# Cryptic variation fuels plant phenotypic change through hierarchical epistasis

https://doi.org/10.1038/s41586-025-09243-0

Received: 21 February 2025

Accepted: 5 June 2025

Published online: 09 July 2025

**Open access** 

Check for updates

Sophia G. Zebell<sup>1,2</sup>, Carlos Martí-Gómez<sup>1</sup>, Blaine Fitzgerald<sup>1,2</sup>, Camila P. Cunha<sup>3</sup>, Michael Lach<sup>3</sup>, Brooke M. Seman<sup>1,2</sup>, Anat Hendelman<sup>1,2</sup>, Simon Sretenovic<sup>4,6</sup>, Yiping Qi<sup>4</sup>, Madelaine Bartlett<sup>5</sup>, Yuval Eshed<sup>3,2</sup>, David M. McCandlish<sup>1,2</sup> & Zachary B. Lippman<sup>1,2,2</sup>

Cryptic genetic variants exert minimal phenotypic effects alone but are hypothesized to form a vast reservoir of genetic diversity driving trait evolvability through epistatic interactions<sup>1-3</sup>. This classical theory has been reinvigorated by pan-genomics, which is revealing pervasive variation within gene families, cis-regulatory regions and regulatory networks<sup>4-6</sup>. Testing the ability of cryptic variation to fuel phenotypic diversification has been hindered by intractable genetics, limited allelic diversity and inadequate phenotypic resolution. Here, guided by natural and engineered cis-regulatory cryptic variants in a paralogous gene pair, we identified additional redundant trans regulators, establishing a regulatory network controlling tomato inflorescence architecture. By combining coding mutations with cis-regulatory alleles in populations segregating for all four network genes, we generated 216 genotypes spanning a wide spectrum of inflorescence complexity and quantified branching in over 35,000 inflorescences. Analysis of this high-resolution genotype-phenotype map using a hierarchical model of epistasis revealed a layer of dose-dependent interactions within paralogue pairs enhancing branching, culminating in strong, synergistic effects. However, we also identified a layer of antagonism between paralogue pairs, whereby accumulating mutations in one pair progressively diminished the effects of mutations in the other. Our results demonstrate how gene regulatory network architecture and complex dosage effects from paralogue diversification converge to shape phenotypic space, producing the potential for both strongly buffered phenotypes and sudden bursts of phenotypic change.

An enduring debate in evolutionary biology concerns the extent to which small-effect genetic variants contribute to expanding phenotypic diversity from developmentally stabilized (canalized) states<sup>7-10</sup>. The most enigmatic cohort among small-effect variants may be cryptic alleles<sup>1,2</sup>. In their simplest form, cryptic alleles have substantial effects on phenotypes only through interactions with environmental factors or with other alleles, including other alleles that may themselves be phenotypically cryptic<sup>3</sup>. Although hidden at the level of organismal phenotype, cryptic alleles can generate molecular phenotypes through altering protein function or gene expression-and are most likely to accumulate and remain hidden in buffered molecular contexts, such as redundancy within gene families and gene regulatory networks<sup>11</sup>. The accumulation of cryptic alleles in buffered contexts may be a major source of genomic variation shaping network architecture and trajectories of phenotypic evolution. Under this hypothesis, epistatic interactions between previously cryptic alleles may result in the sudden appearance of phenotypic variation in previously invariant traits, facilitating both within-species adaptation and macroevolutionary transitions<sup>11-13</sup>.

Demonstrating the contribution of cryptic variation to trait evolution is challenging. Genetic dissection of trait variation is typically confined to single species or a few closely related ones, in which only narrow ranges of phenotypic diversity can be assessed. Moreover, most dissections of natural trait variation expose only major effect variants, as is also true for developmental genetics in model systems, leaving the influence of cryptic alleles on natural populations and gene regulatory networks largely unexplored<sup>1,14</sup>. Importantly, background dependencies—probably stemming in part from cryptic alleles—are common in evolutionary and developmental genetics<sup>15–18</sup>. Yet, despite these clues, efforts to systematically dissect cryptic variation and its role in phenotypic evolution remain hampered by the complex and often poorly characterized structure and redundancy of gene regulatory networks, limited allelic diversity and restricted phenotypic resolution in most systems.

Genome editing in model systems with complex developmental programs offers a powerful approach to examine cryptic variation. Beyond applications in medicine and agriculture, genome editing enables the engineering of customized mutations and allelic series with wide

<sup>1</sup>Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, USA. <sup>2</sup>Howard Hughes Medical Institute, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, USA. <sup>3</sup>Department of Plant and Environmental Sciences, Weizmann Institute of Science, Rehovot, Israel. <sup>4</sup>Department of Plant Science and Landscape Architecture, University of Maryland, College Park, MD, USA. <sup>5</sup>Sainsbury Laboratory Cambridge University (SLCU), Cambridge, UK. <sup>6</sup>Present address: Stanford Genome Technology Center, School of Medicine, Stanford University, Palo Alto, CA, USA. <sup>5</sup>Ce-mail: yuval.eshed@weizmann.ac.il; mccandlish@cshl.edu; lippman@cshl.edu

ranges of phenotypic effects in isogenic backgrounds<sup>19</sup>. This enables deep exploration of gene function and interactions among different classes of mutations, including *cis*-regulatory variants and paralogue duplications and losses that influence dosage but are often cryptic<sup>20-23</sup>. While pairwise interactions are often detectable, how diverse allelic variants interact across larger regulatory and developmental networks remains unexplored. Here we use genome editing in tomato (*Solanum lycopersicum*) to examine cryptic variation in a genetic network.

Natural variation in the architectures of plant reproductive branching systems (inflorescences) within the Solanaceae family, particularly in the Solanum genus, exemplifies how evolutionary processes generate morphological diversity<sup>24,25</sup> (Fig. 1a). Cryptic variation may shape such trait diversity, such as by limiting possible phenotypic states, making tomato an ideal system to test this hypothesis within a controlled genetic framework. Many tomato mutants-often involving cryptic genetic interactions-mirror natural variation across the family, offering a platform to systematically dissect how cryptic variation can influence inflorescence architecture<sup>26</sup>. We identified epistasis between cryptic natural coding and cis-regulatory variants in two MADS-box transcription factor genes of the SEPALLATA (SEP) clade, which have conserved roles in inflorescence development<sup>27,28</sup> (Fig. 1b). Interactions between mutations in the SEP gene JOINTLESS2 (J2), originating from the wild species Solanum cheesmaniae, and a natural cis-regulatory variant in its paralogue ENHANCER OF JOINTLESS2 (EJ2), result in highly branched inflorescences through classical redundancy epistasis<sup>27</sup>. While individual mutations in each paralogue are cryptic on inflorescence branching, different combinations of homozygous and heterozygous genotypes produce varying degrees of branching effects, reflecting a dose-dependent epistatic relationship (Fig. 1b). Notably, while *EJ2* is conserved across the Solanaceae, *J2* is absent in many species and cultivated genotypes (Supplementary Fig. 1), making these species sensitive to changes in inflorescence architecture from variations in EJ2 dosage<sup>4</sup>. The J2-EJ2 paralogue relationship offers a system to study how epistatic interactions between cryptic alleles in a regulatory network influence trait evolvability. However, realizing this potential first requires identifying additional network components.

### Cryptic variants of the MADS-box gene EJ2

To examine whether natural variation in the epistatic relationship between J2 and EJ2 could further contribute to inflorescence architecture diversity, we mined tomato pan-genome data for variation in the EJ2 promoter<sup>5,6</sup>. Through overlap with open chromatin in tomato reproductive meristems and predicted transcription-factor-binding sites (TFBSs)<sup>29</sup> (Fig. 1c), we filtered 629 candidate variants of greater than 5 bp (Supplementary Table 1 and Supplementary Fig. 2) down to two small deletions with nearby single-nucleotide variants (SNVs) that coincided with a cluster of three TFBSs located 6 kb upstream of *EJ2* (Fig. 1c,d and Supplementary Table 2). These variants were found only in the wild species *Solanum habrochaites* and *Solanum pennellii*, which produce weakly branched inflorescences (Fig. 1d,e).

Tomato introgression lines carrying wild-species chromosomal segments with these EJ2 variants in isogenic backgrounds (LA3925, designated  $EJ2^{Sh}$ ; and IL3-4, designated  $EJ2^{Sp}$ ) rarely exhibit branching (Fig. 1d,e). However, branching increased with the addition of the j2 mutation.  $EJ2^{Sh}j2$  plants exhibited a subtle but significant increase in branching compared with the  $EJ2^{Sh}$  genotype (Fig. 1d), and  $EJ2^{Sp}j2$  plants produced an average of five branches per inflorescence (Fig. 1e and Supplementary Tables 3 and 4). As introgression lines have large chromosomal segments carrying additional variants, we used CRISPR to test whether branching specifically resulted from TFBS disruption by attempting to create similar deletions in the j2 mutant background. Owing to the absence of Cas9 gRNA target sites in the 25 bp region, we first used the more permissive Cas9 SpRY variant, which recognizes an expanded protospacer-adjacent motif<sup>30,31</sup>. This approach, which used

three gRNAs and catalytically active and dead versions of Cas9 fused to an adenine base editor, produced three alleles, each with a single SNV within one TFBS. As none of these single-nucleotide changes led to branching (Supplementary Fig. 2a, b and Supplementary Tables 5 and 6), we targeted a 153 bp target region flanking the TFBSs using four conventional Cas9 gRNAs. We recovered seven E/2 promoter (E/2<sup>pro</sup>) alleles (Fig. 2a): five with small indels and SNVs, often at the gRNA target sites, and two with overlapping ~100 bp deletions spanning the entire interval. Notably, none of these alleles exhibited branching in the functional /2 background, but all caused branching in the mutant i2 background, displaying a continuous range of effects (Fig. 2b and Supplementary Tables 7 and 8). Moreover, none of these genotypes exhibited the pleiotropic phenotypes observed in loss of function double mutants, such as enlarged sepals or altered fruit shape<sup>27</sup> (Supplementary Fig. 3 and Supplementary Tables 9-12). We targeted two additional promoter regions with open chromatin or sequence conservation located 1.6 kb and 2.1 kb upstream. Four deletion alleles of varying sizes were generated in each region, but only those affecting the second region, including an approximately 3.3 kb deletion, produced mild branching (1-5 branches on average; Supplementary Fig. 2 and Supplementary Tables 13-16). These findings indicate that several promoter segments regulate EJ2, and the TFBSs disrupted in EJ2<sup>sh</sup> and EJ2<sup>sp</sup>, together with nearby sites removed in our engineered alleles, act collectively to positively regulate EJ2 expression and control inflorescence development.

### PLT paralogues regulate EJ2 and branching

Our finding that multiple cis-regulatory cryptic alleles caused branching, partly due to the disruption of the TFBSs affected by two natural alleles, prompted us to investigate whether the transcription factors predicted to bind to these sites directly regulate E/2 expression and inflorescence architecture. Both the S. habrochaites and S. pennellii EJ2 cis-regulatory alleles are predicted to disrupt binding sites for the DOMAIN OF UNKNOWN FUNCTION (DOF) and the PLETHORA (PLT) transcription factor families. Members of these families have been implicated in meristem development in Arabidopsis thaliana (PLT)<sup>32</sup> and flowering in tomato (DOF)<sup>33</sup>. Using our tomato inflorescence development expression atlas<sup>34</sup>, we searched for DOF and PLT genes expressed during key developmental stages. Among the 34 DOF genes in tomato, SIDOF9 emerged as a primary candidate (Supplementary Fig. 4), because engineered mutants of this gene develop more flowers on inflorescences with weak branching<sup>33</sup>. However, our CRISPR mutants exhibited a substantial change in leaf shape but did not show branching, either alone or in the j2 background (Supplementary Fig. 4 and Supplementary Tables 17 and 18).

We next focused on two closely related PLT paralogues (SIPLT3 and SIPLT7; hereafter, PLT3 and PLT7), expressed in meristems during and after floral transition, similar to J2 and EJ2 (Fig. 3a). These tomato PLT paralogues are orthologues of A. thaliana AtPLT3 and AtPLT7, but arose from independent duplications (Fig. 3b and Extended Data Fig. 1). In Arabidopsis, AtPLT3 and AtPLT7 function in meristem maturation as well as floral organ identity and growth<sup>35</sup>. We tested whether the PLT proteins bind to the EJ2 promoter and activate its expression by performing a heterologous dual-luciferase assay in tobacco (Nicotiana benthamiana) leaves. Although the full-length coding sequence of PLT7 could not be cloned or synthesized, PLT3 strongly activated the intact EJ2 promoter but not the EJ2<sup>pro8</sup>, EJ2<sup>pro-Sh</sup> and EJ2<sup>pro-Sp</sup> alleles, which have mutated PLT-binding sites (Fig. 3c and Supplementary Tables 19 and 20). For normalization, we included ETHYLENE RESPONSIVE12 (ERF12) as a non-binding control, a transcription factor that is expressed in meristems with a single DNA-binding domain that is structurally similar to those in PLTs.

We mutated both *PLT* paralogues using CRISPR–Cas9. *plt7* single mutants appeared like the wild type (WT), whereas *plt3* mutants



**Fig. 1** | **Natural** *cis*-**regulatory variants of the** *SEP* **gene***EJ2***. a**, The increased complexity of inflorescences of *Solanum* species. **b**, The dose-dependent redundancy relationship among the *SEP* paralogues *EJ2* and *J2* in controlling tomato inflorescence architecture. Inset: proliferated *ej2j2* meristem. Scale bar, 1 cm. **c**, A 6 kb region upstream of the *EJ2* transcription start site showing open chromatin (blue), conserved non-coding sequences (CNSs, dark blue), predicted TFBSs (orange) and pan-genome variants (light orange). **d**, The *cis*-regulatory region of *S. lycopersicum* and *S. habrochaites EJ2* with one *DOF* and two *PLT* TFBSs. The overlapping *DOF-PLT* site is disrupted by a 10 bp deletion in *S. habrochaites*. Quantification of inflorescence branching for the three indicated genotypes (middle) is followed by representative images of *S. habrochaites*, the

*EJ2*-containing introgression line (LA3925, *EJ2*<sup>sh</sup>) and *EJ2*<sup>sh</sup> in the *j2* background (*EJ2*<sup>sh</sup> *j2*; right). **e**, The *cis*-regulatory region of *S. pennellii*, showing disruption of all three TFBSs by an 8 bp deletion and linked SNVs, quantification of inflorescence branching for the three indicated genotypes (middle) and representative images of *S. pennelli*, the *EJ2*-containing introgression line (IL3-4, *EJ2*<sup>Sp</sup>) and *EJ2*<sup>Sp</sup> in the *j2* background (*EJ2*<sup>Sp</sup>*j2*; right). For **d** and **e**, the dashed red lines indicate deleted sequences; the blue text shows deletion sizes and SNVs. The area of grey circles shows the number of inflorescences quantified. Data are mean  $\pm$  s.d. *n* values represent the total number of inflorescences. *P* values were calculated using two-sided Dunnett's compare with control test. The red arrowheads in **b**, **d** and **e** mark branch points. Scale bar, 1 cm.

produced inflorescences with few branches. However, double mutants exhibited extreme meristem overproliferation and branching. These mutant alleles also displayed dose-dependent redundancy: *plt3/+ plt7* genotypes showed weak branching, whereas *plt3 plt7/+* genotypes

exhibited moderate branching (Fig. 3d,e and Supplementary Tables 21 and 22). The binding assays, combined with the quantitative effects of the *plt* mutant genotypes and the intermediate branching observed in all  $EJ2^{pro}j2$  genotypes, suggest that the PLTs transcriptionally regulate





**Fig. 2** | **Engineered** *EJ2 cis*-regulatory alleles pinpoint a branching-specific enhancer. a, The 153 bp target region located 5.7 kb upstream of the *EJ2* translation start site showing annotated TFBSs (orange), including the focal *DOF* and *PLT* binding sites, and four CRISPR–Cas9 gRNAs (light blue) used for generation of the *EJ2* promoter (*EJ2<sup>pro2</sup>*) allelic series (left). Two lines with strong phenotypes (*EJ2<sup>pro2</sup> j2* and *EJ2<sup>pro8</sup> j2*) carried overlapping approximately 100 bp deletions that removed the *DOF-PLT* binding sites, as well as other lower-confidence predicted binding sites (Supplementary Tables 2 and 26).

*EJ2* and probably other genes involved in inflorescence development. This aligns with the presence of PLT-binding sites in the *cis*-regulatory regions of *J2* (Supplementary Table 23).

To further characterize the functional relationships between the *PLT* and *SEP* paralogues, we profiled and compared gene expression in proliferating meristems from the two double mutants. A principal component analysis (PCA) of the top 200 maturation marker genes from WT meristem stages<sup>34</sup> revealed that both double mutants cluster closest to the floral meristem maturation stage and also with the *anantha* (*an*) mutant, which overproliferates floral meristems due to a mutation in the orthologue of the *Arabidopsis UNUSUAL FLORAL ORGANS* gene (Fig. 3f). The expression data also showed that *J2* and *EJ2* have very low expression in *plt3 plt7* meristems, whereas both *PLT* paralogues remain expressed in *j2 ej2* meristems, supporting regulation of these *SEP* genes by the PLTs (Fig. 3g).

### A PLT-SEP genotype-phenotype map

Our genetic and molecular analyses identified key components of an inflorescence regulatory network comprising two duplicated transcription factor pairs with dose-dependent epistatic interactions. We generated multiple alleles for these factors in a shared isogenic background, including both promoter regulatory site alleles and strong coding alleles, all of which are cryptic individually. This genetic resource enabled us to systematically explore the phenotypic space and quantify the functional output arising from variations within a simple, rapidly The dashed red lines show deleted sequences; deletion sizes and SNVs are shown in blue font. Right, quantification of inflorescence branching for each genotype. The area of the grey circles shows the number of inflorescences. Data are mean ± s.d. *n* values represent the total number of observed inflorescences. *P* values were calculated using two-sided Dunnett's compare with control test. NS, not significant. **b**, Representative images of *j2* and five of the *EJ2<sup>pro</sup>* alleles in the *j2* background, capturing the range of branching effects. The red arrowheads mark branch points. Scale bars, 1 cm.

evolving network involving the interactions among *PLT3* and *PLT7*, and their downstream targets, *J2* and *EJ2* (Fig. 4a).

To examine the genotype-phenotype map of this tomato inflorescence regulatory network, including the interplay between cryptic dosage effects and paralogous epistatic relationships, we selected six EJ2pro j2 lines spanning a range of branching effects and crossed them with *plt3 plt7/+* plants to generate six  $F_2$  segregating populations (Fig. 4b and Methods). These populations provided 216 genotypic combinations, enabling an in-depth phenotypic and statistical analysis of branching effects from single, double and higher-order mutants, as well as dosage effects from heterozygosity (Fig. 4c). Across four field seasons in two environments, we quantified a total of 35,606 inflorescences (Supplementary Table 24). Preliminary analyses indicated that there was a greater variance in the number of branching events per inflorescence both within plants and within genotypes than would be expected if branching events were Poisson distributed (Supplementary Fig. 5a,b). Consequently, we treated branching events as overdispersed count data in all subsequent analyses, which provided a significantly improved fit relative to a Poisson error model (likelihood ratio test for negative binomial versus Poisson,  $P < 10^{-16}$ ; Methods). For illustration, data from a single population and field season are shown in Fig. 4c, with the full dataset available in Supplementary Table 24.

Using the data from this large set of crosses, we sought to determine how mutations within this genetic network combine to determine the mean number of branching events per inflorescence for any given genotype. We began by comparing a model in which mutations at different



**Fig. 3** |*PLETHORA* transcription factor paralogues control inflorescence architecture upstream of *EJ2/J2*. a, The expression dynamics of *EJ2,J2*, *PLT3* and *PLT7* over five developmental stages (grey): early vegetative meristem (EVM), middle vegetative meristem (MVM), late vegetative meristem (LVM), transition meristem (TM) and floral meristem (FM). b, Pruned phylogenetic tree of PLT3 and PLT7 proteins. The branch thickness reflects bootstrap support: thin lines indicate support of <75; medium lines indicate support of 75–90; and thick lines indicate support of <90. The scale bar represents the number of amino acid substitutions per site; presumed duplication and speciation nodes are denoted by grey and black circles, respectively (the full phylogeny is provided in Extended Data Fig. 1). c, In planta heterologous expression of firefly luciferase driven by WT and *EJ2 cis*-regulatory alleles, co-transfected with PLT3 and normalized to an internal constitutive *Renilla* luciferase and non-binding protein control (NBC). The open circles represent individual biological replicates (*n* = 5). The error bars represent the s.e. *P* values were calculated using two-sided Dunnett's

of branching in double *plt3/7* mutants, including single and homozygous/ heterozygous combinations. The grey circles correspond to the number of inflorescences. Data are mean ± s.d. *n* values represent the number of observations. *P* values were calculated using two-sided Dunnett's compare with control test (blue font). **e**, Representative inflorescences of *plt3* and *plt7* mutant combinations. Inset: double-mutant inflorescence showing meristem overproliferation (×5 magnification). The red arrowheads mark branch points. Scale bars, 1 cm. **f**, PCA of the top 200 genes differentially expressed (RNA sequencing) over meristem maturation in staged WT meristems and proliferating meristems of the mutant genotypes: *ej2j2*, *plt3 plt7* and *an*. **g**, Expression of the paralogues *EJ2*, *J2*, *PLT3* and *PLT7* in proliferating meristems of the mutants *ej2j2* and *plt3 plt7* on the log[1 + transcripts per million (TPM)] scale. The error bars represent the s.e. *n* = 4 (*ej2j2*) and 2 (*plt3 plt7*) samples.

loci combine additively versus a model where mutations at different loci combine multiplicatively (with both models accounting for dominance interactions within each locus). The model with multiplicative effects fit substantially better (59.40% deviance explained by an additive model with a log link versus 46.05% deviance explained by an additive model with an identity link; Supplementary Fig. 5c,d), suggesting an overall tendency for mutations at different loci to interact multiplicatively in this system. Using this multiplicative model as a baseline, we then fit a more

complex model to determine whether epistatic interactions between loci are also present. We found that a pairwise interaction model that included these epistatic interactions significantly outperformed the multiplicative model (83.11% deviance explained; likelihood ratio test,  $P < 10^{-16}$ ) and achieved greater predictive performance for held-out seasons and genotypes (Supplementary Fig. 6; average leave-one-out cross-validated  $R^2$  on held-out seasons was 0.89 for the pairwise model compared to 0.70 for the multiplicative model). This pairwise model



Fig. 4 | See next page for caption.

detected pervasive epistasis across loci, with 43 out of 80 epistatic terms showing *P* values of below 0.05 (Supplementary Table 25). Among the most notable of these interactions were strong, super-multiplicative, synergistic interactions between *plt3* and *plt7* as well as between the *EJ2*<sup>pro</sup> alleles and *j2*, such that combining mutations within a paralogue pair often results in a mean number of branching events several fold (threefold to ninefold) in excess of what would be expected when multiplying the effects of the individual mutations (Table 1). While we observed strong positive interactions within paralogue pairs, consistent with our previous observation regarding at least partial redundancy, we also inferred many negative interactions between non-paralogous pairs (11 out of 14 additive-by-additive epistatic terms between non-paralogous pairs were negative and showed *P* values of below 0.05; Supplementary Table 25). To better understand the structure of the genotype–phenotype relationship implied by these negative interactions, we first calculated maximum-likelihood estimates for the **Fig. 4** | **A hierarchical model of genetic interactions explains** *PLT-SEP* **inflorescence genotype-phenotype map. a**, Simplified architecture of the *PLT3/7-EJ2/J2* genetic and molecular network. The arrows represent positive regulation. The solid arrows are supported by genetic and molecular evidence. The dashed arrows are supported by annotated binding sites. b, Six F<sub>2</sub> populations with different *EJ2<sup>pro</sup>* alleles were phenotyped across multiple seasons and environments, yielding observations of per-inflorescence branching events for 216 different genotypes. c, Representative data for inflorescence branching from a single season of one of six populations, on a log(1+x) scale. The area of grey circles represents the number of inflorescences. Data are mean  $\pm$  s.d. *n* values represent the total number of observations. Genotypes are represented as WT (white), mutant (black) or heterozygous (grey). **d**, **e**, Maximum-likelihood estimates (MLEs) of the mean number of branching events for combinations of mutations in *EJ2* and *J2* (**d**) or *PLT3* and *PLT7* (**e**) versus the WT across different genetic backgrounds of mutations in

average number of branching events per inflorescence across the 470 genotype-season combinations in our dataset. We then plotted how the log-transformed phenotypes conferred by different combinations of j2 and EJ2<sup>pro</sup> mutations are transformed when placed on different plt3 and plt7 mutant backgrounds as compared to a WT PLT3 and PLT7 background. Under a multiplicative model of interaction across paralogue pairs, these quantities are expected to be linearly related with a slope of exactly one and an intercept that varies with the strength of the genetic perturbation at the PLT3 and PLT7 loci. Notably, although we observed that the relationship between backgrounds is always linear, the slope of this linear relationship was not constant, but decreased as mutations accumulated at the PLT3 and PLT7 loci (Fig. 4d and Extended Data Fig. 2). We observed the same pattern when analysing the phenotypes resulting from plt3 and plt7 mutant combinations across different j2 and EJ2pro8 backgrounds, with slopes decreasing as the strength of the EJ2J2 perturbations increased (Fig. 4e). Importantly, this pattern was consistent across all six EJ2pro alleles (Extended Data Fig. 2) and remained robust across other methods of estimating average phenotype (Extended Data Fig. 3). Collectively, these analyses suggest a very simple form of genetic interaction between the paralogue pairs, whereby mutations in one paralogue pair linearly re-scale the effects of mutations in the other. As the slope of the linear relationship reduces almost to zero in highly mutated genetic backgrounds, the negative interaction coefficients between the paralogue pairs reflect a systematic pattern of masking interactions between non-paralogous mutations, whereby the effects of mutations in one pair are diminished when the other pair is highly mutated.

### Modelling hierarchical epistasis

This mode of genetic interaction, in which each mutation systematically re-scales the effects of other mutations, is a defining feature of a classic theoretical model of epistasis known as the multilinear model<sup>36</sup>. To simultaneously capture the supermultiplicative synergistic interactions within paralogue pairs together with the systematic masking interactions between paralogue pairs, we fit an additional model that we call the hierarchical epistasis model because it treats this multilinear interaction between paralogue pairs as an additional layer of epistasis. In this model, similar to the pairwise model, we allowed an arbitrary pattern of dominance and epistasis within each paralogue pair, but the predictions based on the coefficients for each pair of paralogues are combined according to a multilinear interaction. Crucially, because, for a multilinear interaction, mutations in one of the pairs of paralogues simply rescale the effects of mutations in the other pair, this multilinear interaction between paralogue pairs uses only a single parameter to describe the extent to which mutations in one paralogue pair mask the effects of mutations in the other pair, in contrast to our more traditional pairwise interaction model that uses 42 parameters to describe *PLT3* and *PLT7* (**d**) or *EJ2<sup>pro8</sup>* and *J2* loci (**e**). The black dashed lines represent the total least-squares regression lines. n = 10 inflorescences per plant; further details are provided in the Methods. **f**, Comparison of the predicted number of branching events by the hierarchical model and the MLE for each genotype–season combination. The error bars represent 95% confidence intervals. Genotype–season combinations with 95% confidence intervals wider than a thousand-fold range are not shown (**d-f**). **g**, Representation of the inferred hierarchical epistasis model. The surface shows how phenotypes resulting from perturbing each paralogue pair in the WT background are combined to determine the average phenotype of genotypes with mutations in both pairs. The grey dots represent the multiplicative expectation for the designated within-paralogue mutant combinations. **h**, Predicted branching events under the hierarchical model in specific genotypes as a function of their mutational distance to the WT. The grey dots indicate the multiplicative expectation.

interactions between the paralogue pairs (that is, additive-by-additive, dominance-by-dominance and additive-by-dominance parameters for each of the 14 non-paralogous pairs of mutations). Finally, after the within-paralogue pair interactions are combined across paralogue pairs using a multilinear interaction, the results are transformed once more using an exponential function (that is, we again use a log link, identical to that used for the pairwise model; additional technical details on the hierarchical epistasis model are provided in the Methods and Extended Data Fig. 4a). We applied this hierarchical epistasis model to our data and found that it recapitulated the observed phenotypic measurements nearly as well as the full pairwise model (82.14% deviance explained; Fig. 4f), and maintained high predictive power for held-out data from complete seasons and unobserved genotypes (Supplementary Fig. 6; average leave-one-season-out cross-validated  $R^2 = 0.89$ ), despite the substantially reduced number of parameters. Using the Akaike information criterion (AIC) to compare the two models<sup>37,38</sup>, we found that the hierarchical model is roughly 18,358 times more likely than the pairwise model (AIC<sub>hiearchical</sub> – AIC<sub>pairwise</sub> = -19.64).

To provide a more intuitive understanding of the behaviour of this hierarchical model, it is useful to visualize the combination of interaction between paralogue pairs and the exponential mapping as a response surface<sup>39</sup> that depicts the predicted phenotype of a combination of mutations across both paralogue pairs as a function of the phenotypes conferred when mutating each paralogue pair separately (Fig. 4g and Extended Data Fig. 4b). In this response surface, the phenotypes conferred by within-pair mutations are shown on a log scale so that the effects of mutations that combine according to the multiplicative expectation would add along each axis. However, consistent with the pairwise model, interactions within each paralogue pair show a strong pattern of synergy (Fig. 4h (left); see Extended Data Fig. 4c for all other within-paralogue pair interactions), such that double mutants are substantially further displaced along each axis than would be expected from the individual effects of the single mutants (Fig. 4g; multiplicative expectation is shown by grey dots). The key difference from the pairwise model is that the interactions between paralogue pairs, instead of being represented by numerous between-pair interaction terms, are given by a surface (Fig. 4g) of which the shape is controlled by a single parameter that determines the overall strength of the interaction between the paralogue pairs, and that has horizontal and vertical transects that are exponential due to the final exponentiation (Extended Data Fig. 4c). In particular, we see that, as the phenotypic effect of one paralogue pair increases, the corresponding transect for mutations in the other pair becomes progressively flatter, reflecting diminishing phenotypic contributions from additional mutations (Extended Data Fig. 4d). While this masking effect is modest for mutations with moderate phenotypic impacts (for example, the 11.24-fold effect of a homozygous plt3 mutation in a WT background is reduced to 6.18-fold in a j2 background; Fig. 4h (top right)), it becomes far more pronounced

#### Table 1 | Estimates of synergistic epistasis

Genotype	Fold change in excess of multiplicative expectation	95% confidence interval	Р
plt3/+plt7	3.29	2.64-4.10	1.44×10 <sup>-26</sup>
plt3plt7/+	3.27	2.65-4.04	2.95×10 <sup>-28</sup>
EJ2 <sup>pro3</sup> j2	6.81	4.21-11.03	5.16×10 <sup>-15</sup>
EJ2 <sup>pro4</sup> j2	9.32	5.25-16.57	2.74×10 <sup>-14</sup>
EJ2 <sup>pro1</sup> j2	7.05	5.21-9.55	9.91×10 <sup>-37</sup>
EJ2 <sup>pro7</sup> j2	16.97	8.14–35.38	4.16×10 <sup>-14</sup>
EJ2 <sup>pro8</sup> j2	5.95	4.42-8.03	1.13×10 <sup>-31</sup>
EJ2 <sup>pro6</sup> j2	6.49	4.56-9.23	3.10×10 <sup>-25</sup>

Estimates of fold change are shown relative to the expectation under the multiplicative model for combinations of perturbations within each paralogue pair under the pairwise interaction model. *P* values calculated using two-sided *z*-tests for the specified contrasts were estimated after fitting a pairwise interaction model with negative binomial likelihood using statsmodels<sup>55</sup>. The full set of estimated parameters is provided in Supplementary Table 21.

in highly mutated backgrounds (for example, adding *plt3 plt7/+* to a WT *EJ2J2* background has an 81.30-fold effect, but only a 2.49-fold effect in the *EJ2<sup>pro8</sup> j2* background Fig. 4h (bottom right)). Overall, these results show how synergy within gene families can result in an opening of phenotypic space in which previously cryptic mutations have strong effects, but also show how this expansion of phenotypic space begins to close as accumulated mutations in one part of a genetic network increasingly mask the effect of mutations in other parts.

### Discussion

Here we used natural and engineered cryptic *cis*-regulatory variation to identify and functionally dissect transcription factor regulators of tomato inflorescence development in a four-gene regulatory network. Engineering additional cryptic *cis*-regulatory alleles with continuous epistatic effects and densely sampling thousands of inflorescences from hundreds of combinatorial genotypes enabled us to resolve the genetic architecture and genotype–phenotype map of this network. Mutations within this network tended to interact multiplicatively, but with even stronger positive synergistic (that is, super-multiplicative) interactions within paralogous gene pairs, consistent with frequent redundancy between paralogues<sup>40,41</sup>. Notably, we detected dose-dependent masking interactions acting simultaneously between paralogue pairs, whereby mutations in one pair systematically shrink the effects of mutations in the other pair.

Pan-genome sequencing within and across taxa provides a rich resource for understanding how genetic networks are wired. Here, genetic variation in *S. habrochaites* and *S. pennellii* suggested the PLTs as candidate regulators of inflorescence morphology in tomato. However, pan-genome sequencing has also revealed extensive and diverse forms of variation of which the molecular, developmental and evolutionary significance remains unclear. Widespread structural variation in *cis*-regulatory regions and the frequent duplication and loss of both small and large genomic regions drive quantitative expression variance and gene dosage<sup>21,42,43</sup>. The dynamic emergence, divergence and turnover of paralogous genes can alter the architectures and buffering of regulatory networks, therefore perturbing component dosage and potentiating phenotypic change from canalized states when cryptic variants in paralogues converge.

Our detailed dissection of genetic architecture revealed how the *PLT–SEP* network can shift from a canalized state to one poised to release both subtle and substantial phenotypic change. We inferred a hierarchical structure of genetic interactions: classical synergistic interactions within paralogue pairs combine through a multilinear interaction between the two paralogue pairs and then are transformed once

more by an exponential mapping to determine average branches per inflorescence. While our phenomenological modelling is by its nature insufficient to reveal the precise molecular and cellular mechanisms underlying these interactions, we note that each step in our concatenation of simple models is mechanistically plausible, whereby in particular the partial redundancy within paralogue pairs feeds into a regulatory network that requires both PLT paralogues and SEP genes for proper functioning, and gene action in the context of growing cell populations within a developing meristem provides a plausible basis for the overall tendency of mutations to act multiplicatively. Our finding of a multilinear interaction between the two paralogue pairs is notable because, although the multilinear model has been explored theoretically  $^{36,44,45}$ and is often used to estimate how directional selection changes additive genetic variance<sup>46,47</sup>, it has received limited empirical support for capturing real-world patterns of epistasis<sup>46,48</sup>. Nonetheless, the multilinear model can be viewed as a continuous relaxation of Boolean (binary on or off) gene regulatory models<sup>49</sup>, accommodating a spectrum of allelic strengths and dosage effects rather than treating gene expression as a binary. We therefore hypothesize that synergistic interactions within gene families, combined through multilinear interactions that reflect the structure and functional logic of gene regulatory networks, are probably a common genetic architecture governing how phenotypic space is simultaneously expanded and constrained.

Notably, the PLT-SEP genetic network has accumulated extensive genetic variation both within tomato and between Solanaceae species. J2 is a relatively recent duplication missing in many Solanaceae<sup>4,28</sup> (Supplementary Fig. 1). Although the PLT paralogues are broadly retained, their redundancy relationships have probably diverged and vary between genotypes and species. By contrast, in the Brassicaceae, while PLT3 and PLT7 orthologues are conserved, J2 and EJ2 MADS-box paralogues are absent. Notably, Arabidopsis plt mutants do not have altered inflorescence architecture and branching is constrained across species in the family<sup>50</sup>. The architecture of PLT-SEP regulatory networks in the Solanaceae may allow cryptic variation to accumulate more readily than in the Brassicaceae in which the J2/EJ2 subclade is missing, thereby enhancing evolvability of Solanaceae branching. Testing this hypothesis will require broader identification of causal variation across species in these families. Critically, as we show (Figs. 1 and 3), this approach can test specific hypotheses while revealing additional components of conserved and diverged gene regulatory networks. More broadly, the principles identified here-whereby varying paralogue redundancy relationships and presence-absence variation shape trait evolvability through genetic interactions that follow the multilinear model-probably extend across regulatory networks underlying other developmental and physiological programs, influencing the evolutionary trajectories of many traits.

Finally, a key aspect of our study was engineered genetic variation that densely sampled genotypic and phenotypic space in a controlled isogenic background. This approach provided the resolution needed to define the character and quantitative form of gene action and epistasis, represented as a surface illustrating how phenotypic effects from mutations within a paralogue family combine when incorporating mutations across gene families (Fig. 4g). Detailed mapping of genetic interactions in this way could help to reconcile the observation of widespread epistasis in model organisms with the challenge of detecting epistatic effects from allelic variation in natural populations<sup>15-18,51,52</sup>. Placing natural alleles and their combinations onto similar surfaces could reveal how interactions among standing variants push populations into regions of genotypic space in which phenotypic variation is either amplified, suppressed or both. Beyond evolutionary insights, this framework has practical implications in crop engineering<sup>53,54</sup>. Understanding how distinct genetic combinations, along with the specific forms of epistasis they engender, can yield similar phenotypic outcomes may inform targeted editing strategies to predictably 'tune' epistatic interactions. By shifting populations or individuals to advantageous positions on the

genotype-phenotype surface, such strategies could minimize undesirable pleiotropic effects and circumvent genetic constraints imposed by natural alleles in both complex breeding populations and elite genotypes. Realizing these opportunities will hinge on future research that encompasses larger networks of interacting genes, emphasizing how taxon-specific complements of paralogous genes and their variants shape network architecture-and hierarchical epistasis-across broader evolutionary clades.

### **Online content**

Any methods, additional references, Nature Portfolio reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41586-025-09243-0.

- 1. Paaby, A. B. & Rockman, M. V. Cryptic genetic variation: evolution's hidden substrate. *Nat. Rev. Genet.* **15**, 247–258 (2014).
- Gibson, G. & Dworkin, I. Uncovering cryptic genetic variation. Nat. Rev. Genet. 5, 681–690 (2004).
- Mackay, T. F. C. & Anholt, R. R. H. Pleiotropy, epistasis and the genetic architecture of quantitative traits. *Nat. Rev. Genet.* 25, 639–657 (2024).
- Benoit, M. et al. Solanum pan-genetics reveals paralogues as contingencies in crop engineering. Nature 640, 135–145 (2025).
- Alonge, M. et al. Major impacts of widespread structural variation on gene expression and crop improvement in tomato. Cell 182, 145–161 (2020).
- Zhou, Y. et al. Graph pangenome captures missing heritability and empowers tomato breeding. *Nature* 606, 527–534 (2022).
- Waddington, C. H. Canalization of development and genetic assimilation of acquired characters. Nature 183, 1654–1655 (1959).
- Rockman, M. V. The QTN program and the alleles that matter for evolution: all that's gold does not glitter. Evolution 66, 1–17 (2012).
- 9. Lee, Y. W., Gould, B. A. & Stinchcombe, J. R. Identifying the genes underlying quantitative traits: a rationale for the QTN programme. *AoB Plants* **6**, plu004 (2014).
- Raju, A., Xue, B. & Leibler, S. A theoretical perspective on Waddington's genetic assimilation experiments. Proc. Natl Acad. Sci. USA 120, e2309760120 (2023).
- 11. Frank, S. A. Robustness and complexity. Cell Syst. 14, 1015–1020 (2023).
- Schneider, R. F. & Meyer, A. How plasticity, genetic assimilation and cryptic genetic variation may contribute to adaptive radiations. *Mol. Ecol.* 26, 330–350 (2017).
- Ehrenreich, I. M. & Pfennig, D. W. Genetic assimilation: a review of its potential proximate causes and evolutionary consequences. Ann. Bot. 117, 769–779 (2016).
- Sangster, T. A. et al. HSP90 affects the expression of genetic variation and developmental stability in quantitative traits. Proc. Natl Acad. Sci. USA 105, 2963–2968 (2008).
- Nadeau, J. H. Modifier genes in mice and humans. Nat. Rev. Genet. 2, 165–174 (2001).
- Carlborg, O. & Haley, C. S. Epistasis: too often neglected in complex trait studies? Nat. Rev. Genet. 5, 618–625 (2004).
- Mackay, T. F. C. Epistasis and quantitative traits: Using model organisms to study genegene interactions. *Nat. Rev. Genet.* 15, 22–33 (2014).
- Sackton, T. B. & Hartl, D. L. Genotypic context and epistasis in individuals and populations. Cell 166, 279–287 (2016).
- Wolter, F. & Puchta, H. Application of CRISPR/Cas to understand cis- and trans-regulatory elements in plants. Methods Mol. Biol. 1830, 23–40 (2018).
- Liang, Z. & Schnable, J. C. Functional divergence between subgenomes and gene pairs after whole genome duplications. *Mol. Plant* 11, 388–397 (2018).
- Birchler, J. A. & Yang, H. The multiple fates of gene duplications: deletion, hypofunctionalization, subfunctionalization, neofunctionalization, dosage balance constraints, and neutral variation. *Plant Cell* **34**, 2466–2474 (2022).
- Mao, Y. et al. Structurally divergent and recurrently mutated regions of primate genomes. Cell 187, 1547–1562 (2024).
- Jeong, H. et al. Structural polymorphism and diversity of human segmental duplications. Nat. Genet. 57, 390–401 (2025).
- Lemmon, Z. H. et al. The evolution of inflorescence diversity in the nightshades and heterochrony during meristem maturation. *Genome Res* 26, 1676–1686 (2016).
- Hilgenhof, R. et al. Morphological trait evolution in Solanum (Solanaceae): evolutionary lability of key taxonomic characters. TAXON 72, 811–847 (2023).
- Périlleux, C. & Huerga-Fernández, S. Reflections on the triptych of meristems that build flowering branches in tomato. Front. Plant Sci. 13, 798502 (2022).
- Soyk, S. et al. Bypassing negative epistasis on yield in tomato imposed by a domestication gene. Cell 169, 1142–1155 (2017).

- Morel, P. et al. Divergent functional diversification patterns in the SEP/AGL6/AP1 MADS-Box transcription factor superclade. *Plant Cell* **31**, 3033–3056 (2019).
- Hendelman, A. et al. Conserved pleiotropy of an ancient plant homeobox gene uncovered by cis-regulatory dissection. Cell 184, 1724–1739 (2021).
- Walton, R. T., Christie, K. A., Whittaker, M. N. & Kleinstiver, B. P. Unconstrained genome targeting with near-PAMless engineered CRISPR-Cas9 variants. Science 368, 290–296 (2020).
- Ren, Q. et al. PAM-less plant genome editing using a CRISPR-SpRY toolbox. Nat. Plants 7, 25–33 (2021).
- Yamaguchi, N., Jeong, C. W., Nole-Wilson, S., Krizek, B. A. & Wagner, D. AINTEGUMENTA and AINTEGUMENTA-LIKE6/PLETHORA3 induce LEAFY expression in response to auxin to promote the onset of flower formation in *Arabidopsis*. *Plant Physiol*. **170**, 283–293 (2016).
- Hu, G. et al. The auxin-responsive transcription factor SIDOF9 regulates inflorescence and flower development in tomato. *Nat. Plants* 8, 419–433 (2022).
- 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).
- Horstman, A., Willemsen, V., Boutilier, K. & Heidstra, R. AINTEGUMENTA-LIKE proteins: hubs in a plethora of networks. *Trends Plant Sci.* 19, 146–157 (2014).
- Hansen, T. F. & Wagner, G. P. Modeling genetic architecture: a multilinear theory of gene interaction. *Theor. Popul. Biol.* 59, 61–86 (2001).
- Akaike, H. A new look at the statistical model identification. *IEEE Trans. Automat. Contr.* 19, 716–723 (1974).
- Burnham, K. P. & Anderson, D. R. Model Selection and Multimodel Inference (Springer, 2010).
- Domingo, J., Baeza-Centurion, P. & Lehner, B. The causes and consequences of genetic interactions (epistasis). Annu. Rev. Genomics Hum. Genet. 20, 433–460 (2019).
- Diss, G., Ascencio, D., DeLuna, A. & Landry, C. R. Molecular mechanisms of paralogous compensation and the robustness of cellular networks. *J. Exp. Zool. B* 322, 488–499 (2014).
- Iohannes, S. D. & Jackson, D. Tackling redundancy: genetic mechanisms underlying paralog compensation in plants. N. Phytol. 240, 1381–1389 (2023).
- Computational Pan-Genomics Consortium. Computational pan-genomics: status, promises and challenges. Brief. Bioinform. 19, 118–135 (2018).
- Birchler, J. A. & Veitia, R. A. One hundred years of gene balance: How stoichiometric issues affect gene expression, genome evolution, and quantitative traits. Cytogenet. Genome Res. 161, 529–550 (2021).
- Carter, A. J. R., Hermisson, J. & Hansen, T.F. The role of epistatic gene interactions in the response to selection and the evolution of evolvability. *Theor. Popul. Biol.* 68, 179–196 (2005).
- Le Rouzic, A., Álvarez-Castro, J. M. & Hansen, T. F. The evolution of canalization and evolvability in stable and fluctuating environments. *Evol. Biol.* 40, 317–340 (2013).
- 46. Le Rouzic, A. Estimating directional epistasis. Front. Genet. 5, 198 (2014).
- 47. Bourg, S., Bolstad, G. H., Griffin, D. V., Pélabon, C. & Hansen, T. F. Directional epistasis is common in morphological divergence. *Evolution* **78**, 934–950 (2024).
- Pavlicev, M., Le Rouzic, A., Cheverud, J. M., Wagner, G. P. & Hansen, T. F. Directionality of epistasis in a murine intercross population. *Genetics* 185, 1489–1505 (2010).
- Mayer, C. & Hansen, T. F. Evolvability and robustness: a paradox restored. J. Theor. Biol. 430, 78–85 (2017).
- Mudunkothge, J. S. & Krizek, B. A. Three Arabidopsis AIL/PLT genes act in combination to regulate shoot apical meristem function. *Plant J.* 71, 108–121 (2012).
- Omholt, S. W., Plahte, E., Oyehaug, L. & Xiang, K. Gene regulatory networks generating the phenomena of additivity, dominance and epistasis. *Genetics* 155, 969–980 (2000).
- Gjuvsland, A. B., Hayes, B. J., Omholt, S. W. & Carlborg, O. Statistical epistasis is a generic feature of gene regulatory networks. *Genetics* 175, 411–420 (2007).
- Mascher, M., Jayakodi, M., Shim, H. & Stein, N. Promises and challenges of crop translational genomics. *Nature* 636, 585–593 (2024).
- Schreiber, M., Jayakodi, M., Stein, N. & Mascher, M. Plant pangenomes for crop improvement, biodiversity and evolution. *Nat. Rev. Genet.* 25, 563–577 (2024).
- Seabold, S. & Perktold, J. Statsmodels: econometric and statistical modeling with Python. SciPy (2010).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate

credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

© The Author(s) 2025

### Methods

### Motifenrichment and variant discovery

FIMO motif enrichment was performed on the sequence of open chromatin regions in the tomato meristem upstream of *SlEJ2* using the *A. thaliana* non-redundant motif database curated at Plant TFDB (P < 0.00001 and q < 0.01)<sup>56</sup>. The same regions were used to search for insertion–deletion (indel) variants called previously from the tomato pangenome<sup>5</sup>. Indels overlapping with annotated motifs were confirmed to not exist in linkage with previously reported *EJ2* variants (*ej2<sup>w</sup>*, *sb3*) by PCR and then used for subsequent experiments<sup>27,57</sup> (see the 'Plant materials' section below; Supplementary Table 1).

### **Plant materials**

Seeds of WT S. lycopersicum (cultivar M82, LA3475), S. habrochaites (LA1777) and S. pennellii (LA0716) were from our stocks. Introgression line IL3-4 (S. pennellii chromosome 3 introgressed into M82, LA4046) was obtained from the Tomato Genome Resource Center (Department of Plant Sciences, University of California at Davis) and the variant was validated by PCR amplification and Sanger sequencing<sup>58</sup> (a list of all of the primers used in this study is provided in Supplementary Table 27). Two overlapping S. habrochaites chromosome 3 introgression lines LA3925 and LA3926, introgressed into tomato cultivar TA209, were obtained from the Tomato Genome Resource Center (Department of Plant Sciences, University of California at Davis)59. After validation by PCR amplification and Sanger sequencing, only LA3925 contained the ShEJ2pro-3 variant of interest, so LA3926 was used as a control for crosses between M82 and the introgressed region in the TA209 background. Mutants *j2-TE ej2<sup>w</sup>* and *j2 ej2* were from our stocks, as previously described<sup>27</sup>.

#### **Genome editing**

CRISPR-Cas9 mutagenesis and generation of transgenic tomato plants were performed according to our standard protocol<sup>60</sup>. In brief, gRNAs were designed using Geneious Prime (https://www.geneious.com/) (a list of the gRNAs used in this study is provided in Supplementary Table 27). For Cas9 multiplex editing, the Golden Gate cloning system was used to assemble the binary vector containing the Cas9 and the specific gRNAs<sup>60,61</sup>. For SpRY editing, vectors were constructed through a modular Gateway assembly, as described previously (Invitrogen)<sup>62</sup>. The final binary vectors were then transformed into the tomato cultivar M82 by Agrobacterium tumefaciens-mediated transformation through tissue culture<sup>63</sup>. First-generation transgenic plants  $(T_0)$  were genotyped with specific primers surrounding the target sites (a list of all of the primers used in this study is provided in Supplementary Table 27). To purify alleles from potential spontaneous mutations or CRISPR-Cas9 off-target effects after plant transformation, all T<sub>o</sub> transgenic lines were backcrossed (BC1) to parental WT plants. BC1 populations were then screened by PCR and kanamycin herbicide susceptibility for plants lacking the Cas9 transgene, PCR products of the targeted regions were Sanger sequenced to confirm inheritance of alleles and allele-specific genotyping assays were designed for genotyping in subsequent generations. Selected BC1 plants were self-fertilized to generate F<sub>2</sub> populations, and these segregating populations were used to validate the phenotypic effects of each allele by co-segregation. F<sub>2</sub> or F<sub>3</sub> homozygous mutant plants were then used for subsequent crossing and quantitative phenotypic analyses.

### Growth conditions and phenotyping

Seeds were directly sown in soil in 96-cell plastic flats and grown to 4-week-old seedlings in the greenhouse. The seedlings were then transplanted to 4 l pots in the greenhouse for crossing and bulking purposes or directly to the fields at Cold Spring Harbor Laboratory, New York or at The University of Florida Gulf Coast Research and Education Center. Greenhouse conditions are long-day (16 h light, 26–28 °C followed by 8 h dark, 18–20 °C; 40–60% relative humidity) with natural light supplemented with artificial light from high-pressure sodium bulbs (-250  $\mu$ mol m<sup>-2</sup> s<sup>-1</sup>). Plants in the fields were grown under drip irrigation and standard fertilizer regimes, and were used for quantifications of inflorescence branching, fruit shape and sepal length.

To quantify inflorescence branching, inflorescences were counted in order of emergence in two rounds, approximately 60 days after sowing and 75 days after sowing. When available, four primary inflorescences and six axillary inflorescences were counted per plant. 60 or fewer branches were counted, if branching exceeded 60, too many to count (TMTC) was recorded and the number of branching events was treated as 60 for downstream analysis. Proliferated meristem in the place of inflorescence was indicated in the data as proliferated. Occasionally, inflorescences would fail to develop into countable structures, possibly due to stress, in which case, inhibited was recorded.

To quantify fruit shape, ten fruits were collected at the mature green stage, cut in transverse sections and scanned on a single plane. The ratio of maximum height to width, fruit shape index I, was determined from scanned images using Tomato Analyzer<sup>64</sup>. To quantify sepal length, ten closed mature floral buds of similar developmental stage (1–2 days before anthesis, that is, before flower opening) per genotype were collected, length of sepals and petals were manually measured and the sepal/petal ratio was calculated<sup>27</sup>.

#### **Phylogenetic trees**

J2/EJ2 phylogeny was adapted from a previous study<sup>4</sup>. Putative orthologues of SIPLT3/7 and AtPLT3/7 were identified using NCBI BLASTP against proteomes of species selected for taxonomic breadth, representing asterids, rosids, early eudicots and monocots. Retrieved protein sequences were aligned using MAFFT (v.7.505) using the default parameters. An HMM profile was constructed from the alignment using hmmbuild in HMMER (v.3.3.2) and used to search combined species proteomes with hmmsearch ( $E < 1 \times 10^{-5}$ ) to identify additional homologs. All hits were extracted, aligned with MAFFT and manually trimmed when necessary. A maximum-likelihood phylogenetic tree was inferred using IQ-TREE (v.2.2.2) with automatic model selection (-m MFP) and 1,000 ultrafast bootstrap replicates (-bb 1000). The resulting tree was rooted using XP 042461702.1 Zofficinale as an outgroup. Bootstrap support values were used to modulate branch thickness in the visualization: branches with support >90 were plotted thickest, those between 75-90 were medium and those <75 remained thin. The tree was visualized in R using the ape package (v.5.8-1).

### RNA extraction and Illumina sequencing

Inflorescence meristems were collected from n = 4 plants at 8 weeks old under stereoscope magnification. Tissue was frozen, ground with beads and RNA was extracted using TRIzol (Invitrogen) and the Direct-zol RNA Miniprep kit with on-column DNA digestion (Zymo Research). RNA was quantified using the Qubit fluorimeter RNA HS assay kit (Invitrogen). The samples were treated with the Ribo-Zero rRNA removal kit (Epicenter) and the libraries prepared with the TruSeq V2 RNA-seq prep kit (Illumina).

#### **RNA-seq analysis**

Published RNA-sequencing (RNA-seq) data of WT M82, *ej2,j2* and *anantha* mutant meristems were downloaded from Sequence Read Archive (SRA) PRJNA376115 and PRJNA343677 (refs. 24,34). Reads were trimmed with Trimmomatic (ILLUMINACLIP:TruSeq2-PE.fa:2:30:10:1:FALSE LEADING:3 TRAILING:3 SLIDINGWINDOW:4:15 MINLEN:36) and aligned to the cDNA annotation of the reference genome sequence of tomato (SL4.0) using STAR (v.2.6.1.d)<sup>65</sup>. Normalization and quantification of individual transcript expression was done in R by calculating transcripts per million (TPM). Differential expression was calculated in R by DESeq2 time course analysis with LRT and the top 200 most differentially expressed genes ( $log_2$ [FC]) across WT meristem maturation were used for PCA of all meristem samples using Python scikit-learn PCA.transform<sup>66</sup>.

#### **Dual-luciferase assay**

A Gateway-compatible dual-luciferase reporter vector (pSZ106) was assembled using the MoClo GoldenGate assembly system<sup>61,67</sup>. In brief, a Gateway AttR4-AttL1R cassette (Invitrogen) was cloned upstream of a 46 bp minimal 35S promoter driving the Firefly luciferase coding sequence (pICSL80001, pL0\_fLUC-I (CDS1)) with a nopaline synthetase terminator (pICH41421)<sup>61,67</sup>. A cauliflower mosaic virus 35S promoter (pICH51266) was cloned upstream of the coding sequence of Renilla luciferase (pSB123, pL0 rLUC-I (CDS1), Addgene) with a nopaline synthetase terminator (pICH41421)<sup>67</sup>. Both luciferase expression cassettes were cloned into the pICSL4723 binary vector backbone with an NPTII selection cassette. SIEJ2pro-3, ShEJ2pro-3 and SpEJ2pro-3 alleles were cloned into pDONR P4-P1r and introduced into pSZ106 by Gateway cloning (Invitrogen). SIERF12 (Solyc02g077840), SIPLT380 (Solyc05g051380) and SIPLT710-short (Solyc11g010710) were cloned into pDONR207 and introduced into pEAQ-HT-DEST3 by Gateway cloning (Invitrogen).

All binary expression vectors were transformed into *A. tumefaciens* and cultured at 28 °C overnight in selective media. Overnight cultures were diluted and grown at 28 °C to an optical density at 600 nm (OD<sub>600</sub>) of 1 a.u., centrifuged and washed into inductive medium (10 mM MES pH 5.7, 10 mM MgCl<sub>2</sub>, 100  $\mu$ M 3',5'-dimethyoxy-4'-hydroxyacetophenone) at OD<sub>600</sub> 1 a.u. Bacteria was induced for 3 h lying horizontally at the bench, then equal volumes of promoter and TF medium were combined and co-infiltrated into young fully expanded leaves of 4 week old *N. benthamiana* plants grown in long days (16 h–8 h light–dark, 22 °C; 40–60% relative humidity). Plants were returned to the growth chamber and 100 mg tissue was collected and frozen for measurement 3 days after infiltration.

Luciferase activity was measured using the Dual Luciferase Reporter Assay System kit (Promega) as described previously<sup>68</sup>. In brief, tissue was homogenized in the Spex Sample Prep 2010 Geno/Grinder (Cole Parmer) and 10 mg of tissue powder was mixed with 100  $\mu$ l of passive lysis buffer (Promega). Cellular debris was pelleted at 7,500g for 1 min and the supernatant was diluted 40× in passive lysis buffer and 15  $\mu$ l of sample was transferred to three replicate wells of a white flat-bottom Costar 96-well plate (Corning). The assay was measured using a GloMax 96 microplate luminometer, and 75  $\mu$ l per well of luciferase assay reagent and Stop & Glo reagent was added and measured stepwise (Promega).

#### Statistics and reproducibility

All phenotyping and molecular experiments were repeated in at least three seasons with similar results. Transcriptomics was performed once with biological replication. For Fig. 4e, f, n = 10 inflorescences per plant and the number of biologically independent plants per genotype– season combination varies, with quartiles at 2, 5 and 9 plants (source data are provided in Supplementary Table 24).

#### Segregating populations

BC1 inbred plants of the genotype *EJ2pro j2* were crossed to BC1 inbred plants of the genotype *plt3 plt7/+* and genotyped in the  $F_1$  generation by allele-specific PCR to determine the presence of all desired alleles. Segregating  $F_2$  seed was sown in the greenhouse in populations of either 192 or 384 plants, tissue was collected for DNA extraction and plants were transplanted without previous genotyping over the course of four seasons, two in fields at Cold Spring Harbor Laboratory, New York and two at The University of Florida Gulf Coast Research and Education Center. The genotypes of plants were confirmed after phenotyping by allele-specific PCR assays.

#### Linear regression models

Phenotypic data were summarized at the plant level for quantitative modelling, with abnormal inflorescences marked as proliferated or inhibited excluded from further analysis. For each plant *i*, we consider the total number of branching events across all inflorescences  $y_i$  and the number of inflorescences  $t_i$ . The total number of branching events  $y_i$  was modelled as being either Poisson or negative binomially distributed with exposure  $t_i$ .

$$y_i \sim \operatorname{Poisson}(t_i \times f^{-1}(\mu(x_i)))$$

 $y_i$  ~ Negative Binomial  $(t_i \times f^{-1}(\mu(x_i)), \alpha)$ ,

where *f* represents a link function and  $\mu(x_i)$  represents the phenotypic mean of the genotype  $x_i$  of plant *i* and  $\alpha$  is the overdispersion parameter. Var $(y_i) = \mu(x_i) + \alpha \mu(x_i)^2$  so  $\alpha$  reflects the additional variance relative to the expectation under a Poisson model. Under an additive model, the expected mean  $\mu_{add}(x)$  of any given genotype *x* is given by:

$$\mu_{\text{add}}(x_i) = \theta_0 + \sum_l \theta_l^a s_l^a(x_i) + \theta_l^d s_l^d(x_i),$$

where  $s_l^a(x_i) = \{-1, 0, 1\}$  and  $s_l^d(x_i) = \{0, 1, 0\}$  if locus *l* in genotype  $x_i$  is homozygous WT, heterozygous mutant and homozygous mutant, respectively. Thus  $\theta_l^a$  represents the homozygous effects at locus *l*, whereas  $\theta_l^d$  represents the deviation of the heterozygous effect from the semi-dominant expectation.

A basis for pairwise interaction models was built by extending the basis for the additive model with additional basis vectors composed by taking the product of each possible pair of additive and dominance components<sup>69</sup>. Under a pairwise model, the expected mean is given by:

$$\begin{split} \mu_{\rm pw}(x_i) &= \mu_{\rm add}(x_i) + \sum_{l,m} \theta_{l,m}^{a,a} s_l^a(x_i) s_m^a(x_i) + \theta_{l,m}^{a,d} s_l^a(x_i) s_m^d(x_i) \\ &+ \theta_{l,m}^{d,a} s_l^d(x_i) s_m^a(x_i) + \theta_{l,m}^{d,d} s_l^d(x_i) s_m^d(x_i). \end{split}$$

Models were defined and fit using the statsmodels<sup>55</sup> Python package using Poisson and negative binomial likelihoods with the identity (f(x) = x) and log link functions  $(f(x) = \log x)$ . Genotype–season MLEs for the number of branching events were obtained by defining a dummy variable for each genotype–season combination that took a value of 1 for plants of that genotype and 0 otherwise, while assuming that all genotypes share the overdispersion parameter for the negative binomial likelihood function representing plant-to-plant variability that is jointly estimated with the genotype–season means. Confidence intervals for the MLEs were derived using statsmodels with a log-link between model parameters (genotype estimates) and the average number of branching events.

#### **Hierarchical model**

Data were modelled with a negative binomial likelihood function as explained in the 'Linear regression models' section. In this hierarchical model, each paralogue pair has an effect that is modelled separately through a complete pairwise interaction model into  $\varphi_{\text{PLT}}$  and  $\varphi_{\text{SEP}}$ , parametrized by the phenotypic effect between any genotype g (combination of WT, heterozygous or homozygous mutants) and the WT  $\theta_{g}^{\text{PLT}}$  and  $\theta_{g}^{\text{SEP}}$  at the PLTs or SEP pair of loci, respectively:

$$\begin{split} \varphi_{\mathsf{PLT}}(x_i) &= \sum_{g \neq \mathsf{WT}} \, \theta_g^{\mathsf{PLT}} s_g^{\mathsf{PLT}}(x_i), \\ \varphi_{\mathsf{SEP}}(x_i) &= \sum_{g \neq \mathsf{WT}} \, \theta_g^{\mathsf{SEP}} s_g^{\mathsf{SEP}}(x_i), \end{split}$$

where  $s_g^{PLT}(x_i)$  and  $s_g^{PLT}(x_i)$  take value 1 if the genotype at the PLT or SEP loci match g and 0 otherwise. Note that  $\varphi_{PLT}$  takes a different value for every possible combination of mutations in *PLT3* and *PLT7* and  $\varphi_{SEP}$ takes a different value for every possible combination of mutations in *EJ2* and *J2*, so that inferring the values for  $\varphi_{PLT}$  and  $\varphi_{SEP}$  is equivalent to allowing a full set of additive, dominance and pairwise interactions within each of *PLT3/PLT7* and *EJ2/J2*. These two pairwise models are then combined through a multilinear function into the log-transformed average expected number of branching events  $\mu(x_i)$ , given by

 $\mu_{\text{hierarchical}}(x_i) = \theta_{\text{WT}} + \varphi_{\text{PLT}}(x_i) + \varphi_{\text{SEP}}(x_i) - \theta_{\text{Int}}\varphi_{\text{PLT}}(x_i)\varphi_{\text{SEP}}(x_i),$ 

where  $\theta_{WT}$  is the WT log-transformed expected branching events,  $\varphi_{PLT}$  controls the log effect of the relevant combination of mutations in *PLT3* and *PLT7* when placed in a WT *EJ2J2* background,  $\varphi_{SEP}$  controls the log effect of the relevant combination of mutations in *EJ2* and *J2* combinations in a WT *PLT3 PLT7* background and  $\theta_{Int}$  represents the masking interaction between the two phenotypes<sup>36</sup>. Finally, as in the standard linear models from the previous section, the observed number of branching events  $y_i$  for a plant *i* with genotype  $x_i$  and  $t_i$  inflorescences is drawn from a negative binomial distribution with overdispersion parameter  $\alpha$ :

 $y_i \sim \text{Negative Binomial}(t_i \times e^{\mu_{\text{hierarchical}}(x_i)}, \alpha).$ 

In summary, the hierarchical model is equivalent to fitting pairwise interactions within paralogue pairs, then combining the within pair effects through a multilinear interaction across pairs, and then transforming the result through an exponential function. Extended Data Fig. 4a shows a graphical representation of the complete model. This model was coded in PyTorch<sup>70</sup> and the maximum-likelihood solution was found running the Adam optimizer for 10,000 iterations and checking for convergence. Extended Data Fig. 4c shows the inferred model including all the  $EJ2^{pro}$  alleles and illustrates how the different layers of the hierarchical model are applied and combined together to predict the expected number of branching events for a given genotype.

#### **Reporting summary**

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

### Data availability

All data supporting the findings of this study are available within the Article and its Supplementary Information. RNA-seq data generated in this study are available at the Gene Expression Omnibus (https:// www.ncbi.nlm.nih.gov/geo/) under accession GSE289537, and previously published RNA-seq data are available at the SRA under Bio-Projects PRJNA376115 and PRJNA343677. The tomato pangenome is available at the SRA under BioProject PRJNA557253. The *A. thaliana* non-redundant motif database used is available online (https:// planttfdb.gao-lab.org/index.php?sp=Ath). The raw data with the number of branching events for each plant and inflorescence are provided in Supplementary Tables 3, 5, 7, 13, 15, 17, 21 and 24. All unique biological materials used in this Article are available for distribution on request.

### **Code availability**

Code to reproduce the statistical analysis, quantitative modelling and the derived figures is available at GitHub and Zenodo<sup>71</sup> (https://github. com/cmarti/tomato\_branching/tree/master, https://doi.org/10.5281/ zenodo.15552523).

- Jin, J. et al. PlantTFDB 4.0: Toward a central hub for transcription factors and regulatory interactions in plants. *Nucleic Acids Res.* 45, D1040–D1045 (2017).
- Soyk, S. et al. Duplication of a domestication locus neutralized a cryptic variant that caused a breeding barrier in tomato. *Nat. Plants* 5, 471–479 (2019).
- Alseekh, S. et al. Resolution by recombination: breaking up Solanum pennellii introgressions. Trends Plant Sci. 18, 536–538 (2013).
- Monforte, A. J. & Tanksley, S. D. Development of a set of near isogenic and backcross recombinant inbred lines containing most of the *Lycopersicon hirsutum* genome in a *L. esculentum* genetic background: a tool for gene mapping and gene discovery. *Genome* 43, 803–813 (2000).
- 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 (2017).
- Engler, C. et al. A Golden Gate modular cloning toolbox for plants. ACS Synth. Biol. 3, 839–843 (2014).
- Lowder, L. G. et al. A CRISPR/Cas9 toolbox for multiplexed plant genome editing and transcriptional regulation. *Plant Physiol.* 169, 971–85 (2015).
- Van Eck, J., Kirk, D. D. & Walmsley, A. M. Tomato (Lycopersicum esculentum). Methods Mol. Biol. 343, 459–474 (2006).
- Brewer, M. T. et al. Development of a controlled vocabulary and software application to analyze fruit shape variation in tomato and other plant species. *Plant Physiol.* 141, 15–25 (2006).
- 65. Dobin, A. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics*. **29**, 15–21 (2013).
- Pedregosa, F. et al. Scikit-learn: machine learning in Python. J. Mach. Learn. Res. 12, 2825–2830 (2011).
- 67. Mortensen, S. et al. EASI transformation: an efficient transient expression method for analyzing gene function in seedlings. *Front. Plant Sci.* **10**, 755 (2019).
- Moyle, R. L. et al. An optimized transient dual luciferase assay for quantifying microRNA directed repression of targeted sequences. Front. Plant Sci. 8, 1631 (2017).
- Tiwari, H. K. & Elston, R. C. Deriving components of genetic variance for multilocus models. *Genet. Epidemiol.* 14, 1131–1136 (1997).
- Paszke, A. et al. PyTorch: an imperative style, high-performance deep learning library. in Advances in Neural Information Processing Systems 32 (NeurIPS 2019) (eds Wallach H. et al.) 8026–8037 (Curran Associates, 2019).
- Zebel, S. G. et al. Code for 'Cryptic variation fuels plant phenotypic change through hierarchical epistasis'. Zenodo https://doi.org/10.5281/zenodo.15552523 (2025).

Acknowledgements We thank the members of the Lippman laboratory for their support and feedback; S. Hutton for support with field growing support; T. Mulligan, K. Schlecht and S. Qiao for assistance with plant care; J. P. Alvarez for assistance with imaging inflorescences of *Solanum* species; and A. Le Rouzic for conversations. S.G.Z. is supported by the National Institutes of Health grant K99GM149939; Y.E. by the Israel Science Foundation grant 2390/23; C.M.-G. and D.M.M. by NIH grant R35 GM133613 and additional funding from the Simons Center for Quantitative Biology at Cold Spring Harbor Laboratory; Z.B.L. and M.B. by The National Science Foundation Plant Genome Research Program grant IOS-2129189; M.B. is supported by a Gatsby Charitable Foundation grant (GAT3731/GLK); and Z.B.L. is supported by The National Science Foundation Plant Genome Research Program grant IOS-2126612 and the Howard Hughes Medical Institute.

Author contributions Conceptualization: S.G.Z., D.M.M. and Z.B.L. Formal analysis: S.G.Z. and C.M.-G. Funding acquisition: S.G.Z., M.B., Y.Q., Y.E., D.M.M. and Z.B.L. Investigation: S.G.Z., C.M.-G., B.F., C.P.C., M.L., B.M.S., A.H. and S.S. Methodology: S.G.Z., C.M.-G., D.M.M. and Z.B.L. Project administration: S.G.Z., C.M.-G., Y.E., D.M.M. and Z.B.L. Writing—original draft: S.G.Z., C.M.-G., Y.E., D.M.M. and Z.B.L. Writing—review and editing: S.G.Z., C.M.-G., B.F., M.B., Y.E., D.M.M. and Z.B.L. Writing—the editing: S.G.Z., C.M.-G., B.F., M.B., Y.E., D.M.M. and Z.B.L. Writing—the editing: S.G.Z., C.M.-G., B.F., M.B., Y.E., D.M.M. and Z.B.L. Writing—the editing: S.G.Z., C.M.-G., B.F., M.B., Y.E., D.M.M. and Z.B.L. Writing—the editing: S.G.Z., C.M.-G., B.F., M.B., Y.E., D.M.M. and Z.B.L. Writing—the editing: S.G.Z., C.M.-G., B.F., M.B., Y.E., D.M.M. and Z.B.L. Writing—the editing: S.G.Z., C.M.-G., B.F., M.B., Y.E., D.M.M. and Z.B.L. Writing—the editing: S.G.Z., C.M.-G., B.F., M.B., Y.E., D.M.M. and Z.B.L. Writing—the editing: S.G.Z., C.M.-G., B.F., M.B., Y.E., D.M.M. and Z.B.L. Writing—the editing: S.G.Z., C.M.-G., B.F., M.B., Y.E., D.M.M. and Z.B.L. Writing—the edition of the editing: S.G.Z., C.M.-G., S.F., M.B., Y.E., D.M.M. and Z.B.L.

**Competing interests** Z.B.L. is a consultant for and a member of the scientific strategy board of Inari Agriculture. The other authors declare no competing interests.

#### Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41586-025-09243-0.

**Correspondence and requests for materials** should be addressed to Yuval Eshed, David M. McCandlish or Zachary B. Lippman.

Peer review information Nature thanks Mark Cooper, Esther van der Knaap, Mohammad Siddiq and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Reprints and permissions information is available at http://www.nature.com/reprints.



**Extended Data Fig. 1 Phylogenetic tree of PLT3/PLT7.** Phylogenetic tree of PLT3 and PLT7 proteins from species selected for taxonomic breadth, representing asterids, rosids, early eudicots, and monocots, see Methods

for details. Branch thickness reflects bootstrap support: thin lines indicate support <75, medium lines 75–90, and thick lines >90. Scale bar represents number of amino acid substitutions per site.



**Extended Data Fig. 2** | **Mutations in one paralog pair linearly re-scale the effects of mutations in the other paralog pair.** Scatterplots representing the expected number of branching events. Every dot represents a possible combination of mutations in *PLT3* and *PLT7* (first 6 columns) or *EJ2* and *J2* (last column) in a specific season. The value plotted on the y-axis corresponds to the phenotype conferred by the given combinations of mutations either in the designated background at *EJ2* and *J2* (first 6 columns) or the designated background at *PLT3* and *PLT7* (last column); x-axis values are given by the phenotype of each set of mutations in the wild-type background. Dots represent the Maximum Likelihood Estimate (MLE) of the expected number of branching events for genotypes in specific seasons, as estimated under a negative binomial model. Error bars represent the 95% confidence interval for the MLEs. Total least squares regression lines for the log MLEs are represented with black dashed lines. Genotype-season combinations with a 95% confidence interval wider than a thousand-fold range are not shown.





conferred by the given combinations of mutations either in the designated background at *EJ2* and *J2* (first 6 columns) or the designated background at *PLT3* and *PLT7* (last column); x-axis values are given by the sample average of each set of mutations in the wild-type background. Total least squares regression lines for the log sample means are represented with black dashed lines. Genotype-season combinations in which no branching was observed are not shown.



Extended Data Fig. 4 | See next page for caption.

Extended Data Fig. 4 | Understanding synergy and masking under the

inferred hierarchical model. (a) Graphical representation of the hierarchical epistasis model. Equations on the right show the mathematical transformations applied at each layer. See Methods for details and an explanation of the mathematical notation. (b) Different views of the two-dimensional surface representing the expected number of branching events as a function of the mean branching events conferred by placing the combination of mutations in each paralog pair in the wild-type background of the other pair. Points represent the independent maximum likelihood estimates (MLEs) for the expected number of branching events for the measured genotype-season combinations. Error bars represent the 95% confidence interval for the MLEs. n = 10 inflorescences per plant; see Methods for details. (c) Complete representation of the inferred hierarchical model of gene interaction. Left and bottom panels represent the estimated number of branching events for each paralog pair combination in the wild-type background of the other pair as a function of the Hamming distance to the wild-type genotype. Different EJ2pro alleles show phenotypic effects of different magnitude but all interact with J2 in a similar way. The phenotype for any genotype is obtained by combining the independent effect of the mutations in each pair of paralogs through the twodimensional surface in the middle panel. This heatmap represents the inferred multilinear function that quantitatively characterizes epistatic masking between PLT and MADS genes. White dashed lines represent isophenotypic lines, this is, combinations of background phenotypic effects that result in the same number of branching events when combined. Note that the function is linear across any horizontal or vertical transect, as illustrated by the constant distance between isophenotypic lines across any transect. Deviation from a slope of -1 in the shape of the isophenotypic lines, which are scale-independent, indicates the presence of an epistatic interaction between the two pairs of paralogs. Note that distance between isophenotypic lines increases in highly mutated backgrounds, indicating that a larger perturbation is required to achieve the same phenotypic outcome. (d) Estimated number of branching events for all  $E/2^{pro8}/2$  combinations across genetic backgrounds with an increasing number of mutations in PLT3 and PLT7 illustrates how the synergistic interactions between EJ2pro8 and j2 become masked as PLT3 and PLT7 become increasingly mutated.

## nature portfolio

Corresponding author(s):

Zachary B. Lippman, David M. McCandlish, Yuval Eshed

Last updated by author(s): Jun 1, 2025

### **Reporting Summary**

Nature Portfolio 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 Portfolio policies, see our Editorial Policies and the Editorial Policy Checklist.

### **Statistics**

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Cor	firmed
	$\square$	The exact sample size ( $n$ ) for each experimental group/condition, given as a discrete number and unit of measurement
$\boxtimes$		A statement on 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.
	$\boxtimes$	A description of all covariates tested
	$\square$	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
	$\boxtimes$	A full description of the statistical parameters 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)
	$\boxtimes$	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.
$\boxtimes$		For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
	$\square$	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
$\boxtimes$		Estimates of effect sizes (e.g. Cohen's d, Pearson's r), indicating how they were calculated
		Our web collection on statistics for biologists contains articles on many of the points above.

### Software and code

Policy information about availability of computer code

Sequencing data were recorded as fastq files (see Methods), quantitative phenotypic data were collected manually in fields and greenhouses Data collection and recorded in Microsoft Excel version 16.94, Promega Glomax Microplate Reader (default parameters), BioRad Image Lab (v6.1, default parameters) (see Methods). Data analysis Geneious Prime (v2025.0.2), BioRad ImageLab (v6.1, default parameters), MAFFT (v7.505), HMMER (v3.3.2), IQ-TREE (v2.2.2), FastQC (v0.11.9), trimmomatic (v0.39), STAR (v 2.6.1), MEME (v4.11.2). R packages: tidyverse (v2.0.0), ggplot2 (v3.4.3), ggpubr (v0.6.0), plyr (v1.8.9), reshape2 (v1.4.4), dplyr (v1.1.3), DESeq2 (v1.38.), ape (v5.8-1), DescTools(0.99.54). Python packages: Figures 1-3 and associated Supplementary Data: matplotlib (v3.7.0), pandas (v1.5.3), plotly (v5.9.0), scikit-learn (v1.2.1), scipy (v1.10.0), sklearn (v0.0.post5). Figure 4 and associated Supplementary Data: matplotlib (v3.7.5), numpy (v1.24.4), pandas (v2.0.3), scipy (v1.10.1), seaborn (v0.11.1), statsmodels (v0.14.1), torch (v2.4.1).

Code availability: https://github.com/cmarti/tomato\_branching/tree/master, DOI: 10.5281/zenodo.15552524.

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 and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio 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 description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our policy

All data are available within this Article and its Supplementary Information. RNA sequencing data generated in this study is available at the Gene Expression Omnibus (https://www.ncbi.nlm.nih.gov/geo/) under accession GSE289537, and previously published RNA sequencing is available at Sequence Read Archive (SRA, https://ncbi.nlm.nih.gov/sra) under BioProjects PRJNA376115 and PRJNA343677. The tomato pangenome is available at SRA under BioProject PRJNA557253. The Arabidopsis thaliana non-redundant motif database used is available at https://planttfdb.gao-lab.org/index.php?sp=Ath. The raw data with the number of branching events for each plant and inflorescence is provided in Supplementary Tables. All unique biological materials used in this manuscript are available for distribution upon request.

### Research involving human participants, their data, or biological material

Policy information about studies with <u>human participants or human data</u>. See also policy information about <u>sex, gender (identity/presentation)</u>, <u>and sexual orientation</u> and <u>race, ethnicity and racism</u>.

Reporting on sex and gender	NA
Reporting on race, ethnicity, or other socially relevant groupings	NA
Population characteristics	NA
Recruitment	NA
Ethics oversight	NA

Note that full information on the approval of the study protocol must also be provided in the manuscript.

### Field-specific reporting

Please select the one below that is 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/documents/nr-reporting-summary-flat.pdf

### Life sciences study design

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

Sample size	Sample size for RNA-Seq was based on four independent samples, as done previously (Alonge, Wang et al., 2020). No statistic methods were used to predetermine sample size for quantitative phenotypic analyses. Required experimental sample size was estimated based on our past experience performing similar experiments including greenhouse and field tests (see for example Kwon et al., 2022, Alonge, Wang et al., 2020, Rodríguez-Leal et al., 2017).
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 relevant information is presented in figure legends, methods, and supplementary data files. Individual replicates (e.g. tissue samples, plants, shoots, flowers and fruits) are indicated and at least four independent replicates were analyzed for each experiment. Raw phenotypic data are provided in supplementary data files.
Randomization	For the segregating population experiments, F2 population plants were randomly sown and transplanted in an agricultural field and genotyping was completed after phenotypic analysis. These randomized and blinded experiments formed the basis of the model.
Blinding	For the segregating population experiments, F2 population plants were randomly sown and transplanted in an agricultural field and genotyping was completed after phenotypic analysis. These randomized and blinded experiments formed the basis of the model.

### Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems		Methods	
n/a	Involved in the study	n/a	Involved in the study
$\boxtimes$	Antibodies	$\boxtimes$	ChIP-seq
$\boxtimes$	Eukaryotic cell lines	$\boxtimes$	Flow cytometry
$\boxtimes$	Palaeontology and archaeology	$\boxtimes$	MRI-based neuroimaging
$\boxtimes$	Animals and other organisms		
$\boxtimes$	Clinical data		
$\boxtimes$	Dual use research of concern		
	Plants		

### Dual use research of concern

Policy information about <u>dual use research of concern</u>

### Hazards

Could the accidental, deliberate or reckless misuse of agents or technologies generated in the work, or the application of information presented in the manuscript, pose a threat to:

No	Yes
$\ge$	Public health
$\ge$	National security
$\ge$	Crops and/or livestock
$\boxtimes$	Ecosystems
$\boxtimes$	Any other significant area

### Experiments of concern

Does the work involve any of these experiments of concern:

No	Yes
$\boxtimes$	Demonstrate how to render a vaccine ineffective
$\boxtimes$	Confer resistance to therapeutically useful antibiotics or antiviral agents
$\boxtimes$	Enhance the virulence of a pathogen or render a nonpathogen virulent
$\boxtimes$	Increase transmissibility of a pathogen
$\boxtimes$	Alter the host range of a pathogen
$\boxtimes$	Enable evasion of diagnostic/detection modalities
$\boxtimes$	Enable the weaponization of a biological agent or toxin
$\boxtimes$	Any other potentially harmful combination of experiments and agents

### Plants

Seed stocks Novel plant genotypes	Solanum lycopersicum (cultivar M82, LA3475), Solanum habrochaites (LA1777), and Solanum pennellii (LA0716) were from our stocks. Introgression line IL3-4 (Solanum pennellii chromosome 3 introgressed into M82, LA4046) and two overlapping Solanum habrochaites chromosome 3 introgressed into tomato cultivar TA209, were obtained from the verband structures (were achieved interversed into tomato cultivar TA209, were obtained from the verband structures (were achieved interversed into tomato cultivar TA209, were obtained from the verband structures (were achieved interversed into tomato cultivar TA209, were obtained from the verband structures (were achieved interversed into tomato cultivar TA209, were achieved into the structure of the verband structures (were achieved interversed into tomato cultivar TA209, were achieved into the structure of the verband structures (were achieved interversed into tomato cultivar TA209, were achieved into the structure of the structure of the verband structure of the
Authentication	All variants were validated by PCR amplification and Sanger sequencing.