genetic basis underlying the variation of crop SIOS tolerance remains largely unknown, due to the difficulty of quantifying this trait. As SIOS decreases the external osmotic potential and then leads to difficulty of water uptake and physiological drought, the SIOS tolerance could be associated with an enhanced root water uptake or a reduced leaf water evaporation, and to some extent both can be reflected by the leaf WCs of the plants grown under salt environments. Intriguingly, we found that some maize inbred lines maintained greater leaf WCs than the others when subjected to treatment with 300 mM NaCl solution (Fig. 2a); in many cases, this phenomenon was irrelevant to Na^+ contents but, rather, reflected the variation of SIOS tolerance (Fig. S2), indicating that the leaf WC provides a practical measure of SIOS tolerance. We then applied this leaf WC-based assay of SIOS tolerance to a maize population composed of 404 lines and observed large variations of leaf WCs (Fig. 2b,c), and found that the older leaves (1st leaf) showed greater decrease of WCs than the younger leaves (2nd leaf) following the onset of salt (300 mM NaCl) treatment (Fig. 2d). These observations indicated that the natural maize population conferred significant diversities of SIOS tolerance.

We next aimed to identify the metabolite biomarkers of SIOS tolerance. To do so, the 404 lines were categorized into three classes (SIOS-sensitive, moderately SIOS-tolerant and SIOS-tolerant) according to their leaf WCs by a K-means-based clustering analysis (Fig. 2e; see the Materials and Methods section). The inbred lines were divided into training and testing sets according to the presence or absence of the metabolomics profile (Fig. 2e). With the metabolomics data and the SIOS-tolerant grade of the training set samples, we used the maximum MIC analysis to preliminarily screen the metabolites highly associated with leaf WC, and independently identified 200 candidate metabolites from the control and salt datasets (Table S4). Next, we used the random forest algorithm to classify each of the 200 candidate metabolites (see the Materials and Methods section), and subsequently identified 37 metabolites that showed most association with leaf WC (16 from control, 22 from salt, with one overlapped) (Fig. 2f; Table S5). These metabolites were designated as metabolite biomarker of SIOS tolerance 1 (METO1) to METO37, with most of them (32 of 37) being salt-responsive (up- or downregulated by salinity) (Fig. S3). The METOs identified from the control or salt dataset respectively can to some extent discriminate between the SIOS-sensitive and SIOS-tolerant lines, and the 37 METOs together substantially improved the accuracy of the discrimination (Fig. 2f). In addition, PCA with the 37 METOs significantly differentiated the SIOS-sensitive and SIOS-tolerant inbred lines (Fig. 2g). These results indicated that the identified METOs can act as the metabolite biomarkers of maize SIOS tolerance.

Characterization of the candidate genes influencing the METO abundance and SIOS tolerance

The maize population used in this study previously was genotyped by genotyping-by-sequencing (GBS), identifying a total of 11 803 847 SNPs spanning the whole genome (Figs S4, S5; Table S6). The 266 inbred lines with metabolomics profiles were distributed evenly across the main branches of the neighbor-joining tree, and captured 98.74% (11 655 404) of the total numbers of the SNPs (Figs S4, S5). With the genotypes for the 266 inbred lines, we conducted mGWAS assay and identified 22 peak SNPs significantly associated with the abundances of the METOs ($-Log_{10}(P) > 8.4$) (Fig. 3a; Table S7). The majority of the peak SNPs (16 of 22) were identified under both control and salt conditions, with five and one identified only under salt and control conditions, respectively (Fig. 3a). In addition, whereas 13 of the peak SNPs were associated with the abundance of only one METO, we observed that nine of them were associated with the abundances of multiple (up to four) METOs (Table S7).

We next determined whether the METO-associated peak SNPs enable the prediction of SIOS tolerance. With the genotype and SIOS-tolerance grade for the training set lines, we used five-fold cross validation to split the training set five times into a modeling set and validation set, then used logistic regression model with AUROC metrics to assess the capacity of the peak SNPs for predicting SIOS tolerance (see the Materials and Methods section). The results showed that the AUROC in the test (with the 22 peak SNPs) and control (with 22 randomly selected SNPs) assay were 0.81 and 0.59, respectively (Fig. 3b,c); this indicated that the peak SNPs enable the prediction of SIOS
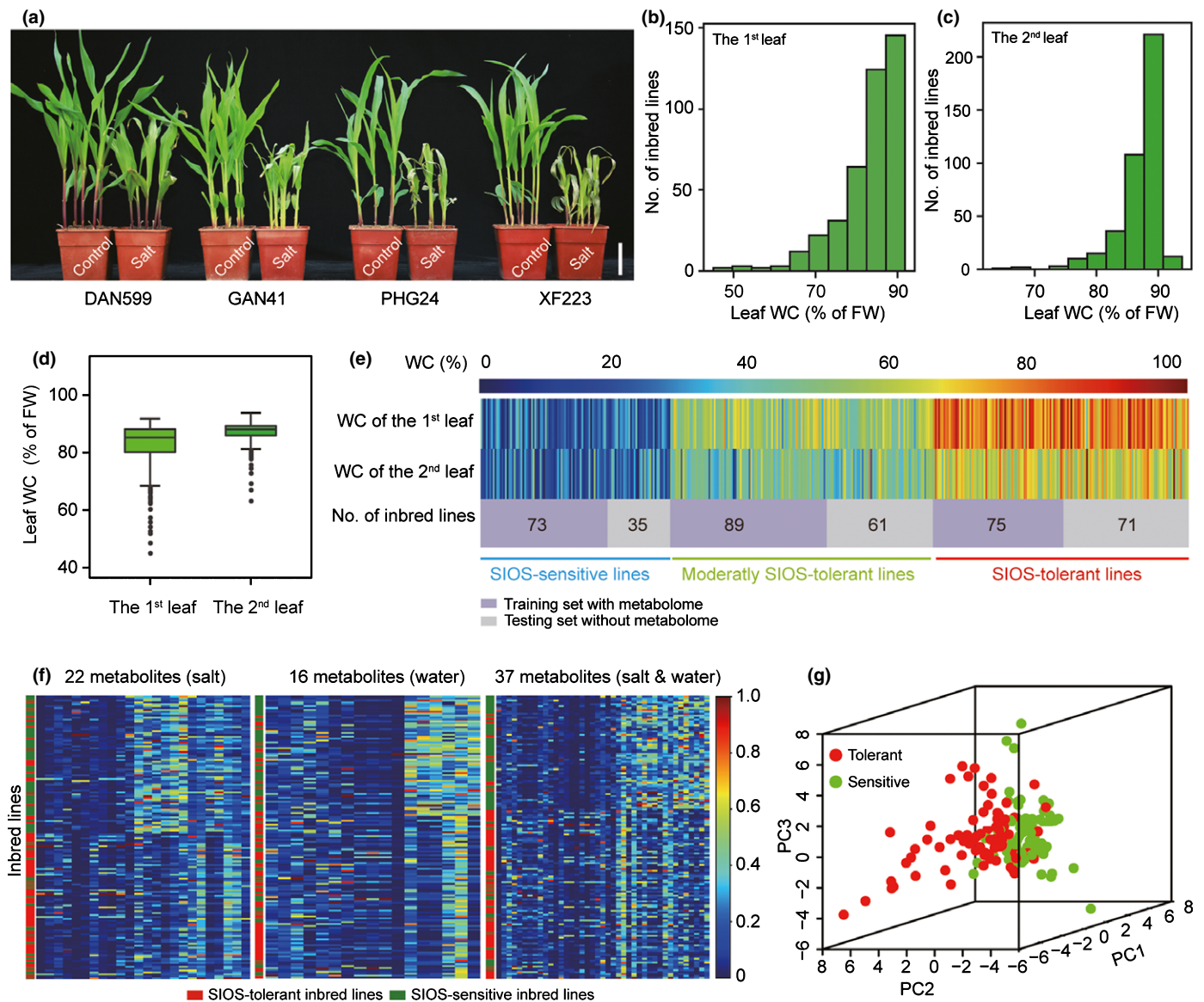
Fig. 2 Identification of metabolite biomarkers of salinity-induced osmotic stress (SIOS) tolerance (METO) in maize. (a) The appearances of the control and salt-treated maize seedlings (genotypes as indicated). 10-d-old seedlings were treated with 300 mM NaCl for 5 d, and then photographed. Bar, 5 cm. (b–d) Distribution of the water content (WC) in the 1st and 2nd leaf across 404 maize inbred lines. The leaf WCs are expressed as %FW. (e) Leaf WC-based grading of the SIOS sensitivity of 404 maize inbred lines. The numbers of lines in each group were shown. (f) K-means clustering heatmap of the abundances of SIOS-tolerance associated metabolites. (g) Principal components analysis (PCA) of maize inbred lines according to the 37 METOs.

tolerance. We then determined the feasibility of making predictions of SIOS tolerance using the testing set samples ($n = 167$) (Fig. S4) (see the Materials and Methods section). With the same logistic regression model, we obtained a predicted average AUROC of 0.79 that was comparable with that of the training set (Fig. 3b), thus provided an independent validation of the prediction model of maize SIOS tolerance.

We next analyzed the prediction effectiveness of the 22 peak SNPs in terms of SIOS tolerance. The software CROPGBM can infer the predictive effectiveness for each SNP and then predict phenotype by adding input SNP one-by-one, ordered by score of each SNP (see the Materials and Methods section). The 10 SNPs with prediction effectiveness greater than zero were retained by the CropGBM model, and the prediction accuracy (expressed as correlation between predicted and observed leaf WC) is 0.57

(Fig. 3d) (see the Materials and Methods section). The prediction accuracy was comparable with the prediction accuracy (0.6) of RR-BLUP using 100 000 SNPs (genome-wide prediction baseline), but significantly greater than the upper bound (0.4) of the random SNP-based prediction (random prediction baseline) (Fig. 3d). Moreover, we observed that the leaf WC predicted by 10-SNP-based CropGBM model and by 100 000-SNP-based RR-BLUP model were significantly correlated ($R^2 = 0.45$). These observations indicated that the 10 SNPs retained by CropGBM were significantly associated with SIOS tolerance, and this was further supported by the observation that the leaf WCs of the inbred lines with different haplotypes at each of these 10 SNPs showed significant differences (Fig. S6). We designated the genes underlying this SNP-SIOS tolerance association as *SIOS tolerance*
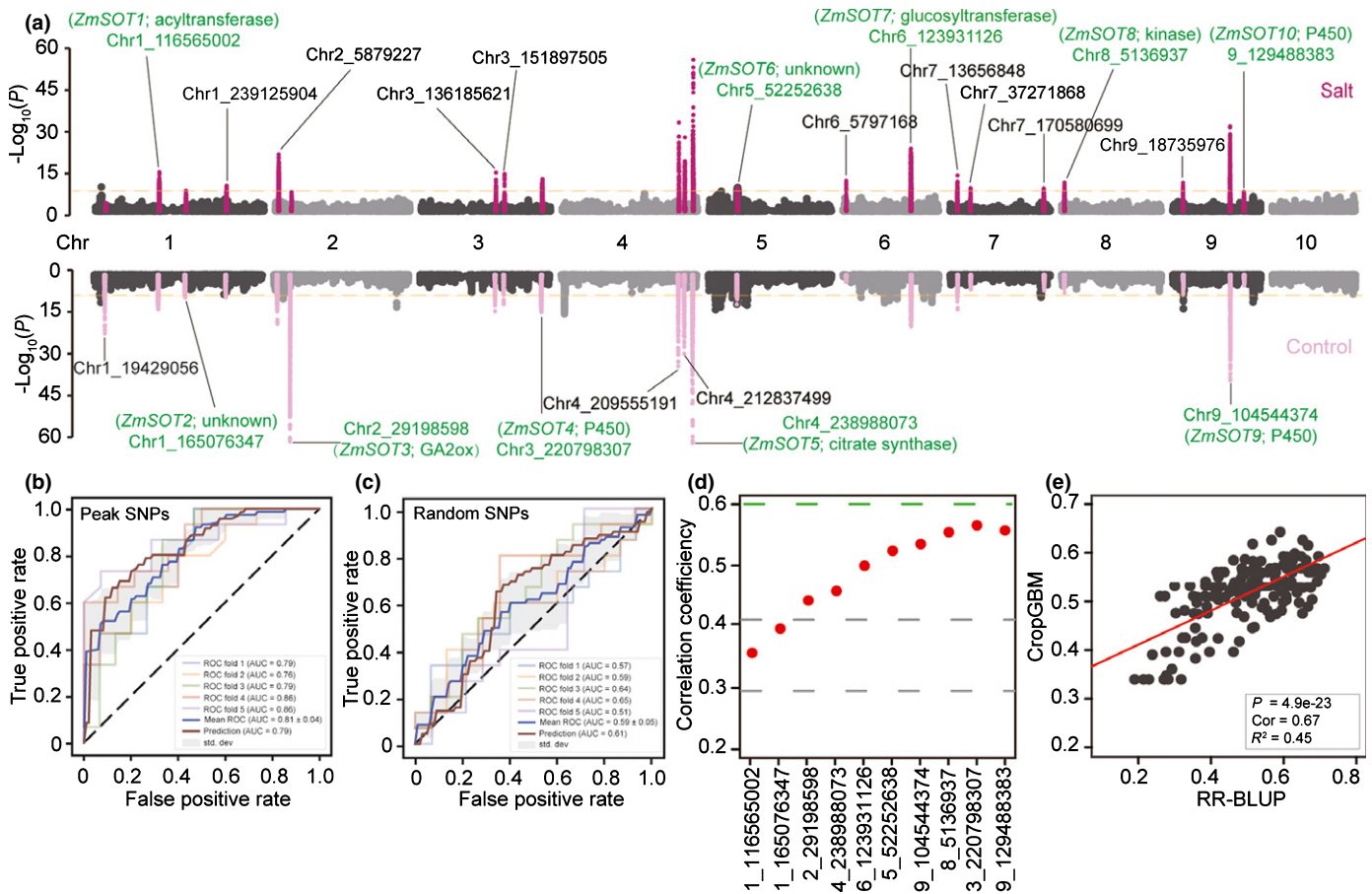
Fig. 3 Prediction model and prediction accuracy of salinity-induced osmotic stress (SIOS) tolerance in maize. (a) Metabolome-based genome-wide association study (mGWAS) results showing 22 significant single nucleotide polymorphisms (SNPs) associated with the abundances of the indicated metabolite biomarkers of SIOS tolerance (METOs) ($-\log_{10}(P) > 8.4$). The associations were plotted against genome location. (b, c) The prediction accuracy (expressed as average area under receiver operating characteristics curve, AUROC) using the 22 peak SNPs (b) or 22 random SNPs (c). The 'ROC fold 1-5' and 'prediction' curves were generated using the training and testing samples, respectively (see the Materials and Methods section). (d) Prediction accuracy of the CropGBM model. The prediction accuracy was expressed as correlation between the predicted and observed leaf water content (WC). The green line shows the prediction accuracy obtained using the RR-BLUP with 100 000 SNPs (genome-wide prediction baseline). The gray lines showed the lower and upper quartiles of random SNPs-based prediction accuracy (random prediction baseline). (e) The correlation between CropGBM-based (10 SNPs) and RR-BLUP-based (100 000 SNPs) prediction accuracy of leaf WC.

1 (*ZmSOT1*) to *ZmSOT10* (Figs 3a, S7). The candidate genes of *ZmSOTs* were annotated with different functions, including three cytochrome P450 (*ZmCYP709B2*, *ZmCYP93D1* and *ZmCYP72A15*), one citrate synthase (*ZmCS3*), one glucosyltransferase (*ZmUGT*), one kinase, one acyl-transferase, one Gibberellin2-oxidase and two unknown genes (Table S8). We experimentally addressed the function of four selected candidates (*ZmSOT5*/*ZmCS3*, *ZmSOT7*/*ZmUGT*, *ZmSOT9*/*ZmCYP93D1* and *ZmSOT10*/*ZmCYP709B2*) in the following sections.

ZmCS3 underlies the ZmSOT5–METO9&28–SIOS tolerance association

ZmCS3 is the candidate of *ZmSOT5*, which encodes an important enzyme catalyzing the oxaloacetate (OAA) to citrate reaction of the tricarboxylic acid (TCA) cycle in mitochondria. The TCA cycle is crucial for various energy (ATP) costly processes in cells (Sweetlove *et al.*, 2010; Vuoristo *et al.*, 2016; Munns *et al.*,

2020). We then determined if ZmCS3 is associated with the METO9&28 abundances and SIOS tolerance. We generated two independent transgenic lines overexpressing *ZmCS3* (Fig. S8), and two independent CRISPR-Cas9 lines knocking out *ZmCS3* (Fig. S9). We then analyzed the abundances of METO9 and METO28 in these genotypes, and observed that the contents of METO9 and METO28 were significantly higher in *ZmCS3*^{OE} plants, and were significantly lower in *ZmCS3*^{crispr} mutants as compared with that of the WT plants (Fig. 4a,b). In addition, we observed that the plants overexpressing *ZmCS3* maintained greater leaf WCs and were more tolerant to salt stress than the WT, whereas the *ZmCS3*^{crispr} mutants showed the opposite phenotype (Fig. 4c,d). These results confirmed that *ZmCS3* is the candidate of *ZmSOT5*, which increases the abundances of METO28 and METO9, and then promotes SIOS tolerance. These results also indicated that the natural variation of maize SIOS tolerance is attributable to the variation of the fundamental energy releasing pathways (TCA cycle). In addition, although a

previous study has shown that the natural variation of *ZmCS3* (*GRMZM2G063909*) is associated with the abundance of various phenylpropanoid hydroxycitric acid ester isomers, the chemical identities of METO28 and METO9 remain to be investigated.

We next investigated the genetic variants underlying the functional variation of ZmCS3. The peak SNP (Chr4_238988073) flanking *ZmSOT5*/*ZmCS3* was located in the 16th intron of *ZmCS3* (Fig. 4e), at which a guanine (G) and an adenine (A) were associated with low and high METO9&28 abundances, respectively (Fig. 4f,g). qRT-PCR assay determined that HapG and HapA inbred lines conferred comparable transcript levels of *ZmCS3* both under control and salinity conditions (Fig. 4h), which suggested that the functional variation of ZmCS3 was unlikely associated with the change of its transcript levels. We also performed transcriptome sequencing of a HapG line (IL11) and a HapA line (CAU284), interestingly, the results indicated that IL11 conferred two splice sites at the left side of the 11th intron (Fig. 4i), with one of them (splice site 2) leaded to a truncation of the open reading frame (Fig. 4j). In addition, we observed that the *ZmCS3*^*Splice site 2*^ transcripts were detected in all of the tested HapG lines, but not in HapA lines (Fig. 4k). To clarify the molecular basis underlying the alternative splicing at the splice site 2, we sequenced the genomic fragment flanking the 11th intron of *ZmCS3* in 20 inbred lines (10 HapA lines and 10 HapG lines), and found that all HapG lines conferred three SNPs (Chr4_238980788-C, Chr4_238980803-G and Chr4_238980808-C) (Fig. S10). NetGene2-based prediction of the splice site showed that these SNPs changed the splice site from site 1 to site 2 (Fig. 4l). Taken together, we concluded that three splice-modifying SNPs leaded to the alternative splicing at the 11th intron of *ZmCS3*, which attenuated the function of ZmCS3, then decreased METO9&28 abundances and SIOS tolerance, hence the allele without these SNPs is the salt (SIOS) tolerance allele of *ZmCS3*.

ZmUGT underlies the ZmSOT7–METO34–SIOS tolerance association

Flavonoids are distributed widely in the plant kingdom; they generally exhibit antioxidative activity involved in response to various environmental stresses, for example, UV stress and water deprivation (Nelson *et al.*, 2004; Yonekura-Sakakibara *et al.*, 2019). A recent study showed that the domestication-associated increase of flavonoid abundance contributes to the improvement of salt tolerance in soybean (Bian *et al.*, 2020). We observed that 17 of the 43 flavonoids identified in this study were upregulated by salt stress, including koparin, ferulic acid and coumaric acid (Fig. S11), indicating that flavonoids likely play important roles in maize SIOS tolerance. Meanwhile, we observed that the candidate gene of *ZmSOT7* encoded a UDP glycosyltransferase (ZmUGT) synthesizing Apigenin di-C-pentoside (Apidi-C-pen) (Fig. 5a). *ZmSOT7* was identified by the GWAS of METO34 abundance, and the MS/MS fragmentation assay confirmed that the chemical identity of METO34 is indeed Apidi-C-pen (Fig. 5b). To further confirm that *ZmSOT7*/*ZmUGT*-mediated synthesis of Apidi-C-pen confers salt (SIOS) tolerance, we generated two independent transgenic lines overexpressing *ZmUGT* (Fig. S12). Follow-up analysis observed that Apidi-C-pen abundances were significantly higher in *ZmUGT*^*OE*^ plants as compared with the WT plants (Fig. 5c), and *ZmUGT*^*OE*^ plants maintained greater leaf WCs and were more tolerant to salt stress than the WT (Fig. 5d,e). These results indicated that *ZmSOT7*/*ZmUGT*-mediated Apidi-C-pen synthesis confers salt (SIOS) tolerance.

A previous study showed that the natural variation of ZmUGT was significantly associated with the abundance of Apidi-C-pen in maize kernels, and that the proposed casual variation is a non-synonymous SNP (cytosine (C) to adenine (A)) causing an amino acid substitution (211^Ala^ to 211^Asp^) (Wen *et al.*, 2012). We analyzed the association of the 35 SNPs in *ZmUGT* with the abundance of METO34 (Table S9) and observed that the previous proposed casual SNP (SNP632 in this study) showed the highest association (Fig. 5f), at which a C and an A were associated with low and high METO34 abundances and SIOS tolerance (leaf WCs), respectively (Fig. 5g,h). These results supported the conclusion that SNP632 is the likely variant underlying the functional variation of ZmUGT. As *ZmUGT*^*OE*^ plants conferred increased Apidi-C-pen abundances and were more tolerant to salt stress than the WT (Fig. 5c–e); thus, we propose that the amino acid change from 211^Ala^ to 211^Asp^ likely increases the ZmUGT activity and then increases the Apidi-C-pen abundance and salt (SIOS) tolerance. This perspective was supported by the observation that the randomly selected HapC and HapA displayed comparable transcript levels of *ZmUGT* (Fig. 5i). Taken together, we concluded that HapA allele is the salt (SIOS) tolerance allele of *ZmUGT*.

ZmCYP709B2 underlies the ZmSOT10–METO31–SIOS tolerance association

Plant P450s play important roles in various biosynthetic and detoxicative processes, for example the biosynthesis of plant hormones and defensive secondary metabolites, and in detoxification of exogenous chemicals (Nelson *et al.*, 2004). Our results above showed that three of the *ZmSOT* candidates were P450 family genes (Table S8), and we conducted functional validation on two of them (*ZmSOT9*/*ZmCYP93D1* and *ZmSOT10*/ *ZmCYP709B2*). We generated the transgenic plant overexpressing the candidates (Fig. S13), and the CRISPR-Cas9 lines knocking out the candidates (Fig. S14), with two independent lines for each construct. We then analyzed the abundances of METOs in these genotypes and subsequently observed that the contents of METO31 in *ZmCYP709B2*^*OE*^ and *ZmCYP709B2*^*crispr*^ plants were higher and lower than that of the WT, respectively (Fig. 6a). Although *ZmCYP709B2*^*OE*^ and WT plants showed comparable salt sensitivity, we observed that the *ZmCYP709B2*^*crispr*^ plants maintained greater leaf WCs and were more tolerant to salt stress than the WT (Fig. 6b,c). These results indicated that *ZmCYP709B2* is the candidate of *ZmSOT10*, which increases the METO31 abundance and then decreases salt (SIOS) tolerance. By contrast, we observed that the contents of METO28 and METO30 in *ZmCYP93D1*^*OE*^, *ZmCYP93D1*^*crispr*^ and WT plants
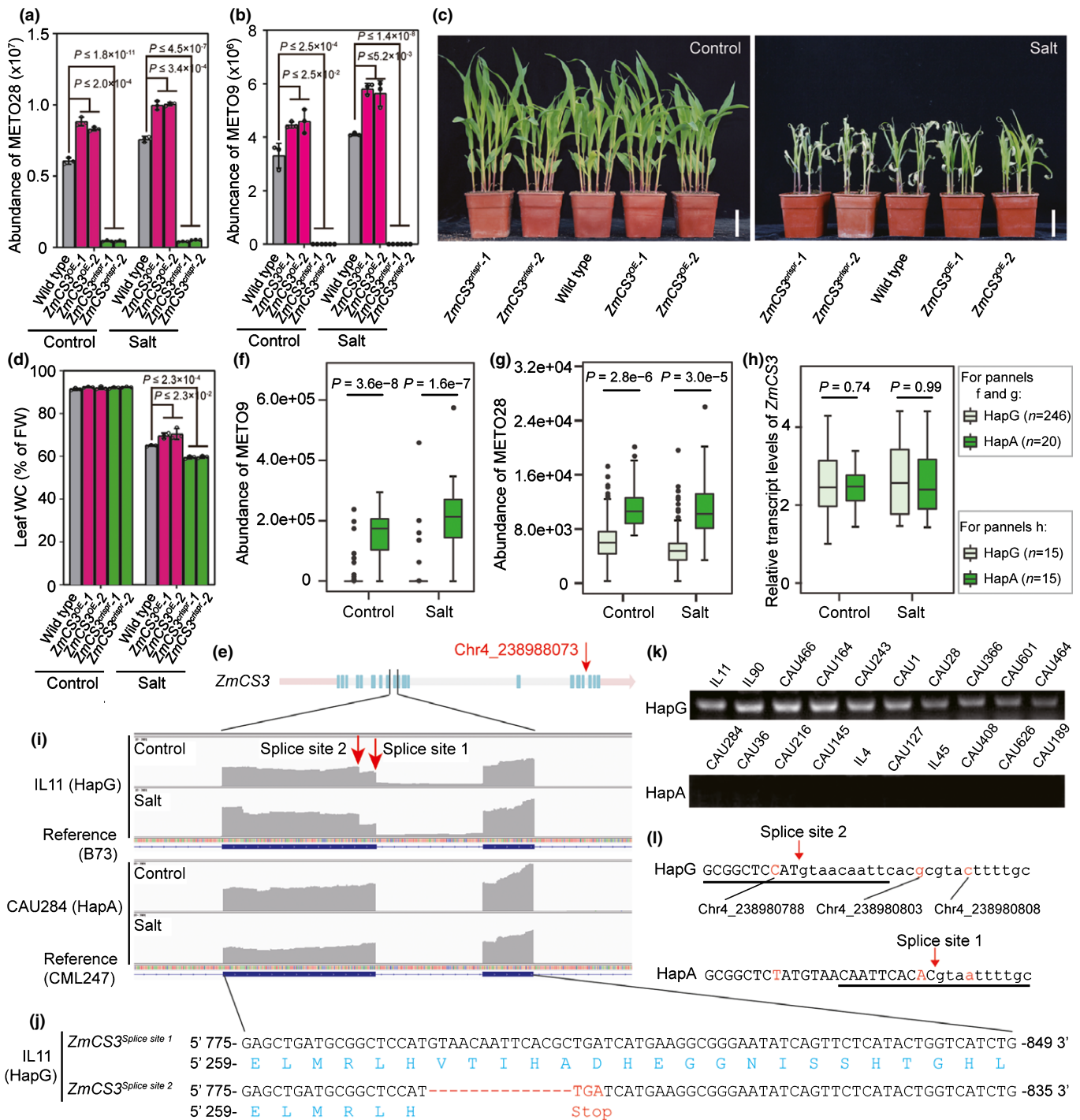
Fig. 4 The citrate synthase *ZmCS3* underlies the variation of metabolite biomarkers of SIOS tolerance (METO)9&28 abundance and salinity-induced osmotic stress (SIOS) tolerance in maize (*Zea mays* ssp. *mays*). (a, b) LC-MS-based analysis of the contents of METO28 (a) and METO9 (b) (genotypes and treatments as indicated). (c, d) The appearances (c) and leaf water contents (WCs) (d) of the control and salt-treated maize plants. 10-d-old seedlings were treated with 300 mM NaCl for 5 d, then the seedlings were photographed and the WCs (%FW) were measured. Bars, 5 cm. (e) The structure of *ZmCS3* and the location of the peak single nucleotide polymorphism (SNP) (Chr4_238988073). Blue boxes, exons; gray boxes, introns; and orange boxes, untranslated regions. (f, g) The distribution of METO9 (f) and METO28 (g) abundances. The samples were grouped based on the haplotypes (HapA and HapG) of the peak SNP Chr4_238988073. (h) The transcript levels of *ZmCS3* in HapA and HapG inbred lines under control and salt-treated conditions. (i, j) Integrated Genome Viewer visualization of the RNA-seq data of IL11 (a HapA line) and CAU284 (a HapG line). Red arrows indicated the alternative splicing occurred at the 11th intron of IL11 (i), with the splice site 2 resulted in a truncation of the open reading frame (j). (k) Reverse transcription-PCR-based analysis of the presence/absence of *ZmCS3*^Splice site 2 transcripts. The primers uniquely amplifying *ZmCS3*^Splice site 2 were used. (l) NetGene2-based prediction of splice site. The predicted splice site of HapA sequence was splice site 1, and the predicted splice site of the HapG sequence (with Chr4_238980788-C, Chr4_238980803-G and Chr4_238980808-C) was splice site 2. Data in (a), (b) and (d) were means ± SD of three independent experiments. Statistical significance was determined by a two-sided Student's *t*-test.
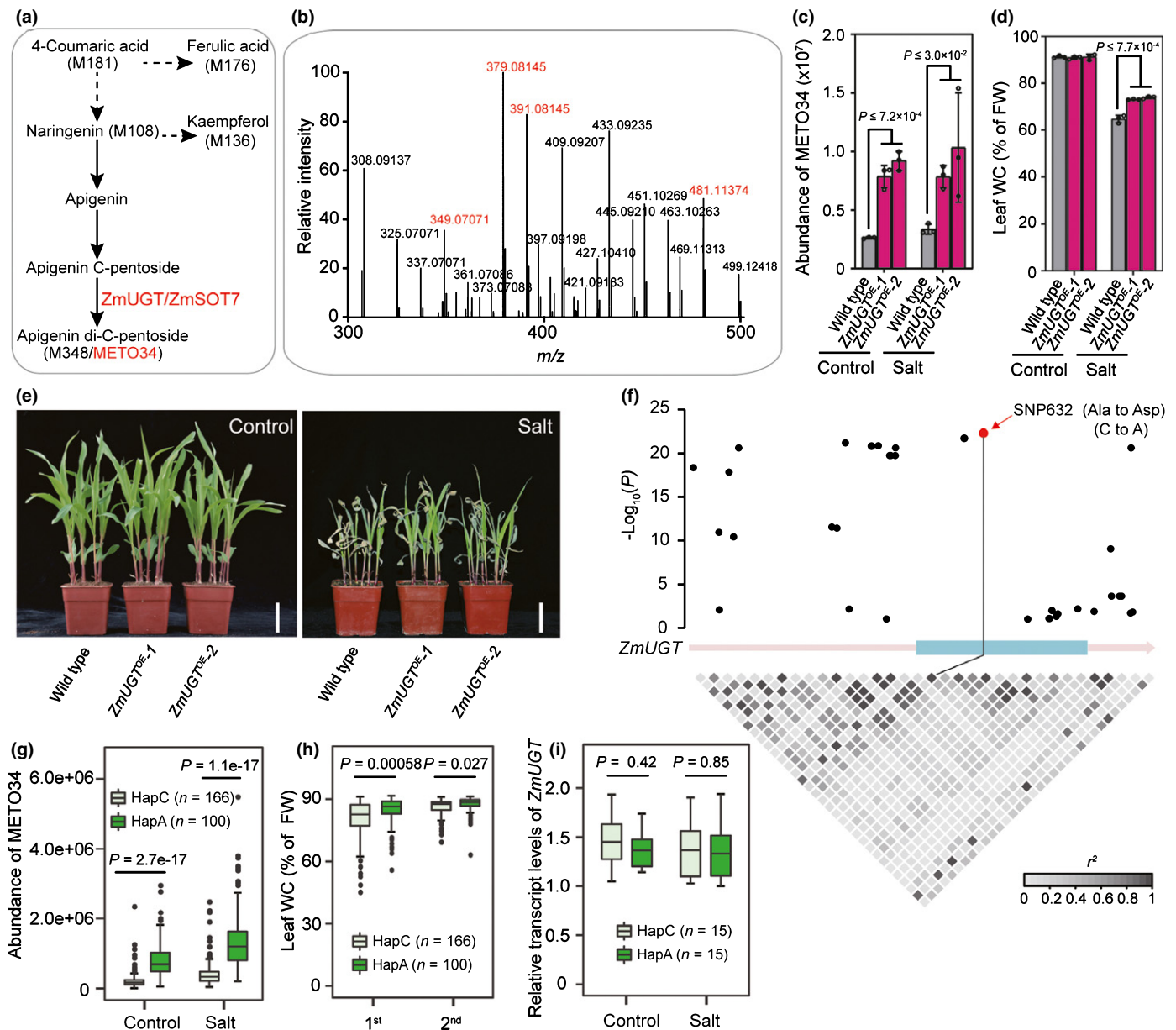
Fig. 5 The glucosyltransferase *ZmUGT* underlies the variation of metabolite biomarkers of SIOS tolerance (METO)34 abundance and salinity-induced osmotic stress (SIOS) tolerance in maize (*Zea mays* ssp. *mays*). (a) Pathway of Apidi-C-pen biosynthesis in maize. (b) MS/MS fragmentation profile of METO34. The fragments highlighted in red matched the MS/MS fragments of Apidi-C-pen. (c) LC-MS-based assay of METO34 abundances (the genotypes and treatments as indicated). (d, e) The leaf water contents (WCs) (d) and appearances (e) of the control and salt-treated maize plants. Plant growth and salt treatment were as described in Fig. 4(c). Bars, 5 cm. (f) The association of the single nucleotide polymorphisms (SNPs) in *ZmUGT* with METO34 abundance. Red dot indicated the SNP (SNP632) with the highest association. Lower panel showed the pattern of the pairwise linkage disequilibrium (LD) of the variants. (g, h) The distributions of METO34 abundances (g) and leaf WCs (h). Samples were grouped based on the haplotypes (HapC and HapA) of SNP632. (i) The transcript levels of *ZmUGT*. Thirty samples (15 HapC lines, 15 HapA lines) were randomly selected to conduct the quantitative reverse transcription (qRT-PCR) as described in the Materials and Methods section. Data in (c) and (d) were means ± SD of three independent experiments. Statistical significance was determined by a two-sided Student's *t*-test.

were comparable (Fig. S15), suggesting that the candidate of *ZmSOT9* is yet to be confirmed.

We next investigated the genetic variants underlying the functional variation of ZmCYP709B2. The peak SNP (Chr9_129488383) flanking *ZmSOT10*/*ZmCYP709B2* was located in the promoter of *ZmCYP709B2* (Fig. S16a), at which a C and a G amino acid were associated with low and high

METO31 abundances, respectively (Fig. S16b). By comparing the genomic region covering *ZmCYP709B2* between a HapG line (IL108) and a HapC line (CAU216), we found that CAU216 conferred a 50-bp deletion (Del382) in the first intron of *ZmCYP709B2* (Table S10). In addition, we determined the presence/absence of Del382 in 200 randomly selected inbred lines by a PCR-based assay (Table S11), and then analyzed the association
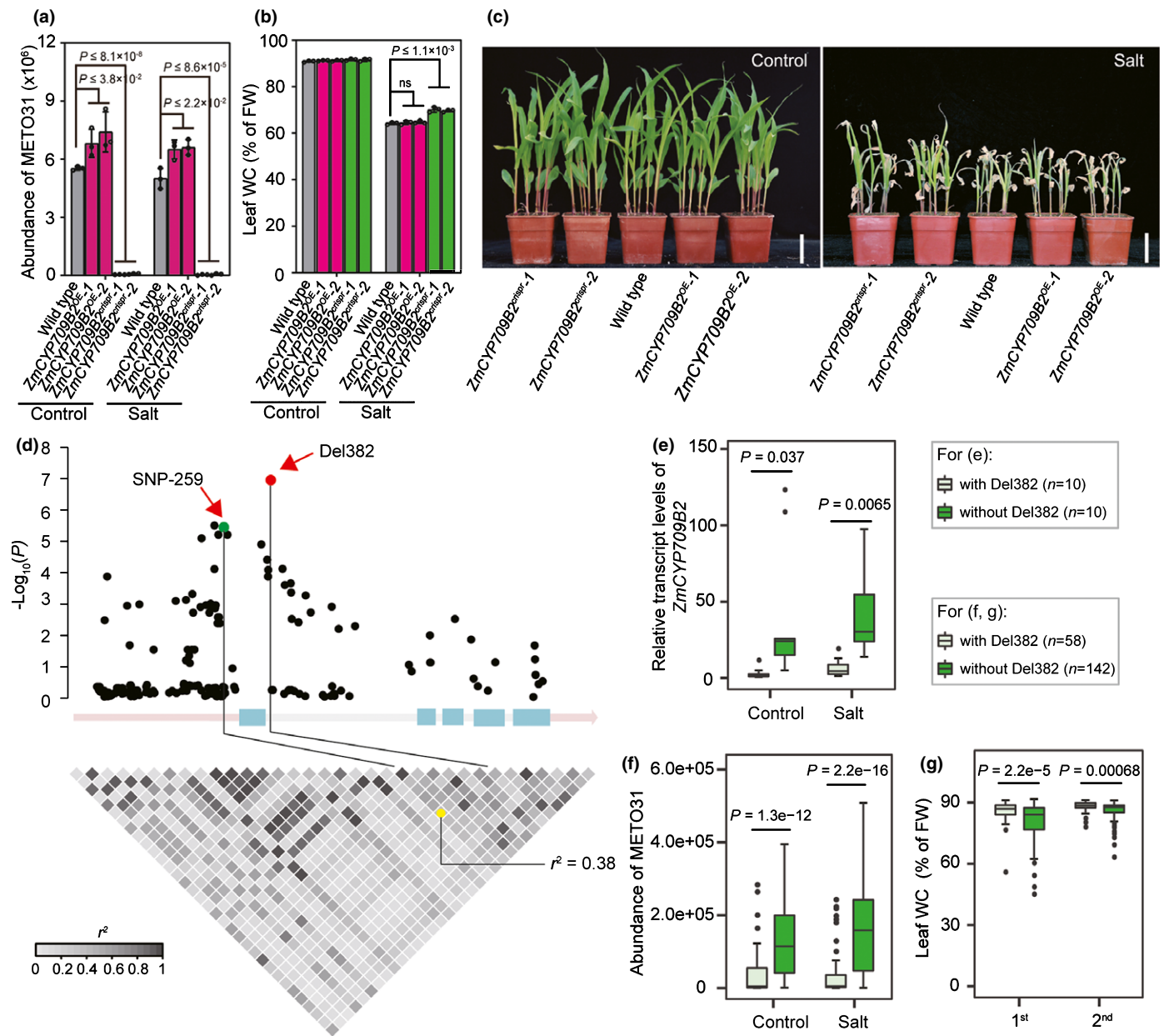
Fig. 6 The cytochrome P450 polypeptide *ZmCYP709B2* underlies the variation of metabolite biomarkers of SIOS tolerance (METO)314 abundance and salinity-induced osmotic stress (SIOS) tolerance in maize (*Zea mays* ssp. *mays*). (a) LC-MS-based analysis of METO31 abundances in the indicated samples. (b, c) The leaf water contents (WCs) (b) and appearances (c) of the control and salt-treated maize plants with indicated genotypes. Plant growth and salt treatment were as described in Fig. 4(c). Bars, 5 cm. ns, no statistical differences. (d) The association of Del382 and single nucleotide polymorphisms (SNPs) in *ZmCYP709B2* with METO31 abundances. Red dot indicates the InDel (Del382) with the highest association. Middle panel showed the structure of *ZmCYP709B2* and the location of Del382. Blue boxes, exons; gray boxes, introns; and orange boxes, untranslated regions. Lower panel showed the pattern of the pairwise linkage disequilibrium (LD) of the variants. The yellow dot highlight the moderate LD between Del382 and the peak SNP (the green dot) identified by the original GWAS. (e–g) The transcript levels of *ZmCYP709B2* (e), the METO31 abundances (f) and leaf WCs (g) of the inbred lines with or without Del382 (treatments as indicated). Data in (a) and (b) are means ± SD of three independent experiments. The statistical significance in (a), (b) and (e–g) were determined by a two-sided Student's *t*-test.

of Del382 and the SNPs in *ZmCYP709B2* with METO31 abundance (Table S12). The results indicated that Del382 showed the highest association with METO31 abundances (Fig. 6d), and that Del382 and the peak SNP (Chr9_129488383) were moderately linkage disequilibrium (LD) together among the 200 tested lines ($r^2 = 0.38$; Fig. 6d). Further analysis showed that the inbred lines with Del382 conferred decreased *ZmCYP709B2* transcript levels and METO31 abundance (Fig. 6e,f), and maintained greater leaf WCs under salt conditions as compared with the lines without Del382 (Fig. 6g). These observations permitted us to conclude that Del382 likely reduces the transcript levels of *ZmCYP709B2*, and then reduces METO31 abundance, which in turn increases SIOS tolerance. Therefore, the allele with Del382 is the salt (SIOS) tolerance allele of *ZmCYP709B2*.

Identification and application of the favorable alleles of *ZmCS3*, *ZmUGT* and *ZmCYP709B2*

The favorable genetic variants provided valuable resources for marker-assisted crop improvement (Huang *et al.*, 2009; Wang *et al.*, 2018). We next investigated whether the favorable alleles of *ZmCS3*, *ZmUGT* and *ZmCYP709B2* can improve the SIOS tolerance of elite maize inbred lines. ZHENG58 and IL108 were the female parents of two commercial maize hybrids (ZD958 and XY335), and they conferred the SIOS-sensitive alleles of *ZmCS3*, *ZmUGT* and *ZmCYP709B2* (Fig. 7a–f). M1016 was a SIOS-tolerant inbred line conferring the SIOS-tolerant alleles of *ZmCS3*, *ZmUGT* and *ZmCYP709B2* (Fig. 7a–f). We then introgressed $ZmCS3^{M1016}$, $ZmUGT^{M1016}$ and $ZmCYP709B2^{M1016}$ into ZHENG58 and IL108 by marker-assisted generation of near-isogenic lines (NILs). We carried out four generations of backcrossing of F_1 plants. For each generation, the plants that were heterozygous at all three target sites were selected and backcrossed. By one generation of self-pollination, we identified $NIL^{ZmCS3-M1016}$, $NIL^{ZmUGT-M1016}$, $NIL^{ZmCYP709B2-M1016}$, $NIL^{ZmCS3-M1016\ ZmUGT-M1016}$, $NIL^{ZmCS3-M1016\ ZmCYP709B2-M1016}$, $NIL^{ZmUGT-M1016\ ZmCYP709B2-M1016}$ and $NIL^{ZmCS3-M1016\ ZmUGT-M1016\ ZmCYP709B2-M1016}$ from the BC4F2 population in ZHENG58 and IL108 background (Fig. S17). Although these NILs and the WT controls (ZHENG58 and IL108) showed similar appearances under control conditions
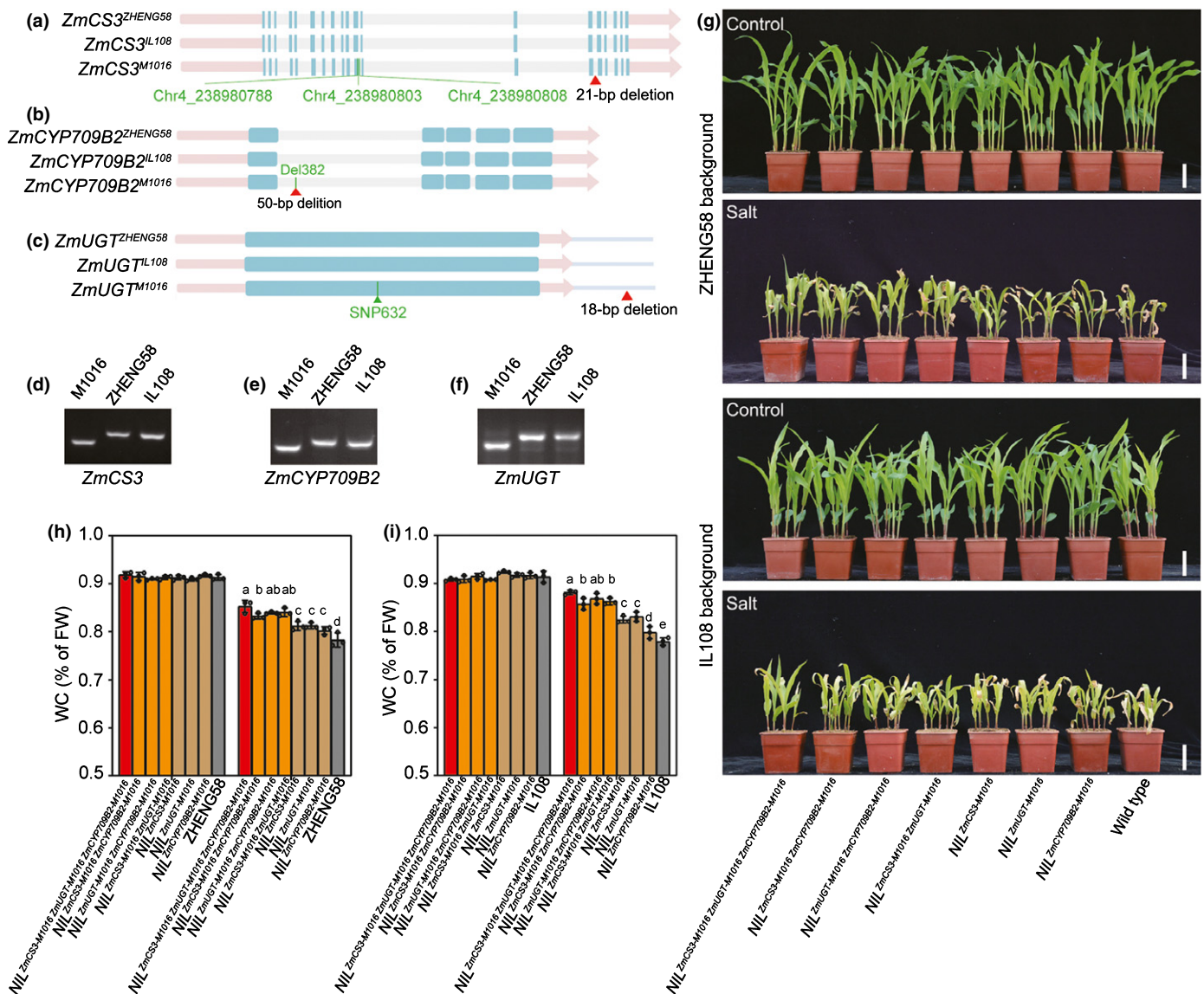

Fig. 7 Characterization and application of the salinity-induced osmotic stress (SIOS)-tolerant sulfotransferase *ZmSOT* alleles in maize (*Zea mays* ssp. *mays*). (a–f) The molecular marker for the SIOS-tolerant *ZmSOT* alleles. The green lines indicate the variants likely associated with the functional variations of *ZmCS3* (a), *ZmCYP709B2* (b) and *ZmUGT* (c). The red arrows in (a)–(c) indicate the location of the InDels that have been used to develop the PCR-based marker for the favorable alleles of *ZmCS3* (d), *ZmCYP709B2* (e) and *ZmUGT* (f). (g–i), The appearances (g) and leaf water contents (WCs) (h–i) of the control and salt-treated maize plants (genotypes as indicated). Data in (h) and (i) are means ± SD of three independent experiments. Statistical significance was determined by one-way ANOVA test. Bars, 5 cm (g).

(Fig. 7g), we observed that, following the onset of salt treatment, $NIL^{ZmCS3-M1016}$, $NIL^{ZmUGT-M1016}$ and $NIL^{ZmCYP709B2-M1016}$ maintained greater leaf WCs than the WT controls, the dual-site modified NILs ($NIL^{ZmCS3-M1016\ ZmUGT-M1016}$, $NIL^{ZmCS3-M1016\ ZmCYP709B2-M1016}$ and $NIL^{ZmUGT-M1016\ ZmCYP709B2-M1016}$) performed better than the single-site modified NILs, and the triple-site modified NIL ($NIL^{ZmCS3-M1016\ ZmUGT-M1016\ ZmCYP709B2-M1016}$) performed the best (Fig. 7g–i). These observations demonstrated that the favorable allele of $ZmCS3$, $ZmUGT$ and $ZmCYP709B2$ additively improve the SIOS tolerance of the elite maize inbred lines ZHENG58 and IL108.

Discussion

Natural variation of salt tolerance is widely present within and among crop species. Comprehensive discovery of the genetic variants associated with the main salt-tolerant physiological processes (sodium (Na^+) homeostasis and salt-induced osmotic stress (SIOS) tolerance) can provide important theoretical and material supports for breeding salt-tolerant crops (Munns & Tester, 2008; Horie *et al.*, 2009; Yang *et al.*, 2018; Wang *et al.*, 2020). Decades of efforts by scientists around the world have made substantial progress of understanding the natural variation of Na^+ homeostasis in various crops (Ren *et al.*, 2005; Munns *et al.*, 2012; C. Zhang *et al.*, 2019; M. Zhang *et al.*, 2019; Wang *et al.*, 2020). However, the mechanism of crops SIOS tolerance remains largely unknown, which has become one of the main theoretical bottlenecks of the genetic improvement of crop salt tolerance. In this study, we conducted a comprehensive analysis of the metabolomics profiles of a maize (*Zea mays* ssp. *mays*) population ($n = 266$) (Fig. 1), by first setting up a leaf water content-based assay of SIOS tolerance, and then establishing a metabolite biomarker-based pipeline for target mining of the genetic variants and genes underlying the natural variations of the SIOS tolerance in maize, in which large-scale metabolomics profiling (Fig. 1), big data mining (Fig. 2), genotype-to-phenotype prediction (Fig. 3), verification of gene function (Figs 4–6) and marker-assisted breeding (Fig. 7) were integrated to expedite the discovery of SIOS-tolerance quantitative trait loci (QTLs) and the genetic improvement of maize salt (SIOS) tolerance.

Salt tolerance is an important agronomic trait for maize breeding. In this study, a citrate synthase ($ZmCS3$), a glucosyltransferase ($ZmUGT$) and a cytochrome P450 ($ZmCYP709B2$) were validated to be associated with the natural variation of maize SIOS tolerance (Figs 4–6), and their favorable alleles additively improve maize SIOS tolerance (Fig. 7). In agreement with the notion that ZmCS3, ZmUGT and ZmCYP709B2 mainly associated with SIOS tolerance but not Na^+ homeostasis, we showed that the salt-grown $ZmCS3^{OE}$, $ZmCS3^{crispr}$, $ZmUGT^{OE}$, $ZmCYP709B2^{OE}$, $ZmCYP709B2^{crispr}$ and the wild-type (WT) plants showed comparable shoot Na^+ contents (Fig. S18). Hence, our functional characterization of $ZmCS3$, $ZmUGT$ and $ZmCYP709B2$ provided insightful understanding of the molecular basis underlying the natural variation of maize SIOS tolerance, as well as valuable genetic resources for improving SIOS tolerance. Because ZmCS3 is a key enzyme of the tricarboxylic

acid (TCA) cycle (Sweetlove *et al.*, 2010; Vuoristo *et al.*, 2016; Zhang & Fernie, 2018), ZmUGT is an enzyme catalyzing the metabolism of flavonoid (Wen *et al.*, 2014), and ZmCYP709B2 is a functionally unknown P450 protein, we suggest that the natural variation of maize SIOS tolerance is attributable to the variation of the fundamental energy releasing pathways (TCA cycle), flavonoid metabolic pathway and other metabolic pathways, which is consistent with the notion that SIOS tolerance is a complex trait regulating by a large number of QTLs each with a small effect.

In this study, we have identified a set of metabolite biomarkers of SIOS tolerance (METO), yet the identities of most of these METOs remain uncharacterized or only putatively identified. Future studies towards the structural identification of these metabolites by a reference compound comparison or stereochemistry NMR analysis will provide further insight into the mechanism of SIOS tolerance mediated by these METOs. Meanwhile, future validation of the rest of the sulfotransferase $ZmSOT$ candidates will facilitate the structural identification of these METOs as well as advance our understanding of the metabolomics and genetic basis underlying the natural variation of maize SIOS tolerance. In the long run, marker-assisted breeding programs could introgress the SIOS-tolerant alleles of $ZmSOTs$ and previously identified genetic variants promoting Na^+ homeostasis into high-yielding maize cultivars to develop commercial hybrids (Zhang *et al.*, 2018; C. Zhang *et al.*, 2019; M. Zhang *et al.*, 2019; Cao *et al.*, 2020).

In summary, we provide a demonstration of using metabolite biomarkers as a chemical fingerprint to discover novel SIOS-tolerant genes, which leads to an insightful understanding and genetic improvement of maize salt (SIOS) tolerance. In the meantime, we present a pipeline of metabolomics-based untargeted mining of novel stress-associated genes. The strategy could be adopted across agronomic trait and crop species to expedite the genetic improvement of a wide range of crops.

Acknowledgements

Author contributions

XL, SL, TW, FL, JL, FQ, ZL, XW and CJ planned and designed the research; XL, TW and FL grew the GWAS population and conducted the metabolomics profiling; XL and JC generated the overexpressing lines and CRISPR/Cas9 knockout lines; XL carried out the functional analysis; SL and XW carried out the bioinformatics analysis; and XL, SL and CJ wrote the manuscript (the other authors contributed). XL, SL and TW contributed equally to this work.

ORCID

Caifu Jiang ![ORCID] https://orcid.org/0000-0003-3032-253X
Feng Qin ![ORCID] https://orcid.org/0000-0001-7197-9197
Xiangfeng Wang ![ORCID] https://orcid.org/0000-0002-0360-0859

Data availability

Source data can be found in the Supporting Information. The data that support the findings of this study are available from the corresponding authors upon request.

References

Bian XH, Li W, Niu CF, Wei W, Hu Y, Han JQ, Lu X, Tao JJ, Jin M, Qin H et al. 2020. A class B heat shock factor selected for during soybean domestication contributes to salt tolerance by promoting flavonoid biosynthesis. *New Phytologist* **225**: 268–283.

Cao Y, Liang X, Yin P, Zhang M, Jiang C. 2019. domestication-associated reduction in K$^+$-preferring HKT transporter activity underlies maize shoot K$^+$ accumulation and salt tolerance. *New Phytologist* **222**: 301–317.

Cao Y, Zhang M, Liang X, Li F, Shi Y, Yang X, Jiang C. 2020. Natural variation of an EF-hand Ca^{2+}-binding-protein coding gene confers saline-alkaline tolerance in maize. *Nature Communications* **11**: 186.

Chen W, Wang W, Peng M, Gong L, Gao Y, Wan J, Wang S, Shi L, Zhou B, Li Z et al. 2016. Comparative and parallel genome-wide association studies for metabolic and agronomic traits in cereals. *Nature Communication* **7**: 12767.

Egamberdieva D, Wirth S, Bellingrath-Kimura SD, Mishra J, Arora NK. 2019. Salt-tolerant plant growth promoting rhizobacteria for enhancing crop productivity of saline soils. *Frontiers in Microbiology* **10**: 2791.

Endelman JB. 2011. Ridge regression and other kernels for genomic selection with R package rrBLUP. *Plant Genome.* **4**: 250–255.

Fu X. 1983. The exploitation and management of saline-alkali land resources in north china plain from the perspective of land planning. *Henan Science Technology* **6**: 12–15.

Ghatak A, Chaturvedi P, Weckwerth W. 2018. Metabolomics in plant stress physiology. *Advances in Biochemical Engineering/Biotechnology* **164**: 187–236.

Hanks RJ, Ashcroft GL, Rasmussen VP, Wilson GD. 1987. Corn production as influenced by irrigation and salinity-Utah studies. *Irrigation Science* **1**: 47–59.

Horie T, Hauser F, Schroeder JI. 2009. HKT transporter-mediated salinity resistance mechanisms in Arabidopsis and monocot crop plants. *Trends in Plant Science* **14**: 660–668.

Huang X, Qian Q, Liu Z, Sun H, He S, Luo D, Xia G, Chu C, Li J, Fu X. 2009. Natural variation at the DEP1 locus enhances grain yield in rice. *Nature Genetics* **41**: 494–497.

Ismail AM, Horie T. 2017. genomics, physiology, and molecular breeding approaches for improving salt tolerance. *Annual Review of Plant Biology* **68**: 405–434.

Jiao Y, Zhao H, Ren L, Song W, Zeng B, Guo J, Wang B, Liu Z, Chen J, Li W et al. 2012. Genome-wide genetic changes during modern breeding of maize. *Nature Genetics* **44**: 812–815.

Li N, Lin B, Wang H, Li X, Yang F, Ding X, Yan J, Chu Z. 2019. Natural variation in ZmFBL41 confers banded leaf and sheath blight resistance in maize. *Nature Genetics* **51**: 1540–1548.

Ligges U, Maechler M. 2003. Scatterplot3d – an R package for visualizing multivariate data. *Journal of Statistical Software* **8**: 1–20.

Liu S, Li C, Wang H, Wang S, Yang S, Liu X, Yan J, Li B, Beatty M, Zastrow-Hayes G et al. 2020. Mapping regulatory variants controlling gene expression in drought response and tolerance in maize. *Genome Biology* **21**: 163.

Luo X, Wang B, Gao S, Zhang F, Terzaghi W, Dai M. 2019. Genome-wide association study dissects the genetic bases of salt tolerance in maize seedlings. *Journal of Integrative Plant Biology* **61**: 658–674.

Munns R, Day DA, Fricke W, Watt M, Arsova B, Barkla BJ, Bose J, Byrt CS, Chen ZH. 2020. Energy costs of salt tolerance in crop plants. *New Phytologist* **225**: 1072–1090.

Munns R, James RA, Xu B, Athman A, Conn SJ, Jordans C, Byrt CS, Hare RA, Tyerman SD, Tester M et al. 2012. Wheat grain yield on saline soils is improved by an ancestral Na$^+$ transporter gene. *Nature Biotechnology* **30**: 360–364.

Munns R, Tester M. 2008. Mechanisms of salinity tolerance. *Annual Review of Plant Biology* **59**: 651–681.

Nakabayashi R, Saito K. 2015. Integrated metabolomics for abiotic stress responses in plants. *Current Opinion in Plant Biology* **24**: 10–16.

Nelson DR, Schuler MA, Paquette SM, Werck-Reichhart D, Bak S. 2004. Comparative genomics of rice and Arabidopsis. Analysis of 727 cytochrome P450 genes and pseudogenes from a monocot and a dicot. *Plant Physiology* **135**: 756–772.

Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, Blondel M, Prettenhofer P, Weiss R, Dubourg V et al. 2011. Scikit-learn: machine learning in python. *Journal of Machine Learning Research* **12**: 2825–2830.

Ren ZH, Gao JP, Li LG, Cai XL, Huang W, Chao DY, Zhu MZ, Wang ZY, Luan S, Lin HX. 2005. A rice quantitative trait locus for salt tolerance encodes a sodium transporter. *Nature Genetics* **37**: 1141–1146.

Sandhu D, Pudussery MV, Kumar R, Pallete A, Markley P, Bridges WC, Sekhon RS. 2020. Characterization of natural genetic variation identifies multiple genes involved in salt tolerance in maize. *Functional & Integrative Genomics* **20**: 261–275.

Saneoka H, Nagasaka C, Hahn DT, Yang WJ, Premachandra GS, Joly RJ, Rhodes D. 1995. Salt tolerance of glycinebetaine-deficient and-containing maize lines. *Plant Physiology* **107**: 631–638.

Schnable JC. 2015. Genome evolution in maize: from genomes back to genes. *Annual Review of Plant Biology* **66**: 329–343.

Sweetlove LJ, Beard KF, Nunes-Nesi A, Fernie AR, Ratcliffe RG. 2010. Not just a circle: flux modes in the plant TCA cycle. *Trends in Plant Science* **15**: 462–470.

Virtanen P, Gommers R, Oliphant TE, Haberland M, Reddy T, Cournapeau D, Burovski E, Peterson P, Weckesser W, Bright J et al. 2020. SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nature Methods* **17**: 261–272.

Vuoristo KS, Mars AE, Sanders JPM, Eggink G, Weusthuis RA. 2016. Metabolic engineering of TCA Cycle for production of chemicals. *Trends in Biotechnology* **34**: 91–197.

Wang M, Li W, Fang C, Xu F, Liu Y, Wang Z, Yang R, Zhang M, Liu S, Lu S et al. 2018. Parallel selection on a dormancy gene during domestication of crops from multiple families. *Nature Genetics* **50**: 1435–1441.

Wang S, Alseekh S, Fernie AR, Luo J. 2019. The structure and function of major plant metabolite modifications. *Molecular Plant* **12**: 899–919.

Wang X, Wang H, Liu S, Ferjani A, Li J, Yan J, Yang X, Qin F. 2016. Genetic variation in ZmVPP1 contributes to drought tolerance in maize seedlings. *Nature Genetics* **48**: 1233–1241.

Wang Z, Hong Y, Li Y, Shi H, Yao J, Liu X, Wang F, Huang S, Zhu G, Zhu JK. 2021. Natural variations in SlSOS1 contribute to the loss of salt tolerance during tomato domestication. *Plant Biotechnology Journal* **19**: 20–22.

Wang Z, Hong Y, Zhu G, Li Y, Niu Q, Yao J, Hua K, Bai J, Zhu Y, Shi H et al. 2020. Loss of salt tolerance during tomato domestication conferred by variation in a Na$^+$/K$^+$ transporter. *EMBO Journal* **39**: e103256.

Weckwerth W. 2003. Metabolomics in systems biology. *Annual Review of Plant Biology* **54**: 669–689.

Weckwerth W, Ghatak A, Bellaire A, Chaturvedi P, Varshney RK. 2020. PANOMICS meets germplasm. *Plant Biotechnology Journal* **18**: 1507–1525.

Weckwerth W, Loureiro ME, Wenzel K, Fiehn O. 2004. Differential metabolic networks unravel the effects of silent plant phenotypes. *Proceedings of the National Academy of Sciences, USA* **101**: 7809–7814.

Wen W, Li D, Li X, Gao Y, Li W, Li H, Liu J, Liu H, Chen W, Luo J et al. 2014. Metabolome-based genome-wide association study of maize kernel leads to novel biochemical insights. *Nature Communications* **5**: 3438.

Wen W, Liu H, Zhou Y, Jin M, Yang N, Li D, Luo J, Xiao Y, Pan Q, Tohge T et al. 2016. Combining quantitative genetics approaches with regulatory network analysis to dissect the complex metabolism of the maize kernel. *Plant Physiology* **170**: 136–146.

Xing HL, Dong L, Wang ZP, Zhang HY, Han CY, Liu B, Wang XC, Chen QJ. 2014. CRISPR/Cas9 toolkit for multiplex genome editing in plants. *BMC Plant Biology* **14**: 327.

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. *The Plant Cell* **31**: 1990–2009.

Yang C, Zhao L, Zhang H, Yang Z, Wang H, Wen S, Zhang C, Rustgi S, von Wettstein D, Liu B. 2014. Evolution of physiological responses to salt stress in hexaploid wheat. *Proceedings of the National Academy of Sciences, USA* **111**: 11882–11887.

Yang M, Lu K, Zhao FJ, Xie W, Ramakrishna P, Wang G, Du Q, Liang L, Sun C, Zhao H et al. 2018. Genome-wide association studies reveal the genetic basis of ionomic variation in rice. *The Plant Cell* **30**: 2720–2740.

Yang Y, Guo Y. 2018. Elucidating the molecular mechanisms mediating plant salt-stress responses. *New Phytologist* **217**: 523–539.

Yonekura-Sakakibara K, Higashi Y, Nakabayashi R. 2019. The origin and evolution of plant flavonoid metabolism. *Frontiers in Plant Science* **10**: 943.

Zhang C, Wang P, Tang D, Yang Z, Lu F, Qi J, Tawari NR, Shang Y, Li C, Huang S. 2019. The genetic basis of inbreeding depression in potato. *Nature Genetics* **51**: 374–378.

Zhang M, Cao Y, Wang Z, Wang Z, Shi J, Liang X, Song W, Chen Q, Lai J, Jiang C. 2018. A retrotransposon in an HKT1 family sodium transporter causes variation of leaf Na$^+$ exclusion and salt tolerance in maize. *New Phytologist* **217**: 1161–1176.

Zhang M, Liang X, Wang L, Cao Y, Song W, Shi J, Lai J, Jiang C. 2019. A HAK family Na$^+$ transporter confers natural variation of salt tolerance in maize. *Nature Plants* **5**: 1297–1308.

Zhang YJ, Fernie AR. 2018. On the role of the tricarboxylic acid cycle in plant productivity. *Journal of Integrative Plant Biology* **60**: 1199–1216.

Zhou S, Kremling KA, Bandillo N, Richter A, Zhang YK, Ahern KR, Artyukhin AB, Hui JX, Younkin GC, Schroeder FC et al. 2019. Metabolome-scale genome-wide association studies reveal chemical diversity and genetic control of maize specialized metabolites. *The Plant Cell* **31**: 937–955.

Zhou X, Stephens M. 2012. Genome-wide efficient mixed-model analysis for association studies. *Nature Genetics* **44**: 821–824.

Zhu G, Gou J, Klee H, Huang S. 2019. Next-gen approaches to flavor-related metabolism. *Annual Review of Plant Biology* **70**: 187–212.

Zhu G, Wang S, Huang Z, Zhang S, Liao Q, Zhang C, Lin T, Qin M, Peng M, Yang C et al. 2018. Rewiring of the fruit metabolome in tomato breeding. *Cell* **172**: 249–261.

Zuo W, Chao Q, Zhang N, Ye J, Tan G, Li B, Xing Y, Zhang B, Liu H, Fengler KA et al. 2015. A maize wall-associated kinase confers quantitative resistance to head smut. *Nature Genetics* **47**: 151–157.

Supporting Information

Additional Supporting Information may be found online in the Supporting Information section at the end of the article.

Fig. S1 Hierarchical clustering result of the concentrations of 33 619 metabolites.

Fig. S2 The correlations between leaf WCs and shoot Na+ contents.

Fig. S3 The contents of the indicated METOs in the control and salt-treated samples.

Fig. S4 The genetic diversity of the inbred lines used in this study.

Fig. S5 The density of SNPs in the indicated sample set.

Fig. S6 Box plots showing leaf WCs of inbred lines with indicated haplotypes at the given positions.

Fig. S7 Characterization of the candidate genes of *ZmSOTs*.

Fig. S8 The transcript levels of *ZmCS3* in the WT and transgenic plants overexpressing ZmCS3.

Fig. S9 Generation of *ZmCS3* mutants via CRISPR-Cas9-based approach.

Fig. S10 DNA sequences flanking the splice site of the 11th intron of *ZmCS3* in 10 randomly selected HapA and HapG lines.

Fig. S11 The abundances and chemical identities of 17 flavonoids in control and salt-treated samples. Statistical significances were determined by a two-sided Student's *t*-test.

Fig. S12 The transcript levels of *ZmUGT* in the WT and transgenic plants overexpressing *ZmUGT*.

Fig. S13 The transcript levels of *ZmCYP93D1* and *ZmCYP709B2* in the WT and transgenic plants overexpressing the indicated genes.

Fig. S14 Generation of *ZmCYP93D1* and *ZmCYP709B2* mutants via the CRISPR-Cas9-based approach.

Fig. S15 The abundances of METO28 and METO30 in plants grown under control and salt conditions.

Fig. S16 GWAS results of METO31 abundance in salt sample.

Fig. S17 Graphic demonstration of the generation of NILs.

Fig. S18 Na$^+$ content in the shoot tissue of the plants grown under control and salt conditions (genotypes as indicated).

Table S1 The list of the 266 inbred lines used for metabolomics profiling.

Table S2 The profiles and normalized abundances of 33 619 metabolites.

Table S3 Detailed information about the annotated metabolites.

Table S4 List of metabolites showing high correlation with leaf WCs.

Table S5 List of the 37 METOs.

Table S6 Details of the 434 maize inbred lines used in this study.

Table S7 List of the 22 peak SNPs associated with the abundances of METOs.

Table S8 Candidate genes of SIOS tolerance 1 (*ZmSOT1*) to *ZmSOT10*.

Table S9 Association between the SNP in *ZmUGT* and the METO34 content.

Table S10 Sequences of the DNA fragment flanking *ZmCYP709B2* in IL108 (with a HapG *ZmCYP709B2*) and CAU216 (with a HapC *ZmCYP709B2*).

Table S11 Situation of the 50-bp deletion (Del382) in 200 maize inbred lines.

Table S12 Association between the genetic variation in *ZmCYP709B2* and the METO31 content.

Table S13 List of the primers used in this study.

Please note: Wiley Blackwell are not responsible for the content or functionality of any Supporting Information supplied by the authors. Any queries (other than missing material) should be directed to the *New Phytologist* Central Office.

About *New Phytologist*

- *New Phytologist* is an electronic (online-only) journal owned by the New Phytologist Foundation, a **not-for-profit organization** dedicated to the promotion of plant science, facilitating projects from symposia to free access for our Tansley reviews and Tansley insights.

- Regular papers, Letters, Viewpoints, Research reviews, Rapid reports and both Modelling/Theory and Methods papers are encouraged. We are committed to rapid processing, from online submission through to publication 'as ready' via *Early View* – our average time to decision is <26 days. There are **no page or colour charges** and a PDF version will be provided for each article.

- The journal is available online at Wiley Online Library. Visit **www.newphytologist.com** to search the articles and register for table of contents email alerts.

- If you have any questions, do get in touch with Central Office (np-centraloffice@lancaster.ac.uk) or, if it is more convenient, our USA Office (np-usaoffice@lancaster.ac.uk)

- For submission instructions, subscription and all the latest information visit **www.newphytologist.com**

See also the Commentary on this article by Fernie, **230**: 2091–2093.