An InDel in the Promoter of AL-ACTIVATED MALATE TRANSPORTER9 Selected during Tomato Domestication Determines Fruit Malate Contents and Aluminum Tolerance

Jie Ye, a Xin Wang, a Tixu Hu, a Fengxia Zhang, a Bing Wang, a Changxin Li, a Tianxia Yang, a Hanxia Li, a Yongen Lu, a James J. Giovannoni, b,1 Yuyang Zhang, a,1 and Zhibiao Ye a,1

a Key Laboratory of Horticultural Plant Biology, Ministry of Education, Huazhong Agricultural University, Wuhan 430070, China
b U.S. Department of Agriculture and Boyce Thompson Institute for Plant Research, Cornell University, Ithaca, New York 14853

ORCID IDs: 0000-0001-7728-3588 (J.Y.); 0000-0002-0972-2515 (J.J.G.); 0000-0002-9602-583X (Z.Y.)

Deciphering the mechanism of malate accumulation in plants would contribute to a greater understanding of plant chemistry, which has implications for improving flavor quality in crop species and enhancing human health benefits. However, the regulation of malate metabolism is poorly understood in crops such as tomato (Solanum lycopersicum). Here, we integrated a metabolite-based genome-wide association study with linkage mapping and gene functional studies to characterize the genomics of malate accumulation in a global collection of tomato accessions with broad genetic diversity. We report that TFM6 (tomato fruit malate 6), which corresponds to AL-ACTIVATED MALATE TRANSPORTER9 (SI-ALMT9 in tomato), is the major quantitative trait locus responsible for variation in fruit malate accumulation among tomato genotypes. A 3-bp indel in the promoter region of SI-ALMT9 was linked to high fruit malate content. Further analysis indicated that this indel disrupts a W-box binding site in the SI-ALMT9 promoter, which prevents binding of the WRKY transcription repressor Sl-WRKY42, thereby alleviating the repression of SI-ALMT9 expression and promoting high fruit malate accumulation. Evolutionary analysis revealed that this highly expressed SI-ALMT9 allele was selected for during tomato domestication. Furthermore, vacuole membrane-localized SI-ALMT9 increases in abundance following Al treatment, thereby elevating malate transport and enhancing Al resistance.

INTRODUCTION

Plants produce numerous metabolites that are important for plant growth and tolerance of environmental stress (Schwab, 2003; Saito and Matsuda, 2010) and also serve as essential sources of fiber, energy, and nutrients in the human diet (Schauer et al., 2006). For example, the water-soluble metabolite ascorbic acid has been reported as a key antioxidant and modulator of plant development through hormone signaling (Pastori et al., 2003). Ascorbic acid can also assist in preventing human diseases such as cancer and diabetes by scavenging reactive oxygen species (Riso et al., 2004). Another organic acid, malic acid (or malate), affects plant growth, stress responses, and the flavor and nutrient quality of fruit. Malate is an intermediate metabolite in the citric acid and glyoxylate cycles (Sweetman et al., 2009) and is formed during carbon fixation in plants exhibiting C4 and crassulacean acid metabolism photosynthesis. High malate contents protect plants from Al toxicity via root malate efflux (Hoekenga et al., 2006). In addition, malate is essential for maintaining cellular osmotic pressure and charge balance in guard cells and thus directly regulates stomatal aperture (Meyer et al., 2010).

In tomato (Solanum lycopersicum) fruits, the primary organic acids are malate, citric acid, and tartaric acid, among which malate is a crucial compound that contributes to fruit flavor and palatability. Enhanced malate concentrations lead to altered starch metabolism and soluble solid contents in tomato, which subsequently affects postharvest fruit softening (Centeno et al., 2011). Malate has also been shown to enhance the perception of sucrose during fruit consumption (Lobit et al., 2006). Thus, consumers and plant breeders have directly or indirectly selected for desirable malate contents during agricultural variety development in the past; however, the specific genotypes associated with malate trait selection remain uncharacterized (Ma et al., 2015a).

Fruit malate content is determined by several processes, including malate synthesis, degradation, and intracellular transport (Sweetman et al., 2009). The malate synthesis and degradation pathways within glycolysis and the tricarboxylic acid cycle are well known, but the role that malate transport plays in final malate concentration remains poorly understood. Al-activated malate transporters (ALMTs) are plant-specific anion channel proteins that are activated by Al and function in the efflux of malate to chelate surplus Al³⁺, thereby protecting plant root tissue and function (Delhaize et al., 2007). The first ALMT gene, Ta-ALMT1, was identified in near-isogenic wheat (Triticum aestivum) lines (Sasaki et al., 2004) and was characterized as a plasma membrane protein that promotes Al tolerance in wheat and barley (Hordeum vulgare; Delhaize et al., 2004; Sasaki et al., 2004; Yamaguchi et al., 2005). In Ta-ALMT1, the Ser-384 residue regulates protein activity depending on its level of phosphorylation, which allows malate transport to be regulated in response to varying Al³⁺ concentration.
(Ligaba et al., 2009). Both plasma membrane (Bn-ALMT1 and Ta-ALMT1; Yamaguchi et al., 2005; Ligaba et al., 2007) and vacuolar membrane (At-ALMT9; Kovermann et al., 2007) ALMTs contain transmembrane N-terminal and hydrophobic C-terminal domains and contribute to the Al stress response (Motoda et al., 2007). The ALMT N terminus contains several transmembrane motifs and is predicted to function in anion transport (Ligaba et al., 2013), whereas the C terminus contains a long hydrophobic region located externally of the membrane that is specifically required for the activation of anion transport by Al (Furuichi et al., 2010). Recently, ALMT2 was presented as a candidate gene corresponding to the Ma locus that affects apple (Malus domestic) fruit acidity (Bai et al., 2012; Khan et al., 2012). A single nucleotide polymorphism (SNP; G to A) in the terminal ALMT2 exon causes premature termination of protein translation, which relocates ALMT2 from the vacuole to the plasma membrane, resulting in decreased fruit malate concentration (Ma et al., 2015b). Recently, it was reported that two tomato ALMTs, namely, Sl-ALMT4 and Sl-ALMT5, are involved in malate transport and that overexpression of Sl-ALMT5 increased both malate and citric acid in transgenic tomato seeds (Sasaki et al., 2016).

Tomato is an important source of organic acids, sugars, and antioxidant nutrients in the human diet. Furthermore, tomato serves as a model system for fruit biology studies. Genetic characterization of the natural variation in tomato fruit malate content therefore contributes to a greater understanding of the molecular regulation of fruit flavor. Such knowledge potentially facilitates the breeding of elite varieties with enhanced flavor and nutritional composition for human consumption. Traditional quantitative trait locus (QTL) linkage mapping is frequently used to investigate the genetic architecture of metabolic traits in tomato (Schauer et al., 2006; Stevens et al., 2007). Although QTL linkage mapping has been used to detect valuable QTLs specific to the parental lines of mapping populations, the efficiency and resolution of this approach are low due to a limited genetic diversity in the analyses and a low recombination frequency (Mauricio, 2001). With the currently available high-throughput genotyping technologies, genome-wide association studies (GWAS) have been developed as a powerful approach to characterize the genetic basis of complex traits and locus-locus interactions in plants (Nordborg and Weigel, 2008). More recent studies combining GWAS with metabolomics (mGWAS) demonstrate a powerful forward genetics strategy that can be used to dissect the genetic and biochemical basis of plant metabolism, including that in model species, such as Arabidopsis thaliana, tomato, rice (Oryza sativa), and Medicago truncatula, and other crop species, such as barley, perilla (Perilla ocymoides), and pepper (Capsicum annuum) (Achnine et al., 2005; Yamazaki et al., 2008; Widodo et al., 2009; Wahyuni et al., 2011; Carreno-Quintero et al., 2013; Luo, 2015). Moreover, recent studies have revealed that linkage disequilibrium mapping in combination with linkage mapping improves the ability to identify important loci. For example, mGWAS on 983 metabolites were performed in 702 maize genomes, which identified a number of significant locus/trait associations that were validated through linkage mapping (Wen et al., 2014). Also, both linkage disequilibrium and linkage mapping indicated that the underlying natural variation in leaf Cd accumulation in the global Arabidopsis population was caused by different Heavy Metal ATPase 3 (At-HMA3) alleles (Chao et al., 2012). Recently, an mGWAS analyzed 28 chemicals specifically associated with consumer preference and flavor intensity, using 2,014,488 common SNPs distributed among 398 modern, heirloom, and wild tomato accessions, which yielded a total of 251 associated loci for 20 traits, including malate content (Tieman et al., 2017).

Previously, Sauvage et al. (2014) conducted an mGWAS study using 19 primary metabolites and 5955 SNPs among 163 tomato accessions. This led to the identification of 44 major polymorphic loci that control the variation of 19 primary metabolites in natural populations. Among these loci, two SNPs located on chromosomes 2 and 6 were associated with fruit malate content; however, only the association with the SNP located on chromosome 6, positioned 7.9 kb from a putative candidate gene (Solyco06g072910.2), was determined as being significant. A limited number of strong associations between loci and fruit malate content may be attributed to the low number of SNPs (5955) used in this study. This is further supported by the observations of Shirasawa et al. (2013), who used a population of 663 accessions and 1536 SNPs and yet identified only a single SNP linked to fruit color, which was associated with Sl-MYB12. By contrast, using a population of 360 accessions and a greater number of SNPs (5 Mb), Lin et al. (2014) identified more than 100 SNPs that were significantly associated with fruit color, and the highest-ranked SNP was significantly linked to Sl-MYB12. Furthermore, the function and influence of the Solyco06g072910.2 candidate gene on fruit malate content variation remains largely unexplored.

In this study, we analyzed an enlarged mapping population of 272 accessions (Lin et al., 2014; Tieman et al., 2017) with GWAS methods to identify reliable and high precision QTLs that contribute to fruit malate content in a diverse tomato population. Furthermore, we integrated the mGWAS approach with analyses of gene expression patterns, genetic variations, and T-DNA (CRISPR/Cas9)-derived mutants to functionally characterize a malate-content-related QTL. We identified a major QTL, TOMATO FRUIT MALATE ON CHROMOSOME 6 (TFM6), and found it to contribute toward high malate content in both fruit and root tissue. Furthermore, the high malate content improved both fruit flavor and Al stress resistance in cultivated tomato plants harboring TFM6.

RESULTS

Combined GWAS and BSA for Efficient Mapping of Fruit Malate Content in Tomato

From more than 11.6 million SNPs identified in 30,945 annotated genes (Lin et al., 2014), 5.5 million high-quality SNPs were selected for this study. These genotyped lines were subsequently processed via a quality control protocol (see Methods). We observed consistent fruit malate contents between two experimental repeats, and the broad-sense heritability (H^2) and coefficient of variation were 19.1% and 37.6%, respectively (Figure 1A; Supplemental Data Set 1). A number of accessions within the BIG and CER variety groups showed high fruit malate content, and several of these accessions also were present in the PIM group (Supplemental Figure 1). In this GWAS, we observed that a single region on chromosome 6 contained multiple SNPs that were highly associated with fruit malate content, consistent with
previous results (Figures 1B and 1C; Sauvage et al., 2014). We therefore designated this locus TFM6. Within TFM6, we observed 83 SNPs that were significantly associated with fruit malate content (P value < $1.82 \times 10^{-7}$, compressed mixed linear model), of which 31 displayed significant association with a more stringent P value (P value < $1 \times 10^{-15}$; Supplemental Table 1). The SNP with the highest association to fruit malate content was located at ch06_41345866 (P value = $1.76 \times 10^{-20}$) and explained 39.1% of the total variance observed in fruit malate contents. This finding suggested that a genomic sequence related to SNP ch06_41345866 forms the major genetic locus responsible for the natural variation in malate contents of cultivated tomato fruit. There was a total of 24 genes within the 50-kb sequence either side of SNP ch06_41345866, which was less than the linkage disequilibrium decay distance (Figure 1C; Supplemental Table 2). The genes in this genomic region included two ALMTs; thus, given that ALMTs have been shown to function as malate transporters (Hoekenga et al., 2006; De Angeli
et al., 2013a, 2013b), these two genes were considered as causal gene candidates for TFM6.

To further genetically characterize the fruit malate content QTL on chromosome 6, we generated an experimental F2 population that displayed segregation of the majority of malate-content-linked SNPs on chromosome 6. For this, the tomato accession TS-66 with low fruit malate content (1.99 μmol/g fresh weight [FW]) was crossed with TS-40, which has high fruit malate content (18.79 μmol/g FW). The F1 generation resulting from this cross displayed fruit malate contents that were significantly lower than those of the TS-40 parent and slightly higher than that of the TS-86 parent (Supplemental Figure 2). A similar distribution of malate contents was observed in the F2 generation, whereby the majority of individuals possessed fruit malate contents closer to that of the TS-66 parent (Figure 1D). Bulk segregant analysis (BSA) was performed using the individuals within the F2 mapping population (Supplemental Data Set 2). We sequenced two bulk populations, namely, the HM-pool and LM-pool, which comprised individuals with extremely high or low fruit malate contents, respectively. Each population consisted of 40 progeny plants from the F2 population of 350 individuals and was sequenced to a depth of 40 genome equivalents (Figure 1E; Supplemental Data Set 2). The HM-pool exhibited an average fruit malate concentration of 12.84 μmol/g FW, which is more than four times that of the LM-pool (3.08 μmol/g FW; Figure 1E). We called SNPs between the two parental genomes and calculated the SNP indices for the high- and low-malate bulk populations, as well as the difference between them (ΔSNP index), using a 1000-kb sliding window with a step size of 10 kb (Figure 1F). Based on the allele frequency differences between the two pools, the causal locus for fruit malate content was mapped to a 9.6-Mb interval on chromosome 6 (~36–45.6 Mb; Figures 1F and 1G), with the peak centered on the mapping interval identified in our GWAS analysis (Figure 1B). Furthermore, several SNPs identified by the GWAS were located within or near the genomic DNA region of the two aforementioned ALMT genes (Solyco06g072910 and Solyco06g072920) and displayed a considerably high ΔSNP index (Figure 1G; Supplemental Data Set 3). Thus, the results of our linkage mapping in the F2 population from TS-40 × TS-66 were consistent with those from our GWAS analysis and further indicated that ALMT on chromosome 6 was a causal gene candidate for the observed natural variation in fruit malate content.

**A 3-bp Deletion in the SI-ALMT9 Promoter Alters Its Expression**

Through NCBI BLASTP analysis and PCR amplification of cDNA, two candidate ALMT genes in the mapped region of chromosome 6, namely, Solyco06g072910 and Solyco06g072920, were found to correspond to a single gene (SI-ALMT9) separated by a 3708-bp intron (Supplemental Figure 3). Furthermore, NCBI BLASTN analysis revealed that the long terminal repeat (LTR) retrotransposon CopiaSL_37 resided in the SI-ALMT9 3708-bp intron (Paz et al., 2017). SI-ALMT9 contained six exons and five introns and encoded a 559-amino acid protein, which was predicted to contain five transmembrane domains and to have a molecular mass of 62.54 kD (Supplemental Figure 4). SI-ALMT9 exhibited the highest amino acid sequence similarity to Ma of apple (58%) and At-ALMT9 of Arabidopsis (57%; Supplemental Figures 5A and 5B and Supplemental Data Set 4).

As both GWAS and linkage mapping indicated that SI-ALMT9 was the most likely causal candidate gene for the natural variation of tomato fruit malate content, we sequenced the genomic region surrounding this locus, including that corresponding to the upstream promoter sequence. Sequencing was performed on 10 accessions with high fruit malate contents (average 9.73 μmol/g FW), which were 64.18% higher than that of a further seven sequenced accessions with low fruit malate contents (average 5.93 μmol/g FW; Figure 2A). Sequence analysis suggested that the SI-ALMT9 genotype can be classified into two different haplotypes in cultivated tomato, namely, SI-ALMT9_hMH for the high malate-content phenotype and SI-ALMT9_LMH for the low malate-content phenotype (Figure 2B; Supplemental Figures 6A and 6B). Comparative analysis of the SI-ALMT9_hMH and SI-ALMT9_LMH sequences in 155 tomato accessions showed that only 8 of 29 polymorphisms were tightly associated with the malate-associated SNP ch06_41345866 (Supplemental Data Set 5), and these included one indel (indel_3, ->GTC variant at position –423 to –421) in the promoter region, five SNPs located in introns (T>C variant at position 872, A>G variant at position 1978, A>G variant at position 2356, C>T variant at position 3357, and G>A variant at position 3941), one nonsynonymous polymorphism in exon 5 (G>A variant at position 6094, with amino acid change from R to H), and one synonymous SNP in exon 6 (G>A variant at position 5185, ch06_41343002). Interestingly, CopiaSL_37 in the second intron was observed in some accessions within BIG and CER variety groups but was not observed in those in the PIM variety group and in wild species, indicating that CopiaSL_37 insertion into SI-ALMT9 likely occurred during the recent domestication of tomato from the PIM to CER variety.

To determine whether the variation in malate contents between the two SI-ALMT9 haplotypes was due to polymorphisms in the promoter or coding regions, we generated transgenic plants using four constructs: SI-ALMT9_hMH pro:SI-ALMT9_hMH, SI-ALMT9_LMH pro:SI-ALMT9_LMH, SI-ALMT9_hMH pro:SI-ALMT9_LMH, and 35S pro:SI-ALMT9_hMH (Figure 2C, constructs II, III, IV, and OX, respectively; see Methods). In comparison to the wild-type TS66, the transgenic lines expressing SI-ALMT9_hMH pro:SI-ALMT9_hMH, SI-ALMT9 hMH pro:SI-ALMT9 hMH, and 35S pro:SI-ALMT9 hMH displayed enhanced SI-ALMT9 expression that was accompanied by increased malate contents, indicating that malate accumulation was governed by SI-ALMT9 expression level as opposed to coding sequence variance. The transgenic lines expressing SI-ALMT9 hMH pro:SI-ALMT9 hMH, SI-ALMT9 hMH pro:SI-ALMT9 hMH, and 35S pro:SI-ALMT9 hMH did not accumulate malate to levels comparable to those in TS66, suggesting that the indel_3 identified in the SI-ALMT9 promoter (~GTC) was the causal QTL for fruit malate contents (Figures 2D to 2F). Furthermore, cosegregation of the indel_3 genotype within the BSA HM-pool (GTC deletion) and LM-pool (GTC insertion) accessions, as revealed by our resequencing results, further indicated that the indel_3 in SI-ALMT9 was strongly associated with fruit malate content (Supplemental Figure 7). Combined, these results identify an indel of the GTC sequence from –423 to –421 bp in the SI-ALMT9 promoter as the causal genetic element for TFM6 and, therefore, the variation in fruit malate content among cultivated tomato. The GTC deletion of this indel results in increased
Figure 2. Comparative Analyses of the Sl-ALMT9 Locus between Low-Malate and High-Malate Haplotypes.

(A) Fruit malate content of the 17 accessions can be grouped into two bulks, namely, that with high-malate contents and that with low-malate contents according to the presence of either the Sl-ALMT9<sup>LMH</sup> or Sl-ALMT9<sup>HMH</sup> haplotype, respectively.

(B) Structural variations of the Sl-ALMT9 haplotypes in low-malate (blue) and high-malate (red) varieties. Bold lines represent the promoter and 3' UTR, boxes represent coding sequences, and the thin lines between boxes represent introns. Nucleotide polymorphisms are indicated at their corresponding positions.

(C) The Sl-ALMT9 expression constructs used to generate experimental transgenic lines. Construct I: Sl-ALMT9 expression cassette as it exists in wild-type TS66. Construct II: the 2.7-kb promoter region from low-malate variety TS66 (Sl-ALMT9<sup>LMH pro</sup>) driving expression of the 1.8-kb ORF from high-malate variety TS40 (Sl-ALMT9<sup>HMH</sup>). Construct III: the 2.7-kb promoter region from high-malate variety TS40 (Sl-ALMT9<sup>HMH pro</sup>) driving expression of the 1.8-kb ORF from low-malate variety TS66 (Sl-ALMT9<sup>LMH</sup>). Construct IV: the 2.7-kb promoter region from high-malate variety TS40 (Sl-ALMT9<sup>HMH pro</sup>) driving expression of the 6.2-kb ORF from high-malate variety TS40 (Sl-ALMT9<sup>HMH</sup>). Construct OX: the CaMV 35S promoter (35S<sub>pro</sub>) driving expression of the 6.2-kb ORF from high-malate variety TS40 (Sl-ALMT9<sup>HMH</sup>). Genomic coordinates are from SL2.4.

(D) Fruit malate content (n = 3) in transgenic plants expressing constructs I to V described in (C) (*P < 0.05, **P < 0.01; t test).

(E) and (F) Comparison of Sl-ALMT9 expression levels in red ripe fruit from transgenic plants expressing constructs I to V described in (C). Data in (E) means ± sd (n = 3). Actin depicted in lower panel of (F) was used as the loading control.
Sl-ALMT9 expression levels, leading to high fruit malate contents in cultivated tomato.

Cosegregation between Malate Content and Sl-ALMT9indel_3 Genetic Markers

A cleaved amplified polymorphic sequence (CAPS) marker was developed based on the ALMT9\textsuperscript{indel,3} polymorphism (Figures 3A and 3B). This marker enabled codominant scoring of commercial tomato cultivars (F1 hybrids) that produce fruit with large variation in malate contents and are popular in certain markets such as China. All cultivars were genotyped using the ALMT9\textsuperscript{indel,3} marker and their ripe fruits were analyzed for malate contents (Supplemental Figure 8). Among the 73 F1 hybrids tested, 11 cultivars were genotyped as almt9/almt9, 21 as ALMT9/ALMT9, and 41 as ALMT9/almt9 (Supplemental Table 3). The complete genotyping accuracy of the CAPS marker was verified by subsequent Sanger sequencing. The CAPS marker also displayed cosegregation with malate content, whereby ALMT9/ALMT9 and ALMT9/almt9 had lower malate contents than the high malate contents of almt9/almt9 (Figure 3C).

Figure 3. Genotyping of Commercial Tomato Cultivars Using an indel_3-Based CAPS Marker.
(A) CAPS genotyping of different accessions based on indel_3. The indel_3-based CAPS marker was developed using sequence polymorphisms between the Sl-ALMT9\textsuperscript{indel,3} of high- and low-malate tomato. Three different genotypes were detected, namely, ALMT9/ALMT9, almt9/almt9, and ALMT9/almt9.
(B) Determination of fruit malate content in varieties and accessions representing the genotypes in (A).
(C) CAPS genotyping of 73 commercial tomato cultivars (F1 hybrids) compared with their malate contents. Cultivars with high fruit malate content were associated with the almt9/almt9 genotype (b). Cultivars with low fruit malate content were associated with either the ALMT9/ALMT9 (a) or ALMT9/almt9 (h) genotypes. Malate content data shown are means ± sd (n = 3).

Sl-ALMT9 Positively Regulates Fruit Malate Content and the Associated Enhanced Fruit Quality and Al Stress Tolerance in Roots

Since tomato is a common model system in fruit biology studies, we investigated the role of Sl-ALMT9 in malate content dynamics during fruit development (Figure 4A). In contrast to the high fruit malate content in TS40, the fruit of TS66 displayed a continuous decline in malate content during each of the nine development stages. At the BR developmental stage, significantly higher malate levels were measured in fruit from the three Sl-ALMT9\textsuperscript{OX} transgenic lines compared with that in TS66, and these differences were maintained in the RR developmental stage (Figure 4B). Considerably higher Sl-ALMT9 expression was measured in fruit of TS40 compared with that of TS66, in particular during later development stages, which reflected the rate of fruit malate accumulation (Figures 4B and 4C). We also investigated the expression levels of fumarase (FUM) and malate dehydrogenase (MDH), which are involved in malate synthesis and degradation, respectively (Supplemental Figure 9). Significantly lower MDH expression was observed in later fruit developmental stages, which may underlie the considerably higher malate content in
Figure 4. Dynamics of Malate Content and SI-ALMT9 Expression during Tomato Fruit Development.

(A) Morphology of fruit from TS66 at nine developmental stages. Numbers above images refer to the number of days after flowering. Br, breaker stage; Br+2, 2 d after breaker stage; YR, yellow ripening; RR, red ripening.

(B) Dynamics of fruit malate content at the developmental stages described in (A) in TS66, TS40, and three SI-ALMT9 OX lines. Data are means ± SD (n = 3).

(C) Relative expression of SI-ALMT9 in TS66 and TS40 at the developmental stages described in (A).

(D) Heatmap showing the percent of genes upregulated and downregulated in TS66 compared to TS40 across different biological processes, molecular functions, and cellular components.

(E-I) Bar graphs showing relative expression of selected genes in TS66 and TS40.
TS40 fruit compared with that in the three SI-ALMT9 OX transgenic lines (Supplemental Figure 9A).

Previous reports indicate that fruit malate content is considerably higher in green-fruited wild species than in red-fruited species (PIM) and cultivars (BIG; Schauer et al., 2005). Therefore, we analyzed the correlation between fruit malate content and SI-ALMT9 expression level in these different tomato lines. Interestingly, we found higher SI-ALMT9 expression in wild species with high fruit malate content (LA0444 and LA0719) compared with that in red-fruited varieties (PIM, CER, and BIG) and the high-level fruit malate content accession TS40 (Supplemental Figure 10). This result demonstrates the high correlation between SI-ALMT9 expression level and malate contents in a wide range of tomato varieties and accessions.

A number of sugars and organic acids were significantly altered in SI-ALMT9 OX fruit at the RR developmental stage, which likely influenced fruit flavor (Supplemental Table 4). Compared with that in TS-66 wild-type fruit, sucrose, inositol, citric acid, and aminobutanoic acid contents were significantly increased in SI-ALMT9 OX lines, whereas galacturonic acid content was significantly decreased (Supplemental Table 4). These metabolite changes suggest that SI-ALMT9 regulates malate accumulation and, in turn, malate plays a central role in fruit primary metabolism in tomato.

RNA-seq analysis of wild-type TS66 and SI-ALMT9 OX lines was employed to investigate the potential SI-ALMT9 regulation network (Figure 4D; Supplemental Data Set 6). Among the 1887 differentially expressed genes, several core genes involved in sugar and organic acid metabolism were differentially expressed in the SI-ALMT9 OX lines (Figure 4E), which partially corresponded with changes in sugar and organic acid contents (Supplemental Table 4). The expression of PR genes was also altered in the OX lines (Figure 4F), which is in agreement with a previous study that suggested tomato fruit malate content is involved in pathogen resistance (Centenno et al., 2011). Furthermore, members of the multidrug and toxin efflux (MATE) protein family, which are alternative malate transporters, displayed either upregulated (Sl-MATE5 and 6) or downregulated (Sl-MATE9) gene expression levels (Figure 4G). Moreover, we observed enhanced gene expression for three sugar transport proteins and four other vacuole/plasma membrane proteins in SI-ALMT9 OX lines (Figures 4H and 4I).

To elucidate how SI-ALMT9 regulates malate content, we mutagenized SI-ALMT9 in vivo using CRISPR/Cas9 in the SI-ALMT9Accession Ailsa Craig (AC; Figure 5A). Deletions in SI-ALMT9 were detected by PCR and further confirmed by DNA sequencing. Using quantitative RT-PCR (RT-qPCR), we found that SI-ALMT9 expression was downregulated in five isolated almt9 mutant lines, and almost no SI-ALMT9 expression was detectable in almt9-5 (Figure 5D). Furthermore, CR-almt9-5 contained a 93-bp deletion in SI-ALMT9 that included the entire 74-bp spacer sequence between the CRISPR/Cas9 target sites (Figures 5B and 5C). Consistent with reduced SI-ALMT9 expression, lower fruit malate contents were observed in the five almt9 mutants compared with that in wild-type AC (Figure 5E). It should be noted that the decreased fruit malate contents in almt9 mutants (AC background) remained higher than that in a number of naturally occurring SI-ALMT9Accession accessions, such as TS66, and that the enhanced fruit malate content following SI-ALMT9 OX in TS66 remained lower than that in a number of SI-ALMT9Accession accessions, such as AC and TS40. These varying fruit malate contents may be attributable to the effect of different genotypic backgrounds.

Phylogenetic analysis has placed SI-ALMT9 into the same clade as the tonoplast-located proteins At-ALMT6, At-ALMT9, and Ma1 (Supplemental Figure 5B and Supplemental Data Set 4), of which Ma1 plays a key role in regulating apple malate content (Ma et al., 2015b). To investigate SI-ALMT9 cellular localization, we created SI-ALMT9-GFP fusion proteins, which were transiently expressed in Nicotiana benthamiana. The SI-ALMT9-GFP fluorescence signal overlapped with that of CBL6-OPF, a marker for the vacuolar membrane (Batistic et al., 2010), suggesting that SI-ALMT9 is located to the tonoplast (Figure 6A).

The SI-ALMT9 expression pattern was investigated in tomato plants carrying SI-ALMT9Accession, whereby expression of the GUS reporter gene was driven by the putative SI-ALMT9 promoter. GUS staining of these transgenic plants revealed high SI-ALMT9 promoter activity in both main and lateral roots following Al treatment under low pH conditions, which was considerably higher than that observed in roots under normal conditions (Figure 6B). The root-specific and Al-responsive expression of SI-ALMT9 suggested that SI-ALMT9 could function in malate transport for the detoxification of Al. Root malate contents were significantly greater in SI-ALMT9 Accession than in wild-type TS66 (Supplemental Figure 11).

The role of SI-ALMT9 in Al stress tolerance was assessed by measuring the root length of tomato plants grown in hydroponic culture supplemented with Al. SI-ALMT9 OX lines and TS40 (the high fruit malate content line from which SI-ALMT9 was originally obtained) displayed robust root growth in hydroponic culture supplemented with Al at a concentration that severely inhibited root growth in wild-type TS66 (Figures 6C and 6D). The extent of malate efflux from the hairy roots of these lines was closely correlated with root growth, with significantly greater root growth in the SI-ALMT9 OX lines and TS40 compared with that measured...
in the wild-type TS66 (Figure 6E). Hematoxylin forms a purple/red complex with Al thus providing an indirect measure of non-complexed Al in root apices; therefore, the intensity of hematoxylin staining is correlated with plant Al sensitivity. The apices of wild-type roots displayed extensive hematoxylin staining, indicating high-level Al accumulation, whereas minimal hematoxylin staining was observed in roots of the Sl-Almt9 lines (Figure 6F).

Overall, these results suggest that Sl-ALMT9 functions in tomato roots as a malate efflux transporter that promotes malate accumulation at the root surface to chelate free Al and protect root tissue, thereby facilitating Al tolerance.

**Sl-ALMT9 Expression Is Directly Regulated by the Transcription Factor Sl-WRKY42**

In plants, WRKYs are reported to directly target a core W-box binding motif [(T)(T)TGAC(C/T)] (Eulgem et al., 2000) present in the promoters of target genes. Previous reports indicated that At-WRKY46 acts as a negative regulator of At-ALM1 by directly binding to its promoter at W-boxes and that mutation of WRKY46 led to increased root malate efflux and reduced Al accumulation in root apices (Ding et al., 2013). Yeast one-hybrid and transient expression analyses indicated that Sl-WRKY42, an At-WRKY46 homolog, is a candidate protein binding to cis-elements in the Sl-ALMT9 promoter. Nine W-boxes are present in the Sl-ALMT9 promoter, located at positions 2363, 2263, 1308, 1274, 701, 506, 463, 420, and 85 bp upstream of the translation initiation codon (Figure 7A; motifs 1–9, respectively). The fruit malate content-associated 3-bp indel (GTC) is located in the eighth W-box (Figures 7A and 7B). Yeast one-hybrid results showed that yeast cells transformed with Sl-ALMT9HMH grown more slowly and with less binding activity than those transformed with Sl-ALMT9LMH, which was more evident with increasing abscisic acid (ABA) concentration (Figure 7C). Taken together, these results suggest that the eighth W-box in the Sl-ALMT9 promoter is critical for interaction with Sl-WRKY42. To further investigate this interaction,
we generated luciferase (LUC) reporter constructs containing a promoter sequence from −1523 to the start codon ATG derived from either Sl-ALMT9LMHpro or Sl-ALMT9HMHpro, which either contained or lacked the eighth W-box, respectively. Alongside effector constructs containing Sl-WRKY42, these promoter-LUC reporter constructs were transiently expressed in N. benthamiana via infiltration with Agrobacterium tumefaciens (GV3101) strains. Analysis of infiltrated tissue revealed that although the use of both Sl-ALMT9LMHpro and Sl-ALMT9HMHpro promoter sequences activated LUC signal, the Sl-ALMT9HMHpro sequence was associated with significantly higher LUC activity (Figure 7D). These results indicate that Sl-WRKY42 can directly bind to the eighth W-box in the Sl-ALMT9 promoter to modify Sl-ALMT9 expression and that Sl-WRKY42 directly affects Sl-ALMT9 expression to negatively regulate tomato fruit malate content.

**Human Selection May Have Altered Sl-ALMT9 in Modern Tomato Varieties**

To examine the evolutionary history of the TFM6 locus, DNA sequence variation in the genomic region spanning part of the Sl-ALMT9 coding region and promoter was investigated. The Sl-ALMT9 region showed reduced nucleotide diversity (π) in BIG (0.0001526 and 0.001353 for the promoter and coding region, respectively) compared with that in CER (0.003168 and 0.002707) and PIM (0.006587 and 0.004115) (Figure 8A). The ratio of nucleotide diversity of PIM to CER (πPIM/πCER) and CER to BIG (πCER/πBIG) at indel_3 of Sl-ALMT9 was 3.545 and 29.56, respectively (Figure 8B), which is greater than 3 and 6.9, and therefore suggestive of a domestication event (Lin et al., 2014). A similar trend was observed for CopiaSL_37 within the Sl-ALMT9 locus, in that πPIM/πCER was higher than 3; thus, this polymorphism was likely created by selection during the domestication of tomato from the PIM variety to the BIG and CER varieties (Figure 8B). These molecular diversity findings are also supported by phylogenetic analysis using Sl-ALMT9 gene sequences, which revealed single-clade grouping of HMH haplotypes, which are CER and BIG. Genotypes containing indel_3 were grouped into a single subclade that included a number of CERCers and cultivated tomato accessions, suggesting that the indel mutation arose in the CER background (Supplemental Figure 12 and Supplemental...
By contrast, the CopiaSL\_37 allele was detected only in wild tomatoes and a number of CERCERs, suggesting that the mutation arose in the PIM background. Moreover, we examined Sl\_ALMT9 evolution by conducting a cross-population composite likelihood ratio test (XP-CLR) on the TFM6 locus between different subspecies (Figure 8C). The value of XP-CLR of PIM to CER and CER to BIG at indel\_3 of Sl\_ALMT9 was considerably higher, specifically 14.03 and 9.26, respectively (Figure 8C). Again, these results are suggestive of a domestication event.

The geographic distribution of a genetic locus may indicate whether a particular allele had been selected for in a certain environment. To further explore the origin of the Sl\_ALMT9\_indel\_3 allele, we genotyped the mutation in 322 diverse accessions covering the proposed trajectory of tomato domestication (Supplemental Data Set 9). The high fruit malate-associated indel\_3 was present at a frequency of 23.2% in CER accessions that were domesticated from the wild progenitor variety PIM. The frequency of indel\_3 presence increased significantly in the BIG landraces from Europe and Asia (Figure 8D). Therefore, selection for higher fruit malate content and Al\(^{3+}\) tolerance likely resulted in the inadvertent selection of Sl\_ALMT9 on chromosome 6, thus representing a genomic signature of modern tomatoes. Moreover, we
Figure 8. Molecular Diversity Analysis of SI-ALMT9 during Domestication and Improvement.

(A) Distribution of nucleotide diversity ($\pi$) of the PIM (S.p, red solid line), CER (S.l.c, red dotted line), and BIG (S.I.I, red dashed line) within the 8-kb region of TFM6 covering the promoter and coding region of SI-ALMT9. Gray lines indicate the locations of indel_3 (GTC/-), LTR retrotransposon (CopiaSL_37), and SNP6 (A/G).

(B) The ratio of nucleotide diversity ($\pi$) is calculated from the SI-ALMT9 sequences of S. pimpinellifolium (S.p) with S. l. cerasiforme (S.l.c) (blue line) and S. l. cerasiforme (S.l.c) with S. l. lycopersicum (red line). Vertical black dashed line in (A) and (B) indicates the dividing of the promoter and the coding region of SI-ALMT9.

(C) The level of XP-CLR within the region of TFM6 between different subspecies (up, S. l. cerasiforme and S. l. lycopersicum; down, S. l. cerasiforme and S. pimpinellifolium). In total, 331 tomato accessions were used for analysis, including 53 S. pimpinellifolium accessions, 112 S. l. cerasiforme accessions, and 166 S. l. lycopersicum accessions.
plotted the geographical distribution of the two indel_3 alleles of SI-ALMT9, which showed that most of the accessions originating from South America, including those from Ecuador and Peru where wild tomato relatives originated, carried the “GTC/GTC” genotype (Figure 8E). The “−/−” genotype frequency increased slightly in the Mesoamerican BIG accession, was further prevalent in the European and Asian BIG landraces, and appeared almost systemic in modern varieties (Figure 8E; Supplemental Data Set 9). This analysis suggests that high fruit malate content selection may not have been performed until recently or that the indel_3 SI-ALMT9 allele was not associated with this trait during early tomato breeding efforts.

DISCUSSION

Species distribution and human selection are key processes in plant domestication and improvement, as they may contribute to geographical differences in preferences for cultivated species and their traits (Gross and Olsen, 2010). Here, using GWAS based on ~5.5 million SNPs obtained from a diverse world-wide collection of 360 tomato accessions (Lin et al., 2014), we successfully identified a single strong peak of SNPs associated with fruit malate accumulation near SI-ALMT9 (Figure 1). To confirm the GWAS result and provide further insight into the genetics of malate accumulation, we performed linkage mapping, transgenic complementation experiments, functional segregation analyses, and molecular biology experiments to demonstrate that polymorphisms of indel_3 in the SI-ALMT9 promoter are the major genetic determinant for the variation we observe in fruit malate content in cultivated tomatoes. Our result demonstrates that mGWAS combined with linkage analysis as well as functional validation provides a powerful tool to quickly identify new genes associated with fruit quality and stress tolerance in plants.

SI-ALMT9 encodes an Al-activated malate transporter with the highest similarity to the apple Ma1, which controls fruit malate transport and accumulation (Bai et al., 2012; Khan et al., 2012; Ma et al., 2015b). This highly conserved function of the ALMT subfamily in regulating malate transport led us to postulate that SI-ALMT9 was the most likely candidate gene for the TF/6 QTL. The proteins in Arabidopsis most similar to SI-ALMT9 are At-ALMT6 and At-ALMT9, which localize to the vacuolar membrane and are expressed in guard cells of leaves (Meyer et al., 2011; De Angeli et al., 2013a). Both At-ALMT6 and At-ALMT9 play a role in vacuolar malate transport in guard cells; At-ALMT6 is a Ca²⁺-activated malate channel that does not function in regulating stoma movement and At-ALMT9 is a malate-activated vacuolar chloride channel that does control stomatal aperture and thereby stomatal conductance and drought tolerance. Other members of this family, e.g., Ta-ALMT1, At-ALMT1, Bn-ALMT1, Bn-ALMT2, and Gm-ALMT1, control leaf malate accumulation, malate exudation in root, and affect Al resistance (Delhaize et al., 2004; Hurth et al., 2005; Hoekenga et al., 2006; Ligaba et al., 2006; Kobayashi et al., 2013; Liang et al., 2013). In this work, SI-ALMT9^AlM^ accessions (e.g., TS40) and SI-ALMT9^LM^ accessions (e.g., TS66) showed significant differences in malate contents in fruits, but contained comparable amounts in leaves. In addition, SI-ALMT9-OX transgenic lines did not exhibit altered leaf malate content. These results point to a divergent regulatory mechanism for ALMT9 in spite of high similarity in conserved domains.

We found that variation in transcript abundance, rather than variation in the protein-coding sequence in SI-ALMT9 was responsible for the SI-ALMT9-dependent variation in fruit malate. This finding is consistent with previous reports that attributed natural variation in tomato fruit size and rice grain size to expression variation in Sl-KLUH and OsSPL13, respectively (Chakrabarti et al., 2013; Si et al., 2016). Genetic complementation tests with four transgene constructs representing four haplotypes (II, III, IV, and OX) showed that the TS66 loss-of-function haplotype cannot be complemented by SI-ALMT9 alleles from haplotype construct II, which indicates that construct II haplotypes also represent loss-of-function alleles. By contrast, the construct III and IV haplotype and SI-ALMT9 overexpression lines were able to complement the loss-of-function allele in TS66, confirming that these two haplotypes represent a functional allele of SI-ALMT9. We conclude that the genetically determined natural variation in fruit malate observed in our worldwide tomato population is largely driven by variation in the expression of SI-ALMT9. The sequence differences between the high-malate pool and low-malate pool tomato haplotypes of SI-ALMT9 allowed us to further conclude that deletion_3 in the promoter of SI-ALMT9 was responsible for the performance of the SI-ALMT9 alleles in haplotype groups III and IV. Furthermore, we confirmed that the indel_3 mutation was located within a putative W-box element (Eulgem et al., 2000), which could be directly targeted by SI-WRKY42 and function as a negative regulator of SI-ALMT9 expression. A similar hypothesis has been proposed to explain the action of Arabidopsis ALMT1 (Ding et al., 2013). Although further evidence is necessary to confirm the biochemical function of SI-WRKY42, our discoveries contribute to the understanding of the functional mechanism of SI-ALMT9 and other Al-activated malate transporters.

Unlike secondary metabolites, which have a less complex genetic architecture but great diversity (Chan et al., 2011; Suhre et al., 2011; Chen et al., 2014), natural variation in primary metabolites and other agronomic traits tends to be controlled by a large number of small-effect loci (Keurentjes et al., 2006; Rowe et al., 2008; Huang et al., 2010, 2011; Chan et al., 2011; Suhre et al., 2011; Joseph et al., 2013). In a linkage mapping analysis of introgression lines in which marker-defined genomic regions of a cultivated S. lycopersicum variety M82 (BIG group) were replaced with homologous intervals of the wild species S. pennellii, malate QTLs were found mainly on chromosomes 1, 2, 5, and 7 but...
not on chromosome 6 (Schauer et al., 2006). However, here, only one major locus spanning Si-ALMT9 on chromosome 6 was identified, while other core genes that have been known to be involved in malate synthesis and degradation were not recovered in our GWAS analysis (Centeno et al., 2011). In the promoter of the S. pennellii allele Sp-ALMT9, indel_3 was neither GTC nor a deletion, but ATG, which might not be targeted by WRKY42 (Supplemental Figure 13A) and might lead to S. pennellii acquiring a higher fruit malate content. While the ALMT9\textsuperscript{indel_3} marker exists in M82 and S. pennellii, additional SNPs and insertions found in ALMT9 and at additional loci throughout the genome may influence its performance in S. pennellii (Supplemental Figures 13B and 13C). Si-ALMT9 showed the highest sequence similarity to Si-ALMT4, but with different expression and protein localization. Specifically, Si-ALMT4 displays constitutive expression and is located in the endoplasmic reticulum (Sasaki et al., 2016), while Si-ALMT9 showed Al-induced expression in the root and vacuole membrane localization (Figures 6A and 6B). Our results indicate that Si-ALMT9 mainly promotes fruit malate accumulation and root Al resistance by active vacuole membrane malate effusion (Supplemental Figure 14). Si-ALMT4 was previously reported to affect malate concentration in seeds; this difference in function thereby promotes seed dispersal.

In summary, we identified the /¬ genotype at indel_3 linked to the Si-ALMT9\textsuperscript{LMH} allele, which tended to be in Europe and Asia (Figure 8C). This finding raises the question of whether the effect of Si-ALMT9\textsuperscript{LMH} is neutral or provides an adaptive benefit to humans or plants under certain environmental conditions. Our findings indicate reduced nucleotide diversity in BIG in the Si-ALMT9 region compared with PIM and CER. Parts of the genomic region around Si-ALMT9 showed significant XP-CLR values, which may suggest an excess of low frequency SNPs in certain regions of Si-ALMT9 in tomato, as a result of a population expansion after a bottleneck or a selective sweep. Taken together, these results support a selective pressure around the TFM6 region, but this selection is likely to occur posterior to domestication because the reduction is found in both CER and BIG but not in PIM. We could speculate that the selection of hyperfunctional Si-ALMT9 is beneficial in areas of elevated Al\textsuperscript{3+} where enhanced vascular transport of Al\textsuperscript{3+} would potentially improve the plant’s tolerance to Al\textsuperscript{3+}. Also, due to the diverse consumption habits of humans and other animals (e.g., birds), the malate content in fruit would help improve fruit quality and flavor that attracts consumption and thereby promotes seed dispersal.

In summary, we identified Si-ALMT9, which is required for malate accumulation in tomato fruit during ripening and in Al-stressed roots. Phylogenetic analysis indicates that the Si-ALMT9\textsuperscript{LMH} genotype originated through three evolutionary events: LTR retrotransposon CopiaSL_37 insertion into the second intron of the wild tomato, producing the Si-ALMT9\textsuperscript{LMH} genotype in most CER accessions. The second step was the removal of indel_3 in the Si-ALMT9\textsuperscript{LMH} genotype under strong human selection, leading to Si-ALMT9\textsuperscript{LMH} genotype prevalence in CER and BIG (Supplemental Figure 15). These results indicate that Si-ALMT9 might have been a critical factor during the process of domestication and improvement in tomato, thus providing the opportunity to improve flavor and resistance qualities for other varieties. Importantly, our work demonstrates that the process of tomato domestication was associated with fixation and extension of favored alleles or mutations that increased malate accumulation to adapt to the needs of humans.

METHODS

Plant Materials and Growth Conditions

The tomato varieties used in this study were Solanum pimpinellifolium (PIM), S. lycopersicum var. cerasiforme (CER), and S. lycopersicum (BIG). GWAS was performed with 272 tomato accessions, which included 144 BIG, 104 CER, and 23 PIM accessions that were selected from the previously described 360 accessions distributed worldwide (Lin et al., 2014). The experimental population was grown in Wuhan, China, over two locations in 2013: open-field cultivation at Huazhong Agricultural University (HZ location) and greenhouse cultivation at Zhongdou Seed Company (ZD location). The F2 population consisting of 350 individuals resulting from TS-40 x TS-66 was grown in open-field cultivation at Huazhong Agricultural University in 2014. The 73 commercial tomato cultivars (F1 hybrids) were germinated and planted in an experimental field in Yichang, China. The 155 individual cultivated and wild tomato lines used in the phylogenetic analysis are listed in Supplemental Data Set 7. For GWAS, BSA and F1 plants were used, and a minimum of three fruits were harvested at the ripe stage from a minimum of three plants for each line. Ripe stage was determined by external and internal visual inspection of fruit, e.g., extent of pigmentation, seed development, and development of locular jelly. Following tissue selection, the outer pericarp of five fruits was pooled and stored at −80°C for metabolic profiling. For fruit development analysis, flowers were tagged at the full-bloom stage to synchronize developmental stages. The fruits were harvested at 7, 14, 21, 28, and 35 d after flowering (DAF), breaker (BR, 38 DAF), 2 d after breaker (BR+2, 40 DAF), yellow ripe (YR, 41 DAF), and red ripe (RR, 44 DAF) developmental stages. Three biological replicates of each developmental stage were analyzed, and each biological replicate contained 5 to 10 individual fruits of the same developmental stage from the same genotype. Only fruits that appeared developmentally equivalent were used for analysis. After fruit selection, whole fruits, including the seeds, pulp, and skin, were combined in developmental stage-specific pools (5–10 fruits each), frozen in liquid nitrogen, and stored at −80°C for further analysis.

Determination of Malate Content

Malate content of 272 tomato accessions was measured by gas chromatography-mass spectrometry with a modified method as described previously (Roessner-Tunali et al., 2003). Tomato pericarp tissue (~100 mg) was homogenized and extracted in 700 μL of methanol, and 31.5 μL of internal standard (0.2 mg ribitol mL\textsuperscript{−1} water) was subsequently added as a quantification standard. The mixture was extracted for 15 min at room temperature and mixed vigorously with 1 volume of water. To separate polar and nonpolar metabolites, 750 μL of chloroform was then added to the mixtures. After centrifugation at 2200g, the upper methanol/water phase was taken and reduced to dryness under vacuum. Residues after reduction were dissolved in 90 μL of 20 mg mL\textsuperscript{−1} methoxamine hydrochloride in pyridine and then derivatized for 90 min at 37°C followed
by treatment with 90 μL [trithiolsilyl] trifluoroacetamide at 37°C for 30 min. Then, 1-μL samples were injected onto the GC column using a hot needle technique (Zamperini et al., 2000). The gas chromatography-mass spectrometry system used comprised an AS2000 autosampler, a TRACE GC Ultra gas chromatograph, and a Voyager quadrupole mass spectrometer (Thermo Finnigan). The mass spectrometer was tuned according to the manufacturer’s recommendations using Tris-(perfluorobutyl)amine(CF43).

GC was performed on a 30 m HP-5 MS column with 0.25 mm*0.25 μm film thickness (Agilent). The injection temperature was set at 230°C, the interface at 250°C, and the ion source adjusted to 200°C. Helium was used as the carrier gas at a flow rate of 1 mL min⁻¹. The analysis was performed under the following temperature program: 5 min of isothermal heating at 70°C, followed by a 5°C min⁻¹ oven temperature ramp to 300°C, and a final 3 min heating at 300°C. The system was then temperature equilibrated for 1 min at 70°C before injection of the next sample. Mass spectra were recorded at 5 scans s⁻¹ with a mass-to-charge ratio of 50 to 600.

Fruit malate determination of transgenic plants, linkage population, and F1 hybrids was performed by a modified method described previously (Nunes-Nesi et al., 2007). Malate was determined using 30 μL extract or standards (ranging from 0–8 μg), which were pipetted into a microplate containing 105 μL 200 mM Tricine/KOH, pH 9.0, 40 μL 20 mM NAD⁺, 20 μL 1 mM methylthiazolyldi-phenyl-tetrazolium bromide, 4 μL 0.4 mM phenazine ethosulfate, and 1 μL Triton X-100. Absorbance at 570 nm was recorded after 5 min. Then, 2.5 units of malate dehydrogenase was added and the absorbance was recorded when the reaction was complete.

Association Mapping

We used 5.5 M high-quality SNPs (MAF > 0.05, the number of varieties with the minor allele ≥ 6) to perform GWAS for fruit malate content in 272 accessions. The association analyses were performed using the compressed mixed linear model (Yu et al., 2006) with TASSEL 4.0 (Bradbury et al., 2007). The P value of each SNP was calculated and significance was defined at a uniform threshold of ≤1.8 × 10⁻⁷ (P = 1/n; n = total number of markers used) (Wen et al., 2014). Among the unique association signals identified, several candidate genes within 50 kb up- and downstream of the lead SNP were validated, among which the closest genes were considered to be the most likely candidates. The physical locations of the SNPs were identified based on the tomato genomic sequence version SL 2.40 (http://solgenomics.net/).

Candidate Gene Resequencing and Analysis

To further detect the causal variant in the significantly associated region (chromosome 6: 41,300,000–41,400,000), the polymorphism of candidate genes (Soly06g072910 and Soly06g072920) was compared between high-malate and low-malate tomatoes by resequencing 10 high-malate and 7 low-malate fruit accessions (release SL2.40) by PCR, sequencing, and comparison against the reference genome. Primers were designed using Primer Premier 5 (http://www.premierbiosoft.com/primerdesign/index.html) to cover the full length of the candidate genes (Supplemental Data Set 10). Sequencing was performed by the Tianyi Huayuan Bioscience and Technology.

Linkage Mapping by Bulked Segregant Analysis

An F2 population of 350 individuals derived from the cross between TS-40 (a high-malate accession) and TS-66 (a low-malate accession) was planted in the spring of 2015 in the experimental station of Huazhong Agricultural University, China. For each individual, the average malate content of three representative fruits was recorded (Figure 2; Supplemental Data Set 2) and genomic DNA was isolated from fresh leaves using the CTAB method. For bulked segregant analysis, bulk DNA samples for high- and low-malate accessions were constructed by mixing equal amounts of DNA from 40 individuals showing an extremely high- and low-malate content, respectively. Roughly 40× genome sequences for each bulk sample (high-malate fruit and low-malate fruit) were generated. Short reads were aligned against the reference genome (release SL2.40) using the Burrows-Wheeler Aligner (BWA) 47, and SNPs were identified using SAMtools48. SNPs between the two parental genomes were identified for further analysis when the base quality value was ≥20 and the SNP quality value was ≥20. Based on these criteria and the number of SNPs with a read depth from 4 to 200, a SNP index was calculated for both bulk samples expressing the proportion of reads harboring SNPs that were identical to those in the high-malate parent (TS-40). A ΔSNP index was obtained by subtracting the SNP index of the low-malate bulk sample from that of the high-malate bulk sample. An average SNP index for the high-malate and low-malate bulk samples was calculated using a 1000-kb sliding window with a step size of 10 kb (Figure 1). The statistical confidence intervals of the ΔSNP index under the null hypothesis of no QTLs were also calculated. For each position, the 95% confidence intervals of the ΔSNP index were obtained following the method described by Takagi et al. (2013).

Gene Cloning, Construct Generation, and Plant Transformation

The 4.5-kb sequence fused by the 2.7-kb promoter from TS66 and 1.8-kb coding sequence from TS40 of Sl-ALMT9 was amplified and cloned into plant binary vector pMv2 (modified from pHELLSGATE2) to generate construct I (Sl-ALMT9HM promoter:Sl-ALMT9MAH). The 2.7-kb promoter region from TS40 was fused with the coding region from TS66 to generate construct II (Sl-ALMT9MAH promoter:Sl-ALMT9MAH). The 8.9-kb genomic region including the 2.7-kb promoter and coding sequence of Sl-ALMT9 was amplified from TS40 to generate construct IV (Sl-ALMT9MAH promoter:Sl-ALMT9MAH). For the overexpression construct, the 6.3-kb fragment including the genomic DNA region of Sl-ALMT9 was amplified and cloned into the pMv2 vector driven by the cauliflower mosaic virus (CaMV) 35S promoter. Control plants were generated by introducing the empty vector pMv2. For the Sl-ALMT9MAH promoter:GUS construct, a genomic DNA sequence (from –2541 to –1 bp) upstream of the Sl-ALMT9 coding sequence was amplified using sequence-specific primers (Supplemental Data Set 10). The resulting fragments were recombinant upstream of the glucuronidase synthase (GUS) coding sequences in the pPVP vector (modified from pHELLSGATE2). The recombinant constructs were transformed into the Agrobacterium tumefaciens strain C58 by electroporation and subsequently transformed into the tomato genome via explants of cotyledon. The constructs with allelic combinations of promoter and genes were transformed into TS-66 (Figure 2C). The Sl-ALMT9MAH promoter:GUS construct was transformed into tomato Ailsa Craig.

CAPS Markers for Fruit Malate in Tomato

Homozygous and heterozygous Sl-ALMT9MAH/3 tomato plants were genotyped using CAPS markers. PCR was performed to amplify a 784-bp fragment of the functional Sl-ALMT9MAH/3 sequence using the following primers: a forward primer, 5'-TGGATGTAATAAAAACAAGGGAA-3', and a reverse primer, 5'-TATCCATAAAATAGCGAAATAGAAAC-3'. The PCR program was conducted as follows: (1) 3 min at 94°C; (2) 34 cycles of 30 s at 94°C, 30 s at 56°C and 50 s at 72°C; (3) 10 min at 72°C. The PCR products were subsequently digested with HinII (Thermo scientific) for 2 h at 37°C in 10 μL reaction volumes containing 3 μL PCR products, 1 μL 10× buffer G, 0.3 μL HinII (5 units/μL), and 5.7 μL double-distilled water. Digested DNA products were then separated by electrophoresis in 1% agarose gels stained with ethidium bromide and visualized by UV light.
CRISPR/Cas9 Gene Editing

The CRISPR/Cas9 binary vectors (pTX) were derived from pBin19, in which the target sequence was driven by the tomato U6 promoter and Cas9 by 2×3SS. The recombinant pTX vector was designed to produce defined deletions within the coding sequence of SI-ALMT9 using two sgRNAs alongside the Cas9 endonuclease gene (Supplemental Data Set 10 for the two sgRNAs used in this study). For genotyping of each first-generation (T0) transgenic line, mixed samples of three different leaves were collected to capture all possible induced mutant alleles due to sectoring (chimerism). Positive detection of each plant was conducted by PCR for the presence of Cas9. The CRISPR/Cas9 T-DNA-positive lines were further genotyped for indel mutations using a forward primer to the left of sgRNA1 and a reverse primer to the right of sgRNA2 (Supplemental Data Set 10).

RNA Extraction and RNA-Seq

The fruits from CK and OX1 harvested at the ripening stage were frozen in liquid nitrogen and kept at −80°C until use. Total RNA extractions were performed essentially as described by Ye et al. (2015). Total RNAs were then sent to the Annoroad Company, where the libraries were produced and sequenced using single-ended sequencing of Illumina HiSeq 2500. The sequencing data can be accessed at the website: http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE89772. Raw sequences were filtered to remove the 3' adaptor sequence, low-quality reads (reads containing sequencing Ns > 5) and short reads (<16 nucleotides), and the resulting sets of clean reads were used for the subsequent analyses, as described previously (Patel and Jain, 2012). All clean reads were mapped to contig assemblies using the TopHat mapping algorithm with version 2.0.4 (http://tophat.cbcb.umd.edu/) allowing no more than 2-nucleotide mismatches. Clean reads were mapped to the genome sequences (SL2.40 version) of S. lycopersicum downloaded from the SOL Genomics Network database (http://solgenomics.net/organism/Solanum_lycopersicum/ genome). The multiple aligned reads were then filtered by TopHat software and the remaining clean reads were designated as unambiguous clean reads. The number of unambiguous clean reads for each gene was calculated and then normalized to reads per kilobase of gene per million reads, a standard unit to calculate UniGene expression (Mortazavi et al., 2008). The software edgeR was used to perform differential expression analysis (Robinson et al., 2010). edgeR can be used to analyze the difference in expression between two or more samples and indices of fold change (log2 ratio) and P value (false discovery rate) provide an indication of whether a gene is differentially expressed. Here, genes with a P value <0.01 and a log2 ratio >1.0 or <0.5 were considered differentially expressed. The differentially expressed tomato genes extracted from ITAG2.4_proteins.fast (ftp://ftp.solgenomics.net/organisms/Solanum_lycopersicum/annotation/ITAG2.4_release/) were used as query to identify Arabidopsis thaliana homologs (TAIR9 version) using an e-value of 1×10−5 by BLASTP (http://blast.ncbi.nlm.nih.gov/Blast.cgi).

Gene Expression Analysis

The expression abundance of Si-ALMT9 was also investigated by RT-qPCR. The sequences of the primer pair (designed using Primer Premier 3.0, http://frodo.wi.mit.edu/primer3) are listed in Supplemental Data Set 10. The cDNA synthesis and RT-qPCR steps were performed as previously described (Liu et al., 2012).

GUS Staining

Slices of roots, stems, and leaves from the transgenic lines transformed with SI-ALMT9 were stained with a GUS staining solution (100 mM sodium phosphate buffer) to evaluate GUS activity. Staining was allowed to proceed for 5 h at 37°C in darkness, and then samples were washed with a graded ethanol series at room temperature for decolorization and observed by light microscopy (Olympus SZX12).

Protein Subcellular Localization

To investigate the subcellular localizations of the SI-ALMT9 proteins, the coding sequence of SI-ALMT9 without the stop codon was amplified by PCR and then cloned into the expression vector pCAMBIA1302:GFP under the control of the CaMV 35S promoter by homologous recombination. The fusion construct combined with CBL6-OFP (Calcineurin B-like protein) as a marker of the tonoplast and the control GFP vector were individually transformed into leaves of Nicotiana benthamiana as described previously (Batistic et al., 2010). GFP and OFP fluorescence were detected and analyzed at 48 h following transfection using Leica Confocal software.

AI Tolerance and Malate Efflux

Plants in hydroponic culture were assessed for AI tolerance by supplementing the hydronic nutrient solution with a range of AlCl3 concentrations. Specifically, tomato seeds (TS66, OX1, OX2, OX5, and TS40) were sown on wet filter paper in a Petri dish and germinated in darkness for 3 d. Subsequently, seedlings were transferred to a floating cystosepiment in a plastic box containing 2 liters of modified one-fifth nutrient solution (Urbanczyk-Wochniak and Fernie, 2005). The Ca(NO3)2 concentration was set to 0.8 mM instead of 1.25 mM (Wang et al., 2001). Following 1 week of growth under 70 to 80% relative humidity and a 12/12-h light/dark (white light with an irradiation intensity of 900 μmol m−2 s−1) regime at 25°C, the seedlings were transferred to hydronic boxes containing 40 liters of full nutrient solution at pH 6.5, which was refreshed every 3 d. Following 1 week of hydroponic growth, AlCl3 at different concentrations (0, 10, and 20 μM) was added to the nutrient solution at pH 4.2. Primary root length was recorded to quantify the effect of Al toxicity on root growth. Relative root growth was calculated as previously described (Delhaize et al., 2004). Each experimental treatment of the different lines was performed in three replicates, each of which contained three plants. Malate efflux from root segments was assayed according to previously described methods (Delhaize et al., 2004) with some modifications. Four 2-cm lengths of root apices were incubated with constant agitation in 1 mL of 1 mM CaCl2, pH 4.2, for 5 h. The apices were then rinsed with the CaCl2 solution three times and subsequently treated with 1 mL of either CaCl2 (control) or treatment solution (1 mM CaCl2 with 100 μM AlCl3, pH 4.2). Following a 3-h incubation, the treatment solution was collected and dried, and the resulting residues were collected and analyzed for malate contents as described above. To characterize the role of SI-ALMT9 in Al detoxification, tomato hairy roots were treated for 3 h with 1 mM CaCl2, pH 4.2, containing 100 μM Al, after which they were then stained with hematoxylin as described previously (Liang et al., 2013).

Yeast One-Hybrid Assay

The yeast one-hybrid assay was performed as described in the Matchmaker One-Hybrid Library Construction and Screening Kit (Clontech). The full-length and SI-ALMT9 open reading frame (ORF) sequences were amplified from tomato cDNA by PCR and cloned into the pGADT7 vector (Clontech). The fragment of the SI-ALMT9 promoter was amplified from TS66 and TS40 genomic DNA and cloned into the pAbAi vector (Clontech). The pAbAi bait vectors were introduced into the GOLD1 yeast and cultured on SD/-Ura. The pGADT7 prey vector was introduced into yeast strains containing pAbAi bait vectors and cultured on SD/-Leu. After 4 d incubation, the positive yeast strains were picked and diluted in double-distilled water to an OD600 of 0.1, and 2 μL of suspension was spotted on SD/-Leu, with or without ABA (0–20 ng/mL) (Sigma-Aldrich), followed by 3 to 7 d incubation at 30°C.
Analysis of Promoter-cis-Element Interaction via Transient Expression in N. benthamiana Leaves

The full-length Sl-WRKY42 ORF was cloned into the effector vector pGreen II 62-SK under the control of CaMV 3SS promoter (Hellens et al., 2005). Sl-ALMT9 promoter fragments with serial deletions from the 5’ end were PCR amplified using specific primers (Supplemental Data Set 10) and cloned into the reporter vector pGreen II 0800-LUC. Individual combinations of effector and reporter vectors were cotransformed into Agrobacterium GV3101 cells alongside the pSoup vector, and the transformed Agrobacterium strains were used to infiltrate young N. benthamiana leaves, in which transient expression was analyzed following a 2-d incubation. Firefly and Renilla luciferase signals were assayed with the dual luciferase assay reagents (Promega) using an Infinite M200 (Tecan).

Molecular Diversity and Phylogenetic Analysis

A contig of 8000 bp comprising the promoter and coding region of Sl-ALMT9 (corresponding to 41,336,000–41,344,000 in Tomato WGS chromosome SL2.40 ch06) was obtained from the 155 accessions (Supplemental Data Sets 7 and 8) and used for the phylogenetic analysis. All sequences were aligned using Geneious4.8.5 with default multiple alignment parameters. The alignment file was imported into MEGA5.2.2 (Tamura et al., 2011) and converted into mega (.meg) file format. The phylogeny was reconstructed using the neighbor-joining method with 1000 bootstrap replications and the maximum composite likelihood model.

For the molecular diversity analysis, two methods, i.e., the \( r^2 \) ratios and the XP-CLR (Chen et al., 2010), were used to identify the selective sweeps in Sl-ALMT9 associated with tomato domestication and improvement events. Briefly, \( r^2 \) ratios (\( \overline{r}^2 \), PIM/\( \overline{r}^2 \), CER; \( \overline{r}^2 \), CER/\( \overline{r}^2 \), BIG) were calculated with a sliding window length of 100 bp and step size of 25 bp. XP-CLR was employed with the following parameters: a window size of 0.005 Mb, no more than 50 SNPs within a window, 200 bp between two grid points, and a correlated level of 0.95 between two SNPs.

Accession Numbers

Sequence data from this article can be found in the Sol Genomics Network or GenBank/EMBL databases under the following accession numbers: Sl-ALMT9 (Soly06g072910 and Soly06g072920; KY944467); Sl-WRKY42 (Soly10g009550; XM_004248403); mitochondrial MDH (AY725474); fumarase (SGN-US70526); Zm-ALMT1, ABC6748.2; Zm-ALMT2, NP_001132468.1; Hv-ALMT1, ACJ15441.1; Ta-ALMT1, AB081803; At-ALMT1, NP_172319; Br-ALMT1, AB194300; Bn-ALMT2, AB194301; At-ALMT6, NP_001325206; At-ALMT9, OAP01562; At-ALMT12, NP_193531; Ma1, XP_008361731.1; and Ma2, XP_008339686.2.

Supplemental Data

Supplemental Figure 1. Spectra of fruit malate content for three tomato variety groups.

Supplemental Figure 2. Malate contents in tomato accessions TS-40 and TS-66, and their F1 progeny.

Supplemental Figure 3. Characterization of Sl-ALMT9 gene structure.

Supplemental Figure 4. Comparison of amino acid sequences between Sl-ALMT9^{HMH} (TS40) and Sl-ALMT9^{LMH} (TS66).

Supplemental Figure 5. Phylogenetic and protein similarity analysis of Sl-ALMT9 orthologs in plants.

Supplemental Figure 6. Association of polymorphism sites in Sl-ALMT9 with malate contents.

Supplemental Figure 7. Comparison of Sl-ALMT9 promoter sequences between HMM pool and LMM pool in BSA analysis.

Supplemental Figure 8. Analysis of indel, 3-based CPAS marker in 73 commercial tomato cultivars (F1 hybrids).

Supplemental Figure 9. The expression levels of the two core genes involved in malate biosynthesis.

Supplemental Figure 10. The malate content and relative Sl-ALMT9 transcript levels in ripe fruits from different tomato species.

Supplemental Figure 11. The malate content in two Sl-ALMT9 overexpression lines (OX1, OX3) compared with that in wild-type TS66.

Supplemental Figure 12. Phylogenetic tree of Sl-ALMT9 in 155 tomato accessions.

Supplemental Figure 13. Alignment of genomic DNA sequences, cDNA sequences, and amino acid sequences of ALMT9 in Solanum lycopersicum cv M82, TS-87, and S. pennelli.

Supplemental Figure 14. Proposed model for the role of TFM6 in regulating fruit malate content and AI resistance in roots.

Supplemental Figure 15. Structural variations of TFM6 between PIM accessions, low-malate accessions, and high-malate accessions.

Supplemental Table 1. List of 83 SNPs significantly associated with the malate accumulation in tomato fruit.

Supplemental Table 2. Predicted genes in a region extending the TFM6 locus within 50 kb from both sides.

Supplemental Table 3. Malate content of ripe fruits in 73 genotyped commercial tomato cultivars (F1 hybrids).

Supplemental Table 4. Multiple metabolites analysis in the overexpression plant of Sl-ALMT9.

Supplemental Data Set 1. Relative content (log2-transformed) of malate in ripening fruit of 272 resequenced tomato accessions.


Supplemental Data Set 3. List of SNPs involved in the significantly linkage region on chromosome 6 by BSA.

Supplemental Data Set 4. Amino acid sequences of 14 Sl-ALMT9 orthologs in plants referred to in Supplemental Figure 5B.

Supplemental Data Set 5. Natural Sl-ALMT9 sequence variation in tomato accessions.

Supplemental Data Set 6. Significantly differentially expressed genes (RPKM >5 at least in OX1 or CK) in red ripe fruits of Sl-ALMT9 overexpression line (OX1) compared with wild-type CK.

Supplemental Data Set 7. List of tomato accessions used in phylogenetic analysis.

Supplemental Data Set 8. Nucleotide sequences of Sl-ALMT9 in 155 accessions refer to Supplemental Figure 12 and Supplemental Data Set 7.

Supplemental Data Set 9. Accessions used to determine origin of ALMT9^{indel,3} mutation.

Supplemental Data Set 10. Primers used in this study.

ACKNOWLEDGMENTS

We appreciate the helpful comments on the manuscript made by J. Yan (Huazhong Agricultural University, China). This work was supported by the National Natural Science Foundation of China (Grant 31230064).
AUTHOR CONTRIBUTIONS

Z.Y., Y.Z., and J.Y. conceived and designed the research. Z.Y., Y.Z., and H.L. supervised the study. J.Y., T.H., F.Z., B.W., and T.Y. performed experiments. J.Y., X.W., C.L., and Y.L. analyzed data. Z.Y., Y.Z., and H.L. collected tomato accessions. J.Y., F.Z., and B.W. generated the F2 population and transgenic plants. J.Y., Y.Z., and Z.Y. wrote the manuscript. J.J.G. revised the manuscript. All of the authors discussed the results and commented on the manuscript.

Received March 20, 2017; revised July 5, 2017; accepted August 11, 2017; published August 16, 2017.

REFERENCES


An InDel in the Promoter of Al-ACTIVATED MALATE TRANSPORTER9 Selected during Tomato Domestication Determines Fruit Malate Contents and Aluminum Tolerance
Jie Ye, Xin Wang, Tixu Hu, Fengxia Zhang, Bing Wang, Changxin Li, Tianxia Yang, Hanxia Li, Yongen Lu, James J. Giovannoni, Yuyang Zhang and Zhibiao Ye

Plant Cell 2017;29:2249-2268; originally published online August 16, 2017;
DOI 10.1105/tpc.17.00211

This information is current as of November 6, 2017