nature plants

SUPPLEMENTARY INFORMATION

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

In the format provided by the authors and unedited.

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]*

[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

Supplementary Information for

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

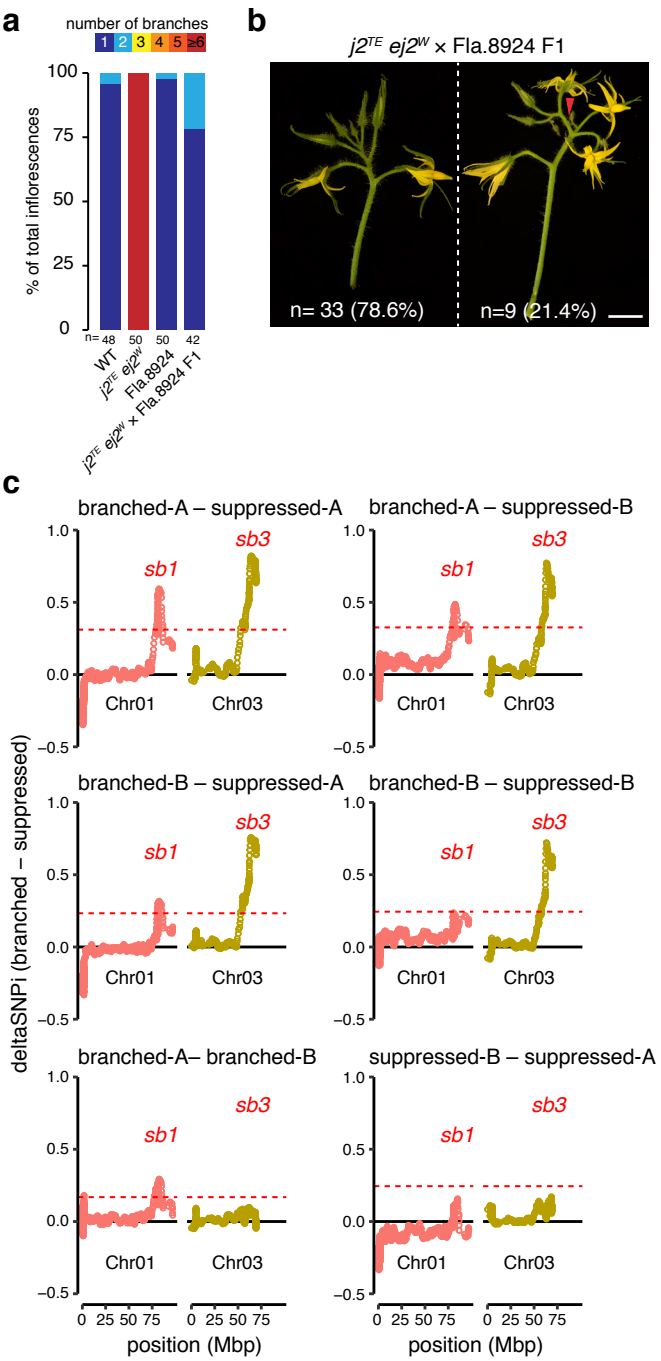
Sebastian Soyk[1,*], Zachary H. Lemmon[1], Fritz J. Sedlazeck[2], José M. Jiménez-Gómez[3], Michael Alonge[4], Samuel Hutton[5], Joyce Van Eck[6,7], Michael C. Schatz[4,8], and Zachary B. Lippman[1,9,*]

* To whom correspondence should be addressed:
lippman@cshl.edu
ssoyk@cshl.edu

This PDF includes

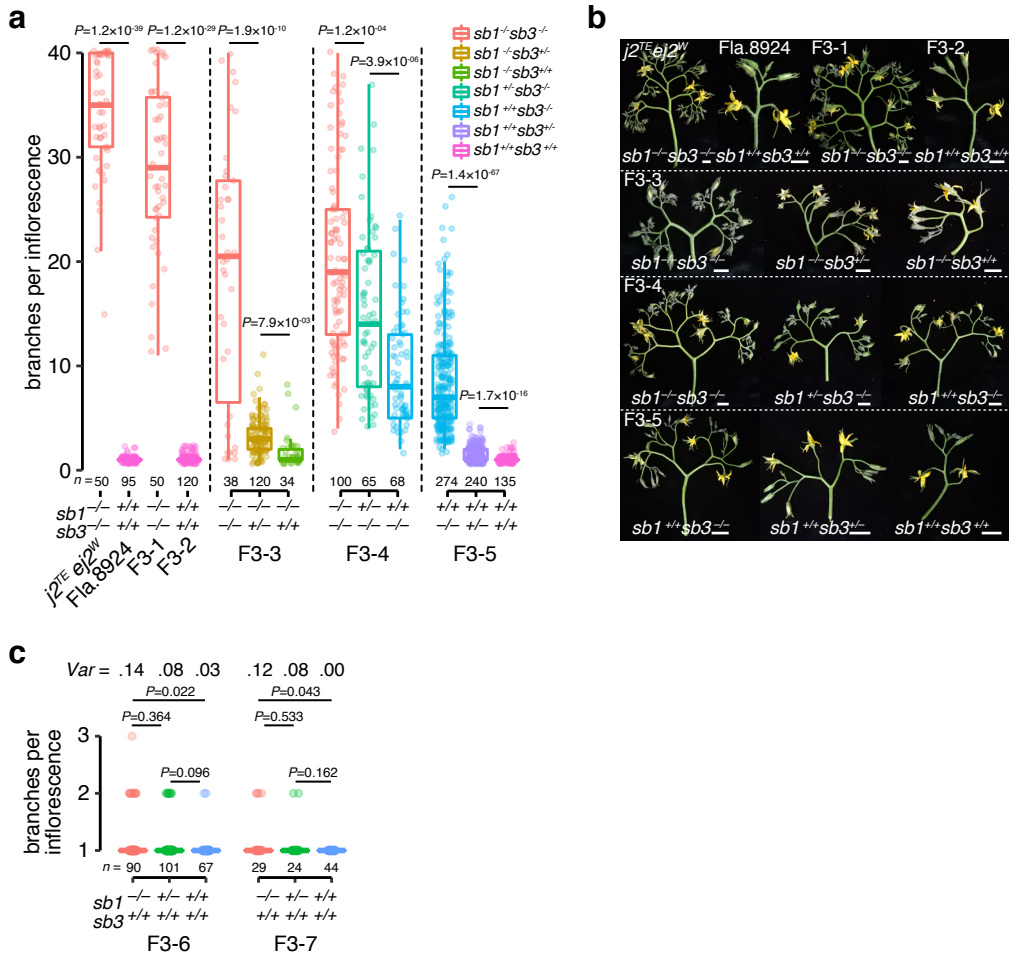
Supplementary Figure 1-5

Supplementary Figure 1



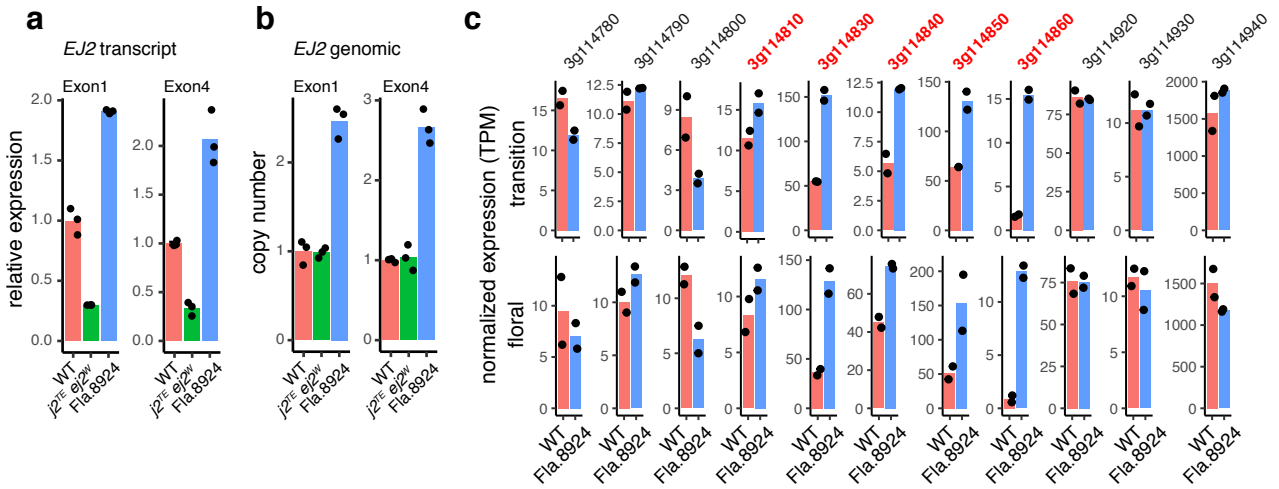
Supplementary Fig. 1 | Excessive inflorescence branching from *j2*^{TE} *ej2*^{W} mutations is strongly suppressed in *j2*^{TE} *ej2*^{W} × Fla.8924 F1 progeny suggesting dominant suppression of negative epistasis from Fla.8924. a, Quantification of branching in WT, *j2*^{TE} *ej2*^{W} and Fla.8924 parents, and the F1 generation. *n* = number of inflorescences. **b**, Representative inflorescences from *j2*^{TE} *ej2*^{W} × Fla.8924 F1 plants. Red arrowheads indicate branch points. n = number of inflorescences; scale bar = 1cm. **c**, QTL-seq of two major *suppressors of branching* (*sb*) loci on chromosome 1 and 3. The difference in SNP index (deltaSNPi) between branched and suppressed pools for chromosome 1 and 3 is shown, suggesting a minor contribution from *sb1* in suppression of branching. Dashed horizontal lines indicate the genome-wide 95% cutoff in SNP index.

Supplementary Figure 2



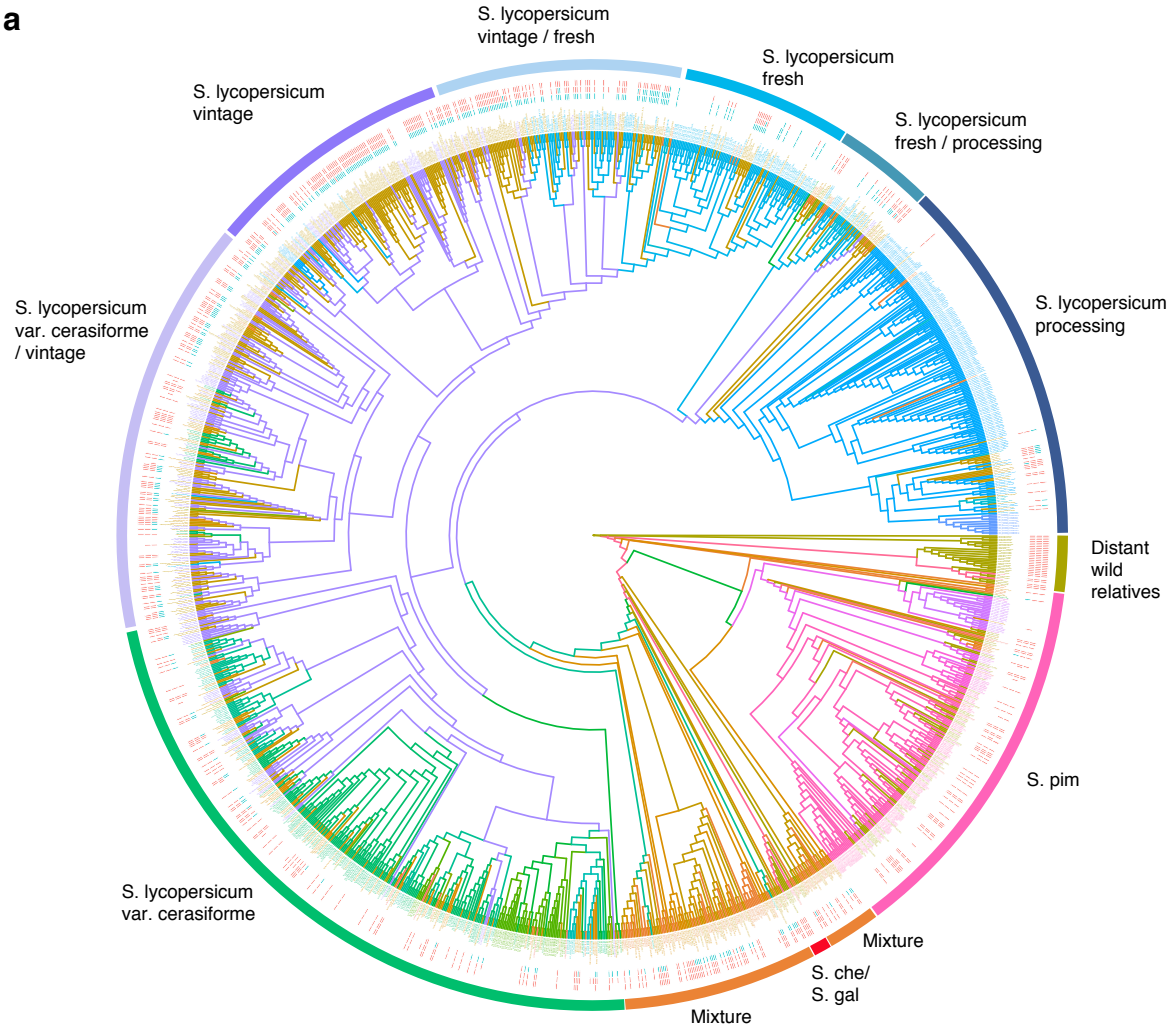
Supplementary Fig. 2 | Individual and combined effects of *sb1* and *sb3* on suppression of *j2*^{TE} *ej2*^{W} inflorescence branching. a, Quantification of inflorescence branching in *j2*^{TE} *ej2*^{W} and Fla.8924 parental genotypes, F3 families homozygous for *sb1* and *sb3* (F3-1, F3-2), and F3 families segregating *sb1* and *sb3* (F3-3 to F3-5). Each data point represents a single inflorescence (*n*). **b**, Representative inflorescences from all genotypes in (**a**) showing different strengths of branching. Scale bars = 1 cm. **c**, F3 families homozygous for *sb3* segregating the minor-effect locus *sb1* (F3-6 to F3-7). Variance (*Var*) is shown at the top. Each data point represents a single inflorescence (*n*). For box plots in (**a**) and (**c**), 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 (**a**) and (**c**) represent two-tailed, two-sample *t* tests.

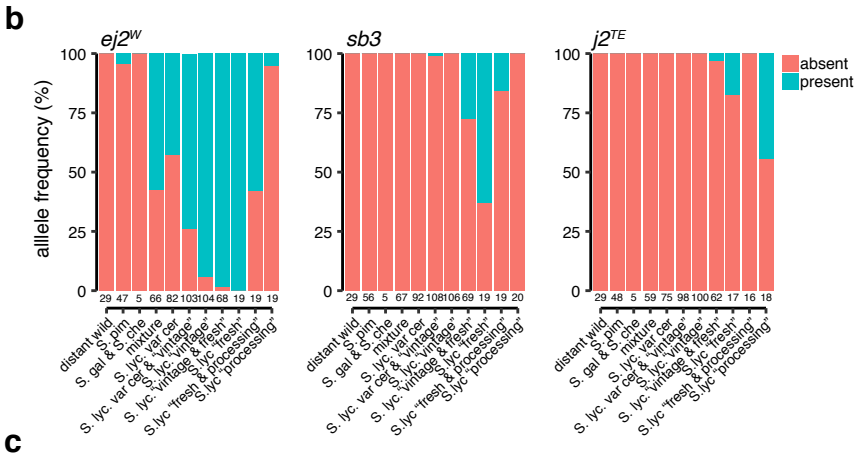
Supplementary Figure 3

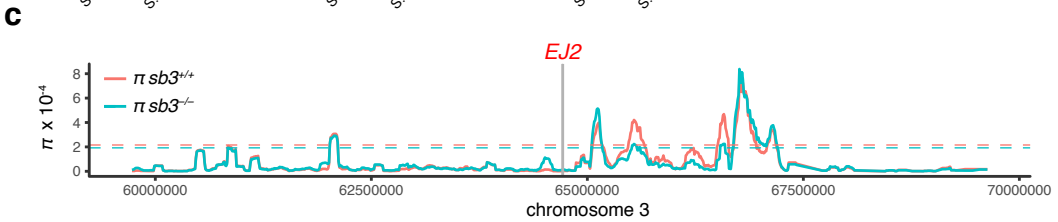


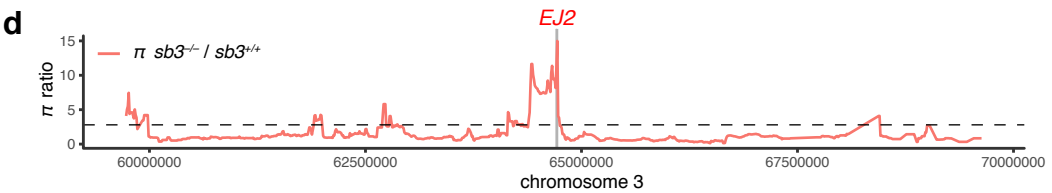
Supplementary Fig. 3 | A duplication containing *ej2*^{w} results in increased *EJ2* expression. a, qRT-PCR reveals *EJ2* expression is increased in floral buds of Fla.8924 compared to floral buds of WT and *j2*^{TE}*ej2*^{w} double mutants. Data is normalized to *UBIQUITIN* (*UBI*) and shown as mean ± s.d.; *n* = 1 pooled tissue sample and 3 technical replicates. **b**, qPCR with genomic DNA confirms a duplication of the *ej2*^{w} allele in Fla.8924. Values are means ± s.d.; *n* = 1 pooled tissue sample and 3 technical replicates. **c**, RNA-seq analysis of meristems at the transition (top) and floral (bottom) stage of meristem maturation from WT and Fla.8924 showing increased expression of each gene in the duplicated region (red font), but not in non-duplicated neighboring genes (black font). Data is shown as means, *n* = 2 biologically independent pooled meristem samples.

Supplementary Figure 4



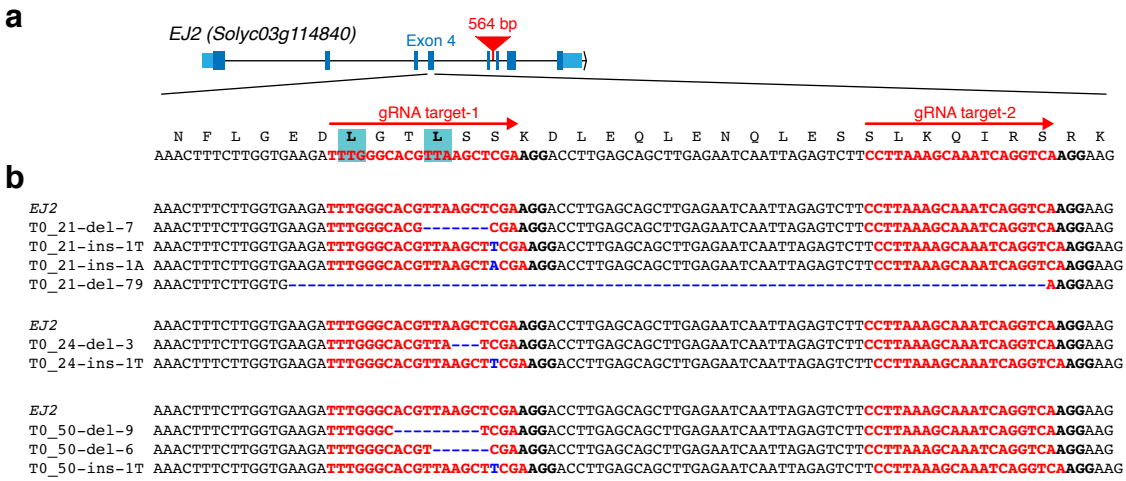






Supplementary Figure 4. Distribution of natural *EJ2* and *J2* alleles and signatures of selection across the tomato germplasm. (**a**) Phylogenetic tree of 1606 accessions constructed with 1812 SNPs, showing that *sb3* is restricted to vintage and modern cultivars. The presence (red) or absence (blue) of *ej2*^{W}, *sb3*, and *j2*^{TE} alleles is indicated by colored lines next to the accession names. (**b**) Allele frequencies of *ej2*^{W}, *sb3*, and *j2*^{TE} in distant wild *Solanum* species, wild tomato relatives (*S. galapagense*, *S. cheesmaniae*, *S. pimpinellifolium*), admixtures, early tomato domesticates (*S. lyc. var. cerasiforme*), and cultivated tomatoes (*S. lycopersicum*). Number of accessions is indicated. (**c**) Nucleotide diversity (pi) plots across a 10Mbp chromosomal region containing *EJ2*, calculated using *n*=100 accessions classified as 'fresh market' tomatoes with (*n*=37, red line) and without (*n*=73, green line) the *sb3* tandem duplication allele. (**d**) The ratio between pi values of accessions without and with the *sb3* tandem duplication. Dashed lines in (c) and (d) indicate the 90% cut-off across the whole chromosome.

Supplementary Figure 5



Supplementary Fig. 5 | Targeting *EJ2* in Fla.8924 by CRISPR-Cas9. a, Targeting of *EJ2* was performed with two single-guide RNA sequences (gRNA, red arrows) in Exon 4. Conserved hydrophobic amino acid residues important for protein-protein interaction are highlighted in turquoise (ref. 70). sgRNA targets and protospacer-adjacent motifs (PAM) are indicated in red and bold font, respectively. **b**, Sequences of CRISPR-Cas9 engineered *ej2*^{CR} alleles in three independent first-generation (T0) plants (#21, #24, and #50) are shown. Deletions are indicated by blue dashes, and sequence gap length (bp) is shown in parentheses. Note in T0 #50 that the in-frame deletions del-6 and del-9 affect a leucine residue important for protein-protein interaction.

Supplementary References

70. Yang, Y., Fanning, L. & Jack, T. The K domain mediates heterodimerization of the Arabidopsis floral organ identity proteins, APETALA3 and PISTILLATA. *Plant J* **33,** 47–59 (2003).