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

Anna N. Glaus[1,2], Marion Brechet[1,2], Gwen Swinnen[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 may allow predictable crop improvement.

39 **MAIN**

40 Deleterious mutations lead to the alteration or loss of gene activity. Crop domestication has
41 been accompanied by an accumulation of potentially deleterious mutations[1,2], a phenomenon
42 described as the genetic cost of domestication[3]. Such potentially harmful variants likely
43 influence many important agricultural traits[4]. For example, harmful recessive alleles can have
44 detrimental effects that are exposed in homozygous progeny during inbreeding[5]. Deleterious
45 mutations are often considered to mainly negatively affect fitness of natural populations but
46 recently, a more nuanced view has been proposed that considers their adaptive value[6,7].
47 Deleterious, loss-of-function mutations may confer an evolutionary advantage during rapid
48 shifts in environmental conditions and the selective pressures thereof[7]. Crop domestication
49 created novel environments under which many traits that were beneficial in the wild likely
50 became neutral or even detrimental. Illustrative examples include loss of photoperiodic
51 flowering and seed shattering. These observations support the "less-is-more" idea, which
52 proposes that selection may favor a less-than-complete repertoire of functional genes[7].
53 Nonetheless, eliminating deleterious variants from domesticated germplasm has been proposed
54 as a major goal in future crop breeding to avert potential harmful effects[4,8]. However, correcting
55 genetic variants by recombination during cross-breeding can be complicated by genetic linkage
56 with beneficial alleles or near fixation in domesticated populations. Recent advances in
57 precision genome editing promise to facilitate the repair of deleterious variants[9]. However, to
58 our knowledge, an experimental demonstration of precision genome editing for the repair of
59 deleterious variants in domesticated germplasm has been lacking.

60 A recurrent target of selection during crop domestication and breeding are alterations in
61 flowering time[9]. Changes in flowering time allowed the adaptation of crops to novel
62 environments and growing seasons different from their wild ancestors' origin. The floral
63 transition also influences plant architecture by balancing vegetative and reproductive growth[10].
64 At the molecular level, flowering occurs when the universal flowering hormone, florigen,
65 reaches a critical level that triggers stem cells in the shoot meristems to switch from vegetative
66 to reproductive growth. In the model crop tomato (*Solanum lycopersicum*), florigen is encoded
67 by *SINGLE FLOWER TRUSS* (*SFT*), a homolog of Arabidopsis *FLOWERING LOCUS T* (*FT*)
68 and member of the *CENTRORADIALIS, TERMINATING FLOWER1, SELF-PRUNING*
69 (*CETS*) gene family[11]. While *SFT* promotes the floral transition, *SELF PRUNING* (*SP*) acts as
70 antiflorigen and opposes the activity of florigen to repress flowering[12]. Evidence from rice and
71 Arabidopsis suggests that florigen protein competes with antiflorigen protein 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 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 var. lycopersicum*) genomes (**Fig. 1a**, **Supplementary Table 1)**[16,17]. 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) (**Supplementary Fig. 1a, b, Supplementary Table 2**)[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 (**Supplementary Fig. 1c**). Next, we focused on core components of the FAC[14] and searched for deleterious variants in *CETS* and Group-A bZIP genes (**Supplementary Fig. 2)**[18]. Among all 12 annotated tomato *CETS* genes, we identified three genes with predicted deleterious variants (**Fig. 1b**). Besides a predicted deleterious variant with low confidence in an uncharacterized *TERMINATING FLOWER1* (*TFL1*)-like gene, we detected high-confidence predicted deleterious variants in an uncharacterized *MOTHER OF FT* (*MFT*)-like gene (Solyc03g119100) and in the known flowering repressor *SELF-PRUNING 5G* (*SP5G*; Solyc05g053850). The deleterious variants in *MFT-like* and *SP5G* were absent from wild tomato genomes but detected in 6 (26%) and 13 (57%) landrace, and 30 (94%) and 32 (100%) domesticated genomes, respectively (**Fig. 1c**, **Supplementary Table 3**). We also detected the *sp-classic* breeding mutation (P76L) that was predicted to not be deleterious but tolerated,

104 which supports a hypomorphic nature of the mutation[19]. Among all 13 annotated tomato
105 Group-A bZIP genes, we identified two uncharacterized abscisic acid responsive element
106 binding factor (*ABF*)-like genes and an SSP-like gene (Solyc02g061990) with high-confidence
107 predicted deleterious mutations (**Fig. 1d and Supplementary Fig. 2**). The predicted
108 deleterious variant in Solyc02g061990 was the most frequent and absent from wild tomato
109 genomes but detected in 5 landrace (22%) and 31 domesticated (97%) genomes (**Fig. 1e**). We
110 concluded from these analyses that several central florigen pathway components have acquired
111 potentially deleterious mutations during tomato domestication.

112

113 **A missense mutation in *SSP2* accumulated during domestication**

114 A phylogenetic analysis comparing group-A bZIP proteins of tomato and Arabidopsis showed
115 that Solyc02g061990 is most closely related to *SSP*, thus we named the gene *SSP2* (**Fig. 2a**
116 and **Supplementary Fig. 2**). SSP and SSP2 form a sister clade to the Arabidopsis proteins FD
117 and FD PARALOG (FDP) [20], with SSP and FD being the more ancient genes. In Arabidopsis,
118 FD and FDP are involved in flowering control and phytohormone responses[21,22]. Expression
119 data from different tomato plant tissues showed that *SSP* and *SSP2* had similar expression
120 patterns, suggesting functional redundancy, most notably in secondary (sympodial) shoot
121 meristems (**Fig. 2b**)[23,24]. The putative deleterious variant in SSP2 causes a serine-to-
122 phenylalanine (S169 to F169) exchange at a conserved residue in the DNA-binding domain
123 (**Fig. 2c**). We analyzed the distribution of the ancestral ($SSP2^{S169}$) and domesticated ($SSP2^{F169}$)
124 variants across 768 re-sequenced tomato accessions and found that the domesticated allele was
125 absent from wild tomato species. The putative deleterious variant first arose in tomato
126 landraces (*S. lycopersicum var. cerasiforme)*, was enriched in domesticated genotypes, and
127 nearly fixed in modern fresh-market and processing types (**Fig. 2d)**. To genetically test if the
128 putative deleterious variant has an effect on the floral transition, we introgressed the ancestral
129 $SSP2^{S169}$ allele into a processing tomato type (cv. M82). We found that near-isogenic lines
130 (NILs) harboring $SSP2^{S169}$ flowered earlier on sympodial shoots and developed shoots that
131 grew more compact compared to the wild-type (WT) controls (**Supplementary Fig. 3a-f**). In
132 addition, we introduced $SSP2^{S169}$ into the hypomorphic ssp^{2129} mutant[15] to test whether
133 $SSP2^{S169}$ acts redundantly with its paralog *SSP*. We found that $SSP2^{S169}$ suppressed late-
134 flowering and indeterminate growth of ssp^{2129} mutants (**Supplementary Fig. 3g, h**), suggesting
135 that the ancestral $SSP2^{S169}$ allele can compensate for reduced *SSP* activity.

136 **Domesticated SSP2 is compromised as a transcription factor**

137 We hypothesized that the loss of the conserved serine residue affects the ability of SSP2 to
138 bind DNA during the regulation of target genes. We modelled the structure of the ancestral
139 ($SSP2^{S169}$) and domesticated ($SSP2^{F169}$) proteins in a homology-based modelling approach[25,26].
140 The model predicted that the conserved serine (S169) most likely forms hydrogen bonds with
141 the phosphate backbone of the DNA target sequence whereas a phenylalanine at this position
142 (F169) might increase the distance between the protein and target DNA due to its larger side-
143 chain and hydrophobicity (**Fig. 2e**). To test whether the amino acid exchange affects the
144 transcription factor function of SSP2, we co-expressed $SSP2^{F169}$, $SSP2^{S169}$ and SSP with SFT
145 in tobacco leaves to quantify their transactivation activity on the upstream regions of
146 *MACROCALYX* (*MC*; Solyc05g056620), *S. lycopersicum FRUITFULL1* (*SlFUL1*,
147 Solyc06g069430), and *SlFUL2* (Solyc03g114830). These genes are homologous to
148 Arabidopsis *APETALA1* and *FRUITFULL*, which have been shown to be activated by FD
149 during the floral transition[13]. None of the effector constructs activated the *MC* reporter, which
150 may result from a non-direct relationship between tomato *MC* and Arabidopsis *AP1*. In
151 addition, we observed only low or nonsignificant transactivation of the *SlFUL1* reporter.
152 However, the *SlFUL2* reporter was significantly activated by SSP and both SSP2 variants, and
153 transactivation was significantly higher by the ancestral $SSP2^{S169}$ compared to the $SSP2^{F169}$
154 variant (**Fig. 2f, Supplementary Fig. 3i**). Together with the structural and genetic analyses,
155 the results from this heterologous expression system suggested that the deleterious variant in
156 *SSP2* disrupts the DNA-binding domain of domesticated $SSP2^{F169}$ and compromises its
157 transcription factor function.

158 To determine whether the deleterious $SSP2^{F169}$ variant affects binding at genome-wide targets,
159 we performed DNA affinity purification sequencing (DAP-seq) with SSP, ancestral $SSP2^{S169}$
160 and domesticated $SSP2^{F169}$ as bait proteins[27]. We identified 14,091 DAP-seq peaks that were
161 significantly enriched ($\log_2FC \geq 3$, FDR ≤ 0.01) compared to the input controls (**Fig. 3a and**
162 **Supplementary Table 4**). The majority (7,388) of peaks were shared between SSP and the
163 ancestral $SSP2^{S169}$ but only 1,285 peaks were also bound by domesticated $SSP2^{F169}$. We
164 analyzed the genome-wide distribution of peaks for all three transcription factors and found
165 more than 50% of peaks within proximal regulatory regions (**Fig. 3b**). *De-novo* motif
166 enrichment analysis identified a G-box motif (CACGTG) with a subtle variation for $SSP2^{F169}$
167 outside the core-motif (**Fig. 3c**). Next, we analyzed genes with proximal peaks (≤ 3 Kbp
168 upstream and ≤ 2 Kbp downstream) and identified 7,688 and 5,010 putative target genes for

169 SSP and SSP^{S169}, of which the majority (4,699 genes) were bound by both proteins (**Fig. 3d**
170 and **Supplementary Table 5**). In contrast, domesticated $SSP2^{F169}$ bound only 1,175 and 1,136
171 of SSP and SSP^{S169} targets, respectively, and 1,616 genes in total. The low number of $SSP2^{F169}$
172 targets and shared targets with SSP and $SSP2^{S169}$ suggested that the ability of $SSP2^{F169}$ to bind
173 its genome-wide targets is compromised. To support this finding, we quantified binding
174 intensity at target regions based on normalized read coverage. While SSP and $SSP2^{S169}$
175 displayed similar binding intensities, $SSP2^{F169}$ binding was strongly reduced (**Fig. 3e, f** and
176 **Supplementary Fig. 4c, e**). Next, we conducted a differential binding analysis between
177 $SSP2^{S169}$ and $SSP2^{F169}$ (see Methods). Of the 9,199 SSP2 peaks, we identified 5,063
178 differentially bound by $SSP2^{S169}$ and $SSP2^{F169}$ ($\log_2 FC \geq 3$, $FDR \leq 0.01$) (**Supplementary Fig.**
179 **5a**). The majority (4,637) of differentially-bound peaks were private to $SSP2^{S169}$ while only
180 426 private peaks were identified for $SSP2^{F169}$ (**Supplementary Fig. 5b**). Private peaks for
181 $SSP2^{S169}$ were enriched for a G-box while a more relaxed TA-rich motif was enriched in
182 $SSP2^{F169}$ private peaks (**Supplementary Fig. 5c**). In addition, $SSP2^{F169}$ private peaks tended
183 to have a lower coverage, suggesting compromised binding ability also at private peaks
184 (**Supplementary Fig. 5d,e**). In support of this, we did not detect any enriched gene ontology
185 (GO) terms for the 230 genes near $SSP2^{F169}$ private peaks while the 3,046 genes near $SSP2^{S169}$
186 private peaks were enriched for genes related to circadian rhythm and rhythmic processes
187 (**Supplementary Fig. 5f,g**). Differential binding of $SSP2^{F169}$ and $SSP2^{S169}$ was also obvious at
188 the level of individual genes. For example, we found that the upstream regions of the two
189 tomato homologs of *GIGANTEA* (*GI*), which regulates photoperiodic flowering in
190 Arabidopsis[28], were bound by SSP and $SSP2^{S169}$ but not by the domesticated $SSP2^{F169}$ variant
191 (**Fig. 3g, h**). Together, our genome-wide binding data demonstrates that SSP and the ancestral
192 $SSP2^{S169}$ variant bind a set of largely shared targets while domesticated $SSP2^{F169}$ is
193 compromised in its ability to bind the targets of the ancestral protein.

194

195 ***SSP2* acts redundant with *SSP* to regulate plant architecture**

196 To genetically explore the function of *SSP2*, we used CRISPR-Cas9 genome editing and
197 generated $ssp2^{CR}$ and ssp^{CR} null mutants in two determinate cultivars (**Supplementary Fig. 6a-**
198 **b)**. The ssp^{CR} mutants flowered later than the WT and developed indeterminate shoots, which
199 confirmed previous findings that *SSP* promotes the floral transition (**Fig. 4a-c** and
200 **Supplementary 6c-d**)[15]. We did not observe obvious differences in flowering time for $ssp2^{CR}$
201 single mutants, which supports a diminished activity of $SSP2^{F169}$ in domesticated tomato (**Fig.**

202 **4a, c** and **Supplementary 6c-d**). However, *ssp*^{CR}*ssp2*^{CR} double mutants tended to flower later
203 than the *ssp*^{CR} single mutant, although at high variability (**Fig. 4c and Supplementary 6c, d**).
204 This phenotypic enhancement became more pronounced on sympodial shoots. Double
205 *ssp*^{CR}*ssp2*^{CR} mutants produced more leaves on sympodial shoots and more flowers on flowering
206 shoots (inflorescences) (**Fig. 4d, e**). We concluded that domesticated *SSP2*^{F169} is a partial loss-
207 of-function allele and that *SSP* and *SSP2* act partially redundant to promote the transition of
208 meristems to reproductive growth.

209 To obtain molecular insights into how *SSP* and *SSP2* promote meristem transitions, we
210 sequenced mRNA from micro-dissected meristems at the transition (TM) stage of meristem
211 maturation of the *ssp*^{CR} and *ssp2*^{CR} single and double mutants, and the WT (in cv. M82)[15].
212 Clustering of samples in a principal component analysis (PCA) was consistent with the mutant
213 phenotypes that indicated a delayed transition of *ssp*^{CR}*ssp2*^{CR} double mutants compared to the
214 *ssp*^{CR} single mutant (**Fig. 4f**). We identified 1,816 differentially expressed genes (DEGs) that
215 changed in expression by more than 1.5-fold in at least one of the mutants compared to the WT
216 (FDR $\leq$ 0.05) (**Supplementary Fig. 6e-f**). Of those, 591 (32.5%) were nearby DAP-seq peaks,
217 indicating that they are direct targets of SSP and/or SSP2 **(Fig. 4g)**. Clustering of the 591
218 putative direct targets revealed two main patterns of gene expression that contained genes
219 either down- or upregulated (de-repressed) in the *ssp*^{CR}*ssp2*^{CR} double mutant (**Fig. 4h,i and**
220 **Supplementary Table 6**). Among the downregulated genes, we found both tomato homologs
221 of the Arabidopsis floral promoter *GI*, and a homolog of its interactor *FLAVIN-BINDING,*
222 *KELCH REPEAT, F-BOX 1* (*FKF1*)[29]. In addition, the MADS-box gene *SlMBP10*, a homolog
223 of the Arabidopsis floral promoter *FUL*, was downregulated in *ssp*^{CR}*ssp2*^{CR} double mutants,
224 while *SlMBP14* and a *FLOWERING LOCUS C* (*FLC*)-like gene were de-repressed in
225 *ssp*^{CR}*ssp2*^{CR}. The MADS-box genes *SlFUL1* and *SlFUL2*, which we used for the initial reporter
226 assays (**Fig. 2f**), were not among the direct targets although both genes were significantly
227 downregulated in the *ssp*^{CR}*ssp2*^{CR} double mutant (**Supplementary Fig. 6g**), suggesting that
228 transactivation in the heterologous tobacco system was indirect. We also identified several
229 putative direct targets involved in phytohormone signaling. Two cytokinin
230 dehydrogenase/oxidase genes (*CKX1a*, *CKX5*) and putative negative regulators of cytokinin
231 levels were downregulated while a cytokinin activating enzyme encoding *SlLONELY GUY1*
232 (*SlLOG1*) gene was de-repressed in *ssp*^{CR}*ssp2*^{CR}. Furthermore, three abscisic acid receptor
233 genes (*PYLs*) were downregulated in the *ssp*^{CR}*ssp2*^{CR} double mutant. These data indicate that

234 *SSP* and *SSP2* redundantly regulate the expression of central regulators of the floral transition
235 and phytohormone responses, and guide meristem transitions towards floral fate.

236 We revisited the hypothesis that SSP2^{F169} neofunctionalized and compared the transcriptomes
237 of the WT (*SSP2^{F169}*) to *ssp2^{CR}* single mutants (*SSP2^{F169}* loss-of-function) but found only one
238 (Solyc11g071290, alcohol dehydrogenase) of the 230 private SSP2^{F169} DAP-seq targets
239 **(Supplementary Table 7**) to be differentially expressed (**Supplementary Fig. 5h**).
240 Furthermore, Solyc11g071290 was also differentially expressed in the *ssp^{CR}* mutant, indicating
241 that misexpression was not specific to loss of *SSP2^{F169}* activity (**Supplementary Fig. 5i**). The
242 finding that loss of *SSP2^{F169}* activity does not lead to specific misexpression of private SSP2^{F169}
243 targets in the *ssp2^{CR}* mutant suggests that the SSP2^{F169} variant did not neofunctionalize but
244 rather compromised its function at targets that are redundant with SSP.

245

246 ***SSP2* was lost during the evolution of *Physalis grisea***

247 To determine whether genetic redundancy between *SSP* and *SSP2* is evolutionary conserved,
248 we inspected orthologs across eudicots (**Supplementary Fig. 7**). Surprisingly, our
249 phylogenetic analyses indicated that tomato *SSP*/*SSP2* and Arabidopsis *FD*/*FDP* resulted from
250 independent duplication events that occurred after the *Solanacaeae* and *Brassicaceae* lineages
251 diverged (**Supplementary Fig. 7**). When we inspected protein sequences of SSP-like
252 transcription factors in the *Solanaceae*, we identified an additional missense mutation (alanine
253 to serine) in a conserved residue in the DNA-binding domain of the potato (*S. tuberosum*)
254 StSSP ortholog, which occurred after the divergence of tomato and potato (**Fig. 2c)**.
255 Furthermore, we found only one SSP-like ortholog in *Physalis grisea* (*PgSSP;*
256 Phygri02g013770), a relative of tomato in the *Solanoideae* subfamily[30]. Phylogenetic and
257 synteny analyses supported an evolutionary scenario in which the ortholog of *SSP2* was lost in
258 *P. grisea* (**Fig. 5a and Supplementary Fig. 8a-c**). To obtain experimental evidence for the
259 loss of redundancy in *P. grisea*, we mutated *PgSSP* by CRISPR-Cas and quantified effects on
260 shoot architecture (**Fig. 5b**). Wild-type *P. grisea* plants produce seven leaves on the primary
261 shoot before terminating in a single-flowered inflorescence (**Fig. 5c**). As in pepper (*Capsicum*
262 *spp*)[31], growth continues from two sympodial meristems that are released in the axil of the last
263 two leaves preceding the flower on the primary shoot. Each sympodial meristem gives rise to
264 one sympodial shoot unit, which results in a bifurcation of the shoot. The leaf subtending the
265 sympodial meristem is carried upwards due to strong internode elongation below the sympodial

266 unit and displaced above the preceding flower[31]. Each sympodial meristem produces two leaves
267 and one flower, and in turn releases two additional sympodial shoots. We observed alterations
268 to this pattern in two independent *Pgssp*^{CR} mutant lines, which produced an additional
269 sympodial shoot at the first bifurcation and grew less compact than the WT (**Fig. 5c, d-g**). The
270 additional sympodial shoot on *Pgssp*^{CR} mutants resulted from a sympodial meristem in the axil
271 of an extra leaf that was produced before flowering, which indicated that loss of *PgSSP* leads
272 to a mild but highly significant delay in flowering time (**Fig. 5e, h**). Together, these results
273 suggest that *PgSSP* regulates the transition of primary and sympodial meristems in the absence
274 of a direct paralog. However, *PgSSP* may act redundantly with other bZIP genes to regulate
275 the floral transition as recently reported in Arabidopsis[32].

276

277 **Repairing *SSP2* by base-editing leads to earlier fruit yield**

278 Our findings in tomato show that *SSP2* acts partially redundant with *SSP* to promote the
279 transition to flowering on sympodial shoots (**Fig. 4a, d**). We asked whether restoring the
280 activity of *SSP2* in domesticated tomato by correcting the deleterious variant would accelerate
281 the floral transition. We tested this hypothesis by repairing the deleterious variant in
282 domesticated tomato by CRISPR-Cas base editing. The critical non-synonymous mutation
283 results from a TCC (Ser) to TTC (Phe) codon exchange (**Fig. 6a**). The correction of this
284 mutation requires a A-to-G transition on the reverse strand, which can be induced with an
285 adenine base editor (ABE)[33]. Since none of the nearby canonical PAMs (NGG) allowed us to
286 position the target nucleotide into the high-activity editing window (A4-A8) of the protospacer,
287 we used the PAM-less Cas9 variant SpRY fused to ABE8e (**Fig. 6a**)[34]. We edited *SSP2* in the
288 domesticated and double-determinate S100 background[35] and observed high editing efficiency
289 with edits at the target adenine in 37.5 % (3 of 8) second-generation (T1) transgenic families
290 (**Fig. 6b** and **Supplementary Fig. 9a**). In one T1 family we also detected editing at the
291 bystander T position (**Supplementary Fig. 9a**). To determine whether the base-edited (be)
292 *ssp2*^{S169be} allele affected flowering time and shoot architecture, we generated a segregating (F4)
293 population descending from single, heterozygous F3 individual, and selected homozygous
294 (*ssp2*^{S169be}/*ssp2*^{S169be}) and heterozygous (*ssp2*^{S169be}/*SSP*^{F169}) individuals for the repaired allele,
295 and WT siblings (*SSP*^{F169}/*SSP*^{F169}) as controls by genotyping (**Fig. 6c**). We found that plants
296 homozygous or heterozygous for the repaired *ssp2*^{S169be} allele did not flower earlier than their
297 WT siblings (**Fig. 6d-e**). However, they developed less sympodial shoot units and less flowers
298 per inflorescence compared to WT siblings homozygous for the domesticated (*SSP2*^{F169}) allele,

299 which resulted in an overall more compact architecture (**Fig. 6d, f-g**). To assess if repair of
300 *SSP2* could compensate for the loss of *SSP*, we introduced the repaired *ssp2*^{S169be} allele into the
301 *ssp*^{CR} null mutant (in cv. S100). We found that *ssp2*^{S169be} did not suppress late flowering and
302 indeterminate growth of *ssp*^{CR} (**Fig. 6h-j**). However, we observed a partial and significant
303 suppression of late flowering on sympodial shoots (**Fig. 6h, k**). Moreover, *ssp*^{CR} *ssp2*^{S169be}
304 plants developed shorter inflorescences compared to *ssp*^{CR} mutants and WT (*SSP2*^{F169}) plants
305 (**Fig. 6l**). Together, these results demonstrate that functional *SSP2* accelerates the reproductive
306 transition of meristems on sympodial shoots in partial redundancy with *SSP*.

307 Tomato production was revolutionized during the 20th century by the *self-pruning* mutation,
308 which confers determinate growth and facilitates mechanical harvesting. Our findings showed
309 that a functional *SSP2* allele accelerates sympodial shoot flowering and thus suggested an
310 agronomic value for this allele regarding earliness for yield. To test whether accelerated
311 flowering from the repaired *ssp2*^{S169be} allele leads to earlier yield, we quantified fruit production
312 in segregating (F4) populations under experimental greenhouse conditions (see **Online**
313 **Methods**). We found that total fruit yields, harvest index, and fruit size for *ssp2*^{S169be} plants
314 were comparable to the WT sibling controls (**Fig. 6n-o** and **Supplementary Fig. 9b-e**).
315 Notably, *ssp2*^{S169be} homozygotes displayed an 8% increase in the proportion of ripe fruits
316 compared to WT siblings, which was likely due to precocious flowering and termination of
317 sympodial shoots (**Fig. 6m, p**). Thus, compact growth from repairing the deleterious *SSP2*
318 mutation by base editing can confer earliness for fruit yields and represents a new target for
319 customizing tomato shoot architecture. However, *ssp2*^{S169be} fruits had a reduced sugar content
320 (brix) (by 11%) (**Supplementary Fig. 9f**), suggesting trade-offs between compact growth with
321 soluble sugar content of fruits.

322

323 **DISCUSSION**

324 Here, we investigated the load of deleterious mutations that accumulated during domestication
325 and improvement of tomato. Within genes central to flowering time control, we discovered a
326 deleterious variant in the previously uncharacterized bZIP transcription factor gene *SSP2*. The
327 deleterious variant results in the exchange of a conserved serine to a phenylalanine in the DNA-
328 binding domain of the transcription factor. Our results from structural modelling, genome-wide
329 DNA binding assays, and genetic analyses indicate that the domesticated SSP2^{F169} variant
330 partially lost its ability to bind and regulate target genes that are largely shared between the

331 ancestral SSP2^{S169} variant and its paralog SSP. However, we cannot fully rule out that
332 domesticated SSP2^{F169} neonfunctionalized given its 230 private targets genes and ability to
333 bind both a G-box and an alternative TA-rich target motif, although at low affinity.
334 Interestingly, in the yeast bZIP factor Pap1, the equivalent serine-to-phenylalanine exchange
335 contributes to a similar change in binding specificity for TTACGTAA/TTAGTAA
336 sequences[36]. However, we found only very few target genes with this alternative motif bound
337 by SSP2^{F169} and none of them was specifically differentially expressed in the *ssp2^{CR}* mutant,
338 in which SSP2^{F169} activity is lost. We conclude from this data that the deleterious variant in
339 *SSP2* led to loss of genetic redundancy between *SSP* and *SSP2*, a pair of paralogs that is widely
340 conserved in flowering plants. In Arabidopsis, it was shown that *FD* and *FDP* act redundantly
341 during phytohormone responses while only *FD* affects the floral transition, suggesting
342 functional divergence of *FD*[21]. In contrast, our findings in tomato indicate that *SSP* and *SSP2*
343 act also partially redundant during the floral transition. Notably, our phylogenetic analyses
344 suggest that paralogs of *SSP* and *FD* arose independently in *Solanaceae* and *Brassicaceae*,
345 which could explain species-specific divergence of this paralogous pair. The complete loss of
346 a *PgSSP* paralog in *Physalis grisea* further supports dynamic evolution of the paralog pair.
347 Deleterious mutations and gene loss have been proposed as an important mechanism of
348 adaptation[6,7]. However, the benefit of the deleterious *SSP2^{F169}* variant during domestication
349 remains speculative. Our genetic data demonstrates that domesticated *SSP2^{F169}* delays
350 meristem transitions on shoots and inflorescences. Notably, the domesticated *SSP2^{F169}*
351 genotype develops more flowers per inflorescence than the ancestral *SSP2^{S169}* genotype.
352 Although flower number correlates with fruit yield, the number of flowers per inflorescence in
353 general decreased during tomato domestication, likely due to source-sink imbalances driven
354 by dramatic increases in fruit size[37]. This overall decrease in flower number during tomato
355 domestication suggests that effects from *SSP2^{F169}* on flower number were rather minor and
356 difficult to select. Furthermore, the deleterious *SSP2^{F169}* variant could have hitchhiked near
357 QTLs that were selected during domestication and improvement, which is a common scenario
358 in crops with a narrow genetic base such as tomato[38]. However, the closest known improvement
359 sweep on chromosome 2 with five fruit-weight QTLs is more than 5 Mbp away from *SSP2*,
360 rendering linkage unlikely[38]. Finally, we cannot exclude that *SSP2^{F169}* is adaptive under
361 specific conditions or developmental processes that were absent from our experiments given
362 that *CETS* genes play roles beyond flowering time control[39].

363 Whether *SSP2*^{F169} was nearly fixed in cultivated tomato due to selection or drift remains
364 therefore an open question. Yet, the loss of genetic redundancy caused by a deleterious
365 mutation may reflect a common feature during the selection of crops in human-made
366 environments. The less-is-more idea proposes the accumulation of loss-of-function mutations
367 as a driver of rapid evolutionary change[7], and gene loss may be even more frequent during the
368 intense artificial selection in domesticated environments. A reduced genetic repertoire in
369 domesticated genomes could result in lower genetic redundancy compared to their ancestral
370 states and, as a consequence, facilitate the exposure and selection of novel mutations, which
371 are otherwise masked by redundant paralogs. Our data shows that the ancestral *SSP2*^{S169} allele
372 can suppress effects of *ssp* mutations, which allow tuning of shoot architecture and
373 optimization of tomato yields[15]. Intriguingly, the deleterious *SSP2*^{F169} mutation, which broke
374 redundancy with the paralog *SSP*, may have been a prerequisite for the identification of the
375 *ssp*^{2129} breeding mutation. This illustrates how standing variants can become adaptive due to
376 genetic interactions with mutations that are introduced or arose during breeding.

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

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, G.S., 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.

COMPETING INTERESTS

The authors declare no competing interests.

414 **FIGURE LEGENDS**

415

416 **Figure 1: Predicting the load of deleterious variants along the domestication history of**
417 **tomato. a**, Number of predicted deleterious mutations in a panel of 82 tomato genomes,
418 including wild species (*S. pimpinellifolium*, green), landraces (*S. lycopersicum var.*
419 *cerasiforme*, orange), and cultivars (*S. lycopersicum*, purple). **b**, Prediction of deleterious
420 variants across all annotated *CETS* genes. The dashed red line indicates the threshold for
421 deleterious prediction (SIFT-score<0.05). Dot size scales with the number of genomes that
422 carry the variant. Black and grey font indicates genes with predicted deleterious and tolerated
423 mutations, respectively. **c**, Distribution of predicted deleterious *CETS* gene variants across wild
424 (*S. pimpinellifolium*; SP), landrace (*S. lyc var. cerasiforme*; SLC) and domesticated (*S.*
425 *lycopersicum*, SLL) genomes. Black and grey font indicate high and low prediction confidence,
426 respectively. **d,e**, Prediction of deleterious variants in all annotated Group-A bZIP genes (d)
427 and their distribution across different genome categories (e) as in (b, c).

428

429 **Figure 2: A deleterious mutation in *SSP2* reduces its transcription factor activity. a**,
430 Maximum-likelihood tree of A-group bZIP proteins in tomato (red font) and Arabidopsis (blue
431 font). Red arrowhead marks SSP2. Numbers represent bootstrap values from 1,000 replicates
432 and scale bar indicates the average number of substitutions per site. **b**, Normalized gene
433 expression (TPM) for *SSP* and *SSP2* in different tissues and developmental stages (veg.
434 earl./mid./late, stand for early, middle and late vegetative meristem stage). **c**, Partial alignment
435 of SSP-like bZIP proteins from Arabidopsis, domesticated tomato (*S. lycopersicum*; *Slyc*),
436 close wild tomato relative (*S. pimpinellifolium*; *Spim*), distant wild tomato relative (*S. pennellii*;
437 *Spen*), potato (*S. tuberosum*; *St*), and *Physalis grisea* (*Pg*). Red arrowheads mark conserved
438 DNA-binding residues. **d**, Distribution of ancestral ($SSP2^{S169}$) and derived ($SSP2^{F169}$) *SSP2*
439 alleles in distant wild tomato relatives, wild relatives (*S. galapagense* / *S. cheesmaniae*), wild
440 progenitor species (*S. pimpinellifolium*), landraces (*S. lyc var. cerasiforme*), and cultivars (*S.*
441 *lycopersicum*). n=number of accessions. **e**, Predicted structures of ancestral $SSP2^{S169}$ and
442 derived $SSP2^{F169}$ proteins on target DNA determined by homology modelling. Insets show a
443 magnified view of the serine/phenylalanine residue at position 169. **f**, Reporter assays in
444 tobacco leaves using SSP, $SSP2^{F169}$, and $SSP2^{S169}$ as effectors and firefly Luciferase (fLuc)
445 driven by upstream sequences of *MC* (*pMC::fLUC*), *SlFUL1* (*pSlFUL1::fLUC*), and *SlFUL2*

446 (*pSlFUL1::fLUC*) as reporter. Numbers indicate biological replicates (transformed leaves).
447 Ctrl indicates the no effector control (35S::YFP). Letters represent the results from pairwise
448 comparisons of means using one-way ANOVA and post-hoc Tukey's HSD test with 95%
449 confidence level. For box plots in (f), the bottom and top of boxes represent the first and third
450 quartile, respectively, the middle line the median and the whiskers the maximum and minimum
451 values.

452

453 **Figure 3: Domesticated SSP2^{F169} shows reduced binding at genome-wide target loci. a**,
454 Overlap of significant ($\log_2$FC ≥ 3, FDR ≤ 0.01) SSP, SSP2^{F169}, and SSP2^{S169} DAP-seq peaks
455 (n=14'091). **b**, Distribution of significant SSP, SSP2^{F169}, and SSP2^{S169} DAP-seq peaks across
456 gene features. **c**, Most-significant motifs identified by *de-novo* motif enrichment analysis of
457 SSP, SSP2^{F169}, and SSP2^{S169} DAP-seq peak regions. Grey box delimits region with motif
458 variation outside the core-motif. **d**, Overlap of genes with significant DAP-seq peaks ≤ 3 Kbp
459 upstream and ≤ 2 Kbp downstream of the transcriptional start site (n=8'395). **e**, Profiles of
460 normalized read coverage at significant SSP, SSP2^{F169}, and SSP2^{S169} peaks. **f**, Comparison of
461 SSP, SSP2^{F169}, and SSP2^{S169} DAP-seq peaks relative to the transcriptional start (TSS) and end
462 (TES) site of nearby genes (n=8'395). **g-h**, Browser view of SSP, SSP2^{F169}, and SSP2^{S169} DAP-
463 seq peaks at *SlGIGANTEA-LIKE1* (g) and *SlGIGANTEA-LIKE2* (h). Normalized coverage
464 (CPM) is shown in yellow, green and blue. Significant peak regions are indicated by red boxes.

465

466 **Figure 4: *SSP* and *SSP2* act partially redundant to regulate the transition to flowering. a**,
467 Representative images of wild-type S100, *ssp^{CR}* and *ssp2^{CR}* single mutants, and *ssp ssp2^{CR}*
468 double mutants. Digits in superscript indicate the number of deleted (CR-) or inserted (CR+)
469 nucleotides. All allele sequences are found in **Supplementary Figure 6**. L= leaf number,
470 arrowheads mark the last leaf before flowering. Determinate (D) and indeterminate (ID) shoots
471 are indicated. Scale bars represent 7.5 cm. **b**, Schematic depiction of tomato shoot architecture.
472 Different shades of green delimit primary and sympodial shoots. **c-e**, Quantification of the
473 floral transition (number of leaves before flowering) on the primary (c) and secondary (d)
474 shoots, and the number of flowers per inflorescence (e) for genotypes shown in (a). The number
475 of plants (c,d) and inflorescences (e) are indicated. Letters represent the results from pairwise
476 comparisons of means using one-way ANOVA and post-hoc Tukey's HSD test with 95%
477 confidence level. **f**, Principal component analysis of 22'726 expressed genes in transition

478 meristems of the WT, *ssp, ssp2*, and *ssp ssp2*, determined by RNA-seq. **g**, Overlap of genes
479 differentially expressed ($\log_2$FC ≥ 0.58, FDR ≤ 0.05) in *ssp*, *ssp*2, and/or *ssp ssp2* with genes
480 at SSP, SSP2^{F169}, and SSP2^{S169} DAP-seq peaks. **h**, Heatmap depicting expression of 591
481 putative SSP/SSP2 target genes. **i**, Normalized expression (TPM, transcript per million) for
482 selected putative direct targets in 3 independent biological replicates (black dots) and mean
483 values (colored dots). Error bars indicate standard error. Genes are color coded based on the
484 biological pathway. For box plots in (c-e), the bottom and top of the boxes represent the first
485 and third quartile, respectively, the middle line the median and the whiskers the maximum and
486 minimum values.

487

488 **Figure 5: The genome of *Physalis grisea* encodes a single direct *SSP* ortholog that**
489 **regulates meristem transitions. a**, Schematic phylogenetic tree of tomato and closely related
490 *Solanaceae* species. Filled circles, empty circles or stars show presence, absence, or missense
491 mutation, respectively, of *SSP*/*SSP2* or *FD/ FDP* in these species. Full tree is displayed in
492 **Supplementary Fig. 7**. **b**, CRISPR-Cas9 targeting of *PgSSP* in *P. grisea*. Blue boxes, black
493 lines, and grey boxes represent exonic, intronic, and untranslated regions, respectively. Single
494 guide RNAs (sgRNAs) are indicated with red arrowheads. PAM and sgRNA sequences are
495 indicated in black and red bold letters, respectively; deletions are indicated with blue dashes;
496 sequence gap length is given in parenthesis. Insertions are indicated by blue letters. **c**, Model
497 of the sympodial growth habit of *P. grisea* WT and *Pgssp^{CR}* plants. Different shades of green
498 delimit primary, first sympodial, and second sympodial shoots. The color of leaves on the
499 primary (LP) and sympodial (LS) shoots corresponds with the shoot of origin **d**, Representative
500 pictures illustrating differences in number of sympodial shoots in WT and *Pgssp* mutants. Last
501 leaf before shoot bifurcation (LP5) and displaced leaves (LP6-8) are indicated (L5). White
502 arrowheads indicate individual sympodial units (SU). Scale bar represents 7.5 cm. **e**,
503 Representative stereoscope images of the shoot apex of WT and *Pgssp* mutants. Upper images
504 show the apex with a terminal flower (*). Lower images show the same view with the flower
505 removed. The sympodial meristems (SYMs) and position of removed primary shoot leaves
506 (LP) are delimited by dashed and dotted lines, respectively and numbered in developmental
507 order. Scale bar represents 100 µm. **f-h**, Quantification of the number of sympodial shoots at
508 the first (f) and second (g) bifurcation, and flowering time (number of leaves before the first
509 inflorescence) (h). Number of plants is indicated at the bottom of plots. Letters represent results
510 from pairwise comparisons of means using one-way ANOVA and post-hoc Tukey's HSD test

with 95% confidence level. For box plots in (f-h), the bottom and top of boxes represent the first and third quartile, respectively, the middle line the median and the whiskers the maximum and minimum values.

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 *SSP2* mutation 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 the deaminated adenine. **b**, Validation of editing in a chimeric first-generation (T0) transgenic and corresponding T1 progeny by Sanger sequencing. The target nucleotide is marked by a red arrowhead. **c**, Crossing scheme to generate the segregating $ssp2^{S169be}$ F4 population. **d**, Representative pictures showing the number of sympodial units on WT and $ssp2^{S169be}$ plants. Terminal inflorescences of sympodial units are indicated by white arrowheads. **e-g**, Quantification of the number of leaves before the first inflorescence (flowering time) (e), sympodial shoots (f), and flowers per truss (g) 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 (i) (as in (e)), number of determinate plants (j), leaves per sympodial unit (SU) (k), and flowers per truss (l) 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 complete 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 pairwise comparisons of means using one-way ANOVA and post-hoc Tukey's HSD test with 95% confidence level**.** Scale bars represent 10 cm (d) and 7.5 cm in (h,m). For box plots in (e-g, i, k, l, n-p), the bottom and top of boxes represent the first and third quartile, respectively, the middle line the median and the whiskers the maximum and minimum values.

REFERENCES

Koenig, D. *et al.* Comparative transcriptomics reveals patterns of selection in domesticated and wild tomato. *Proceedings of the National Academy of Sciences* **110**, E2655–E2662 (2013).

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

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

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. *Annu Rev Genet* **52**, 421–444 (2018).

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

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).

Olson, M. V. When Less Is More: Gene Loss as an Engine of Evolutionary Change. *The American Journal of Human Genetics* **64**, 18–23 (1999).

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

Gaarslev, N., Swinnen, G. & Soyk, S. Meristem transitions and plant architecture—learning from domestication for crop breeding. *Plant Physiol* **187**, 1045–1056 (2021).

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).

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* **5**, 465 (2014).

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).

Zhu, Y. *et al.* TERMINAL FLOWER 1-FD complex target genes and competition with FLOWERING LOCUS T. *Nat Commun* **11**, 5118 (2020).

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

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

574. Alonge, M. *et al.* Major Impacts of Widespread Structural Variation on Gene Expression and Crop Improvement in Tomato. *Cell* **182**, 145-161.e23 (2020).

576. Vaser, R., Adusumalli, S., Leng, S. N., Sikic, M. & Ng, P. C. SIFT missense predictions for genomes. *Nat Protoc* **11**, 1–9 (2016).

578. Soyk, S. *et al.* Variation in the flowering gene SELF PRUNING 5G promotes day-neutrality and early yield in tomato. *Nat Genet* **49**, 162–168 (2017).

580. 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).

583. Abe, M. *et al.* FD, a bZIP protein mediating signals from the floral pathway integrator FT at the shoot apex. *Science* **309**, 1052–6 (2005).

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

587. Wigge, P. A. *et al.* Integration of spatial and temporal information during floral induction in Arabidopsis. *Science* **309**, 1056–9 (2005).

589. 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).

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

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

596. Tahirov, T. H. *et al.* Structural Analyses of DNA Recognition by the AML1/Runx-1 Runt Domain and Its Allosteric Control by CBFβ. *Cell* **104**, 755–767 (2001).

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

28. Fowler, S. 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* **318**, 261–5 (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. Elitzur, T. *et al.* Co-ordinated regulation of flowering time, plant architecture and growth by FASCICULATE: the pepper orthologue of SELF PRUNING. *J Exp Bot* **60**, 869–880 (2009).

32. Martignago, D. *et al.* The bZIP transcription factor AREB3 mediates FT signalling and floral transition at the Arabidopsis shoot apical meristem. *PLoS Genet* **19**, e1010766 (2023).

33. Richter, M. F. *et al.* Phage-assisted evolution of an adenine base editor with improved Cas domain compatibility and activity. *Nat Biotechnol* **38**, 883–891 (2020).

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

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

36. 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* **7**, 889–893 (2000).

37. Grandillo, S. & Tanksley, S. D. QTL analysis of horticultural traits differentiating the cultivated tomato from the closely related species Lycopersicon pimpinellifolium. *Theoretical and Applied Genetics* **92**, 935–951 (1996).

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

39. Shalit-Kaneh, A. *et al.* The flowering hormone florigen accelerates secondary cell wall biogenesis to harmonize vascular maturation with reproductive development. *Proceedings of the National Academy of Sciences* **116**, 16127–16136 (2019).

629 **ONLINE METHODS**

630 Plant material, growth conditions, and phenotyping

631 Seeds of *S. lycopersicum* cv. M82 (LA3475), *S. lycopersicum* cv. Sweet-100 (S100) double-
632 determinate[35], *S. pimpinellifolium* (LA1589), *P. grisea,* and *N. benthamiana* were from our
633 own stocks. Tomato seeds were germinated on soil in 96-cell plastic flats. *P.grisea* seeds were
634 incubated at 50°C for 3 days prior to sowing to increase germination rates. Plants were grown
635 under long-day conditions (16-h light/ 8-h dark) in a greenhouse supplemented with artificial
636 light from high-pressure sodium bulbs (~250umol m-2s-1). Temperature was 25°C and relative
637 humidity was 50-60%. Tomato plants were grown in 5L pots (2 plants per pot) and physalis in
638 0.36 l pots (one plant per pot) under drip irrigation and standard fertilizer regimes. Tomato
639 plants were pruned down to the primary and proximal axillary shoots. Phenotypic data was
640 collected from the F3 and T4 generation for ssp^{CR} $ssp2^{CR}$ plants in the S100 background, the
641 F7 (ssp^{CR} and $ssp2^{CR}$) and F4 (ssp^{CR} $ssp2^{CR}$) generation in the M82 background, and the T3
642 generation for $PgSSP^{CR}$ in *Physalis* and the F4 generation in *ssp* $ssp2^{S169be}$ plants. Data for
643 flowering time, sympodial shoot number, number of leaves per sympodial shoot, and number
644 of flowers per inflorescence were collected from primary and proximal shoots. To assess
645 tomato yield components under experimental greenhouse conditions, mature plants were
646 harvested 79 days after transplanting. For data collection, plants and fruits were manually
647 removed from the soil and the plant, respectively. Total fruit yield was defined as the sum of
648 red and green fruits from each plant. Harvest index was calculated by dividing total fruit yield
649 by plant weight (i.e., vegetative biomass after fruit removal). Ten fruits per plant were
650 randomly selected to measure average fruit weight and total soluble sugar content (brix). Brix
651 was quantified using a digital Brix refractometer (HANNA® instruments, HI96801). Statistical
652 analyses of phenotyping data were conducted in R[40].

653 *N. benthamiana* (tobacco) seeds were sown on soil in square pots. Seedlings were grown under
654 long-day conditions (16-h light/ 8-h dark) in a growth room under LED light (~120 umol m-
655 2s-1) and constant temperature (25°C). Approximately one week after germination, tobacco
656 seedlings were singled out into square pots.

657

658 LA1589 de novo genome assembly

659 Nanopore long read sequences for the *S. pimpinellifolium* accession LA1589 were previously
660 generated[16,41]. Basecalling was performed using Guppy v3.1.5. Illumina sequencing data were

661 previously generated[24]. We assembled the Nanopore and Illumina sequences together with
662 MaSuRCA (v3.4.1)[42]. Resulting contigs were scaffolded against the Heinz 4.0 reference
663 genome using RaGOO (v1.1)[43]. Gaps were closed with LR_Gapcloser (v3)[44] and the assembly
664 was polished with 3 rounds of Pilon (v1.23)[45]. Assembly statistics are found in **Supplementary**
665 **Table 8**. We used liftoff[46] to annotate the LA1589 assembly with ITAG4.0 gene models and
666 tomato pan-genome genes as previously described[35].

667

668 Genome-wide prediction of deleterious variants

669 Illumina raw reads from 27 *S. pimpinellifolium*, 23 *S. lycopersicum* var. *cerasiforme*, and 32 *S.*
670 *lycopersicum* accessions (**Supplementary Table 1**) were retrieved from public repositories as
671 described before[47]. Reads were aligned to the *S. pimpinellifolium* reference genome
672 (LA1589v0.1) using BWA-MEM (v0.7.17) using default parameters. Alignments were sorted
673 and duplicates marked with PicardTools (v2.26.2) and indexed using samtools (v1.15.1)[48].
674 Variants were called with bcftools (v.1.15.1, parameters mpileup --no-BAQ --ignore-RG -d
675 1000000 -Q0 --annotate FORMAT/AD,FORMAT/DP). Variants were filtered with vcftools
676 (v0.1.14, parameters --min-alleles 2 --max-alleles 2 --minQ 30 --minDP 5 --maxDP 50 --mac
677 2 --recode --recode-INFO-all). Filtered variant call format (vcf) files were used to predict
678 deleterious mutations using SIFT-4G[17]. A SIFT database was built from the *S. pimpinellifolium*
679 genome sequence (SpimLA1589_v0.1) and annotation (SolpimLA1589_v0.2) using the SIFT
680 instructions and default parameters. The LA1589 SIFT database contained SIFT scores for
681 70% of genes (21578 of 30808), SIFT scores for 83% of positions (56424493/67919880), and
682 confident scores for 73% of positions (41083097/56424493). SIFT was used to determine the
683 effect of coding sequence variants on protein sequence, and to predict deleterious missense
684 variants. Variant types and SIFT scores were plotted in R using the ggplot2 package.

685

686 Phylogenetic analyses and sequence alignments

687 Protein sequences of tomato and Arabidopsis bZIP family members were obtained from the
688 Plant Transcription Factor Database (PlantTFDB, v5.0)[49]. Physalis bZIP protein sequences
689 were identified in a BLAST search on the Phygri1.3.1 protein annotation[30] using SSP protein
690 as query. Full-length sequences of 70 tomato, 74 Arabidopsis, 58 Physalis, and yeast Pap1
691 (SPAC1783.07c.1) bZIP proteins were aligned with MAFFT (v7.481) using default
692 parameters[50]. Maximum likelihood phylogenetic trees were constructed in IQ-Tree (v2.2.0.5;

693 parameters -m MFP -bb 1000 -bnni -redo)[51] and visualized in FigTree (v1.4.4;
694 http://tree.bio.ed.ac.uk/software/figtree/). Average number of substitutions per site is indicated
695 by scale bars. Specific bZIP groups were assigned according to their Arabidopsis homologs[52].

696 To reconstruct the phylogenetic tree of the bZIP family in eudicots we used the OMA
697 browser's[53] 07/2023 release to collect homologs for tree building. Hierarchical Orthologous
698 Groups (HOGs) were identified by searching for the SSP gene identifier (Solyc02g083520) for
699 the initial HOG and adding additional closely related HOGs, inferred to be closely related as
700 they share many predicted orthologs. The following HOGs were downloaded: D0228852,
701 D0178917, D0181214, D0210160, D0214417, D0216285, D0223413 (accessed 23 Jan 2024).
702 Additionally, through BLAST searches, we incorporated the bZIP gene of *Amborella*
703 *trichopoda* and closely related bZIP genes from eight *Solanaceae* species: *Nicotiana*
704 *benthamiana*, *Nicotiana tabacum*, *Physalis grisea*, *Petunia axillaris*, *Petunia inflata*, *Solanum*
705 *tuberosum*, *Capsicum annuum*, and *Capsicum chinense*. The final dataset comprised 128 genes
706 from 51 plant species. These protein sequences were aligned using the approach described in
707 the PhylomeDB pipeline[54]. Briefly, we obtained alignments in forward and reverse directions
708 using three programs (MUSCLE v3.8.1551[55], MAFFT v7.490[50], and Kalign v3.3.5[56]). Then,
709 the six alignments were combined using M-COFFEE v13.46.0.919e8c6b[57]. The phylogenetic
710 tree was reconstructed using a maximum likelihood approach in IQ-TREE v2.2.2.6[58], using the
711 best-fit model identified by ModelFinder[59] (JTT+F+I+R5) and 1000 ultrafast bootstrap
712 replicates. The tree was manually rooted using *Amborella trichopoda* as outgroup. Duplication
713 events were inferred using ETE v4.0[60] and the species overlap method[61].

714

715 Homology modelling

716 The HHpred server was used to find suitable templates for SSP2 protein modeling[62]. The final
717 templates were chosen based on sequence similarity in the area of protein-DNA interaction.
718 The 50 homology models of wild tomato protein $SSP2^{S169}$ dimers were calculated using
719 Modeller 9v18[25] and CCAAT/enhancer-binding protein beta (C/EBP beta) as a template. The
720 crystal structure of human C/EBP beta in complex with DNA is stored under 1HJB code in the
721 Protein Data Bank[26]. The target and template sequence shared 26% of sequence identity. The
722 best model in terms of its DOPE score[63] was chosen. Analogically, the 50 homology models
723 of domestic tomato $SSP2^{F169}$ protein dimers were calculated based on the structure of Pap1 as
724 a template and the best model, according to DOPE score, was chosen. The crystal structure of

725 Pap1 is stored in the PDB under 1GD2 code and shares 24% of sequence identity with the
726 SSP2^{F169} protein[36]. For both SSP2 proteins the DNA molecule from the template structure was
727 included in the models. The DNA sequence was changed to the SSP2 recognition motif with
728 UCSF Chimera tool that was also used for visualization of the models[64].

729

730 Molecular cloning

731 Binary vectors for CRISPR-Cas9 mutagenesis in tomato were assembled using the Golden
732 Gate cloning system as previously described[35,65]. For CRISPR-Cas9 mutagenesis, a new Level
733 (L) 1 part pICH47742_SpCas9-P2A-GFP was cloned by amplifying SpCas9 from
734 pICH47742::35S::Cas9 (Addgene no. 49771) using primers P94 and P129. The fragments were
735 cloned into the L0 acceptor pAGM1287 to generate pAGM1287-SpCas9. P2A-GFP was
736 amplified from pGG-D-P2A-GFP-NLS-E[66] using primer P96 and P97 and cloned into the L0
737 acceptor pAGM1301 to generate pAGM1301_P2A-GFP. The pAGM1287_SpCas9 and
738 pAGM1301_P2A-GFP parts were combined with pICH51288 (2Xp35S) and pICH41421
739 (nosT) in pICH47742 (L1 acceptor) to generate pICH47742_SpCas9-P2A-GFP. For CRISPR-
740 Cas base editing, the PAM-less adenosine base editor ABE8e-SpRY[34] was domesticated by
741 amplifying four fragments using the primer pairs P576/ P577, P578/ P579, P580/ P581,
742 P582/P583 on the template pYPQ262B[34]. Fragments were cloned into the L-1 acceptor
743 pAGM1311 and combined in the L0 acceptor pAGM1287 to generate pAGM1287_ABE8e-
744 SpRY. pAGM1287_ABE8e-SpRY was combined with pAGM1301_P2A-GFP, pICH51288
745 (2Xp35S), and pICH41421 (nosT) in the L1 acceptor pICH47742 to generate
746 pICH47742_SpRY-ABE8e-P2A-GFP. Constructs for transactivation assays were cloned using
747 the MoClo kit[65]. The p19 construct for silencing suppression was assembled with the L1
748 acceptor pICH47742 and L0 parts pICH85281 (pMas), pICH44022 (p19), and pICH77901
749 (tMas). The YFP construct was assembled with the L1 acceptor pICH47742 and L0 parts
750 pICH51266 (p35S), pICSL80014 (YFP), and pICH41414 (t35S). To clone the SFT co-effector
751 and the SlycSSP2 effector constructs, the coding sequences of SFT and SlycSSP2 were
752 amplified from *S. lycopersicum* (cv. M82) transition meristem cDNA with gene specific primer
753 pairs (SFT: SFT_F/SFT_R, SlycSSP2: SSP2_F/SSP2_R). To clone the SpimSSP2 effector
754 construct, the coding sequence of SpimSSP2 was amplified from *S. pimpinellifolium* (LA1589)
755 transition meristem cDNA with the primer pair SSP2_F/SSP2_R. The amplicons were cloned
756 into the L0 acceptor pICH41308. Individual L0 effector parts (SlycSSP2, and SpimSSP2) were
757 combined with pICSL13001 (p35S), pICSL30009 (Myc-tag), and pICH41414 (t35S) in the L1

758 acceptor pICH47772. The L0 co-effector part (SFT) was combined with pICSL13001 (p35S),
759 pICSL30008 (HA-tag) and pICH41414 (t35S) in the L1 acceptor pICH47761. To clone the
760 SSP effector construct and the 35S::YFP no-effector construct the coding sequence of SSP was
761 amplified from *S. lycopersicum* (cv. M82) transition meristem cDNA with the primer pair
762 SSP_F/SSP_R and the coding sequence of 35S::YFP was amplified from pENTR_L1-Turbo-
763 YFP-NLS-STOP-L2 (addgene no. 127361) with the primer pair YFP_F/YFP_R. Both
764 amplified coding sequences were cloned into the L1 acceptor pICH47772 that was amplified
765 with the primer pair TEA_F/TEA_R with the NEBuilder HiFi DNA Assembly Cloning Kit
766 (NEB #E5520). To clone the luciferase reporter constructs, upstream regions of pMC, pFUL,
767 and pFUL2 were amplified from *S. lycopersicum* (cv. M82) gDNA in multiple fragments gene-
768 specific primer pairs (pMC: pMC_F1/pMC_R1 and pMC_F2/pMC_R2; pFUL1:
769 pFUL1_F3/pFUL1_R3 pFUL1_F2p/FUL1_R2p, and FUL1_F1/pFUL1_R1; pFUL2:
770 pFUL2_F1/pFUL2_R1 and pFUL2_F2/pFUL2_R2) and cloned into the L-1 acceptor
771 pAGM1311. The pMC construct contained 2170 bp genomic sequence including upstream
772 region, the 5'UTR, and the first exon. The pFUL1 and pFUL2 constructs contained 2640 bp
773 and 2040 bp genomic sequence, respectively, including upstream regions and the 5'UTR. The
774 L-1 parts were cloned into the L0 acceptor pICH41295. Individual L0 effector parts (pMC,
775 pFUL1, and pFUL2) were combined with pICSL80001 (fLuc) and pICH41432 (tOCS) in the
776 L1 acceptor pICH47751. All primers and gRNA sequences are listed in **Supplementary Table**
777 **9 and 10.**

778

779 <u>CRISPR/Cas9 genome editing, plant transformation and identification of mutant alleles</u>

780 CRISPR-Cas9 mutagenesis in tomato and physalis was performed as described
781 previously[35,67,68]. Briefly, guide RNAs (gRNAs) were designed using the CRISPOR tool and
782 the M82v1.0, Sweet-100v2.0 or Phygriv1.0 genome assemblies. Final vectors were
783 transformed into the tomato cultivar M82, LA1589 or double-determinate Sweet-100, or into
784 *P. grisea* by A*grobacterium tumefaciens*-mediated transformation. CRISPR-Cas9 editing in
785 tomato and physalis was verified by genotyping or amplicon sequencing as described[35]. Base
786 editing was quantified in first-generation (T0) transgenics using EditR v1.0.10[69] and in the T1
787 generation with a CAPS marker (**Supplementary Table 9)**.

788

789

790 Generation of near-isogenic lines (NILs)

791 Near-isogenic *SSP2*^{S169} lines in the M82 background were generated by crossing the *S.*
792 *pimpinellifolium* accession LA1589 with *S. lycopersicum* cv. M82, and backcrossing F2
793 individuals homozygous for *SSP2*^{S169} to the recurrent parent (*S. lyc.* cv. M82) over 4 (BC4) to
794 5 (BC5) generations. Presence *SSP2*^{S169} allele was confirmed by CAPS marker
795 (**Supplementary Table 9**).

796

797 Transactivation assays

798 Transient transactivation assays with luciferase reporter constructs were conducted in *N.*
799 *benthamiana* leaves as previously described[70]. In brief, leaves of 34 and 36 days old plants
800 were infiltrated with mixtures of *A. tumefaciens* (strain GV3101) cultures containing effector,
801 co-effector, luciferase reporter, transfection control, and silencing inhibitor vectors. Effector
802 constructs contained the coding sequence (CDS) of *SSP*, *SSP2*^{F169} or *SSP2*^{S169} with an N-
803 terminal Myc tag and driven by the CaMV 35S promoter. The co-effector construct contained
804 the CDS of *SFT* with an N-terminal HA tag and driven by CMV 35S promoter. The luciferase
805 reporter constructs contained the CDS of fLUC driven by the upstream regions of *MC*, *SlFUL*,
806 or *SlFUL2*. The transfection control was pGREENII-0800-LUC, which contains the CDS of
807 rLUC driven by the CMV 35S promoter. A 35S::YFP construct was used as no effector control.
808 A p19 construct was used to suppress silencing. Liquid cultures were grown in 10 ml YEB in
809 50 mL Falcon tubes for 20 hrs at 28°C, shaking. Agrobacteria were harvested by centrifugation
810 at 1932 x g and resuspended in infiltration buffer (50 mM MES pH 5.7 and 10 mM $MgCl_2$, 200
811 µM Acetosyringone) to an OD_{600} = 0.5. Before leaf infiltration, individual cultures were
812 incubated 3 hrs at RT and combined to obtain mixtures with effectors, reporters (fLUC), and
813 transfection control (pGREEN 35S:rLUC), and silencing inhibitor (p19) plasmids at final
814 OD_{600} of 0.1, 0.1, 0.1, and 0.05. Agrobacteria mixtures were infiltrated into the 4th, 5th and 6th
815 leaf using a needleless syringe, with three different plants being infiltrated for each
816 combination. Leaf disks of 0.8 cm diameter were harvested 3 days after infiltration and flash-
817 frozen in liquid nitrogen before grinding in a mix mill (twice 20 s^{-1} for 30s). Luciferase assays
818 were performed using the Dual-Luciferase Reporter Assay System (Promega) and a Hidex
819 Sense Microplate Reader. In short, leaf powder was extracted in 300 µl of 1x PLB and
820 vigorously vortexed for 30 s. Volumes of 10 µl protein extracts were mixed with 40 µl
821 luciferase reagent in 96-well microplates and incubated for 10 min at RT. Firefly luciferase

822 (fLUC) activity was quantified (9 s integration time). Reactions were mixed with 25 µl Stop &
823 Glo. Renilla luciferase (rLUC) activity was measured (9 s integration time). Transactivation
824 by the effectors was determined by calculating the fLUC/rLUC ratios.

825

826 DAP-seq

827 Myc-tagged coding sequences of *SSP*, $SSP2^{F169}$ and $SSP2^{S169}$ were amplified from effector
828 constructs used in the transactivation assay. The pTnT™ vector, and the $SSP2^{F169}$ and $SSP2^{S169}$
829 inserts were digested using XhoI (NEB) and NotI-HF (NEB) and combined using T4 Ligase
830 (NEB). The Myc-tagged coding sequence for *SSP* was cloned into pTnT™ vectors with the
831 NEBuilder HiFi DNA Assembly Cloning Kit (NEB #E5520). Plasmid DNA was isolated from
832 100 ml bacterial cultures using the PureYield™ Plasmid Midiprep System (Promega, A2492).
833 Two replicates of SSP and SSP2 proteins were expressed *in-vitro* in the TnT® SP6 High-Yield
834 Wheat Germ Protein Expression System (Promega, L3260) from 3.5 µg plasmid DNA per
835 reaction. Protein expression and solubility was verified by western blotting using primary anti-
836 Myc (Merck, clone 4A6, 1:5000) and secondary anti-mouse (Amersham ECL NA931,
837 1:10,000) antibodies, according to the manufacturer's instructions. High molecular weight
838 DNA for genomic library construction was isolated from inflorescence meristem tissue of the
839 *anantha* mutant in Sweet-100 using a CTAB protocol as described previously[35]. DAP-seq was
840 performed as previously described with minor modifications[27,71]. The DNA-library was
841 prepared according to Franco-Zorilla & Prat (2021) with minor modifications. The gDNA
842 library was purified using SPRI beads (B23317, Beckman Coulter). Adaptor ligation was
843 verified by qPCR with index-specific primers (**Supplementary Table 9**) and KAPA standards
844 20, 2 and 0.2 nM (Roche) in 10 µl reactions. DNA affinity-purification steps were performed
845 according to Bartlett et al. (2017) with 75 ng of gDNA input library per replicate. Eluted
846 libraries were single-indexed (**Supplementary Table 9**). Eight uniquely-indexed libraries
847 were produced, two replicate libraries per protein (SSP, $SSP2^{F169}$, $SSP2^{S169}$) and two replicates
848 of the input library as negative control. Indexed libraries were purified individually with the
849 Monarch® PCR&DNA Cleanup Kit (NEB, T1030S). Individual indexed libraries were
850 analyzed on a Fragment Analyzer (Agilent), purified with SPRI beads and pooled at equimolar
851 (10 nM) concentrations. Pooled libraries were sequenced on one Illumina NovaSeq6000 lane
852 at the Genome Technology Facility (GTF) of the University of Lausanne. A total of
853 753'327'838 PE150 reads (between 64'808'988 and 144'444'123 per sample) were generated.

854 Raw read quality was assessed using FastQC (v0.11.9;
855 http://www.bioinformatics.babraham.ac.uk/projects/fastqc/). Adapter sequences were trimmed
856 with NGmerge (v0.3, parameters -g -d -a)[72]. Reads were aligned to the SollycSweet-100v2.0
857 reference [35] with hisat2 (v2.2.0, default parameters)[73], and alignments were sorted and indexed
858 using samtools (v1.15.1)[48].

859 Peak calling and differential binding (DB) analysis was performed with the Bioconductor csaw
860 package (v1.301)[74]. Peak calling in csaw is based on count-based models, in particular the
861 quasi-likelihood (QL) framework of the edgeR package is used[75]. The counts are modelled
862 using NB distributions that account for overdispersion between biological replicates[76]. Each
863 window can then be tested for significant DB between conditions. We used a window width of
864 10 bp and an estimated fragment length of 213 bp. Prior to counting, repeats were blacklisted
865 from the genome using the SollycSweet-100v2.0 TE annotation[35]. To filter regions and
866 windows, we used the global enrichment approach of csaw. After variance stabilization,
867 windows with significant differential binding with respect to specific factors of interest in our
868 design matrix were identified using the glmQLFTest function. Bins of 10000 bp were used for
869 global background estimation. The median of the average abundances across all 10000 bp bins
870 was used as the global background coverage estimate. We only retained windows with at least
871 a 4-fold change from the global background coverage. We counted the reads into large bins
872 and normalized with the wrapper function normFactors, which uses trimmed mean of M-values
873 (TMM) method. The false discovery rate (FDR) control was obtained by applying the
874 Benjamini-Hochberg (BH) method. In csaw, the FDR control was applied to all detected DB
875 regions or peaks. Regions were obtained by grouping windows into regions using the csaw
876 mergeWindows function. The minimum distance at which two binding events were treated as
877 separate sites was set to 1000 bp. Significant regions were identified with the csaw
878 makeContrasts function (FDR ≤0.01). Gene-based annotation of differentially-bounds regions
879 was performed using the detailRanges function of csaw (3 Kbp upstream and 2 Kbp
880 downstream of TSS) and annotation file SollycSweet-100_genes_v2.1.1.gff3[35]. BED files with
881 significant regions and BigWig files with normalized read coverage were exported via the
882 *export* function of the rtracklayer package[77] in R. *De-novo* motif discovery was performed with
883 the 1000 most significant peaks (by FDR) for each sample by analysing genomic sequences
884 from position -100 to +100 relative to the peak center using MEME (v 5.3.3; parameters -dna
885 -mod zoops -nmotifs 3 -minw 6 -maxw 15 -maxsites 1000 -objfun classic -revcomp -
886 markov_order 0)[78].

887 Genome-wide distribution of peaks was determined using ChIPSeeker (v1.32.0)[79] by
888 annotating regions +/- 5 Kbp around the TSS with the function annotatePeak (parameters
889 tssRegion=c(-5000, 5000)). Peak intensity profiles and peak heatmaps were generated using
890 the computeMatrix, plotHeatmap, and plotProfile functions in deepTools[80]. Most-enriched
891 motifs for SSP, SSP2^{F169}, and SSP2^{S169} were mapped to the SollycSweet-100v2.0 reference[35]
892 with FIMO/MEME Suite[78]. Browser shots of peak coverage, peak regions and binding motifs
893 at putative direct targets were generated in jbrowse2[81].

894

895 Gene Ontology enrichment analyses

896 Gene Ontology (GO) enrichment analyses for biological processes were performed using a
897 functional annotation database for the ITAG4.0 tomato genome from PLAZA[82] and the
898 ClusterProfiler package in R[83] with a pvalueCutoff = 0.05 and pAdjustMethod = BH.

899

900 RNA-seq

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

914 The quality of raw reads was assessed using FastQC (v0.11.9;
915 http://www.bioinformatics.babraham.ac.uk/projects/fastqc/). Raw reads were aligned to the
916 genome reference M82v1.0[35] using STAR[84] (v2.7.6a; parameters --runMode alignReads --
917 outFilterType BySJout --outFilterMultimapNmax 20 --outMultimapperOrder Random --
918 alignSJoverhangMin 8 --alignSJDBoverhangMin 1 --alignIntronMin 20 --alignIntronMax

919 1000000 --alignMatesGapMax 1000000). Alignments were sorted and indexed using
920 samtools[48] and gene expression was quantified as unique read pairs aligned to reference
921 annotated gene features (M82v1.1.1) using HTSeq-count (v0.11.2; parameter --order=pos --
922 stranded=no --type=exon --idattr=Parent)[85].

923 All statistical analyses of gene expression were conducted in R[40]. Differentially expressed
924 genes (DEGs) between the mutants *ssp*, *ssp2*, *ssp ssp2*, and the WT were determined with
925 DESeq2 (v1.34.0)[86]. Raw count data was transformed in DESeqp2 by variant stabilizing
926 transformation (VST). Reproducibility of biological replicates was assessed by hierarchical
927 clustering (method ward.D) and principle component analysis (PCA) using the PCAtools
928 package (v2.6.0) in R[40]. Significantly differentially expressed genes (DEGs) were identified in
929 *ssp* (n=686), *ssp2* (n=180), and *sspssp2* (n=1507) genes with a 1.5-fold change ($\log_2 FC \geq 0.58$,
930 compared to the WT) and adjusted *p*-value ≤ 0.05 cutoff. Gene normalized z-scores were
931 visualized in heatmaps using pheatmap (v1.10.12) and normalized expression of individual
932 transcripts in transcripts per million (TPM) was plotted using ggplot2.

933

934 Microscopy

935 Physalis shoot apical meristems manually dissected with a sharp needle and imaged under a
936 Leica stereomicroscope (M205FCA, camera DFC7000T) using the Leica LAS X software
937 (v3.7.0.20979).

938

939 Statistics

940 Statistical tests and results are reported in Figures and Figure Legends. Significant changes in
941 transactivation assays and in growth phenotypes were determined by one-way ANOVAs
942 followed by post-hoc Tukey's HSD test at 95% confidence level. Significant differentially
943 expressed genes were determined by Wald's tests and corrected using the Benjamini and
944 Hochberg (BH) method in DESeq2. Significant differentially bound DAP-seq regions were
945 determined by the Benjamini-Hochberg (BH) method in csaw. Significant changes in
946 normalized read coverage at private DAP-seq regions were determined by performing two-
947 tailed, two-sample t-test. All statistical tests were performed in R.

948

949

DATA AVAILABILITY

The LA1589 genome assembly is available at the Solanaceae Genomics Network (https://solgenomics.net/ftp/genomes/Solanum_pimpinellifolium/LA1589/2020/) and on NCBI under the GenBank accession JBIENU000000000. Raw Nanopore sequence data is available on NCBI SRA under the BioProjects PRJNA1169794, PRJNA607731, and PRJNA557253. Raw Illumina sequence data is available on NCBI SRA under the BioProject PRJNA1069353. The LA1589 SIFT database, deleterious variant predictions, and LA1589 liftoff annotation have been deposited on Zenodo (https://doi.org//10.5281/zenodo.13787654). Sequences of tomato and Arabidopsis bZIP proteins are available on the Plant Transcription Factor Database (PlantTFDB, v5.0, https://planttfdb.gao-lab.org/)[49]. Seeds are available on request from S. Soyk. Data for Supplementary Information is provided as Supplementary Data. Source data are provided with this paper.

CODE AVAILABILITY

All original code has been deposited on GitHub (https://github.com/soyklab/delvar-2024) and Zenodo (https://doi.org//10.5281/zenodo.13805190)[87] is publicly available as of the date of publication.

METHODS-ONLY REFERENCES

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

41. Wang, X. *et al.* Genome of Solanum pimpinellifolium provides insights into structural variants during tomato breeding. *Nat Commun* **11**, 5817 (2020).

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

43. Alonge, M. *et al.* RaGOO: fast and accurate reference-guided scaffolding of draft genomes. *Genome Biol* **20**, 224 (2019).

44. 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).

979 45. Walker, B. J. *et al.* Pilon: An Integrated Tool for Comprehensive Microbial Variant
980 Detection and Genome Assembly Improvement. *PLoS One* **9**, e112963 (2014).

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

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

985 48. Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**,
986 2078–2079 (2009).

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

989 50. Katoh, K. & Standley, D. M. MAFFT Multiple Sequence Alignment Software Version
990 7: Improvements in Performance and Usability. *Mol Biol Evol* **30**, 772–780 (2013).

991 51. Minh, Q. B. *et al.* IQ-TREE 2: New Models and Efficient Methods for Phylogenetic
992 Inference in the Genomic Era. *Molecular Biologoy and Evolution* **37**, 1530–1534
993 (2020).

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

996 53. Altenhoff, A. M. *et al.* OMA orthology in 2024: improved prokaryote coverage,
997 ancestral and extant GO enrichment, a revamped synteny viewer and more in the OMA
998 Ecosystem. *Nucleic Acids Res* **52**, D513–D521 (2024).

999 54. Huerta-Cepas, J. *et al.* PhylomeDB v3.0: an expanding repository of genome-wide
1000 collections of trees, alignments and phylogeny-based orthology and paralogy
1001 predictions. *Nucleic Acids Res* **39**, D556–D560 (2011).

1002 55. Edgar, R. C. MUSCLE: multiple sequence alignment with high accuracy and high
1003 throughput. *Nucleic Acids Res* **32**, 1792–1797 (2004).

1004 56. Lassmann, T. & Sonnhammer, E. L. L. Kalign – an accurate and fast multiple sequence
1005 alignment algorithm. *BMC Bioinformatics* **6**, 298 (2005).

1006 57. Wallace, I. M. M-Coffee: combining multiple sequence alignment methods with T-
1007 Coffee. *Nucleic Acids Res* **34**, 1692–1699 (2006).

1008 58. Minh, B. Q. *et al.* IQ-TREE 2: New Models and Efficient Methods for Phylogenetic
1009 Inference in the Genomic Era. *Mol Biol Evol* **37**, 1530–1534 (2020).

1010 59. Kalyaanamoorthy, S., Minh, B. Q., Wong, T. K. F., von Haeseler, A. & Jermiin, L. S.
1011 ModelFinder: fast model selection for accurate phylogenetic estimates. *Nat Methods* **14**,
1012 587–589 (2017).

1013 60. Huerta-Cepas, J., Serra, F. & Bork, P. ETE 3: Reconstruction, Analysis, and
1014 Visualization of Phylogenomic Data. *Mol Biol Evol* **33**, 1635–1638 (2016).

1015 61. Huerta-Cepas, J., Dopazo, H., Dopazo, J. & Gabaldón, T. The human phylome. *Genome*
1016 *Biol* **8**, R109 (2007).

1017 62. Gabler, F. *et al.* Protein Sequence Analysis Using the MPI Bioinformatics Toolkit. *Curr*
1018 *Protoc Bioinformatics* **72**, e108 (2020).

1019 63. Shen, M. & Sali, A. Statistical potential for assessment and prediction of protein
1020 structures. *Protein Science* **15**, 2507–2524 (2006).

1021 64. Pettersen, E. F. *et al.* UCSF Chimera—A visualization system for exploratory research
1022 and analysis. *J Comput Chem* **25**, 1605–1612 (2004).

1023 65. Engler, C. *et al.* A Golden Gate Modular Cloning Toolbox for Plants. *ACS Synth Biol* **3**,
1024 839–843 (2014).

1025 66. Decaestecker, W. *et al.* CRISPR-TSKO: A Technique for Efficient Mutagenesis in
1026 Specific Cell Types, Tissues, or Organs in Arabidopsis. *Plant Cell* **31**, 2868–2887
1027 (2019).

1028 67. Swartwood, K. & Van Eck, J. Development of plant regeneration and Agrobacterium
1029 tumefaciens-mediated transformation methodology for Physalis pruinosa. *Plant Cell,*
1030 *Tissue and Organ Culture (PCTOC)* **137**, 465–472 (2019).

1031 68. Brooks, C., Nekrasov, V., Lippman, Z. B. & Van Eck, J. Efficient Gene Editing in
1032 Tomato in the First Generation Using the Clustered Regularly Interspaced Short
1033 Palindromic Repeats/CRISPR-Associated9 System. *Plant Physiol* **166**, 1292–1297
1034 (2014).

1035 69. Kluesner, M. G. *et al.* EditR: A Method to Quantify Base Editing from Sanger
1036 Sequencing. *CRISPR J* **1**, 239–250 (2018).

1037 70. Galvão, V. C. *et al.* PIF transcription factors link a neighbor threat cue to accelerated
1038 reproduction in Arabidopsis. *Nat Commun* **10**, 4005 (2019).

1039 71. Hutin, S. *et al.* Identification of Plant Transcription Factor DNA-Binding Sites Using
1040 seq-DAP-seq. *Methods Mol Biol* **2698**, 119–145 (2023).

1041 72. Gaspar, J. M. NGmerge: merging paired-end reads via novel empirically-derived models
1042 of sequencing errors. *BMC Bioinformatics* **19**, 536 (2018).

1043 73. Kim, D., Paggi, J. M., Park, C., Bennett, C. & Salzberg, S. L. Graph-based genome
1044 alignment and genotyping with HISAT2 and HISAT-genotype. *Nat Biotechnol* **37**, 907–
1045 915 (2019).

1046 74. Lun, A. T. L. & Smyth, G. K. csaw: a Bioconductor package for differential binding
1047 analysis of ChIP-seq data using sliding windows. *Nucleic Acids Res* **44**, e45 (2016).

1048 75. Lund, S. P., Nettleton, D., McCarthy, D. J. & Smyth, G. K. Detecting differential
1049 expression in RNA-sequence data using quasi-likelihood with shrunken dispersion
1050 estimates. *Stat Appl Genet Mol Biol* **11**, e45–e45 (2012).

1051 76. Robinson, M. D. & Smyth, G. K. Small-sample estimation of negative binomial
1052 dispersion, with applications to SAGE data. *Biostatistics* **9**, 321–332 (2007).

1053 77. Lawrence, M., Gentleman, R. & Carey, V. rtracklayer: an R package for interfacing with
1054 genome browsers. *Bioinformatics* **25**, 1841–1842 (2009).

1055 78. Bailey, T. L., Johnson, J., Grant, C. E. & Noble, W. S. The MEME Suite. *Nucleic Acids*
1056 *Res* **43**, W39–W49 (2015).

1057 79. Yu, G., Wang, L.-G. & He, Q.-Y. ChIPseeker: an R/Bioconductor package for ChIP
1058 peak annotation, comparison and visualization. *Bioinformatics* **31**, 2382–2383 (2015).

1059 80. Ramírez, F., Dündar, F., Diehl, S., Grüning, B. A. & Manke, T. deepTools: a flexible
1060 platform for exploring deep-sequencing data. *Nucleic Acids Res* **42**, W187–W191
1061 (2014).

1062 81. Diesh, C. *et al.* JBrowse 2: a modular genome browser with views of synteny and
1063 structural variation. *Genome Biol* **24**, 74 (2023).

1064 82. Van Bel, M. *et al.* PLAZA 5.0: extending the scope and power of comparative and
1065 functional genomics in plants. *Nucleic Acids Res* **50**, D1468–D1474 (2022).

1066 83. Yu, G., Wang, L.-G., Han, Y. & He, Q.-Y. clusterProfiler: an R Package for Comparing
1067 Biological Themes Among Gene Clusters. *OMICS* **16**, 284–287 (2012).

1068 84. Dobin, A. *et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21
1069 (2013).

1070 85. Anders, S., Pyl, P. T. & Huber, W. HTSeq—a Python framework to work with high-
1071 throughput sequencing data. *Bioinformatics* **31**, 166–169 (2015).

1072 86. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion
1073 for RNA-seq data with DESeq2. *Genome Biol* **15**, 550 (2014).

1074 87. Soyk, S. soyklab/delvar-2024-v1.0.0 (v1.0.0), Zenodo
1075 https://doi.org//10.5281/zenodo.13805190

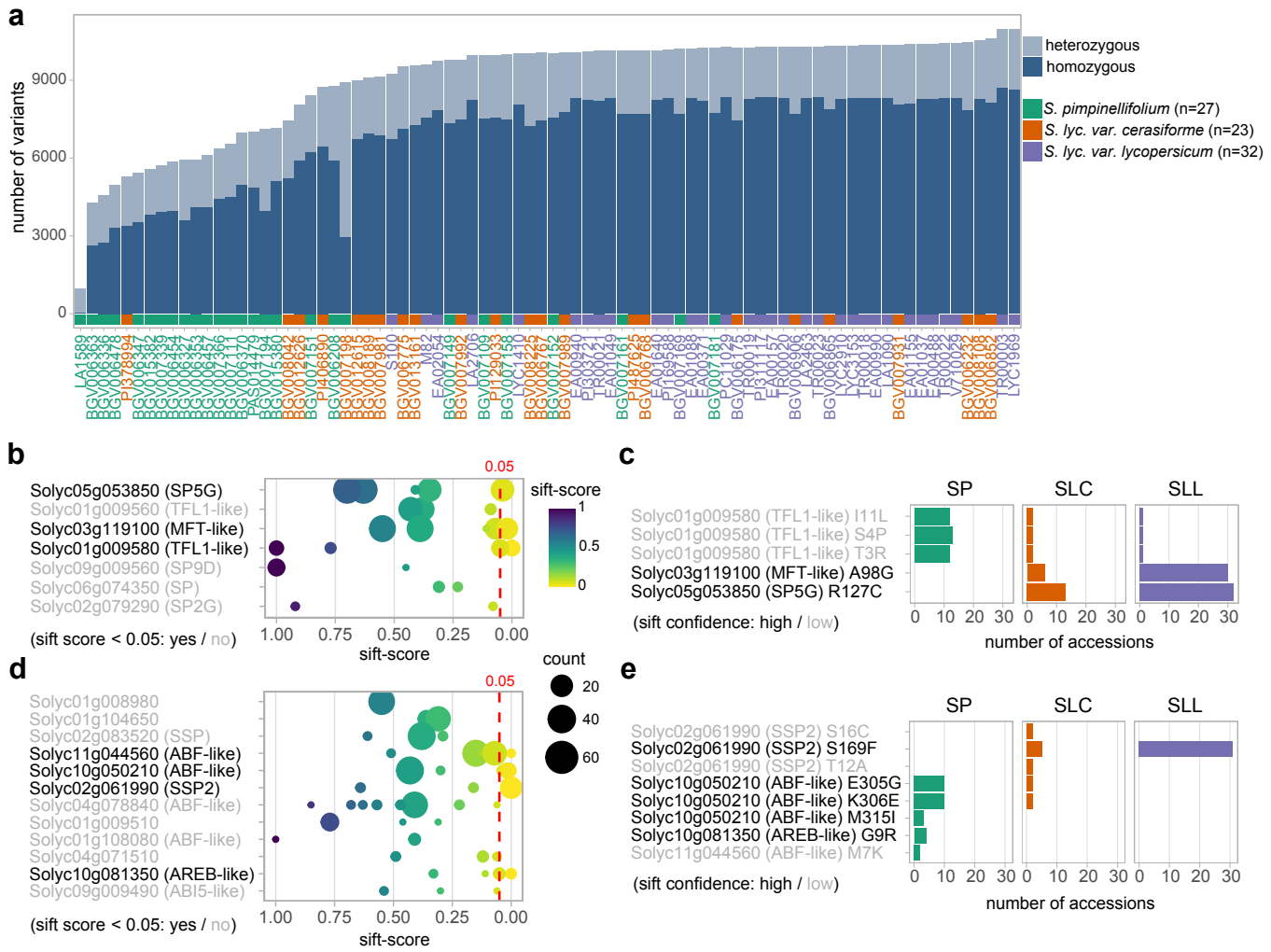

Figure 1: Predicting the load of deleterious variants along the domestication history of tomato. a, Number of predicted deleterious mutations in a panel of 82 tomato genomes, including wild species (*S. pimpinellifolium*, green), landraces (*S. lycopersicum var. cerasiforme*, orange), and cultivars (*S. lycopersicum*, purple). **b**, Prediction of deleterious variants across all annotated *CETS* 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. Black and grey font indicates genes with predicted deleterious and tolerated mutations, respectively. **c**, Distribution of predicted deleterious *CETS* gene variants across wild (*S. pimpinellifolium*; SP), landrace (*S. lyc var. cerasiforme*; SLC) and domesticated (*S. lycopersicum*, SLL) genomes. Black and grey font indicate high and low prediction confidence, respectively. **d,e**, Prediction of deleterious variants in all annotated Group-A bZIP genes (d) and their distribution across different genome categories (e) as in (b, c).

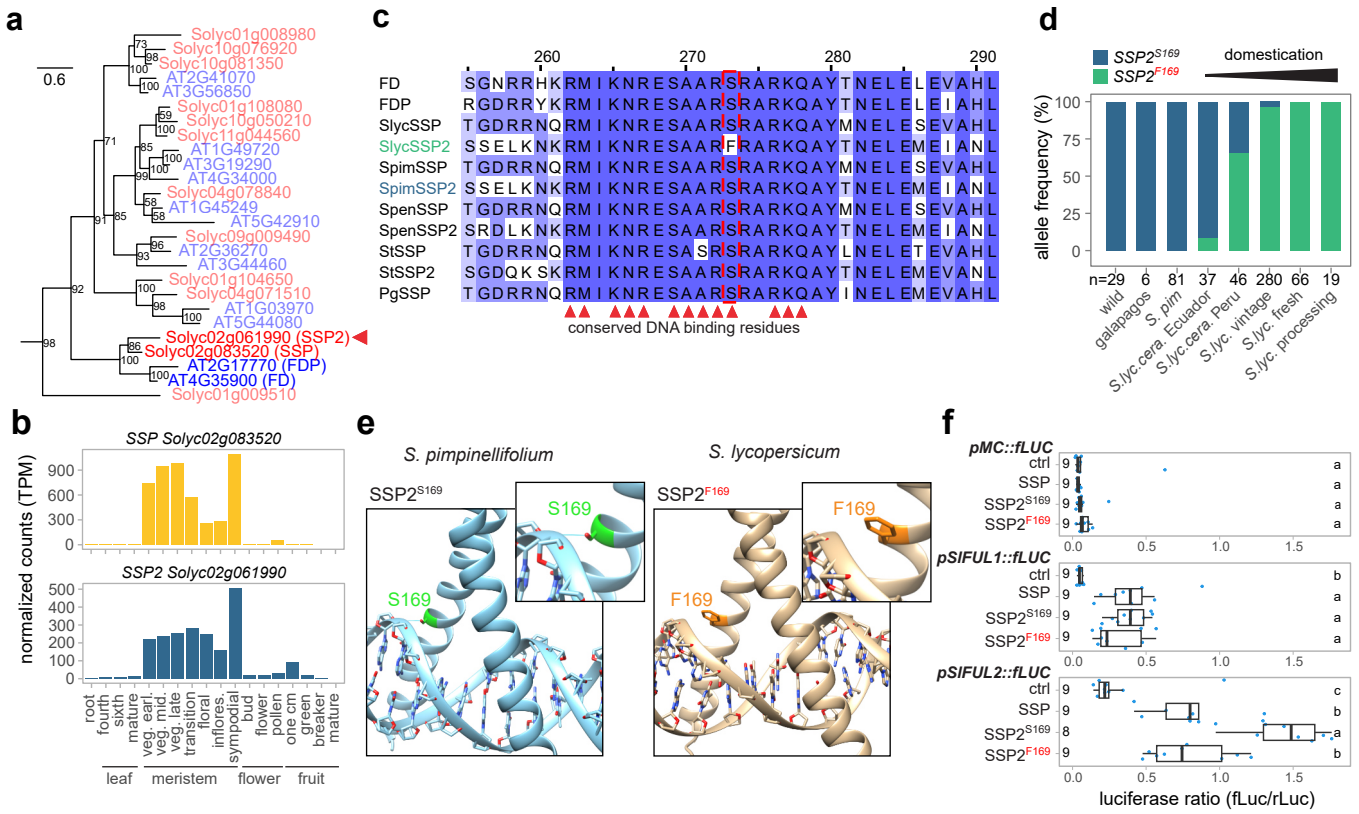

Figure 2: A deleterious mutation in *SSP2* reduces its transcription factor activity. a, Maximum-likelihood tree of A-group bZIP proteins in tomato (red font) and Arabidopsis (blue font). Red arrowhead marks SSP2. Numbers represent bootstrap values from 1,000 replicates and scale bar indicates the average 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 biological replicates (transformed leaves). Ctrl indicates the no effector control (35S::YFP). Letters represent the results from pairwise comparisons of means using one-way ANOVA and post-hoc Tukey's HSD test with 95% confidence level. For box plots in (f), the bottom and top of boxes represent the first and third quartile, respectively, the middle line the median and the whiskers the maximum and minimum values.

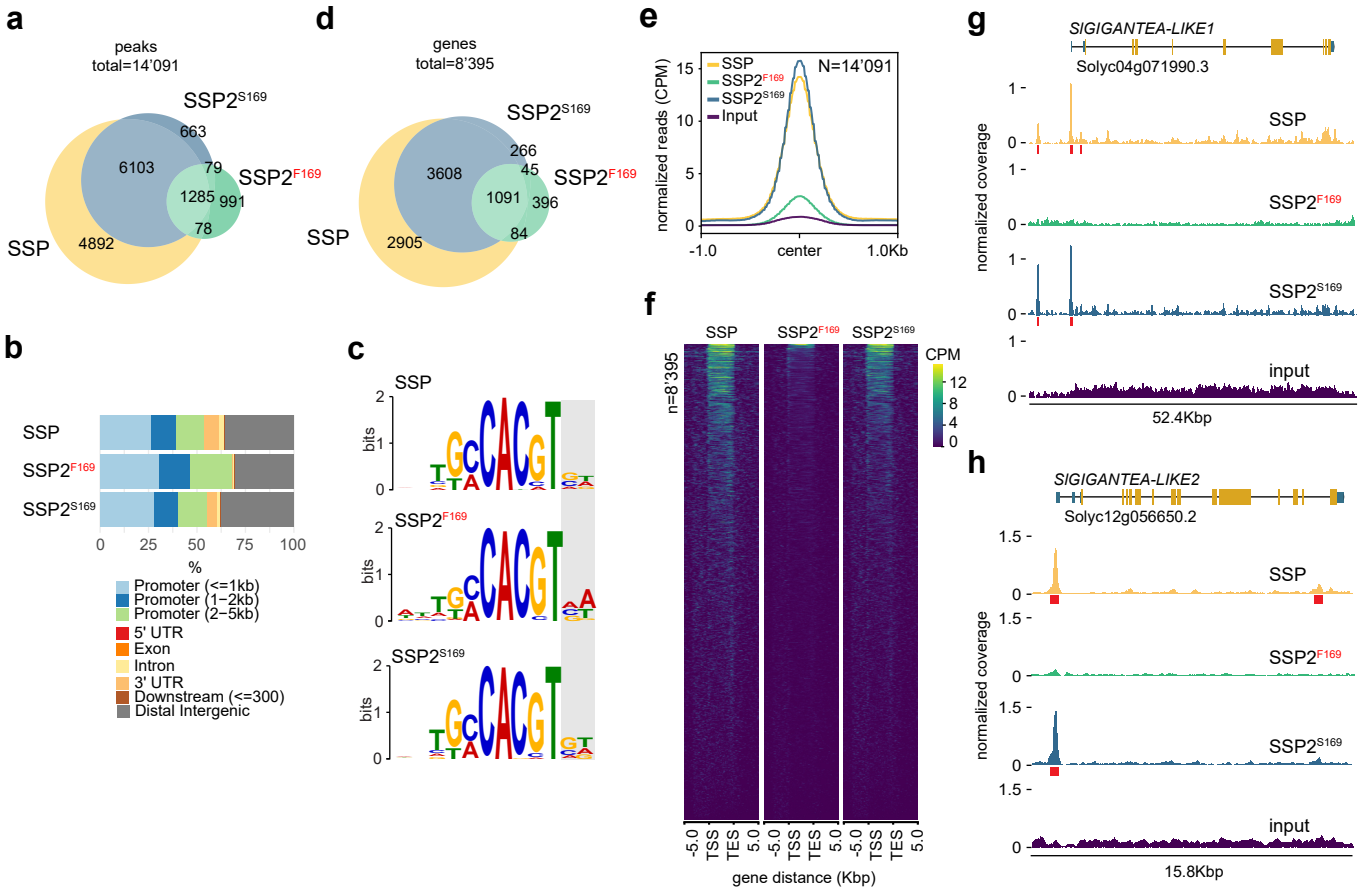

Figure 3: Domesticated SSP2^{F169} shows reduced binding at genome-wide target loci. a, Overlap of significant ($\log_2$FC $\geq$ 3, FDR $\leq$ 0.01) SSP, SSP2^{F169}, and SSP2^{S169} DAP-seq peaks (n=14'091). **b**, Distribution of significant SSP, SSP2^{F169}, and SSP2^{S169} DAP-seq peaks across gene 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 $\leq$ 3 Kbp upstream and $\leq$ 2 Kbp downstream of the transcriptional start site (n=8'395). **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=8'395). **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.

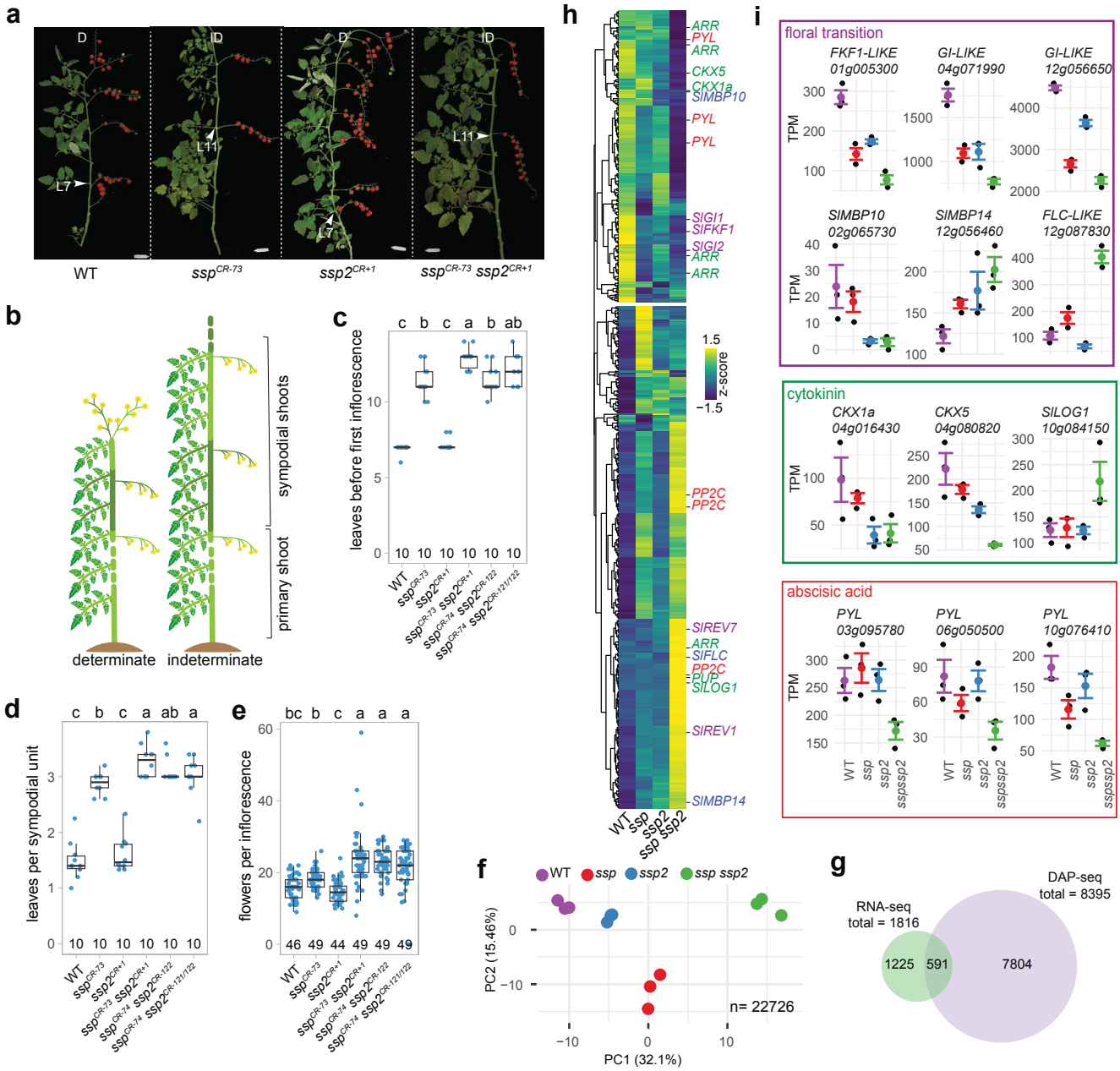

Figure 4: *SSP* and *SSP2* act partially redundant to regulate the transition to flowering. a, Representative images of wild-type S100, *ssp*^{CR} and *ssp2*^{CR} single mutants, and *ssp ssp2*^{CR} double mutants. Digits in superscript indicate the number of deleted (CR-) or inserted (CR+) nucleotides. All allele sequences are found in **Supplementary Figure 6**. 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 the results from pairwise comparisons of means using one-way ANOVA and post-hoc Tukey's HSD test 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 591 putative SSP/SSP2 target genes. **i**, Normalized expression (TPM, transcript per million) for selected putative direct targets in 3 independent biological replicates (black dots) and mean values (colored dots). Error bars indicate standard error. Genes are color coded based on the biological pathway. For box plots in (c-e), the bottom and top of the boxes represent the first and third quartile, respectively, the middle line the median and the whiskers the maximum and minimum values.

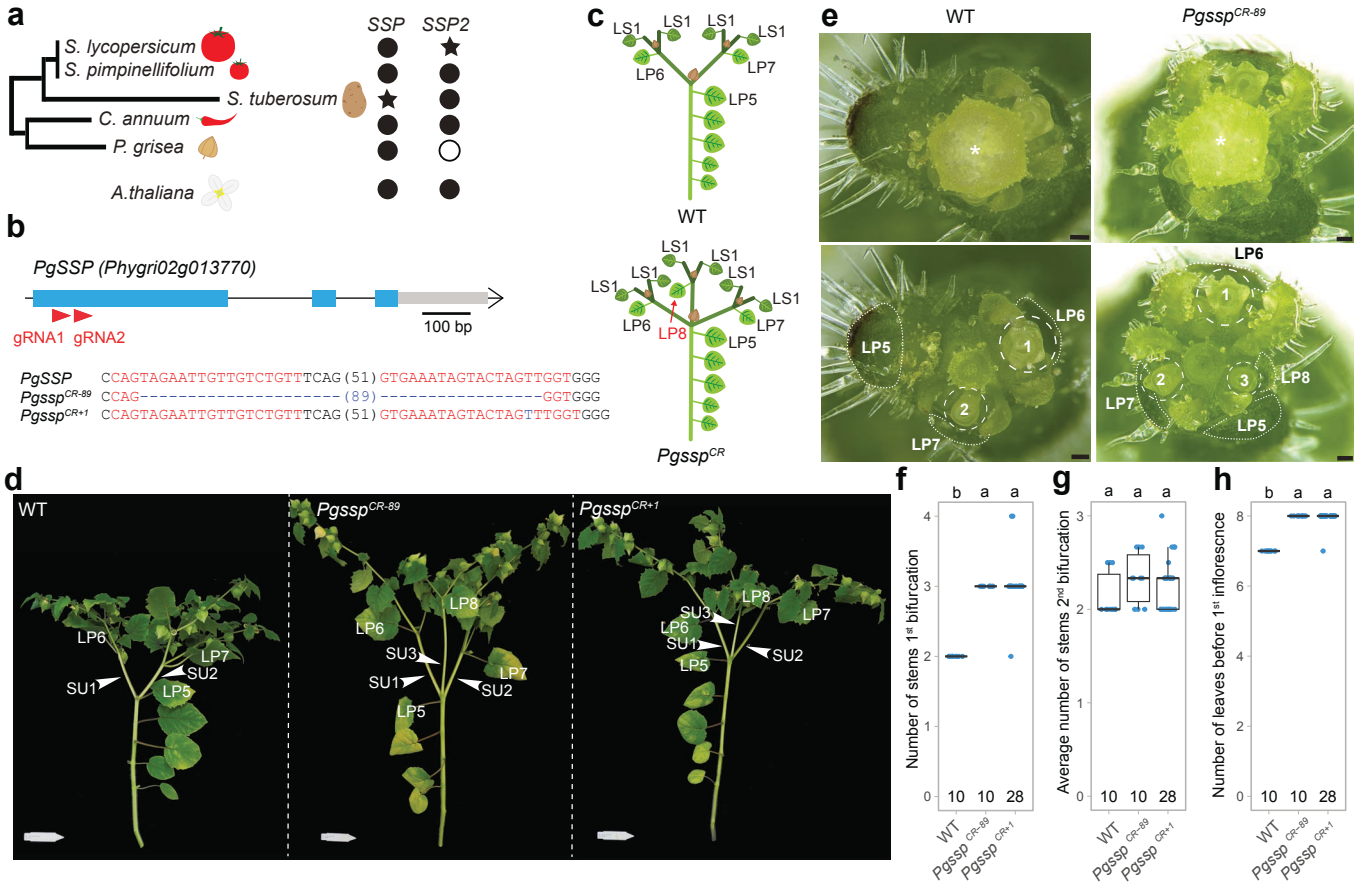

Figure 5: The genome of *Physalis grisea* encodes a single direct *SSP* ortholog that regulates meristem transitions. a, Schematic phylogenetic tree of tomato and closely related *Solanaceae* species. Filled circles, empty circles or stars show presence, absence, or missense mutation, respectively, of *SSP/ SSP2* or *FD/ FDP* in these species. Full tree is displayed in **Supplementary Fig. 7**. **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 sympodial 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 on the primary (LP) and sympodial (LS) shoots corresponds with the shoot of origin **d**, Representative pictures illustrating differences in number of sympodial shoots in WT and *Pgssp* mutants. Last leaf before shoot bifurcation (LP5) and displaced leaves (LP6-8) are indicated (L5). White arrowheads indicate individual sympodial units (SU). Scale bar represents 7.5 cm. **e**, Representative stereoscope images of the shoot apex of WT and *Pgssp* mutants. Upper images show the apex with a terminal flower (*). Lower images show the same view with the flower removed. The sympodial meristems (SYMs) and position of removed primary shoot leaves (LP) are delimited by dashed and dotted lines, respectively and numbered in developmental order. Scale bar represents 100 µm. **f-h**, Quantification of the number of sympodial shoots at the first (f) and second (g) bifurcation, and flowering time (number of leaves before the first inflorescence) (h). Number of plants is indicated at the bottom of plots. Letters represent results from pairwise comparisons of means using one-way ANOVA and post-hoc Tukey's HSD test with 95% confidence level. For box plots in (f-h), the bottom and top of boxes represent the first and third quartile, respectively, the middle line the median and the whiskers the maximum and minimum values.

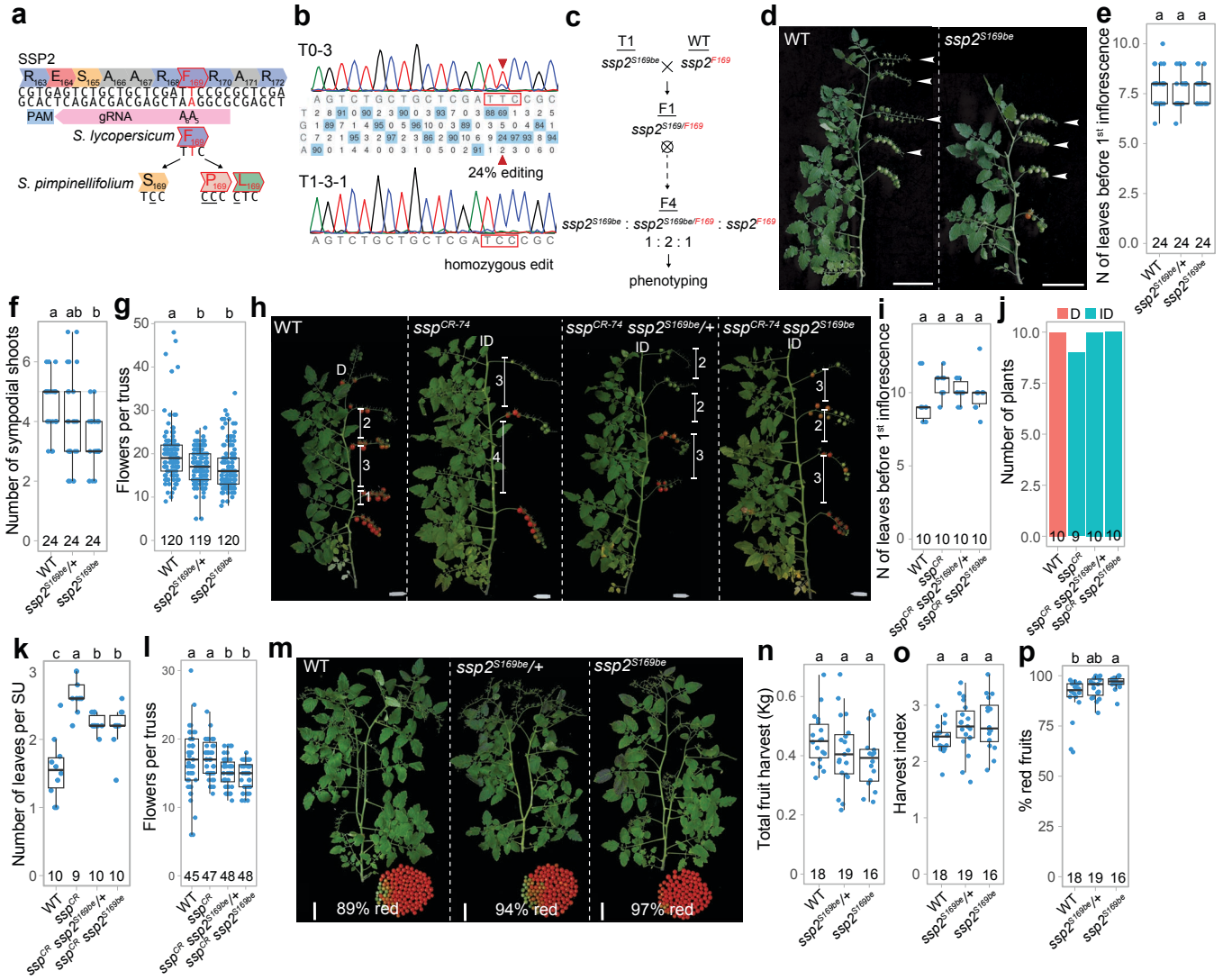

Figure 6: Repairing the deleterious *SSP2* mutation in domesticated tomato by base-editing leads to compact growth and earliness for yield. a, Base-editing strategy to correct the *SSP2* mutation using an adenosine base editor (ABE) and a PAM-less Cas9 variant. The target adenine in *SSP2* (A5) is at 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 the deaminated adenine. **b**, Validation of editing in a chimeric first-generation (T0) transgenic and corresponding T1 progeny by Sanger sequencing. The target nucleotide is marked by a red arrowhead. **c**, Crossing scheme to generate the segregating $ssp2^{S169be}$ F4 population. **d**, Representative pictures showing the number of sympodial units on WT and $ssp2^{S169be}$ plants. Terminal inflorescences of sympodial units are indicated by white arrowheads. **e-g**, Quantification of the number of leaves before the first inflorescence (flowering time) (e), sympodial shoots (f), and flowers per truss (g) 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 (i) (as in (e)), number of determinate plants (j), leaves per sympodial unit (SU) (k), and flowers per truss (l) 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 complete 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 pairwise comparisons of means using one-way ANOVA and post-hoc Tukey's HSD test with 95% confidence level**.** Scale bars represent 10 cm (d) and 7.5 cm in (h,m). For box plots in (e-g, i, k, l, n-p), the bottom and top of boxes represent the first and third quartile, respectively, the middle line the median and the whiskers the maximum and minimum values.