doi:10.1093/plphys/kiab230 PLANT PHYSIOLOGY 2021: 186: 2078–2092

Plant Physiology®

Research Article

Genome-wide association study reveals the genetic architecture of 27 agronomic traits in tomato

Jie Ye,[1,2,#] Xin Wang,[2,#] Wenqian Wang,[1,3,#] Huiyang Yu,[1] Guo Ai,[1] Changxing Li,[1] Pengya Sun,[1] Xianyu Wang,[4] Hanxia Li,[1] Bo Ouyang,[1] Junhong Zhang,[1] Yuyang Zhang,[1] Heyou Han,[3] James J. Giovannoni,[2,5] Zhangjun Fei[2,5] and Zhibiao Ye[1,*,†]

1 Key Laboratory of Horticultural Plant Biology, Ministry of Education, Huazhong Agricultural University, Wuhan 430070, China
2 Boyce Thompson Institute for Plant Research, Cornell University, Ithaca, New York 14853, USA
3 College of Food Science and Technology, Huazhong Agricultural University, Wuhan 430070, China
4 College of Agriculture, Guangxi University, Nanning 530004, China
5 U.S. Department of Agriculture-Agricultural Research Service, Robert W. Holley Center for Agriculture and Health, Ithaca, New York 14853, USA

*Author for communication: zbye@mail.hzau.edu.cn
†Senior author.
#These authors contributed equally to the article.

Z.Y., Z.F., and J.Y. designed the experiments and managed the project. H.L., J.J.G., X.W., Y.Z., B.O., J.Z., and H.H. contributed to the original concept of the project. J.Y., W.W., Y. W., and P.S. collected samples and performed phenotyping. J.Y., X.W., H.Y., G.A., and C.L. performed the data analyses. J.Y. wrote the manuscript. Z.Y., Z.F., and X.W. revised the manuscript.

The author responsible for distribution of materials integral to the findings presented in this article in accordance with the policy described in the Instructions for Authors (https://academic.oup.com/plphys/pages/general-instructions) is: Zhibiao Ye (zbye@mail.hzau.edu.cn).

Abstract

Tomato (*Solanum lycopersicum*) is a highly valuable fruit crop, and yield is one of the most important agronomic traits. However, the genetic architecture underlying tomato yield-related traits has not been fully addressed. Based on ~4.4 million single nucleotide polymorphisms obtained from 605 diverse accessions, we performed a comprehensive genome-wide association study for 27 agronomic traits in tomato. A total of 239 significant associations corresponding to 129 loci, harboring many previously reported and additional genes related to vegetative and reproductive development, were identified, and these loci explained an average of ~8.8% of the phenotypic variance. A total of 51 loci associated with 25 traits have been under selection during tomato domestication and improvement. Furthermore, a candidate gene, *SI-ACTIVATED MALATE TRANSPORTER15*, that encodes an aluminum-activated malate transporter was functionally characterized and shown to act as a pivotal regulator of leaf stomata formation, thereby affecting photosynthesis and drought resistance. This study provides valuable information for tomato genetic research and breeding.

Introduction

Tomato (*Solanum lycopersicum*) is considered the leading fruit crop, with a global production of 182.3 million tons in 2018 (FAOSTAT; http://www.fao.org/faostat), and has served as a model system for fleshy fruit biology. Tomatoes originated in South America (Wang et al., 2020), and there is a long history of "wild" tomato distribution in the mountainous area of southwest Guangxi (GX) of China. The natural barrier formed by the steep terrain ensures the independent evolution of these tomatoes, which contain valuable genetic resources of fruit quality and biotic/abiotic stress tolerance (Shu et al., 1995).

Received January 19, 2021. Accepted May 3, 2021. Advance access publication June 17, 2021

Traits directly related to yield, such as fruit size, have been well studied (Frary et al., 2000; Nesbitt and Tanksley, 2001; Chakrabarti et al., 2013; Li et al., 2018). Other yield-related traits, such as inflorescence location, inflorescence architecture, flower development, and leaf stoma density, are also vital to tomato fruit production. Most tomato varieties exhibit a sympodial growth habit instead of monopodial branching (Pnueli et al., 1998), and domesticated tomatoes generate zigzag inflorescences with various flower numbers and inflorescence branches (MacAlister et al., 2012). Taken together, flowering pattern, branching style, and inflorescence polymorphism contribute to total production in tomato.

Because of their importance, a number of genes controlling yield-related traits have been identified in tomato. Tomato *SELF PRUNING* (*SP*) gene, which is homologous to *TERMINAL FLOWER 1* and *FLOWERING LOCUS T* of Arabidopsis (*Arabidopsis thaliana*), and *CENTRORADIALIS* of *Antirrhinum majus*, prevent flowering in the sympodial shoots and reduces the number of leaves only in the sympodial segments of tomato (Pnueli et al., 1998). *LOCULE NUMBER* (*LC*), a *WUSCHEL* gene also named *lcn2.1*, has a major effect on the LC of tomato fruit by controlling stem cell fate in the apical meristem (Munos et al., 2011). *FASCIATED* (*FAS*), encoding a secreted peptide (CLV3) modified with sugars, is considered a major gene and has the strongest effect in increasing the number of locules (from two to more than six) in tomato (Xu et al., 2015). Moreover, *ANANTHA* and *S* (*COMPOUND INFLORESCENCE*) regulate inflorescence branching (Lippman et al., 2008), *SP5G* promotes day-neutrality and early yield in tomato (Soyk et al., 2017a, 2017b), *SINGLE FLOWER TRUSS* (*SFT*) promotes flowering and attenuates apical meristem growth through systemic *SFT* signals (Lifschitz et al., 2006), *TERMINATING FLOWER* (*TMF*) has a key role in determining simple versus complex inflorescences (MacAlister et al., 2012), *FALSIFLORA* causes highly branched inflorescences (Molinero-Rosales et al., 1999), *Style 2.1* controls the style length in cultivated tomatoes (Chen et al., 2007), *BLADE-ON-PETIOLE* (*SlBOP1/2/3*) promotes inflorescence complexity by interacting with *TMF* (Xu et al., 2016a, 2016b), and *Jointless2* (*j2*) and *Enhancer-of-j2* are two homologs of the Arabidopsis floral organ identity MADS-box gene *SEPALLATA4* and regulate the formation of the flower abscission zone (Soyk et al., 2017a, 2017b).

In addition to the reproductive traits described above related to inflorescence and fruit development, vegetative traits also contribute to crop yield. A goal of modern agriculture is to improve plant drought tolerance and production per amount of water used, referred to as water use efficiency (Yoo et al., 2010). Stomata, epidermal valves that modulate CO_2 and water vapor exchange between plants and the atmosphere, play critical roles in primary productivity and in plant adaptation to the global climate. Positively acting transcription factors and negatively acting mitogen-activated protein kinase signaling control stomatal development in Arabidopsis (Serna and Fenoll, 2000; Lampard et al., 2008; Sugano et al., 2010; Pillitteri and Torii, 2012); only a few regulatory genes of stomata formation have been functionally identified in tomato by reverse genetics (Morales-Navarro et al., 2018; Ortega et al., 2019).

Genome-wide association study (GWAS) is an effective approach to investigate the genetic architecture of complex agronomic traits (Huang et al., 2010; Li et al., 2014; Cao et al., 2016; Varshney et al., 2017, 2019). Recently, GWAS has been employed to explore genetic loci associated with 15 agronomic traits in 163 tomato accessions (Mata-Nicolas et al., 2020). In this study, an improved tomato haplotype map was constructed by including sequencing data of 66 GX tomatoes, and using this haplotype map GWAS for 27 agronomic traits was performed to identify loci potentially associated with tomato production. Several genomic loci underlying these agronomic traits are consistent with previous reports and many additional loci are identified in this study. A candidate gene underlying the GWAS signal of leaf stomatal density was further functionally verified.

Results

Sequencing, variants, and population structure of GX tomatoes

In this study, a total of 605 tomato accessions were used for genotyping and subsequent GWAS analysis. Among these accessions, a diverse global collection of 539 accessions were genotyped in previous studies (Lin et al., 2014; Tieman et al., 2017; Zhu et al., 2018). The remaining 66 newly sequenced accessions were collected from the mountainous areas of GX Province, China, and these accessions had small leaves and small red fruits with thin skins (Supplemental Table S1; Supplemental Figure S1). A total of 6.25 billion 100-bp paired-end reads were obtained for these GX accessions, representing a mean depth of 9.85× coverage of the tomato genome (Supplemental Table S1). A total of 4,412,112 million high-quality single nucleotide polymorphisms (SNPs) were called from the sequencing data of the 605 accessions.

To determine the population structure of the GX tomatoes, we performed clustering analysis of the 605 tomato accessions using the genome-wide SNPs. Largely consistent with the previously reported results (Lin et al., 2014), these tomatoes were divided into six groups (Figure 1A), which was further supported by population structure analyses (Figure 1, B and C). Interestingly, GX tomato accessions clustered together and divided the *S. lycopersicum* var. *cerasiforme* (SLC) group into two subgroups: South American SLC and Non-South American SLC, suggesting that GX tomatoes fall within the SLC lineage, and they could be derived from South American SLC.

The nucleotide diversity decreased from the *S. pimpinellifolium* (SPIM) group ($\pi = 2.61 \times 10^{-3}$) to the SLC group ($\pi = 1.3 \times 10^{-3}$) and to the *S. lycopersicum* var. *lycopersicum* (SLL) group ($\pi = 0.57 \times 10^{-3}$) and GX group ($\pi = 0.5 \times 10^{-3}$), indicating that a large amount of genetic diversity has been lost in GX accessions, possibly because of geographical isolation and narrow ancestral genetic

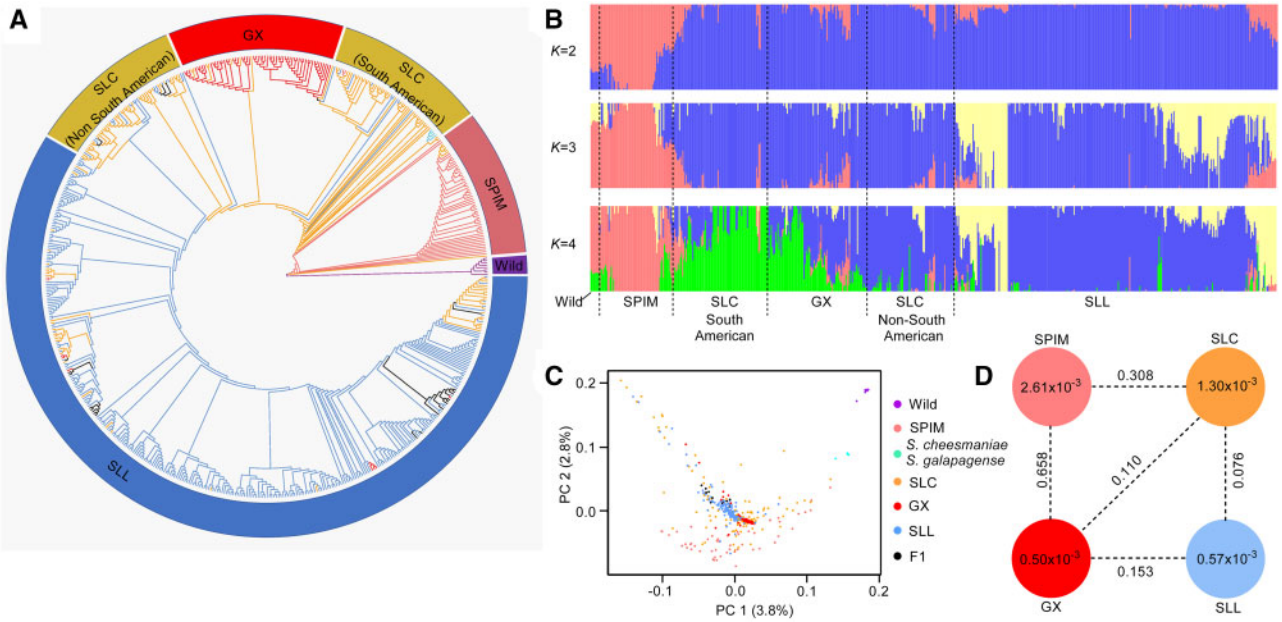

Figure 1 Population diversity of the tomato accessions. A, Neighbor-joining clustering analysis of 605 tomato accessions. The outer ring indicates different groups. Colors of branches on the tree indicate different groups: SPIM (pink), SLC (orange), GX (red), SLL (blue), and other wild (purple). B, Population structure of tomato accessions with different numbers of clusters ($K = 2$, 3, and 4). The orders and positions of the accessions on the *x* axis are consistent with those in the neighbor-joining tree. C, PCA plot of the tomato accessions. The dot color scheme is same as in a. PC1, first principal component; PC2, second principal component. D, Nucleotide diversity (π) and population divergence (F_{ST}) among the four groups. For each group, π is shown inside the circle. F_{ST} between the two groups is shown on the dotted line.

background. Large population divergences between the GX group and the other three groups were observed (Figure 1D), suggesting that GX tomatoes in China might have accumulated some unique genetic variants by selection and/or genetic drift after their introduction.

Phenotypic variation in the tomato population

A total of 27 agronomic traits, including fruit number on the second truss (FRNS), fruit number on the third truss (FRNT), flower number on the second inflorescence (FLNS), flower number on the third inflorescence (FLNT), sepal number (SN), petal number (PN), first inflorescence node (FIN), ovary transverse diameter (OTD), ovary longitudinal diameter (OLD), ratio of OTD to OLD (OTLD), stigma exsertion (SE), stigma shape (SS), ratio of sepal length (SL) to petal length (PL; SPR), stomatal density (SD), first to second inflorescence node (FSIN), internode length (IL), PL, stamen length (STAL), stigma length (STIL), SL, indeterminate or determinate meristem (IDM), ratio of STAL to (STIL + OLD) (SSR), fruit stalk diameter (FSD), fruit stalk length (FSL), inflorescence type (IT), and FAS flower (FL), were investigated during the whole growth period of tomato with three replications (Supplemental Figure S2; Supplemental Note). These traits were classified into three categories: four organ location traits, nine organ number traits and 14 organ size traits (Supplemental Tables S2 and S3). For the majority of the agronomic traits, abundant variation was detected, with the coefficients of variation ranging from 0.12 for PN to 2.92 for FLNS (Supplemental Table S2). All traits determined in the diverse global collection of tomato accessions (Supplemental Table S3) displayed a broad-sense heritability ($H2$) greater than 0.7, and 14 had heritability >0.9 (Supplemental Table S2), suggesting these traits were primarily determined by genotype. Most of the traits showed a normal distribution, while PN, SN, FSIN, FSD, FSL, and OTD showed a skewed distribution (Supplemental Figure S3). The phenotypic values of all agronomic traits, except IL, were significantly different among the three subgroups of tomato (SPIM, SLC, and SLL;Supplemental Figure S4). For example, ovary size (OLD and OTD) increased during tomato domestication and improvement, while flower number (FLNS and FLNT), and fruit number (FRNS and FRNT) declined (Supplemental Figure S4). Many of the 27 traits were correlated (Supplemental Figure S5), and the correlation coefficients between some traits were very high, such as SL and PL ($r = 0.85$), IL and FSIN ($r = 0.88$), and OTD and SN ($r = 0.83$). Negative correlation between ovary size (OLD and OTD) and flower and fruit number (FLNS, FLNT, FRNS, and FRNT) suggested that fruit size was positively selected (Lin et al., 2014), while the number of fruits decreased during the process of tomato breeding (Supplemental Figures S4 and S5).

GWAS of 27 agronomic traits

To reveal the genetic architecture of the vegetative and reproductive organ development in tomato, GWAS was performed on the 27 agronomic traits in the 605 tomato accessions using the genome-wide SNPs. The Manhattan plots of GWAS for all 27 traits are shown in Supplemental Figures S6–S32, and detailed information

about all significant associations is summarized in Supplemental Table S4.

In this study, an association locus has been defined as a chromosomal region in which the distance between the adjacent pairs of associated SNPs is <200 kb (Ye et al., 2019). According to this definition, a total of 239 suggestive associations (including 148 significant associations) corresponding to 129 loci were identified (Figure 2; Supplemental Table S4). For organ location, organ number, and organ size traits, 28, 85, and 126 associated loci were identified, respectively (Table 1). On average, each trait had around nine identified associated loci. Four potential GWAS hotspots (density > 0.03) were identified, which aligned perfectly with known QTLs involved in the regulation of the growth and development of tomatoes on chromosomes 2 (*LC*), 3 (*FA*), 6 (*SP*), and 11 (*FAS*; Supplemental Figure S33). For example, the locus overlapping with *LC* was simultaneously detected for 11 different traits including FLNS, FRNS, FLNT, FRNT, PN, SN, FSD, OTD, OTLD, SS, and SSR. The locus overlapping with *FAS* was detected for nine different traits including FSIN, IL, IDM, IT, PN, SN, FL, OTD, and SS. *LC* and *FAS* are the main loci regulating the size and shape of tomato fruit (Barrero et al., 2006; Conget al., 2008; Munos et al., 2011).

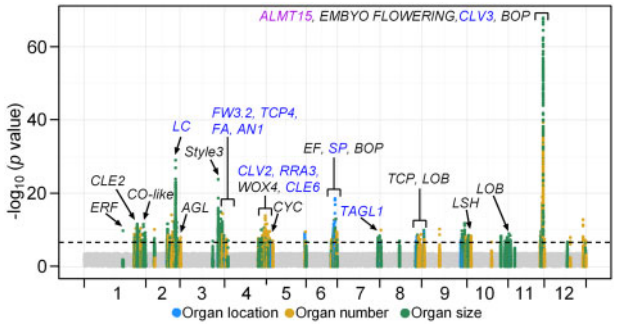

Figure 2 Manhattan plot of GWAS with identified genetic associations. The significance of the associations with agronomic traits is indicated as the negative logarithm of the *P*-values. All trait-SNP associations with *P*-values $<2.4 \times 10^{-7}$ are plotted against the genome at 1-Mb intervals. Black horizontal dashed line indicates the genome-wide significance threshold (*P*-values $<2.4 \times 10^{-7}$). Associations with organ location traits are indicated by light blue circles, organ number traits by yellow circles, and organ size traits by green circles. Candidate genes including those with known functions (blue text) and unknown functions (black text) are indicated. Function of *ALMT15* in the purple text was validated in this study.

These results are consistent with that many of the agronomic traits were highly correlated (Supplemental Figure S5). The percentage of phenotypic variation explained by each locus ranged from 1.6% to 69.4% in organ location traits, from 1.2% to 78.5% in organ number traits, and from 1.4% to 75.3% in organ size traits, with mean values of 9.8%, 8.0%, and 8.6%, respectively (Table 1). Although some traits were controlled by one major locus that explained over 60% of the natural variation, such as SN, IDM, IT, and FL, most traits were determined by multiple moderate-effect loci.

Key candidate genes involved in vegetative and reproductive development

We searched for candidate genes responsible for the variation in tomato agronomic traits based on the information of gene annotation, phylogenetic analysis of candidate genes with their homologs with known functions, and cross-referencing with results from previous linkage mapping. We were able to identify several plausible candidate genes and possible causative SNPs underlying the agronomic traits (Supplemental Table S5). Taking the 18 associated loci of PN as an example, in addition to previously reported large-effect genes including *CLV3*, *LC*, *SlWOX4*, and *SlRRA3* (Ji et al., 2010; Xu et al., 2015), several new minor-effect candidate genes were also identified such as *SlCLE2* and *SlCLE6* belonging to the CLV3/EMBRYO-SURROUNDING REGION family that play an important role in regulating stem cell proliferation and differentiation of plant development (Zhang et al., 2014; Supplemental Figure S34). Two major haplotypes at each of the two lead SNPs (SL2.50ch01_89460876 and SL2.50ch05_1427678 in *SlCLE2* and *SlCLE6*, respectively) were significantly associated with different PNs in tomato (Supplemental Figure S34, B and C). The loci identified here provide valuable candidates for future studies that can improve our understanding of the genetic regulation of these traits.

For FSIN, IL, PL, STAL, and IDM, the association signal at the end of chromosome 6 showed a subtle zigzag pattern, suggesting multiple trait-associated genes present in this small region (Figure 3A). Three significantly associated SNPs (SL2.50ch06_43765964, SL2.50ch06_44230173 and SL2.50ch06_45972263) corresponding to genes *Solyc06g071140*, *Solyc06g071830*, and *Solyc06g074350*, respectively, were detected in this region (Figure 3B). *Solyc06g074350* corresponds to the *SP* gene (Pnueli et al., 1998). *Solyc06g071140*

Table 1 Summary of significant locus-trait associations identified in GWAS

Item	Organ Location Traits	Organ Number Traits	Organ Size Traits
Number of Traits	4	9	14
Number of Suggestive SNPs (Significant SNPs)[a]	28 (18)	85 (49)	126 (81)
Number of Loci Per Trait	4–11	2–18	1–20
Lead SNPs Explaining >15% of Variation[b]	7	10	14
Maximum Explained Variation (%)	69.4	78.5	75.3
Explained Variation per SNP (%)	9.8	8.0	7.9

aSuggestive SNPs, $P \leq 2.4 \times 10^{-7}$; significant SNPs, $P \leq 1.2 \times 10^{-8}$.
bLead SNPs are those with the lowest *P*-values in the defined association loci.

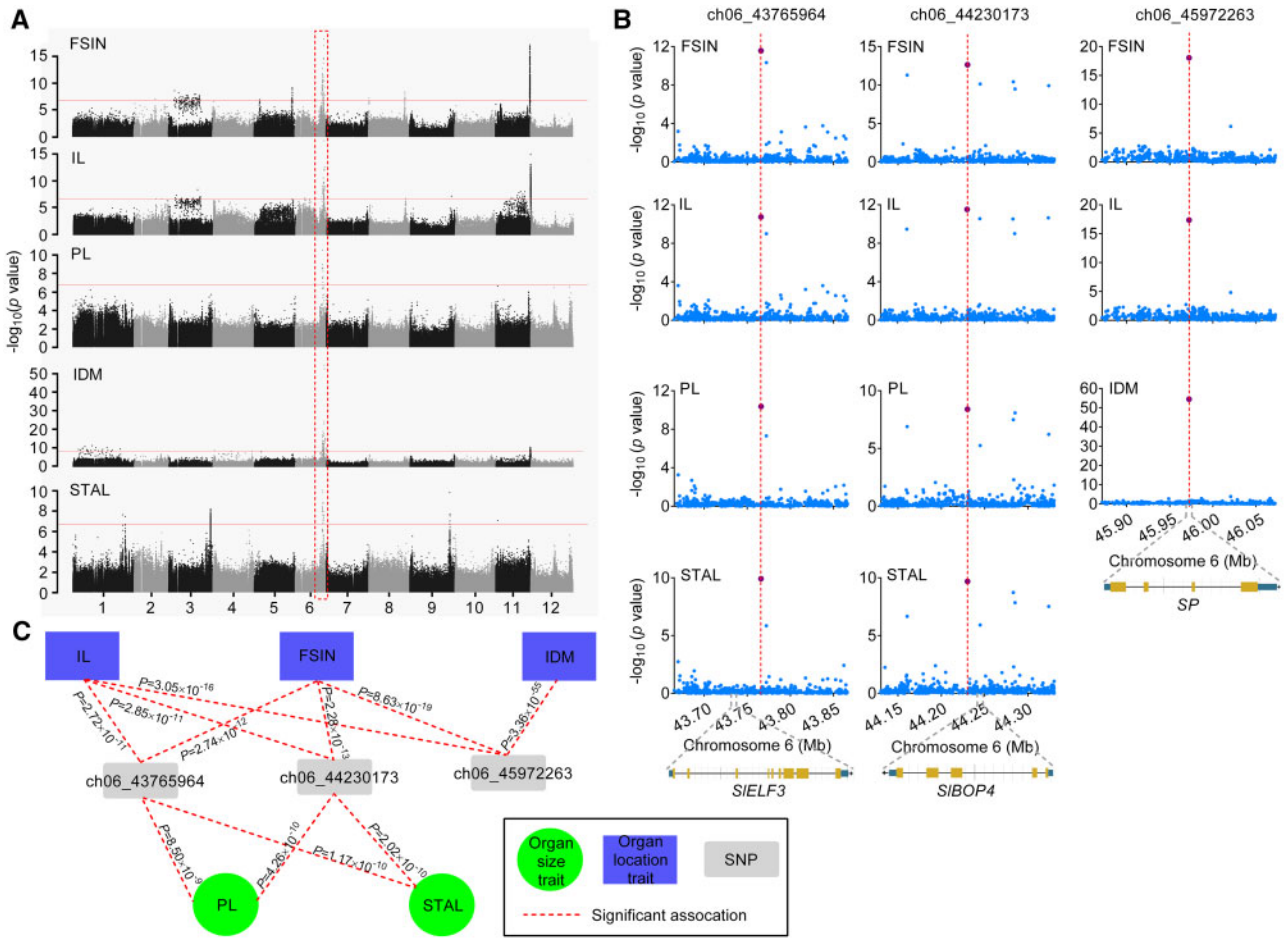

Figure 3 Genome-wide associations for FSIN, IL, PL, IDM, and STAL. A, Manhattan plots of GWAS for FSIN, IL, PL, IDM, and STAL. The common association signal is highlighted with a dashed red box. B, Regional (200-kb) association plots of GWAS for FSIN, IL, PL, STAL, and IDM. Three common significant association SNPs (SL2.50ch06_43765964, SL2.50ch06_44230173 and SL2.50ch06_45972263) are highlighted with dashed lines. Corresponding candidate genes *SlELF3* (*Solyc06g071140*), *SlBOP4* (*Solyc06g071830*), and *SP* (*Solyc06g074350*) are indicated. C, Association network summarizing GWAS results shown in (B). GWAS results of certain trait containing any of the three significant SNPs are treated as nodes and are connected. Green circle indicates organ size traits (PL and STAL) and purple rectangle indicates organ location traits (IL, FSIN, and IDM).

(*SlELF3*), located 28.7 kb downstream of SL2.50ch06_43765964, was annotated as a homolog of the *EARLY FLOWERING* gene that regulates circadian clock function and flowering in Arabidopsis (Hicks et al., 2001; Herrero et al., 2012), rice (Zhao et al., 2012), and pea (Rubenach et al., 2017; Supplemental Figure S35A). *Solyc06g071830* (*SlBOP4*), located 15.3-kb downstream of SL2.50ch06_44230173, encoded a BTB/POZ protein that is paralogous to three other SlBOPs (SlBOP1, SlBOP2, and SlBOP3) known to control inflorescence architecture and flower production in tomato (Xu et al., 2016a, 2016b; Supplemental Figure S35B), suggesting that *Solyc06g071830* is likely the candidate gene underlying this locus. To compare association results across different traits, we constructed an association network to visualize the relationships of complex traits (Figure 3C). The SNP SL2.50ch06_45972263 near the *SP* gene was the most significantly associated SNP for the IDM trait ($P = 3.36 \times 10^{-55}$) and was repeatedly detected for IL ($P = 3.05 \times 10^{-16}$) and FSIN ($P = 8.63 \times 10^{-19}$). The SNP SL2.50ch06_44230173 near *Solyc06g071830* was the second most significant SNP of FSIN ($P = 2.28 \times 10^{-13}$) and was repeatedly detected for IL ($P = 2.85 \times 10^{-11}$), STAL ($P = 2.02 \times 10^{-10}$), and PL ($P = 4.26 \times 10^{-10}$). The SNP SL2.50ch06_43765964 near *Solyc06g071140* was the third most significant SNP for FSIN ($P = 2.74 \times 10^{-12}$) and was repeatedly detected for IL ($P = 2.72 \times 10^{-11}$), STAL ($P = 1.17 \times 10^{-10}$), and PL ($P = 8.5 \times 10^{-9}$). Co-localization between the two organ size traits (PL and STAL) and the three organ location traits (IL, FSIN, and IDM) suggests a possible common genetic basis between flowering time and flower size in tomato.

SE, defined as the pistil longer than the stamen (Supplemental Note), has been reported as a key determinant of the plant mating system (DePaoli et al., 2011; Zhou et al., 2017). *Style2.1*, the major QTL responsible for SE in cultivated tomatoes, has been identified on chromosome 2 (Chen et al., 2007). However, Vosters et al. (2014) determined that the *Style2.1* association with stigma exertion in domestic tomatoes was likely due to shared phylogenetic relatedness rather than being the causal

variant. In this study, SE and SSR were significantly associated with SL2.50ch01_84029382 ($P = 2.78 \times 10^{-12}$) and SL2.50ch03_60427735 ($P = 1.29 \times 10^{-16}$) on chromosome 1 and 3, respectively (Supplemental Figures S26 and S30). Genotyping analysis revealed that all SE and stigma flush accessions (52 SLC, 22 SLL, 18 SPIM, and 11 other wild accessions) exhibited the C allele, while the stigma inside accessions (6 SLC and 37 SLL accessions) exhibited the T allele of the lead SNP (SL2.50ch03_60427735; Supplemental Figure S36c; Supplemental Table S6). SL2.50ch03_60427735 is located within *Solyc03g098070* that encodes a C2H2L domain class transcription factor. The Arabidopsis homologue of *Solyc03g098070*, *SGR5*, has been reported to regulate the gravitropism of inflorescence stems (Morita et al., 2006). *Solyc03g098070* was highly expressed in flower tissues, especially in the style (Supplemental Figure S36d). These results provide the first evidence that *Solyc03g098070* may be involved in determining the style length.

Despite the widely reported genetic mapping of the fruit size trait, genes responsible for early fruit development remain largely unexplored in tomato (Frary et al., 2000; Causse et al., 2004; Chakrabarti et al., 2013). For the GWAS of OTD, OLD, and OTLD, four clear signals, SL2.50ch01_84023965 corresponding to *fw1.2*, SL2.50ch02_47188498 corresponding to *lc*, SL2.50ch03_64734105 corresponding to *fw3.2* and SL2.50ch11_55052389 corresponding to *fas*, were identified for OTD and OTLD, but not for OLD, indicating that these loci mainly regulate the lateral development of ovary in early fruit stage of tomato (Supplemental Figures S23, S24).

Functional characterization of a candidate gene regulating SD

The phenotype values of SD presented a normal distribution in the natural tomato population investigated here (Supplemental Figure S3). Two significant loci associated with SD in tomato leaf on chromosome 3 and 11 were obtained (Supplemental Figure S17). The significant association ($P = 5.59 \times 10^{-8}$) between SNP SL2.50ch11_53544569 and SD suggested that a genomic sequence related to SNP SL2.50ch11_53544569 forms the major genetic locus (explain 26.2% of the variation) responsible for the natural variation in stomatal formation of tomato leaves (Figure 4A). Two major genotypes, C and T, at the lead SNP (SL2.50ch11_53544569) of the association signal were associated with high- and low-density stomatal phenotypes in tomato, respectively (Figure 4B).

There were 22 genes within the 200-kb sequences flanking the lead SNP (100 kb on either side; Supplemental Table S7). Pairwise linkage disequilibrium (LD) analysis within the 400-kb interval centered on the lead SNP showed that SNPs with high LD to the lead SNP fell into a 70-kb region from 53.51 to 53.58 Mb (Figure 4A). *Solyc11g068970*, encoding an aluminum-activated malate transporter (ALMT; Figure 4C), was the closest gene to the lead SNP (1.8 kb downstream), and the gene and the SNP were in the same LD block (Supplemental Figure S37). Previous studies have shown that many ALMTs are expressed in guard cells and contribute to stomatal closure in plants (Meyer et al., 2011; De AngeLi et al., 2013); therefore, *Solyc11g068970* (which we named *Sl-ACTIVATED MALATE TRANSPORTER15* [*SlALMT15*]) was considered as the causal candidate gene for controlling SD in tomato.

A total of 98 orthologs with high amino acid similarity (>50%) to SlALMT15 were identified in different plant species (Supplemental Figure S38). The SlALMT15 protein was predicted to contain five transmembrane helices and a long C-terminal domain that harbored a conserved WEP-motif (Figure 4E). To investigate functional allelic variation at the *SlALMT15* locus, we analyzed the nucleotide sequence of *SlALMT15* in 13 tomato accessions with diverse stomatal densities, which revealed 17 polymorphisms including five indels and 12 SNPs in the promoter region, and no polymorphism in the gene region (Supplemental Figure S39). Except two indels, the remaining 15 polymorphisms led to 36 possible *cis*-regulatory element changes in the promoter of *SlALMT15* according to PLACE (https://www.dna.affrc.go.jp/PLACE/; Supplemental Table S8). The spatial and temporal expression patterns of *SlALMT15* in high-density stomata accessions (Ts-9 and Ts-53) and low-density stomata accessions (Ts-55 and Ts-52) were then investigated. *SlALMT15* showed high expression levels in stem, flower, and leaf, but low in fruit and root, with the transcript levels higher in most tissues of Ts-9 and Ts-53 than in Ts-55 and Ts-52 (Figure 4F), supporting a role of *SlALMT15* in positively regulating SD in tomato.

To further functionally characterize the role of *SlALMT15* and stomatal formation, we mutated *SlALMT15* in vivo using CRISPR/Cas9 in the high-density stomatal accession Ts-9 (Figure 4D). CRISPR/Cas9-induced knockout mutations (deletions) in *SlALMT15* were detected by PCR and further confirmed by DNA sequencing (Figure 4G). The three investigated mutant lines developed significantly less stomata than Ts-9 in leaves (Figure 4, H and I). To investigate whether *SlALMT15* affects drought stress tolerance by affecting SD, 6-week-old seedlings from the four studied mutant lines and the wild-type were challenged with drought stress by withholding water for 8 d. Dehydration symptoms (leaf wilting) were observed in both mutants and wild-type plants, but the wilting was significantly more severe in the wild-type plants. Changes in drought-related physiological indicators, including net photosynthetic rate, stomatal conductance, transpiration rate, and malondialdehyde (MDA) content (an indicator of cellular membrane integrity and frequently used to evaluate plant drought tolerance), also supported the different degrees of wilting in the mutants and the wild-type plants (Supplemental Figure S40). Together, these results strongly support that SlALMT15 functions in the stomata formation and further affects drought stress tolerance in tomato.

Selective sweeps related to agronomic traits

Long-term domestication and improvement have brought many morphological changes to tomato, such as larger

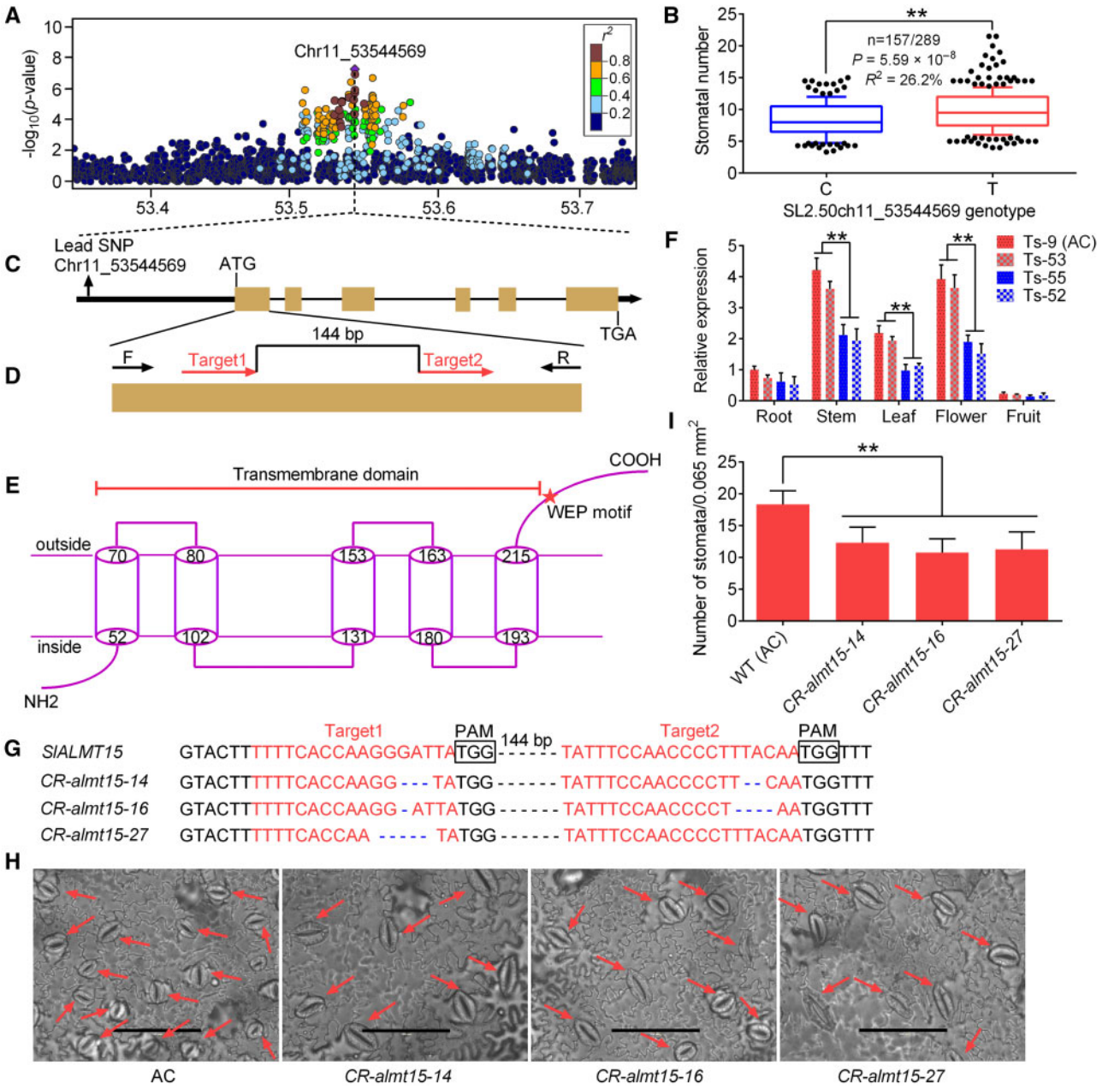

Figure 4 Genetic associations with SD and functional characterization of *SlALMT15*. a, Locus on chromosome 11 associated with SD identified through GWAS. Regional Manhattan plot of a genomic region spanning 400 kb centered at the peak SNP is shown. Lead SNP is indicated in purple. A representation of pairwise r^2 values (a measure of LD) between the lead SNP and all other SNPs in this 400-kb region is shown, where the color of each dot corresponds to the r^2 value according to the color scale. b, Box plot of stomatal densities in tomato accessions with different alleles (C or T) at SNP ch11_53544569. For each box plot, the horizontal line in the box indicates the median value, the box height indicates the 25th to 75th percentiles of the total data, the whiskers indicate the interquartile range, and the outer dots indicate outliers. C, Gene structure of *SlALMT15*. Filled black, filled orange and black lines represent promoter, coding sequence and introns, respectively. D, Schematic illustration of the two sgRNA target sites (red arrows) in *SlALMT15*. Black arrows represent the location of PCR genotyping primers. E, Proposed topology for SlALMT15. The N-terminal contains five transmembrane domains (cylinder). The position of the highly conserved WEP-motif at the C-terminal is indicated. F, Transcript levels of *SlALMT15* in different tomato tissues (root, stem, leaf, flower, and red fruit). Ts-9 (AC) and Ts-53 are high-density stomata accessions, whereas Ts-52 and Ts-55 are low-density stomata accessions. G, Mutated alleles identified from three T_2 *CR-almt15* mutant lines. Red letters indicate sgRNA target sequences, and black boxes indicate protospacer-adjacent motif sequences. H, Images showing the abaxial epidermis of leaves from wild-type and three *CR-almt15* mutant lines. Red arrows point to stomata. Scale bar, 100 μm. I, Number of stomata per unit area in leaves of wild-type and three *CR-almt15* lines. Stomatal number was counted in each field of view (×200, ~0.065 mm^2) of three plants. The third leaf from the top of 6-week-old transgenic and non-transgenic plants was used for stomata number analysis. Data shown in (F) and (I) are means ± SD ($n = 3$). Asterisks indicate significant differences by *t* test: $^{**}P < 0.01$.

flower and fruit (Frary et al., 2000), stronger stem (Ye et al., 2020), embedded stigma (Chen et al., 2007), and so on. To investigate how artificial selection underlies these changes, we searched for signatures of selection in the tomato genome. Based on the phylogenetic analysis, we combined the GX group and the SLC group into one group (SLC_GX) for selective sweep identification. In total, 128 selective sweeps exhibiting lower nucleotide diversity in SLC_GX compared with SPIM were identified, covering 59.85 Mb and harboring 2,492 genes ($\pi_{SPIM}/\pi_{SLC\_GX} > 2.76$; Figure 5A; Supplemental Figure S41; Supplemental Table S9). Comparison between SLL and SLC_GX identified 204 selective sweeps with a cumulative size of 68.22 Mb and harboring 4,959 genes ($\pi_{SLC\_GX}/\pi_{SLL} > 5.38$; Figure 5B; Supplemental Table S10). Collectively, there were 2,132 and 4,599 genes only involved in the domestication or improvement, respectively, and 360 genes in both (Supplemental Table S11). We found that 62% of genes in domestication sweeps (1,545 out of 2,492) and 63% in improvement sweeps (3,122 out of 4,959) detected in our study were also detected in a previous study (Lin et al., 2014; Figure 5C).

To determine the genetic and phenotypic targets of tomato breeding, we compared selective sweeps with the 129 GWAS loci identified in our study (Supplemental Table S4), and observed that 51 out of 129 (39.5%) GWAS loci overlapped with selective sweeps, including three overlapping only with domestication sweeps, 43 only with improvement sweeps, and five with both (Figure 5D; Supplemental Table S12). These 51 loci corresponded to 92.6% of traits (25 out of 27, except SD and STIL) investigated in our study, suggesting that most of these agronomic traits might have been under artificial selection, especially during tomato improvement, consistent with the phenotypic difference of most traits among SPIM, SLC, and SLL (Supplemental Figure S4). Moreover, five loci associated with SSR, flower branch (FB), FSIN, and IDM were located within the domestication and improvement sweeps, indicating a continuous selection of these agronomic traits (Figure 5D). However, in addition to human selection, we cannot rule out other factors such as genetic drift that may account for some phenotypic differences of these traits.

To further look for evidence of artificial selection and take into account population structure, we also used XP-CLR (Chen et al., 2010) to detect selective sweeps and identified 235 domestication sweeps (XP-CLR score > 9.58) and 188 improvement sweeps (XP-CLR score > 17.16), respectively (Supplemental Table S13; Supplemental Figure S42). Surprisingly, very few overlaps were found between sweeps identified based on the nucleotide diversity analysis and those using XP-CLR (7.23% for domestication sweeps and 12.23% for improvement sweeps), and only 4 out of 129 (3.1%) GWAS loci overlapped with selective sweeps identified using XP-CLR.

Discussion

Dissection of the genetic architecture underlying complex agronomic traits among a large number of tomato accessions is helpful to improve the utilization of these germplasms, and provides a foundation for marker-assisted selection in tomato breeding programs. In this study, we evaluated 27 agronomic traits in a collection of 605 tomato accessions. High heritability of these traits suggested that they are mainly regulated by genetic factors, and abundant

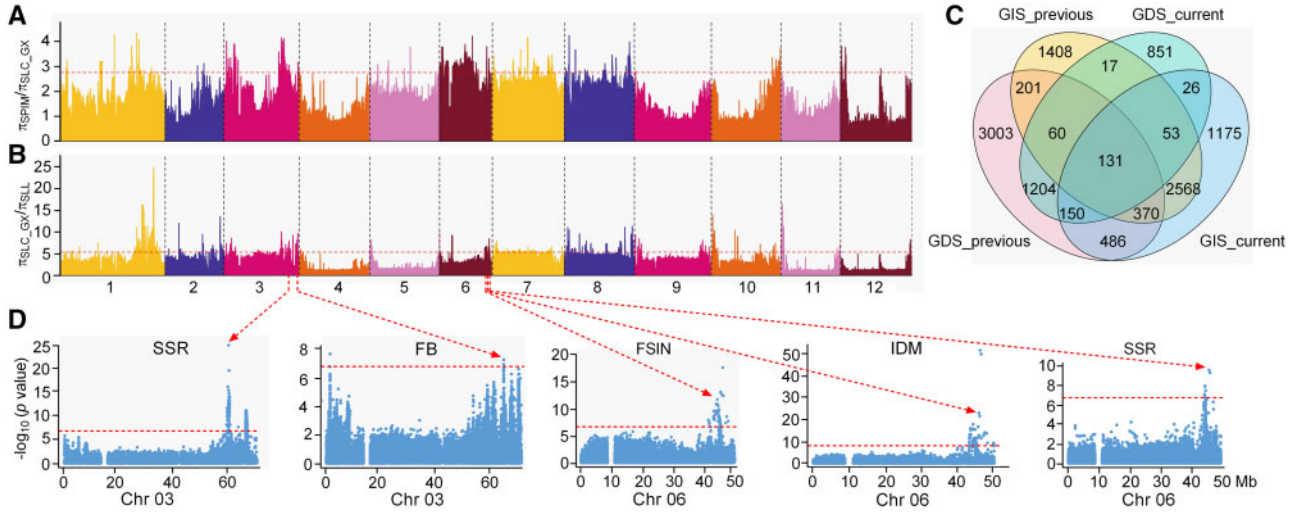

Figure 5 Genome-wide screen of selective sweeps during tomato domestication and improvement. A and B, Selection signals during tomato domestication (A) and improvement (B). The horizontal red dashed lines indicate the genome-wide threshold for domestication sweeps ($\pi_{SPIM}/\pi_{SLC\_GX} > 2.76$) and improvement sweeps ($\pi_{SLC\_GX}/\pi_{SLL} > 5.38$), respectively. C, Comparison of genes within the putative domestication and improvement sweeps in our study with those reported in Lin et al. (2014). GDS_previous and GDS_current: genes in domestication sweeps identified in Lin et al. and in this study, respectively. GIS_previous and GIS_current: genes in improvement sweeps identified in Lin et al. and in this study, respectively. D, Five GWAS association loci that overlapped with both domestication and improvement sweeps are shown. The Bonferroni significance threshold (2.4×10^{-7}) is indicated by the red horizontal dashed lines.

variation in this collection makes it suitable for GWAS (Supplemental Table S2). Significant variation for each of these traits was observed between different groups, suggesting that most of these agronomic traits were likely selected during the tomato evolutionary transition from SPIM to SLL (Supplemental Figure S4), which was further supported by the selective sweep analysis (Supplemental Table S12).

Association mapping has identified candidate trait-associated genes in various crops such as rice (*Oryza sativa*; Yang et al., 2015), maize (*Zea mays*; Li et al., 2013), peach (*Prunus persica*; Cao et al., 2016), watermelon (*Citrullus lanatus*; Guo et al., 2019), and tomato (Tieman et al., 2017). In this study, a total of 239 significant signals were identified, corresponding to 129 genome loci associated with the 27 assessed agronomic traits. Of these, many significant signals for organ size, organ number, and organ location traits were co-located with candidate genes identified by a previous linkage analysis (Zhang et al., 2018; Supplemental Tables S4 and S5). In addition, many additional candidate genes were also identified by our GWAS. Furthermore, one candidate gene (*SlALMT15*) for SD and the other two candidate genes (*SlELF3* and *SlBOP4*) for five traits (FSIN, IL, PL, IDM, and STAL), were successfully mapped (Figures 3 and 4). On the other hand, GWAS results for SE were not consistent with those obtained through QTL mapping (Chen and Tanksley, 2004; Chen et al., 2007). The possible reason for this inconsistency could be that the polymorphism of *Style2.1* was only found between wild and cultivated tomatoes, and the genotype of *Style2.1* has been fixed in cultivated species. Moreover, the association of *Style2.1* with SE was likely due to shared phylogenetic relatedness rather than being the causal variant (Vosters et al., 2014). Here, the variation of *Style3* (SL2.50ch01_84029382) is more abundant in cultivated tomatoes and thus easier to be detected by GWAS (Supplemental Figure S36; Supplemental Table S6).

In this study, several candidate genes for tomato yield-related traits identified through GWAS were homologous to those previously identified in Arabidopsis and tomato (Supplemental Table S5). We found that CLV3/EMBRYO-SURROUNDING REGION genes *CLE2* and *CLE6* were significantly associated with the PN. It has been reported that the CLE gene family plays important roles in cell-to-cell communication to control the balance between stem cell proliferation and differentiation in plant development (Zhang et al., 2014). An acyltransferase gene (*Solyc02g081740*) was identified to be associated with FRNS, OTD, OTLD, and SN. Recently, an acyl-CoA N-acyltransferase, MANY NODED DWARF1, was reported to regulate meristem phase change and plant architecture in barley (*Hordeum vulgare*; Walla et al., 2020). In this study, several LOB domain genes were identified as candidates to regulate various agronomic traits in tomato, similar to those reported previously (Xu et al., 2016a, 2016b).

In this study, *SlALMT15* was functionally verified to regulate SD, and tomato accessions with high *SlALMT15* expression showed high SD (Figure 4). The 36 *cis*-regulatory element changes in the promoter of *SlALMT15* promoter could lead to differences in binding by upstream stomatal-related transcription factors (Supplemental Table S8; Nadeau and Sack, 2002). In addition to *SlLMT15*, a significant association ($P = 1.49 \times 10^{-9}$) between SD and SNP SL2.50ch03_67193998 that was located 1,098 bp upstream of *Solyc03g118280* (encoding a pentatricopeptide repeat-containing protein; PPR), was also detected (Supplemental Table S4). PPR proteins function in multiple aspects of organelle RNA metabolism, such as RNA splicing, editing, degradation, and translation (Schmitz-Linneweber and Small, 2008; Yin et al., 2013), while their involvement in regulating stomatal development has not been reported. Therefore, candidate genes underlying this locus need further characterization.

Two approaches, nucleotide diversity analysis and XP-CLR (Chen et al., 2010), were used to identify selective sweeps (Supplemental Tables S9–S12). Sweeps identified through nucleotide diversity analysis had a high degree of agreement with those reported in a previous study (Lin et al., 2014); however, XP-CLR analysis identified many additional selective sweeps that were not consistent with our nucleotide diversity analysis and the previous results (Lin et al., 2014; Tieman et al., 2017; Zhu et al., 2018). Unexpectedly, using XP-CLR we failed to detect several QTLs related to fruit size, such as *fw2.2* (Frary et al., 2000), *fw3.2* (Chakrabarti et al., 2013), and *fw11.3* (Mu et al., 2017), that are known to have been selected during domestication and improvement of tomato. Therefore, we speculate that the cross-population composite likelihood ratio test implemented in XP-CLR might not be suitable for selective sweep detection when comparing different tomato populations.

Our study provides phenotypic and genetic insights into the variation of yield-related traits in tomato. It also provides a powerful resource for genetic improvement of tomato and other *Solanaceae* crops. Genes and the possible causative SNPs identified here could be used as potential targets for marker-assisted breeding and/or engineering of tomato with enhanced yield and stress tolerance.

Methods

Primers

Primers used in this study are listed in Supplemental Table S14.

Plant materials and sequencing

A diverse worldwide collection of 605 tomato accessions, including 54 *Solanum pimpinellifolium* (SPIM), 140 *S. lycopersicum* var *cerasiforme*, 333 *S. lycopersicum*, 66 Guanxi (GX), 2 *S. habrochaites*, 3 *S. cheesmaniae*, 1 *S. neorickii*, 4 *S. peruvianum*, 1 *S. galapagense*, and 1 *S. corneliomuelleri*, was used for phenotypic and genotypic survey and GWAS analysis. The 66 GX accessions were collected from northwest mountainous region of GX, China, while the remaining 539 accessions and their genotypes were obtained from previous studies (Lin et al., 2014; Tieman et al., 2017; Zhu et al., 2018). Genomic DNA of the 66 GX accessions was isolated

from young leaves using the CTAB method (Murray and Thompson, 1980). Paired-end libraries with insert sizes of ~450–500 bp were constructed using Illumina TruSeq DNA Sample Prep kit according to the manufacturer's instructions, and sequenced on the Illumina HiSeq 2000 platform. Raw reads were processed to remove adaptor and low-quality sequences (base quality of more than 50% bases ≤5), yielding 7.6–10.9 Gb sequences for each of the 66 GX accessions. All library construction, sequencing and sequence processing were carried out by BGI-Shenzhen, China.

Phenotyping

The 605 tomato accessions were first grown in 15 greenhouses at the Wuhan Academy of Agricultural Sciences, China, with three biological replicates for each accession (five greenhouses for each biological replicates). Seeds were sown in 50-hole trays containing the soil, peat, and vermiculite (1:1:1) in early March 2016 and then transplanted to the field in mid-April 2016. For each biological replicate, the accessions were grown in a randomized design, with 12 plants for each accession. Field management, including irrigation, fertilizer application, and pest control, followed essentially normal agricultural practice.

The 27 agronomic traits were classified into three categories: organ location traits (FIN, FSIN, IL, and IDM), organ number traits (FB, FLNS, FLNT, FRNS, FRNT, IT, PN, SD, and SN), and organ size traits (FL, FSD, FSL, OLD, OTD, OTLD, PL, SE, SL, SPR, SS, SSR, STAL, and STIL). All four organ location traits, all nine organ number traits, and three organ size traits (FL, SE, and SS) were evaluated by manual observation and counting. The eight organ size traits (FSD, FSL, OLD, OTD, PL, SL, STAL, and STIL) were analyzed by vernier calipers in ten fruits or flowers, which were randomly picked from six plants for each accession in each biological replicate. OTLD was calculated as OTD/OLD, SPR was calculated as SL/PL, and SSR was calculated as STAL/(STIL+OLD) (Supplemental Note).

For SD investigation, seeds were sown in 50-hole trays filled with a 1:1:1 mixture of peat, vermiculite, and soil. Subsequently, seedlings were grown at 28°C/20°C under a 16-h d/8-h night photoperiod of natural light in a greenhouse. The third fully expanded leaves from the top were sampled after 5 weeks of growth. The lower epidermal strips of leaf blade apical were peeled, stained with toluidine blue and then loaded onto slides. The number of stomata was observed and counted under an Olympus BHS/BHT microscope (BH-2) at a field of 200×: ≈0.065 mm^2. Three uniformly growing plants were observed separately as three biological replicates. For each leaf, the average stomata number at the leaf blade apical under three random fields was used as one biological replicate.

Phenotypic data processing and statistical analysis

Frequency distribution analysis of phenotypic values of each agronomic trait was performed using EXCEL 2010. The coefficient of variation was calculated independently for each agronomic trait as σ/μ, where σ and μ are the standard deviation and mean of each agronomic trait in the association panel, respectively. Phenotypic data collected from the three biological sample sets were used to calculate H^2 as previously described (Ye et al., 2019). Significance analysis of difference in the agronomic traits among the three subgroups (SPIM, SLC, and SLL) was conducted using the ANOVA (Analysis of Variance) test. A post-ANOVA Tukey's honestly significant difference test was then used to test for significant differences between pairwise means among the three subgroups. Spearman's test was used to evaluate the pairwise correlation between phenotypes.

SNP identification and population analyses

The cleaned paired-end reads of the 66 GX accessions were aligned to the tomato reference genome (Heinz 1706; version SL2.5; Tomato Genome Consortium, 2012) using BWA (Li and Durbin, 2009) with default parameters. The SNP dataset of 539 previously reported tomato accessions was downloaded from the Sol Genomics Network (https://solgenomics.net/). The SNP calling of all the 605 accessions was performed by adding the allele information of GX accessions to the SNP dataset of 539 accessions using GATK (McKenna et al., 2010).

For the phylogenetic analysis, SNPs of all accessions were first filtered with a missing data rate <15% and minor allele frequency (MAF) > 0.05. The resulting SNPs at fourfold degenerated sites (23,635) were then used to construct a neighbor-joining tree using the IQTREE software (Nguyen et al., 2015) with 500 bootstrap replicates. Principal component analysis (PCA) was performed with PLINK version 1.9 (Purcell et al., 2007). The population structure of the tomato accessions was inferred using fastStructure (Raj et al., 2014) with all SNPs for each K ($K = 2–4$). Genome-wide π and F_{ST} were calculated for each subpopulation with VCFtools version 0.1.15 (Danecek et al., 2011) using 100-kb sliding windows with a step size of 10 kb.

Genome-wide association study

SNPs with the MAF ≥ 0.05 and miss rate <0.15 were used to perform the GWAS. Associations between SNPs and the 27 agronomic traits were detected using a linear mixed model (Zhang et al., 2010) implemented in GEMMA (Zhou and Stephens, 2012). To control false associations, the K matrix (BN matrix calculated by *emmax-kin*) modeled population structure as a random effect and the Q matrix (the top 10 PCs of PCA) as a fixed effect. The K and Q matrix was calculated with 23,635 SNPs (same as those used for phylogenetic analyses). The genome-wide suggestive and significance thresholds of associations were set at $P = 1/n$ and $P = 0.05/n$, respectively, where n is the effective number of independent SNPs (Ye et al., 2019), which corresponded to $P = 2.4 \times 10^{-7}$ and $P = 1.2 \times 10^{-8}$. Pairwise LD between the suggestive/significant SNPs for each agronomic trait was calculated using the PLINK software version 1.9 (Purcell et al., 2007). All 239 significant lead SNPs on the tomato genome were integrated into different loci by dividing the whole genome into 200-kb partitions, and the number of

significant loci was counted. We further made use of LD estimates for grouping associated SNPs into loci.

Identification of domestication and improvement sweeps

For selective sweep detection, we combined SLC and GX groups into a single group (SLC_GX) to exclude the potential effect of genetic drift. To identify genomic regions affected by domestication and improvement, we first measured the level of nucleotide diversity (π) within 100-kb sliding windows with a step size of 10 kb in SPIM, SLC_GX, and SLL. Candidate domestication and improvement sweeps were identified with the top 5% largest π_{SPIM}/π_{SLC_GX} (2.76) and π_{SLC_GX}/π_{SLL} (5.33) values, respectively. Finally, windows that were $\leq$100 kb apart were merged into a single selected region.

We also identified domestication and improvement sweeps using XP-CLR (Chen et al., 2010) with the following parameters: "-w1 0.0005 100 100 1 -p0 0.7." Genome regions with the top 5% largest XP-CLR scores were considered as the potential selective sweeps, and that $\leq$100 kb apart were merged into a single selected region.

Candidate gene sequencing

To identify the genotype of SL2.50ch03_60427735 in the tomato population, DNA fragments containing this SNP were amplified by PCR in 146 tomato accessions. To detect the variation in the *SlALMT15* gene region, DNA sequences of *SlALMT15* in 13 tomato accessions (TS-9, TS-53, TS-67, TS-91, TS-572, TS-577, and TS-604 with high-density stomata; TS-52, TS-55, TS-210, TS-531, np3, and db6 with low-density stomata) were amplified by PCR using primers listed in Supplemental Table S14. The PCR products were sequenced and compared against the reference genome for polymorphism analysis.

RNA isolation and expression analysis

Total RNA was isolated from various tissue of Ts-9, Ts-52, Ts-53, and Ts-55 using TRIZOL reagent (Invitrogen, Carlsbad, CA, USA). cDNAs were synthesized from the total RNA using HiScriptII Reverse Transcriptase (Vazyme, Miramar Beach, FL, USA), according to the manufacturer's protocol. Gene expression was quantified using RT-qPCR as previously described (Liu et al., 2012) using primer pairs listed in Supplemental Table S14. Three biological replicates were conducted for each sample. The *Actin* gene (*Solyc11g008430*) was used as an internal standard. The relative expression of specific genes was quantified using the comparative C_T method.

CRISPR/Cas9 construct design and transformation

CRISPR/Cas9 binary vectors (pTX; Ye et al., 2017), in which the target sequence was driven by the tomato U6 promoter and Cas9 by 2 × 35S, were used for editing of the *SlALMT15* gene. Two *SlALMT15*-specific target sites (sgRNA1 and sgRNA2) were manually selected. The recombinant pTX vector was designed to produce defined deletions within the coding sequence of *SlALMT15* using two sgRNAs alongside the *Cas9* endonuclease gene. The plasmid with the correct sgRNA insertion was introduced into *Agrobacterium tumefaciens* strain C58 by electroporation and subsequently transformed into the tomato genome via explants of cotyledon. The high-density stomata accession TS-9 (Ailsa Craig) was used for transformation. Positive detection of T_0 plants was conducted by PCR using Cas9-specific primers (Supplemental Table S14). The CRISPR/Cas9-induced mutations were further genotyped by PCR sequencing using CRISPR/Cas9 detection primers.

Drought tolerance assays and measurement of physiological indexes

Three T_2 CR-*almt15* lines and wild-type plants were grown at 28°C/20°C (day/night) with a 16 h photoperiod in 10 cm (diameter) plastic pots in a greenhouse with a medium-light intensity (~160 μmoles of photons m^{-2} s^{-1}). For drought tolerance testing, 5-week-old seedling plants were fully watered. Water was then withheld from the seedlings in drought treatment group for 8 d. The control group of seedlings was watered every 2 d. The phenotype was investigated and recorded at the end of the drought treatment. The third fully expanded leaves from the control and stressed plants were used for the determination of physiological indicators.

Net photosynthetic rate, transpiration rate, and stomatal conductance were measured using the LI-6400XT Portable Photosynthesis System (LI-COR, Lincoln, NE, USA) following the manufacturer's instructions. The values of photosynthesis at 400 μmol·mol^{-1} minus the values of photosynthesis at 0 μmol·mol^{-1} were used to approximately represent the photosynthesis capacity (Perry et al., 1983). At least five replicated plants were used for each measurement.

The MDA levels were measured as previously described (Yu et al., 2018). Briefly, 200 mg of ground leaf samples were homogenized using 3 mL of 5 (w/v) trichloroacetic acid (TCA) and incubated for 20 min. After centrifuged at 5,500 rpm for 25 min, 2 mL of supernatant was mixed with 2 mL of 0.67% thiobarbituric acid in 10% TCA. After 30 min incubating in boiling water, the mixture was centrifuged and the absorbance of the supernatant was determined spectrophotometrically at 450, 532, and 600 nm. The content of MDA (μmol/g) was calculated as $(6.45 \times (A_{532}-A_{600})-0.56\times A_{450} \times V)/(1000 \times W)$, where V represents the volume of the extraction buffer (mL), and W represents the weight of the sample (g). The significant differences between CRISPR/Cas9-induced mutations and wild-type were evaluated using *t* test.

Data availability

Raw sequences of the 66 GX accessions have been deposited in the Sequence Read Archive of National Center for Biotechnology Information under the accession number PRJNA666021. The SNPs have also been deposited into the Figshare database (10.6084/m9.figshare.13019237).

Accession numbers

Sequence data from this article can be found in the Sol Genomics Network under accession numbers Solyc06g074350 (*SP*), Solyc06g071140 (*SlELF3*), Solyc06g071830 (*SlBOP4*), Solyc03g098070 (*Style3*), and Solyc11g068970 (*SlALMT15*).

Supplemental data

The following materials are available in the online version of this article.

Supplemental Note. Methods and description for phenotype evaluation.

Supplemental Figure S1. Field phenotypes of some GX tomatoes.

Supplemental Figure S2. Diagram of the measured 27 agronomic traits in this study.

Supplemental Figure S3. Frequency distribution of phenotypic values of 27 agronomic traits.

Supplemental Figure S4. Phenotypic distribution of 27 agronomic traits in different tomato subgroups.

Supplemental Figure S5. Correlations between the analyzed phenotypes.

Supplemental Figure S6. GWAS for FIN.

Supplemental Figure S7. GWAS for FSIN.

Supplemental Figure S8. GWAS for IL.

Supplemental Figure S9. GWAS for IDM.

Supplemental Figure S10. GWAS for FB.

Supplemental Figure S11. GWAS for FLNS.

Supplemental Figure S12. GWAS for FLNT.

Supplemental Figure S13. GWAS for FRNS.

Supplemental Figure S14. GWAS for FRNT.

Supplemental Figure S15. GWAS for IT.

Supplemental Figure S16. GWAS for PN.

Supplemental Figure S17. GWAS for SD.

Supplemental Figure S18. GWAS for SN.

Supplemental Figure S19. GWAS for FL.

Supplemental Figure S20. GWAS for FSD.

Supplemental Figure S21. GWAS for FSL.

Supplemental Figure S22. GWAS for OLD.

Supplemental Figure S23. GWAS for OTD.

Supplemental Figure S24. GWAS for OTLD.

Supplemental Figure S25. GWAS for PL.

Supplemental Figure S26. GWAS for SE.

Supplemental Figure S27. GWAS for SL.

Supplemental Figure S28. GWAS for SPR.

Supplemental Figure S29. GWAS for SS.

Supplemental Figure S30. GWAS for SSR.

Supplemental Figure S31. GWAS for STAL.

Supplemental Figure S32. GWAS for STIL.

Supplemental Figure S33. Suggestive loci ($P < 2.4 \times 10^{-7}$) for the GWAS results associated with 27 tomato traits.

Supplemental Figure S34. Candidate genes identified by GWAS for PN.

Supplemental Figure S35. Phylogenetic analysis of ELFs in tomato and Arabidopsis BOPs and LOBs in tomato.

Supplemental Figure S36. GWAS for SE and expression profiles of the candidate gene.

Supplemental Figure S37. Gene structure and LD blocks surrounding *SlALMT15*.

Supplemental Figure S38. Phylogenetic analyses of SlALMT15.

Supplemental Figure S39. Variations in the *SlALMT15* promoter among 13 tomato accessions with different stomata densities.

Supplemental Figure S40. CRISPR/Cas9-engineered mutations in *SlALMT15* result in enhanced drought tolerance in tomato.

Supplemental Figure S41. Distribution of nucleotide diversity (π) for SP (light pink), SLC (orange), GX (red), SLC_GX (black), and SLL (light blue) across the 12 chromosomes.

Supplemental Figure S42. Genome-wide distribution of selective sweeps in tomato detected by XP-CLR analysis.

Supplemental Table S1. Tomato accessions used in this study.

Supplemental Table S2. Summary statistics of the 27 agronomic traits.

Supplemental Table S3. Phenotypic data of 27 agronomic traits assessed in the 605 tomato accessions.

Supplemental Table S4. List of 239 detected associated lead SNPs by GWAS.

Supplemental Table S5. List of lead SNPs significantly associated with 27 agronomic traits and the related candidate genes.

Supplemental Table S6. PCR sequencing identified an association between SL2.50ch03_60427735 and stigma exertion in 146 tomato accessions.

Supplemental Table S7. Genes within 100 kb of the most highly associated SNP with SD on chromosome 11.

Supplemental Table S8. Predicted cis-acting elements in the SlALMT15 promoter.

Supplemental Table S9. Putative domestication sweeps (from SPIM to SLC_GX) detected by nucleotide diversity analysis.

Supplemental Table S10. Putative improvement sweeps (from SLC_GX to SLL) detected by nucleotide diversity analysis.

Supplemental Table S11. Genes within the putative domestication and improvement sweeps.

Supplemental Table S12. List of 51 GWAS loci under selection during tomato domestication and/or improvement.

Supplemental Table S13. Putative domestication (from SPIM to SLC_GX) and improvement (from SLC_GX to SLL) sweeps detected by the XP-CLR method.

Supplemental Table S14. Primers used in this study.

Funding

This work was supported by grants from the National Key Research and Development Program of China (grant no. 2017YFD0101902 to Z.Y.), the National Natural Science Foundation of China (grant nos. 31872118 to Z.Y. and 31801861 to J.Y.), HZAU Research Start Fund for High-Level talents (grant no. 105/11042010004 to J.Y.), China

Agricultural Research System (CARS-23-A-03), and the US National Science Foundation (grant no. IOS-1855585 to Z.F. and J.J.G.).

Conflict of interest statement: The authors declare no conflict of interest.

References

Barrero LS, Cong B, Wu F, Tanksley SD (2006) Developmental characterization of the fasciated locus and mapping of Arabidopsis candidate genes involved in the control of floral meristem size and carpel number in tomato. Genome **49:** 991–1006

Cao K, Zhou Z, Wang Q, Guo J, Zhao P, Zhu G, Fang W, Chen C, Wang X, Wang X, Tian Z, Wang L (2016) Genome-wide association study of 12 agronomic traits in peach. Nat Commun **7:** 13246

Causse M, Duffe P, Gomez MC, Buret M, Damidaux R, Zamir D, Gur A, Chevalier C, Lemaire-Chamley M, Rothan C (2004) A genetic map of candidate genes and QTLs involved in tomato fruit size and composition. J Exp Bot **55:** 1671–1685

Chakrabarti M, Zhang N, Sauvage C, Munos S, Blanca J, Canizares J, Diez MJ, Schneider R, Mazourek M, McClead J, et al. (2013) A cytochrome P450 regulates a domestication trait in cultivated tomato. Proc Natl Acad Sci USA **110:** 17125–17130

Chen H, Patterson N, Reich D (2010) Population differentiation as a test for selective sweeps. Genome Res **20:** 393–402

Chen KY, Cong B, Wing R, Vrebalov J, Tanksley SD (2007) Changes in regulation of a transcription factor lead to autogamy in cultivated tomatoes. Science **318:** 643–645

Chen KY, Tanksley SD (2004) High-resolution mapping and functional analysis of se2.1: a major stigma exsertion quantitative trait locus associated with the evolution from allogamy to autogamy in the genus Lycopersicon. Genetics **168:** 1563–1573

Cong B, Barrero LS, Tanksley SD (2008) Regulatory change in YABBY-like transcription factor led to evolution of extreme fruit size during tomato domestication. Nat Genet **40**: 800–804

Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, Handsaker RE, Lunter G, Marth GT, Sherry ST, et al. (2011) The variant call format and VCFtools. Bioinformatics **27:** 2156–2158

De Angeli A, Zhang J, Meyer S, Martinoia E (2013) AtALMT9 is a malate-activated vacuolar chloride channel required for stomatal opening in Arabidopsis. Nat Commun **4:** 1804

DePaoli HC, Brito MS, Quiapim AC, Teixeira SP, Goldman GH, Dornelas MC, Goldman MH (2011) Stigma/style cell cycle inhibitor 1 (SCI1), a tissue-specific cell cycle regulator that controls upper pistil development. New Phytol **190:** 882–895

Frary A, Nesbitt TC, Grandillo S, Knaap E, Cong B, Liu J, Meller J, Elber R, Alpert KB, Tanksley SD (2000) fw2.2: a quantitative trait locus key to the evolution of tomato fruit size. Science **289:** 85–88

Guo S, Zhao S, Sun H, Wang X, Wu S, Lin T, Ren Y, Gao L, Deng Y, Zhang J, et al. (2019) Resequencing of 414 cultivated and wild watermelon accessions identifies selection for fruit quality traits. Nat Genet **51:** 1616–1623

Herrero E, Kolmos E, Bujdoso N, Yuan Y, Wang M, Berns MC, Uhlworm H, Coupland G, Saini R, Jaskolski M, et al. (2012) EARLY FLOWERING4 recruitment of EARLY FLOWERING3 in the nucleus sustains the Arabidopsis circadian clock. Plant Cell **24:** 428–443

Hicks KA, Albertson TM, Wagner DR (2001) *EARLY FLOWERING3* encodes a novel protein that regulates circadian clock function and flowering in Arabidopsis. Plant Cell **13:** 1281–1292

Huang X, Wei X, Sang T, Zhao Q, Feng Q, Zhao Y, Li C, Zhu C, Lu T, Zhang Z, et al. (2010) Genome-wide association studies of 14 agronomic traits in rice landraces. Nat Genet **42:** 961–967

Ji J, Strable J, Shimizu R, Koenig D, Sinha N, Scanlon MJ (2010) WOX4 promotes procambial development. Plant Physiol **152:** 1346–1356

Lampard GR, Macalister CA, Bergmann DC (2008) Arabidopsis stomatal initiation is controlled by MAPK-mediated regulation of the bHLH SPEECHLESS. Science **322:** 1113–1116

Li H, Durbin R (2009) Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics **25:** 1754–760

Li H, Peng Z, Yang X, Wang W, Fu J, Wang J, Han Y, Chai Y, Guo T, Yang N, et al. (2013) Genome-wide association study dissects the genetic architecture of oil biosynthesis in maize kernels. Nat Genet **45:** 43–50

Li M, Wang X, Li C, Li H, Zhang J, Ye Z (2018) Silencing GRAS2 reduces fruit weight in tomato. J Integr Plant Biol **60:** 498–513

Li YH, Zhou G, Ma J, Jiang W, Jin LG, Zhang Z, Guo Y, Zhang J, Sui Y, Zheng L et al. (2014) De novo assembly of soybean wild relatives for pan-genome analysis of diversity and agronomic traits. Nat Biotechnol **32:** 1045–1052

Lifschitz E, Eviatar T, Rozman A, Shalit A, Goldshmidt A, Amsellem Z, Alvarez JP, Eshed Y (2006) The tomato FT ortholog triggers systemic signals that regulate growth and flowering and substitute for diverse environmental stimuli. Proc Natl Acad Sci USA **103:** 6398–6403

Lin T, Zhu G, Zhang J, Xu X, Yu Q, Zheng Z, Zhang Z, Lun Y, Li S, Wang X, et al. (2014) Genomic analyses provide insights into the history of tomato breeding. Nat Genet **46:** 1220–1226

Lippman ZB, Cohen O, Alvarez JP, Abu-Abied M, Pekker I, Paran I, Eshed Y, Zamir D (2008) The making of a compound inflorescence in tomato and related nightshades. PLoS Biol **6:** e288

Liu H, Ouyang B, Zhang J, Wang T, Li H, Zhang Y, Yu C, Ye Z (2012) Differential modulation of photosynthesis, signaling, and transcriptional regulation between tolerant and sensitive tomato genotypes under cold stress. PLoS One **7:** e50785

MacAlister CA, Park SJ, Jiang K, Marcel F, Bendahmane A, Izkovich Y, Eshed Y, Lippman ZB (2012) Synchronization of the flowering transition by the tomato *TERMINATING FLOWER* gene. Nat Genet **44:** 1393–1398

Mata-Nicolas E, Montero-Pau J, Gimeno-Paez E, Garcia-Carpintero V, Ziarsolo P, Menda N, Mueller LA, Blanca J, Canizares J, van der Knaap E, et al. (2020) Exploiting the diversity of tomato: the development of a phenotypically and genetically detailed germplasm collection. Hortic Res **7:** 66

McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, Garimella K, Altshuler D, Gabriel S, Daly M, et al. (2010) The genome analysis toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res **20:** 1297–1303

Meyer S, Scholz-Starke J, De Angeli A, Kovermann P, Burla B, Gambale F, Martinoia E (2011) Malate transport by the vacuolar AtALMT6 channel in guard cells is subject to multiple regulation. Plant J **67:** 247–257

Molinero-Rosales N, Jamilena M, Zurita S, Gomez P, Capel J, Lozano R (1999) FALSIFLORA, the tomato orthologue of FLORICAULA and LEAFY, controls flowering time and floral meristem identity. Plant J **20:** 685–693

Morales-Navarro S, Perez-Diaz R, Ortega A, de Marcos A, Mena M, Fenoll C, Gonzalez-Villanueva E, Ruiz-Lara S (2018) Overexpression of a SDD1-Like Gene from wild tomato decreases stomatal density and enhances dehydration avoidance in Arabidopsis and cultivated tomato. Front Plant Sci **9:** 940

Morita MT, Sakaguchi K, Kiyose S, Taira K, Kato T, Nakamura M, Tasaka M (2006) A C2H2-type zinc finger protein, SGR5, is involved in early events of gravitropism in Arabidopsis inflorescence stems. Plant J **47:** 619–628

Mu Q, Huang Z, Chakrabarti M, Illa-Berenguer E, Liu X, Wang Y, Ramos A, van der Knaap E (2017). Fruit weight is controlled by *Cell Size Regulator* encoding a novel protein that is expressed in maturing tomato fruits. PLoS Genet **13:** e1006930

Munos S, Ranc N, Botton E, Berard A, Rolland S, Duffe P, Carretero Y, Le Paslier MC, Delalande C, Bouzayen M, et al. (2011) Increase in tomato locule number is controlled by two

single-nucleotide polymorphisms located near *WUSCHEL*. Plant Physiol **156:** 2244–2254

Murray MG, Thompson WF (1980) Rapid isolation of high molecular-weight plant DNA. Nucleic Acids Res **8:** 4321–4325

Nadeau JA, Sack FD (2002) Stomatal development in Arabidopsis. Arabidopsis Book **1:** e0066

Nesbitt TC, Tanksley SD (2001) fw2.2 directly affects the size of developing tomato fruit, with secondary effects on fruit number and photosynthate distribution. Plant Physiol **127:** 575–583

Nguyen LT, Schmidt HA, von Haeseler A, Minh BQ (2015) IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. Mol Biol Evol **32:** 268–274

Ortega A, de Marcos A, Illescas-Miranda J, Mena M, Fenoll C (2019) The tomato genome encodes SPCH, MUTE, and FAMA candidates that can replace the endogenous Functions of their Arabidopsis orthologs. Front Plant Sci **10:** 1300

Perry SW, Krieg DR, Hutmacher RB (1983) Photosynthetic rate control in cotton: photorespiration. Plant Physiol **73:** 662–665

Pillitteri LJ, Torii KU (2012) Mechanisms of stomatal development. Annu Rev Plant Biol **63:** 591–614

Pnueli L, Carmel-Goren L, Hareven D, Gutfinger T, Alvarez J, Ganal M, Zamir D, Lifschitz E (1998) The *SELF-PRUNING* gene of tomato regulates vegetative to reproductive switching of sympodial meristems and is the ortholog of *CEN* and *TFL1*. Development **125:** 1979–1989

Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, Maller J, Sklar P, de Bakker PI, Daly MJ, et al. (2007) PLINK: a tool set for whole-genome association and population--based linkage analyses. Am J Hum Genet **81:** 559–575

Raj A, Stephens M, Pritchard JK (2014) fastSTRUCTURE: variational inference of population structure in large SNP data sets. Genetics **197:** 573–U207

Rubenach AJ, Hecht V, Vander Schoor JK, Liew LC, Aubert G, Burstin J, Weller JL (2017) *EARLY FLOWERING3* redundancy fine-tunes photoperiod sensitivity. Plant Physiol **173:** 2253–2264

Schmitz-Linneweber C, Small I (2008) Pentatricopeptide repeat proteins: a socket set for organelle gene expression. Trends Plant Sci **13:** 663–670

Serna L, Fenoll C (2000) Stomatal development in Arabidopsis: how to make a functional pattern. Trends Plant Sci **5:** 458–460

Shu B, Xie G, Guo Y (1995) Preliminary report on the discovery and observation of wild tomato in China. J Beijing Agric Univ **1**: 30

Soyk S, Lemmon ZH, Oved M, Fisher J, Liberatore KL, Park SJ, Goren A, Jiang K, Ramos A, van der Knaap E, et al. (2017a) Bypassing negative epistasis on yield in tomato imposed by a domestication gene. Cell **169:** 1142–1155 e1112

Soyk S, Muller NA, Park SJ, Schmalenbach I, Jiang K, Hayama R, Zhang L, Van Eck J, Jimenez-Gomez JM, Lippman ZB (2017b) Variation in the flowering gene *SELF PRUNING 5G* promotes day-neutrality and early yield in tomato. Nat Genet **49:** 162–168

Sugano SS, Shimada T, Imai Y, Okawa K, Tamai A, Mori M, Hara-Nishimura I (2010) Stomagen positively regulates stomatal density in Arabidopsis. Nature **463:** 241–244

Tieman D, Zhu G, Resende MF, Jr, Lin T, Nguyen C, Bies D, Rambla JL, Beltran KS, Taylor M, Zhang B, et al. (2017) A chemical genetic roadmap to improved tomato flavor. Science **355:** 391–394

Tomato Genome Consortium (2012) The tomato genome sequence provides insights into fleshy fruit evolution. Nature **485:** 635–641

Varshney RK, Saxena RK, Upadhyaya HD, Khan AW, Yu Y, Kim C, Rathore A, Kim D, Kim J, An S, et al. (2017) Whole-genome resequencing of 292 pigeonpea accessions identifies genomic regions associated with domestication and agronomic traits. Nat Genet **49:** 1082–1088

Varshney RK, Thudi M, Roorkiwal M, He W, Upadhyaya HD, Yang W, Bajaj P, Cubry P, Rathore A, Jian J, et al. (2019) Resequencing of 429 chickpea accessions from 45 countries provides insights into genome diversity, domestication and agronomic traits. Nat Genet **51:** 857–864

Vosters SL, Jewell CP, Sherman NA, Einterz F, Blackman BK, Moyle LC (2014) The timing of molecular and morphological changes underlying reproductive transitions in wild tomatoes (*Solanum* sect. *Lycopersicon*). Mol Ecol **23:** 1965–1978

Walla A, Wilma van Esse G, Kirschner GK, Guo G, Brunje A, Finkemeier I, Simon R, von Korff M (2020) An acyl-CoA N-acyltransferase regulates meristem phase change and plant architecture in barley. Plant Physiol **183:** 1088–1109

Wang X, Gao L, Jiao C, Stravoravdis S, Hosmani PS, Saha S, Zhang J, Mainiero S, Strickler SR, Catala C, et al. (2020) Genome of *Solanum pimpinellifolium* provides insights into structural variants during tomato breeding. Nat Commun **11:** 5817

Xu C, Liberatore KL, MacAlister CA, Huang Z, Chu YH, Jiang K, Brooks C, Ogawa-Ohnishi M, Xiong G, Pauly M, et al. (2015) A cascade of arabinosyltransferases controls shoot meristem size in tomato. Nat Genet **47:** 784–792

Xu C, Luo F, Hochholdinger F (2016a) LOB domain proteins: beyond lateral organ boundaries. Trends Plant Sci **21:** 159–167

Xu C, Park SJ, Van Eck J, Lippman ZB (2016b) Control of inflorescence architecture in tomato by BTB/POZ transcriptional regulators. Genes Dev **30**: 2048–2061.

Yang W, Guo Z, Huang C, Wang K, Jiang N, Feng H, Chen G, Liu Q, Xiong L (2015) Genome-wide association study of rice (*Oryza sativa* L.) leaf traits with a high-throughput leaf scorer. J Exp Bot **66:** 5605–5615

Ye J, Li W, Ai G, Li C, Liu G, Chen W, Wang B, Wang W, Lu Y, Zhang J, et al. (2019) Genome-wide association analysis identifies a natural variation in basic helix-loop-helix transcription factor regulating ascorbate biosynthesis via D-mannose/L-galactose pathway in tomato. PLoS Genet **15:** e1008149

Ye J, Tian R, Meng X, Tao P, Li C, Liu G, Chen W, Wang Y, Li H, Ye Z, et al. (2020) Tomato *SD1*, encoding a kinase-interacting protein, is a major locus controlling stem development. J Exp Bot **71:** 3575–3587

Ye J, Wang X, Hu T, Zhang F, Wang B, Li C, Yang T, Li H, Lu Y, Giovannoni JJ, et al. (2017) An InDel in the promoter of *Al-ACTIVATED MALATE TRANSPORTER9* selected during tomato domestication determines fruit malate contents and aluminum tolerance. Plant Cell **29:** 2249–2268

Yin P, Li Q, Yan C, Liu Y, Liu J, Yu F, Wang Z, Long J, He J, Wang HW, et al. (2013) Structural basis for the modular recognition of single-stranded RNA by PPR proteins. Nature **504:** 168–171

Yoo CY, Pence HE, Jin JB, Miura K, Gosney MJ, Hasegawa PM, Mickelbart MV (2010) The Arabidopsis GTL1 transcription factor regulates water use efficiency and drought tolerance by modulating stomatal density via transrepression of *SDD1*. Plant Cell **22:** 4128–4141

Yu C, Song L, Song J, Ouyang B, Guo L, Shang L, Wang T, Li H, Zhang J, Ye Z (2018) ShCIGT, a Trihelix family gene, mediates cold and drought tolerance by interacting with SnRK1 in tomato. Plant Sci **270:** 140–149

Zhang S, Yu H, Wang K, Zheng Z, Liu L, Xu M, Jiao Z, Li R, Liu X, Li J, et al. (2018) Detection of major loci associated with the variation of 18 important agronomic traits between *Solanum pimpinellifolium* and cultivated tomatoes. Plant J **95:** 312–323

Zhang Y, Yang S, Song Y, Wang J (2014) Genome-wide characterization, expression and functional analysis of CLV3/ESR gene family in tomato. BMC Genomics **15:** 827

Zhang Z, Ersoz E, Lai CQ, Todhunter RJ, Tiwari HK, Gore MA, Bradbury PJ, Yu J, Arnett DK, Ordovas JM, et al. (2010) Mixed linear model approach adapted for genome-wide association studies. Nat Genet **42:** 355–360

Zhao J, Huang X, Ouyang X, Chen W, Du A, Zhu L, Wang S, Deng XW, Li S (2012) *OsELF3-1*, an ortholog of Arabidopsis *EARLY FLOWERING 3*, regulates rice circadian rhythm and photoperiodic flowering. PLoS One **7:** e43705

Zhou H, Li P, Xie W, Hussain S, Li Y, Xia D, Zhao H, Sun S, Chen J, Ye H, et al. (2017) Genome-wide association analyses reveal the genetic basis of stigma exsertion in rice. Mol Plant **10:** 634–644

Zhou X, Stephens M (2012) Genome-wide efficient mixed-model analysis for association studies. Nat Genet **44:** 821–824

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 e212