Leaf form diversification in an ornamental heirloom tomato results from alterations in two different HOMEOBOX genes

Hokuto Nakayama
Department of Plant Biology, University of California, Davis, One Shields Avenue, Davis, CA 95616, USA.

Steven D. Rowland
Department of Plant Biology, University of California, Davis, One Shields Avenue, Davis, CA 95616, USA.

See next page for additional authors

Let us know how access to this document benefits you

Copyright ©2021 Nakayama et al

This work is licensed under a Creative Commons Attribution-NonCommercial-No Derivative Works 4.0 International License.

Follow this and additional works at: https://scholarworks.uni.edu/facpub

Part of the Plant Sciences Commons

Recommended Citation
Nakayama, Hokuto; Rowland, Steven D.; Cheng, Zizhang; Zumstein, Kristina; Kang, Julie; Kondo, Yohei; and Sinha, Neelima R., "Leaf form diversification in an ornamental heirloom tomato results from alterations in two different HOMEOBOX genes" (2021). Faculty Publications. 4.
https://scholarworks.uni.edu/facpub/4

This Article is brought to you for free and open access by UNI ScholarWorks. It has been accepted for inclusion in Faculty Publications by an authorized administrator of UNI ScholarWorks. For more information, please contact scholarworks@uni.edu.
Leaf form diversification in an ornamental heirloom tomato results from alterations in two different HOMEBOX genes

**Graphical abstract**

- **Highlights**
  - Two HOMEBOX genes are responsible for the leaf shape in an heirloom tomato, SiFT
  - BIP regulates leaf complexity; SIWOX1 regulates leaflet width and vascular density
  - SIWOX1 is mutated in the classical tomato mutant, solanifolia
  - The bip mutation in SiFT arose de novo during the breeding process

**Authors**
Hokuto Nakayama, Steven D. Rowland, Zizhang Cheng, Kristina Zumstein, Julie Kang, Yohei Kondo, Neelima R. Sinha

**Correspondence**
nrsinha@ucdavis.edu

**In brief**
Nakayama et al. identify two HOMEBOX genes responsible for leaf form diversification in an ornamental heirloom tomato. Those genes regulate leaf complexity, leaflet width, and vascular density. New morphological traits can be introduced into a crop during the ongoing improvement processes after domestication.
Leaf form diversification in an ornamental heirloom tomato results from alterations in two different HOMEBOX genes

Hokuto Nakayama,1,4 Steven D. Rowland,1 Zizhang Cheng,1 Kristina Zumstein,1 Julie Kang,2 Yohei Kondo,3 and Neelima R. Sinha1,5,6,*

1Department of Plant Biology, University of California, Davis, One Shields Avenue, Davis, CA 95616, USA
2Biological Department, University of Northern Iowa, 144 McCollum Science Hall, Cedar Falls, IA 50614, USA
3Division of Quantitative Biology, Okazaki Institute for Integrative Bioscience, National Institute for Basic Biology, National Institutes of Natural Sciences, Myodaiji, Higashiyama 5-1, Okazaki, Aichi 444-8787, Japan
4Present address: Graduate School of Science, Department of Biological Sciences, The University of Tokyo, Science Build. No. 2, 7-3-1 Hongo Bunkyo-ku, Tokyo 113-0033, Japan
5Twitter: @PlantDevo
6Lead contact
7Correspondence: nrsinha@ucdavis.edu
https://doi.org/10.1016/j.cub.2021.08.023

SUMMARY

Domesticated plants display diverse phenotypic traits. However, the influence of breeding effort on this phenotypic diversity remains unknown. Here, we demonstrate that a single nucleotide deletion in the homeobox motif of BIPINNATA, a BEL-LIKE HOMEODOMAIN gene, led to a highly complex leaf phenotype in an heirloom tomato (Solanum lycopersicum), Silvery Fir Tree (SiFT), which is used as a landscaping and ornamental plant. A comparative gene network analysis revealed that repression of SOLANIFOLIA, the ortholog of WUSCHEL RELATED HOMEobox 1, caused the narrow leaflet phenotype seen in SiFT. Comparative genomics indicated that the bip mutation in SiFT likely arose de novo and is unique to SiFT and not introgressed from other tomato genomes. These results provide new insights into the natural variation in phenotypic traits introduced into crops during improvement processes after domestication and establish homeobox genes as evolutionary hotspots.

INTRODUCTION

Domestication and subsequent improvement processes have made animals and plants more suitable for agriculture and have improved their usability, quality, and yield.1 In contrast to domestication, which usually occurs only once for crops, selection for improvement occurs multiple times and in numerous locations, leading to varieties adapted to local conditions and needs.2 Consequently, many crops exhibit morphological diversity. Darwin focused on this morphological diversity more than 150 years ago and postulated that knowledge of the mechanisms underlying diversity generated under human selection would provide the general principles for understanding the process of evolution under natural selection.3 The domesticated tomato, Solanum lycopersicum L. (Solanaceae), exhibits tremendous morphological variation because of its long breeding history.4 With varieties passed down through several generations within a family or specific region, many heirloom tomatoes had an interesting breeding history before 1940, when commercial hybrids became available. These heirloom cultivars show morphological variation and flavor profiles that are often favored by gardeners.5 Many heirloom tomatoes contain genetic loci to improve various phenotypes of the modern commercial tomato,6 making them an interesting research topic regarding tomato improvement. The Silvery Fir Tree (SiFT) is a traditional Russian heirloom tomato.5 In addition to a pale leaf color, SiFT has a highly complex leaf phenotype, with narrower leaflets than those seen in processing tomatoes, such as M82 (Figures 1A–1D). SiFT is sometimes used as an ornamental and landscaping plant rather than a crop because of its unique leaf shape, although SiFT produces edible fruit.6 Introduced tomatoes were cultivated first as ornamental plants in Europe in the 16th century, instead of the more commonly used crop plant today.7,8 Knowledge regarding the leaf development of ornamental tomatoes, such as SiFT, provides a unique opportunity for understanding the molecular mechanism underlying leaf-form diversification during crop domestication and the subsequent improvement processes. However, the genetic basis underlying this unique leaf morphology and how it was introduced remains unknown because of a lack of information about heirloom tomatoes.

In the present study, we used a cross between SiFT and M82 to generate a mapping population and identified a single nucleotide deletion in the homeobox motif of the BEL-LIKE HOMEODOMAIN gene, BIPINNATA (BIP), leading to a premature stop codon and a highly complex leaf phenotype in SiFT. The bip mutation affects KNOX1 expression, which is known to be involved in intra- and interspecific leaf-form variations in various plant species. Comparative genomics indicated that
the bip mutation in SiFT is a de novo mutation that was not introgressed from other cultivars or wild species. We used a combination of deep-learning-based leaf shape analysis, gene co-expression network analysis, and CRISPR-Cas9 knockout mutants to show that reduced expression of the WUSCHEL RELATED HOMEOBOX 1 (WOX1) ortholog was the cause of the narrower leaflet phenotype and reduced leaf vascular density (LVD) in SiFT. We demonstrated that solanifolia, a well-studied classic leaf mutation with a previously unknown causative gene, is a wox1 mutant. These results provide insights into the
natural variation in phenotypic traits introduced into heirloom tomatoes during improvement after domestication and the repeated evolution of specific gene modules.

RESULTS

SIFT has a mutation in a BEL-LIKE HOMEODOMAIN gene, BIPINNATA

Leaf complexity (LC) in SIFT was higher than that in M82 (Figures 1A–1C). The difference started with the first-formed leaves (Figure 1D). The leaf primordia from plastochron 1 (P1) to P3 stages were not markedly different between the cultivars; however, from the P4 stage onward, differences in the number of leaflet primordia were consistently observed (Figures 1E and S1A). Thus, the SIFT leaf primordia in P4 and older stages were more active in generating leaflets than M82 leaves at the same developmental stage. Previous studies have shown variations in LVD among cultivars and mutants.\(^{10}\) Although no difference was observed in leaf anatomy around the midvein (Figure S1B), LVD was different between M82 and SIFT (Figure 1F). Therefore, SIFT differed from M82 in terms of LC, developmental trajectory, and vascular density.

To identify the genes involved in the regulation of LC, bulked segregant analysis (BSA) on an F2 population (198 individuals) derived from a cross between M82 and SIFT was performed (Figure S1C). Two phenotypically defined bulks showing a difference in LC (high-LC bulk and low-LC bulk; Figure S1D) were used to detect the most obvious difference between the two bulks, a locus between 45,000,000 and 55,000,000 bp on chromosome 2, which controlled LC (Figures 1G, top, and S1E). Whole-genome sequencing (WGS) of SIFT was used to define the sequence variants in the genome, including the region defined by BSA. We detected more than 100 variants in this region of the SIFT genome (Figure 1G, middle). To condense the number of candidates, we used a protein variation effect analyzer (PROVEAN), which allowed us to predict whether an amino acid substitution or indel influenced the biological function of a protein (Figure S1C).\(^{11}\) The PROVEAN analysis found only one deleterious variant in the region (Figure 1G, bottom), located in the BIPINNATA (BIP) (Solvyc04g077210), the ortholog of Arabidopsis KNOX1 gene BREVIPEDICELLUS, was increased in the bip mutant.\(^{12}\) Quantitative RT-PCR (qPCR) was used to detect elevated levels of Tkn1 expression in SIFT compared to M82 (Figure 2E). KNOX overexpression is known to increase LC.\(^{13}\) These data suggest that the highly complex leaf phenotype seen in SIFT is caused by the high expression of Tkn1 facilitated by the bip mutation.

Although the LC in bip3 is quite similar to that of SIFT, the two genotypes have distinctly different leaflet shapes. Deep-learning-based nonlinear principal component analysis (PCA) with leaflet shapes in M82, bip3, and SIFT suggested that the bip3 leaf shape was different from that of M82; however, it was not the same as SIFT (Figure 2F). This trend was confirmed by different methods (Figure S3A). The SIFT leaflets were narrower than those of bip3 (Figure 2G). Additionally, LVD in bip3 was similar to that of M82 and differed from that of SIFT (Figures 2H and 2I). Thus, the mutation at the BIP locus was not sufficient to explain all leaf phenotypes observed in SIFT.

Gene co-expression network analysis indicates a role for WOX1 in regulating leaf phenotypes

To investigate the molecular basis for leaf phenotypes seen in SIFT, we collected mRNAs (meristem + P1–P2; P3; P4) and performed RNA sequencing (RNA-seq) to compare differentially expressed genes (DEGs) between M82 and SIFT. Tkn1 was detected as a DEG, suggesting the robustness of the analysis (Data S1A–S1C). However, the large number of DEGs precluded the identification of genes critical for generating differences in leaf shape between the two genotypes. The gene co-expression network (GCN) analysis can reveal biologically relevant information to identify the molecular mechanisms underlying biological processes.\(^{15}\) Therefore, we constructed GCNs with the RNA-seq data of M82 and SIFT and compared these networks to identify key genes responsible for the leaf phenotypes seen in SIFT. The genes used for the network analysis included a set of literature-curated genes involved in leaf development.\(^{16}\) The GCN for M82 showed differences in network structure, edge number, node number, and average degree between genes compared to the GCN for SIFT (Figure 3A; Table S1), indicating that many genes involved in leaf development were differentially expressed between the two genotypes. The community structure in the networks was analyzed based on the fast greedy modularity optimization algorithm, and gene ontology (GO) enrichment analysis by community was performed.\(^{16}\) Two communities (community 1 [C1] and community 2 [C2]) predominated in both networks (Figure 3A), and GO enrichment analysis by community showed that C1 was enriched for the same GO terms between M82 and SIFT networks; in particular, GO terms with higher fold enrichment (>50; full result of the GO enrichment analysis: Data S1D). In contrast, C2 was different (Figure 3B; Data S1D). In C2 of the M82 GCN, GO terms, such as “cytokinin biosynthetic processes” (GO: 0009691), “cytokinin metabolic process” (GO: 0009690), “regulation of cell cycle arrest” (GO: 0071156), and “cellular hormone metabolic process” (GO: 0034754; Data S1D), were enriched by more than 100-fold. However, there

Highly complex leaf phenotype seen in SIFT is caused by the bip mutation

To verify the effect of the bip mutation on leaf phenotypes, we investigated the morphology and early development of leaves in bip3, a bip mutant in the M82 background.\(^{12}\) The LC in bip3 was similar to that of SIFT (Figures 2A and 2B). Additionally, we confirmed that bip3 leaf primordia were active in generating multiple leaflets at the P4 stage, as seen in SIFT (Figure 2C). Although the BIP gene has been studied in Arabidopsis and the tomato, the expression pattern of the BIP gene in leaf primordia remains unknown.\(^{12,13}\) We developed and performed whole-mount in situ hybridization and detected BIP gene expression in the proximal part of the leaf primordium, where leaflet primordia emerged (Figure 2D). A previous study showed that the expression of TOMATO KNOTTED-1 (Tkn1; Solyc04g077210), the ortholog of Arabidopsis KNOX1 gene BREVIPEDICELLUS, was increased in the bip mutant.\(^{12}\) Quantitative RT-PCR (qPCR) was used to detect elevated levels of Tkn1 expression in SIFT compared to M82 (Figure 2E). Tkn1 overexpression is known to increase LC.\(^{13}\) These data suggest that the highly complex leaf phenotype seen in SIFT is caused by the high expression of Tkn1 facilitated by the bip mutation.
were no enriched GO terms with high fold enrichment in C2 of the SIFT GCN (Figure 3B; Data S1D), suggesting that C2 genes might be crucial for explaining differences in leaf phenotype between M82 and SIFT. To determine the statistically significant differences between the two networks, we performed comparative network analysis using the R package "DiffCorr." The analysis identified 160 DiffCorr genes, which are differentially correlated genes between the two networks (Data S1E). The 160 DiffCorr genes had distinct expression profiles between M82 and SIFT (Figure 3C), suggesting that these DiffCorr genes were responsible for the difference between the two networks. DiffCorr analysis revealed a WOX-like gene (Solyc03g118770) was the most significantly different between the M82 and SIFT GCNs (Data S1E), and the gene was located in C2 of the M82 GCN (Data S1F). Based on the phylogenetic analyses and alignments, the WOX-like gene is the tomato ortholog of Arabidopsis WOX1 (SlWOX1; Figures S3B and S3C) and is located on the long arm of chromosome 3 ([https://solgenomics.net/feature/17777644/details](https://solgenomics.net/feature/17777644/details)).

To understand the role of SlWOX1 in leaf development, we focused on the SlWOX1 sub-network, which comprises genes that are directly connected to SlWOX1. This sub-network showed that the SlWOX1 gene was connected to many genes involved in leaf development in the M82 GCN (Figure 3D). qPCR revealed that the SlWOX1 expression level in the SIFT leaf primordia was lower than that in the M82 samples (Figure 3E). Additionally, SlWOX1 was expressed at the margins of the leaf and leaflet primordia in M82 (Figures 3F and 3G). This expression was similar to that seen in Tomato eFP Browser (Figure S3D; [http://bar.utoronto.ca/efp_tomato/cgi-bin/efpWeb.cgi](http://bar.utoronto.ca/efp_tomato/cgi-bin/efpWeb.cgi)) and the pattern unaltered in SIFT (even though levels of expression were reduced; Figure S3E). SlWOX1 is expressed in leaf primordia and is involved in leaf lamina expansion in various plant species, including tomato. In the Medicago truncatula wox1 mutant, leaf vein density was lower than that in the wild type (WT). Therefore, SlWOX1 is likely to be a candidate gene for controlling both leaflet width and leaf vein density in the tomato.

**SlWOX1 is involved in leaf lamina expansion and leaf vascular development**

We mutated SlWOX1 in M82 using the CRISPR-Cas9 system and obtained a null mutation, referred to here as **Cr-wox1-1**...
Figure 3. GCN analysis with M82 and SiFT RNA-seq data
(A) Gene co-expression networks for genes involved in leaf development. Each node represents genes. Only nodes with at least one edge are represented. Left: M82 is shown; right: SiFT is shown.
(B) An overview of the enriched GO terms visualized by a bubble plot. The analysis was performed by the community in each network (C1 and C2). Each bubble represents a GO term, and only GO terms with higher fold enrichment (>50) are represented. For the full results of the GO enrichment analysis, see Data S1D.
(C) A profile of 160 DiffCorr genes. Top plot: number of differential correlations of each DiffCorr gene is shown. A higher number means a greater difference between the M82 and SiFT networks. Bottom heatmap: comparison of expression levels of each DiffCorr gene between M82 and SiFT is shown. Each expression level is shown as a blue-to-yellow-colored scale. The 160 DiffCorr genes were sorted by the number of differential correlations (left: low; right: high). The position of each gene is the same between the top and bottom panels.
(D) The Sl WOX1 gene network from M82 shown in (A). This network comprises only genes showing a direct connection to Sl WOX1.
(E) Sl WOX1 expression level in leaf primordia (n = 4). p = 0.011 (Welch’s t test). GAPDH was used as a reference gene.
(F and G) Whole-mount in situ localization of Sl WOX1 transcripts in (F) leaf primordia and (G) leaflet primordia of M82. Left: sense probe is shown; right: antisense probe in each panel is shown. White arrowheads indicate expression of Sl WOX1 (the purple color is the hybridization signal). Scale bars: 100 μm in (F) and (G).
See also Figure S3 and Table S1.
CR-woxl-1 plants showed narrower leaves (Figures 4B and 4C) and lower LVD than M82 (Figure 4D), matching the phenotype described in wox1 mutants in other species\(^{18,19}\) and tomato mutants, *unfused flower* (*uf*), and *SILAM1*.\(^{21-23}\) We noticed that these phenotypes were similar to those of the classical tomato mutant, *solanifolia* (*sf*). *sf* is known to have narrower leaflets and reduced vascular density,\(^{24}\) and our rough mapping of the *sf* mutation identified a genomic location close to that known for the WOX1 locus (at the end of chromosome 3; Figure S4A). We obtained the two known alleles of the mutation (*sf* and *sf^wl*) and another allele (e1862) from the Tomato Genetics Resource Center and Genes that make Tomatoes, respectively. *sf*, *sf^wl*, and e1862 arose from the Pearson, ROMA, and M82 backgrounds, respectively. *sf* has a 1-bp substitution (G to A) at position 230 in Sl WOX1, resulting in an amino acid swap from arginine to histidine in the conserved homeodomain. *sf^wl* and e1862 have a 1-bp substitution (G to A) at the splice site between intron 3 and exon 4, which results in a 10-bp shift to the next splice site. Therefore, these mutants have a 10-bp deletion from position 595 to 604, resulting in a premature stop codon, which truncates the WOX1 protein so that it lacks the conserved WOX domain (Figures 4A and S4B–S4E). All mutants showed narrower leaflets and lower LVD than their background genotypes (Figures 4B–4D). Hence, the phenotypes seen in these classic mutants were the same as those of CR-woxl-1. Hereafter, we refer to the SI WOX1 gene (previously known as *UF* and *SlLAM1*) as *SOLANIFOLIA* (*SF*). These results confirm the role of reduced expression of *SF* in conferring a narrower leaflet and lower LVD phenotypes in SiFT. Additionally, we have shown that *sf*, whose causative gene was previously unknown, is a wox1 mutant. None of the coding and promoter sequences of *SF* in SiFT showed SNPs compared to the reference tomato genome (Figures S4F and S4G), coinciding with the fact that the pattern of expression of *SF* was unaltered in SiFT compared to M82 (Figure S3E). Therefore, changes in trans-factors may be the cause of this repression.

**bip sf double mutant exhibits highly complex, narrow leaves and low LVD**

SiFT has a *bip* mutation and *SF* repression; therefore, we generated a *bip sf* double mutant to investigate the leaf phenotypes. Leaves in the double mutant between *bip3* and e1862 had more leaflets than those in e1862, and the leaflets were narrower than those of *bip3* (Figure 5A). Sometimes, secondary leaflets were observed on the 4th leaf (Figure 5B), and the secondary leaflets became more obvious in higher order leaves in the double mutant (Figure 5C). Additionally, LVD in the double mutant was lower than that in *bip3* (Figure 5D), and the *SF* expression level in *bip3* leaf primordia was similar to that of M82 (Figure S5A). These results suggest that these phenotypes were additive. These trends were confirmed by another double mutant with different combinations of mutants (Figures S5B–S5D). However, leaf morphology in the double mutant was not the same as that of the SiFT. Leaflets in SiFT leaves had many lobes (Figure 1B);
however, leaves of the double mutant did not have obvious lobes (Figures 5A and S5B). This may be due to the difference between the reduced expression and complete loss of function of the SF in the two genotypes. sf single mutants with truncated WOX1 protein, such as CR-wox1-1, sf^w1, and e1862, do not have any lobes on their leaves, whereas a weaker phenotype mutant, sf, had lobed leaves (Figure 4B), suggesting that sf regulated tomato leaf development in a dosage-dependent manner. Although a previous study showed that sf was involved in floral organ growth with mutants showing severe defects in fruit development, SiFT produced normal edible fruits (Figure S5E). This difference in fruit development might be related to the SF level. Therefore, a mutation at bip and SF repression created highly complex and narrower leaves with reduced leaf vein density in SiFT (Figure 5E).

The bip mutation in SiFT likely arose de novo and is unique to SiFT

Although previous studies constructed phylogenies with heirloom tomatoes, they were generated using whole-genome sequencing data. To understand the history of the bip mutation on chromosome 2, we first constructed a phylogenetic tree based on whole-genome sequencing data from 106 heirloom tomatoes (http://www.tomatogenome.net/accessions.html) to determine the relationship among them (Figure S6A). Subsequently, we constructed a phylogenetic network using the “PhyloNetworks” package in Julia to estimate whether various biological processes, such as hybridization, introgression, or horizontal gene transfer, occurred in SiFT. We used sequences around the BIP locus on chromosome 2 from 32 representative tomatoes based on the phylogeny of 106 heirloom tomatoes. S. pimpinellifolium, which is thought to be the progenitor wild species of the domesticated tomato, was used as an outgroup in the heirloom tomato phylogeny. S. cheesemaniae and S. lycopersicum var. cerasiforme were also used (Figure 6A). The network indicated that hybridization and introgression did not occur in SiFT, and a US heirloom tomato, Giant Oxheart (GiO), was sister to SiFT. However, GiO did not show the highly complex leaf phenotype characteristic of SiFT and lacked the BIP mutation (Figures 6B and 6C; Table 1). The phylogeny suggests that Druzba results from the cross between an ancestor of Glacier and an ancestor of GiO and SiFT. However, the bip mutation seen in SiFT did not exist in Druzba (Figures 6B and 6C). Additionally, the wild species, S. pimpinellifolium (Figure 6D), and the other tomato varieties did not harbor the SiFT-specific mutation at BIP (Figures S6B and S6C). These data suggest that the bip mutation in SiFT is likely a de novo mutation, instead of standing genetic variation, and is unique in SiFT.

DISCUSSION

We found that SiFT, an heirloom tomato, has a highly complex leaf phenotype with narrow leaflets. Genomic analysis showed...
Figure 6. Reconstruction of breeding history and comparison of SNPs data
(A) PhyloNetwork constructed using sequences around the BIP locus on chromosome 2. The network describes various biological processes, such as hybridizations or introgressions (green lines). The bootstrap values (those with >50%) are indicated on the tree, and the percentage inheritance is indicated on the branches.
(B) Magnified view of the network shown in (A) focusing on SIFT.
(C and D) Comparison of SNPs data around the BIP locus (Solyc02g089940) with the heirloom tomato (C) and a wild species: S. pimpinellifolium (D). Each vertical black line indicates a SNP.
(E) Simplified phylogeny in the genus Solanum with leaf morphologies. Circles indicate presumed key events during evolution and breeding. See text for details.
that the cultivar carries a mutation in the BIP gene, which encodes a BLH protein. Our experimental analyses demonstrated that the bip mutation found in SiFT is the likely cause of an increase in Tkn1 expression, prolonging morphogenesis, and increasing complexity of the leaf primordia. A double mutant in saw1 saw2 (the bip orthologs in Arabidopsis), with ectopic KNOX1 expression, has increased leaf serrations, indicating that BLH genes act to limit leaf margin complexity via KNOX1 repression. With a few exceptions, such as some legume species, KNOX1 is involved in intra- and interspecific leaf-form variations in different plant lineages. Even with different species, genetic backgrounds, and population histories, similar phenotypes are sometimes caused by mutations in the same pathway or loci. These genes or pathways are known evolutionary hotspots. Previous studies have suggested that hubs in a regulatory network connecting upstream activators and downstream effector genes can be hotspots. KNOX1 is thought to be a hub gene with a few nodes with maximal information exchange with other nodes. Therefore, leaf morphology in SiFT is shaped by the alteration in an evolutionary hotspot via the bip mutation, indicating that KNOX1 is not only involved in intra- and interspecific morphological variation caused by natural selection but is also responsible for variation in crop varieties sculpted by artificial selection.

Deep-learning-based nonlinear PCA and leaf-shape analysis indicated that the leaflet shape in SiFT was narrower than that in bip3. Additionally, the leaf-vein density in SiFT was lower than that in bip3. We used comparative GCN analysis to identify the genetic alterations beyond BIP that explain the rest of the leaf phenotypes seen in SiFT. The WOX1 ortholog had the most altered correlations between the M82 and SiFT co-expression networks. The expression level of the SiFT leaf was lower than that in M82. The CRISPR-Cas9 wox1 mutant showed narrower leaves and lower vascular density. WOX genes promote lamina outgrowth by controlling auxin biosynthesis and regulating cell proliferation in cells expressing WOX1 and their surrounding cells, and leaf vascular development is influenced by marginal blade outgrowth. A recent study demonstrated that slami1 failed to establish auxin response maxima along the leaf margin. Hence, we proposed that SF functions in leaf lamina outgrowth and couples this growth feature with vascular patterning. WOX1 is present in the early diverging angiosperm Amborella trichopoda, but its function in basal angiosperms remains unknown. Additionally, no WOX1 homologs have been identified in monocots. In contrast, wox1 mutants in Arabidopsis and Medicago have narrower leaves than the WT. Therefore, function of this WOX1 gene in leaf development might be conserved across eudicots. A recent study demonstrated that Arabidopsis WOX1, 3, and 5 genes redundantly performed this function, suggesting that SF is a major regulator of leaf lateral blade expansion in tomato, unlike Arabidopsis. Leaves of bip and sf double mutants are more complex than those of wox1 and narrower than those of bip3, confirming that SF promotes secondary leaflet formation, as suggested by previous studies. Moreover, the double mutant showed low LVD. A recent study suggested that the regulation of local growth and differentiation in leaf primordia leads to diversity in leaf shape. BIP and SF regulate local growth and differentiation in leaf primordia: BIP functions in the proximal part of leaf primordia and SF functions in the marginal part. Therefore, we conclude that the highly complex and narrow leaf with reduced leaf-vein density seen in SiFT is caused by a combination of a mutation at bip and another unknown second-site mutation that leads to SF repression (Figure 6E).

A phylogenetic tree constructed with WGS data and a phylogenetic network constructed with sequences around the BIP locus revealed that German Red Strawberry and Giant Oxheart are sisters to SiFT. However, they lack the bip mutation and have a regular leaf shape. Moreover, none of the other varieties or wild species harbored the same mutation at the BIP seen in SiFT. The wild tomato species S. galapagense has increased LC; however, this is linked to promoter alterations in an atypical KNOX1 gene, PETROSELENUM, not in BIP. These data indicate that the BIP mutation seen in SiFT is a de novo mutation that occurs during breeding and is not likely to be an introgression from other varieties or wild species. These results are consistent with the fact that no cultivated tomatoes show a SiFT-like leaf shape. This uniqueness of leaf shape in SiFT likely led to the use of this variety as an ornamental and landscaping plant. It has been postulated that more complex toothed leaves have an advantage in gas exchange with decreasing temperature. SiFT is a Russian heirloom tomato; therefore, a highly complex leaf phenotype might be advantageous for growth in a low-temperature environment and, combined with the uniqueness of its leaf shape, might have helped to popularize the cultivar in these areas. Emerging data suggest that leaflet shape and leaf-vein density affect fruit sugar content in tomatoes. The fruit sugar content was different between M82 and SiFT (Figure 6F). Therefore,

See also Figure S6.
identification of these novel mutations provides new insights into the breeding history of heirloom tomatoes and suggests potential targets for enhancing fruit quality in tomatoes.

**STAR METHODS**

Detailed methods are provided in the online version of this paper and include the following:

- **KEY RESOURCES TABLE**
- **RESOURCE AVAILABILITY**
  - **Materials available**
  - **Lead contact**
  - **Data and code availability**
- **EXPERIMENTAL MODEL AND SUBJECT DETAILS**
  - **Plant material and growth conditions**
- **METHOD DETAILS**
  - **Morphological observations**
  - **Mutation calling and allele frequency analysis with whole-genome sequencing data**
  - **Phylogenetic analyses of isolated genes**
  - **Whole-mount in situ hybridization**
  - **Phylogenetic analysis with whole-genome sequencing data**
  - **Quantitative real-time PCR**
  - **Deep learning-based nonlinear PCA**
  - **DNA-seq and RNA-seq library preparation and sequencing**
  - **SNP calling and allele frequency analysis with DNA-seq data and phylogenetic analysis with whole-genome sequencing data**
  - **Mapping, normalization, and network analysis with RNA-seq data**
  - **CRISPR-Cas9 mutagenesis and plant transformation**
  - **Mapping sf**
  - **PhyloNetwork analyses**
- **QUANTIFICATION AND STATISTICAL ANALYSIS**

**SUPPLEMENTAL INFORMATION**

Supplemental information can be found online at https://doi.org/10.1016/j.cub.2021.08.023.

**ACKNOWLEDGMENTS**

This work was supported by grants from the USDA NIFA (2014-67013-21700), NSF, IOS (1558900), and JSPS KAKENHI (19K23742 and 20K06682 to H.N.). H.N. was a recipient of a JSPS Fellowship (13J00161). This work was supported by grants from the USDA NIFA (2014-67013-21700), NSF, IOS (1558900), and JSPS KAKENHI (19K23742 and 20K06682 to H.N.). H.N. was a recipient of a JSPS Fellowship (13J00161).

**AUTHOR CONTRIBUTIONS**

H.N. initiated the project. N.R.S. supervised the project. H.N. and N.R.S. designed the experiments. H.N. performed the majority of the experiments and analyses and prepared the figures. S.D.R., Z.C., K.Z., J.K., and Y.K. performed additional experiments. H.N. and S.D.R. analyzed the data. H.N. and N.R.S. wrote the paper with input from all authors.

**DECLARATION OF INTERESTS**

The authors declare no competing interests.

Received: April 14, 2021
Revised: June 17, 2021
Accepted: August 5, 2021
Published: September 1, 2021

**REFERENCES**


## STAR METHODS

### KEY RESOURCES TABLE

<table>
<thead>
<tr>
<th>REAGENT or RESOURCE</th>
<th>SOURCE</th>
<th>IDENTIFIER</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Bacterial and virus strains</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Escherichia coli</em> DH5α</td>
<td>Widely distributed</td>
<td>N/A</td>
</tr>
<tr>
<td><em>Escherichia coli</em> One Shot Mach1</td>
<td>Thermo Fisher Scientific</td>
<td>Cat# C862003</td>
</tr>
<tr>
<td><em>Agrobacterium tumefaciens</em> EHA105</td>
<td>Widely distributed</td>
<td>N/A</td>
</tr>
<tr>
<td><strong>Critical commercial assays</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TOPO TA Cloning Kit</td>
<td>Thermo Fisher Scientific</td>
<td>Cat# K202020</td>
</tr>
<tr>
<td>DIG RNA Labeling Kit</td>
<td>Roche</td>
<td>Cat# 11175025910</td>
</tr>
<tr>
<td>T7 RNA polymerase</td>
<td>Roche</td>
<td>Cat# 10881767001</td>
</tr>
<tr>
<td>NBT-BCIP Stock Solution</td>
<td>Roche</td>
<td>Cat# 11681451001</td>
</tr>
<tr>
<td>Transcriptor High Fidelity cDNA Synthesis Kit</td>
<td>Roche</td>
<td>Cat# 5081955001</td>
</tr>
<tr>
<td>Phusion High-Fidelity Polymerase</td>
<td>Thermo Fisher Scientific</td>
<td>Cat# F530L</td>
</tr>
<tr>
<td>Apex Taq DNA Polymerase Master Mix</td>
<td>Genesee Scientific</td>
<td>Cat# 42-132</td>
</tr>
<tr>
<td>iTaq Universal SYBR Green Supermix</td>
<td>BIO-RAD</td>
<td>Cat# 1725121</td>
</tr>
<tr>
<td>GeneJET Plant Genomic DNA Purification Mini Kit</td>
<td>Thermo Fisher Scientific</td>
<td>Cat# K0792</td>
</tr>
<tr>
<td><strong>Deposited data</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RNA-seq and DNA-seq data files</td>
<td>BioProject</td>
<td>PRJDB8552</td>
</tr>
<tr>
<td><strong>Experimental models: organisms/strains</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>M82</td>
<td>TGRC</td>
<td>LA3475</td>
</tr>
<tr>
<td>sF/wl</td>
<td>TGRC</td>
<td>LA2012</td>
</tr>
<tr>
<td>sF</td>
<td>TGRC</td>
<td>LA2311</td>
</tr>
<tr>
<td>Pearson</td>
<td>TGRC</td>
<td>LA0012</td>
</tr>
<tr>
<td>e1862</td>
<td>genes that make tomato</td>
<td><a href="http://zamir.sgn.cornell.edu/mutants/">http://zamir.sgn.cornell.edu/mutants/</a></td>
</tr>
<tr>
<td><strong>Oligonucleotides</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>M13_F (-20) GTAAACGACGCTGAC</td>
<td>Widely distributed</td>
<td>N/A</td>
</tr>
<tr>
<td>M13_R CAGGAAACAGCTATGAC</td>
<td>Widely distributed</td>
<td>N/A</td>
</tr>
<tr>
<td>Tkm1_RT_f: GAAGGTTGCATTGGCAGAAT</td>
<td>This paper</td>
<td>N/A</td>
</tr>
<tr>
<td>Tkm1_RT_r: TGATGAAGAGACATGG</td>
<td>This paper</td>
<td>N/A</td>
</tr>
<tr>
<td>SlWOX1_RT_f: TCCACTAATCGCAGAGGTG</td>
<td>This paper</td>
<td>N/A</td>
</tr>
<tr>
<td>SlWOX1_RT_r: TGCTACTCATAGCAAGGTC</td>
<td>This paper</td>
<td>N/A</td>
</tr>
<tr>
<td>SlGAPDH_F: GGTGCTGACTTCGTTGTTG</td>
<td>This paper</td>
<td>N/A</td>
</tr>
<tr>
<td>SlGAPDH_R: GCTCTGGCTTGTATTCATTTC</td>
<td>This paper</td>
<td>N/A</td>
</tr>
<tr>
<td><strong>Recombinant DNA</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>pCR2.1</td>
<td>Thermo Fisher Scientific</td>
<td>Cat# K202020</td>
</tr>
<tr>
<td>pCR2.1_BIP_CDS</td>
<td>This paper</td>
<td>N/A</td>
</tr>
<tr>
<td>pCR2.1_WOX1_CDS</td>
<td>This paper</td>
<td>N/A</td>
</tr>
<tr>
<td>pEn_Chimera</td>
<td>Fauser et al.41</td>
<td>RRID: Addgene_61432</td>
</tr>
<tr>
<td>pMR290</td>
<td>Bari et al.42</td>
<td>N/A</td>
</tr>
<tr>
<td><strong>Software and algorithms</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ImageJ</td>
<td>Schneider et al.43</td>
<td><a href="https://imagej.nih.gov/ij/">https://imagej.nih.gov/ij/</a></td>
</tr>
<tr>
<td>R v0.4.2</td>
<td>R Development Core Team44</td>
<td><a href="http://www.R-project.org">http://www.R-project.org</a></td>
</tr>
<tr>
<td>R package MOMOCS</td>
<td>Bonhomme et al.45</td>
<td><a href="https://github.com/MomX/Momocs">https://github.com/MomX/Momocs</a></td>
</tr>
<tr>
<td>MEGA6</td>
<td>Tamura et al.46</td>
<td><a href="https://www.megasoftware.net/">https://www.megasoftware.net/</a></td>
</tr>
<tr>
<td>CLC Genomics Workbench 11.0</td>
<td>QIAGEN</td>
<td>N/A</td>
</tr>
</tbody>
</table>

(Continued on next page)
RESOURCE AVAILABILITY

Lead contact
Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact, Neelima R. Sinha (nrsinha@ucdavis.edu).

Materials availability
Plasmids and plant materials generated in this research are all available from the Lead Contact upon request. Please note that the distribution of transgenic plants will be governed by material transfer agreements (MTAs) and will be dependent on appropriate import permits acquired by the receiver.

Data and code availability
All data used to determine the conclusions of the study are presented in the paper or supporting information. All data are available in the main text or supporting information. All DNA-seq and RNA-seq raw data are deposited on DDBJ DRA009167-009182 (Bio-Project: PRJDB8552). Source data files for all main and Supporting Information are available in the online version of the paper. All additional datasets are available from the corresponding author upon request. All R scripts and packages for analysis were deposited on GitHub. R script for RNA-Seq GCN analysis was deposited on GitHub (Link: https://github.com/Hokuto-GH/gene-coexpression-network-script).

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Plant material and growth conditions
We used *S. lycopersicum* M82 (M82) as the control genotype in the present study. The detailed origin of SiFT remains unknown; however, M82 is widely used to map populations, perform molecular biological experiments, and generate transgenics. The following lines and mutants were provided by the Tomato Genetics Resource Center, University of California, Davis (USA; https://tgrc.ucdavis.edu/): M82 (LA3475), *sf*^wl^ (LA2012), *sf* (LA2311), and Pearson (LA0012). SiFT and ROMA were from our own stocks. *bip3* (e1444m2) was obtained from the saturated mutation library of tomato.53 e1862 was obtained from Prof. Dani Zamir (https://plantscience.agri.huji.ac.il/dani-zamir). Tomato seeds were soaked in 50% bleach for 10 min, rinsed three times with water, and placed on a water-dampened paper towel in Phytatrays (Sigma Aldrich). The seeds were incubated in the dark at room temperature for 3 days and then transferred to a growth chamber set at 22 °C under long-day conditions (16/8 hour light/dark cycle and light intensity of 150-200 μmol/m²/s) for 4 days. After approximately 7 days, the seedlings had expanded cotyledons. These were then transplanted to 24-cell seedling propagation trays and grown in the walk-in chamber (MTPC432, Conviron) for 35 days as described previously54 and arranged in a randomized block design. The shoots or leaf primordia were frozen in liquid nitrogen immediately after sampling and stored at –80 °C until subsequent use for DNA and RNA extractions. All F2 plants were grown in a field at the University of California, Davis, with an interplant spacing of 30 x 30 cm² for transplanting (April-August, 2016).
METHOD DETAILS

Morphological observations
For morphological observations and tissue collection for RNA extraction, shoots and leaves were dissected under a dissection microscope (Discovery V12; Zeiss). To determine the vascular density of leaves, the 6th leaves from the first set of cotyledons were used and cleared using ethanol and 50% bleach following the methodology as described previously.\(^5\) The samples were then photographed under a microscope (ECLIPSE E600; Nikon), and vascular length per unit area was determined using ImageJ software\(^5\) (https://rsb.info.nih.gov/ij/) (n = 4). LC was determined by counting the number of leaflets and intercalary leaflets on a fully developed leaf (n = 29). LC and leaflet shape were analyzed for the leaves collected from the chamber. LC measures included all leaflets present on the leaves. After LC was obtained, the primary leaflets were removed and used for imaging and analysis of shape and size. The intercalary and secondary/tertiary leaflets were not used for shape analysis because of their smaller size and irregular shape. The binary images were then processed in R\(^4\) using MOMOC3, a shape analysis package.\(^4\)

Phylogenetic analyses of isolated genes
The predicted amino acid sequences of the isolated genes were aligned using ClustalW, and then readjusted manually if needed. Phylogenetic trees were reconstructed using MEGA6\(^4\) with the neighbor-joining method.\(^5\) Bootstrap values were derived from 1000 replicate runs. The ML phylogenetic tree with the highest log-likelihood is presented. Initial trees for the heuristic search were obtained automatically, and neighbor-join and BioNJ algorithms were applied to a matrix of pairwise distances estimated with MCL, and then the topology with a superior log-likelihood value was selected. The tree was drawn to scale, with branch lengths measured as the number of substitutions per site.

Whole-mount in situ hybridization
Portions of genes isolated in pCR 2.1 (Invitrogen) were amplified by PCR using the universal primers M13_F (-20) and M13_R. The amplified fragments were then used to produce digoxigenin (DIG)-labeled sense and antisense RNA probes using a DIG RNA Labeling Kit (Roche). Whole-mount in situ hybridization was performed as previously described.\(^5\) Shoots were fixed in 1 x PBST containing 4% (w/v) paraformaldehyde and 1% (w/v) glutaraldehyde for 2 h. The fixed samples were then dehydrated in an ethanol series (50%, 70%, 90%, and 100%; 10 min each). The dehydrated samples were stored in 100% methanol at -20°C until subsequent use. The dehydrated samples were treated with 1% (w/v) H2O2 for 30 min. Then, samples were rehydrated in an ethanol series (100%, 75%, 50%, and 25%; 5 min each). After proteinase K (80 mg/l) and 1 x PBST containing 4% (w/v) paraformaldehyde treatments, the samples were used for the hybridization with DIG-labeled sense and antisense RNA probes. The probes were synthesized using T7 RNA polymerase (Roche). For immunological detection, the samples were incubated in detection buffer containing NBT-BCIP (Roche) at 25°C for several hours or at 4°C overnight. Photographs were taken using an ECLIPSE E600 microscope (Nikon). The experiments were performed at least three times with multiple samples.

Quantitative real-time PCR
Total RNA was extracted from the leaf primordia of plants grown for a month. RNA concentration was determined using NanoDrop 1000 Spectrophotometer (Thermo Scientific). 10 ng total RNA was used to synthesize cDNA with the Transcriptor High Fidelity cDNA Synthesis Kit (Roche). Two 5 μM Anchored-oligo(dT) primer was used for the cDNA synthesis. GAPDH was used as a reference gene as previously described.\(^5\) The quantitative RT-PCR analysis was conducted using the following gene-specific primer pairs: Tkn1_RT_f and Tkn1_RT_r, SIW01_RT_f and SIW01_RT_r, and S1GAPDH_F and S1GAPDH_R. We confirmed that there was no difference between M82 and SIFT in the region used to design the primers, then M82 sequence was used to design the primers. Real-time PCR amplification was performed using iTaq Universal SYBR Green Supermix (Bio-Rad) in an iQ5 Real-Time PCR Detection System (Bio-Rad) in a total reaction volume of 10 μL (5 μL of iTaq Universal SYBR Green Supermix, 2.4 μL of sterile water, 0.8 μL each of 10 μM forward and reverse primers, and 1 μL of cDNA) for 45 cycles. Melting curves were analyzed to check primer specificity. Normalization was done against the average of the reference gene Sl GAPDH control (STAR Methods). Experiments were performed in triplicate from three independent tissue RNA extractions.

Deep learning-based nonlinear PCA
For nonlinear PCA on image data, we used leaflet images from M82, bip3, and SIFT for the analysis (4th leaf; n < 55) and adopted a pre-trained neural network with the ImageNet dataset, VGG19,\(^5\) as the feature extractor. Instead of the original scanned images, binary silhouette images were fed into the network to avoid the effects of non-morphogenetic features such as leaf color. We extracted image features from an intermediate layer, “block4_pool,” using the Keras 2.3.1 library (https://keras.io). Subsequently, linear PCA was applied to the image features. We performed no training of the neural network with our data; therefore, the feature extraction was completely agnostic on which genotypes the leaves came from.

DNA-seq and RNA-seq library preparation and sequencing
DNA-Seq libraries for BSA were prepared as previously described.\(^5\) DNA was extracted using the GeneJET Plant Genomic DNA Purification Mini Kit (Thermo Scientific) from plants grown for a month. DNA-Seq libraries for phylogenetic analysis were prepared based on BrAD-seq\(^4\) with the following modifications. After DNA fragmentation with Covaris E220 (Covaris), the fragmented DNA
was end-repaired, A-tailed, and adaptor ligated with a Y-adaptor. Enrichment PCR was then performed with the adaptor-ligated product as previously described. After the final library cleanup with AMPure beads (Beckman Coulter, Brea, CA, USA), RNA-seq libraries were prepared following the methodology described by Townsley et al. from four biological replicates of leaf primordia at four developmental stages (meristem + P1-P2; P3; P4). DNA-Seq libraries were sequenced on Novogene (Novogene, Sacramento, CA, USA). RNA-Seq libraries were sequenced at the University of California Berkeley Vincent J. Coates Genomics Sequencing Laboratory (QB3 genomics) using the HiSeq 2000 platform (Illumina).

SNP calling and allele frequency analysis with DNA-seq data
To detect SNPs in the SiFT genome and perform phylogenetic analysis, all variants were detected by the CLC Genomics Workbench 11.0 (CLC Bio, a QIAGEN Company, Aarhus, Denmark). After read mapping and local realignment, the fixed Ploidy variant detection function was used to calculate allele frequency. For phylogenetic analysis, the data were exported as vcf files. The SNPRelate package for R was used to determine the variant positions that overlapped between cultivars, and then all sequences were combined into a single gds file. This file was run through SNPhylo with the following parameters: the linkage disequilibrium was set to 1.0, as we wanted to exclude as few variants as possible based on this factor, the minor allele frequency was set to 0.05, and the missing rate was set to 0.1. One thousand bootstraps were performed for confidence intervals and significance. S. pimpinellifolium was used as an outer group. The output bootstrapped tree was displayed in MEGA. To detect SNPs in the SiFT genome and perform phylogenetic analysis, all variants were detected by the CLC Genomics Workbench 11.0 (CLC Bio, a QIAGEN Company, Aarhus, Denmark). After read mapping and local realignment, the fixed Ploidy variant detection function was used to calculate allele frequency. For phylogenetic analysis, the data were exported as vcf files. The SNPRelate package for R was used to determine the variant positions that overlapped between cultivars, and then all sequences were combined into a single gds file. This file was run through SNPhylo with the following parameters: the linkage disequilibrium was set to 1.0, as we wanted to exclude as few variants as possible based on this factor, the minor allele frequency was set to 0.05, and the missing rate was set to 0.1. One thousand bootstraps were performed for confidence intervals and significance. S. pimpinellifolium was used as an outer group. The output bootstrapped tree was displayed in MEGA.

Mapping, normalization, and network analysis with RNA-seq data
The 50 bp single-end sequence reads obtained were quality-trimmed and parsed to individual libraries using custom Perl scripts. All reads were mapped to the ITAG2.4 genome build (https://solgenomics.net/itag/release/2.4/list_files) using RSEM/txExpress with default parameters. The uniquely mapped read data were normalized using the Bioconductor package EdgeR ver. 2.11, with default parameters.49 The uniquely mapped read data were normalized using the Bioconductor package EdgeR ver. 2.11, with the trimmed mean of the M-values method. Bioinformatic and statistical analyses were performed on the iPLANT (Cyverse) Atmosphere cloud server. GCN analysis was performed following16 using the R script. The R script for RNA-Seq GCN analysis was deposited on GitHub (https://github.com/Hokuto-GH/gene-coexpression-network-script). For GO enrichment analysis, we used GENEONTOLOGY enrichment analysis tools (http://geneontology.org/docs/go-enrichment-analysis/). DiffCorr analysis was performed as previously described.17 Normalized count data from M82 and SiFT were used for the analysis. DiffCorr genes were then analyzed to identify the most different genes between the two genotypes at a 0.005 FDR cut-off. Cytoscape was used to analyze and visualize DiffCorr genes (https://cytoscape.org/). The number of edges of each DiffCorr gene was calculated by analyzing the network function in Cytoscape. The numbers were then compared to determine the most different gene between the two genotypes.

CRISPR-Cas9 mutagenesis and plant transformation
For CRISPR-Cas9 mutagenesis, guide (g) RNAs for SF/SIWOX1 (Solyco3 g118770) were designed using CCTop (https://cctop.cos.uni-heidelberg.de:8043/help.html) and two gRNAs were chosen (Figure S3). The tandem tRNA–gRNA array for R was synthesized (Genescript) and was cloned into pEn_Chiperma13 by Golden Gate reaction using BbsI followed by a Gateway LR reaction into pMR29052 to generate the final binary vector. This vector was transformed into the tomato cultivar M82 by Agrobacterium tumefaciens (EHA105)-mediated transformation. The transformation was performed at the Ralph M. Parsons Foundation Plant Transformation Facility (University of California, Davis). Cotyledon pieces were soaked in the Agrobacterium solution for five minutes. Then the cotyledon pieces were co-cultivated on medium consisting of Murashige and Skoog minimal organics medium (MSO) modified with 20 g/L glucose, 0.75mg/L trans-zeatin, 1.0 mg/L IAA, and 200 uM acetosyringone, and incubated at 23°C for 2-3 days in the dark. The cotyledon pieces were transferred to induction medium consisting of MSO modified with 20 g/L glucose, 0.75mg/L trans-zeatin, 1.0 mg/L IAA, 400 mg/L carbenicillin, 150 mg/L timentin, and 100mg/L kanamycin sulfate. After 10 days, the cotyledon pieces were transferred to MSO modified with 20 g/L glucose, 1mM MES, 0.75mg/L trans-zeatin, 0.0 mg/L IAA, 400 mg/L carbenicillin, 150 mg/L timentin, and 100mg/L kanamycin sulfate until shoots began to form. The plates were incubated at 26°C under 16hr light 8 hr dark and 30 µmol/m²/s of light. Once shoots began to form, the developing shoots and callus were transferred to elongation medium consisting of MSO modified with 20 g/L glucose, 0.1mg/L trans-zeatin, 400 mg/L carbenicillin, 150 mg/L timentin, and 100mg/L kanamycin sulfate. Once a shoot reached 2-4 cm in height, it was transferred to rooting medium consisting of MSO modified with 20 g/L glucose, 0.2mg/L IBA, 400 mg/L carbenicillin, 150 mg/L timentin, and 100 mg/L kanamycin. First-generation (T0) transgenics were genotyped using GT-seq following Campbell et al. It revealed a single nucleotide substitution (C to A) in the gRNA2 (g2) region. There were no T0 transgenics with mutations in the gRNA1 (g1) region. After genotyping and self-pollination in a greenhouse, we obtained T1 plants with the mutated sf/slwox1 gene. First, we screened these plants by leaf phenotypes because wox1 mutants have narrower leaflets than WT based on previous studies with various kinds of plant species. We then performed genotyping by sequencing to confirm whether each individual had the sf/slwox1 mutation.

Mapping sf
A total of 23 cleaved amplified polymorphic sequence markers were used to map solanifolia. The markers spanned 53.50 cM to 167.50 cM (Data S1G; Figure S4A), on chromosome 3. The CAP markers were derived from Tomato-EXPEN 2000, amplified using PCR, and digested with restriction enzymes. The PCR conditions were as follows: denaturation at 94°C for 5 min, followed by 35
cycles of 94°C for 30 s, 50–55°C for 45 s, and 72°C for 1 min, with a final extension at 72°C for 5 min. Restriction digestion was performed at 37°C for 1.5 h. The products were then separated on an agarose gel (2% or 3%) and stained with ethidium bromide.

**PhyloNetwork analyses**

To perform phylogenetic analysis, all SNPs detected by CLC Genomics Workbench 11.0 (CLC Bio, QIAGEN Company, Aarhus, Denmark) from whole-genome sequencing obtained from the 360 genomes project were exported as a vcf. file. The VCFtools package was used to convert the vcf. files to fasta files, and these sequences were aligned using ClustalW. All aligned SNPs from the two megabase regions surrounding the *BIP* gene for 32 cultivars were run through the TICR pipeline. They were then analyzed using PhyloNetworks with default settings with the following exceptions: number of runs was set to 10 and Nfail was set to 10. After the hybrid network was obtained, bootstrap analysis was performed in PhyloNetworks using default settings with the following exceptions: number of runs was set to 10 and Nfail was set to 10. These adjustments were made to decrease the processing time. The bootstrapped tree was output using a Dendroscope.

**QUANTIFICATION AND STATISTICAL ANALYSIS**

All statistical analyses were performed using JMP software (JMP Pro 14.0.0, 2018). To determine the statistical significance, measurements were modeled using a general linear regression model and tested by one-way ANOVA followed by Tukey’s HSD test and Welch’s t test (p < 0.05), if necessary.