Repairing a deleterious domestication variant in a floral regulator of tomato by base editing

Anna N. Glaus[1,2], Marion Brechet[1,2], Ludivine Lebeigle[2], Justyna Iwaszkiewicz[3], Giovanna Ambrosini[4,5], Irene Julca[6], Jing Zhang[7], Robyn Roberts[7], Christian Iseli[4,5], Nicolas Guex[4,5], José Jiménez-Gómez[8], Natasha Glover[6], Gregory B. Martin[7,9], Susan Strickler[7,10,11], and Sebastian Soyk[1,2] *

[1]Department of Plant Molecular Biology, University of Lausanne, 1015 Lausanne, Switzerland

[2]Center for Integrative Genomics, University of Lausanne, 1015 Lausanne, Switzerland

[3]Molecular Modeling Group, Swiss Institute of Bioinformatics, 1015 Lausanne, Switzerland

[4]Bioinformatics Competence Centre, University of Lausanne, 1015 Lausanne, Switzerland

[5]Bioinformatics Competence Centre, École Polytechnique Fédérale de Lausanne, 1015 Lausanne, Switzerland

[6]SIB Swiss Institute of Bioinformatics, 1015 Lausanne, Switzerland

[7]Boyce Thompson Institute for Plant Research, Ithaca, NY 14853, USA

[8]Centro de Biotecnología y Genómica de Plantas (CBGP), Madrid, Spain

[9]Plant Pathology and Plant-Microbe Biology Section, School of Integrative Plant Science, Cornell University, Ithaca, NY 14853, USA

[10]Plant Science and Conservation Chicago Botanic Garden, Glencoe, IL 60022, USA.

[11]Plant Biology and Conservation Program Northwestern University, Evanston, IL 60208, USA.

*Correspondence: sebastian.soyk@unil.ch

ABSTRACT

Crop genomes accumulated deleterious mutations, a symptom known as the cost of domestication. Precision genome editing has been proposed to eliminate such potentially harmful mutations, however, experimental demonstration is lacking. Here, we identified a deleterious mutation in the tomato transcription factor *SUPPRESSOR OF SP2* (*SSP2*), which became prevalent in the domesticated germplasm and diminished DNA-binding to genome-wide targets. We found that *SSP2* acts partially redundant with its paralog *SSP* to regulate shoot and inflorescence architecture. However, redundancy was compromised during tomato domestication and completely lost in the closely-related species *Physalis grisea*, in which a single ortholog regulates shoot branching. We applied base editing to directly repair the deleterious mutation in cultivated tomato and obtained plants with compact growth that provide an early fruit yield. Our work shows how deleterious variants sensitized modern genotypes for phenotypic tuning and illustrates how repairing deleterious mutations with genome editing allows for predictable crop improvement.

INTRODUCTION

Deleterious mutations lead to the alteration or loss of gene activity. Crop domestication has been accompanied by an accumulation of potentially deleterious mutations[1,2], a phenomenon described as the genetic cost of domestication[3]. Such potentially harmful variants likely influence many important agricultural traits[4]. For example, harmful recessive alleles can have detrimental effects that are exposed in homozygous progeny during inbreeding[5]. Deleterious mutations are often considered to mainly negatively affect fitness of natural populations but recently, a more nuanced view has been proposed that considers their adaptive value[6,7]. Deleterious, loss-of-function mutations may confer an evolutionary advantage during rapid shifts in environmental conditions and the selective pressures thereof[7]. Crop domestication created novel environments under which many traits that were beneficial in the wild likely became neutral or even detrimental. Illustrative examples include loss of photoperiodic flowering and seed shattering. These observations support the "less-is-more" idea, which proposes that selection may favor a less-than-complete repertoire of functional genes[7]. Nonetheless, eliminating deleterious variants from domesticated germplasm has been proposed as a major goal in future crop breeding to avert potential harmful effects[4,5]. However, correcting genetic variants by recombination during cross-breeding can be complicated by genetic linkage with beneficial alleles or near fixation in domesticated populations. Recent advances in precision genome editing promise to facilitate the repair of deleterious variants[8]. However, to our knowledge, an experimental demonstration of precision genome editing for the repair of deleterious variants in domesticated germplasm has been lacking.

A recurrent target of selection during crop domestication and breeding are alterations in flowering time[9]. Changes in flowering time allowed the adaptation of crops to novel environments and growing seasons different from their wild ancestors' origin. The floral transition also influences plant architecture by balancing vegetative and reproductive growth[10]. At the molecular level, flowering occurs when the universal flowering hormone, florigen, reaches a critical level that triggers stem cells in the shoot meristems to switch from vegetative to reproductive growth. In the model crop tomato (*Solanum lycopersicum*), florigen is encoded by *SINGLE FLOWER TRUSS* (*SFT*), a homolog of Arabidopsis *FLOWERING LOCUS T* (*FT*) and member of the *CENTRORADIALIS, TERMINATING FLOWER1*, *SELF-PRUNING* (*CETS*) gene family[11]. While *SFT* promotes the floral transition, *SELF PRUNING* (*SP*) acts as antiflorigen and opposes the activity of florigen to repress flowering[12]. Evidence from rice and Arabidopsis suggests that

florigen protein competes with antiflorigen for Group-A basic region/leucine zipper (bZIP) transcription factors to form the Florigen Activation Complex (FAC)[12–14]. In tomato, the bZIP transcription factor SUPPRESSOR OF SP (SSP) is a functional FAC component and *ssp* mutations have been used to fine-tune plant architecture for optimized fruit productivity[15]. In other crops, mutations in central florigen pathway components have been also selected to change flowering time and shoot architecture[9]. Yet, how deleterious mutations affected key components of the florigen pathway during crop domestication has not been systematically studied.

RESULTS

Prediction of deleterious variants in central components of the florigen pathway. To determine the mutational load in domesticated tomato, we generated a chromosome-scale genome assembly for the closely-related wild tomato species *S. pimpinellifolium* (accession LA1589) (see **Online Methods**). We used this wild tomato genome as a reference to identify nonsynonymous mutations across a collection of 82 genomes along the domestication history of tomato, including 27 wild tomato species (*S. pimpinellifolium*), 23 landrace (*S. lyc. var. cerasiforme*), and 32 domesticated (*S. lycopersicum*) genomes (**Fig. 1a**, **Table S1)**[8,16]. We predicted deleterious variants by amino acid conservation modelling and identified 39,132 (23.1 %) nonsynonymous variants with a putative deleterious effect (SIFT-score < 0.05) (**Fig. S1a, b, Table S2**)[17]. This analysis indicated that wild species, landrace, and domesticated tomato genomes contain on average 5,114, 7,131, and 8,233 homozygous deleterious variants, respectively (**Fig. S1c**). Next, we focused on core components of the FAC[14] and searched for deleterious variants in *CETS* and Group-A bZIP genes (**Fig. S2)**[18]. Among all 12 tomato *CETS* genes, we identified three genes with predicted deleterious variants (**Fig. 1b**). Besides two uncharacterized *TERMINATING FLOWER1* (*TFL1*)-like and *MOTHER OF FT* (*MFT*)-like genes, we found the known flowering repressor *SELF-PRUNING 5G* (*SP5G*; Solyc05g053850), which contained a predicted deleterious variant in 45 of the genomes (54.9%) (**Table S3**). We also detected the *sp-classic* breeding mutation (P76L) that was predicted to not be deleterious but tolerated, which supports a hypomorphic nature of the mutation[19]. Among all 13 tomato Group-A bZIP genes, we identified four uncharacterized abscisic acid responsive element binding factor (*ABF*)-like genes with predicted deleterious mutations (**Fig. 1c and Fig. S2**). The most frequent predicted deleterious variant affected the bZIP gene

Solyc02g061990 and was detected in 36 genomes (43.9%). We concluded from these analyses that several central florigen pathway components have acquired potentially deleterious mutations during tomato domestication.

A missense mutation in the transcription factor *SSP2* was enriched during domestication. A phylogenetic analysis comparing group-A bZIP proteins of tomato and Arabidopsis showed that Solyc02g061990 is most closely related to *SSP*, thus we named the gene *SSP2* (**Fig. 2a** and **Fig. S2**). SSP and SSP2 form a sister clade to the Arabidopsis proteins FD and FD PARALOG (FDP) [20], with SSP and FD being the more ancient genes. In Arabidopsis, FD and FDP are involved in flowering control and phytohormone responses[21,22]. Expression data from different tomato plant tissues showed that *SSP* and *SSP2* had similar expression patterns, suggesting functional redundancy, most notably in secondary (sympodial) shoot meristems (**Fig. 2b**)[23,24]. The putative deleterious variant in SSP2 causes a serine-to-phenylalanine (S169 to F169) exchange at a conserved residue in the DNA-binding domain (**Fig. 2c**). We analyzed the distribution of the ancestral ($SSP2^{S169}$) and domesticated ($SSP2^{F169}$) variants across 768 re-sequenced tomato accessions and found that the domesticated allele was absent from wild tomato species. The putative deleterious variant first arose in tomato landraces (*S. lycopersicum var. cerasiforme)*, was enriched in domesticated genotypes, and nearly fixed in modern fresh-market and processing types (**Fig. 2d)**. To genetically test if the putative deleterious variant has an effect on the floral transition, we introgressed the ancestral $SSP2^{S169}$ allele into a processing tomato type (cv. M82). We found that near-isogenic lines (NILs) harboring $SSP2^{S169}$ flowered earlier on sympodial shoots and developed shoots that grew more compact compared to the wild-type (WT) controls (**Fig. S3a-f**). In addition, we introduced $SSP2^{S169}$ into the hypomorphic ssp^{2129} mutant[15] to test whether $SSP2^{S169}$ acts redundantly with its paralog *SSP*. We found that $SSP2^{S169}$ suppressed late-flowering and indeterminate growth of ssp^{2129} mutants (**Fig. S3g, h**), suggesting that the ancestral $SSP2^{S169}$ allele can compensate for reduced *SSP* activity.

Domesticated $SSP2^{F169}$ is compromised in its function as a transcription factor. We hypothesized that the loss of the conserved serine residue affects the ability of SSP2 to bind DNA during the regulation of target genes. We modelled the structure of the ancestral ($SSP2^{S169}$) and domesticated ($SSP2^{F169}$) proteins in a homology-based modelling approach[25,26]. The model predicted that the conserved serine (S169) most likely forms hydrogen bonds with the phosphate backbone of the DNA target sequence whereas a phenylalanine at this position (F169) might

increase the distance between the protein and target DNA due to its larger side-chain and hydrophobicity (**Fig. 2e**). To test whether the amino acid exchange affects the transcription factor function of SSP2, we co-expressed $SSP2^{F169}$, $SSP2^{S169}$ and SSP with SFT in tobacco leaves to quantify their transactivation activity on the upstream regions of *MACROCALYX* (*MC*; Solyc05g056620), *S. lycopersicum FRUITFULL1* (*SlFUL1*, Solyc06g069430), and *SlFUL2* (Solyc03g114830). These genes are homologous to Arabidopsis *APETALA1* and *FRUITFULL*, which have been shown to be activated by FD during the floral transition[13]. None of the effector constructs activated the *MC* reporter, which may result from a non-direct relationship between *MC* and Arabidopsis *AP1*. However, the *SlFUL1* and *SlFUL2* reporters were significantly activated by both SSP and ancestral $SSP2^{S169}$ while the level of transactivation by $SSP2^{F169}$ was not significant (**Fig. 2f**). Together, these results suggest that the deleterious variant in *SSP2* disrupts the DNA-binding ability of domesticated $SSP2^{F169}$ and compromises its transcription factor function.

To determine how the deleterious $SSP2^{F169}$ variant affects binding at genome-wide targets, we performed DNA affinity purification sequencing (DAP-seq) with SSP, ancestral $SSP2^{S169}$ and domesticated $SSP2^{F169}$ as bait proteins[27]. We identified 14,091 DAP-seq peaks that were significantly enriched ($\log_2FC \geq 3$, $FDR \leq 0.01$) compared to the input controls (**Fig. 3a and Table S4**). The majority (7,388) of peaks were shared between SSP and the ancestral $SSP2^{S169}$ but only 1,285 peaks were also bound by domesticated $SSP2^{F169}$. We analyzed the genome-wide distribution of peaks for all three transcription factors and found more than 50% of peaks within proximal regulatory regions (**Fig. 3b**). *De-novo* motif enrichment analysis identified a G-box motif (CACGTG) with a subtle variation for $SSP2^{F169}$ outside the core-motif (**Fig. 3c**). Next, we analyzed genes with proximal peaks (≤ 3 Kbp upstream and ≤ 2 Kbp downstream) and identified 6,485 and 4,229 putative target genes for SSP and SSP^{S169}, of which the majority (3,953 genes) were bound by both proteins (**Fig. 3d** and **Table S5**). In contrast, domesticated $SSP2^{F169}$ bound only 984 and 952 of SSP and SSP^{S169} targets, respectively, and 1,377 genes in total. The low number of $SSP2^{F169}$ targets and shared targets with SSP and $SSP2^{S169}$ suggested that the ability of $SSP2^{F169}$ to bind its genome-wide targets is compromised. To support this finding, we quantified binding intensity at target regions based on normalized read coverage. While SSP and $SSP2^{S169}$ displayed similar binding intensities, $SSP2^{F169}$ binding was strongly reduced (**Fig. 3e, f** and **Fig. S4c, e**). Furthermore, diminished binding of $SSP2^{F169}$ at $SSP2^{S169}$ and SSP targets was also obvious at the level of individual genes. For example, we found that the upstream regions of the two tomato

homologs of *GIGANTEA* (*GI*), which regulates flowering in Arabidopsis[28], were bound by SSP and SSP2^{S169} but not by the domesticated SSP2^{F169} variant (**Fig. 3g, h**). Together, our genome-wide binding data demonstrates that SSP and the ancestral SSP2^{S169} variant bind a set of largely shared targets while domesticated SSP2^{F169} is compromised in its ability to bind the targets of the ancestral protein.

***SSP2* acts partially redundant with *SSP* to regulate shoot and inflorescence architecture.** To genetically explore the function of *SSP2*, we used CRISPR-Cas9 genome editing and generated *ssp2^{CR}* and *ssp^{CR}* null mutants in two determinate cultivars (**Fig. S5a-b)**. The *ssp^{CR}* mutants flowered later than the WT and developed indeterminate shoots, which confirmed previous findings that *SSP* promotes the floral transition (**Fig. 4a-c** and **S5c-e**)[15]. We did not observe obvious differences in flowering time for *ssp2^{CR}* single mutants, which supports a diminished activity of *SSP2^{F169}* in domesticated tomato (**Fig. 4a, c** and **S5c-d**). However, *ssp^{CR}ssp2^{CR}* double mutants tended to flower later than the *ssp^{CR}* single mutant, although at high variability (**Fig. 4c and S5c, d**). This phenotypic enhancement became more pronounced on sympodial shoots. Double *ssp^{CR}ssp2^{CR}* mutants produced more leaves on sympodial shoots and more flowers on flowering shoots (inflorescences). We concluded that domesticated *SSP2^{F169}* is a partial loss-of-function allele and that *SSP* and *SSP2* act partially redundant to promote the transition of meristems to reproductive growth (**Fig. 4d, e)**.

To obtain molecular insights into how *SSP* and *SSP2* promote meristem transitions, we sequenced mRNA from micro-dissected meristems at the transition (TM) stage of meristem maturation of the *ssp^{CR}* and *ssp2^{CR}* single and double mutants, and the WT (in cv. M82)[15]. Clustering of samples in a principal component analysis (PCA) was consistent with the mutant phenotypes that indicated a delayed transition of *ssp^{CR}ssp2^{CR}* double mutants compared to the *ssp^{CR}* single mutant (**Fig. 4f**). We identified 1,832 differentially expressed genes (DEGs) that changed in expression by more than 1.5-fold in at least one of the mutants compared to the WT (FDR $\leq$ 0.05) (**Fig. S5e-f**). Of those, 520 (28.6%) were nearby DAP-seq peaks, indicating that they are direct targets of SSP and/or SSP2 **(Fig. 4g)**. Clustering of the 520 putative direct targets revealed two main patterns of gene expression that contained genes either down- or upregulated (de-repressed) in the *ssp^{CR}ssp2^{CR}* double mutant (**Fig. 4h-i and Table S6**). Among the downregulated genes, we found both tomato

homologs of the Arabidopsis floral promoter *GI*, and a homolog of its interactor *FLAVIN-BINDING, KELCH REPEAT, F-BOX 1* (*FKF1*)[29]. In addition, the MADS-box gene *SlMBP10*, a homolog of the Arabidopsis floral promoter *FUL*, was downregulated in $ssp^{CR}ssp2^{CR}$ double mutants, while *SlMBP14* and a *FLOWERING LOCUS C* (*FLC*)-like gene were de-repressed in $ssp^{CR}ssp2^{CR}$. We also identified several putative direct targets involved in phytohormone signaling. Two cytokinin dehydrogenase/oxidase genes (*CKX1a*, *CKX5*) and putative negative regulators of cytokinin levels were downregulated while a cytokinin activating enzyme encoding *SlLONELY GUY1* (*SlLOG1*) gene was de-repressed in $ssp^{CR}ssp2^{CR}$. Furthermore, three abscisic acid receptor genes (*PYLs*) were downregulated in the $ssp^{CR}ssp2^{CR}$ double mutant. These data indicate that *SSP* and *SSP2* redundantly regulate the expression of central regulators of the floral transition and phytohormone responses, and guide meristem transitions towards floral fate.

***SSP2* was lost during the evolution of *Physalis grisea*.** To determine whether genetic redundancy between *SSP* and *SSP2* is evolutionary conserved, we inspected orthologs across eudicots (**Fig. S6**). Surprisingly, our phylogenetic analyses indicated that tomato *SSP*/*SSP2* and Arabidopsis *FD*/*FDP* resulted from independent duplication events in the *Solanacaeae* and *Brassicaceae* families (**Fig. S6**). When we inspected protein sequences of SSP-like transcription factors in the *Solanaceae*, we identified a missense mutation in a conserved residue in the DNA-binding domain of the potato SSP ortholog (**Fig. 2c)**. Furthermore, we found only one SSP-like ortholog in *Physalis grisea* (*PgSSP;* Phygri02g013770), a relative of tomato in the *Solanoideae* subfamily[30]. Phylogenetic and synteny analyses supported an evolutionary scenario in which the ortholog of *SSP2* was lost in *P. grisea* (**Fig. 5a and Fig. S7a-c**). To obtain experimental evidence for the loss of redundancy in *P. grisea*, we mutated *PgSSP* by CRISPR-Cas and quantified effects on shoot architecture (**Fig. 5b**). Wild-type *P. grisea* plants produce seven leaves on the primary shoot before terminating in a single-flowered inflorescence (**Fig. 5c**). Growth continues from two sympodial meristems that each produce one sympodial unit, which results in a bifurcation of the shoot. Each sympodial meristem produces two leaves and one flower, and in turn releases two additional sympodial shoots. We observed alterations to this pattern in two independent $Pgssp^{CR}$ mutant lines, which produced an additional sympodial shoot at the first bifurcation and grew less compact than the WT (**Fig. 5c, d-g**). The additional sympodial shoot on $Pgssp^{CR}$ mutants resulted from a sympodial meristem in the axil of an extra leaf that was produced before flowering, which

indicated that loss of *PgSSP* leads to a flowering delay (**Fig. 5e, h**). Together, these results suggest that *PgSSP* regulates the transition of primary and sympodial meristems in the paralog-free context of *Physalis* in which *SSP2* is dispensable.

Repairing *SSP2^{F169}* by base-editing in cultivated tomato leads to compact growth and earlier yield. Our findings in tomato show that *SSP2* acts partially redundant with *SSP* to promote the transition to flowering on sympodial shoots (**Fig. 4a, d**). We asked whether restoring the activity of *SSP2* in domesticated tomato by correcting the deleterious variant would accelerate the floral transition. We tested this hypothesis by repairing the deleterious variant in domesticated tomato by CRISPR-Cas base editing. The critical non-synonymous mutation results from a TCC (Ser) to TTC (Phe) codon exchange (**Fig. 6a**). The correction of this mutation requires a A-to-G transition on the reverse strand, which can be induced with an adenine base editor (ABE)[31]. Since none of the nearby canonical PAMs (NGG) allowed us to position the target nucleotide into the high-activity editing window (A4-A8) of the protospacer, we used the PAM-less Cas9 variant SpRY fused to ABE8e (**Fig. 6a**)[32]. We edited *SSP2* in the domesticated and double-determinate S100 background[33] and observed high editing efficiency with edits at the target adenine in 37.5 % (3 of 8) second-generation (T1) transgenic families (**Fig. 6b** and **Fig. S8a**). In one T1 family we also detected editing at the bystander T position (**Fig. S8a**). To determine whether the base-edited (be) *ssp2^{S169be}* allele affected flowering time and shoot architecture, we generated segregating (F4) populations and selected homozygous (*ssp2^{S169be}/ssp2^{S169be}*) and heterozygous (*ssp2^{S169be}/SSP^{F169}*) individuals for the repaired allele, and WT siblings (*SSP^{F169}/SSP^{F169}*) as controls by genotyping (**Fig. 6c**). We found that plants homozygous or heterozygous for the repaired *ssp2^{S169be}* allele did not flower earlier than their WT siblings (**Fig. 6d-e**). However, they developed less sympodial shoot units and less flowers per inflorescence compared to WT siblings homozygous for the domesticated (*SSP2^{F169}*) allele, which resulted in an overall more compact architecture (**Fig. 6d,f-g**). To assess if repair of *SSP2* could compensate for the loss of *SSP*, we introduced the repaired *ssp2^{S169be}* allele into the *ssp^{CR}* null mutant (in cv. S100). We found that *ssp2^{S169be}* did not suppress late flowering and indeterminate growth of *ssp^{CR}* (**Fig. 6h-j**). However, we observed a partial and significant suppression of late flowering on sympodial shoots (**Fig. 6h, k**). Moreover, *ssp^{CR} ssp2^{S169be}* plants developed shorter inflorescences compared to *ssp^{CR}* mutants and WT (*SSP2^{F169}*)

plants (**Fig. 6l**). Together, these results demonstrate that functional *SSP2* accelerates the reproductive transition of meristems on sympodial shoots in partial redundancy with *SSP*.

Tomato production was revolutionized during the 20th century by the *self-pruning* mutation, which confers determinate growth and facilitates mechanical harvesting. Our findings showed that a functional *SSP2* allele accelerates sympodial shoot flowering and thus suggested an agronomic value for this allele regarding earliness for yield. To test whether accelerated flowering from the repaired *ssp2*^{S169be} allele leads to earlier yield, we quantified fruit production in segregating (F4) populations under experimental greenhouse conditions (see **Online Methods**). We found that total fruit yields, harvest index, and fruit size for *ssp2*^{S169be} plants were comparable to the WT sibling controls (**Fig. 6n-o** and **Fig. S8b-e**). However, *ssp2*^{S169be} fruits had a reduced sugar content (brix) (by 11%) (**Fig. S8f**). Notably, *ssp2*^{S169be} homozygotes displayed an 8% increase in the proportion of ripe fruits compared to WT siblings, which was likely due to precocious flowering and termination of sympodial shoots (**Fig. 6m, p**). Thus, compact growth from repairing the deleterious *SSP2* mutation by base editing can confer earliness for fruit yields and represents a promising new target for customizing tomato shoot architecture.

DISCUSSION

Here, we investigated the load of deleterious mutations that accumulated during domestication and improvement of tomato. Within genes central to flowering time control, we discovered a deleterious variant in the previously uncharacterized bZIP transcription factor gene *SSP2*. The deleterious variant results in the exchange of a conserved serine to a phenylalanine in the DNA-binding domain of the transcription factor. Our results from structural modelling, genome-wide DNA binding assays, and genetic analyses indicate that the domesticated SSP2^{F169} variant partially lost its ability to bind and regulate target genes that are largely shared between the ancestral SSP2^{S169} variant and its paralog SSP. However, we cannot fully rule out that domesticated SSP2^{F169} nonfunctionalized given its 353 private target genes and a subtle variation near the G-box target motif. Interestingly, in the yeast bZIP factor Pap1, the equivalent serine-to-phenylalanine exchange contributes to a similar change in binding specificity[34]. Nevertheless, our data shows that the deleterious variant in *SSP2* led to loss of genetic redundancy between *SSP* and *SSP2*, a pair of paralogs that is widely conserved in flowering plants. In Arabidopsis, it was shown

that *FD* and *FDP* act redundantly during phytohormone responses while only *FD* affects the floral transition, suggesting functional divergence of *FD*[21]. In contrast, our findings in tomato indicate that *SSP* and *SSP2* act also partially redundant during the floral transition. Notably, our phylogenetic analyses suggest that paralogs of *SSP* and *FD* arose independently in *Solanaceae* and *Brassicaceae*, which could explain species-specific divergence of this paralogous pair. The complete loss of a *PgSSP* paralog in *Physalis grisea* further supports dynamic evolution of the paralog pair. Deleterious mutations and gene loss have been proposed as an important mechanism of adaptation[6,7]. However, the benefit of the deleterious *SSP2*^{F169} variant during domestication remain speculative. Our genetic data demonstrates that domesticated *SSP2*^{F169} delays meristem transitions on shoots and inflorescences. Notably, the domesticated *SSP2*^{F169} genotype develops more flowers per inflorescence than the ancestral *SSP2*^{S169} genotype. Although flower number correlates with fruit yield, the number of flowers per inflorescence in general decreased during tomato domestication, likely due to source-sink imbalances driven by dramatic increases in fruit size[35]. This overall decrease in flower number during tomato domestication suggests that effects from *SSP2*^{F169} on flower number were rather minor and difficult to select. Furthermore, the deleterious *SSP2*^{F169} variant could have hitchhiked near QTLs that were selected during domestication and improvement, which is a common scenario in crops with a narrow genetic base such as tomato[8]. However, the closest known improvement sweep on chromosome 2 with five fruit-weight QTLs is more than 5 Mbp away from *SSP2*, rendering linkage unlikely[36]. Finally, we cannot exclude that *SSP2*^{F169} is adaptive under specific conditions that were absent from our experiments. Whether *SSP2*^{F169} was nearly fixed in cultivated tomato due to selection or drift remains therefore an open question. Yet, the loss of genetic redundancy caused by a deleterious mutation may reflect a common feature during the selection of crops in human-made environments. The less-is-more idea proposes the accumulation of loss-of-function mutations as a driver of rapid evolutionary change[7], and gene loss may be even more frequent during the intense artificial selection in domesticated environments. A reduced genetic repertoire in domesticated genomes could result in lower genetic redundancy compared to their ancestral states and, as a consequence, facilitate the exposure and selection of novel mutations, which are otherwise masked by redundant paralogs. Our data shows that the ancestral *SSP2*^{S169} allele can suppress effects of *ssp* mutations, which allow tuning of shoot architecture and optimization of tomato yields[15]. Intriguingly, the deleterious *SSP2*^{F169} mutation, which broke redundancy with the paralog *SSP*,

may have been a prerequisite for the identification of the *ssp*^{2129} breeding mutation. This illustrates how standing variants can become adaptive due to genetic interactions with mutations that are introduced or arose during breeding.

Correcting deleterious variants with genome editing in crops has been proposed as major strategy for future crop breeding[4]. To our knowledge, we present here the first example of a direct repair of a deleterious mutation in a crop using base editing. We show that repairing the deleterious *SSP2* variant in tomato leads to precocious flowering on sympodial shoots and an overall more compact plant architecture. Notably, precocious flowering and compact growth of base-edited plants was associated with earliness for yield, with repaired plants displaying an 8% increase in ripe fruits at harvest. Such earliness for fruit yield is a highly desirable trait for customizing shoot architecture for specific environments. Our work shows that base editing provides a promising approach for correcting deleterious variants that accumulated during domestication and improvement in crops. However, our study also emphasizes that deleterious mutations are not unfavorable *per se* and may have adaptive roles that are only exposed in specific genetic backgrounds or environmental conditions.

FIGURE LEGENDS

Figure 1: Predicting the load of deleterious variants along the domestication history of tomato. a, Number of predicted deleterious mutations in a panel of 82 tomato genomes, including wild species (*S. pimpinellifolium*, green), landraces (*S. lycopersicum var. cerasiforme*, orange), and cultivars (*S. lycopersicum*, purple). **b-c**, Prediction of deleterious variants across all CETS (b) and Group-A bZIP (c) genes. The dashed red line indicates the threshold for deleterious prediction (SIFT-score<0.05). Dot size scales with the number of genomes that carry the variant. Red font indicates genes with predicted deleterious mutations.

Figure 2: A deleterious mutation in *SSP2* reduces its transcription factor activity. a, Maximum-likelihood tree of A-group bZIP proteins in tomato (red font) and Arabidopsis (blue font). Red arrowhead marks SSP2. Numbers represent bootstrap values from 1,000 replicates and scale bar indicates the average number of substitutions per site. **b**, Normalized gene expression (TPM) for *SSP* and *SSP2* in different tissues and developmental stages (veg. earl./mid./late, stand for early, middle and late vegetative meristem stage). **c**, Partial alignment of SSP-like bZIP proteins from Arabidopsis, domesticated tomato (*S. lycopersicum*; *Slyc*), close wild tomato relative (*S. pimpinellifolium*; *Spim*), distant wild tomato relative (*S. pennellii*; *Spen*), potato (*S. tuberosum*; *St*), and *Physalis grisea* (*Pg*). Red arrowheads mark conserved DNA-binding residues. **d**, Distribution of ancestral (*SSP2*^{S169}) and derived (*SSP2*^{F169}) *SSP2* alleles in distant wild tomato relatives, wild relatives (*S. galapagense* / *S. cheesmaniae*), wild progenitor species (*S. pimpinellifolium*), landraces (*S. lyc var. cerasiforme*), and cultivars (*S. lycopersicum*). n=number of accessions. **e**, Predicted structures of ancestral SSP2^{S169} and derived SSP2^{F169} proteins on target DNA determined by homology modelling. Insets show a magnified view of the serine/phenylalanine residue at position 169. **f**, Reporter assays in tobacco leaves using SSP, SSP2^{F169}, and SSP2^{S169} as effectors and firefly Luciferase (fLuc) driven by upstream sequences of *MC* (*pMC::fLUC*), *SlFUL1* (*pSlFUL1::fLUC*), and *SlFUL2* (*pSlFUL1::fLUC*) as reporter. Numbers indicate technical replicates. Ctrl indicates no effector control. Letters represent post-hoc Tukey's HSD tests results with 95% confidence level.

Figure 3: Domesticated SSP2^{F169} shows reduced binding at genome-wide target loci. a, Overlap of significant ($log_2FC \geq 3$, FDR ≤ 0.01) SSP, SSP2^{F169}, and SSP2^{S169} DAP-seq peaks (n=14'091). **b**, Distribution of significant SSP, SSP2^{F169}, and SSP2^{S169} DAP-seq peaks across gene features. **c**, Most-significant motifs identified by *de-novo* motif enrichment analysis of SSP, SSP2^{F169}, and SSP2^{S169} DAP-seq peak regions. Grey box delimits region with motif variation outside the core-motif. **d**, Overlap of genes with significant DAP-seq peaks ≤ 3 Kbp upstream and ≤ 2 Kbp downstream of the transcriptional start site (n=7'114). **e**, Profiles of normalized read coverage at significant SSP, SSP2^{F169}, and SSP2^{S169} peaks. **f**, Comparison of SSP, SSP2^{F169}, and SSP2^{S169} DAP-seq peaks relative to the transcriptional start (TSS) and end (TES) site of nearby genes (n=7'114). **g-h**, Browser view of SSP, SSP2^{F169}, and SSP2^{S169} DAP-seq peaks at *SlGIGANTEA-LIKE1* (g) and *SlGIGANTEA-LIKE2* (h). Normalized coverage (CPM) is shown in yellow, green and blue. Significant peak regions are indicated by red boxes.

Figure 4: *SSP* and *SSP2* act partially redundant to regulate the transition to flowering. a, Representative images of wild-type S100, *ssp^{CR}* and *ssp2^{CR}* single mutants, and *ssp ssp2^{CR}* double mutants. L= leaf number, arrowheads mark the last leaf before flowering. Determinate (D) and indeterminate (ID) shoots are indicated. Scale bars represent 7.5 cm. **b**, Schematic depiction of tomato shoot architecture. Different shades of green delimit primary and sympodial shoots. **c-e**, Quantification of the floral transition (number of leaves before flowering) on the primary (c) and secondary (d) shoots, and the number of flowers per inflorescence (e) for genotypes shown in (a). The number of plants (c,d) and inflorescences (e) are indicated. Letters represent post-hoc Tukey's HSD tests results with 95% confidence level. **f**, Principal component analysis of 22'726 expressed genes in transition meristems of the WT, *ssp*, *ssp2*, and *ssp ssp2*, determined by RNA-seq. **g**, Overlap of genes differentially expressed ($log_2FC \geq 0.58$, FDR ≤ 0.05) in *ssp*, *ssp*2, and/or *ssp ssp2* with genes at SSP, SSP2^{F169}, and SSP2^{S169} DAP-seq peaks. **h**, Heatmap depicting expression of 520 putative SSP/SSP2 target genes. **i**, Normalized expression levels for selected putative direct targets. Genes are color coded based on the biological pathway.

Figure 5: The genome of *Physalis grisea* encodes a single direct *SSP* ortholog that regulates meristem transitions. a, Scheme of the phylogenetic tree of tomato and closely related

Solanaceae species. Filled circles, empty circles or star show presence, absence, or missense mutation, respectively, of *SSP*/*SSP2* or *FD/ FDP* in these species. Full tree is displayed in Fig. S6. **b**, CRISPR-Cas9 targeting of *PgSSP* in *P. grisea*. Blue boxes, black lines, and grey boxes represent exonic, intronic, and untranslated regions, respectively. Single guide RNAs (sgRNAs) are indicated with red arrowheads. PAM and sgRNA sequences are indicated in black and red bold letters, respectively; deletions are indicated with blue dashes; sequence gap length is given in parenthesis. Insertions are indicated by blue letters. **c**, Model of the growth habit of *P. grisea* WT and $Pgssp^{CR}$ plants. Different shades of green delimit primary, first sympodial, and second sympodial shoots. The color of leaves corresponds with the shoot of origin. Note that the last leaf of each shoot is displaced upwards during shoot development. **d**, Representative pictures illustrating the difference in number of sympodial shoots in WT and *Pgssp* mutant plants. Last leaf before the shoot bifurcation is indicated (L5). White arrowheads indicate individual sympodial shoots. Scale bar represents 7.5 cm. **e**, Representative stereoscope images of the shoot apex of WT and *Pgssp* mutant plants. Upper images show the apex with a terminal flower (*). Lower images show the same view with the flower removed. The sympodial meristems (SYMs) are delimited by a dashed line and numbered in developmental order. Scale bar represents 100 µm. **f-h**, Quantification of the number of sympodial shoots at the first and second bifurcation, and flowering time (number of leaves before the first inflorescence). Number of plants is indicated at the bottom of the plots. Letters represent post-hoc Tukey's HSD tests results with 95% confidence level.

Figure 6: Repairing the deleterious *SSP2* mutation in domesticated tomato by base-editing leads to compact growth and earliness for yield. a, Base-editing strategy to correct the deleterious *SSP2* mutation in domesticated tomato using an adenosine base editor (ABE) and a PAM-less Cas9 variant. The target adenine in *SSP2* (A5) is at position 5 of the protospacer with a bystander adenine (A6) at position 6. Editing of the target codon (TTC) can lead to three different outcomes depending on which adenine is deaminated. Only editing the target nucleotide (A5) alone reverts the phenylalanine codon (TTC) back to the ancestral serine (TCC). **b**, Validation of editing in a chimeric first-generation (T0) transgenic and the corresponding T1 progeny by Sanger sequencing. The target nucleotide is indicated by a red arrowhead. **c**, Crossing scheme to generate the segregating $ssp2^{S169be}$ F4 population. **d**, Representative pictures showing the total number of sympodial units on WT and $ssp2^{S169be}$ plants. Terminal inflorescences of each sympodial unit are

indicated by a white arrow. **e-g**, Quantification of flowering time (number of leaves before the first inflorescence), number of sympodial shoots, and number of flowers per truss of WT, *ssp2*^{S169be}/+ and *ssp2*^{S169be} plants. **h**, Representative pictures showing the number of leaves per sympodial unit and determinacy of WT, *ssp*^{CR}, *ssp*^{CR} *ssp2*^{S169be}/+ and *ssp*^{CR} *ssp2*^{S169be} plants. **i-l**, Quantification of flowering time (as in (e)), number of determinate plants, number of leaves per sympodial unit (SU), and number of flowers per truss of WT, *ssp*^{CR}, *ssp*^{CR} *ssp2*^{S169be}/+ and *ssp*^{CR} *ssp2*^{S169be} plants. Determinate (D) and indeterminate (ID) shoots are indicated. **m**, Representative images showing the full harvest of individual WT, *ssp2*^{S169be}/+ and *ssp2*^{S169be} plants. Percentage of red fruits is indicated. **n-p**, Quantification of total fruit yield (n), harvest index (total fruit yield / plant weight) (o), and percentage of red fruits. Number of plants are indicated in the plots for (e-g), (i-k) and (l-o). Letters on top of the plots represent post-hoc Tukey's HSD tests results with 95% confidence level**.** Scale bars represent 10 cm (d) and 7.5 cm in (h,m).

SUPPLEMENTARY FIGURE LEGENDS

Figure S1: Prediction of deleterious variants in tomato. a, Number of coding sequence variants across a panel of 82 genomes. **b**, Number of non-synonymous variants predicted to be tolerated (sift-score ≥ 0.05), deleterious (sift-score < 0.05), or without prediction (na). Color code indicates confidence of SIFT prediction. **c**, Number of heterozygous and homozygous predicted deleterious mutations in wild (*S. pimpinellifolium*, n=27, in green), landrace (*S. lyc. var. cerasiforme*, n=23, in orange), and domesticated (*S. lycpersicum*, n=32, in purple) tomato genomes.

Figure S2: Phylogenetic analysis of the bZIP transcription factor family in Arabidopsis and tomato. Maximum-likelihood phylogenetic tree constructed with full-length bZIP protein sequences from Arabidopsis (n=74) and tomato (n=70). Arabidopsis and tomato proteins are indicated in black and red font, respectively. The yeast protein Pap1 was used as an outgroup (blue font). Proteins were classified into 13 groups (A-K, M, S) according to the Arabidopsis nomenclature [37]. Numbers represent bootstrap values from 1000 replicates, and scale bar indicates the average number of substitutions per site.

Figure S3: Introgression of ancestral *SSP2^{S169}* into domesticated tomato suppresses late flowering and indeterminate growth of *ssp* mutants. a, Representative image of greenhouse-grown wild-type (WT) and *SSP2^{S169}*-NIL individual in the determinate M82 background. **b-d**, Quantification of the floral transition (the number of leaves before flowering) on primary (b) and sympodial shoots (c), and the number of sympodial shoot units (d). **e, f,** Representative images of field-grown WT and *SSP2^{S169}*-NIL plants at flowering (c) and fruiting (d) stage. **g**, Representative images of detached WT, *ssp^{2129}* and *ssp^{2129} SSP2^{S169}*-NIL shoots (in the determinate M82 background). D, determinate; ID, indeterminate; L, leaves. **h**, Quantification of the floral transition on the primary shoot for genotypes shown in (e). Numbers at the bottom and letters at the top of the plots of (b) and (f) represent the number of replicate plants and post hoc Tukey's HSD test results with 95% confidence level, respectively. Scale bars indicate 10 cm (a, e, f) and 1 cm (g).

Figure S4: Identification of SSP, SSP2^{F169}, and SSP2^{S169} genome-wide binding sites by DAP-seq. a, Overlap of SSP, SSP2^{F169}, and SSP2^{S169} DAP-seq peaks at different significant thresholds ($log_2FC \geq 2, 3, 4$). **b**, Distribution of SSP, SSP2^{F169}, and SSP2^{S169} DAP-seq peaks across gene features at different significant thresholds as in (a) **c**, Profiles of normalized read coverage at SSP, SSP2^{F169}, and SSP2^{S169} peaks at different significant thresholds as in (a). **d**, Overlap of genes with DAP-seq peaks ≤ 3 Kbp upstream and ≤ 2 Kbp downstream of the transcriptional start site, at different significant thresholds as in (a). **e**, Comparison of SSP, SSP2^{F169}, and SSP2^{S169} DAP-seq peaks relative to the transcriptional start (TSS) and end (TES) site of nearby genes, at different significant thresholds as in (a). Top and bottom panels show coverage profiles and heatmaps, respectively.

Figure S5: Targeting *SSP* and *SSP2* in two tomato cultivars by CRISPR-Cas9. a,b CRISPR-Cas9 targeting of *SSP* and *SSP2* in *S. lycopersicum* cv. S100 (a) and cv. M82 (b). Orange boxes, black lines, and grey boxes represent exonic, intronic, and untranslated regions, respectively. Single guide RNAs (sgRNAs) are indicated with red arrowheads. PAM and protospacer sequences are indicated in black and red bold letters, respectively; deletions are indicated with blue dashes; sequence gap length is given in parenthesis. **c**, Representative images WT S100, *ssp^{CR}* and *ssp2^{CR}* single mutants, and *ssp ssp2^{CR}* double mutants. L= leaf number, white arrowheads mark inflorescences. Determinate (D) and indeterminate (ID) shoots are indicated. Scale bars represents 1 cm. **d**, Quantification of the floral transition on the primary shoot for genotypes in (c). N, number of plants. Letters represent post hoc Tukey's HSD tests. **e**, Volcano plots showing differentially expressed genes (log_2 FC > 0.58, FDR < 0.05) in *ssp^{CR}* and *ssp2^{CR}* single mutants, and *ssp ssp2^{CR}* double mutants compared to WT (cv. M82). **f**, Heatmap of z-scores showing expression pattern for 1'832 genes that are differentially expressed (log_2 FC > 0.58, FDR < 0.05) in *ssp^{CR}*, *ssp2^{CR}* single mutants, and/or *ssp ssp2^{CR}* double mutants in M82.

Figure S6: Phylogenetic analysis of SSP homologs in eudicots. Maximum-likelihood phylogenetic tree constructed with 128 full-length bZIP protein sequences from 51 eudicot species. Tomato, Arabidopsis, and Physalis proteins are highlighted in red, blue, and orange font, respectively. Red branches indicate duplication events, and the two separate duplication events in

the *Solanaceae* and *Brassicaceae* are highlighted with stars. Numbers represent bootstrap values from 1000 replicates, and scale bar indicates the average number of substitutions per site.

Figure S7: The ortholog of *SSP2* in *Physalis grisea* was lost during evolution. a. Maximum-likelihood phylogenic tree of the group A bZIP transcription factor family of *A. thaliana*, *S. lycopersicum* and *P. grisea*. Numbers represent bootstrap values from 1000 replicates, and scale bar indicates the average number of substitutions per site. **b,c**, Browser view of synteny analysis of *SSP* (b) and *SSP2* (c) between tomato (cv. S100) and *P. grisea*. Yellow rectangles show annotated genes and yellow streaks link them with their syntenic counterpart. *SSP* and *SSP2* genes are indicated in red. Note the lack of a unique syntenic block for *SSP2* in *P. grisea* in (c).

Figure S8: Base-editing of *SSP2* in domesticated tomato and its effect on different tomato yield components. a, CRISPR base-editing sequencing result of three T0 individuals (upper row) and their T1 progeny (lower row). Note that the target edit was detected in only one T0 individual (T0-3) but in three T1 families. One T1 individual (T1-9-17) was also edited at the bystander adenine. The edited nucleotides are indicated by a red arrowhead. **b-f**, Quantification of the vegetative biomass (b), total red and green fruit harvest (c,d), average fruit weight (e), and average soluble sugar content (brix) (f). The number of plants are indicated in the plots. Letters on top of the plots represent post-hoc Tukey's HSD tests results with 95% confidence level.

SUPPLEMENTARY TABLES

Table S1: List of accessions for deleterious variant analyses

Table S2: Number of predicted deleterious variants

Table S3: SIFT-score predictions for non-synonymous variants in CETS and Group-A bZIP genes

Table S4: List of significant SSP and SSP2 DAP-seq peaks

Table S5: List of genes associated with significant SSP and SSP2 DAP-seq peaks

Table S6: List of putative SSP/SSP2 target genes

Table S7: Assembly statistics

Table S8: List of primers used in this study

Table S9: List of gRNA sequences used in this study

ONLINE METHODS

Plant material, growth conditions, and phenotyping

Seeds of *S. lycopersicum* cv. M82 (LA3475), *S. lycopersicum* cv. Sweet-100 (S100) double-determinate[33], *S. pimpinellifolium* (LA1589), *P. grisea,* and *N. benthamiana* were from our own stocks. Tomato seeds were directly sown and germinated in soil in 96-cell plastic flats. The *P.grisea* seeds were incubated at 48°C for 3 days prior to sowing to increase germination rates. Plants were grown under long-day conditions (16-h light/ 8-h dark) in a greenhouse under natural light supplemented with artificial light from high-pressure sodium bulbs (~250umol m-2s-1). Temperature was 25°C and relative humidity was 50-60%. Plants were grown in 5L pots (2 plants per pot) under drip irrigation and standard fertilizer regimes. Tomato plants were pruned and only the primary shoot and the proximal axillary shoot were kept. Phenotypic data was collected from the F3 and T4 generation for ssp^{CR} $ssp2^{CR}$ plants in the S100 background, the F7 (ssp^{CR} and $ssp2^{CR}$) and F4 (ssp^{CR} $ssp2^{CR}$) generation in the M82 background, and the T3 generation for $PgSSP^{CR}$ in *Physalis* and the F4 generation in *ssp* $ssp2^{S169be}$ plants. Data for flowering time, sympodial shoot number, per sympodial shoot, and number of flowers per inflorescence were collected from the primary shoot and the proximal shoot. To assess different tomato yield components under experimental greenhouse conditions, mature plants were harvested 79 days after transplanting. For data collection, plants and fruits were manually removed from the soil and the plant, respectively. The total fruit yield was defined as the sum of red and green fruits from each plant. The harvest index was calculated by dividing the total fruit yield by the plant weight (i.e., the vegetative biomass after the removal of fruits). Ten fruits from each plant were randomly selected to measure average fruit weight and total soluble sugar content (brix) in fruit juice. Brix was quantified using a digital Brix refractometer (HANNA® instruments, HI96801). All statistical analyses of phenotyping data were conducted in R[38].

N. benthamiana (tobacco) seeds were directly sown on soil in square pots. Seedlings were grown under long-day conditions (16-h light/ 8-h dark) in a plant growth room under LED light panels (~100umol m-2s-1) and constant temperature (22°C). Approximately one week after germination, tobacco seedlings were singled out into individual square pots and grown for an additional 2-3 weeks before leaf infiltration.

LA1589 de novo genome assembly

Nanopore long read sequences for the *S. pimpinellifolium* accession LA1589 were previously generated[16,39]. Basecalling was performed using Guppy v3.1.5. Illumina sequencing data were previously generated[24]. We assembled the Nanopore and Illumina sequences together with MaSuRCA (v3.4.1)[40]. The resulting contigs were then scaffolded against the Heinz 4.0 reference genome using RaGOO (v1.1)[41]. Gaps were closed with LR_Gapcloser (v3)[42] and the assembly was polished with 3 rounds of Pilon (v1.23)[43]. Assembly statistics can be found in **Table S7**. We used liftoff[44] to annotate the LA1589 assembly with ITAG4.0 gene models and tomato pan-genome genes as previously described[33].

Genome-wide prediction of deleterious variants

Illumina raw reads from 27 *S. pimpinellifolium* and 28 *S. lycopersicum* accessions (**Table S1**) were retrieved from public repositories as described before[45] (Gao et al). Reads were aligned to the *S. pimpinellifolium* reference genome (LA1589v0.1) using BWA-MEM (v0.7.17) using default parameters. Alignments were sorted and duplicates marked with PicardTools (v2.26.2) and indexed using samtools (v1.15.1)[46]. Variants were called with bcftools (v.1.15.1, parameters mpileup --no-BAQ --ignore-RG -d 1000000 -Q0 --annotate FORMAT/AD,FORMAT/DP). Variants were filtered with vcftools (v0.1.14, parameters --min-alleles 2 --max-alleles 2 --minQ 30 --minDP 5 --maxDP 50 --mac 2 --recode --recode-INFO-all). Filtered variant call format (vcf) files were then used to predict deleterious mutations using SIFT-4G[17]. A custom SIFT library was built from the *S. pimpinellifolium* reference genome sequence (SpimLA1589_v0.1) and annotation (SolpimLA1589_v0.2) using the SIFT instructions and default parameters. The LA1589 SIFT library contained SIFT scores for 70% of genes (21578 of 30808), SIFT scores for 83% of positions (56424493/67919880), and confident scores for 73% of positions (41083097/56424493). SIFT was used to determine the effect of coding sequence variants on protein sequence, and to predict deleterious missense variants. Variant types and SIFT scores were plotted in R using the ggplot2 package.

Phylogenetic analyses and sequence alignments

Protein sequences of tomato and Arabidopsis bZIP family members were obtained from the Plant Transcription Factor Database (PlantTFDB, v5.0)[47]. Physalis bZIP protein sequences were identified in a BLAST search on the Phygri1.3.1 protein annotation[30] using the SSP protein sequence as query. Full-length amino acid sequences of 70 tomato, 74 Arabidopsis, 58 Physalis, and yeast Pap1 (SPAC1783.07c.1) bZIP proteins were aligned using MAFFT (v7.481) using default parameters[48]. Maximum likelihood phylogenetic trees were constructed in IQ-Tree (v2.2.0.5; parameters -m MFP -bb 1000 -bnni -redo)[49] and visualized in FigTree (v1.4.4; http://tree.bio.ed.ac.uk/software/figtree/). Average number of substitutions per site are indicated by the scale bars. Specific bZIP groups were assigned according to their Arabidopsis homologs[37].

To reconstruct the phylogenetic tree of the bZIP family in eudicots we used the OMA browser's[50] July 2023 release to collect a pool of homologs for tree building. The Hierarchical Orthologous Groups (HOGs) were identified by searching for the tomato SSP gene's identifier (Solyc02g083520) for the initial HOG and then adding additional closely related HOGs, inferred to be closely related as they share many predicted orthologs. The following HOGs were downloaded: D0228852, D0178917, D0181214, D0210160, D0214417, D0216285, D0223413 (accessed 23 Jan 2024). Additionally, through BLAST searches, we incorporated the bZIP gene of *Amborella trichopoda* and closely related bZIP genes from eight *Solanaceae* species: *Nicotiana benthamiana*, *Nicotiana tabacum*, *Phylloscopus griseolus*, *Petunia axillaris*, *Petunia inflata*, *Solanum tuberosum*, *Capsicum annuum*, and *Capsicum chinense*. The final dataset comprised 128 genes from 51 plant species. These protein sequences were aligned using the approach described in the PhylomeDB pipeline[51]. Briefly, we obtained alignments in forward and reverse directions using three programs (MUSCLE v3.8.1551[52], MAFFT v7.490[48], and Kalign v3.3.5[53]). Then, the six alignments were combined using M-COFFEE v13.46.0.919e8c6b[54]. The phylogenetic tree was reconstructed using a maximum likelihood approach as implemented in IQ-TREE v2.2.2.6[55], using the best-fit model identified by ModelFinder[56] (JTT+F+I+R5) and 1000 ultrafast bootstrap replicates. The tree was manually rooted using Amborella trichopoda as the outgroup. Duplication events were inferred using ETE v4.0[57] using the species overlap method[58].

Homology modelling

The HHpred server was used to find suitable templates for SSP2 protein modeling[59]. The final templates were chosen based on the sequence similarity in the area of protein-DNA interaction, not on the highest sequence identity to the target.

The 50 homology models of wild tomato protein $SSP2^{S169}$ dimers were calculated using Modeller 9v18[25] and CCAAT/enhancer-binding protein beta (C/EBP beta) as a template. The crystal structure of human C/EBP beta in complex with DNA is stored under 1HJB code in the Protein Data Bank[26]. The target and template sequence shared 26% of sequence identity. The best model in term of its DOPE score[60] was chosen.

Analogically, the 50 homology models of domestic tomato $SSP2^{F169}$ protein dimers were calculated based on the structure of Pap1 transcription factor as a template and the best model, according to DOPE score, was chosen. The crystal structure of Pap1 factor is stored in the PDB under 1GD2 code and shares 24% of sequence identity with the SSP2F169 protein[34]. For both SSP2 proteins the DNA molecule from the template structure was included in the models. The DNA sequence was changed to the SSP2 recognition motif with UCSF Chimera tool that was also used for visualization of the models[61].

Molecular cloning

Binary vectors for CRISPR-Cas9 mutagenesis in domesticated tomato were assembled using the Golden Gate cloning system as previously described[33,62]. For CRISPR-Cas9 mutagenesis in *S. pimpinellifolium* and Sweet-100, a new Level (L) 1 part pICH47742_SpCas9-P2A-GFP was cloned by amplifying the coding sequence of SpCas9 from pICH47742::35S::Cas9 (Addgene no. 49771) using primers P94 and P129. The fragments were cloned into the L0 acceptor pAGM1287 to generate pAGM1287-SpCas9. P2A-GFP was amplified from pGG-D-P2A-GFP-NLS-E[63] using primer P96 and P97 and cloned into the L0 acceptor pAGM1301 to generate pAGM1301_P2A-GFP. The pAGM1287_SpCas9 and pAGM1301_P2A-GFP parts were combined with pICH51288 (2Xp35S) and pICH41421 (nosT) in pICH47742 (L1 acceptor) to generate pICH47742_SpCas9-P2A-GFP. For CRISPR-Cas base editing, the PAM-less adenosine base editor ABE8e-SpRY[32] was domesticated by amplifying four fragments using the primer pairs P576/ P577, P578/ P579, P580/ P581, P582/P583 on the template pYPQ262B[32]. Fragments were cloned into the L-1 acceptor pAGM1311 and combined in the L0 acceptor pAGM1287 to generate

pAGM1287_ABE8e-SpRY. pAGM1287_ABE8e-SpRY was combined with pAGM1301_P2A-GFP, pICH51288 (2Xp35S), and pICH41421 (nosT) in the L1 acceptor pICH47742 to generate pICH47742_SpRY-ABE8e-P2A-GFP. Constructs for transactivation assays were cloned using the Golden Gate MoClo kit[62]. The p19 construct for silencing suppression was assembled with the L1 acceptor pICH47742 and the L0 parts pICH85281 (pMas), pICH44022 (p19), and pICH77901 (tMas). The YFP construct was assembled with the L1 acceptor pICH47742 and the L0 parts pICH51266 (p35S), pICSL80014 (YFP), and pICH41414 (t35S). To clone the SFT co-effector and the SlycSSP2 effector constructs, the coding sequences of SFT and SlycSSP2 were amplified from *S. lycopersicum* (cv. M82) transition meristem cDNA with gene specific primer pairs (SFT: SFT_F/SFT_R, SlycSSP2: SSP2_F/SSP2_R). To clone the SpimSSP2 effector construct, the coding sequence of SpimSSP2 was amplified from *S. pimpinellifolium* (LA1589) transition meristem cDNA with the primer pair SSP2_F/SSP2_R. The amplicons were cloned into the L0 acceptor pICH41308. To clone SSP effector construct, the coding sequences of SSP2 was amplified from *S. lycopersicum* (cv. M82) transition meristem cDNA in two fragments with the primer pairs SSP_F1/SSP_R1 and SSP_F2/SSP_R2 and cloned into the L-1 acceptor pAGM1311. The L-1 parts were cloned into the L0 acceptor pICH41308. Individual L0 effector parts (SSP, SlycSSP2, and SpimSSP2) were combined with pICSL13001 (p35S), pICSL30009 (Myc-tag), and pICH41414 (t35S) in the L1 acceptor pICH47772. The L0 co-effector part (SFT) was combined with pICSL13001 (p35S), pICSL30008 (HA-tag) and pICH41414 (t35S) in the L1 acceptor pICH47761. To clone the luciferase reporter constructs, the upstream regions of pMC, pFUL, and pFUL2 were amplified from *S. lycopersicum* (cv. M82) gDNA in multiple fragments gene-specific primer pairs (pMC: pMC_F1/pMC_R1 and pMC_F2/pMC_R2; pFUL1: pFUL1_F3/pFUL1_R3 pFUL1_F2p/FUL1_R2p, and FUL1_F1/pFUL1_R1; pFUL2: pFUL2_F1/pFUL2_R1 and pFUL2_F2/pFUL2_R2) and cloned into the L-1 acceptor pAGM1311. The pMC construct contained 2170 bp genomic sequence including upstream region, the 5'UTR, and the first exon. The pFUL1 and pFUL2 constructs contained 2640 bp and 2040 bp genomic sequence, respectively, including upstream regions and the 5'UTR. The L-1 parts were cloned into the L0 acceptor pICH41295. Individual L0 effector parts (pMC, pFUL1, and pFUL2) were combined with pICSL80001 (fLuc) and pICH41432 (tOCS) in the L1 acceptor pICH47751. All primers and gRNA sequences used for cloning are listed in **Table S8 and S9.**

673 CRISPR/Cas9 genome editing, plant transformation and identification of mutant alleles

674 CRISPR-Cas9 mutagenesis in tomato and physalis was performed as described previously[33,64,65].
675 Briefly, guide RNAs (gRNAs) were designed using the CRISPOR tool and the M82v1.0, Sweet-
676 100v2.0 or Phygriv1.0 genome assemblies. Final vectors were transformed into the tomato cultivar
677 M82, LA1589 or double-determinate Sweet-100, or into *P. grisea* by A*grobacterium tumefaciens*-
678 mediated transformation. CRISPR-Cas9 editing in tomato and physalis was verified by genotyping
679 or amplicon sequencing as described[33]. Base editing was quantified in first-generation (T0)
680 transgenics using EditR v1.0.10[66] and in the T1 generation with a CAPS marker. All primer
681 sequences are listed in **Table S8**.

682

683 Generation of near-isogenic lines (NILs)

684 Near-isogenic *SSP2*^{S169} lines in the domesticated M82 background were generated by crossing the
685 *S. pimpinellifolium* accession LA1589 with *S. lycopersicum* cv. M82), and backcrossing F2
686 individuals homozygous for *SSP2*^{S169} to the recurrent parent (*S. lyc.* cv. M82) over 4 (BC4) to 5
687 (BC5) generations. Presence *SSP2*^{S169} allele was confirmed by genotyping using a CAPS marker
688 (**Table S8**).

689

690 Transactivation assays

691 Transient transactivation assays with luciferase reporter constructs were conducted in *N.*
692 *benthamiana* leaves as previously described[67]. In brief, leaves of 3-4 week old plants were
693 infiltrated with mixtures of *A. tumefaciens* (strain GV3101) cultures containing effector, co-
694 effector, luciferase reporter, transfection control, and silencing inhibitor vectors. Effector
695 constructs contained the coding sequence (CDS) of *SSP*, *SSP2*^{F169} or *SSP2*^{S169} with an N-terminal
696 Myc tag and driven by the CaMV 35S promoter. The co-effector construct contained the CDS of
697 *SFT* with an N-terminal HA tag and driven by CMV 35S promoter. The luciferase reporter
698 constructs contained the CDS of fLUC driven by the upstream regions of *MC*, *SlFUL*, or *SlFUL2*.
699 The transfection control was pGREENII-0800-LUC, which contains the CDS of rLUC driven by
700 the CMV 35S promoter. A p19 construct was used to suppress silencing. Liquid cultures were
701 grown in 4 ml LB in 15 mL round-bottom Falcon tubes for 36 hrs at 30°C and 220 rpm.

Agrobacteria were harvested by centrifugation at 3000 rpm and resuspended in infiltration buffer (50 mM MES pH 5.7 and 10 mM $MgCl_2$) to an OD_{600} = 1. Before leaf infiltration, individual cultures were incubated up to 3 hrs at RT and combined to obtain mixtures with effectors, reporters (fLUC), and transfection control (pGREEN 35S:rLUC), and silencing inhibitor (p19) plasmids at final OD_{600} of 0.1, 0.1, 0.1, and 0.05. Agrobacteria mixtures were infiltrated into the 5th leaf using a needleless syringe, with four to twelve different plants being infiltrated for each combination. Leaf disks of 0.8 cm diameter were harvested 3 days after infiltration and flash-frozen in liquid nitrogen before grinding in a mix mill (twice 15 s^{-1} for 30s). Luciferase assays were performed using the Dual-Luciferase Reporter Assay System (Promega) and a Tecan Saphire plate reader. In short, leaf powder was extracted in 300 µl of 1x PLB and vigorously vortexed for 30 s. Volumes of 10 µl protein extracts were mixed with 40 µl luciferase reagent in 96-well microplates and incubated for 10 min at RT. Firefly luciferase (fLUC) activity was quantified with a 10 s integration time. Afterwards, reactions were mixed with 30 µl Stop & Glo and incubated for 10 min before Renilla luciferase (rLUC) activity was measured with a 10 s integration time. Transactivation activity of the effectors was determined by calculating the fLUC/rLUC ratios and statistically significant differences were determined using one-factor ANOVAs followed by Tukey tests.

DAP-seq

Myc-tagged coding sequences of *SSP*, *SSP2*^{F169} and *SSP2*^{S169} were amplified from effector constructs used in the transactivation assay. The pTnT™ vector, and the *SSP2*^{F169} and *SSP2*^{S169} inserts were digested using XhoI (NEB) and NotI-HF (NEB) and combined using T4 Ligase (NEB). The Myc-tagged coding sequence for *SSP* was amplified from M82 cDNA and cloned into pTnT™ vectors with the NEBuilder HiFi DNA Assembly Cloning Kit (NEB #E5520). Plasmid DNA was isolated from 100 ml bacterial cultures using the PureYield™ Plasmid Midiprep System (Promega, A2492). Two replicates of SSP and SSP2 proteins were expressed *in-vitro* in the TnT® SP6 High-Yield Wheat Germ Protein Expression System (Promega, L3260) from 3.5 µg plasmid DNA per reaction. High molecular weight DNA for genomic library construction was isolated from inflorescence meristem tissue of the *anantha* mutant in the Sweet-100 genotype using a CTAB protocol as described previously[33]. DAP-seq was performed as previously described with

minor modifications[27,68]. The DNA-library was prepared according to Franco-Zorilla & Prat (2021) with minor modifications. The gDNA library was purified using SPRI beads (B23317, Beckman Coulter). Adaptor ligation was verified by qPCR with primers specific for the indices (**Table S8**) and the KAPA standards 20, 2 and 0.2 nM (Roche) in 10 µl reaction volumes. DNA affinity-purification steps were performed according to Bartlett et al. (2017) with 75 ng of gDNA input library per replicate. Eluted libraries were single-indexed (**Table S8**). Eight uniquely indexed libraries were produced, two replicate libraries per protein (SSP, $SSP2^{F169}$, $SSP2^{S169}$) and two replicates of the input library as negative control. Indexed libraries were purified individually with the Monarch® PCR & DNA Cleanup Kit (NEB, T1030S). Individual indexed libraries were analyzed on a Fragment Analyzer (Agilent), purified with SPRI beads and pooled at equimolar (10 nM) concentrations. The pooled libraries were sequenced on 1 Illumina NovaSeq6000 lane at the Genome Technology Facility (GTF) of the University of Lausanne. A total of 753'327'838 PE150 reads (between 64'808'988 and 144'444'123 per sample) were generated.

Raw read quality was assessed using FastQC (v0.11.9; http://www.bioinformatics.babraham.ac.uk/projects/fastqc/). Adapter sequences were trimmed with NGmerge (v0.3, parameters -g -d -a)[69]. Reads were aligned to the SollycSweet-100v2.0 reference [33] with hisat2 (v2.2.0, default parameters)[70], and alignments were sorted and indexed using samtools (v1.15.1)[46]. Differential binding (DB) analysis was performed with the Bioconductor csaw package (v1.301)[71]. We used a window width of 10 bp and an estimated fragment length of 213 bp. Prior to counting, repeats were blacklisted from the genome using the SollycSweet-100v2.0 TE annotation[33]. To filter regions and windows, we used the global enrichment approach of the csaw module. Bins of 10000 bp were used for global background estimation. The median of the average abundances across all 10000 bp bins was used as the global background coverage estimate. We only retained windows with at least a 4-fold change from the global background coverage. We counted the reads into large bins and normalized with the wrapper function normFactors, which uses trimmed mean of M-values (TMM) method. Significant regions were identified with the csaw makeContrasts function (FDR ≤0.01). Gene-based annotation of differentially-bounds regions was performed using the detailRanges function of csaw (3 Kbp upstream and 2 Kbp downstream of TSS) and annotation file SollycSweet-100_genes_v2.1.1.gff3[33]. BED files with significant regions and BigWig files with normalized read coverage were exported via the *export* function of the rtracklayer package[72] in R. *De-novo*

motif discovery was performed with the 1000 most significant peaks (by FDR) for each sample by analysing genomic sequences from position -100 to +100 relative to the peak center using MEME (v 5.3.3; parameters -dna -mod zoops -nmotifs 3 -minw 6 -maxw 15 -maxsites 1000 -objfun classic -revcomp -markov_order 0)[73].

Genome-wide distribution of peaks was determined using ChIPSeeker (v1.32.0)[74] by annotating regions +/- 5 Kbp around the TSS with the function annotatePeak (parameters tssRegion=c(-5000, 5000)). Peak intensity profiles and peak heatmaps were generated using the computeMatrix, plotHeatmap, and plotProfile functions in deepTools[75]. The most-enriched motifs for SSP, $SSP2^{F169}$, and $SSP2^{S169}$ were mapped to the SollycSweet-100v2.0 reference[33] with the FIMO tool of the MEME Suite[73]. Browser shots of peak coverage, peak regions and binding motifs at putative direct targets were generated in jbrowse2[76].

RNA-seq

Meristem staging, collection, RNA extraction, and library construction for the $ssp^{CR\text{-}181}$ (188 bp deletion allele), $ssp2^{CR\text{-}122}$ (122 bp deletion allele) and $ssp^{CR\text{-}181}ssp2^{CR\text{-}122}$ mutants, and the WT in the genetic background of cv. M82 was performed as previously described[23]. In brief, seedlings shoot apices were collected at the transition (TM) stage of meristem maturation, and immediately submerged in ice-cold acetone. Shoot apices were manually dissected under a stereoscope and three biological replicates consisting of 14-22 meristems were collected per genotype from individual seedlings. Total RNA was extracted with the Arcturus Pico-Pure RNA Extraction kit (Thermo). We prepared indexed libraries using the TruSeq Stranded mRNA Library Prep kit from Illumina according to the manufacturer's instructions. Fragment size and concentration were assessed with a Bioanalyzer. Libraries were sequenced on 2 Illumina NovaSeq6000 lanes at the Genome Technology Facility (GTF) of the University of Lausanne. A total of 187'907'134 SE100 reads (between 14'133'226 and 17'789'680 per sample) were generated.

The quality of raw reads was assessed using FastQC (v0.11.9; http://www.bioinformatics.babraham.ac.uk/projects/fastqc/). Raw reads were aligned to the genome reference M82v1.0[33] using STAR[77] (v2.7.6a; parameters --runMode alignReads --outFilterType BySJout --outFilterMultimapNmax 20 --outMultimapperOrder Random --alignSJoverhangMin 8 --alignSJDBoverhangMin 1 --alignIntronMin 20 --alignIntronMax

1000000 --alignMatesGapMax 1000000). Alignments were sorted and indexed using samtools[46] and gene expression was quantified as unique read pairs aligned to reference annotated gene features (M82v1.1.1) using HTSeq-count (v0.11.2; parameter --order=pos --stranded=no --type=exon --idattr=Parent)[78].

All statistical analyses of gene expression were conducted in R[38]. Differentially expressed genes (DEGs) between the mutants *ssp*, *ssp2*, *ssp ssp2*, and the WT were determined with DESeq2 (v1.34.0)[79]. Raw count data was transformed in DESeqp2 by variant stabilizing transformation (VST). Reproducibility of biological replicates was assessed by hierarchical clustering (method ward.D) and principle component analysis (PCA) using the PCAtools package (v2.6.0) in R[38]. Significantly differentially expressed genes (DEGs) were identified in *ssp* (n=686), *ssp2* (n=180), and *sspssp2* (n=1507) genes with a 1.5-fold change ($\log_2FC \geq 0.58$, compared to the WT) and adjusted *p*-value ≤ 0.05 cutoff. Gene normalized z-scores were visualized in heatmaps using pheatmap (v1.10.12) and normalized expression of individual transcripts in transcripts per million (TPM) was plotted using ggplot2.

DATA AVAILABILITY

The LA1589 genome assembly is available at the Solanaceae Genomics Network (https://solgenomics.net/ftp/genomes/Solanum_pimpinellifolium/LA1589/2020/). Raw Nanopore sequence data is available on SRA under the BioProjects PRJNA607731 and PRJNA557253. Raw Illumina sequence data will be made available on SRA under the BioProject PRJNA1069353 upon publication. Seeds are available on request from S. Soyk.

ACKNOWLEDGEMENTS

We thank all members of the Soyk lab, Y. Eshed, and C. Fankhauser for helpful discussions; J. Marquis and J. Weber for support with sequencing; B. Tissot, L. Nerny, V. Vashanthakumar, Y. Emmenegger, A. Chatillon, L. Keel, and T. Stupp for support with plant care; G. Ghazi Soltani and S. Mainiero for support with experiments; J. M. Franco-Zorrilla for advice with DAP-seq; J. van Eck and K. Swartwood for advice with plant transformation; Z. Lippman, Y. Qi, and T. Jacobs for providing materials. This work was supported by the University of Lausanne, the European

Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (ERC Starting Grant "EPICROP" Grant No. 802008) to S.So., the Swiss National Science Foundation (SNSF) under an Eccellenza Professorial Fellowship (Grant No. PCEFP3_181238) and Project Grant (Grant No. 310030_212218) to S.So., and an UNIL Interdisciplinary Project Grant to N.Gl. and S.So., and an National Science Foundation Grant (IOS-1546625) to G.B.M and S.St..

AUTHOR CONTRIBUTIONS

A.N.G., S.St., and S.So. conceived the project and designed and planned experiments

A.N.G., M.B, L.L., J.I., I.J., J.Z., S.So. performed experiments and collected data

A.N.G., M.B, J.I., G.A., I.J., J.Z., R.R., C.I., N.Gu., J.J.-G., N.Gl., S.St., S.So. analysed data

N.Gl., G.B.M., S.St., S.So. aquired project funding.

A.N.G. and S.So. wrote the first draft of the manuscript

All authors read, edited, and approved the manuscript.

REFERENCES

1. Koenig, D. *et al.* Comparative transcriptomics reveals patterns of selection in domesticated and wild tomato. *Proc Natl Acad Sci U S A* **110**, E2655–E2662 (2013).

2. Renaut, S. & Rieseberg, L. H. The Accumulation of Deleterious Mutations as a Consequence of Domestication and Improvement in Sunflowers and Other Compositae Crops. *molecubar biology evolution* **32**, 2273–2283 (2015).

3. Moyers, B. T., Morrell, P. L. & McKay, J. K. Genetic Costs of Domestication and Improvement. *Journal of Heredity* **109**, 103–116 (2018).

4. Wallace, J. G., Rodgers-Melnick, E. & Buckler, E. S. On the Road to Breeding 4.0: Unraveling the Good, the Bad, and the Boring of Crop Quantitative Genomics. *Annual reviews genetics* **52**, 421–444 (2018).

5. Zhang, C. *et al.* The genetic basis of inbreeding depression in potato. *Nat Genet* **51**, 374–378 (2019).

6. Monroe, J. G., McKay, J. K., Weigel, D. & Flood, P. J. The population genomics of adaptive loss of function. *Heredity (Edinb)* **126**, 383–395 (2021).

7. Olson, M. V. MOLECULAR EVOLUTION ’99 When Less Is More: Gene Loss as an Engine of Evolutionary Change. *Am. J. Hum. Genet* **64**, 18–23 (1999).

8. Gao, C. Genome engineering for crop improvement and future agriculture. *Cell* **184**, 1621–1635 (2021).

9. Gaarslev, N., Swinnen, G. & Soyk, S. Meristem transitions and plant architecture—learning from domestication for crop breeding. *Plant Physiol* 1–12 (2021) doi:10.1093/plphys/kiab388.

10. Shalit, A. *et al.* The flowering hormone florigen functions as a general systemic regulator of growth and termination. *Proceedings of the National Academy of Sciences* **106**, 8392–8397 (2009).

11. Lifschitz, E., Ayre, B. G. & Eshed, Y. Florigen and anti-florigen – a systemic mechanism for coordinating growth and termination in flowering plants. *Front Plant Sci* **0**, 465 (2014).

12. Lifschitz, E. & Eshed, Y. Universal florigenic signals triggered by FT homologues regulate growth and flowering cycles in perennial day-neutral tomato. *J Exp Bot* **57**, 3405–3414 (2006).

13. Zhu, Y. *et al.* TERMINAL FLOWER 1-FD complex target genes and competition with FLOWERING LOCUS T. *Nature Communications 2020 11:1* **11**, 1–12 (2020).

14. Taoka, K. I. *et al.* 14-3-3 proteins act as intracellular receptors for rice Hd3a florigen. *Nature* **476**, 332–335 (2011).

15. Park, S. J. *et al.* Optimization of crop productivity in tomato using induced mutations in the florigen pathway. *Nat Genet* **46**, 1337–1342 (2014).

16. Alonge, M. *et al.* Major Impacts of Widespread Structural Variation on Gene Expression and Crop Improvement in Tomato. *Cell* **182**, 145–161 (2020).

17. Vaser, R., Adusumalli, S., Ngak Leng, S., Sikic, M. & Ng, P. C. SIFT missense predictions for genomes. *Nat Protoc* **11**, 1073–1081 (2016).

18. Soyk, S. *et al.* Variation in the flowering gene SELF PRUNING 5G promotes day-neutrality and early yield in tomato. *Nature Genetics 2016 49:1* **49**, 162–168 (2017).

19. Rodríguez-Leal, D., Lemmon, Z. H., Man, J., Bartlett, M. E. & Lippman, Z. B. Engineering Quantitative Trait Variation for Crop Improvement by Genome Editing. *Cell* **171**, 470-480.e8 (2017).

20. Abe, M. *et al.* FD, a bZIP protein mediating signals from the floral pathway integrator FT at the shoot apex. *Science (1979)* **309**, 1052–1056 (2005).

21. Romera-Branchat, M. *et al.* Functional Divergence of the Arabidopsis Florigen-Interacting bZIP Transcription Factors FD and FDP. *Cell Rep* **31**, 107717 (2020).

22. Wigge, P. A. *et al.* Integration of spatial and temporal information during floral induction in Arabidopsis. *Science (1979)* **309**, 1056–1059 (2005).

23. Park, S. J., Jiang, K., Schatz, M. C. & Lippman, Z. B. Rate of meristem maturation determines inflorescence architecture in tomato. *Proceedings of the National Academy of Sciences* **109**, 639–644 (2012).

24. Consortium, T. T. G. The tomato genome sequence provides insights into fleshy fruit evolution. *Nature* **485**, 635–641 (2012).

25. Šali, A. & Blundell, T. L. Comparative Protein Modelling by Satisfaction of Spatial Restraints. *J Mol Biol* **234**, 779–815 (1993).

26. Tahirov, T. H. *et al.* Structural analyses of DNA recognition by the AML1/Runx-1 Runt domain and its allosteric control by CBFbeta. *Cell* **104**, 755–767 (2001).

27. Bartlett, A. *et al.* Mapping genome-wide transcription-factor binding sites using DAP-seq. *Nat Protoc* **12**, 1659–1672 (2017).

28. Fowler, S. *et al.* GIGANTEA: a circadian clock-controlled gene that regulates photoperiodic flowering in Arabidopsis and encodes a protein with several possible membrane-spanning domains. *EMBO J* **18**, 4679–4688 (1999).

29. Sawa, M., Nusinow, D. A., Kay, S. A. & Imaizumi, T. FKF1 and GIGANTEA complex formation is required for day-length measurement in Arabidopsis. *Science (1979)* **318**, 261–265 (2007).

30. He, J. *et al.* Establishing Physalis as a Solanaceae model system enables genetic reevaluation of the inflated calyx syndrome. *Plant Cell* **35**, 351–368 (2023).

31. Richer; Michelle F. *et al.* Phage-assisted evolution of an adenine base editor with improved Cas domain compatibility and activity. *Nature Biotechnology* **38**, 883–891 (2020).

32. Ren, Q. *et al.* PAM-less plant genome editing using a CRISPR–SpRY toolbox. *Nat Plants* **7**, 25–33 (2021).

33. Alonge, M. *et al.* Automated assembly scaffolding using RagTag elevates a new tomato system for high-throughput genome editing. *Genome Biol* **23**, 258 (2022).

34. Fujii, Y., Shimizu, T., Toda, T., Yanagida, M. & Hakoshima, T. Structural basis for the diversity of DNA recognition by bZIP transcription factors. *Nat Struct Biol.* **10**, 889–893 (2000).

35. Grandillo, S. & Tanksley, S. D. *QTL Analysis of Horticultural Traits Differentiating the Cultivated Tomato from the Closely Related Species Lycopersicon Pimpinellifolium*. *Theor Appl Genet* vol. 92 (1996).

36. Lin, T. *et al.* Genomic analyses provide insights into the history of tomato breeding. *Nat Genet* **46**, 1220–1226 (2014).

37. Dröge-Laser, W., Snoek, B. L., Snel, B. & Weiste, C. The Arabidopsis bZIP transcription factor family — an update. *Curr Opin Plant Biol* **45**, 36–49 (2018).

38. R Core Team. R: A language and environment for statistical computing. Preprint at (2021).

39. Wang, X. *et al.* Genome of Solanum pimpinellifolium provides insights into structural variants during tomato breeding. *Nat Commun* **11**, (2020).

40. Zimin, A. V *et al.* Genome analysis The MaSuRCA genome assembler. *Bioinformatics* **29**, 2669–2677 (2013).

41. Alonge, M. *et al.* RaGOO: Fast and accurate reference-guided scaffolding of draft genomes. *Genome Biol* **20**, 1–17 (2019).

42. Xu, G. C. *et al.* LR_Gapcloser: a tiling path-based gap closer that uses long reads to complete genome assembly. *Gigascience* **8**, 1–14 (2019).

43. Walker, B. J., Abeel, T. ¤, Shea, T., Priest, M. & Abouelliel, A. Pilon: An Integrated Tool for Comprehensive Microbial Variant Detection and Genome Assembly Improvement. *PLoS One* **9**, 112963 (2014).

44. Shumate, A. & Salzberg, S. L. Liftoff: accurate mapping of gene annotations. *Bioinformatics* **37**, 1639–1643 (2021).

45. Gao, L. *et al.* The tomato pan-genome uncovers new genes and a rare allele regulating fruit flavor. *Nat Genet* **51**, 1044–1051 (2019).

46. Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).

47. Jin, J. *et al.* PlantTFDB 4.0: toward a central hub for transcription factors and regulatory interactions in plants. *Nucleic Acids Res* **45**, D1040–D1045 (2017).

48. Katoh, K. & Standley, D. M. MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability. *Mol Biol Evol* **30**, 772–780 (2013).

945 49. Minh, Q. B. *et al.* IQ-TREE 2: New Models and Efficient Methods for Phylogenetic
946 Inference in the Genomic Era. *Molecular Biologoy and Evolution* **37**, 1530–1534 (2020).

947 50. Altenhoff, A. M. *et al.* OMA orthology in 2024: improved prokaryote coverage, ancestral
948 and extant GO enrichment, a revamped synteny viewer and more in the OMA Ecosystem.
949 *Nucleic Acids Res* **52**, 513–521 (2024).

950 51. Huerta-Cepas, J. *et al.* PhylomeDB v3.0: an expanding repository of genome-wide
951 collections of trees, alignments and phylogeny-based orthology and paralogy predictions.
952 *Nucleic Acids Res* **39**, D556–D560 (2011).

953 52. Edgar, R. C. MUSCLE: multiple sequence alignment with high accuracy and high
954 throughput. *Nucleic Acids Res* **32**, 1792–1797 (2004).

955 53. Lassmann, T. & Sonnhammer, E. L. L. Kalign - An accurate and fast multiple sequence
956 alignment algorithm. *BMC Bioinformatics* **6**, 1–9 (2005).

957 54. Wallace, I. M., O'Sullivan, O., Higgins, D. G. & Notredame, C. M-Coffee: combining
958 multiple sequence alignment methods with T-Coffee. *Nucleic Acids Res* **34**, 1692–1699
959 (2006).

960 55. Minh, B. Q. *et al.* IQ-TREE 2: New Models and Efficient Methods for Phylogenetic
961 Inference in the Genomic Era. *Mol Biol Evol* **37**, 1530–1534 (2020).

962 56. Kalyaanamoorthy, S., Minh, B. Q., Wong, T. K. F., Von Haeseler, A. & Jermiin, L. S.
963 modelfinder: fast model selection for accurate phylogenetic estimates. *Nat Methods* **14**,
964 (2017).

965 57. Huerta-Cepas, J., Serra, F. & Bork, P. ETE 3: Reconstruction, Analysis, and Visualization
966 of Phylogenomic Data. *Mol Biol Evol* **33**, 1635–1638 (2016).

967 58. Huerta-Cepas, J., Dopazo, H., Dopazo, J. & Gabaldón, T. The human phylome. *Genome*
968 *Biol* **8**, 1–16 (2007).

969 59. Gabler, F. *et al.* Protein Sequence Analysis Using the MPI Bioinformatics Toolkit. *Curr*
970 *Protoc Bioinformatics* **72**, e108 (2020).

971 60. Shen, M. & Sali, A. Statistical potential for assessment and prediction of protein structures.
972 *Protein Sci* **15**, 2507 (2006).

61. Pettersen, E. F. *et al.* UCSF Chimera—A visualization system for exploratory research and analysis. *J Comput Chem* **25**, 1605–1612 (2004).

62. Engler, C., Youles, M. & Gruetzner, R. A Golden Gate Modular Cloning Toolbox for Plants. *ACS Synth Biol* **3**, 839–843 (2014).

63. Decaestecker, W. *et al.* CRISPR-TSKO: A Technique for Efficient Mutagenesis in Specific Cell Types, Tissues, or Organs in Arabidopsis. *Plant Cell* **31**, 2868–2887 (2019).

64. Swartwood, K., Joyce, · & Eck, V. Development of plant regeneration and Agrobacterium tumefaciens-mediated transformation methodology for Physalis pruinosa. *Plant Cell Tissue Organ Cult* **137**, 465–472 (2019).

65. Brooks, C., Nekrasov, V., Lipppman, Z. B. & Van Eck, J. Efficient gene editing in tomato in the first generation using the clustered regularly interspaced short palindromic repeats/CRISPR-associated9 system. *Plant Physiol* **166**, 1292–1297 (2014).

66. Kluesner, M. G. *et al.* EditR: A Method to Quantify Base Editing from Sanger Sequencing. *CRISPR J* **1**, 239 (2018).

67. Galvão, V. C. *et al.* PIF transcription factors link a neighbor threat cue to accelerated reproduction in Arabidopsis. *Nat Commun* **10**, 1–10 (2019).

68. Franco-Zorrilla, J. M. & Prat, S. DAP-Seq Identification of Transcription Factor-Binding Sites in Potato. in *Methods in Molecular Biology* vol. 2354 123–142 (Humana Press Inc., 2021).

69. Gaspar, J. M. NGmerge: Merging paired-end reads via novel empirically-derived models of sequencing errors. *BMC Bioinformatics* **19**, 1–9 (2018).

70. Kim, D., Paggi, J. M., Park, C., Bennett, C. & Salzberg, S. L. Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nature Biotechnology 2019 37:8* **37**, 907–915 (2019).

71. Lun, A. T. L. & Smyth, G. K. csaw: a Bioconductor package for differential binding analysis of ChIP-seq data using sliding windows. *Nucleic Acids Res* **44**, 45 (2015).

72. Lawrence, M., Gentleman, R. & Carey, V. rtracklayer: an R package for interfacing with genome browsers. *BIOINFORMATICS APPLICATIONS NOTE* **25**, 1841–1842 (2009).

73. Bailey, T. L., Johnson, J., Grant, C. E. & Noble, W. S. The MEME Suite. *Nucleic Acids Res* **43**, 39–49 (2015).

74. Yu, G., Wang, L.-G. & He, Q.-Y. ChIPseeker: an R/Bioconductor package for ChIP peak annotation, comparison and visualization. *Bioinformatics* **31**, 2382–2383 (2015).

75. Ramírez, F., Dündar, F., Diehl, S., Grüning, B. A. & Manke, T. deepTools: a flexible platform for exploring deep-sequencing data. *Nucleic Acids Res* **42**, W187–W191 (2014).

76. Diesh, C. *et al.* JBrowse 2: a modular genome browser with views of synteny and structural variation. *Genome Biol* **24**, 1–21 (2023).

77. Dobin, A. *et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).

78. Anders, S., Pyl, P. T. & Huber, W. HTSeq—a Python framework to work with high-throughput sequencing data. *Bioinformatics* **31**, 166–169 (2015).

79. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* **15**, 1–21 (2014).

Figure 1

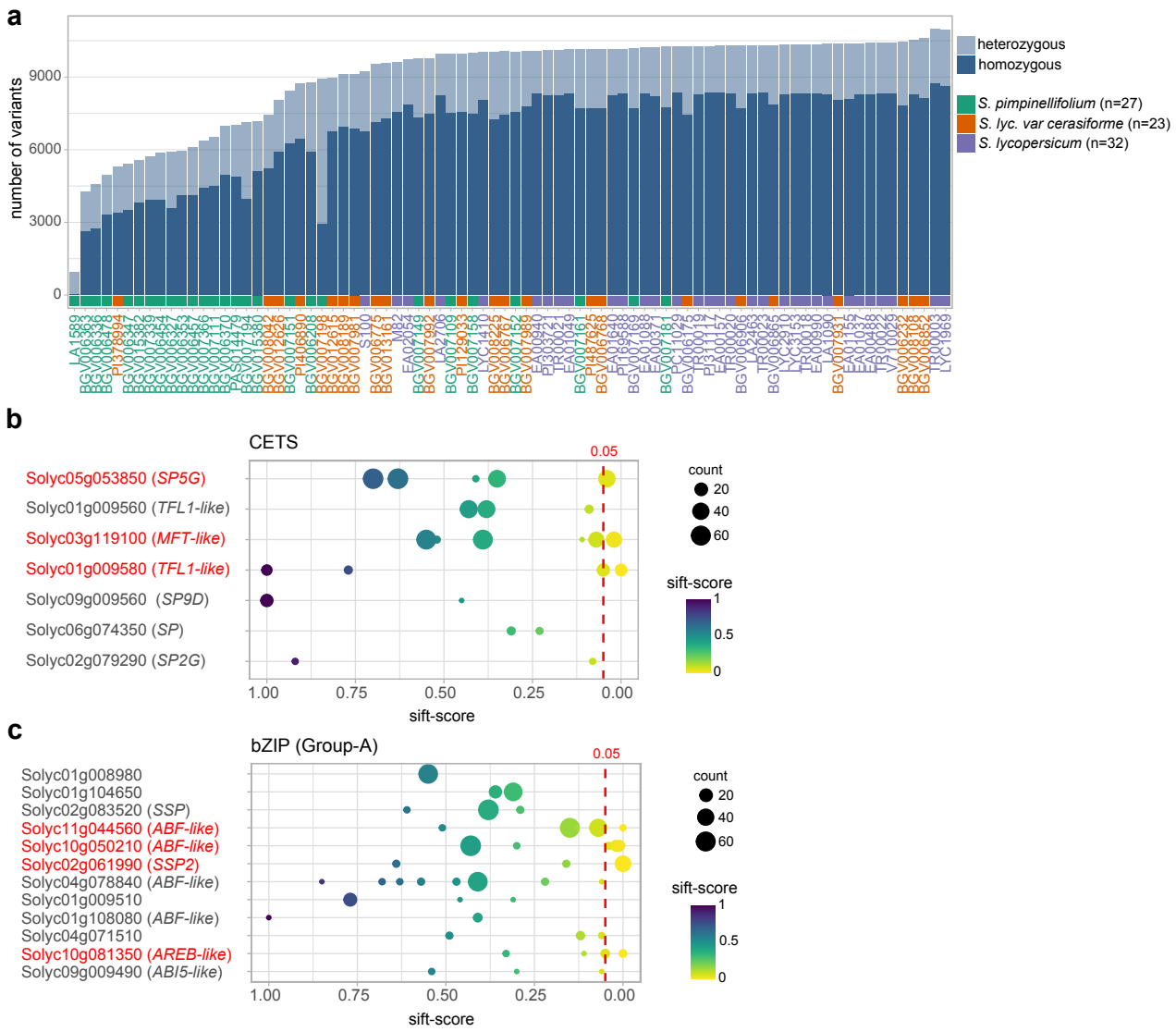

Figure 1: Predicting the load of deleterious variants along the domestication history of tomato. a, Number of predicted deleterious mutations in a panel of 82 tomato genomes, including wild species (*S. pimpinellifolium*, green), landraces (*S. lycopersicum var. cerasiforme*, orange), and cultivars (*S. lycopersicum*, purple). **b-c**, Prediction of deleterious variants across all CETS (b) and Group-A bZIP (c) genes. The dashed red line indicates the threshold for deleterious prediction (SIFT-score<0.05). Dot size scales with the number of genomes that carry the variant. Red font indicates genes with predicted deleterious mutations.

Figure 2

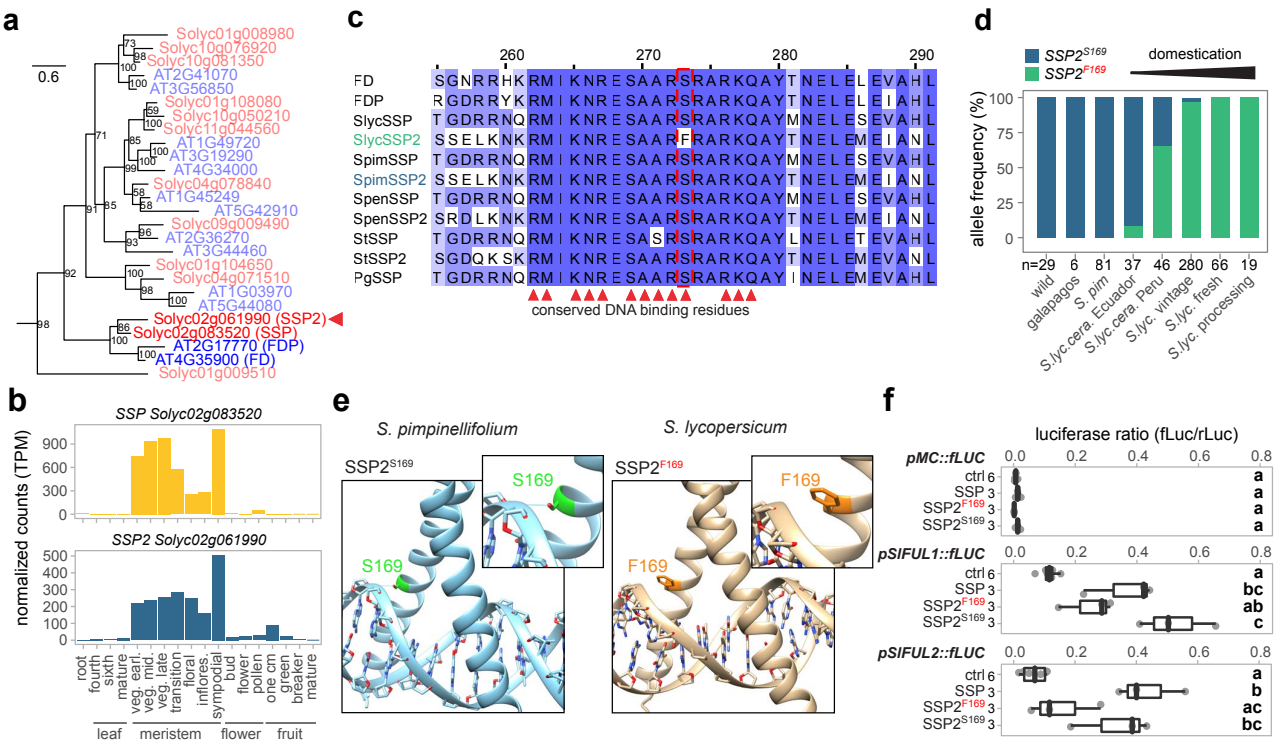

Figure 2: A deleterious mutation in *SSP2* reduces its transcription factor activity. a, Maximum-likelihood tree of A-group bZIP proteins in tomato (red font) and Arabidopsis (blue font). Red arrowhead marks SSP2. Numbers represent bootstrap values from 1,000 replicates and scale bar indicates the average number of substitutions per site. **b**, Normalized gene expression (TPM) for *SSP* and *SSP2* in different tissues and developmental stages (veg. earl./mid./late, stand for early, middle and late vegetative meristem stage). **c**, Partial alignment of SSP-like bZIP proteins from Arabidopsis, domesticated tomato (*S. lycopersicum*; *Slyc*), close wild tomato relative (*S. pimpinellifolium*; *Spim*), distant wild tomato relative (*S. pennellii*; *Spen*), potato (*S. tuberosum*; *St*), and *Physalis grisea* (*Pg*). Red arrowheads mark conserved DNA-binding residues. **d**, Distribution of ancestral (*SSP2^{S169}*) and derived (*SSP2^{F169}*) *SSP2* alleles in distant wild tomato relatives, wild relatives (*S. galapagense* / *S. cheesmaniae*), wild progenitor species (*S. pimpinellifolium*), landraces (*S. lyc var. cerasiforme*), and cultivars (*S. lycopersicum*). n=number of accessions. **e**, Predicted structures of ancestral SSP2^{S169} and derived SSP2^{F169} proteins on target DNA determined by homology modelling. Insets show a magnified view of the serine/ phenylalanine residue at position 169. **f**, Reporter assays in tobacco leaves using SSP, SSP2^{F169}, and SSP2^{S169} as effectors and firefly Luciferase (fLuc) driven by upstream sequences of *MC* (*pMC::fLUC*), *SlFUL1* (*pSlFUL1::fLUC*), and *SlFUL2* (*pSlFUL1::fLUC*) as reporter. Numbers indicate technical replicates. Ctrl indicates no effector control. Letters represent post-hoc Tukey's HSD tests results with 95% confidence level.

Figure 3

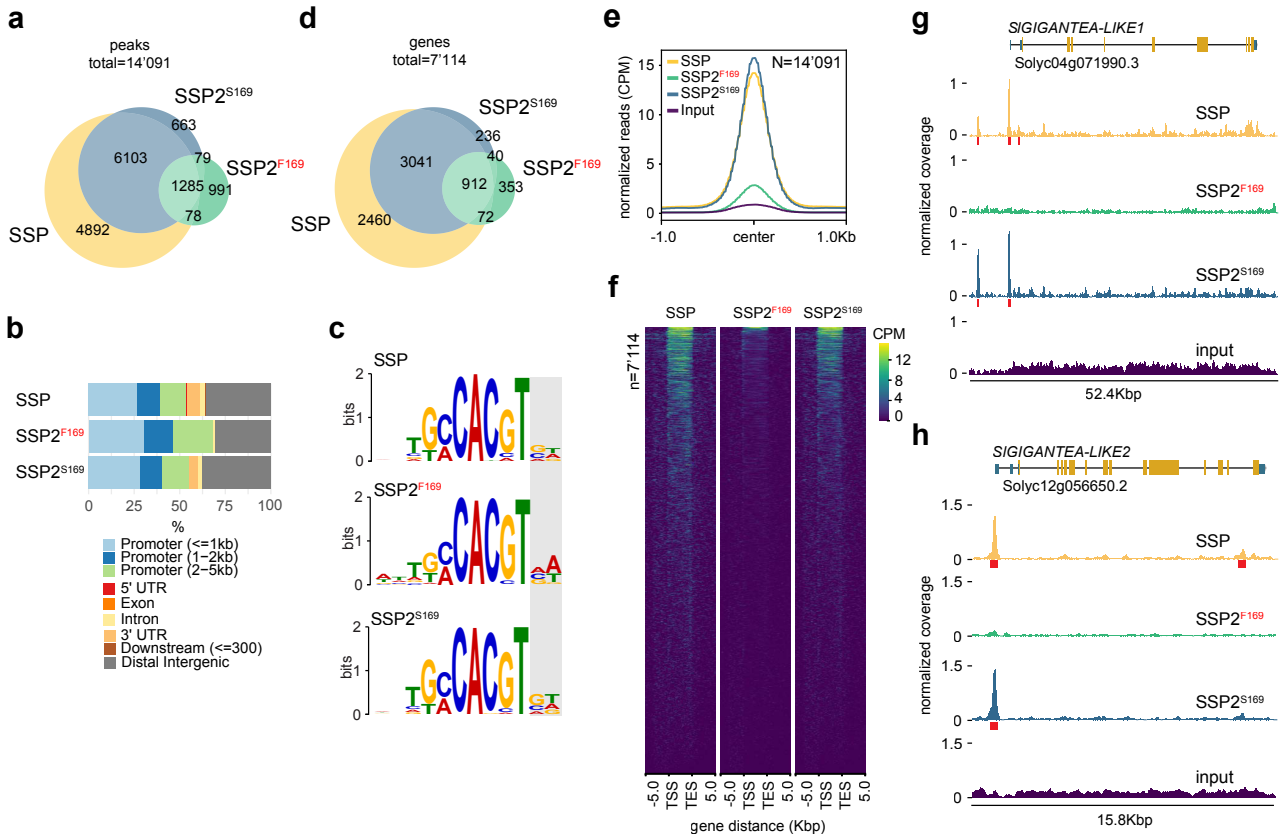

Figure 3: Domesticated SSP2^{F169} shows reduced binding at genome-wide target loci. a, Overlap of significant ($log_2FC \geq 3$, FDR ≤ 0.01) SSP, SSP2^{F169}, and SSP2^{S169} DAP-seq peaks (n=14'091). **b**, Distribution of significant SSP, SSP2^{F169}, and SSP2^{S169} DAP-seq peaks across gene features. **c**, Most-significant motifs identified by *de-novo* motif enrichment analysis of SSP, SSP2^{F169}, and SSP2^{S169} DAP-seq peak regions. Grey box delimits region with motif variation outside the core-motif. **d**, Overlap of genes with significant DAP-seq peaks ≤ 3 Kbp upstream and ≤ 2 Kbp downstream of the transcriptional start site (n=7'114). **e**, Profiles of normalized read coverage at significant SSP, SSP2^{F169}, and SSP2^{S169} peaks. **f**, Comparison of SSP, SSP2^{F169}, and SSP2^{S169} DAP-seq peaks relative to the transcriptional start (TSS) and end (TES) site of nearby genes (n=7'114). **g-h**, Browser view of SSP, SSP2^{F169}, and SSP2^{S169} DAP-seq peaks at *SlGIGANTEA-LIKE1* (g) and *SlGIGANTEA-LIKE2* (h). Normalized coverage (CPM) is shown in yellow, green and blue. Significant peak regions are indicated by red boxes.

Figure 4

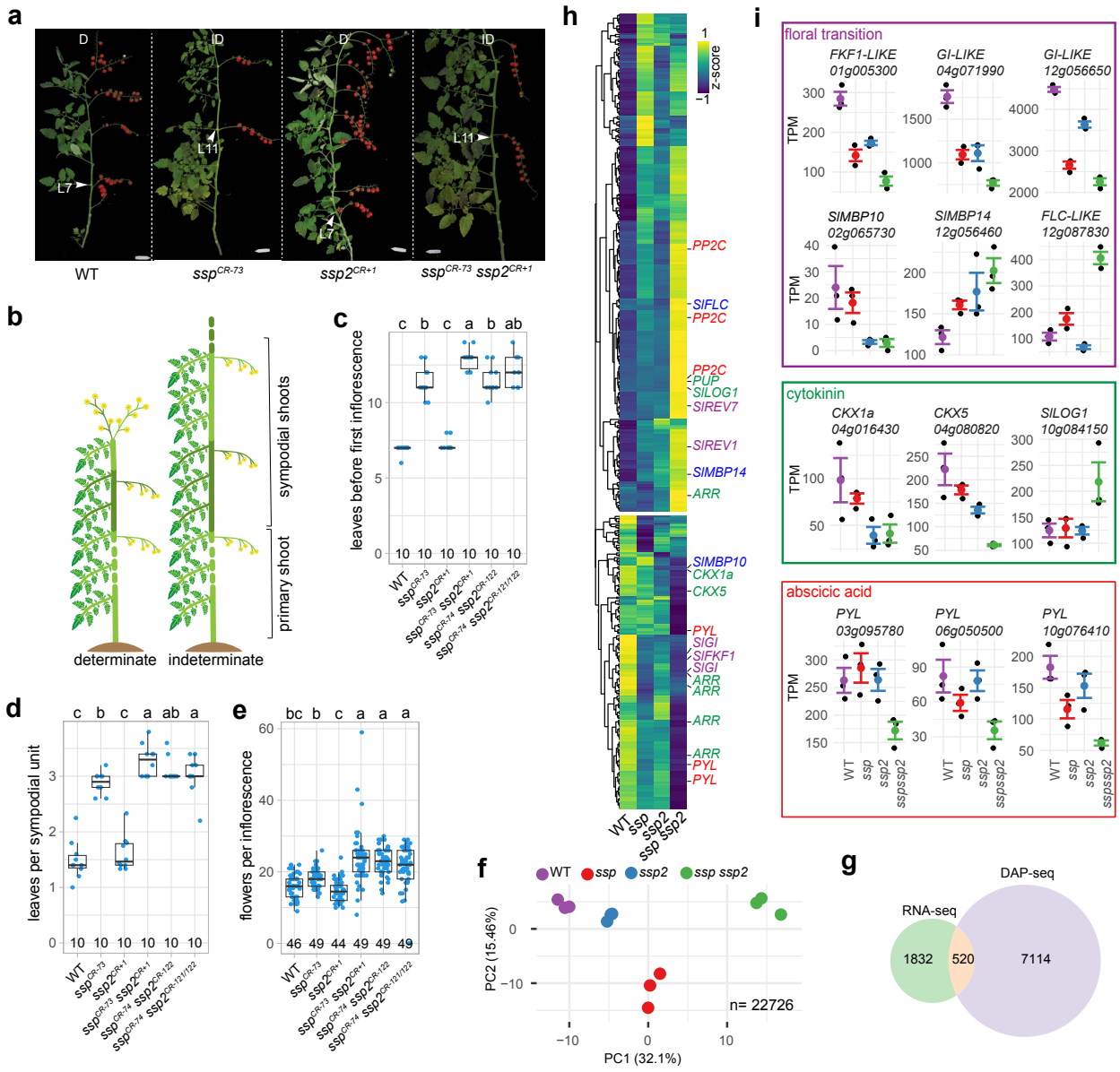

Figure 4: *SSP* and *SSP2* act partially redundant to regulate the transition to flowering. a, Representative images of wild-type S100, *ssp*^{CR} and *ssp2*^{CR} single mutants, and *ssp ssp2*^{CR} double mutants. L= leaf number, arrowheads mark the last leaf before flowering. Determinate (D) and indeterminate (ID) shoots are indicated. Scale bars represent 7.5 cm. **b**, Schematic depiction of tomato shoot architecture. Different shades of green delimit primary and sympodial shoots. **c-e**, Quantification of the floral transition (number of leaves before flowering) on the primary (c) and secondary (d) shoots, and the number of flowers per inflorescence (e) for genotypes shown in (a). The number of plants (c,d) and inflorescences (e) are indicated. Letters represent post-hoc Tukey's HSD tests results with 95% confidence level. **f**, Principal component analysis of 22'726 expressed genes in transition meristems of the WT, *ssp, ssp2*, and *ssp ssp2*, determined by RNA-seq. **g**, Overlap of genes differentially expressed ($\log_2$FC $\geq$ 0.58, FDR $\leq$ 0.05) in *ssp*, *ssp*2, and/or *ssp ssp2* with genes at SSP, SSP2^{F169}, and SSP2^{S169} DAP-seq peaks. **h**, Heatmap depicting expression of 520 putative SSP/SSP2 target genes. **i**, Normalized expression levels for selected putative direct targets. Genes are color coded based on the biological pathway.

Figure 5

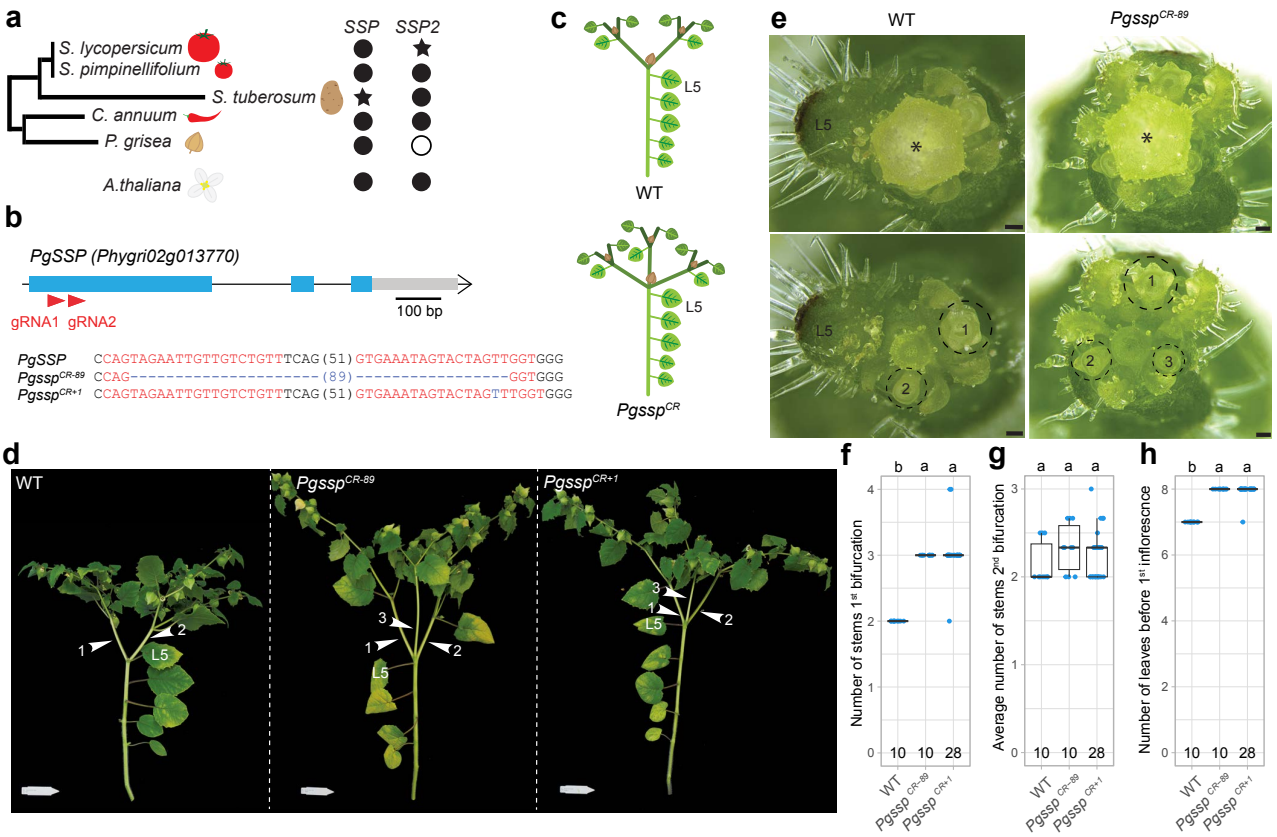

Figure 5: The genome of *Physalis grisea* encodes a single direct *SSP* ortholog that regulates meristem transitions. **a**, Scheme of the phylogenetic tree of tomato and closely related *Solanaceae* species. Filled circles, empty circles or star show presence, absence, or missense mutation, respectively, of *SSP*/*SSP2* or *FD/ FDP* in these species. Full tree is displayed in Fig. S6. **b**, CRISPR-Cas9 targeting of *PgSSP* in *P. grisea*. Blue boxes, black lines, and grey boxes represent exonic, intronic, and untranslated regions, respectively. Single guide RNAs (sgRNAs) are indicated with red arrowheads. PAM and sgRNA sequences are indicated in black and red bold letters, respectively; deletions are indicated with blue dashes; sequence gap length is given in parenthesis. Insertions are indicated by blue letters. **c**, Model of the growth habit of *P. grisea* WT and *Pgssp*^{CR} plants. Different shades of green delimit primary, first sympodial, and second sympodial shoots. The color of leaves corresponds with the shoot of origin. Note that the last leaf of each shoot is displaced upwards during shoot development. **d**, Representative pictures illustrating the difference in number of sympodial shoots in WT and *Pgssp* mutant plants. Last leaf before the shoot bifurcation is indicated (L5). White arrowheads indicate individual sympodial shoots. Scale bar represents 7.5 cm. **e**, Representative stereoscope images of the shoot apex of WT and *Pgssp* mutant plants. Upper images show the apex with a terminal flower (*). Lower images show the same view with the flower removed. The sympodial meristems (SYMs) are delimited by a dashed line and numbered in developmental order. Scale bar represents 100 µm. **f-h**, Quantification of the number of sympodial shoots at the first and second bifurcation, and flowering time (number of leaves before the first inflorescence). Number of plants is indicated at the bottom of the plots. Letters represent post-hoc Tukey's HSD tests results with 95% confidence level.

Figure 6

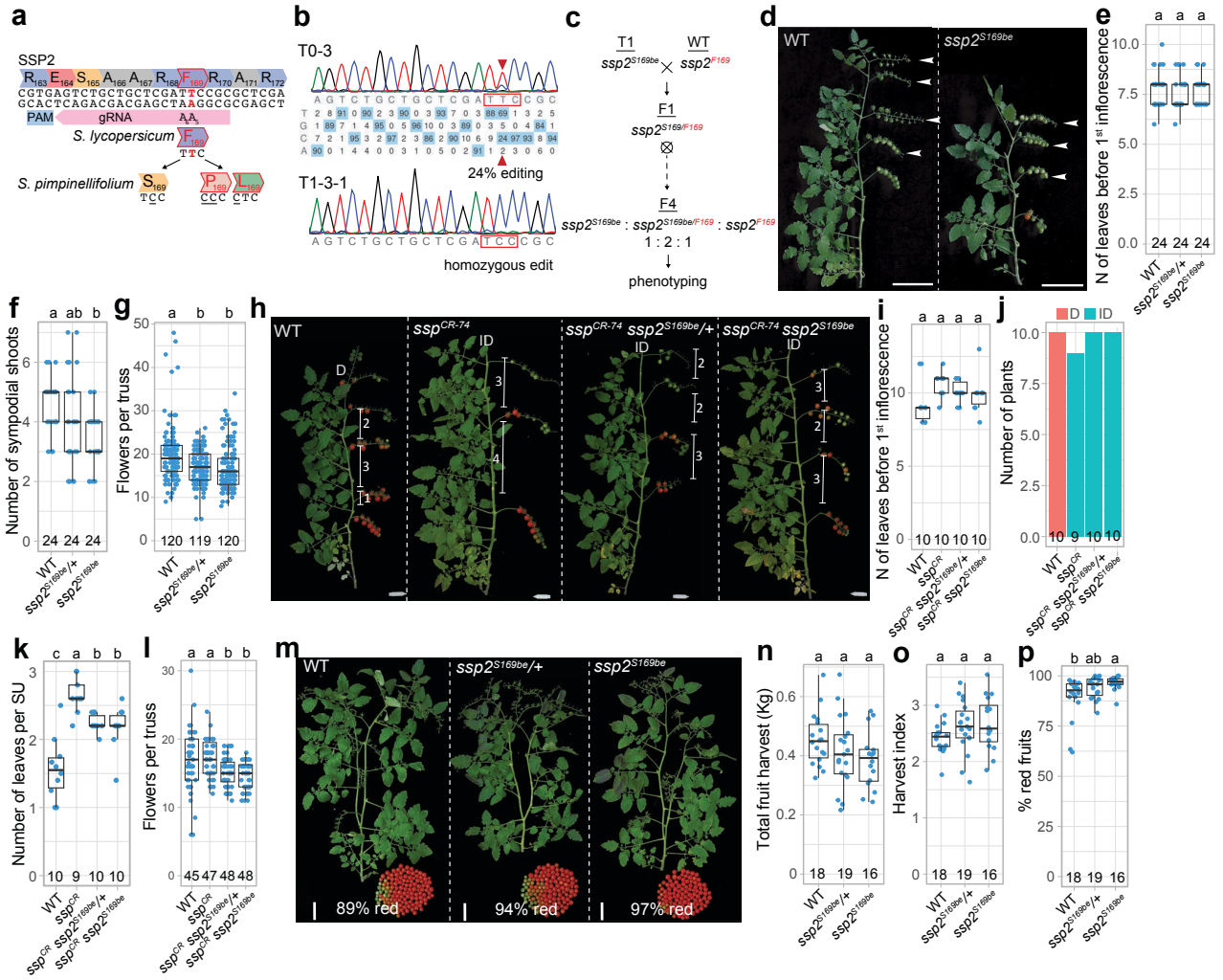

Figure 6: Repairing the deleterious *SSP2* mutation in domesticated tomato by base-editing leads to compact growth and earliness for yield. a, Base-editing strategy to correct the deleterious *SSP2* mutation in domesticated tomato using an adenosine base editor (ABE) and a PAM-less Cas9 variant. The target adenine in *SSP2* (A5) is at position 5 of the protospacer with a bystander adenine (A6) at position 6. Editing of the target codon (TTC) can lead to three different outcomes depending on which adenine is deaminated. Only editing the target nucleotide (A5) alone reverts the phenylalanine codon (TTC) back to the ancestral serine (TCC). **b**, Validation of editing in a chimeric first-generation (T0) transgenic and the corresponding T1 progeny by Sanger sequencing. The target nucleotide is indicated by a red arrowhead. **c**, Crossing scheme to generate the segregating *ssp2*^{S169be} F4 population. **d**, Representative pictures showing the total number of sympodial units on WT and *ssp2*^{S169be} plants. Terminal inflorescences of each sympodial unit are indicated by a white arrow. **e-g**, Quantification of flowering time (number of leaves before the first inflorescence), number of sympodial shoots, and number of flowers per truss of WT, *ssp2*^{S169be}/+ and *ssp2*^{S169be} plants. **h**, Representative pictures showing the number of leaves per sympodial unit and determinacy of WT, *ssp*^{CR}, *ssp*^{CR} *ssp2*^{S169be}/+ and *ssp*^{CR} *ssp2*^{S169be} plants. **i-l**, Quantification of flowering time (as in (e)), number of determinate plants, number of leaves per sympodial unit (SU), and number of flowers per truss of WT, *ssp*^{CR}, *ssp*^{CR} *ssp2*^{S169be}/+ and *ssp*^{CR} *ssp2*^{S169be} plants. Determinate (D) and indeterminate (ID) shoots are indicated. **m**, Representative images showing the full harvest of individual WT, *ssp2*^{S169be}/+ and *ssp2*^{S169be} plants. Percentage of red fruits is indicated. **n-p**, Quantification of total fruit yield (n), harvest index (total fruit yield / plant weight) (o), and percentage of red fruits. Number of plants are indicated in the plots for (e-g), (i-k) and (l-o). Letters on top of the plots represent post-hoc Tukey's HSD tests results with 95% confidence level**.** Scale bars represent 10 cm (d) and 7.5 cm in (h,m).

Figure S1

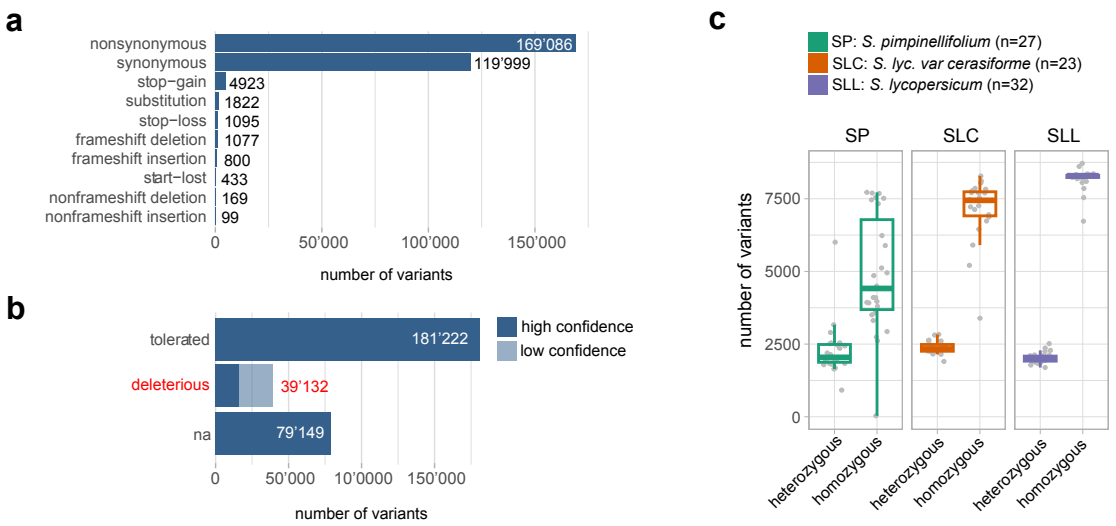

Figure S1: Prediction of deleterious variants in tomato. a, Number of coding sequence variants across a panel of 82 genomes. **b**, Number of non-synonymous variants predicted to be tolerated (sift-score ≥ 0.05), deleterious (sift-score < 0.05), or without prediction (na). Color code indicates confidence of SIFT prediction. **c**, Number of heterozygous and homozygous predicted deleterious mutations in wild (*S. pimpinellifolium*, n=27, in green), landrace (*S. lyc. var. cerasiforme*, n=23, in orange), and domesticated (*S. lycpersicum*, n=32, in purple) tomato genomes.

Figure S2

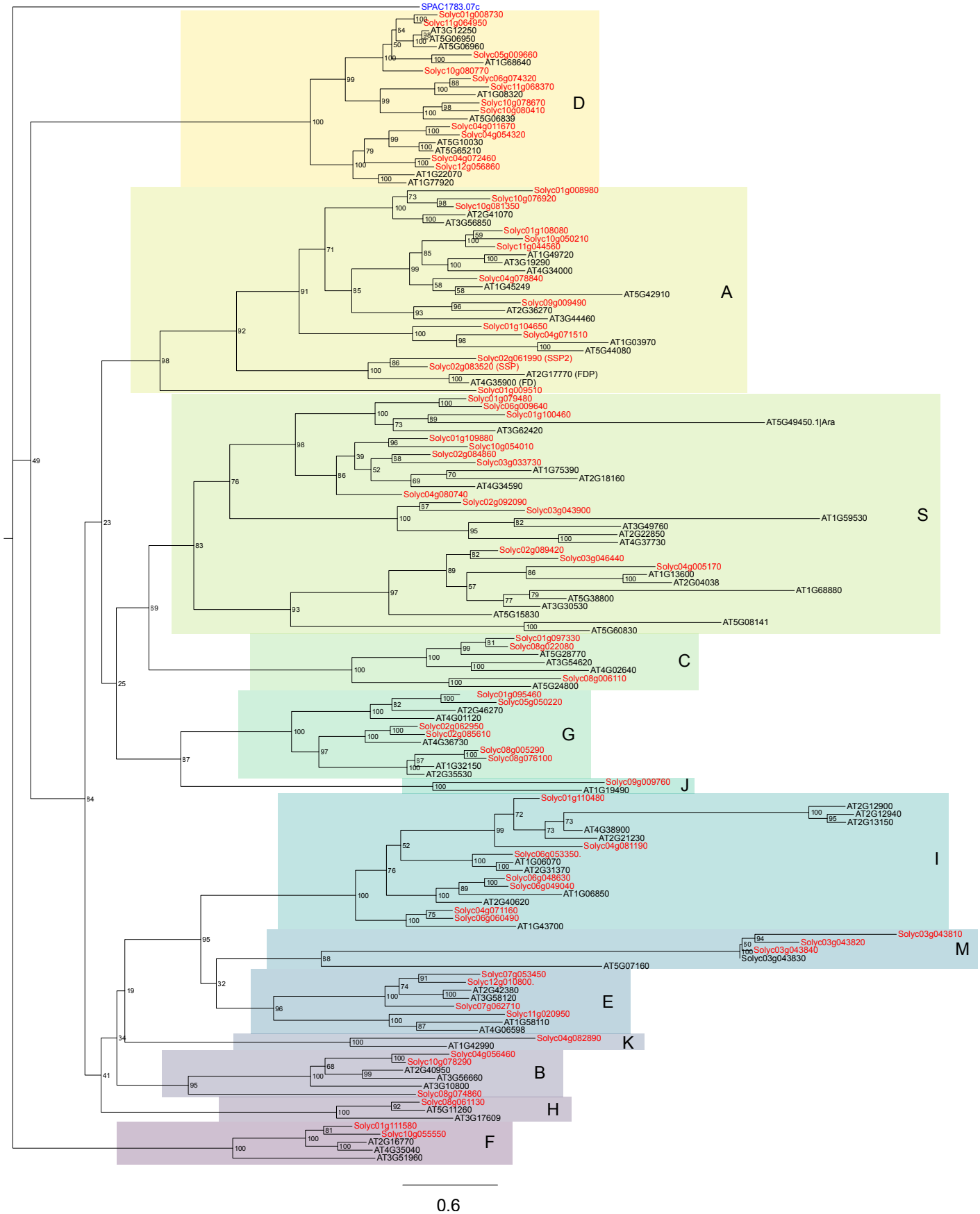

Figure S2: Phylogenetic analysis of the bZIP transcription factor family in Arabidopsis and tomato. Maximum-likelihood phylogenetic tree constructed with full-length bZIP protein sequences from Arabidopsis (n=74) and tomato (n=70). Arabidopsis and tomato proteins are indicated in black and red font, respectively. The yeast protein Pap1 was used as an outgroup (blue font). Proteins were classified into 13 groups (A-K, M, S) according to the Arabidopsis nomenclature [37]. Numbers represent bootstrap values from 1000 replicates, and scale bar indicates the average number of substitutions per site.

Figure S3

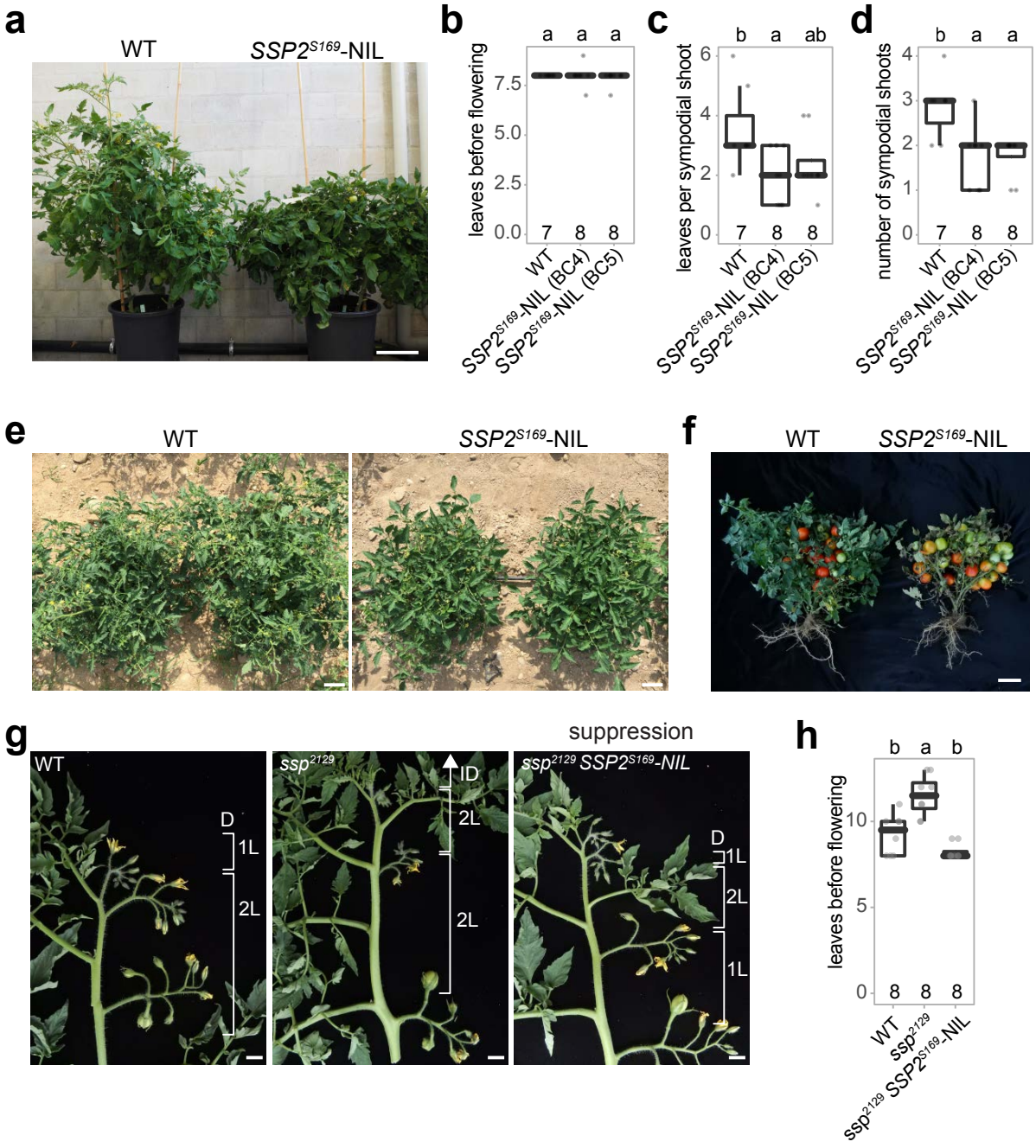

Figure S3: Introgression of ancestral *SSP2^{S169}* into domesticated tomato suppresses late flowering and indeterminate growth of *ssp* mutants. a, Representative image of greenhouse-grown wild-type (WT) and *SSP2^{S169}*-NIL individual in the determinate M82 background. **b-d**, Quantification of the floral transition (the number of leaves before flowering) on primary (b) and sympodial shoots (c), and the number of sympodial shoot units (d). **e, f,** Representative images of field-grown WT and *SSP2^{S169}*-NIL plants at flowering (c) and fruiting (d) stage. **g**, Representative images of detached WT, *ssp^{2129}* and *ssp^{2129} SSP2^{S169}*-NIL shoots (in the determinate M82 background). D, determinate; ID, indeterminate; L, leaves. **h**, Quantification of the floral transition on the primary shoot for genotypes shown in (e). Numbers at the bottom and letters at the top of the plots of (b) and (f) represent the number of replicate plants and post hoc Tukey's HSD test results with 95% confidence level, respectively. Scale bars indicate 10 cm (a, e, f) and 1 cm (g).

Figure S4

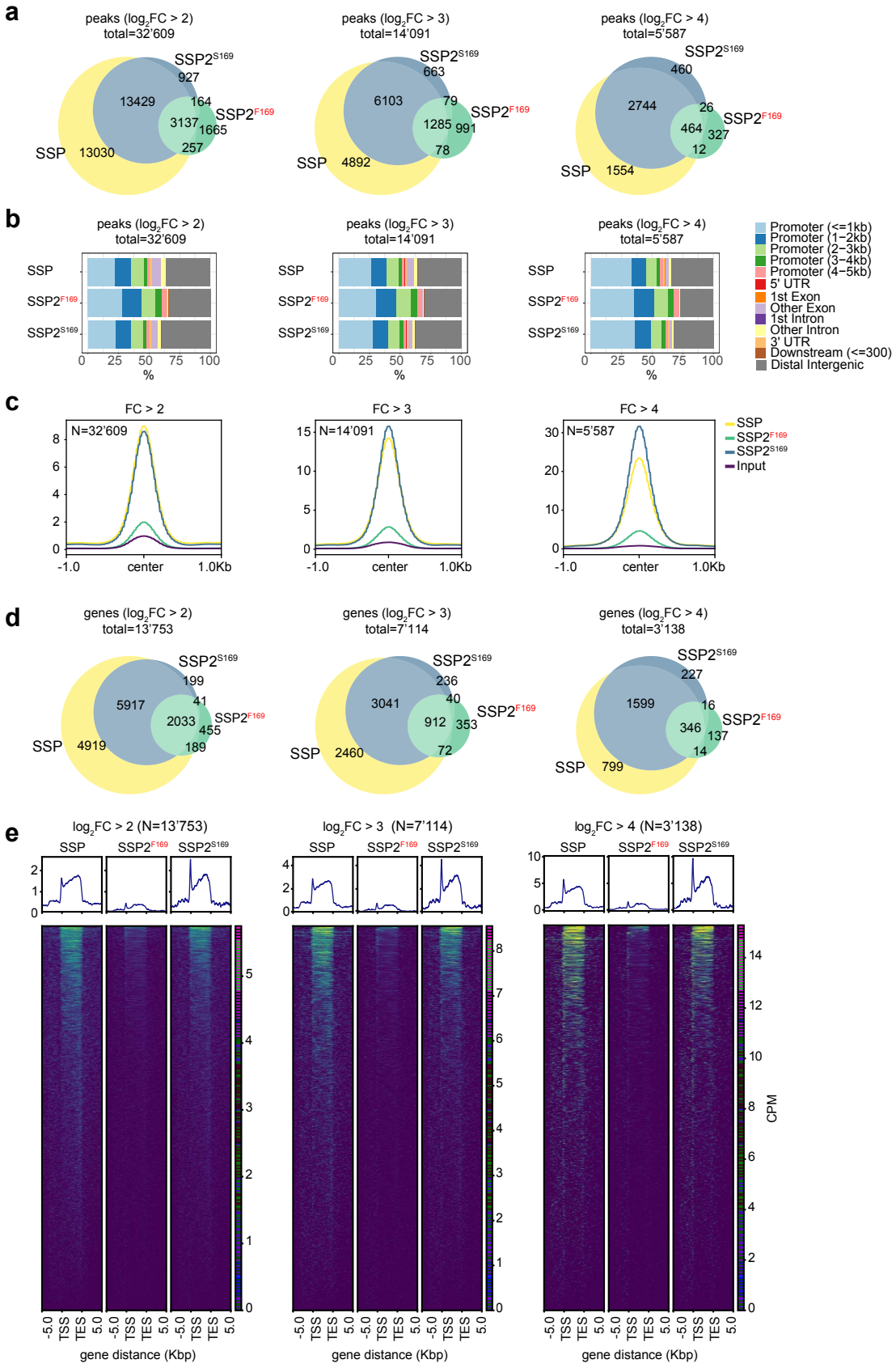

Figure S4: Identification of SSP, SSP2^{F169}, and SSP2^{S169} genome-wide binding sites by DAP-seq. a, Overlap of SSP, SSP2^{F169}, and SSP2^{S169} DAP-seq peaks at different significant thresholds ($log_2FC \geq 2$, 3, 4). **b**, Distribution of SSP, SSP2^{F169}, and SSP2^{S169} DAP-seq peaks across gene features at different significant thresholds as in (a) **c**, Profiles of normalized read coverage at SSP, SSP2^{F169}, and SSP2^{S169} peaks at different significant thresholds as in (a). **d**, Overlap of genes with DAP-seq peaks ≤ 3 Kbp upstream and ≤ 2 Kbp downstream of the transcriptional start site, at different significant thresholds as in (a). **e**, Comparison of SSP, SSP2^{F169}, and SSP2^{S169} DAP-seq peaks relative to the transcriptional start (TSS) and end (TES) site of nearby genes, at different significant thresholds as in (a). Top and bottom panels show coverage profiles and heatmaps, respectively.

Figure S5

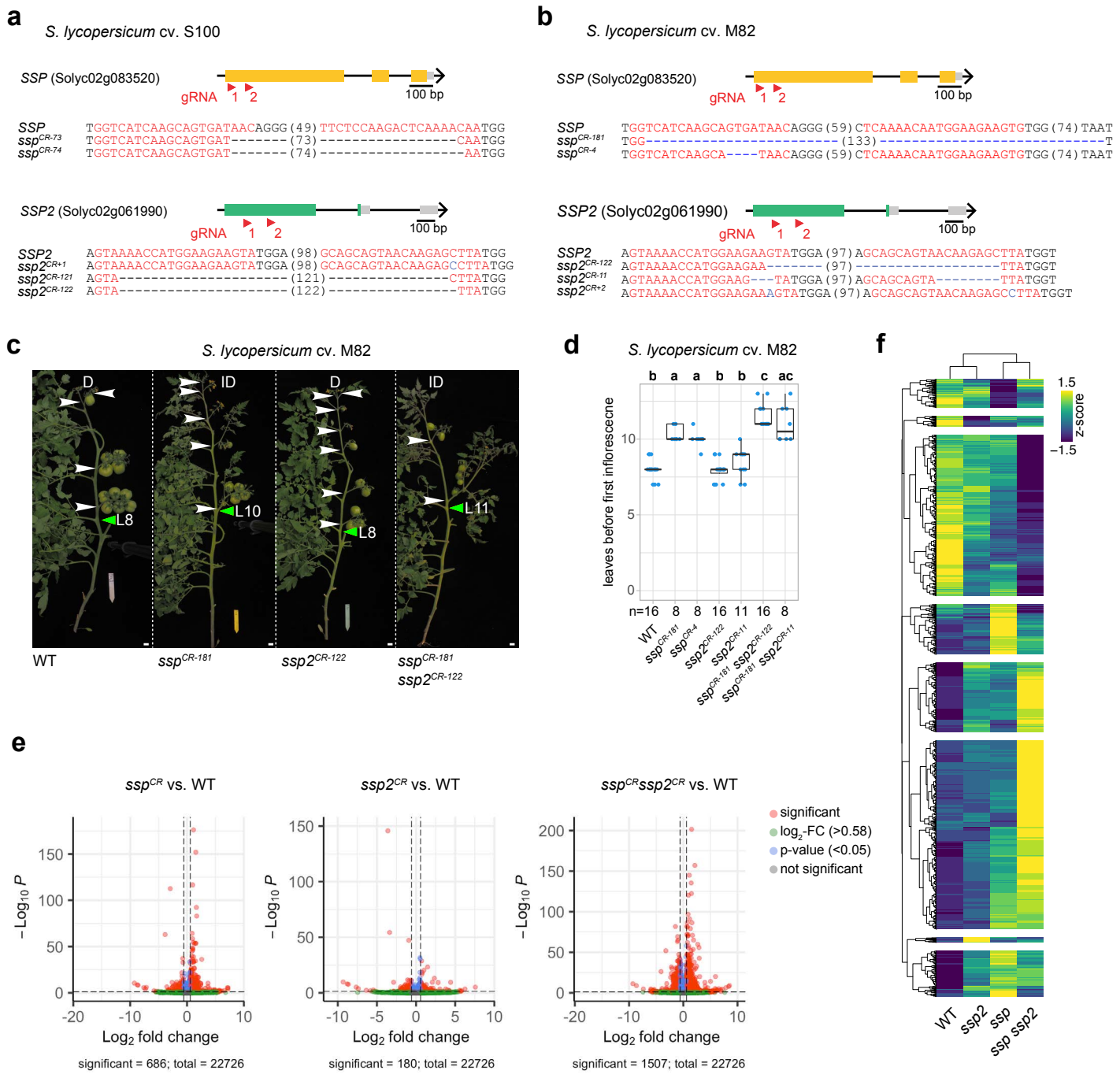

Figure S5: Targeting *SSP* and *SSP2* in two tomato cultivars by CRISPR-Cas9. a,b CRISPR-Cas9 targeting of *SSP* and *SSP2* in *S. lycopersicum* cv. S100 (a) and cv. M82 (b). Orange boxes, black lines, and grey boxes represent exonic, intronic, and untranslated regions, respectively. Single guide RNAs (sgRNAs) are indicated with red arrowheads. PAM and protospacer sequences are indicated in black and red bold letters, respectively; deletions are indicated with blue dashes; sequence gap length is given in parenthesis. **c**, Representative images WT S100, *ssp*^{CR} and *ssp2*^{CR} single mutants, and *ssp ssp2*^{CR} double mutants. L= leaf number, white arrowheads mark inflorescences. Determinate (D) and indeterminate (ID) shoots are indicated. Scale bars represents 1 cm. **d**, Quantification of the floral transition on the primary shoot for genotypes in (c). N, number of plants. Letters represent post hoc Tukey's HSD tests. **e**, Volcano plots showing differentially expressed genes ($\log_2$ FC > 0.58, FDR < 0.05) in *ssp*^{CR} and *ssp2*^{CR} single mutants, and *ssp ssp2*^{CR} double mutants compared to WT (cv. M82). **f**, Heatmap of z-scores showing expression pattern for 1'832 genes that are differentially expressed ($\log_2$ FC > 0.58, FDR < 0.05) in *ssp*^{CR}, *ssp2*^{CR} single mutants, and/or *ssp ssp2*^{CR} double mutants in M82.

Figure S6

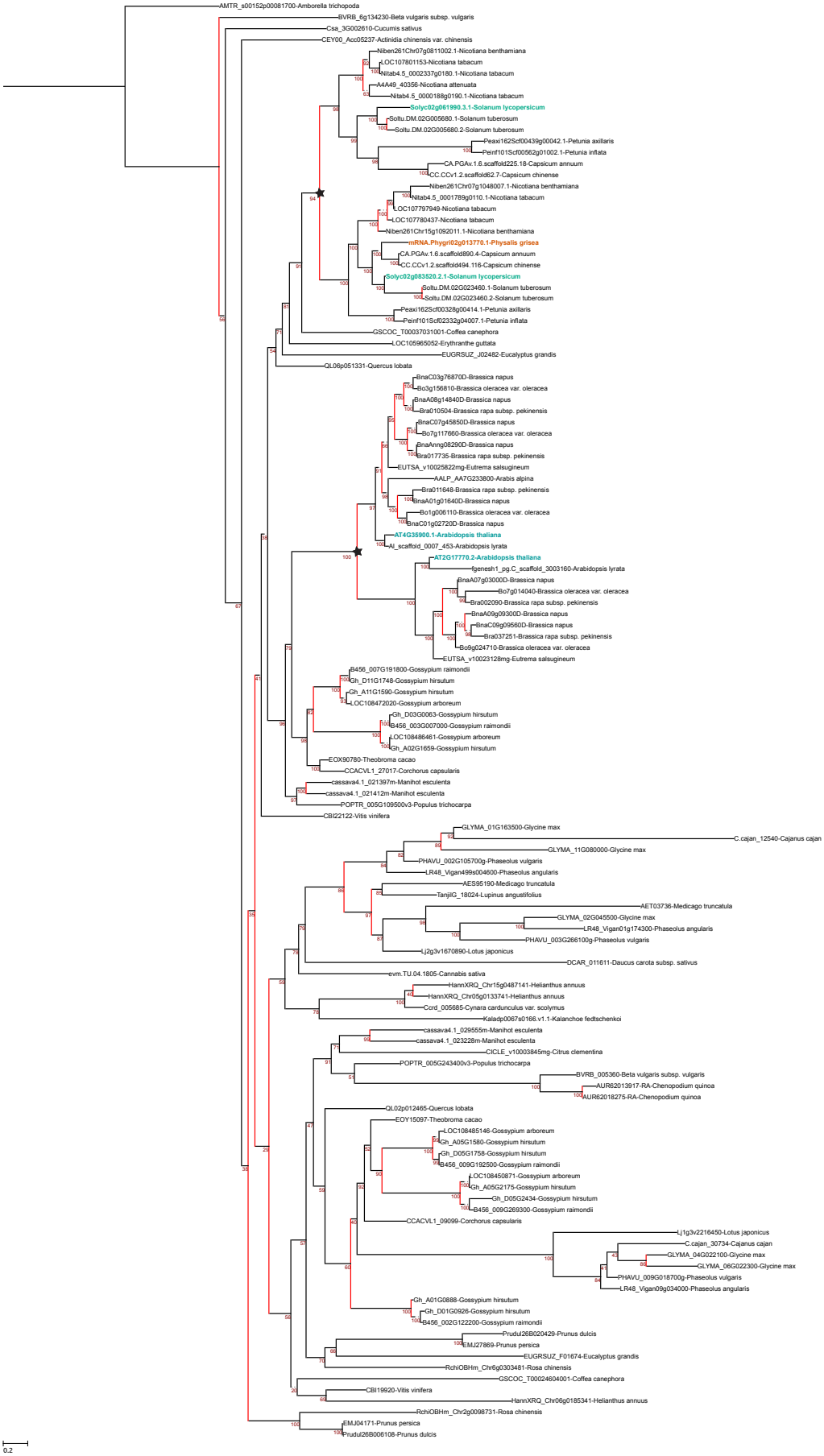

Figure S6: Phylogenetic analysis of SSP homologs in eudicots. Maximum-likelihood phylogenetic tree constructed with 128 full-length bZIP protein sequences from 51 eudicot species. Tomato, Arabidopsis, and Physalis proteins are highlighted in red, blue, and orange font, respectively. Red branches indicate duplication events, and the two separate duplication events in the *Solanaceae* and *Brassicaceae* are highlighted with stars. Numbers represent bootstrap values from 1000 replicates, and scale bar indicates the average number of substitutions per site.

Figure S7

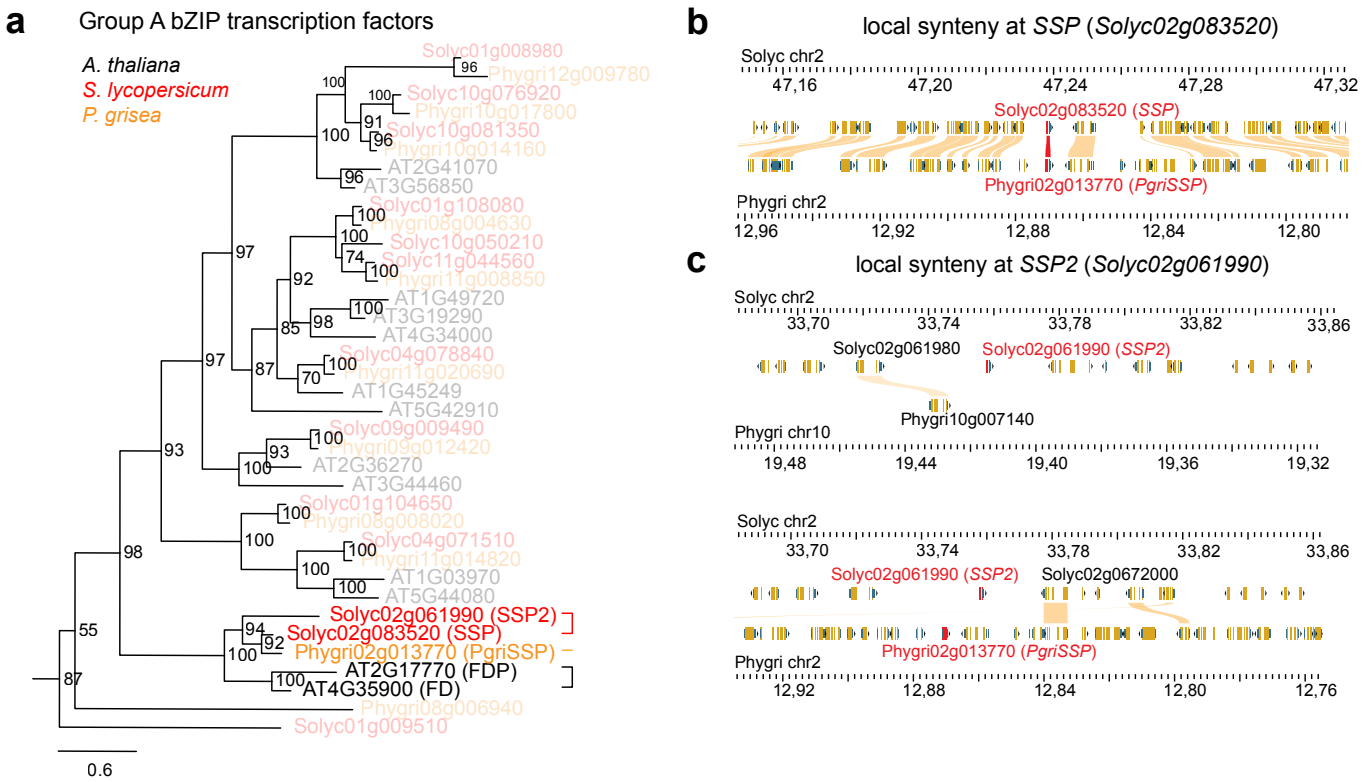

Figure S7: The ortholog of *SSP2* in *Physalis grisea* was lost during evolution. a. Maximum-likelihood phylogenic tree of the group A bZIP transcription factor family of *A. thaliana*, *S. lycopersicum* and *P. grisea*. Numbers represent bootstrap values from 1000 replicates, and scale bar indicates the average number of substitutions per site. **b,c**, Browser view of synteny analysis of *SSP* (b) and *SSP2* (c) between tomato (cv. S100) and *P. grisea*. Yellow rectangles show annotated genes and yellow streaks link them with their syntenic counterpart. *SSP* and *SSP2* genes are indicated in red. Note the lack of a unique syntenic block for *SSP2* in *P. grisea* in (c).

Figure S8

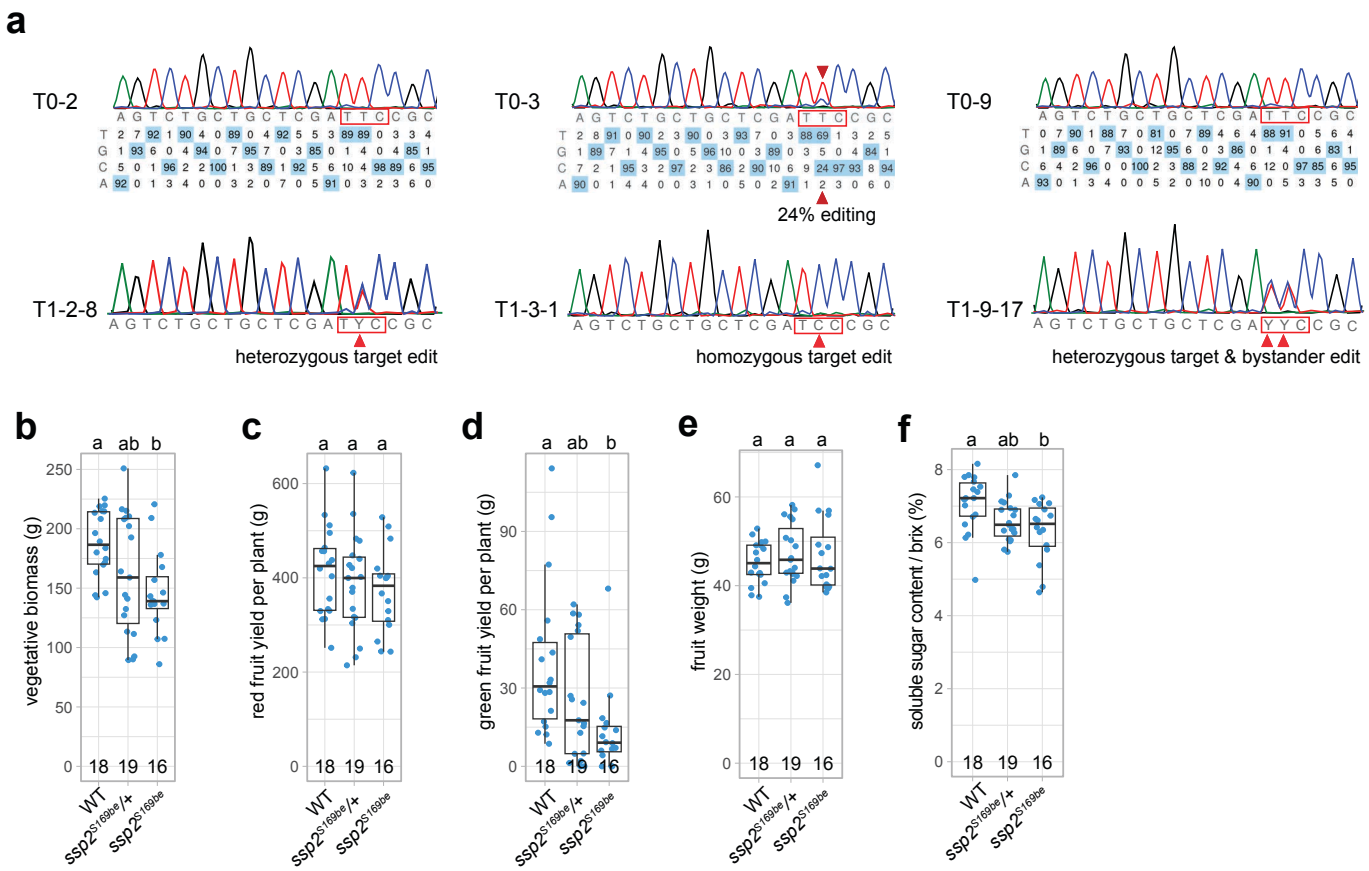

Figure S8: Base-editing of *SSP2* in domesticated tomato and its effect on different tomato yield components. a, CRISPR base-editing sequencing result of three T0 individuals (upper row) and their T1 progeny (lower row). Note that the target edit was detected in only one T0 individual (T0-3) but in three T1 families. One T1 individual (T1-9-17) was also edited at the bystander adenine. The edited nucleotides are indicated by a red arrowhead. **b-f**, Quantification of the vegetative biomass (b), total red and green fruit harvest (c,d), average fruit weight (e), and average soluble sugar content (brix) (f). The number of plants are indicated in the plots. Letters on top of the plots represent post-hoc Tukey's HSD tests results with 95% confidence level.