nature plants

LETTERS

https://doi.org/10.1038/s41477-019-0422-z

Duplication of a domestication locus neutralized a cryptic variant that caused a breeding barrier in tomato

Sebastian Soyk[1*], Zachary H. Lemmon[1], Fritz J. Sedlazeck[2], José M. Jiménez-Gómez[3], Michael Alonge[4], Samuel F. Hutton[5], Joyce Van Eck[6,7], Michael C. Schatz[4,8] and Zachary B. Lippman[1,9*]

Genome editing technologies are being widely adopted in plant breeding[1]. However, a looming challenge of engineering desirable genetic variation in diverse genotypes is poor predictability of phenotypic outcomes due to unforeseen interactions with pre-existing cryptic mutations[2–4]. In tomato, breeding with a classical MADS-box gene mutation that improves harvesting by eliminating fruit stem abscission frequently results in excessive inflorescence branching, flowering and reduced fertility due to interaction with a cryptic variant that causes partial mis-splicing in a homologous gene[5–8]. Here, we show that a recently evolved tandem duplication carrying the second-site variant achieves a threshold of functional transcripts to suppress branching, enabling breeders to neutralize negative epistasis on yield. By dissecting the dosage mechanisms by which this structural variant restored normal flowering and fertility, we devised strategies that use CRISPR–Cas9 genome editing to predictably improve harvesting. Our findings highlight the under-appreciated impact of epistasis in targeted trait breeding and underscore the need for a deeper characterization of cryptic variation to enable the full potential of genome editing in agriculture.

Cryptic variation consists of naturally occurring mutations that have little or no phenotypic consequences unless exposed to additional genetic or environmental interactions[2,3]. Genome sequencing projects have revealed widespread allelic variation between even closely related genotypes[9–12]. The proportion of genetic variation that is cryptic is potentially vast, providing mutations that could contribute to evolutionary adaption[13,14]. More important over shorter time scales is the impact on plant and animal breeding, where intense selection pressure combines diverse alleles to enhance productivity[1]. However, cryptic variation can cause unpredictable and sometimes detrimental outcomes due to epistatic interactions with beneficial genetic variation, which then must be overcome to achieve desirable phenotypes[15–18]. How such neutralization of cryptic variation is achieved has not been investigated.

In tomato, natural mutations in the MADS-box transcription factor gene *JOINTLESS2* (*J2*) have the potential to improve harvestability through a modification of flower development that eliminates the abscission zone on the stems of fruits[6,7,19]. However, the *j2* breeding mutation, caused by a transposon insertion ($j2^{TE}$), results in undesirable branching of flower-bearing shoots (inflorescences) in genetic backgrounds that also carry a cryptic variant for the close homologue *ENHANCER OF J2* ($ej2^{W}$), which was selected during domestication[5,8]. This combination of loss-of-function alleles ($j2^{TE}$ $ej2^{W}$) results in excessive flower production and low fertility due to poor fruit set, which has prevented widespread use of the jointless trait[20]. We discovered a small group of elite large-fruited varieties that carry both mutations but produce normal, unbranched inflorescences with high fertility and regular fruit set (Fig. 1a–c). This discrepancy suggested breeders selected additional cryptic variants that suppressed negative epistasis on yield.

To find the genetic basis for this suppression, we crossed branched $j2^{TE}$ $ej2^{W}$ double mutants with an unbranched $j2^{TE}$ $ej2^{W}$ breeding line (Fla.8924) and found F_1 hybrids produced mostly unbranched inflorescences, indicating suppression is partially dominant[8,21] (Fig. 1a–c and Supplementary Fig. 1). A large F_2 population of 1,536 plants produced a range of inflorescence complexities, suggesting that multiple quantitative trait loci (QTLs) underlie suppression. To dissect the genetic architecture of suppression, we selected 70 F_2 plants that captured the phenotypic extremes of inflorescence branching (Fig. 1d). We subcategorized branched and suppressed groups into 'A' and 'B' classes on the basis of expressivity and sequenced pools of DNA from each group. Comparing SNP-ratios between the branched and suppressed phenotypic classes showed two major QTLs on chromosomes 1 and 3, which we designated *suppressor of branching1* (*sb1*) and *sb3*, respectively (Fig. 1d,e). Whereas both QTLs were found when comparing A-classes, only *sb3* appeared in the B-class comparison, suggesting a stronger contribution from *sb3* (Fig. 1f).

To resolve the individual and combined effects from *sb1* and *sb3* we analysed inflorescence complexity from F_3 families segregating for both QTLs. Similar to the branched $j2^{TE}$ $ej2^{W}$ double-mutant parent, F_3 plants lacking both Fla.8924 suppressor QTLs ($sb1^{-/-}sb3^{-/-}$) developed strongly branched inflorescences. Notably, *sb1* partially suppressed branching in a dose-dependent additive manner, with heterozygotes ($sb1^{+/-}sb3^{-/-}$) and homozygotes ($sb1^{+/+}sb3^{-/-}$) showing

[1]Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, USA. [2]Human Genome Sequencing Center, Baylor College of Medicine, Houston, TX, USA. [3]Institut Jean-Pierre Bourgin, INRA, AgroParisTech, CNRS, Université Paris-Saclay, Versailles, France. [4]Department of Computer Science, Johns Hopkins University, Baltimore, MD, USA. [5]Horticultural Sciences Department, University of Florida, Wimauma, FL, USA. [6]The Boyce Thompson Institute, Ithaca, NY, USA. [7]Plant Breeding and Genetics Section, School of Integrative Plant Science, Cornell University, Ithaca, NY, USA. [8]Department of Oncologye, Johns Hopkins Medicine, Baltimore, MD, USA. [9]Howard Hughes Medical Institute, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, USA. *e-mail: ssoyk@cshl.edu; lippman@cshl.edu

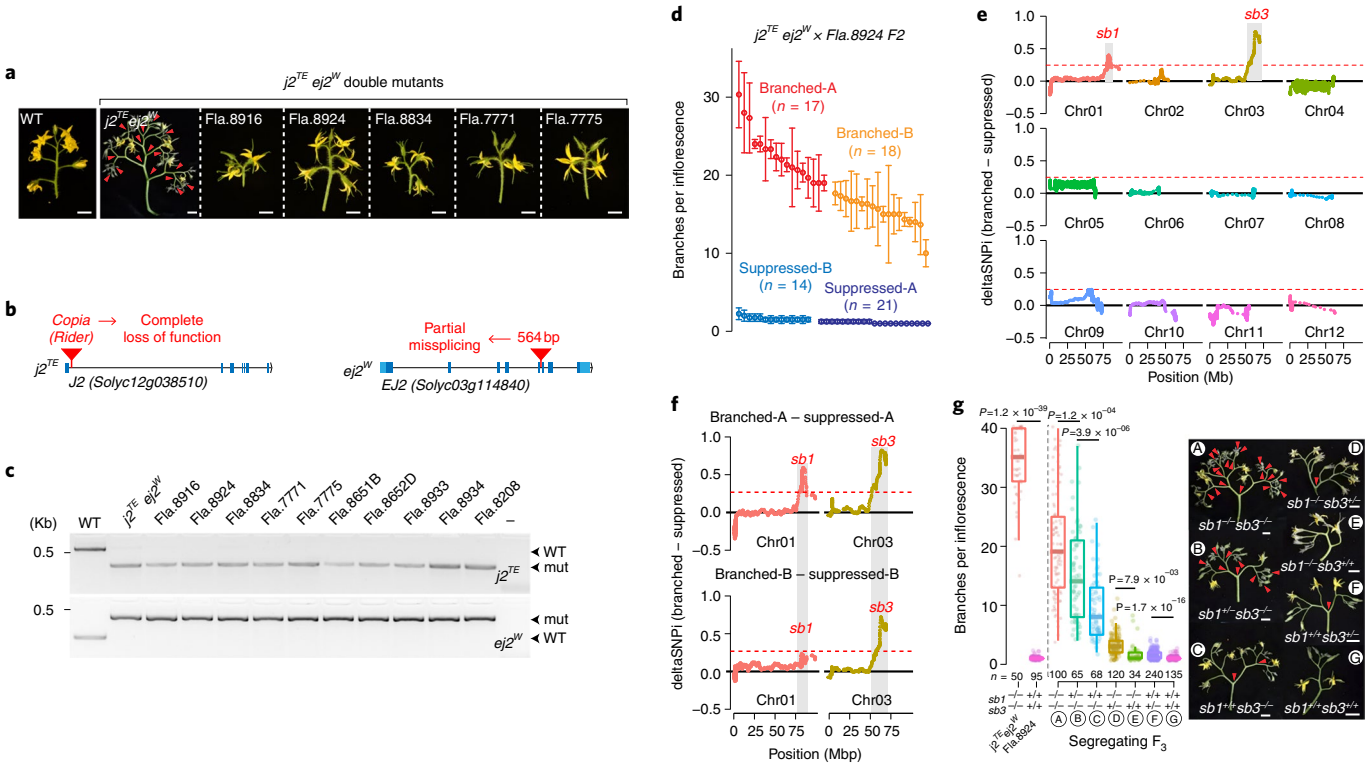

Fig. 1 | Two QTLs suppress undesirable inflorescence branching in tomato jointless breeding lines. a, Natural mutations in the related *SEPALLATA4* MADS-box genes *J2* and *EJ2* ($j2^{TE}$ $ej2^{W}$) lead to excessively branched inflorescences in the $j2^{TE}$ $ej2^{W}$ double mutant in the reference cultivar M82 but not in selected elite breeding lines. Representative images from ten independently repeated experiments with similar results are shown. **b**, *J2* (*Solyc12g038510*) and *EJ2* (*Solyc03g114840*) gene structures showing the natural intronic transposon insertion that results in complete loss of gene function in the $j2^{TE}$ allele and the naturally occurring intron insertion that causes partial mis-splicing in the $ej2^{W}$ allele, both of which are responsible for negative epistasis when combined. **c**, PCR genotyping for $j2^{TE}$ and $ej2^{W}$ in WT, the branched $j2^{TE}$ $ej2^{W}$ double mutant and ten unbranched $j2^{TE}$ $ej2^{W}$ breeding lines. Representative gel from three independently repeated experiments with similar results is shown. **d**, Quantification of inflorescence branches in an F_2 population from a cross between $j2^{TE}$ $ej2^{W}$ and the unbranched inbred Fla.8924. Plants were grouped into 'A' and 'B' categories, depending on phenotypic strength; *n* = number of plants, means ± s.d. from 3–4 inflorescences per plant. **e**,**f**, QTL-seq using bulked segregants of branched and unbranched plants shown in **e** showed two *suppressors of branching* (*sb*) loci from Fla.8924 on chromosomes 1 and 3. Only *sb3* is detected when comparing the 'B' categories, suggesting a major effect QTL (**f**). Differences in SNP index (deltaSNPi) between branched and suppressed pools is shown. Red dashed horizontal lines indicate genome-wide 95% cut-off in SNP index. **g**, Quantification of inflorescence branching in *j2 ej2* and Fla.8924 and segregating F_3 progeny families. Each data point is a single inflorescence (*n*). Representative inflorescences with different strengths of branching are shown on the right. For box plots in **g**, the bottom and top of boxes represent the first and third quartile, respectively, the middle line is the median and the whiskers represent the maximum and minimum values. *P* values in **g** represent two-tailed, two-sample *t*-tests. Scale bars, 1 cm.

weak and moderate suppression of branching, respectively (Fig. 1g and Supplementary Fig. 2a,b). Along with environmental influence, *sb1* additivity probably explains the range of inflorescence complexity in the F_2 population. In contrast, both *sb3* heterozygosity ($sb1^{-/-}sb3^{-/+}$) and homozygosity ($sb1^{-/-}sb3^{+/+}$) nearly completely suppressed branching, supporting that *sb3* is partially dominant. Interestingly, *sb3* homozygosity suppressed branching comparable to Fla.8924 and F_3 families homozygous for both suppressors ($sb1^{+/+}sb3^{+/+}$), indicating that *sb3* drives suppression. However, variance in branching was higher in $sb1^{-/-}sb3^{+/+}$ plants compared to $sb1^{+/+}sb3^{+/+}$, suggesting that *sb1* stabilizes unbranched inflorescence architecture in Fla.8924 (Supplementary Fig. 2c).

We focused on *sb3* for further dissection and inspected the 14.6-megabase pair (Mb) mapping interval on chromosome 3 for candidate genes, which included *EJ2* (Fig. 2a). The *EJ2* variant in both the $j2^{TE}$ $ej2^{W}$ branched and Fla.8924 unbranched genotypes is a weak allele ($ej2^{W}$) due to an insertion in the fifth intron[8]. This insertion causes a partial loss of functional transcripts from mis-splicing, resulting in enlarged leaf-like organs (sepals) on flowers and fruits—a trait that may have been selected during domestication. However, $ej2^{W}$ is a cryptic variant in the context of inflorescence architecture until exposed in *j2* mutant backgrounds. Notably, *j2* mutants that are also heterozygous for $ej2^{W}$ ($j2^{TE-/-}ej2^{W+/-}$) produce weakly branched inflorescences compared to strongly branched inflorescences in $j2^{TE}$ $ej2^{W}$ double mutants, due to having one fully functional copy of *EJ2*. This dosage relationship with *j2*, along with partial mis-splicing from the $ej2^{W}$ variant, led us to propose that the dominant effect of *sb3* could be on the basis of higher *EJ2* expression that then reaches a threshold of correctly spliced transcripts. Consistent with this, we found *EJ2* expression was increased more than about twofold in floral tissues of Fla.8924 compared to WT (Supplementary Fig. 3a).

One explanation for this higher expression could be a mutation in a linked trans-acting factor that represses *EJ2*, but this seemed unlikely. Alternatively, nearby cis-regulatory mutations could be responsible but we found no variants 28 kilobase (kb) upstream or 56 kb downstream of $ej2^{W}$ in our Fla.8924 genome sequencing[21]. A third possibility that would also explain the around twofold increased expression is an additional linked copy of *EJ2* exists in Fla.8924, possibly from a duplication event. Inspecting the mapping

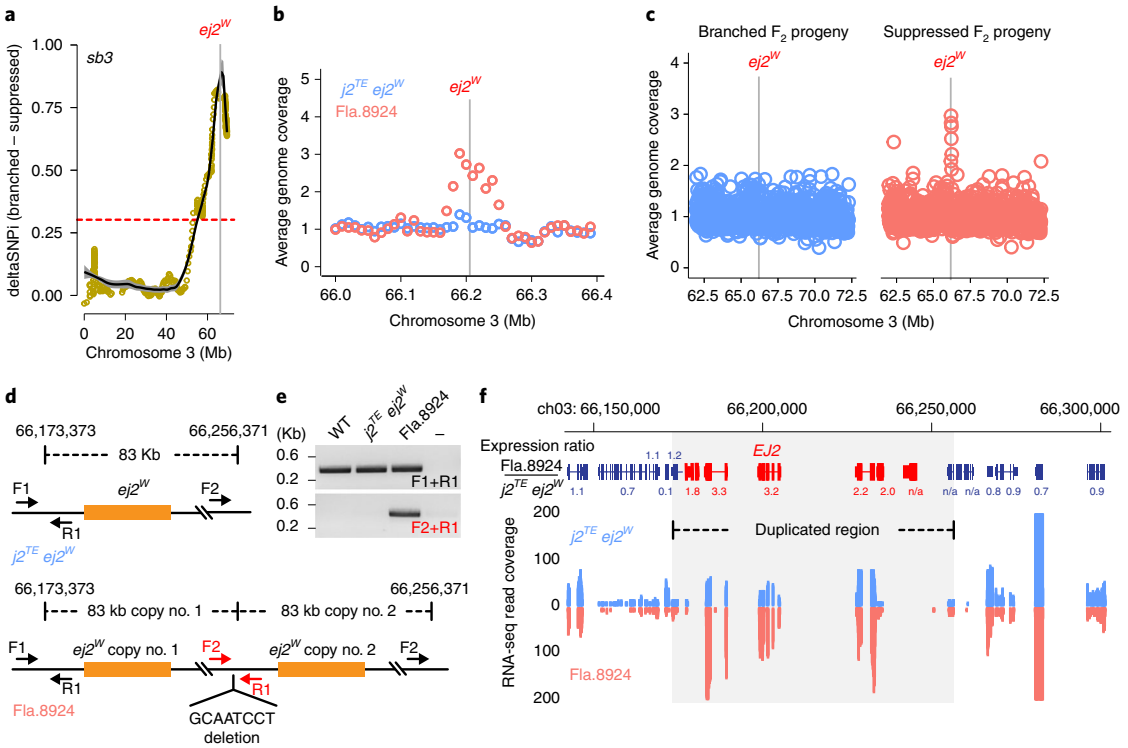

Fig. 2 | A tandem duplication at *sb3* underlies suppression of branching. a, The *sb3* QTL contains the cryptic *ej2*^{W} variant. The delta SNP index (deltaSNPi) between branched and suppressed pools is shown; dashed red line, genome-wide 99% cut-off; black line, locally weighted smoothing regression. **b**, Increased coverage of genomic sequencing reads at *sb3* in Fla.8924 suggests copy number variation for *EJ2* and surrounding genes. Open circles reflect average genome coverage in 10-kb windows across chromosome 3. **c**, Increased coverage of genomic sequencing reads at *EJ2* is detected in the suppressed F_2 progeny from the *j2*^{TE} *ej2*^{W} × Fla.8924 F_2 population. **d**, Model of the *EJ2* locus in *j2*^{TE} *ej2*^{W} and Fla.8924, with the latter having a ~83-kb tandem duplication harbouring *ej2*^{W}. **e**, PCR validation of the tandem duplication. Primers flanking the tandem duplication junction site (F2 + R1) amplifies a product in Fla.8924, but not in WT or the *j2*^{TE} *ej2*^{W} double mutant. Representative gel from three independently repeated experiments with similar results is shown. **f**, RNA-seq from reproductive meristems shows increased expression from genes in the *sb3* duplication compared to no duplication in *j2*^{TE} *ej2*^{W}. Gene position and expression ratios (top) and raw read coverage (bottom) at the floral stage of meristem maturation are shown. Numbers represent normalized expression ratios (TPM).

interval for variation in genome sequencing coverage showed an 83-kb window surrounding *EJ2* with about twofold higher genomic read coverage in Fla.8924 compared to *j2*^{TE} *ej2*^{W}, supporting a duplication (Fig. 2b). We confirmed higher genome coverage at *EJ2* in Fla.8924 by quantitative PCR and defined the borders of the duplication by PCR and Sanger sequencing (Fig. 2d,e and Supplementary Fig. 3b). We also detected the same structural variant by Nanopore long-read sequencing[22], confirming an 83-kb tandem duplication at *sb3* containing *EJ2* and ten additional genes (Supplementary Table 1). Importantly, this duplication co-segregated with suppression of branching in our F_2 mapping population (Fig. 2c).

To determine if increased *EJ2* dosage underlies *sb3* suppression we performed a series of molecular and genetic experiments. We first sequenced RNA from reproductive meristems of WT, *j2*^{TE} *ej2*^{W} double mutants and Fla.8924 and found that *EJ2* and four additional genes on the duplication were expressed significantly higher in Fla.8924 compared to *j2*^{TE} *ej2*^{W} double mutants, including a second MADS-box gene, the fruit ripening regulator *FRUITFULL2* (*FUL2*, ref. [23]; Fig. 2f, Supplementary Fig. 3c). Several MADS-box genes increase in expression during tomato meristem maturation[24]. However, among all genes in the *sb3* duplication only *EJ2* was upregulated from the transition to early floral stages of meristem maturation, which define a critical window during which inflorescence architecture is established[24–27] (Fig. 3a). We found that the levels of correctly spliced *EJ2* transcripts doubled from ~30% in *j2*^{TE} *ej2*^{W} to ~60% in Fla.8924 (*j2*^{TE} *sb3*) relative to WT (Fig. 3b–d), supporting that increased *EJ2* dosage from the *sb3* duplication overcomes insufficient levels of functional transcripts caused by the *ej2*^{W} variant.

To validate genetically that two copies of *ej2*^{W} explain *sb3* suppression and exclude a role for *FUL2* or other duplicated genes, we took advantage of a collection of *ej2* loss-of-function alleles to quantitatively modify *EJ2* dosage in isogenic hybrid plants (Fig. 3b). In a *j2* mutant background we found that heterozygotes of *sb3* with a CRISPR–Cas9-generated *EJ2* null mutation (*ej2*^{CR}) developed highly branched inflorescences similar to those of the *ej2*^{W}/*ej2*^{W} homozygous controls, indicating that two *ej2*^{W} copies in cis (in *sb3*/*ej2*^{CR}) and in trans (in *ej2*^{W}/*ej2*^{W}) had comparable effects on branching from the same dosage of the *ej2*^{W} allele (Fig. 3c). Quantitative differences in branching between these genotypes were probably attributable to additional unknown modifier loci that distinguish Fla.8924 and the donor genotype carrying *ej2*^{W} and the *ej2*^{CR} null allele (M82) (Fig. 3d). Importantly, higher *EJ2* gene dosage in *sb3*/*EJ2* hybrids resulted in a near-complete suppression of branching (Fig. 3c,d). Because *sb3*/*ej2*^{CR} and *sb3*/*EJ2* hybrids differ only in *EJ2* dosage and no other genes in the duplication, these results confirm that the additional copy of *ej2*^{W} is responsible for suppression.

Our findings indicate that the *sb3* duplication facilitated use of *j2* mutations in tomato breeding but only for a narrow germplasm pool, suggesting *sb3* is a rare structural variant. To determine when *sb3* arose and was combined with *j2* during breeding, we used re-sequencing data from 590 diverse tomato genomes to analyse *sb3* allele frequency[11,28,29] and then compared allele distributions for

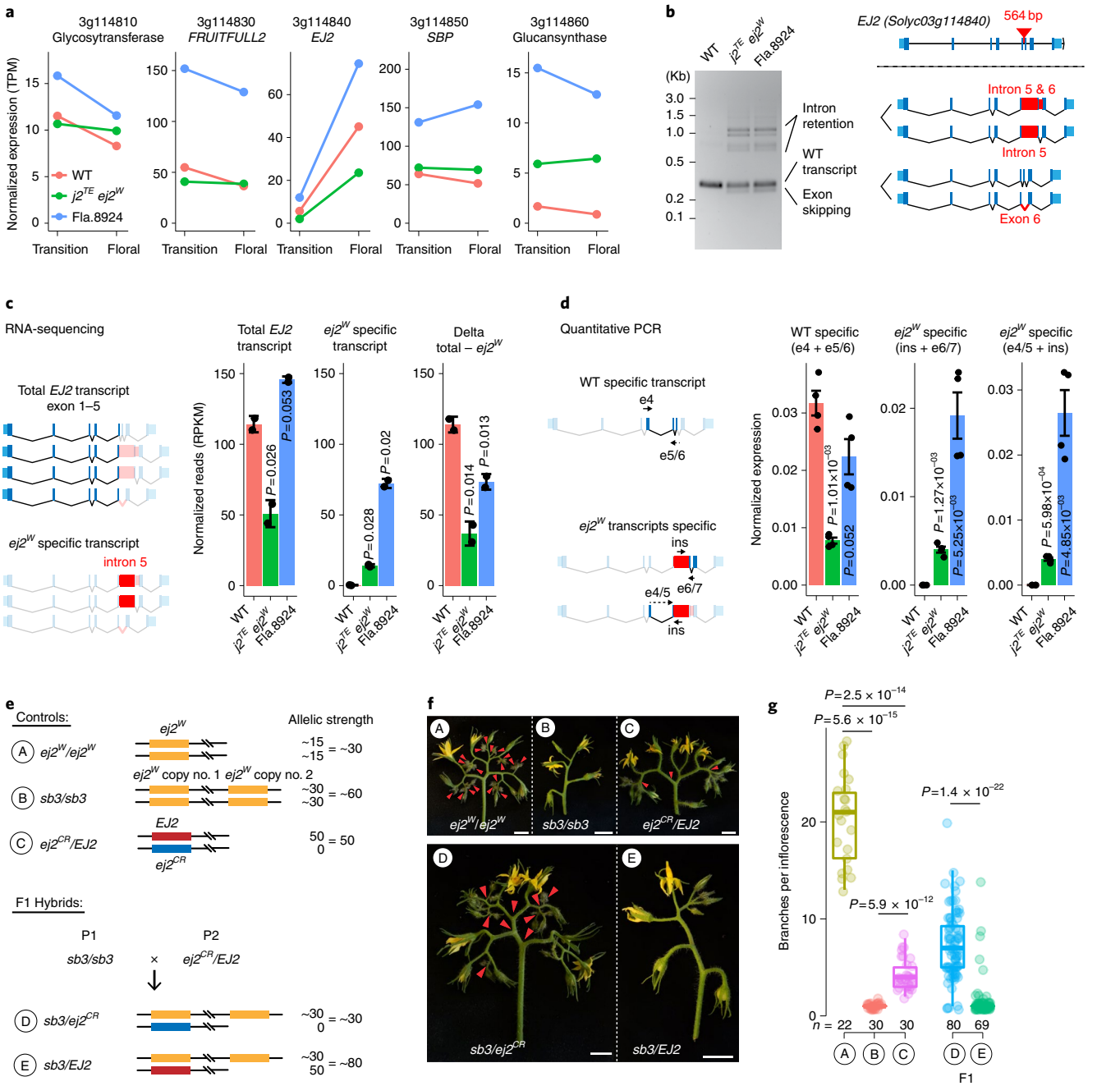

Fig. 3 | Duplication of the weak *ej2*^{W} allele causes a dose-dependent suppression of undesirable inflorescence branching in the *j2* background. **a**, RNA-seq analysis of reproductive meristems at the transition and floral stage from the WT, *j2*^{TE} *ej2*^{W} and Fla.8924, showing dynamic expression of *EJ2* (*3g114840*) but not the neighbouring MADS-box gene *FRUITFULL2* (*3g114830*). Data are shown as means; $n = 2$ biologically independent pooled meristem samples. **b**, PCR and cloning of *EJ2* complementary DNA (cDNA) from WT, *j2*^{TE} *ej2*^{W} and Fla.8924 shows four alternatively spliced transcripts from *ej2*^{W}. Representative gel from two independently repeated experiments with similar results is shown. **c**, Quantification of alternative *EJ2* transcripts in floral meristems by RNA sequencing. *EJ2* exons 1–5 were used as a readout for the total amount of *EJ2* transcript and intron 5 with the *ej2*^{W} insertion for quantifying *ej2*^{W} specific transcripts. Functional *EJ2* transcript (delta total-*ej2*^{W}) is increased in unbranched Fla.8924 compared to branched *j2*^{TE} *ej2*^{W}. Data are shown as means; $n = 2$ biologically independent pooled meristem samples. **d**, Quantification of alternative *EJ2* transcripts in floral meristems by quantitative PCR with reverse transcription (RT-qPCR). *EJ2* exon 5/6 border was used as readout for the WT-specific transcript and intron 5 with the *ej2*^{W} insertion for *ej2*^{W}-specific transcripts. Data is normalized to *UBIQUITIN* and shown as means ± s.d.; $n = 2$ biologically independent pooled meristem samples and two technical replicates. **e**, Genetic manipulation of *EJ2* gene dosage using a collection of homozygous and heterozygous alleles (*ej2*^{W}, weak; *sb3*, weak and duplicated; *ej2*^{CR}, CRISPR-Cas9-generated null; *EJ2*, WT). Schematics of alleles, genetic crosses to create an *ej2*^{w} copy number dosage series and the resulting genotypes analysed in **f** and **g**. Allelic strength is colour-coded (red, WT; orange, weak *ej2*^{w}; blue, null *ej2*) and estimated allelic strength on the basis of functional transcripts is indicated as arbitrary values. **f**, Representative images of inflorescences from isogenic genotypes having different *EJ2* gene dosage (**e**). Red arrowheads indicate inflorescence branch points. Scale bars, 1 cm. **g**, Quantification of inflorescence branching of genotypes in **e**. Each data point represents a single inflorescence (n). For box plots in **g**, the bottom and top of boxes represent the first and third quartile, respectively, the middle line is the median and the whiskers represent the maximum and minimum values. P values in **c**, **d** and **g** represent two-tailed, two-sample t-tests.

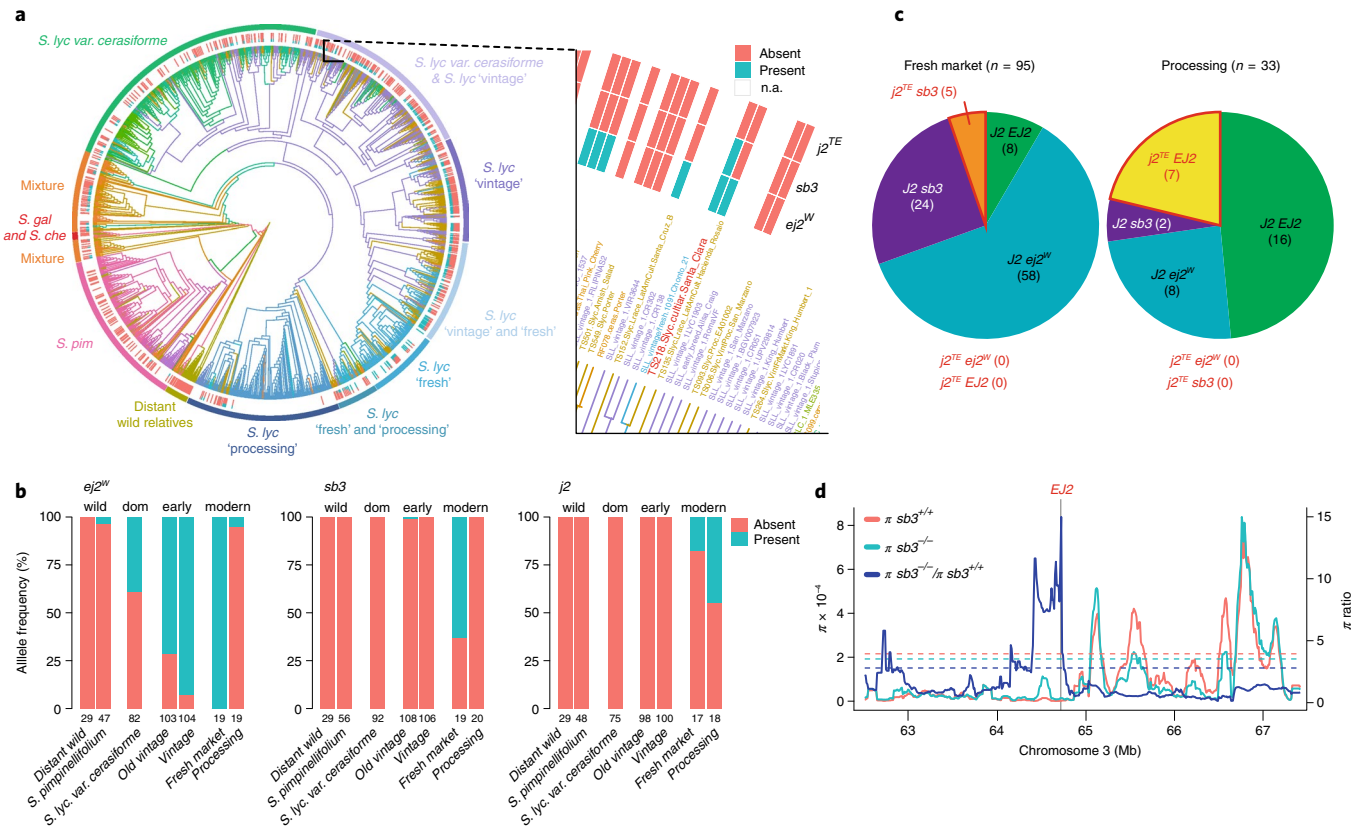

Fig. 4 | The *sb3* duplication is a rare variant that arose recently in a 'vintage' tomato cultivar and was selected to overcome *j2 ej2*^{W} negative epistasis during the development of large fresh-market varieties. a, Phylogeny of 1,606 tomato accessions using 1,812 genome-wide SNPs showing that *sb3* is restricted to vintage and modern cultivars. Coloured lines at the ends of each branch indicate the presence (blue) or absence (red) of the *ej2*^{W}, *sb3* and *j2*^{TE} alleles. Zoom-in of the probable founder genotype of *sb3* ('Santa Clara') is shown at right. **b**, Allele frequencies of *ej2*^{W}, *sb3* and *j2*^{TE} in accessions classified as wild *Solanum* species (distant relatives and *S. pimpinellifolium*, the wild progenitor of domesticated tomato), early domesticates and cultivars (*S. lycopersicum* var. *cerasiforme* and *S. lycopersicum* vintage) and modern cultivars (fresh-market and processing), see also Supplementary Fig. 4b. Number of accessions is indicated below each bar. **c**, Distribution of *J2 EJ2* genotypes in accessions classified as fresh-market ($n = 95$) and processing tomato types ($n = 33$). Jointless (*j2*^{TE}) genotypes are highlighted in red. **d**, Nucleotide diversity (π) plots across a 10-kb chromosomal region containing *EJ2*, calculated using $n = 107$ accessions classified as 'fresh-market' genotypes with ($n = 37$, red line) and without ($n = 73$, green line) the *sb3* locus. The blue line shows the pi-ratio between the accessions with and without the *sb3* locus. Dashed lines indicate the 90% cut-off across the whole chromosome.

j2^{TE}, *ej2*^{W} and *sb3* with a phylogeny of 1,606 wild, early domesticated and cultivated accessions[30] (Fig. 4a, Supplementary Fig. 4a and Supplementary Table 2). We did not detect *sb3* in distantly related wild *Solanum* species, the progenitor species of domesticated tomato (*S. pimpinellifolium*), nor the first domesticated types (*S. lycopersicum* var. *cerasiforme*) (Fig. 4b, Supplementary Fig. 4b and Supplementary Table 3). Instead, we found the *sb3* duplication arose in the 'vintage' accessions, which comprise cultivars developed approximately 75 years ago[30], suggesting *sb3* emerged as a cryptic variant during the earliest stages of modern breeding. In contrast, *ej2*^{W} arose in *S. lycopersicum* var. *cerasiforme* and reached near-fixation in vintage cultivars, indicating that *sb3* evolved in a cultivar that already carried the cryptic *ej2*^{W} variant. Notably, *j2*^{TE} was only detected in modern breeding cultivars, consistent with *j2*^{TE} being introduced later into rare *sb3* genotypes.

Over the last 75 years, breeding efforts took two major directions: 'processing' types with 'square round' fruits that withstand machine harvesting and 'fresh-market' types with large round fruits that are harvested by hand[31]. Interestingly, we found *ej2*^{W} and *sb3* enriched in fresh-market but not in processing cultivars. This suggested that using the jointless trait in fresh-market and processing breeding programmes involved distinct solutions (Fig. 4b and Supplementary Fig. 4b). In support of this, we observed only *j2*^{TE} *sb3* genotypes in fresh-market and *j2*^{TE} *EJ2* genotypes in processing lines, showing that breeders used *j2*^{TE} by either avoiding the cryptic *ej2*^{W} variant (processing) or selecting for the *sb3* duplication (fresh-market) (Fig. 4c). Consistent with this, a ~350-kb region surrounding *EJ2* showed low levels of genetic diversity (π) in fresh-market cultivars that carry *sb3* (*sb3*$^{+/+}$) compared to those without (*sb3*$^{-/-}$), indicating the *sb3* duplication was actively selected to permit breeding of the jointless trait into large-fruited fresh-market cultivars (Fig. 4d and Supplementary Fig. 4c,d).

We sought to leverage our discovery of *sb3* and its mechanism of suppression to develop strategies that allow predictable breeding for the jointless trait. On the basis of our findings, higher *EJ2* dosage from *sb3* suppresses negative epistasis between *j2* and *ej2*^{W} mutations, thereby protecting plants from undesirable inflorescence branching. In support of this, targeting both copies of *ej2*^{W} on the *sb3* duplication by CRISPR–Cas9 resulted in strongly branched inflorescences in Fla.8924 (Fig. 5a). This suggested that genotypes carrying *sb3* could be exploited to maintain normal inflorescence architecture when generating jointless types by genome editing. We engineered *j2* mutations using CRISPR–Cas9 in three jointed fresh-market Florida breeding lines that carry *sb3* (refs. [32–34]; Fig. 5b,c) and obtained jointless fruit pedicels without disrupting inflorescence development (Fig. 5d). Together, our results show that the *sb3*

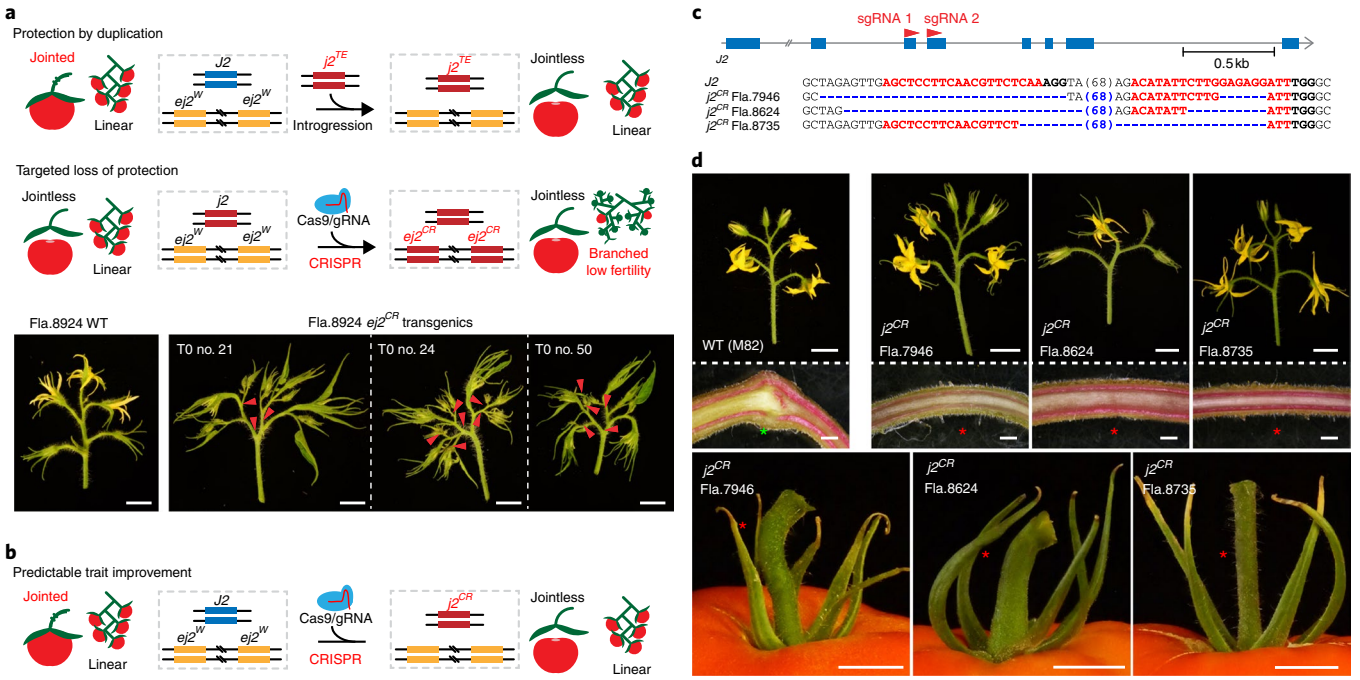

Fig. 5 | The *sb3* duplication enables predictable breeding for the jointless trait by genome editing. a, Schematic showing the *sb3* duplication protects against undesirable inflorescence branching upon introduction of *j2* mutations. Targeting *ej2*^{W} in the *sb3* duplication using CRISPR–Cas9 eliminates protection, resulting in inflorescence branching in *j2* backgrounds. Three independent chimeric first-generation (T_0) *ej2*^{CR} transgenics in Fla.8924 develop branched inflorescences (see also Supplementary Fig. 5). Red arrowheads indicate inflorescence branch points. Representative inflorescences of *ej2*^{CR} transgenics and the WT from three independent transformation experiments are shown. **b**, The *sb3* duplication allows predictable breeding for the jointless trait. **c**, Targeting of *J2* was performed with two single-guide RNAs (sgRNA, red arrowheads). Sequences of CRISPR–Cas9-engineered *j2*^{CR} null alleles in three different *sb3* large-fruited breeding lines are shown.The sgRNA targets and protospacer-adjacent motifs (PAM) are indicated in red and bold font, respectively. Deletions are indicated by blue dashes and sequence gap length (bp) is shown in parentheses. **d**, *j2*^{CR} mutations in three breeding lines containing the *sb3* tandem duplication result in desirable jointless pedicels on flowers and fruits and produce normal unbranched inflorescences. Representative inflorescences of non-transgenic *j2*^{CR} F_2 mutant plants and WT from four independently repeated experiments with similar results are shown. Lignified cells were stained using phloroglucinol to show loss of the abscission zone. Scale bars represent 1 cm in **a** and the top and bottom panel of **d** and 1 mm in the middle panel of **d**. Green and red asterisks mark the presence and absence of a pedicel abscission zone, respectively.

duplication neutralizes negative epistasis between *j2* and the cryptic *ej2*^{W} variant. Thus, introgression of either the wild-type allele of *EJ2* or the *sb3* duplication into diverse genetic backgrounds can now provide predictable breeding for the jointless trait. Although current genome editing technologies do not yet permit creating precise deletions, we expect it will soon be possible to simultaneously mutate *J2* and eliminate the *ej2*^{W} insertion to rapidly engineer improved harvestability in any genotype.

We have shown that selection of a rare tandem duplication during tomato breeding neutralized a deleterious cryptic variant for inflorescence development. The importance of structural variants—particularly those that affect gene copy number—in the domestication and breeding of both plants and animals is becoming more apparent[35]. Gene duplications, and the increase in expression and protein dosage they confer, have been instrumental in modifying productivity traits in maize[36], rice[37] and wheat[38]. Notably, experimental evolution studies in yeast[39], nematodes[40] and plants[41] have shown that gene copy number variation becomes especially important for rapid adaption under strong selective pressure, which mirrors the intense selection imposed during domestication and breeding[35]. Over longer time scales, these and similar as yet uncharacterized duplications and copy number variants in agricultural systems have the potential to further evolve if selection pressure is relaxed. An enlightening example involves evolution of leaf complexity in the Brassicaceae, where a homeodomain gene was tandemly duplicated and one duplicate copy was subsequently dampened in function by a coding sequence mutation[42,43]. This haplotype provided a subtle dosage change and facilitated neofunctionalization from an additional cis-regulatory change that changed expression domains. Similar suppressing or enhancing combinations of coding and regulatory mutations arising after gene duplications may have also been important for achieving more subtle molecular outputs and possibly new crop phenotypes and adaptations. Greatly increasing reference genomes for many related model and crop species, enabled by long-read sequencing technologies, will be critical for exposing the full repertoire of haplotype complexity involving both structural variation and SNPs, as well as for studying the impact of specific haplotypes over the different time scales of evolution, domestication and breeding.

Neutralizing the negative epistasis between *j2*^{TE} and *ej2*^{W} depended not only on the increased gene dosage from the *sb3* duplication but also the second suppressor QTL *sb1*, which is also a cryptic modifier of inflorescence architecture that stabilized normal inflorescence development, flower production and fertility. This genetic complexity, in which epistatic and additive effects from multiple loci and cryptic alleles are involved, probably reflects many as-yet-undetected cases where the exposure and selection for or against cryptic variants have shaped quantitative traits in other crops. The impact of cryptic variation and epistasis on crop domestication and improvement has not been deeply explored[2,15]. Although standing variation comprises an untapped source of alleles to provide gains in productivity and environmental adaptations, our work serves as a cautionary tale for how cryptic variants can complicate predictable breeding, since seemingly neutral mutations can become

adaptive or deleterious when genetically reconfigured[1,44]. Such genetic background dependencies, which are frequently subtle in effect, are pervasive across systems[4,45]. With the rapid advancement of genome editing technologies, it is now feasible to begin exposing epistatic background effects at the population level, by systematically engineering mutations with known phenotypic consequences into tens or even hundreds of related genotypes. Such an endeavour will become especially informative for traits that are controlled by functionally related genes and gene families where epistatic interactions are probably more abundant[46]. Combined with advanced quantitative genetic approaches[47], such strategies could show how genotypic context contributes to phenotypic variability (penetrance and expressivity) and also aid in resolving the responsible cryptic variants to the nucleotide level, many of which could be structural alleles. Future dissection of cryptic genetic variation can help predict how complex genetic architectures shape quantitative trait variation, which is critical for creating superior genotypes in crop and livestock breeding and also for guiding decisions in personalized medicine[4].

Methods

Plant material, growth conditions and phenotyping. Seeds of the *S. lycopersicum* processing cultivar M82 (LA3475) and Florida fresh-market breeding lines Fla.7946, Fla.8624, Fla.8735 and Fla.8924 were from our own stocks. Seeds were either pre-germinated on moistened Whatman paper at 28 °C in complete darkness or directly sown and germinated in soil in 96-cell plastic flats. Plants were grown under long-day conditions (16 h light/8 h dark) in a greenhouse under natural light supplemented with artificial light from high-pressure sodium bulbs (~250 μmol m^{-2} s^{-1}). Daytime and night-time temperatures were 26–28 °C and 18–20 °C, respectively, with a relative humidity of 40–60%. Analyses of inflorescence architecture were conducted on plants grown in the fields at Cold Spring Harbor Laboratory (July 2017), the Cornell Long Island Horticultural Experiment Station (July 2017) and the fields of the Gulf Coast Research and Education Center (March 2018). Seeds were germinated in 96-cell flats and grown for 32 d in the greenhouse before being transplanted to the field. Plants were grown under drip irrigation and standard fertilizer regimes. Damaged or diseased plants were marked throughout the season and were excluded from the analyses.

For quantitative analyses of inflorescence complexity, we counted the number of branching events on 3–5 inflorescences from at least four replicate plants per genotype. Lignified cells in the pedicel abscission zones of flowers and young fruits were stained using a Phloroglucinol-HCl solution (two volumes 2% (w/v) phloroglucinol in 95% (v/v) ethanol and one volume of concentrated HCl). The phloroglucinol-HCl solution was directly applied to pedicels that were freshly cut along the longitudinal axis using a razor blade. Stained pedicles were immediately imaged with a Nikon SMZ1500 stereomicroscope (Nikon).

QTL-sequencing. To map the loci underlying suppression of of *j2^{TE} ej2^{W}* branching in the Fla.8924 breeding line, we generated an F_2 segregating population by crossing a branched *j2^{TE} ej2^{W}* double mutant in the M82 background with the unbranched *j2^{TE} ej2^{W}* mutant in the Fla.8924 background. We followed an established standard QTL-seq protocol[48] and focused on F_2 plants with extreme phenotypes. From a total of 1,536 *j2^{TE} ej2^{W}* × Fla.8924 F_2 plants grown at the Cornell Long Island Horticultural Experiment Station, we selected 35 plants with excessively branched inflorescences (6–36 branches) and 35 clearly suppressed plants (1–4 branches). We further subcategorized these suppressed and branched classes on the basis of average number of branches into 'A' and 'B' classes, branched-A (19–30 branches), branched-B (10–18 branches), suppressed-B (1.5–2.25 branches) and suppressed-A (1–1.25 branches) classes. An equal amount of tissue from each plant (~0.2 g) was pooled for DNA extraction using standard protocols. Libraries were prepared with the Illumina TruSeq DNA PCR-free prep kit from 2 μg genomic DNA sheared to 550 bp insert size. We sequenced all DNA libraries on an Illumina NextSeq platform at the Cold Spring Harbor Laboratory Genome Center.

Genomic DNA reads were trimmed by quality using Trimmomatic[49] and paired reads mapped to the reference tomato genome (SL3.00) using BWA-MEM (refs. [50,51]). Alignments were then sorted with samtools and duplicates marked with PicardTools (ref. [52]; http://broadinstitute.github.io/picard). SNPs were called with samtools/bcftools[52,53] using read alignments for the various genomic DNA sequencing pools from this project in addition to reference M82 (ref. [54]) and Fla.8924 (ref. [21]) reads. Called SNPs were then filtered for bi-allelic high-quality SNPs at least 100 bp from a called indel using bcftools[52]. Following read alignment and SNP calling, all statistics and calculations were done in R (ref. [55]). Read depth for each allele at segregating bi-allelic SNPs in 100-kb sliding windows (by 10 kb) was summed for the various sequencing pools and allele frequencies were calculated. Finally, the difference in allele frequency (SNP index) between sequencing pools was calculated for all pairwise comparison and plotted across the 12 tomato chromosomes. This analysis disclosed two genomic regions that exceeded a genome-wide 95% cut-off in SNP index on chromosomes 1 and 3. Both of these regions exhibited high frequencies of Fla.8924 alleles in the unbranched but not in the branched pool, indicating these two loci underlie the suppression of inflorescence branching in Fla.8924.

To test for copy number variations we determined regions with differences in genome coverage between *j2^{TE}ej2^{W}* and Fla.8924. For this, we calculated genome coverage from Illumina data using bedtools multicov only counting properly paired reads (v.2.26.0) in 10-kb windows across chromosome 3 for *j2^{TE}ej2^{W}*, Fla.8924 and the reference cultivar M82. Depth in the two mutant genotypes was normalized by dividing by the average depth in M82 using R.

QTL analysis of *sb1* and *sb3* in F_3 families. We genotyped F_2 plants derived from a cross between *j2^{TE} ej2^{W}* and Fla.8924 for *sb1* and *sb3* using linked PCR markers (see Supplementary Table 4 for marker information). DNA was extracted from leaf tissue using a standard cetrimonium bromide (CTAB) protocol and PCR was performed using Taq polymerase (NEB) according to the manufacturer's instructions. Two *sb1$^{-/-}$ sb3$^{-/-}$*, two *sb1$^{+/+}$ sb3$^{+/+}$*, one *sb1$^{-/-}$ sb3$^{+/-}$*, two *sb1$^{+/-}$ sb3$^{-/-}$*, one *sb1$^{+/+}$ sb3$^{+/-}$* and one *sb1$^{+/-}$ sb3$^{+/+}$* F_2 plant were selected. These nine F_2 plants were self-pollinated and the F_3 progeny was PCR genotyped for *sb1* and *sb3* with linked markers (see Supplementary Table 4). The progeny was grouped by *sb1 sb3* genotypes and 10–28 plants per genotype were phenotyped for inflorescence branching (number branches per inflorescence; 3–5 inflorescences per plant) in agricultural fields.

Meristem transcriptome profiling. Meristem collection, RNA extraction and library preparation for *s2* mutant plants were performed as previously described[24]. Briefly, we collected seedling shoots at the transition meristem and floral meristem stages of meristem maturation and immediately fixed them in ice-cold acetone. Meristems were manually dissected under a stereoscope and two biological replicates consisting of 20–30 meristems from independent plants were generated. Total RNA was extracted with the PicoPure RNA Extraction kit (Arcturus) and messenger RNA was purified with Dynabeads mRNA Purification kits (Thermo Fisher). Barcoded libraries were prepared using the NEBNext Ultra RNA library prep kit for Illumina according to the manufacturer's instructions and assessed for size distribution and concentration with a Bioanalyzer 2100 (Agilent) and the Kapa Library quantification kit (Kapa Biosystems), respectively. Libraries were sequenced on Illumina NextSeq platform at the Genome Center of Cold Spring Harbor Laboratories. Reads were trimmed by quality using Trimmomatic[49] and aligned to the reference genome sequence of tomato (SL3.0, ref. [56]) using Tophat2 (ref. [57]). Alignments were sorted with samtools[53] and gene expression quantified as unique read pairs aligned to reference annotated gene features (ITAG3.2) using HTSeq-count (ref. [58]). All statistical analyses of gene expression were conducted in R (ref. [55]). Expression of individual genes is shown as transcripts per million (TPM). Significant differential expression between meristem stages in wild-type tomato cultivar M82 was identified with edgeR (ref. [59]) using twofold change, average 1 CPM and FDR ≤ 0.10 cut-offs[27].

Alternative *EJ2* splicing variants, determined by cloning and sequencing RT–qPCR products from *ej2^{W}* and WT, were quantified using RNA-seq data from meristems of the WT (M82), the *j2^{TE}ej2^{W}* double mutant and Fla.8924 at the floral meristem stage of meristem maturation. Reads were trimmed using Trimmomatic[49] and aligned to the reference genome sequence (SL3.0, ref. [56]), a sequence comprising *EJ2* exons 1 to 5 ('total-EJ2 transcript') and a sequence comprising *EJ2* intron 5 including the *ej2^{W}* 564 bp insertion ('*ej2^{W}* transcript') using Tophat2 (ref. [57]). Reads mapping to the total *EJ2* and *ej2^{W}* transcripts were counted using samtools[53] and reads per kilobase of transcript, per million mapped reads (RPKM) values were calculated in R (ref. [55]). The relative amount of functional *EJ2* transcript (delta total-ej2^{W}) for each genotype was calculated by subtracting *ej2^{W}* RPKM from total transcript RPKM values.

Detection of structural variants. To detect the *ej2^{W}* and *j2^{TE}* insertion alleles and the *sb3* duplication using Illumina re-sequencing data[28,60], we adopted a method recently described[61]. In short, for each variant we first used blastn (v.2.2.29+) to detect similar sequences of the variant across the genome. We subsequently extracted the coordinates and masked the region of the reference with N's using bedtools (v.2.17.0). This modified reference was extended with the variant sequence as an extra contig. Next, we extracted the reads and their pairs that are mapped in the region where the variant was expected ±3 kb and all the unmapped pairs per sample. These reads were then mapped to the modified reference genome using BWA-MEM (v.0.7.10–r789). Samtools view (v.0.1.19–44428 cd) was used to count the reads that mapped to the inserted variant sequence represented as the contigs and their pairs to the chromosome of the expected variant site.

To validate the *sb3* duplication in Fla.8924, we sequenced the genome using Oxford Nanopore sequencing according to established protocols[62]. For Nanopore long-read sequencing, high molecular weight DNA was obtained from 21-day-old seedlings that were dark-treated for two days before tissue collection. DNA was extracted from isolated nuclei using a modified CTAB protocol that ensures DNA integrity by minimizing shearing at all steps. Libraries were prepared for MinION

flow cell sequencing according to standard protocols and seven flow cells generated a total of 36.6 giga base pairs (Gb) of data with an average mean read length of 20 kb. Reads were mapped with NGMLR and SVs were called with Sniffles[22], confirming the *sb3* duplication at exactly the same coordinates.

CRISPR–Cas9 mutagenesis, plant transformation and selection of mutant alleles. CRISPR–Cas9 mutagenesis and generation of transgenic plants was performed following our standard protocol[32]. Guide (g)RNAs were designed using the CRISPR-P tool (http://cbi.hzau.edu.cn/cgi-bin/CRISPR) and vectors were assembled using the Golden Gate cloning system as described[63]. We have described selection of gRNA sequences and cloning of binary vectors for targeting *J2* and *EJ2* before[8]. Final binary vectors were transformed into the tomato cultivars M82, Fla.7946, Fla.8624, Fla.8735 and Fla.8924 by *Agrobacterium tumefaciens*-mediated transformation[64]. After in-vitro regeneration of transgenic plants, culture medium was washed from the root system and plants transplanted into soil. Plants were covered with transparent plastic domes and maintained in a shaded area for five days.

In general, first-generation (T_0) transgenics were genotyped for induced lesions using a forward primer 5′ of the sgRNA-1 and a reverse primer 3′ of the sgRNA-2 recognition site. PCR products were separated in agarose gels and selected products were cloned into pSC-A-amp/kan vector (Stratagene). At least five clones per PCR product were sequenced using M13-F and M13-R primer. When targeting *EJ2* on the *sb3* duplication in Fla.8924, we analysed the effect of $ej2^{CR}$ mutations on inflorescence branching directly in chimeric $ej2^{CR}$ T_0 plants. For this, we cloned and sequenced 5–8 *EJ2* alleles from DNA extracted from sepals of three independent $ej2^{CR}$ T_0 plants, which developed branched inflorescences. When targeting *J2* in Fla.7946, Fla.8624 and Fla.8735, we both self-pollinated and backcrossed independent $j2^{CR}$ T_0 plants with lesions to the respective WT parental line. The T_1 and F_1 generation was genotyped for deletion alleles and the absence of the CRISPR–Cas9 transgene using primer binding the 3′ of the 35S promoter and the 5′ of the Cas9 transgene, respectively. T_1 and F_1 plants carrying the engineered $j2^{CR}$ mutant alleles and lacking the transgene were self-pollinated to isolate homozygous, non-transgenic $j2^{CR}$ null mutants from the T_2 and F_2 generation. Phenotypes were observed and documented in T_0, T_2 and F_2 generation plants that lacked WT alleles according to amplicon sequencing. All primers are listed in Supplementary Table 4.

Phylogenetic analyses and measuring signatures of selection. Illumina re-sequencing data from more than 600 tomato accessions[28,60] was downloaded from NCBI-SRA and aligned to the tomato genome reference v.2.50 using Bowtie2 with default parameters. The resulting alignment files were filtered to remove reads mapping to multiple locations using samtools[53] with parameter -q 5 and to remove duplicated reads with Picard MarkDuplicates with default parameters (http://broadinstitute.github.io/picard, parameter REMOVE_DUPLICATES = true). Finally, indels were realigned using GATK RealignerTargetCreator and IndelRealigner successively with default parameters[65].

Alignment files were used to call SNPs at 8,760 positions genotyped in the SolCAP Infinium Chip SNP microarray as indicated in the tomato annotation (ITAG2.4_solCAP.gff3). For this, we ran GATK's UnifiedGenotyper with default parameters in all 602 accessions simultaneously[65]. The SNP matrix obtained was merged with a previous published matrix containing data for 1,008 accessions genotyped at 7,720 positions with the SolCAP SNP array[30]. After filtering accessions not suitable for our analysis and SNPs that did not agree between the two datasets, 5,856 SNPs remained in 1,606 accessions. We obtained a final matrix of 1,812 SNPs by removing SNPs in linkage disequilibrium using PLINK with parameters—mind 0.1—geno 0.1—indep 50 5 3.5 (ref. [66]). A phylogenetic tree was estimated from the final matrix using the ape package in R and the neighbour-joining method including *S. pennellii* LA0716 as a root[67]. The resulting tree was plotted using the ggtree package in R (ref. [68]). Tomato accessions in the tree were classified manually taking into account previously described classifications[30] and their positions in the tree relative to known classifications of species and type (Supplementary Table 2).

To identify signatures of selection in the chromosomal region of *EJ2*, we analysed bi-allelic SNPs in all re-sequenced accessions belonging to the phylogenetic groups with fresh tomatoes ('*S. lycopersicum* fresh/processing', '*S. lycopersicum* fresh' and '*S. lycopersicum* vintage/fresh' in Supplementary Table 2). For this, we first generated gvcf files for each of the 100 accessions classified as fresh tomatoes using GATK HaplotypeCaller with default parameters. Next, we called SNPs in chromosome 3 in all accessions simultaneously using GATK GenotypeGVCFs and we filtered the resulting table to stay only with bi-allelic SNPs with a minor allele frequency greater than 0.05 using vcftools[69] with parameters—remove-indels—maf 0.05—min-alleles 2—max-alleles 2. We finally calculated nucleotide diversity (pi) separately for all accessions with and without the *sb3* duplication using vcftools[69] in windows of 100 kb with steps of 10 kb. We considered a region to be under positive selection when the ratio between the pi values of the accessions without the *sb3* duplication and the pi values of the accessions with the duplication exceeded the top 10% values for the whole chromosome.

Statistical analyses. For quantitative analyses of inflorescence branching, 3–5 inflorescences on at least four plants were analysed per genotype and exact numbers of plants (*n*) are presented in all figures. For expression analyses using RT–qPCR, at least three plants were pooled per tissue sample and at least two RT–qPCR reactions (technical replicates) were performed. The exact number of replicates is given in figure legends. Statistical calculations were performed using R (ref. [62]) and Microsoft Excel. Mean values for each measured parameter were compared using two-way analysis of variance (ANOVA) or two-tailed, two-samples Student's *t*-test, whenever appropriate, with statistical tests listed for each experiment in figure legends.

Reporting Summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability

The DNA sequencing data used to map branching QTLs in the *S. lycopersicum* $j2\ ej2^{W}$ × *S. lycopersicum* Fla.8924 F_2 population and the RNA sequencing data of transition and floral meristem stages for *S. lycopersicum* M82, *S. lycopersicum j2* $ej2^{W}$ and *S. lycopersicum* Fla.8924 has been deposited in SRA (http://ncbi.nlm.nih.gov/sra) under the accession code PRJNA509653.
Source Data files for all main and supplementary figures are available in the online version of the paper. All additional data sets are available from the corresponding author on request.

Received: 7 January 2019; Accepted: 2 April 2019;
Published online: 06 May 2019

References

1. 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).
2. Gibson, G., Dworkin, I. & Hall, G. Uncovering cryptic genetic variation. *Nat. Rev. Genet.* **5**, 1–10 (2004).
3. Paaby, A. B. & Rockman, M. V. Cryptic genetic variation: evolution's hidden substrate. *Nat. Rev. Genet.* **15**, 247–258 (2014).
4. Sackton, T. B. & Hartl, D. L. Genotypic context and epistasis in individuals and populations. *Cell* **166**, 279–287 (2016).
5. Reynard, G. B. New source of the *j2* gene governing jointless pedicel in tomato. *Science* **134**, 4–6 (1961).
6. Rick, C. M. A new jointless gene from the Galapagos *L. pimpinellifolium*. *TGC Rep.* **6**, 23 (1956).
7. Zahara, M. B. & Scheuerman, R. W. Hand-harvesting jointless vs. jointed-stem tomatoes. *Calif. Agric.* **42**, 14–14 (1988).
8. Soyk, S. et al. Bypassing negative epistasis on yield in tomato imposed by a domestication gene. *Cell* **169**, 1142–1155 (2017).
9. Alonso-Blanco, C. et al. 1,135 Genomes reveal the global pattern of polymorphism in *Arabidopsis thaliana*. *Cell* **166**, 481–491 (2016).
10. Auton, A. et al. A global reference for human genetic variation. *Nature* **526**, 68–74 (2015).
11. Aflitos, S. et al. Exploring genetic variation in the tomato (*Solanum* section *Lycopersicon*) clade by whole-genome sequencing. *Plant J.* **80**, 136–148 (2014).
12. Lin, T. et al. Genomic analyses provide insights into the history of tomato breeding. *Nat. Genet.* **46**, 1220–1226 (2014).
13. Le Rouzic, A. & Carlborg, Ö. Evolutionary potential of hidden genetic variation. *Trends Ecol. Evol.* **23**, 33–37 (2008).
14. McGuigan, K. & Sgrò, C. M. Evolutionary consequences of cryptic genetic variation. *Trends Ecol. Evol.* **24**, 305–311 (2009).
15. Lauter, N. & Doebley, J. Genetic variation for phenotypically invariant traits detected in teosinte: implications for the evolution of novel forms. *Genetics* **342**, 333–342 (2002).
16. Mcguigan, K., Nishimura, N., Currey, M., Hurwit, D. & Cresko, W. A. Cryptic genetic variation and body size evolution in threespine stickleback. *Evolution* **65**, 1203–1211 (2011).
17. Pires, N. D. et al. Genetic variation involved in the paternal regulation of seed development. *PLoS Genet.* **12**, e1005806 (2016).
18. Monniaux, M. et al. The role of APETALA1 in petal number robustness. *eLife* **7**, 1–22 (2018).
19. Reynard, G. B. New source of the *j2* gene governing jointless pedicel in tomato. *Science* **134**, 2102 (1961).
20. Boiteux, L. S., Giordano, L., de, B., Furumoto, O. & Aragao, F. A. S. Estimating the pleiotropic effect of the jointless-2 gene on the processing and agronomic traits of tomato by using near-isogenic lines. *Plant Breed.* **114**, 457–459 (1995).
21. Lee, T. G., Shekasteband, R., Menda, N., Mueller, L. A. & Hutton, S. F. Molecular markers to select for the *j-2* –mediated jointless pedicel in tomato. *HortScience* **53**, 153–158 (2018).
22. Sedlazeck, F. J. et al. Accurate detection of complex structural variations using single-molecule sequencing. *Nat. Methods* **15**, 461–468 (2018).
23. Bemer, M. et al. The tomato FRUITFULL homologs TDR4/FUL1 and MBP7/FUL2 regulate ethylene-independent aspects of fruit ripening. *Plant Cell* **24**, 4437–4451 (2012).

24. Park, S. J., Jiang, K., Schatz, M. C. & Lippman, Z. B. Rate of meristem maturation determines inflorescence architecture in tomato. *Proc. Natl Acad. Sci. USA* **109**, 639–644 (2012).
25. Park, S. J., Eshed, Y. & Lippman, Z. B. Meristem maturation and inflorescence architecture - lessons from the Solanaceae. *Curr. Opin. Plant Biol.* **17**, 70–77 (2014).
26. Kyozuka, J., Tokunaga, H. & Yoshida, A. Control of grass inflorescence form by the fine-tuning of meristem phase change. *Curr. Opin. Plant Biol.* **17**, 110–115 (2014).
27. Lemmon, Z. H. et al. The evolution of inflorescence diversity in the nightshades and heterochrony during meristem maturation. *Genome Res.* **26**, 1676–1686 (2016).
28. Zhu, G. et al. Rewiring of the fruit metabolome in tomato breeding. *Cell* **172**, 249–261 (2018).
29. Jeffares, D. C. et al. Transient structural variations have strong effects on quantitative traits and reproductive isolation in fission yeast. *Nat. Commun.* **8**, 1–11 (2017).
30. Blanca, J. et al. Genomic variation in tomato, from wild ancestors to contemporary breeding accessions. *BMC Genom.* **16**, 257 (2015).
31. Rick, C. M. The tomato. *Sci. Am.* **239**, 76–87 (1978).
32. Brooks, C., Nekrasov, V., Lippman, Z. B. & Van Eck, J. Efficient gene editing in tomato in the first generation using the CRISPR/Cas9 system. *Plant Physiol.* **166**, 1292–1297 (2014).
33. Scott, J. W. Fla. 7946 tomato breeding line resistant to *Fusarium oxysporum* f.sp. *lycopersici* races 1, 2, and 3. *HortScience* **39**, 440–441 (2004).
34. Scott, J. W., Hutton, S. F. & Freeman, J. H. Fla. 8638B and Fla. 8624 tomato breeding lines with begomovirus resistance genes Ty-5 plus Ty-6 and Ty-6, respectively. *HortScience* **50**, 1405–1407 (2015).
35. Lye, Z. N. & Purugganan, M. D. Copy number variation in domestication. *Trends Plant Sci.* **24**, 352–365 (2019).
36. Maron, L. G. et al. Aluminum tolerance in maize is associated with higher MATE1 gene copy number. *Proc. Natl Acad. Sci. USA* **110**, 5241–5246 (2013).
37. Wang, Y. et al. Copy number variation at the GL7 locus contributes to grain size diversity in rice. *Nat. Genet.* **47**, 944–948 (2015).
38. Würschum, T., Boeven, P. H. G., Langer, S. M., Longin, C. F. H. & Leiser, W. L. Multiply to conquer: copy number variations at Ppd-B1 and Vrn-A1 facilitate global adaptation in wheat. *BMC Genet.* **16**, 1–8 (2015).
39. Gresham, D. et al. The repertoire and dynamics of evolutionary adaptations to controlled nutrient-limited environments in yeast. *PLoS Genet.* **4**, e1000303 (2008).
40. Farslow, J. C. et al. Rapid Increase in frequency of gene copy-number variants during experimental evolution in *Caenorhabditis elegans*. *BMC Genom.* **16**, 1–18 (2015).
41. Debolt, S. Copy number variation shapes genome diversity in Arabidopsis over immediate family generational scales. *Genome Biol. Evol.* **2**, 441–453 (2010).
42. Vlad, D. et al. Leaf shape evolution through duplication, regulatory diversification, and loss of a homeobox gene. *Science* **343**, 780–783 (2014).
43. Vuolo, F. et al. Coupled enhancer and coding sequence evolution of a homeobox gene shaped leaf diversity. *Genes Dev.* **30**, 2370–2375 (2016).
44. Hickey, J. M., Chiurugwi, T., Mackay, I. & Powell, W. Genomic prediction unifies animal and plant breeding programs to form platforms for biological discovery. *Nat. Genet.* **49**, 1297–1303 (2017).
45. Hou, J., van Leeuwen, J., Andrews, B. J. & Boone, C. Genetic network complexity shapes background-dependent phenotypic expression. *Trends Genet.* **34**, 578–586 (2018).
46. Van Leeuwen, J. et al. Exploring genetic suppression interactions on a global scale. *Science* **354**, aag0839 (2016).
47. Bazakos, C., Hanemian, M., Trontin, C., Jiménez-Goméz, J. M. & Loudet, O. New strategies and tools in quantitative genetics: how to go from the phenotype to the genotype. *Annu Rev. Plant Biol.* **68**, 435–455 (2017).
48. Takagi, H. et al. QTL-seq: rapid mapping of quantitative trait loci in rice by whole genome resequencing of DNA from two bulked populations. *Plant J.* **74**, 174–183 (2013).
49. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2120 (2014).
50. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).
51. Li, H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. Preprint at https://arxiv.org/abs/1303.3997 (2013).
52. Li, H. A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics* **27**, 2987–2993 (2011).
53. Li, H. et al. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).
54. Bolger, A. et al. The genome of the stress-tolerant wild tomato species *Solanum pennellii*. *Nat. Genet.* **46**, 1034–1038 (2014).
55. R Core Team. *R: A Language and Environment for Statistical Computing* (R Foundation for Statistical Computing, 2013); http://www.R-project.org/
56. Tomato Genome Consortium.The tomato genome sequence provides insights into fleshy fruit evolution. *Nature* **485**, 635–641 (2012).
57. Kim, D. et al. TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol.* **14**, R36 (2013).
58. Anders, S., Pyl, P. T. & Huber, W. HTSeq-A Python framework to work with high-throughput sequencing data. *Bioinformatics* **31**, 166–169 (2015).
59. Robinson, M. D., McCarthy, D. J. & Smyth, G. K. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**, 139–140 (2009).
60. Aflitos, S. A. et al. Introgression browser: high-throughput whole-genome SNP visualization. *Plant J.* **82**,174–182 (2015).
61. Dennenmoser, S. et al. Genome-wide patterns of transposon proliferation in an evolutionary young hybrid fish. *Mol. Ecol.* **28**, 1491–1505 (2018).
62. Schmidt, M. H. et al. De novo assembly of a new *Solanum pennellii* accession using nanopore sequencing. *Plant Cell* **29**, 2336–2348 (2017).
63. Werner, S., Engler, C., Weber, E., Gruetzner, R. & Marillonnet, S. Fast track assembly of multigene constructs using Golden Gate cloning and the MoClo system. *Bioeng. Bugs* **3**, 38–43 (2012).
64. van Eck, J., Tjahjadi, P. & Keen, M. *Agrobacterium tumefaciens*-mediated transformation of tomato. *Methods Mol. Biol.* **1864**, 225–234 (2019).
65. McKenna, A. et al. The genome analysis toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* **20**, 1297–1303 (2010).
66. Chang, C. C. et al. Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience* **4**, 7 (2015).
67. Paradis, E., Claude, J. & Strimmer, K. APE: Analyses of Phylogenetics and Evolution in R language. *Bioinformatics* **20**, 289–290 (2004).
68. Yu, G., Smith, D. K., Zhu, H., Guan, Y. & Lam, T. T.-Y. ggtree: an R package for visualization and annotation of phylogenetic trees with their covariates and other associated data. *Methods Ecol. Evol.* **8**, 28–36 (2016).
69. Danecek, P. et al. The variant call format and VCFtools. *Bioinformatics* **27**, 2156–2158 (2011).

Acknowledgements

We thank all members of the Lippman laboratory for valuable discussions. We thank A. Krainer, J. Dalrymple, G. Robitaille and J. Kim for technical support. We thank K. Swartwood for assistance with tomato transformation. We thank T. Mulligan, S. Vermylen, A. Krainer, S. Qiao and K. Schlecht, from CSHL, and staff from Cornell University's Long Island Horticultural Research and Extension Center, for assistance with plant care. We thank S. Goodwin, S. Muller, R. Wappel and E. Ghiban from the CSHL Genome Center for sequencing support. We thank D. Zamir (Hebrew University, Israel) and E. van der Knaap (University of Georgia) for providing seed. This research was supported by an EMBO Long-Term Fellowship (no. ALTF 1589-2014) to S.S., a National Science Foundation Postdoctoral Research Fellowship in Biology Grant (no. IOS-1523423) to Z.H.L., a National Institute of Health Research Project with Complex Structure Cooperative Agreement (3UM1HG008898-01S2) to F.J.S., the ANR grant tomaTE (no. ANR-17-CE20-0024-02) to J.M.J.-G., a Research Grant from BARD (no. IS-4818-15), the United States–Israel Binational Agricultural Research and Development Fund, to Z.B.L., an Agriculture and Food Research Initiative competitive grant no. 2016-67013-24452 of the USDA National Institute of Food and Agriculture to S.H and Z.B.L. and the National Science Foundation Plant Genome Research Program (no. IOS-1732253) to J.V.E., M.C.S. and Z.B.L.

Author contributions

S.S., Z.H.L, F.J.S., J.M.J.-G., S.H., J.V.E., M.C.S. and Z.B.L. designed and planned experiments. S.S., Z.H.L, F.J.S., J.M.J.-G., S.H. and Z.B.L. performed experiments and collected the data. S.S., Z.H.L, F.J.S., J.M.J.-G., M.A., M.C.S. and Z.B.L. analysed the data. S.S., Z.H.L. and Z.B.L. designed the research. S.S. and Z.B.L. wrote the paper with the input from all authors.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information is available for this paper at https://doi.org/10.1038/s41477-019-0422-z.

Reprints and permissions information is available at www.nature.com/reprints.

Correspondence and requests for materials should be addressed to S.S. or Z.B.L.

Journal peer review information: *Nature Plants* thanks Allen Van Deynze, Jianbing Yan and the other anonymous reviewer(s) for their contribution to the peer review of this work.

nature research

Corresponding author(s): S. Soyk and Z.B. Lippman

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see Authors & Referees and the Editorial Policy Checklist.

Statistical parameters

When statistical analyses are reported, confirm that the following items are present in the relevant location (e.g. figure legend, table legend, main text, or Methods section).

n/a	Confirmed	
☐	☒	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒	An indication of whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☒	☐	A description of all covariates tested
☒	☐	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☒	☐	A full description of the statistics including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☒	☐	For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☒	☐	Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated
☐	☒	Clearly defined error bars *State explicitly what error bars represent (e.g. SD, SE, CI)*

Our web collection on statistics for biologists may be useful.

Software and code

Policy information about availability of computer code

Data collection	no software was used for data collection
Data analysis	BWA MEM (v 0.7.10-r789), samtools (v0.1.19-44428cd), blastn (v2.2.29+), bedtools (v2.17.0), R (v3.4.3), RStudio (v1.1.383), Microsoft Excel (v14.4.1), Bcftools (v1.7), GATK (v3.7-0), Trimmomatic (v0.32), Tophat2 (v2.0.12), HTSeq count (v0.9.1), edgeR (v3.20.9), PicardTools (v2.9.4).

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers upon request. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

Figure 1, Figure 3, Supplementary Figure 1, Supplementary Figure 2, Supplementary Figure 3. No restrictions on data availability. All raw data is made available in a Supplementary File "Soyk-etal-SourceData".

Field-specific reporting

Please select the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/authors/policies/ReportingSummary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	Sample sizes were defined based on previous work that found to be sufficient for determining statistically significant results. At least 3 independent replicates analyzed in each experiment. n values and statistical tests are presented
Data exclusions	Mechanically damaged and diseased plants were excluded from the analyses to minimize environmental effects and focus on the genetic control of the observed developmental phenotypes.
Replication	All presented in figure legends and methods. Individual replicates (e.g. tissue samples, plants, shoots, flowers and fruits) are indicated and at least 3 independent replicates analyzed for each experiment
Randomization	For the QTL-sequencing experiment, a segregating F2 mapping population of 1536 individual plants was randomly sown and transplanted in an agricultural field. This randomized and blinded experiment allowed the identification of the suppressor loci sb1 and sb3, which were confirmed in two field-grown experiment that were grouped and labelled by genotype.
Blinding	For the QTL-sequencing experiment, we selected phenotypic extremes from a randomized F2 mapping population of 1536 individual plants and phenotyped these individuals for inflorescence branching. This randomized and blinded experiment allowed the identification of the suppressor loci sb1 and sb3, which were confirmed in two field-grown experiment that were grouped and labelled by genotype.

Reporting for specific materials, systems and methods

Materials & experimental systems

n/a	Involved in the study
☒	☐ Unique biological materials
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology
☐	☒ Animals and other organisms
☒	☐ Human research participants

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging

Animals and other organisms

Policy information about studies involving animals; ARRIVE guidelines recommended for reporting animal research

Laboratory animals	this study did not involve laboratory animals
Wild animals	this study did not involve wild animals

Field-collected samples	Phenotypic data and tissue samples for QTL-sequencing were collected between 11 am and 1 pm on July 8, 2017 in the fields of the Cornell Long Island Horticultural Experiment Station in Riverhead, New York.