

Bucknell University

Bucknell Digital Commons

Faculty Journal Articles

Faculty Scholarship

2023

Nuclear phylogeny and insights into whole-genome duplications and reproductive development of Solanaceae plants

Jie Wang

Weibin Xu

Junwen Zhai

Yi Hu

Jing Guo

See next page for additional authors

Follow this and additional works at: https://digitalcommons.bucknell.edu/fac_journ



Part of the [Biology Commons](#), [Botany Commons](#), [Evolution Commons](#), [Genomics Commons](#), and the [Molecular Genetics Commons](#)

Recommended Citation

Huang, J., W. Xu, J. Zhai, Y. Hu, J. Guo, C. Zhang, Y. Zhao, L. Zhang, C.T. Martine, H. Ma, C. Huang. 2023. Nuclear phylogeny and insights into whole-genome duplications and reproductive development of Solanaceae plants. *Plant Communications* <https://doi.org/10.1016/j.xplc.2023.100595>.

This Article is brought to you for free and open access by the Faculty Scholarship at Bucknell Digital Commons. It has been accepted for inclusion in Faculty Journal Articles by an authorized administrator of Bucknell Digital Commons. For more information, please contact dcadmin@bucknell.edu.

Authors

Jie Wang, Weibin Xu, Junwen Zhai, Yi Hu, Jing Guo, Caifei Zhang, Yiyong Zhao, Lin Zhang, Christopher T. Martine, Hong Ma, and Chien-Hsun Huang

Nuclear phylogeny and insights into whole-genome duplications and reproductive development of Solanaceae plants

Jie Huang^{1,2}, Weibin Xu², Junwen Zhai³, Yi Hu⁴, Jing Guo¹, Caifei Zhang¹, Yiyong Zhao¹, Lin Zhang¹, Christopher Martine⁵, Hong Ma^{4,*} and Chien-Hsun Huang^{1,*}

¹State Key Laboratory of Genetic Engineering and Collaborative Innovation Center of Genetics and Development, Ministry of Education Key Laboratory of Biodiversity and Ecological Engineering, Institute of Plant Biology, Center of Evolutionary Biology, School of Life Sciences, Fudan University, Shanghai 200438, China

²Guangxi Key Laboratory of Plant Conservation and Restoration Ecology in Karst Terrain, Guangxi Institute of Botany, Guangxi Zhuangzu Autonomous Region and Chinese Academy of Sciences, Guilin 541006, China

³Key Laboratory of National Forestry and Grassland Administration for Orchid Conservation and Utilization, Fujian Agriculture and Forestry University, Fuzhou 350002, China

⁴Department of Biology, the Huck Institutes of Life Sciences, the Pennsylvania State University, University Park, State College, PA 16802, USA

⁵Department of Biology, Bucknell University, Lewisburg, PA 17837, USA

*Correspondence: Hong Ma (hxm16@psu.edu), Chien-Hsun Huang (huang_ch@fudan.edu.cn)

<https://doi.org/10.1016/j.xplc.2023.100595>

ABSTRACT

Solanaceae, the nightshade family, have ~2700 species, including the important crops potato and tomato, ornamentals, and medicinal plants. Several sequenced Solanaceae genomes show evidence for whole-genome duplication (WGD), providing an excellent opportunity to investigate WGD and its impacts. Here, we generated 93 transcriptomes/genomes and combined them with 87 public datasets, for a total of 180 Solanaceae species representing all four subfamilies and 14 of 15 tribes. Nearly 1700 nuclear genes from these transcriptomic/genomic datasets were used to reconstruct a highly resolved Solanaceae phylogenetic tree with six major clades. The Solanaceae tree supports four previously recognized subfamilies (Goetzeioideae, Cestroioideae, Nicotianoideae, and Solanoideae) and the designation of three other subfamilies (Schizanthoideae, Schwenckioideae, and Petunioideae), with the placement of several previously unassigned genera. We placed a Solanaceae-specific whole-genome triplication (WGT1) at ~81 million years ago (mya), before the divergence of Schizanthoideae from other Solanaceae subfamilies at ~73 mya. In addition, we detected two gene duplication bursts (GDBs) supporting proposed WGD events and four other GDBs. An investigation of the evolutionary histories of homologs of carpel and fruit developmental genes in 14 gene (sub)families revealed that 21 gene clades have retained gene duplicates. These were likely generated by the Solanaceae WGT1 and may have promoted fleshy fruit development. This study presents a well-resolved Solanaceae phylogeny and a new perspective on retained gene duplicates and carpel/fruit development, providing an improved understanding of Solanaceae evolution.

Key words: Solanaceae, phylogeny, transcriptome, carpel and fruit development, whole-genome duplication

Huang J., Xu W., Zhai J., Hu Y., Guo J., Zhang C., Zhao Y., Zhang L., Martine C., Ma H., and Huang C.-H. (2023). Nuclear phylogeny and insights into whole-genome duplications and reproductive development of Solanaceae plants. *Plant Comm.* 4, 100595.

INTRODUCTION

Whole-genome duplications (WGDs) contribute to genomic novelty due to the retention of a subset of gene duplicates and have been implicated in functional complexity and adaptation, speciation, and organismal diversity (Stebbins, 1940; Levin, 1983; Soltis et al., 2004; Rieseberg and Willis, 2007; Maere and Van de Peer, 2010; Mayrose et al., 2011; Arrigo and Barker, 2012). Genetic and

functional changes in the retained gene duplicates may enable organisms to take advantage of new ecological opportunities or cope with new environmental challenges (Ohno, 1970; Hahn,

Published by the Plant Communications Shanghai Editorial Office in association with Cell Press, an imprint of Elsevier Inc., on behalf of CSPB and CEMPS, CAS.

Plant Communications

2009; Maere and Van de Peer, 2010; Schranz et al., 2012; Fawcett et al., 2013). However, as more WGD events are supported by comparative genomic analyses, some of them seem not to be followed closely by increases in species richness. An explanation for this delay was provided by Schranz et al. (2012), who proposed that the positive impact of a WGD on diversification is manifest after a lag period. This lag could involve multiple factors, including genome stabilization through diploidization, an increase in genetic diversity among related species associated with differential retention of gene duplicates, and generation of key novel trait(s) following divergence of WGD-derived paralogs. Tank et al. (2015) used a dated phylogeny of angiosperm families to assess the correlation between WGDs and the increase in diversification rate. They compared the phylogenetic positions of nine well-characterized WGDs at the family level and above (including α , β , γ , ϵ , ρ , σ , τ , and those at the most recent common ancestor [MRCA] of Magnoliales and Laurales, MRCA of Ranunculales, and MRCA of Asteraceae) with the positions of 15 upshifts in diversification rate and found that only one such increase coincided with the location of a WGD. However, they found that four of six WGDs were linked to upshifts with statistical support when they examined upshifts one to three nodes downstream of six WGDs (excluding three of nine WGDs located at tip lineages).

However, even with a lag, the observed association between WGD and increased diversification rate could be due to the greater availability of genomic data in species-rich clades, which allows for reliable detection of WGDs. An examination of >50 WGDs and 84 families with various levels of species richness across angiosperms suggested that species-rich families have a greater tendency to have undergone WGD, but some small families also had WGDs, suggesting that WGDs do not always lead to greater species richness (Ren et al., 2018). Another study of >100 WGDs across angiosperms by Landis et al. (2018) found that 70 of 99 WGDs were accompanied by higher species richness compared with sister clades. However, the positions of WGDs detected in these large-scale phylogenies are often ambiguous because of limited sampling of taxa related to the WGDs. Therefore, investigations with representatives of major lineages in plant groups are needed to detect WGDs with more precise phylogenetic locations. Solanaceae include sister groups with vastly different numbers of species (Solanoideae, ~1940 species and Nicotianoideae, ~125 species) (Schranz et al., 2012), providing an excellent group for investigating evolutionary events related to WGDs and the lag-time hypothesis.

Solanaceae are a cosmopolitan, moderately large angiosperm family with tremendous economic importance, containing approximately 100 genera and ~2700–3000 species (Särkinen et al., 2013; Christenhusz and Byng, 2016; Knapp and Peralta, 2016; Raj et al., 2022). One family member, potato (*Solanum tuberosum*), is the fourth most important crop worldwide (Ezekiel et al., 2013) after wheat, rice, and maize, with an annual yield of 368 million tons (FAO, 2020). Solanaceae also include many vegetables, such as tomato (*Solanum lycopersicum*), aubergine/eggplant (*Solanum melongena*), pepper (*Capsicum annuum* and allied species), and tomatillo (*Physalis*). In addition, tobacco (*Nicotiana tabacum*) is widely grown for its leaves. Popular ornamentals include butterfly flower (*Schizanthus*), bedding plants (*Petunia* × *hybrida*), and

Genome duplications and development in Solanaceae

trumpet flower (*Brugmansia*); other species (e.g., mandrake [*Mandragora*] and henbane [*Hyoscyamus niger*]) have been used as medicinal herbs since ancient times. Several economically important Solanaceae species are also widely used as experimental model organisms, including *Petunia*, tobacco, tomato, and potato.

Solanaceae members are distributed across all continents except Antarctica in diverse habitats from deserts to tropical rainforests (Knapp, 2002b; Dupin et al., 2017). The family exhibits a great variety of life forms, including herbs, shrubs, and even a few trees, with diverse flower, fruit, seed, and embryo morphologies (Knapp, 2002a, 2002b; Barboza et al., 2016). Barboza et al. (2016) summarized Solanaceae classification with four subfamilies (Goetzeoideae, Cestroideae, Nicotianoideae, and Solanoideae), 15 tribes, and 96 genera (Figure S1A). However, some tribes/genera remained *incertae sedis* because of the uncertainties in Solanaceae phylogeny and were not assigned to a subfamily and/or tribe in Barboza et al. (2016) (Figures S1A and S1B). In this study, we primarily follow the classification of Barboza et al. (2016), with some discussion of subfamily and tribe designations proposed by others in the context of results here, including the rank of subfamilies for *Schizanthus* (Schizanthoideae), *Schwenckia* (Schwenckioideae), and *Petunia* (Petunioideae) (Olmstead and Palmer, 1992; Olmstead et al., 1999; Martins and Barkman, 2005; Olmstead and Bohs, 2007). Also, some unassigned Solanoideae genera in Barboza et al. (2016) were classified into (sub)tribes in other studies, such as *Latua* (Latueae), *Jaborosa* (Jaboroseae), *Solandra* (subtribe Solandrinae; tribe Solandreae), and *Mandragora* (Mandragoreae) (Olmstead and Palmer, 1992; Olmstead et al., 1999, 2008; Martins and Barkman, 2005; Dillon et al., 2007; Olmstead and Bohs, 2007; Tu et al., 2010; Orejuela et al., 2017; Finot et al., 2018).

The phylogenetic relationships of Solanaceae have been revealed in previous studies (Figures S1B–S1D) (Olmstead et al., 2008; Särkinen et al., 2013; Ng and Smith, 2016). Although these studies have made great contributions, the number of genes used was relatively small due to technical limitations, and fewer than 10 genes (chloroplast genes alone or with nuclear genes) were used in these previous analyses. The limited phylogenetic signals may have resulted in uncertainties in the positions of some subfamilies/tribes/genera, including the branching order of Schizanthoideae, Goetzeoideae, and Schwenckioideae (subfamily level), the relationships among Solanoideae lineages (such as Datureae, Juanulloae, Lycieae, *Latua*, *Jaborosa*, and *Nicandra*), and other controversial positions of some tribes/genera (such as Petunioideae, Schwenckioideae, and *Reyesia*).

Recently, single and low-copy genes have been shown to be effective for species phylogenetics at different scales (Zhang et al., 2012; Yang and Smith, 2014), and hundreds to thousands of low-copy nuclear genes have been obtained from non-model organisms and used in multiple phylogenetic analyses (Yang et al., 2015; Huang et al., 2016a, 2016b, 2022; Xiang et al., 2017; Qi et al., 2018; Leebens-Mack et al., 2019; Zhang et al., 2020b, 2021, 2022; Guo et al., 2020; Zhao et al., 2021). A well-resolved Solanaceae phylogeny using multiple nuclear genes can provide an important reference for comparative and evolutionary analyses of Solanaceae.

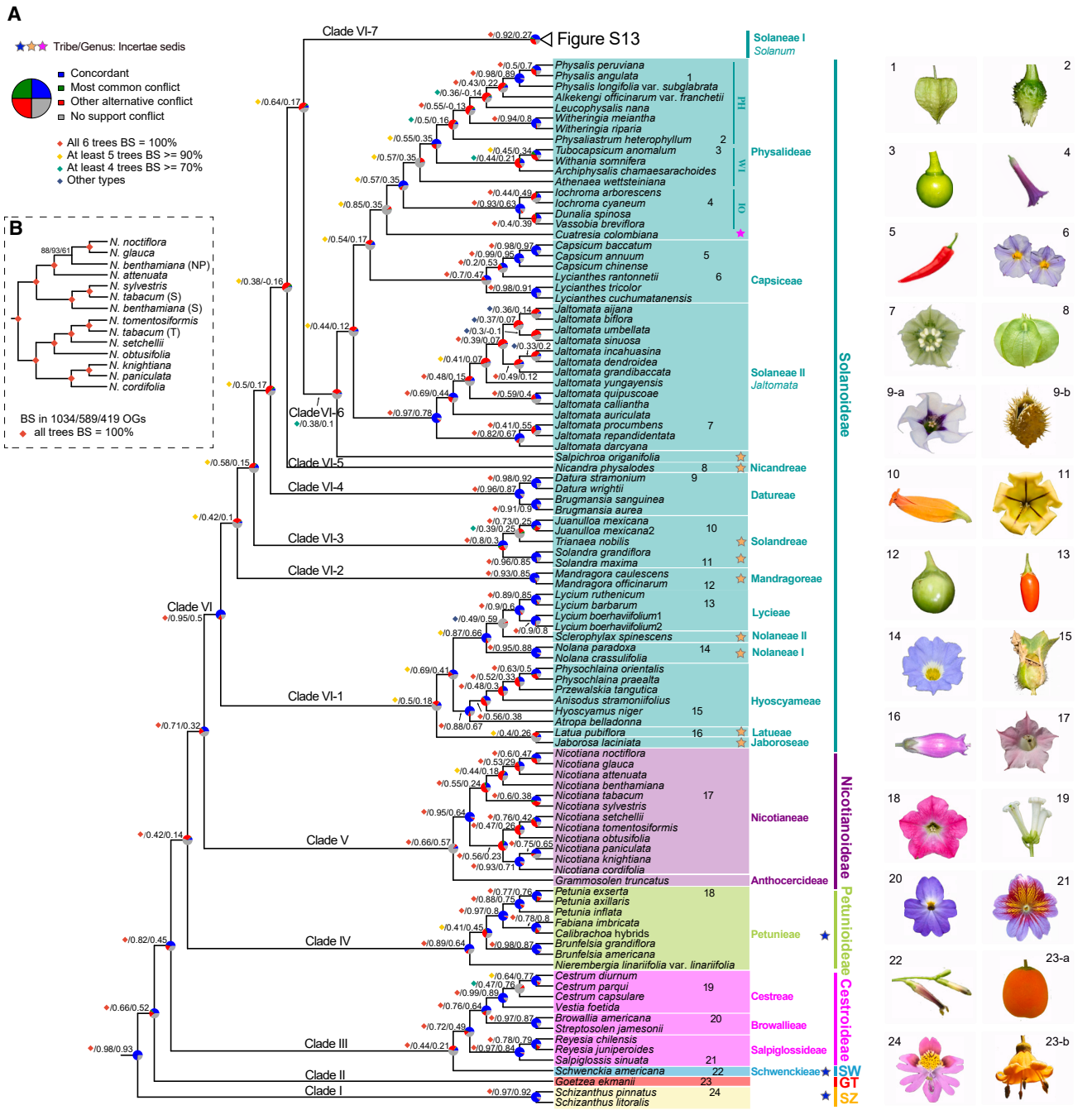


Figure 1. Summary of Solanaceae phylogeny.

(A) Pie charts represent the gene–tree conflict at each node of the 1699 OGs, with colors as follows: blue (proportion of gene trees in concordance), green (in conflict with the dominant alternative topology), red (in conflict with all other topologies), and gray (unsupported, with less than 50% BS). Three measurements for each node are presented at the left and correspond to BS/quartet support (QS)/ICA values, respectively. Six coalescence trees were reconstructed from 1699, 1263, 1034, 857, 589, and 419 OGs (see section “methods”) inferred by ASTRAL. Four levels of BS values for the six gene trees are denoted with solid diamonds in four colors: red, 100% BS in all six trees; yellow, 100% BS in fewer than six trees and ≥ 90% BS in at least five trees; green, ≥ 90% BS in fewer than five trees and ≥ 70% BS in at least four trees; dark blue, alternative topologies in our results. Species names are shown as tip labels on the tree, and species in different subfamilies and some tribes not assigned to a subfamily are highlighted with distinct colored backgrounds, with the names of the subfamilies/tribes to the right. Tribes/genera remaining *incertae sedis* in Barboza et al. (2016) are marked with colored stars: blue, tribes/genera not assigned to a subfamily; orange, genera (in the subfamily Solanoideae) not assigned to a tribe; purple, genera (in the tribe Physalidae) not assigned to a subtribe. SZ, Schizanthoideae; GT, Goetzeoideae; SW, Schwenckioideae; PH, Physalinae; WI, Withaninae; IO, lochrominae. Photographs are not scaled and are provided by Jie Huang and Xinxin Zhu or downloaded from Web sites (Table S2). (1) *P. angulata*; (2) *Physalium heterophyllum*; (3) *Tubocapsicum anomalum*; (4) *L. cyaneum*; (5) *C. annuum*; (6) *L. rantonnetii*; (7) *Jaltomata procumbens*; (8) *Nicandra physalodes*; (9) *D. stramonium*; (10) *Juanulloa mexicana*; (11) *Solandra maxima*; (12) *Mandragora officinarum*; (13) *L. barbarum*; (14) *Nolana paradoxa*; (legend continued on next page)

Plant Communications

Carpels are the female reproductive structures of angiosperms and develop into fruits, which protect young seeds and help to disperse mature seeds. Most Solanaceae plants produce either dry dehiscent or fleshy fruits, and tomato is an excellent system for studying the development and evolution of fleshy fruits (Azzi et al., 2015; Ortiz-Ramirez et al., 2018). Molecular genetic analyses of *Arabidopsis* regulatory networks of carpel and fruit development have facilitated comparative studies of other fruit crops (Seymour et al., 2013; Pfannebecker et al., 2016, 2017; Ortiz-Ramirez et al., 2018). Specifically, MADS-box genes are important for fruit development (Ocares and Mejia, 2016; Liu et al., 2020). MADS-box genes represent a group of transcription factors that control different aspects of organism development and cell differentiation. The name MADS-box comes from the first letters of four transcription factors that were initially identified in this family: *MINICHROMOSOME MAINTENANCE 1* (*MCM1*), *AGAMOUS* (*AG*), *DEFICIENS* (*DEF*), and *SERUM RESPONSE FACTOR* (*SRF*) (Schwarz-Sommer et al., 1990; Shore and Sharrocks, 1995). For example, the *AG* family member *AGL11/STK* is required for normal *Arabidopsis* ovule development (Favaro et al., 2003; Pinyopich et al., 2003), and loss of *AGL11* function triggers the seedless phenotype in fleshy grape fruits (Ocares and Mejia, 2016). Furthermore, seven tomato MADS-box genes were highly expressed during early fruit development (Victoria et al., 2003). Other genes for carpel development include *LEUNIG* (*LUG*), *SEUSS* (*SEU*), *bHLH*, and *YABBY* genes (Seymour et al., 2013; Pfannebecker et al., 2016, 2017). However, molecular evolutionary analyses of fruit-related genes in Solanaceae have been limited. More specifically, an examination of the retention and loss patterns of gene duplicates derived from WGD/whole-genome triplication (WGT) can provide clues about the impact of WGD/WGT on the functional evolution of genes for ovary/ovule and fruit development.

In this study, we generated 93 transcriptomic and shotgun genomic datasets of Solanaceae species and combined them with 87 other public datasets to identify six sets of nuclear genes from 180 species for use in reconstruction of a highly resolved Solanaceae phylogeny. Using this robust phylogenetic framework, we further investigated the evolutionary history of Solanaceae with divergence time estimation and WGD. We found that the MRCA of Solanaceae may have originated in the Late Cretaceous and diversified before the Cretaceous–Paleogene (K–Pg) boundary after a WGT event (WGT1). We also investigated the retention or loss of likely duplicates from this WGT for homologs of key genes for carpel and fruit development in 14 gene (sub)families and found that 21 gene clades have retained such duplicates.

RESULTS AND DISCUSSION

A well-resolved Solanaceae phylogeny with six major clades

To obtain nuclear genes for phylogenomic analyses, we sequenced 93 transcriptomes/shotgun genomes of Solanaceae species and

Genome duplications and development in Solanaceae

one outgroup species (*Dichondra*, Convolvulaceae). These, in combination with publicly available Solanaceae datasets (39 genomes and 48 transcriptomes), provide gene sequences of 180 Solanaceae species representing all four subfamilies (Solinoideae, Nicotianoideae, Cestroideae, and Goetzieoideae), 14 of 15 (93%) tribes as recognized by Barboza et al. (2016) (including Petunieae and Schwenckieae), and 51 (51%) genera (including 10 genera [*Schizanthus* and others with an orange star to the right of species names in Figure 1] not assigned to a tribe/subfamily by Barboza et al., 2016). The unsampled tribe Benthamielleae (*incertae sedis*) contains 15 species in three genera; another lineage not sampled here is represented by a single species, *Duckeodendron cestroides*, and was among possible sisters of most other Solanaceae (Olmstead et al., 2008; Särkinen et al., 2013; Ng and Smith, 2016). In addition, 43 datasets for outgroups were obtained from public resources. More detailed sampling information is provided in the section “methods” (Table S1). An initial set of 1699 ortholog groups (OGs) with single/low-copy genes were selected from sequences of nine representative Solanaceae species, and five smaller subsets (with 1263, 1034, 857, 589, and 419 OGs) of these genes were selected using a procedure with several filtering thresholds, such as length of aligned matrices and species coverage (section “methods”; Figure S2), to reconstruct the Solanaceae phylogeny (Figures S3–S9).

To assess the phylogenetic conflict between the coalescent trees and gene trees, we calculated the concordant bipartitions of 1699 single-gene trees using ASTRAL-Pro (Zhang et al., 2020a) (Figures S10 and S11). ASTRAL-Pro provided the quartet support (QS) and local posterior probability (LPP) for each branch to evaluate the concordance of the topology. A QS value greater than 0.33 and a high LPP value (maximum value is 1) represent relatively strong support for the proposed phylogeny. In addition, the phylogeny of each node was also evaluated by the number of supported gene trees and the internode certainty (ICA) values using PhyParts (Smith et al., 2015) (Figure S12 and Table S4). ICA metric was used to measure the degree of certainty for individual internal edges by looking at the frequency of conflicting bipartitions. The ICA value quantifies the degree of conflict at a given internode. An ICA value close to 1 indicates strong concordance in the bipartition of interest, whereas an ICA value close to 0 indicates equal support for one or more conflicting bipartitions. A negative value indicates that the internode of interest is in conflict with one or more bipartitions that have a higher frequency, and an ICA value close to -1 indicates the absence of concordance for the bipartition of interest (Smith et al., 2015). These analyses showed that all the nodes above the tribe level in our proposed phylogeny have QS values >0.4 (Figure S10) and LPP scores ~ 1 (Figure S11), suggesting high support for concordance. In addition, ICA scores of these nodes have an average of ~ 0.5 , which suggests that most relationships in our phylogeny are relatively robust (Table S4).

(15) *H. niger*; (16) *Latua pubiflora*; (17) *N. tabacum*; (18) *Petunia* × *hybrida*; (19) *Cestrum diurnum*; (20) *Browallia americana*; (21) *Salpiglossis sinuata*; (22) *Schwenckia americana*; (23) *Goetzea ekmanii*; (24) *S. pinnatus*.

(B) Relationships of *Nicotiana* species, with separate subgenomes of the two tetraploids *N. tabacum* and *N. benthamiana*. The summarized phylogeny is from three trees (1034 OGs, 589 OGs, and 419 OGs; see section “methods”) inferred by ASTRAL. Numbers at nodes are BS supports obtained from the three datasets, respectively, and red diamonds represent 100% BS in all three trees.

The coalescent trees reconstructed from the six datasets (Figures S3–S8) are highly resolved and consistent for most relationships, as summarized in Figure 1 (major clades in Solanaceae) and Figure S13 (subclades for the *Solanum* taxa sampled here), with only a few differences in support values and a few species relationships. We also concatenated the 419 genes into a supermatrix to reconstruct a phylogeny with RAxML (Figure S9), which was largely consistent with the results of coalescent methods. Solanaceae are highly supported (100/0.98/1/0.93, values corresponding to bootstrap [BS]/QS score/LPP value/ICA value, respectively) as monophyletic, with six major clades (Clade I to Clade VI) that have a consistent order of divergence and relationships. Clade I–V form a grade outside the large Clade VI as successive sister lineages of the remaining Solanaceae (Figure 1): Clade I and II are, respectively, the genus *Schizanthus* and the subfamily Goetzeoideae, whereas Clade III includes the tribe Schwenckieae and the subfamily Cestroideae (with three tribes here); Clade IV and V are the tribe Petunieae (or subfamily Petunioideae by some) and the subfamily Nicotianoideae, respectively; Clade VI is the subfamily Solanoideae, with seven subclades, Clade VI-1 to Clade VI-7, for the sampled taxa (Figures 1 and S13).

Our Solanaceae phylogeny places the genus *Schizanthus* (Clade I) as sister to other sampled members of Solanaceae and supports the designation of *Schizanthus* as a subfamily (Schizanthoideae). On the other hand, our sampling did not include another early branching lineage (*D. cestroides*) of Solanaceae, leaving the earliest branch of the family still uncertain. In Clade III, Cestroideae is monophyletic (100/0.72/1/0.49), with the tribe Salpiglossideae (*Salpiglossis* and *Reyesia*) being sister to a clade of two other sampled tribes (Cestreae and Browallieae, 100/0.76/1/0.64). The unsampled Benthanielleae was previously placed close to Cestroideae (Särkinen et al., 2013) (Figure S1C). The tribe Petunieae (Clade IV) occupies a lineage separate from other subfamilies in all trees here (100/0.89/1/0.64), supporting its designation as a subfamily (Petunioideae) as proposed previously (Olmstead et al., 1999; Martins and Barkman, 2005).

The subfamily Nicotianoideae (Clade V) is monophyletic (100/0.66/1/0.57) with both recognized tribes sampled here (Nicotianeae and Anthocercideae). Nicotianeae contains only the genus *Nicotiana* and is monophyletic (100/0.95/1/0.64). Among the sampled *Nicotiana* species, *N. tabacum* and *Nicotiana benthamiana* are tetraploids, whereas the others are diploids (Leitch et al., 2008; Siervo et al., 2013; Edwards et al., 2017; Schiavinato et al., 2020). The tetraploid nature of *N. tabacum* and *N. benthamiana* could have affected the selection of putative orthologs, thereby affecting the phylogenetic placement of these two species; therefore, we examined the topologies of the trees from the 1699 OGs and identified the *Nicotiana* species to which most of the genes from the tetraploids were similar, as support for possible parental lineage(s). Among the 1699 trees, 829 (48.8%) had *N. tabacum* genes being sister to *Nicotiana sylvestris* (proposed to be from a subgenome designated S), and 642 (37.8%) had *N. tabacum* genes being sister to *Nicotiana tomentosiformis* (T subgenome). For *N. benthamiana*, 714 (42%) trees had the sister group(s) with *Nicotiana noctiflora*, *Nicotiana glauca*, or *Nicotiana attenuata* (NP subgenome), whereas in 411 (24.2%) trees, the *N. benthamiana* sequences were close to those from *N. tabacum* and/or *N. sylvestris* (S subgenome). With the putative subgenome

information and taxon coverage, three smaller gene sets were selected (1034 OGs, 589 OGs, and 419 OGs; see section “methods”) and used to reconstruct the phylogeny of *Nicotiana* with separate putative subgenomes of the two tetraploids (Figure 1B). As expected, *N. tabacum* (S) and *N. tabacum* (T) are grouped with *N. sylvestris* and *N. tomentosiformis*, respectively. In addition, *N. benthamiana* (S) is sister to the clade consisting of *N. tabacum* (S) and *N. sylvestris*, whereas *N. benthamiana* (NP) is sister to a clade of *N. noctiflora* and *N. glauca* with moderate to strong support (BS = 61–93%), consistent with the previous hypothesis that a species related to *N. noctiflora* and *N. glauca* was a progenitor of *N. benthamiana* (Leitch et al., 2008; Schiavinato et al., 2020).

The largest subfamily, Solanoideae (Clade VI), is monophyletic (100/0.95/1/0.5), covering over 80% of Solanaceae species (Barboza et al., 2016); the topology here includes seven subclades (Clade VI-1 to Clade VI-7; Figure 1), with the first six subclades forming a grade outside Clade VI-7 (*Solanum*; Figure S13). Clade VI-1 (98–99/0.5/1/0.18) includes tribes Hyoscyameae and Lycieae, and also four unassigned genera (*Jaborosa*, *Latua*, *Nolana*, and *Sclerophylax*).

Several Solanoideae genera with previously uncertain positions (orange star in Figure 1) have well-supported placements here among the next few subclades. *Mandragora* (Clade VI-2) is monophyletic here (100/0.93/1/0.85). In Clade VI-3 (100/0.8/1/0.3), *Solandra* and *Trianaea* are successive sisters of *Juanulloa*. The results here support the proposed inclusion of *Juanulloa*, represented by *Juanulloa* in Barboza et al. (2016), in Solandreae by Orejuela et al. (2017). The next two lineages, Clade VI-4 and Clade VI-5, respectively, are the tribe Datureae (100/0.96/1/0.87) and the small genus *Nicandra* (once named *Nicandreae*; Olmstead et al., 1999).

The remainder of Solanoideae forms two highly diverse clades (Clade VI-6 and Clade VI-7). Clade VI-6 comprises four lineages (78–90/0.38/0.99/0.1): the genera *Salpichroa* and *Jaltomata* (part of tribe Solaneae; Olmstead et al., 1999; Olmstead et al., 2008) and the tribe Capsiceae occupy the positions of successive sisters of a clade with Physalideae. To probe the factors affecting the phylogenetic position of *Jaltomata*, we performed further analyses with evidence (Figures S14 and S15) for possible hybridization (see the section “comparison of Solanaceae phylogeny with previous studies”). In tribe Capsiceae, *Lycianthes*, a large genus with ~150 species, is paraphyletic and split into two clades, in agreement with the paraphyly of the genus in other analyses (Olmstead et al., 2008; Särkinen et al., 2013; Spalink et al., 2018). The fourth lineage of Clade VI-6 is the tribe Physalideae (=Physaleae), which includes three subtribes, Iochrominae, Withaninae, and Physalidinae (=Physalinae), and the genus *Cuatresia*, which is not assigned to a subtribe (Barboza et al., 2016).

Clade VI-7 contains the sampled species of *Solanum* (see Figure S13), the largest and most diverse Solanaceae genus (with ~1400 species). Early studies with a large number of *Solanum* species defined 12 major clades, including *Thelopodium* as sister to other *Solanum* species, which form two large clades, Clades I and II (Bohs, 2005; Weese and Bohs, 2007; Särkinen et al., 2013; Tepe et al., 2016; Gagnon et al., 2022). Clade I includes the Clade

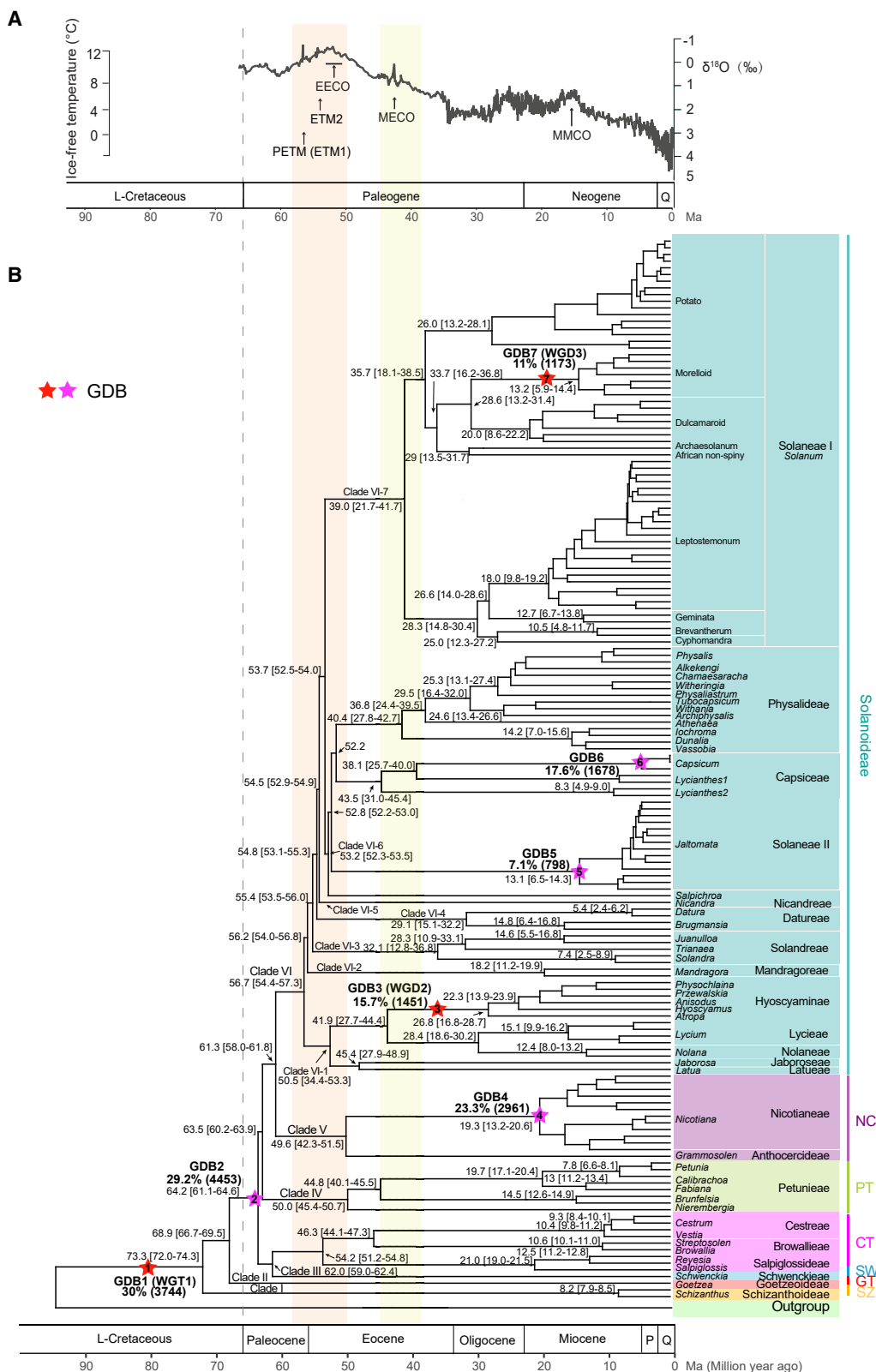


Figure 2. Divergence times and large-scale GDBs of Solanaceae.

The chronogram was estimated with treePL based on the ML tree of 419 concatenated genes. Numbered red and purple stars indicate the large-scale GDBs identified in this study, with percentages and numbers of duplicated gene families shown for each node.

(legend continued on next page)

M (with the African non-spiny, Normania, Archaeosolanum, Valdiviense, Dulcamaroid, and Morelloid subclades) and the Potato clade (including the Regmandra, Petota, Tomato, and some other minor clades), whereas Clade II contains the very large Leptostemonum clade (including Acanthophora, Torva, and Old World/Eastern Hemisphere Spiny) and several smaller clades such as Cyphomadra, Brevantherum, Geminata, and Allophyllum/Wendlandii (Bohs, 2005; Weese and Bohs, 2007; Särkinen et al., 2013; Tepe et al., 2016; Gagnon et al., 2022) (see also Figure S13 and the related supplemental information). Our limited phylogenetic analyses included 81 *Solanum* species (Clade VI-7; Figure S13) with an emphasis on close relatives of tomato (the Tomato subclade with 13 taxa) and potato (with 22 taxa of the Petota subclade), and two other subclades of the Potato clade (Basarthurum, Etuberosum) that are relatively close to Petota/Tomato. The sampling also included members of Clade II, such as those of Cyphomadra, Brevantherum, Geminata, as well as Leptostemonum with several subclades (indicated on the far right in Figure S13). However, our sampling lacked species of the Thelopodium clade and some major clades of Clade I (such as Normania and Regmandra) and II (such as Wendlandii-Allophyllum). Previous *Solanum* phylogenetic studies with >80 species used several plastid or nuclear genes (Särkinen et al., 2013; Ng and Smith, 2016; Tepe et al., 2016) (see Figure S13 for more information). In addition, *Solanum* phylogenies were reconstructed using nuclear genes from 39 and 29 species, respectively (Gagnon et al., 2022; Tang et al., 2022). The analyses here using 1699 nuclear genes from 81 *Solanum* transcriptomic/genomic datasets provide a valuable additional set of relationships for comparison (see more detailed description and discussion of *Solanum* in Figure S13 and its figure legend).

An estimated Solanaceae origin before the K-Pg boundary and rapid divergence of major clades

The molecular divergence times of Solanaceae were estimated with treePL (Smith and O'Meara, 2012) and BEAST (Drummond et al., 2012) (see section "methods") with 10 fossil calibrations (Tables S5–S7). We used two previously reported alternative calibration points of each of the two Solanaceae fossils (Wilf et al., 2017; Dupin and Smith, 2018; Deanna et al., 2020) and devised four calibration strategies for time estimation (see section "methods"). In addition to the Solanaceae taxa included in the phylogenetic analyses, 38 additional outgroup species (Table S1) were included to accommodate fossil calibrations outside Solanaceae. A comparison of divergence times using different calibrations and methods is shown in Table S6 (also Figures S16–S20). The treePL age estimates showed two patterns (calibrations 1 and 4 are similar, and calibration 2 is similar to 3); in addition, the BEAST estimates are ~10 million years earlier than those of treePL with the same calibration. For example, the estimated crown age of our sampled Solanaceae species (MRCA of Schizanthoideae and Solanoideae) is 73.3 million years ago (mya) (treePL, calibrations

1 and 4; Figures S16 and S19), 65.1 mya (treePL, calibrations 2 and 3; Figures S17 and S18), or 83.3 mya (BEAST, calibration 1; Figure S20). The age differences among estimates using the four calibrations are mainly due to alternative placements of fossil 2 (assigned as stem group of either Physalideae or Solanoideae), which led to a variance of 1–28 million years; by contrast, there are only small differences (<1 million years) between the results from different placements of fossil 1 (assigned as stem or crown group of *Datura*).

For convenience, we describe in detail only the results of calibration 1 from treePL in subsequent sections, in part because these results are closest to the average of all four different calibrations and methods. The resulting chronogram with the estimated origin of the family and its lineages (Figures 2 and S16) provides the crown age of the sampled Solanaceae (lacking the *Duckeodendron* branch) at 73.3 mya with a 95% confidence interval (CI) of 72.0–74.3, before the K–Pg boundary with mass extinctions. Subsequently, major Solanaceae clades (Clade II to VI, at the subfamily level) diverged within 12 million years. The divergence time of Clade II from the remaining Solanaceae is 68.9 (CI, 66.7–69.5). The stem and crown ages, respectively, of the remaining clades, are 64.2 (CI, 61.1–64.6) and 62.0 (CI, 59.0–62.4) (Clade III), 63.5 (CI, 60.2–63.9) and 50.0 (CI, 45.5–50.7) (Clade IV), 61.3 (CI, 58.0–61.8) and 49.6 (CI, 42.3–51.5) (Clade V), and 61.3 (CI, 58.0–61.8) and 56.7 (CI, 54.4–57.3) (Clade VI). During the Eocene, Solanoideae (Clade VI) members diverged into subclades in less than 3 million years (56.7–53.7 mya) and contributed to making it the most diverse clade, with ~80% of all Solanaceae species. The divergences and expansion of Clade III to Clade VI occurred mainly during the period from the Paleocene to the early Eocene, including the Paleocene–Eocene Thermal Maximum (PETM) and Early Eocene Climatic Optimum (EECO) (Figure 2). The radiation with rapid divergences of major Solanoideae lineages seems to coincide with a dramatic climate shift, the PETM (~56 mya).

Our estimated crown age of Solanaceae (lacking the *Duckeodendron* branch) (73.3 mya) is older than those (61.9–30.4 mya) from previous analyses (Dillon et al., 2009; Janssens et al., 2009; Särkinen et al., 2013; Zamora-Tavares et al., 2016). One of the main reasons for the age differences is likely that the fossil calibration here, but not in most previous studies, included a recently discovered and well-preserved 52.2-million-year-old fossil of lantern fruits from terminal-Gondwanan Patagonia (Wilf et al., 2017; Deanna et al., 2020). Following this discovery, Dupin and Smith (2018) used the new fossil to constrain the minimum stem age of Solanoideae and found that the crown age of Solanoideae is 54 mya, which is closer to our result than most other results. However, owing to a lack of taxon samples, they did not estimate the divergence time of Solanaceae. The inclusion of this fossil and a well-resolved phylogeny likely provide a relatively accurate estimate of the divergence times of Solanaceae lineages.

(A) Global climate curve over the last 65 million years (modified from Zachos et al., 2008). Major paleo-climatic events during this period are indicated: PETM (ETM1), Paleocene–Eocene Thermal Maximum 1; ETM2, Eocene Thermal Maximum 2; EECO, Early Eocene Climatic Optimum; MECO, Mid-Eocene Climatic Optimum; MMCO, Mid-Miocene Climatic Optimum. Geological timescales are in million years. L-Cretaceous, Late Cretaceous; P, Pliocene; Q, Quaternary.

(B) Divergence times of Solanaceae. The mean ages of major clades are shown to the left of the corresponding nodes, with the 95% CI in square brackets. Names of six subfamilies are abbreviated: SZ, Schizanthoideae; GT, Goetzeoideae; SW, Schwenckioideae; CT, Cestroideae; PT, Petunioideae; NC, Nicotianoideae.

Plant Communications

WGD events near the origin of Solanaceae and during their history

The newly reconstructed Solanaceae phylogeny and numerous newly generated gene sequence datasets provide opportunities to detect evidence for WGD using a phylogenomic approach of gene tree reconciliation as described in previous studies (Jiao et al., 2011; Yang et al., 2015, 2018; Huang et al., 2016a, 2020; Xiang et al., 2017; Zhang et al., 2020b, 2021, 2022; Guo et al., 2020; Zhao et al., 2021). Most of the genes that are duplicated in a WGD event likely return to single copies because of the loss of one duplicate, and the detection of hundreds of gene duplications (GDs) at the same node on the phylogeny is taken as evidence for WGD. Such GD bursts (GDBs) also provide an estimated position of the WGD; to be conservative, we chose to describe GDBs that are shared by three or more of the species sampled here to minimize the chance of mistaking a cluster of small-scale duplications shared by two species for WGDs. Sequences from 152 Solanaceae species and three outgroups with transcriptomes or publicly available genomes were clustered using OrthoMCL (Li et al., 2003), resulting in 23 243 OGs. Among these OGs, 18 204 OGs with ≥ 15 taxa and alignment length ≥ 300 bp were used to reconstruct gene family trees whose topologies were then compared with the species tree to identify GDs. When the duplicate clades with $\geq 80\%$ BS were counted, seven nodes on the species tree that each had a GDB with GDs ≥ 300 (and accounting for $\geq 7\%$ of OGs for the node) were deemed as having sufficiently strong evidence for a proposed WGD (Figures 2 and S21; Table S8).

Two GDBs occurred along the Solanaceae backbone: in the MRCA of extant Solanaceae species (GDB1: 3744 GDs, 30%) and in the MRCA of Clade III through Clade VI (GDB2: 4453 GDs, 29.2%). Among the GDs of GDB1, almost 34.9% (898 GDs) had at least 30% of species coverage in each of the duplicate clades; however, only about 16.4% of the GDs in GDB2 (591 GDs) met a similar 30% species coverage threshold (Figures S21 and S22). We also detected a GDB in the MRCA of Hyoscyameae (GDB3: 1451 GDs, 15.7%) and four others at or within the genus level: in *Nicotiana* (GDB4, 2961 GDs, 23.3%), *Jaltomata* (GDB5, 798 GDs, 7.1%), *Capsicum* (GDB6, 1678 GDs, 17.6%), and the Morelloid clade of *Solanum* (GDB7, 1173 GDs, 11%). We further verified the seven GDBs using MAPS (Li et al., 2018) with the procedure from the 1KP study (Leebens-Mack et al., 2019). For each GDB, we selected a group of species in a ladder relationship to compare the number and proportion of duplicate genes in real and simulated data. Six of the GDBs (GDB1–4 and GDB6–7) were statistically supported by MAPS, as they have higher percentages of real duplications than those of their null simulations (Figure S23; Table S9). The MAPS result for GDB5 (Figure S23D) did not show higher performance compared with its null simulations, suggesting that this may not be a WGD.

The GDBs (GDB1, GDB2, GDB4, GDB5, and GDB6) that are shared by species with sequenced genomes were further examined for gene collinearity (synteny) as genomic support for WGDs (Table S10). The synteny analyses provided evidence for both GDB1 and GDB2 in the Solanaceae genomes; for

Genome duplications and development in Solanaceae

example, among the duplicated genes detected in *S. lycopersicum* from phylogenomic analysis for GDB1 (1424) and GDB2 (1131), 394 and 145 are located in syntenic blocks, respectively (see Figure S24 for examples). Similarly, a fraction of the phylogenomically detected gene pairs were located in syntenic blocks for GDB1 and GDB2 in *Petunia axillaris* (syntenic/total gene pairs: 253/1844 and 93/2242), *S. tuberosum* (325/3355 and 142/2710), *C. annuum* (222/2764 and 87/2719), *N. attenuata* (64/1424 and 32/1189), and *Jaltomata sinuosa* (155/1245 and 49/1605). However, GDB1 and GDB2 are very close to each other in the species phylogeny (phyto_151 and phyto_149 in Figure S22), with only three species being different (two species of *Schizanthus* and one of *Goetzea*) for the corresponding nodes. Therefore, if the genes of these three species are lost in duplicated subclades, a duplication placed at the GDB2 node might have occurred at the GDB1 node. Indeed, an examination of collinear segments in the *S. lycopersicum* genome for their gene pairs revealed that a large fraction of the paralogs (288/394 = 73.1%) from the three consecutive nodes (phyto_151, 150, and 149) are located in the same collinear segments (Figures S24A and S24B). Synonymous substitution rate (Ks) plots of the paralogs in *S. lycopersicum* corresponding to the three mapped nodes showed a peak at ~ 0.7 (Figure S24C). These results support the hypothesis that the duplicated genes of the three nodes likely originated from the same (genome) duplication event (GDB1) at the MRCA of Solanaceae sampled here, whereas the placement of some gene pairs at later nodes could be due to loss (or incomplete sequencing) of genes in species of Clade I and Clade II (Figure S21). A WGD event in Solanaceae has been proposed (Schlueter et al., 2004; Soltis et al., 2009), but its precise phylogenetic placement was previously uncertain in the absence of sequenced genomes for several subfamilies (Schizanthoideae, Goetzoideae, and Cestroideae). Analyses of sequenced Solanaceae genomes have led to a proposed WGT event shared by tomato, potato, pepper, *Nicotiana*, and *Petunia* (Xu et al., 2011, 2017; Sato et al., 2012; Qin et al., 2014; Bombarely et al., 2016). Combined with our phylogenomic/phylotranscriptomic results, these findings allow us to infer that GDB1 probably corresponds to the WGT event (Figures 2 and S21) and occurred at the MRCA of Solanaceae, as suggested by syntenic evidence and other pieces of supporting evidence (Figures S23 and S24). In contrast to the strong evidence for a WGT shared by all Solanaceae, smaller numbers of gene pairs were detected in syntenic blocks for the other GDBs: GDB4 (10/852 in *N. attenuata*), GDB5 (1/75 in *J. sinuosa*), and GDB6 (97/4995 in *C. annuum*); thus, it is possible that other mechanisms for gene duplication were responsible for many of the retained duplicates detected at these three nodes, although the lack of syntenic evidence could also be due to extensive genome rearrangements that disrupted syntenic relationships between gene duplicates.

We also estimated the number of GDs in the GDBs that probably resulted from tandem duplications. A low number and proportion of tandem GDs were detected in all GDBs (GDB1 [212 GDs, 5.7%]; GDB2 [285 GDs, 6.4%]; GDB4 [128 GDs, 4.3%]; GDB5 [19 GDs, 2.4%]) except GDB6. Among the 1678 GDs in GDB6 shared by *Capsicum*, 833 GDs (49.6%) were regarded as tandem duplicates rather than duplicates generated by genome-wide

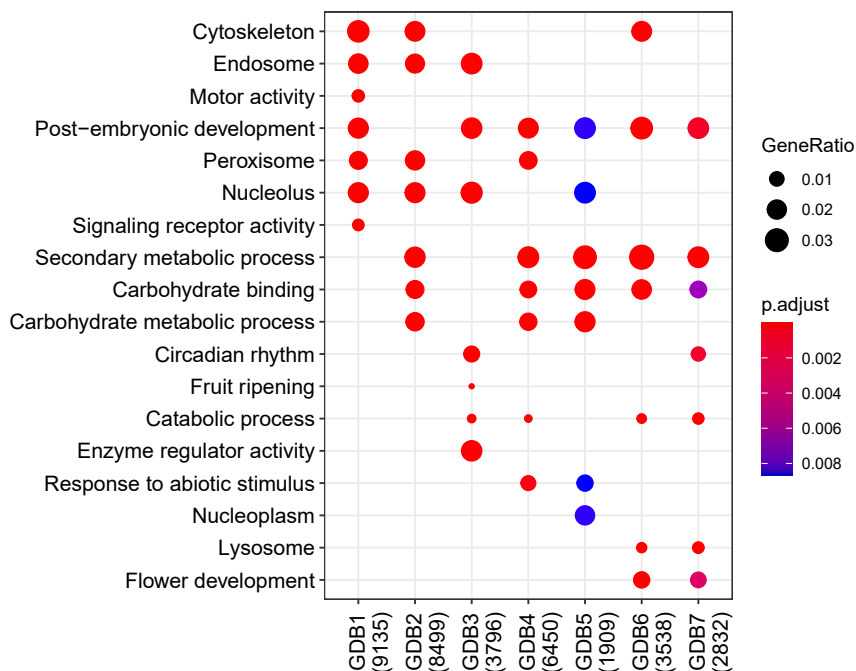


Figure 3. GO enrichment for paralogs of each GDB.

Size of circles indicates the proportion of genes in the GDBs, and color represents the significance level of the enrichment result. The x axis shows the seven detected GDBs in Solanaceae, and the y axis represents the categories obtained from GO enrichment.

duplication. This result is also supported by previous findings of the lack of a *Capsicum*-specific WGD after its divergence from *Solanum* (Qin et al., 2014). For GDB4, we further examined the GDs at the MRCA of *Nicotiana* and found that 85% of the GDs (2529/2961) were supported by duplicates in *N. tabacum*, *N. benthamiana*, *N. tomentosiformis*, and *N. sylvestris*. *N. tabacum* and *N. benthamiana* were reported to be tetraploids (Sierro et al., 2013; Edwards et al., 2017; Schiavinato et al., 2020), and with further support from analyses described in the phylogeny section (Figure 1B), many of the GDs in GDB4 can thus be explained by tetraploidy, especially because the MRCA of these species coincides with the MRCA of the *Nicotiana* species sampled here. In addition, GDB5 in *Joltomata* may have been partially affected by gene topologies with multiple sampled species here, owing to a large number of incomplete lineage sorting and introgression events in *Joltomata* (Wu et al., 2018).

Previous studies support complex polyploid histories of potatoes and their wild relatives (Hawkes, 1990; Hijmans et al., 2007). Our WGD analyses included *S. tuberosum* and *Solanum chacoense*, which were reported as tetraploid and diploid, respectively (Xu et al., 2011; Cai et al., 2012). Although we detected a GD cluster at the node of *S. tuberosum* + *S. chacoense*, this did not meet the criterion for a WGD event (at least three species sharing the GD cluster). Further analyses with more samples of the Petota clade are needed to provide greater support for WGD detection.

In conclusion, the three GDBs in this study are probably WGT/WGDs (indicated by red stars in Figure 2), referred to as WGT1 (GDB1), WGD2 (GDB3), and WGD3 (GDB7). Their average Ks peak values were estimated as 0.65 (WGT1), 0.11 (WGD2), and 0.085 (WGD3), respectively (Figure S25). We further estimated the time of the three WGT/WGDs with the Ks peak values and Ks values of closely related Solanaceae ortholog

metabolic processes, circadian rhythm, fruit development, and resistance mechanisms (peroxisome, response to abiotic stimulus).

Duplication of key genes associated with carpel and fruit development

Polyploidizations/WGDs generate duplicates of all genes, providing great potential for neo/subfunctionalization (Otto, 2007). On the other hand, most return to single-copy genes through loss of one of the duplicates, likely because the two duplicates continue to provide the same function as the ancestral gene and the maintenance of two unnecessarily redundant copies is costly. It is thought that two copies are retained because of selective advantages, including functional divergence of the two copies and stronger functions (such as robustness) from the two copies. Gene losses over time are supported by the observation that older WGDs generally correspond to fewer genes with two retained copies (Ren et al., 2018). Therefore, the retained genes, especially those with special functions (such as fruit/carpel development), are worthy of examination and discussion because they may have undergone functional divergence and contributed to new traits. In the following discussion, when duplication of a specific gene was first detected, it was not necessarily clear whether such a duplication had been part of a WGD. When there is evidence linking the specific gene duplication to a WGD, then it is understood that the GD represents a retained gene pair from the WGD when there is gene phylogeny and/or syntenic support.

Fruits are the harvested organs of several economically important Solanaceae species such as tomato, eggplant, and pepper. Tomato is also a model for studying the molecular mechanisms that regulate carpel and fruit development. Multiple members of the MADS-box gene family and other regulators of carpel development have been identified in previous work (Busi et al., 2003; Victoria et al., 2003; Wang et al., 2019), but when the

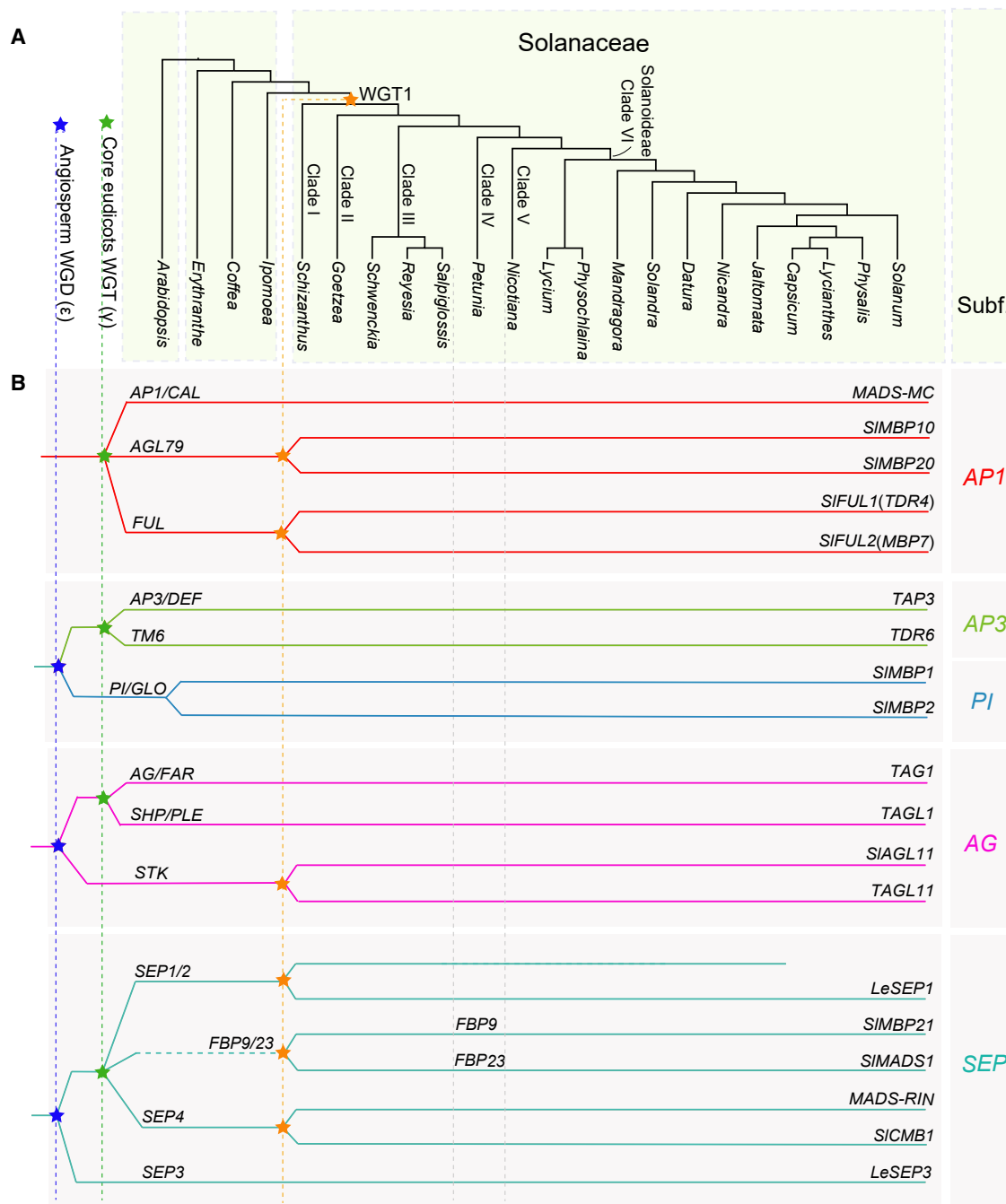


Figure 4. A summary of evolutionary histories of MADS-box genes in Solanaceae.

(A) A simplified phylogeny used in gene family analysis. Stars indicate WGDs/WGTs shared by angiosperms (WGD ϵ , blue), core eudicots (WGT γ , green), and Solanaceae (WGT1, orange). Subf. represents subfamily and indicates the MADS-box subfamily in **(B)**.

(B) The evolution of MADS-box genes in Solanaceae. Gene names are labeled as found in *Arabidopsis* (left) and tomato (right). Gene subfamily names are on the far right. Colored stars indicate GDs from WGDs corresponding to **(A)**. The dashed line of *FBP9/23* indicates gene loss in *Arabidopsis*, and the gene lineage is named as found in *Petunia*.

GDs occurred during Solanaceae evolution is not clear. For example, a GD event was suggested to have occurred within Solanaceae from analyses of a few species (*Capsicum*, *Nicotiana*, *Petunia*, and/or *Solanum*), generating the duplicate copies of *FBP9/23* (Zahn et al., 2005) and/or the *FUL1/FUL2* (Shan et al., 2007) gene lineages. For carpel-related

genes, the duplications in Solanaceae were supported by analyses of only two species (*S. lycopersicum* and *S. tuberosum*) (Pfannebecker et al., 2016, 2017).

To obtain clues about the possible molecular basis of morphological changes in Solanaceae, we examined the evolutionary

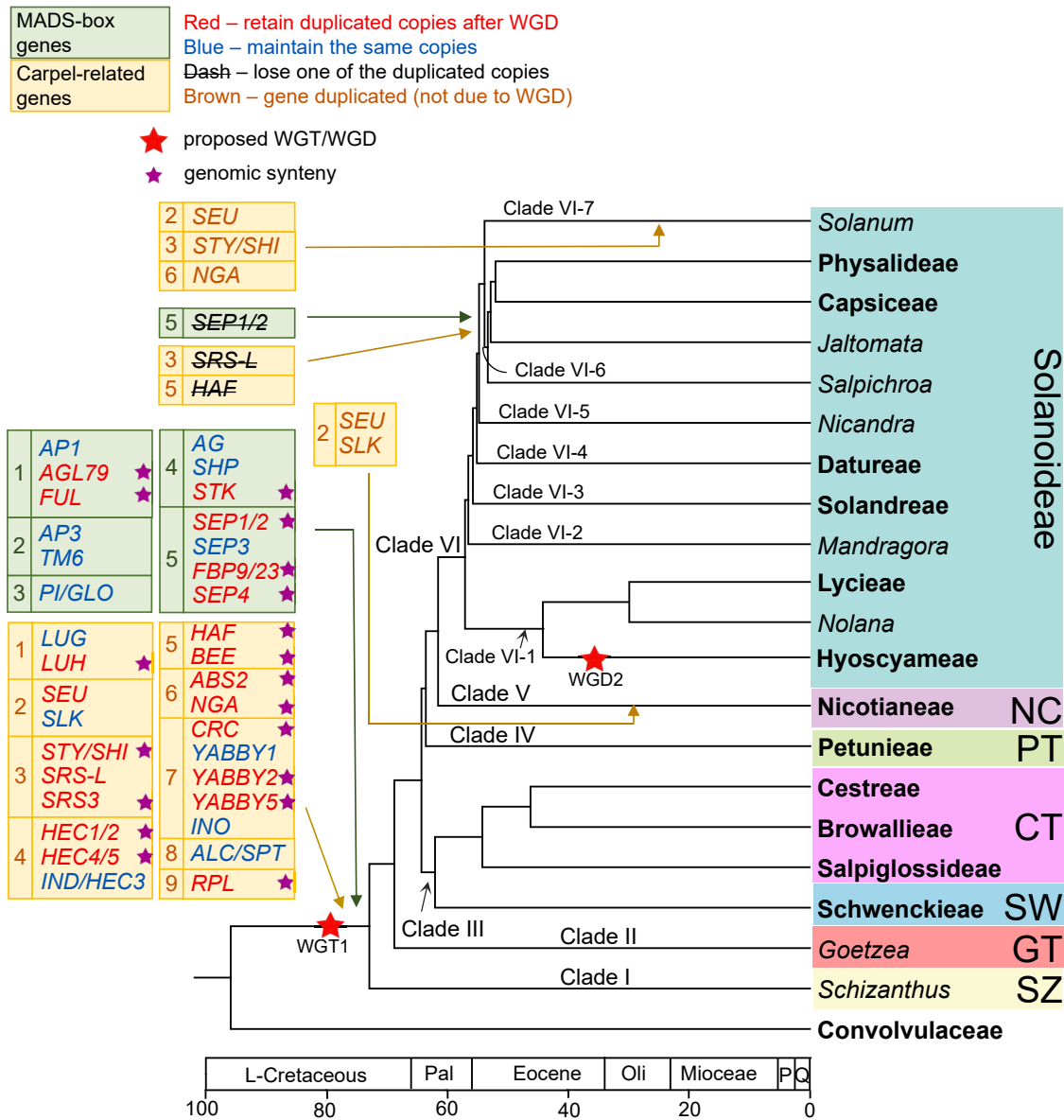


Figure 5. An overview of evolutionary histories of genes related to carpel and fruit development in Solanaceae.

A simplified Solanaceae phylogenetic tree shows retained duplicates or losses after WGDs. Genes in the rectangle with light green backgrounds are MADS-box transcription factors, and those with yellow backgrounds are carpel developmental control genes. Genes from the same gene (sub)family are in a rectangular frame labeled with numbers. Numbers in the MADS-box genes (light green background) indicate these five subfamilies: 1 (*AP1*), 2 (*AP3*), 3 (*PI*), 4 (*AG*), 5 (*SEP*); numbers in the carpel-related genes (yellow background) indicate the following nine gene families: 1 (*LUG*-like), 2 (*SEU*-like), 3 (*STY/SHI/SRS*-like), 4 (*HEC/IND*-like), 5 (*HAF*-like), 6 (*RAV/NGA*-like), 7 (*YABBY*-like), 8 (*ALC*), and 9 (*RPL*). Red letters represent retention of duplicated copies after WGD; letters with a dash indicate retention of duplicated copies after WGD but later loss of one gene copy; blue letters indicate maintenance of the same copy number; and brown letters indicate other gene duplication (not due to WGD). Red stars indicate the WGT/WGD events proposed here. Purple stars represent duplicated copies with collinearity support in model plants (*C. annuum*, *N. tabacum*, *P. axillaris*, *P. inflata*, and/or *S. lycopersicum*). Tip labels of the phylogenetic tree indicate the genus (italic) or tribe (bold). Tribes and subfamilies are highlighted with different colors, with their names shown on the right. SZ, Schizanthoideae; GT, Goetzieoideae; SW, Schwenckioideae; CT, Cestroideae; PT, Petunioideae; NC, Nicotianoideae. Branch lengths of the tree are proportional to ages.

patterns of homologs from 14 gene (sub)families with fruit and carpel development genes (Figures S26–S55, summarized in Figures 4, 5, and S38, and Table S11). These included members of five MADS-box subfamilies, namely *AGAMOUS* (*AG*), *APETALA1* (*AP1*), *APETALA3* (*AP3*), *PISTILLATA* (*PI*), and *SEPALLATA* (*SEP*), and the carpel/fruit regulators *LEUNIG* (*LUG*), *SEUSS* (*SEU*), *SHORT INTERNODE/STYLISH* (*SHI/STY*), *HECATE* (*HEC*), *INDE-*

HISCENT (*IND*), *HALF FILLED* (*HAF*), *NGATHA* (*NGA*), *CRABS CLAW* (*CRC*), *ALCATRAZ* (*ALC*), *SPATULA* (*SPT*), and *REPLUMLESS* (*RPL*) (Seymour et al., 2013; Pfannebecker et al., 2016, 2017). The homologs were retrieved from 41 Solanaceae representatives (including all sampled species from lineages of Schizanthoideae, Goetzieoideae, Schwenckioideae, and Cestroideae) and four outgroups (*Pomoea triloba*, *Coffea*

Plant Communications

canephora, *Erythranthe guttata*, and *A. thaliana*) and included in phylogenetic analyses (Figures S26–S37 and S39–S55; the phylogenetic tree for each gene is shown separately for gene families with multiple genes). We defined gene duplication/retention based on the following criteria: (1) the duplicated/retained gene copies were detected first using published genome sequences of Solanaceae species; (2) the detected gene duplicates were further analyzed using collinearity evidence in at least one of five sequenced genomes (*C. annuum*, *N. tabacum*, *P. axillaris*, *Petunia inflata*, and/or *S. lycopersicum*); and (3) the gene duplicates and their homologs from representative Solanaceae transcriptome datasets were analyzed to determine the phylogenetic positions of duplication/retention using the Solanaceae phylogeny as the reference. Gene loss was detected using sequences from 15 species with published genomes without considering sequences from transcriptome datasets.

Among the MADS-box genes examined here that are important for floral identity and/or fruit development, some (*AP1*, *AP3*, *PI*, *AG*, *SHP*, *SEP3*) remained as single copies during Solanaceae evolution, whereas others (*FUL*, *AGL79*, *STK*, *SEP1/2*, *FBP9/23*, *SEP4*) were duplicated (Figures S26–S37, summarized in Figures 4 and 5). The *FUL* gene of the *AP1* subfamily is necessary for fruit valve differentiation (Ferrándiz et al., 2000), and its tomato homologs (Figure S26, *SIFUL1* and *SIFUL2*) promote fruit ripening (Bemer et al., 2012). The *SHP1/2* and *STK/AGL11* genes in the *AG* subfamily are required for carpel and ovule development (Favaro et al., 2003; Pinyopich et al., 2003). Our results indicate that the *STK* lineage was duplicated in Solanaceae, with *SIAGL11* and *TAGL11* (Figure S29); furthermore, the function of *AGL11* in tomato fruit/seed development is supported by the seedless fruit phenotype that results when *AGL11* expression is suppressed (Ocares and Mejia, 2016). The *SEP1–4* genes confer the E function required for development of multiple flower organs, including carpels (Pelaz et al., 2000; Ditta et al., 2004; Zahn et al., 2006); they underwent duplication in the MRCA of Brassicaceae, generating *SEP1* and *SEP2*; however, the *FBP9/23* lineage was lost in *Arabidopsis* (Malcomber and Kellogg, 2005). The homologs of *SEP1/2*, *FBP9/23*, and *SEP4* underwent duplication in Solanaceae (Figures S32–S34), with subsequent loss of a *SEP1/2* copy probably at the MRCA of *Solanum* and *Jaltomata*. Homologs of *SEP1/2/4* and *FBP9* have been found to promote fleshy fruit development in several plants, including apple (*SEP1/2*) (Ireland et al., 2013), strawberry (*SEP1/2*) (Seymour et al., 2011), and tomato (*SEP1*) (Ampomah-Dwamena et al., 2002). A *SEP4* homologous duplicate in tomato, *RIN*, regulates fruit ripening and has >200 possible direct target genes (Vrebalov et al., 2002; Fujisawa et al., 2011, 2013); the other *SEP4*-like duplicate, *SICMB1*, is important for inflorescence development and sepal size (Zhang et al., 2018).

Among other (non-MADS-box) genes for carpel development, some of their homologs have also been duplicated in Solanaceae (*LUH*, *SEU*, *STY1/SHI*, *SRS-L*, *SRS3*, *HAF*, *BEE*, *ABS2*, *NGA*, *HEC1/2*, *HEC4/5*, *RPL*, *CRC*, *YABBY2*, and *YABBY5*; Figure 5), whereas others have remained as single copies (*LUG*, *SLK*, *IND/HEC3*, *ALC/SPT*, *INO*, and *YABBY1*) (Figures S39–S55, summarized in Figures 5 and S38). *LUG* and its paralogs (*LUHs*) interact with *YABBY* genes and recruit *SEUs* to regulate *AG*

Genome duplications and development in Solanaceae

and fruit development (Franks et al., 2006; Stahle et al., 2009; Shrestha et al., 2014). The tomato *LUG* homologs also regulate tomato fruit development, with duplicate *LUH1* and *LUH2* genes having higher levels of expression than *LUG* during fruit development (Guan et al., 2018). Three retained *SEU* duplicates were detected at the MRCAs of Solanaceae, genus *Nicotiana*, and genus *Solanum*, respectively (Figure S41). The *NGA* gene was reported to have a conserved function in style and stigma specification within eudicots (Fourquin and Ferrandiz, 2014), and the duplication of *NGA* (Figure S48) may have strengthened this function. The *YABBY* gene family contains nine members clustered into five lineages, three of which (*CRC*, *YABBY2*, and *YABBY5*) have duplicate copies in Solanaceae (Figures S38 and S49–S53). *CRC* regulates *Arabidopsis* carpel and nectary development (Alvarez and Smyth, 2002), and the two tomato *CRC* homologs *SICRCa* and *SICRCb* are expressed only in reproductive organs (Huang et al., 2013). One of the two tomato *YABBY2* homologs, *FAS* (*SIYABBY2b*), controls carpel number, thereby affecting tomato fruit shape or size (Cong et al., 2008; Rodriguez et al., 2011); the other *YABBY2* homolog, *SIYABBY2a*, is highly expressed in reproductive organs (Huang et al., 2013). The *RPL* gene is involved in formation of the fruit dehiscence region (between the two valves of the silique, the dry fruit in Brassicaceae), as supported by a study in *Arabidopsis* (Seymour et al., 2013); the duplication of *RPL* in Solanaceae suggests possible functional diversification (Figure S55).

In conclusion, homologs of six gene lineages in MADS-box subfamilies (*FUL*, *AGL79*, *STK*, *SEP1/2*, *SEP4*, and *FBP9/23*) (Figures 4 and 5) and homologs of 15 genes for regulation of carpel development (*LUH*, *SEU*, *STY/SHI*, *SRS-L*, *SRS3*, *HEC1/2*, *HEC4/5*, *BEE*, *HAF*, *ABS2*, *NGA*, *CRC*, *YABBY2*, *YABBY5*, and *RPL*) (Figures 5 and S38) likely underwent Solanaceae-specific GDs (WGT1). The notion that the gene copies are retained duplicates from WGT1 was further supported by the finding that 19 of the 21 duplicated gene pairs are located in syntenic genomic regions in at least one of the examined genomes (*C. annuum*, *N. tabacum*, *P. axillaris*, *P. inflata*, and/or *S. lycopersicum*). The loss of three genes (*SEP1/2*, *SRS-L*, *HAF*) after duplication was also observed. According to their gene family trees (Figures S32, S43, and S46), we found that six species with sequenced genomes (*Nicotiana* spp. and *Petunia* spp.) each retained the two copies, whereas the other nine species with sequenced genomes (*Solanum* spp., *Jaltomata* spp., and *Capsicum* spp.) had each lost one of the copies. We conservatively estimated that this gene loss possibly occurred at the MRCA of *Solanum* and *Jaltomata* (Figures 4, 5, and S38).

Comparison of the Solanaceae phylogeny with previous studies

Considering the highly resolved Solanaceae relationships in the context of extensive previous studies, it would be beneficial to compare some key findings with relevant reported relationships. Clade I and II: the genus *Schizanthus* (Clade I) and subfamily Goetzeoideae (Clade II) are two successive branching clades sister to other Solanaceae. The position of *Schizanthus* supports the proposed subfamily Schizanthoideae (Olmstead et al., 1999, 2008). However, another study using five plastid regions and two nuclear sequences with >1000 Solanaceae species (Särkinen

et al., 2013) did not resolve the relationships of *Schizanthus* with three other lineages (including the subfamily Goetzeoideae and the two genera *Duckeodendron* and *Reyesia*), which together are outside a clade of the remaining Solanaceae. *Duckeodendron* was not sampled here and was placed as a separate lineage and potentially as one of the sisters to the rest of Solanaceae in several studies (Olmstead et al., 2008; Särkinen et al., 2013; Ng and Smith, 2016). Here *Reyesia* is strongly supported as part of the tribe Salpiglossideae of Cestroideae (see below), as proposed previously (Hunziker, 2001; Dillon, 2005). The uncertain placements of Goetzeoideae and *Duckeodendron* relative to most other Solanaceae were also reported by Olmstead et al. (2008); thus, further analyses that include *Duckeodendron* and species of tribe Benthamiellae are needed to resolve their relationships.

Clade III (Cestroideae + Schwenckieae): in Clade III, the sister relationship between the subfamily Cestroideae and the tribe Schwenckieae supports the previous designation of the latter as the subfamily Schwenckioideae (Olmstead et al., 1999); however, previous phylogenies did not group Schwenckieae with Cestroideae (Olmstead et al., 2008; Särkinen et al., 2013). The topology for Cestroideae (Salpiglossideae [Cestreae, Browallieae]) is the same as that reported previously (Olmstead et al., 2008), although *Reyesia* was not sampled. The sister relationship of *Salpiglossis* and *Reyesia* is highly supported here (100/0.97/1/0.84) and was also found by Ng and Smith (2016) (BS = 74%), supporting the placement of *Salpiglossis* and *Reyesia* in the same tribe (Salpiglossideae); this relationship is further supported by their morphological similarities, as both are herbs with zygomorphic flowers and capsules (Hunziker, 2001). By contrast, Goetzeoideae and *Duckeodendron* are trees or shrubs with actinomorphic flowers and non-capsular fruits (Barboza et al., 2016). The tribe Browallieae (including *Browallia* and *Streptosolen*) is consistently monophyletic (100/0.97/1/0.87), as proposed by Hunziker (2001) and strongly supported previously (Olmstead et al., 2008).

Clade IV (Petunieae) and Clade V (Nicotianoideae): the separation of Petunieae from other clades provides strong support for recognition of the subfamily Petunioideae (Olmstead et al., 1999; Martins and Barkman, 2005), unlike its designation as a tribe in Barboza et al. (2016). Within Petunioideae/Petunioideae, *Nierembergia* and *Brunfelsia* occupy two successive sister branches of a previously found clade ([*Calibrachoa*, *Fabiana*], *Petunia*) (Olmstead et al., 2008; Alaria et al., 2022; Wheeler et al., 2022). Clade V with two tribes (Nicotianeae and Anthocercideae) is consistent with previous analyses (Olmstead et al., 2008).

Clade VI (Solanoideae, with seven subclades [Clade VI-1 to Clade VI-7]). Clade VI-1: species in this clade were informally named Atropina (BS = 78%) with unresolved relationships among *Jaborosa*, *Latua*, Hyoscyameae, and a clade containing Lycieae, *Nolana*, and *Sclerophylax* (Olmstead et al., 2008). The positions of *Latua* and *Jaborosa* here (92–99/0.4/1/0.26) outside a clade with both Hyoscyameae and Lycieae support the designation of these genera as tribes Latueae and Jaboroseae, respectively (Olmstead et al., 1999; Finot et al., 2018). In Hyoscyameae (100/0.88/1/0.67), the highly supported relationships among the five genera sampled here are generally consistent with those reported previously (Olmstead et al., 2008). Our results also

support *Nolana* and *Sclerophylax* as successive sisters (98–99/0.87/1/0.66 and 64–73/0.49/0.7/0.59, respectively) of Lycieae (100/0.9/1/0.6), in agreement with these three groups being in a clade previously (BS = 86%; Olmstead et al., 2008). Unlike Lycieae, which have berries, *Nolana* species have unusual fruits called mericarps (schizocarps), and *Sclerophylax* also has an atypical fruit morphology for Solanaceae (Olmstead et al., 2008; Barboza et al., 2016). *Nolana* and *Sclerophylax* were placed in the tribe Nolaneae (Olmstead et al., 1999; Martins and Barkman, 2005; Dillon et al., 2007), but the phylogeny here suggests that the proposed tribe Nolaneae would be paraphyletic (marked as Nolaneae I and II, respectively, in Figure 1) and should include the tribe Lycieae (Dillon et al., 2007).

Clade VI-2 (*Mandragora*): previous phylogenetic analyses using morphological and cytological characters placed *Mandragora* either in or outside of Hyoscyameae (Hoare and Knapp, 1997; Tu et al., 2005); however, analyses using plastid sequences supported a closer relationship of *Mandragora* with Solaneae and Physalideae (Tu et al., 2010; Volis et al., 2018). Our results in which *Mandragora* is separate from both Hyoscyameae and Solaneae + Physalideae support the previously proposed tribe Mandragoreae (Olmstead et al., 1999; Tu et al., 2010).

Clade VI-4 (Datoreae) and Clade VI-5 (*Nicandra*): our results support *Nicandra* as the sister (93–99/0.38/1/–0.16) to the clade containing Physalideae and Solaneae (see below for Clade VI-6 and Clade VI-7) but with a negative ICA score (Table S4). The phylogenetic position and relationship of Datureae and *Nicandra* have been uncertain in previous studies; some placed Datureae as sister to a clade containing Solaneae, Capsiceae, and Physalideae (BS = 90%, Olmstead et al., 2008; BS < 75%, Ng and Smith, 2016), but others suggested that the sister group of Datureae was *Nicandra* (BS < 80%, Särkinen et al., 2013; Dupin et al., 2017; posterior probabilities [PP] = 0.97, Dupin and Smith, 2018).

Clade VI-6 (including *Salpichroa*, *Jaltomata*, Capsiceae and Physalideae) and Clade VI-7 (*Solanum*): *Salpichroa* was once assigned to the tribe Physalideae (subtribe Salpichroinae), but with low support (BS = 45%; Olmstead et al., 1999). In our results, *Salpichroa* did not cluster with other members of Physalideae (Figure 1); thus, we suggest that *Salpichroa* should be excluded from the tribe Physalideae. Previously, *Salpichroa* and the monotypic *Nectouxia* (not sampled here) formed a well-supported clade (Salpichroina) with uncertain relationships to Capsiceae and Physalideae (Olmstead et al., 2008). *Jaltomata* is monophyletic (100/0.97/1/0.78, marked as Solaneae II in Figure 1) and constitutes the second divergent branch in Clade VI-6, whereas *Jaltomata* was recognized as the sister group of *Solanum* (Clade VI-7 here) in previous studies (Olmstead et al., 2008; Särkinen et al., 2013; Ng and Smith, 2016). To gain further understanding of the controversial phylogenetic position of *Jaltomata*, we examined the 1699 gene trees and identified those with different placements of *Jaltomata* (topology 1, as sister to Physalideae + Capsiceae; topology 2, as sister to *Solanum*). Topology 1 was found in 149 (8.8%) gene trees with an ICA value of 0.12, whereas topology 2 was found in 74 (4.4%) gene trees with an ICA value of –0.17 (Figure S14). In another analysis, we counted the number of sister clades of *Jaltomata* in 6326 gene trees (allowing multiple homologs in one species) inferred from 11 representative species (see

Plant Communications

section “methods”). As a result (Figure S15), the sister clade of *Jaltomata* was detected as Physalideae and/or Capsiceae in 2458 (38.9%) trees, whereas 1477 (23.3%) trees supported an alternative relationship in which *Jaltomata* was sister to *Solanum*. A recent analysis of the *J. sinuosa* genome also supported a closer relationship with *Capsicum* than *Solanum* (Wu et al., 2019) and polyphyly of the previously defined Solaneae (indicated in Figure 1 as Solaneae I and II). These various results suggest that there may have been hybridization in the ancestor of *Jaltomata*, involving possible parental lineages related to Physalideae/Capsiceae and *Solanum*, with fewer genes from the latter retained in extant *Jaltomata*.

In the tribe Physalideae, lochrominae is represented here by three genera. Withaninae is represented by *Athenaea*, *Archiphysalis*, *Tubocapsicum*, and *Withania*. However, our phylogeny placed *Athenaea* as sister to a clade consisting of the subtribe Physalidinae and other members of Withaninae, causing Withaninae to be polyphyletic. Physalidinae is monophyletic and represented here by five genera (*Alkekengi*, *Physalis*, *Leucophysalis*, *Witheringia*, and *Physalistrum*). *Physalis* includes >100 species (e.g., Chinese lantern [*P. alkekengi*], tomatillo, and various groundcherries [*Physalis angulata* and *Physalis peruviana*]) that are characterized by a persistent and enlarged calyx outside the mature fruit (Whitson and Manos, 2005; Barboza et al., 2016; Deanna et al., 2019). *Archiphysalis* and *Physalistrum* have both been recognized as genera synonyms of *Withania* in Barboza et al. (2016). However, our results clearly support *Archiphysalis* and *Physalistrum* as different genera from *Withania*, and they should be placed in Withaninae and Physalidinae, respectively. The same relationships have also been proposed by Li et al. (2013).

A WGT at the MRCA of Solanaceae during the late Cretaceous

Evidence for past polyploidization has been detected in several Solanaceae genomes, such as potato for WGD (Xu et al., 2011) and tomato, pepper, and tobacco for WGT (Sato et al., 2012; Qin et al., 2014; Xu et al., 2017). The possibility of shared WGT(s) in Solanaceae has also been suggested from analyses of fewer than five species (Blanc and Wolfe, 2004; Ren et al., 2018). A comparative analysis of the *Petunia* genome with those of other Solanaceae supported a polyploidization (probably triplication) event shared by *Petunia*, *Solanum*, *Capsicum*, and *Nicotiana* (Bombarely et al., 2016), suggesting that this WGT could have occurred earlier than the MRCA of Petunioideae and Solanoideae. Our analyses here with a large number of Solanaceae species, including representatives of Schizanthoideae, Goeteoideae, and Cestroideae, allowed more precise placement of the WGT event at the MRCA of Solanaceae (WGT1; Figure 2). The estimated age of WGT1 is ~81 mya (Figure 2), older than the age (71 mya), but within the CI (71 [±19.4] mya), estimated from the synonymous rate (Ks) distribution of relevant paralogs in tomato, as well as the age (67 mya) estimated from 4DTv (4-fold degenerate sites) of all duplicate pairs in potato (Xu et al., 2011; Sato et al., 2012).

Evolution of carpel and fruit developmental genes in Solanaceae

Changes in fruit type from a dry capsule to a fleshy berry appear to have happened three separate times in Solanaceae (Knapp,

Genome duplications and development in Solanaceae

2002a). Our analysis indicated that several genes related to carpel and fruit development were duplicated at the MRCA of Solanaceae (Figures 4, 5, and S26–S55), including *AGL79*, *FUL*, *STK*, *SEP1/2*, *FBP9/23*, *SEP4*, *LUH*, *SEU*, *STY/SHI*, *SRS-L*, *SRS3*, *HEC1/2*, *HEC4/5*, *HAF*, *BEE*, *ABS2*, *NGA*, *CRC*, *YABBY2*, *YABBY5*, and *RPL*. It is possible that, in early Solanaceae, the recently duplicated genes had not accumulated enough mutations to cause functional differentiation, as suggested previously (Knapp, 2002a; Wang et al., 2015). Both dry and fleshy fruits have evolved independently multiple times throughout angiosperm lineages, suggesting that the conversion between dry and fleshy fruit types may not require complex genetic changes (Berner et al., 2012).

Alternatively, it is also possible that fleshy fruit may not have conferred a significant advantage when duplicate copies were generated at the MRCA of Solanaceae; for example, fleshy-fruit-eating frugivores may not have been present near the early Solanaceae. In addition, Solanaceae species produce alkaloids that can lead to poisoning and indigestion in frugivores; such alkaloid effects could have reduced the evolutionary advantages of fleshy fruits in early Solanaceae history (Milner et al., 2011; Chowanski et al., 2016). It may have taken a period of time after the WGT1 for Solanaceae plants to evolve additional regulation of alkaloid production, enabling the reduction of alkaloids in mature fruit while maintaining high levels in immature fruits (McKey, 1979; Herrera, 1982). For example, the high content of glycoalkaloids (e.g., R-tomatine 3) in tomatoes is generally associated with immature fruits, and the level of R-tomatine 3 is reduced upon fruit maturation (Milner et al., 2011); in addition, α -tomatine, the primary alkaloid in tomato fruit, is converted to esculosides and lycopersides during fruit development and ripening (Szymanski et al., 2020). Frugivores have also evolved a sensory system to identify mature fruit through color vision, olfaction, and other processes (Valenta et al., 2013). These changes might indicate co-evolution of Solanaceae (especially Solanoideae) species and frugivores, providing another excellent example of the co-evolution of fruiting plants and frugivores, further promoting seed dispersal, adaptation, and occupation of new ecological niches by Solanaceae species.

METHODS

Taxa sampling and data collection

A total of 180 Solanaceae species were included in our study, representing all four subfamilies, 14 out of 15 tribes, and 51 genera of Solanaceae (Table S1), classified according to Barboza et al. (2016). In particular, among the 81 *Solanum* species sampled here, 22 *Petota* and 13 *Tomato* members (Tang et al., 2022; Zhou et al., 2022) were included to explore the relationships of potato (*S. tuberosum*) and tomato (*S. lycopersicum*) with their respective closely related species. Two *Convolvulaceae* (*Convolvulus arvensis* and *Dichondra repens*) species were included as outgroups to reconstruct the phylogeny of Solanaceae. Further analyses contained 43 additional outgroups, 38 of which were used to constrain fossil calibrations in divergence time estimation. In addition, three genomes (*Ipomoea nil*, *Ipomoea trifida*, and *E. guttata*) were included for genome duplication analyses, and four datasets (*A. thaliana*, *I. triloba*, *C. canephora* and *E. guttata*) were included for phylogenetic analyses of gene families (see Table S1). Names of Solanaceae subfamilies/tribes/genera followed the classification in Barboza et al. (2016).

Plant materials were obtained from several sources: field investigations, living plants from botanic gardens, seedlings from purchased seeds, and herbarium vouchers. Fresh plant materials (mainly leaves and buds) were preserved to isolate total RNA. Herbarium vouchers were used to extract genomic DNA following a modified 2× CTAB protocol (Doyle and Doyle, 1987). RNA and DNA were sequenced on the Illumina HiSeq 3000 platform, and the sequencing data were subsequently assembled with Trinity (Grabherr et al., 2011) and SOAPdenovo2 (Luo et al., 2012), respectively. The coding sequences of each gene/transcript were identified using TransDecoder v3.0 (Haas et al., 2013) with default settings, and redundant contigs were removed with CD-HIT (Fu et al., 2012) using an identity threshold of 0.98. Publicly available Sequence Read Archive (SRA) data were downloaded from GenBank and processed using the same methods described above. Coding sequences derived from publicly available genomes were downloaded from Phytozome (<https://phytozome.jgi.doe.gov/>), the Sol Genomics Network (<http://solgenomics.net/>), and other completely sequenced genomes (Bombarely et al., 2016; Kim et al., 2017; Wu et al., 2019).

Selection of candidate OGs and phylogenomic reconstruction

Sequence datasets from nine species of Solanaceae (*Schizanthus pinnatus*, *Cestrum parqui*, *P. inflata*, *N. attenuata*, *Nolana crassulifolia*, *Datura wrightii*, *Physalis longifolia* var. *subglabrata*, *S. lycopersicum*, and *Solanum betaceum*) were used to infer low-copy orthologous genes with OrthoMCL v1.4 (Li et al., 2003). The resulting 1699 low-copy OGs were then used as queries to search for homologous sequences in the 180 Solanaceae species and all outgroups using HaMStR (Ebersberger et al., 2009) with an e-value cutoff of 10^{-20} . Protein sequences of each OG were aligned with MUSCLE (Edgar, 2004) and used to generate the corresponding nucleotide alignment with PAL2NAL (Suyama et al., 2006). Poorly aligned regions in nucleotide alignments were trimmed (-automated1) using trimAl v1.4 (Capella-Gutierrez et al., 2009).

Because missing data, short sequences, insufficiently informative sites, and other factors may result in biased phylogenetic inference, we selected five subsets of the 1699 OGs by successively reducing the number of genes according to the following criteria (Figure S2): (1) genes with alignment length ≥ 600 bp, resulting in 1263 OGs; (2) species coverage $\geq 90\%$, leading to 1034 OGs; (3) taxon coverage including representative species from each tribe/subtribe, leading to retention of 857 OGs; (4) an average BS $\geq 65\%$ for each single gene tree, resulting in 589 OGs; (5) genes with alignment length ≥ 900 bp, leading to 419 OGs. These six gene sets (1699, 1263, 1034, 857, 589, and 419 OGs) were used for phylogenetic analyses. The nucleotide alignment of each OG was reconstructed by RAXML (Stamatakis, 2006) with 100 replicates under the GTRGAMMAI model. Then, ASTRAL v5.6 (Mirarab et al., 2014) was used to reconstruct the Solanaceae phylogeny for each of the six OG sets with the 100 replicates from RAXML to obtain the BS values of all nodes.

Assessment of phylogenetic conflicts

To investigate phylogenetic conflicts, each of the 1699 nuclear gene trees was rooted with outgroups. Then, each tree with BS support $\geq 0\%$, $\geq 50\%$, and $\geq 70\%$ for the corresponding node was mapped against the species tree using PhyParts (Smith et al., 2015) to calculate the numbers of concordant bipartitions and the ICA values. For relatively reliable results, we mentioned in the main text only the numbers of concordant trees with $\geq 50\%$ BS and ICA values from the node of trees with BS $\geq 50\%$. In addition, the QS and LPPs of different topologies (the main and the other two alternatives) were estimated with ASTRAL-Pro (Zhang et al., 2020a).

Analyses of gene tree topologies for placement of *Jaltomata*

To further investigate the phylogenetic position of *Jaltomata*, nine species of *Jaltomata* and related genera in Clade VI-6 and VI-7 (*J. sinuosa*, *Jaltomata repandidentata*, *S. lycopersicum*, *Solanum pimpinellifolium*, *P. peruviana*, *Lycianthes rantonnetii*, *C. annum*, *Lochroma cyaneum*, and *Sal-*

pichroa organiflora) and two outgroups (*Lycium barbarum* and *P. axillaris*) were used to detect the sister clades of *Jaltomata* (Figure S15). An all-by-all BLASTP of coding sequences from each species was performed using OrthoFinder (Emms and Kelly, 2019) with default settings to infer closely related homologs within an orthogroup. Orthogroups containing 1–4 sequences from each of the 11 taxa were further aligned in MAFFT (Katoh and Standley, 2013) and used to infer a gene tree with FastTree (Price et al., 2009). The resulting 6326 gene trees were then mapped onto the simplified species tree (Figure S15A) to detect the sister clades of *Jaltomata* with Tree2GD (<https://github.com/Dee-chen/Tree2gd>).

Divergence time estimation

The molecular clock analyses included 10 fossils for calibration, two of which were reported as within Solanaceae (see Table S5). Fossil 1 was a macrofossil (seed) identified as *Datura* cf. *stramonium* (Velichkevich and Zastawniak, 2003), and fossil 2 has lantern fruits and was recently described as a *Physalis* species with an estimated minimum age of 52.2 million years (Wilf et al., 2017; Deanna et al., 2020). Owing to differences in the suggested placements of these two fossils (fossil 1, stem or crown group of *Datura*; fossil 2, stem group of tribe Physalideae or subfamily Solanoideae) (Wilf et al., 2017; Dupin and Smith, 2018; Deanna et al., 2020), we used four calibration strategies (Figures S16–S20; Tables S5 and S6) for divergence time estimation: (1) calibration 1, with fossil 1 assigned to the stem group of *Datura* and fossil 2 assigned to the stem group of Physalideae; (2) calibration 2, with fossil 1 assigned to the crown group of *Datura* and fossil 2 assigned to the stem group of Solanoideae; (3) calibration 3, with fossil 1 assigned to the stem group of *Datura* and fossil 2 assigned to the stem group of Solanoideae; and (4) calibration 4, with fossil 1 assigned to the crown group of *Datura* and fossil 2 assigned to the stem group of Physalideae. Assignments and ages of the remaining eight fossils were in accordance with Smith et al. (2010) and Magallón et al. (2015).

Two methods, treePL (Smith and O'Meara, 2012) and BEAST (Drummond et al., 2012), were used to estimate divergence times (Figures S16–S20; Tables S5 and S6). The 419-gene coalescent tree with 38 more outgroups was used for the analysis; the topology was generally consistent with our Solanaceae phylogeny, and the relationships among outgroups were also consistent with previous studies (Huang et al., 2016a; Chase et al., 2016). The concatenation of the alignments of the 419 OGs was used for branch length calculation with IQ-TREE (Nguyen et al., 2014), using the 419-gene coalescent tree as the constraint tree. For treePL, the cross-validation tested a set of smoothing parameters for each calibration, and the lowest chi-square value in the resulting cvoutfile was suggested as the best smoothing value for subsequent analyses. Optimal smoothing values of 0.1, 0.1, 1000, and 0.1 were selected for our four calibration strategies (calibration 1 to calibration 4, respectively). Default settings were used for all other parameters in treePL analyses. CIs were summarized from 1000 RAXML BS trees using TreeAnnotator (BEAST package). For the BEAST analysis (Drummond et al., 2012), considering the computation time, we selected the top 30 (Table S7) of the 419 OGs suggested by clock-likeness methods (Smith et al., 2018) as input sequences. We used the GTR+G model for site substitution, the birth–death model for the tree prior, and the uncorrelated relaxed lognormal model for the clock prior. For all calibrations, we used a lognormal prior distribution with a mean of 0 and a standard deviation of 0.5. The offset value (age) for each calibration is listed in Table S5. We fixed the tree topology using our coalescent tree of 419 OGs and performed two independent runs with 30 million generations. The effective sample size was evaluated using Tracer v1.7 (Rambaut et al., 2018), and independent runs with the same settings (calibration sets) were combined using LogCombiner (Drummond et al., 2012; Bouckaert et al., 2014). A chronogram with mean ages and 95% highest posterior distribution of each node was generated using TreeAnnotator (Bouckaert et al., 2014) with the first 20% of trees discarded as burn-in.

Plant Communications

Candidate WGD detection

Allmost all Solanaceae species (except five poor-quality genomes sequenced from herbarium vouchers) and three outgroup species with sequenced genomes (*I. trifida*, *I. nil*, and *E. guttata*) were used for genome duplication analyses (Table S1). For coding sequences of each species, an all-by-all BLASTP was performed in DIAMOND (Buchfink et al., 2015) with an e-value cutoff of 10^{-5} to find homologous sequences (query cover = 50%, subject cover = 50%). To generate homolog trees using the putatively homologous sequences, we followed a pipeline modified from Yang et al. (2015). We performed MCL (Enright et al., 2002) clustering by setting a hit-fraction cutoff of 0.4, an inflation value of 1.4, and a minimum log-transformed e-value of 10^{-20} . The resulting 23 243 gene clusters were then used to create an alignment matrix in MAFFT (Kato and Standley, 2013); the alignment was trimmed with trimAl v1.4 (Capella-Gutierrez et al., 2009) and used to infer a tree in FastTree (Price et al., 2009). Putative paralogous sequences were removed from the alignment matrix as described in Yang et al. (2015), using 0.3 and 0.5 as relative and absolute cutoffs for trimming tips. Subsequently, the procedures were repeated, and trees were inferred with IQ-TREE (Nguyen et al., 2014). The gene trees were inferred from nucleotide alignments that were generated according to the corresponding aligned and trimmed protein sequences. Alignments with fewer than 15 taxa and/or shorter than 300 bp were discarded, leading to retention of 18 204 gene clusters. The gene trees were mapped onto the species tree to locate putative duplication events using Tree2GD (<https://github.com/Dee-chen/Tree2gd>), with duplicate clades having $\geq 80\%$ BS. The results of all nodes with values of GD ≥ 300 , the ratio of gene duplication and gene family (GD ratio) $\geq 7\%$, and the percentage of GD with both lineages A and B (ABAB type of GD) $\geq 50\%$ were proposed.

Putative duplication events were then filtered using the rePhylo R package (<https://github.com/Chien-Hsun/rePhylo/>) with a minimum requirement of 30% species coverage in each of the duplicated subclades (filtering results for the candidate duplication events are presented in Table S8). The resulting seven GDBs were further tested using the MAPS pipeline (Li et al., 2018; Leebens-Mack et al., 2019), and five species with ladder relationships (seven species for GDB1 and GDB2) were selected for each GDB in MAPS analysis. Two-hundred resampled sets of null simulations for each group were performed with background gene birth and death rates estimated in CAFE (De Bie et al., 2006). The gene birth rate (λ) and gene death rate (μ) for each group were estimated with CAFE and are shown in Table S9. The support for WGD at a specific GDB node was then assessed by comparing the percentage of subtrees from observed data and null simulations. The GDBs shared by representative genomes (*S. lycopersicum*, *S. tuberosum*, *C. annuum*, *N. attenuata*, *J. sinuosa*, and *P. axillaris*) were further examined with synteny analyses using MCScanX (Wang et al., 2012) (Table S10). The synonymous substitution rate (Ks) of each gene pair was estimated in PAML (Yang, 2007) with the YN method. Ks peaks were identified from histograms of the Ks distribution values of the corresponding GDBs in R.

Estimation of candidate WGD ages

The occurrence times of WGDs were estimated based on the assumption that synonymous mutations accumulated at a constant rate. Each WGD was located between two adjacent species divergence events in the species tree; the species divergence time before the WGD was taken as the upper limit of the WGD age (indicated as T^{prior}), and the divergence time after the WGD was considered to be its lower limit (indicated as T^{post}). The divergence time for each lineage on the species tree was obtained from our chronogram tree in the section “divergence time estimation.” The age of WGD (T^{WGD}) was calculated according to the following formula modified from Ren et al. (2018):

$$T^{\text{WGD}} = T^{\text{prior}} - \frac{Ks^{\text{prior}} - Ks^{\text{WGD}}}{Ks^{\text{prior}} - Ks^{\text{post}}} \times (T^{\text{prior}} - T^{\text{post}})$$

Genome duplications and development in Solanaceae

Ks^{prior} and Ks^{post} represent the Ks peaks of orthologous genes from different lineage species in the divergence events before and after the WGD, respectively. Orthologous genes for Ks calculations were obtained by all-against-all BLAST to identify the best-matched genes of the paired species. Ks^{WGD} represented the Ks peak for gene pairs that were derived from the corresponding WGD.

Identification of tandem repeats and GO enrichment analysis of duplicates

The number and proportion of duplicate genes contributed by tandem repeats at these nodes were also determined. For species with whole-genome sequencing data, we retrieved chromosomal positions of the duplicated gene pairs. If the two genes were separated by 10 or fewer genes on the same chromosome, they were defined as tandemly duplicated genes from the same gene cluster in this analysis. Functional annotations of replicated genes were searched for homologs in *A. thaliana*, and enrichment of gene sets was performed in R with clusterProfiler packages (Yu et al., 2012).

Phylogenetic analyses of gene families

A total of 41 representative Solanaceae species and four outgroups (*I. triloba*, *C. canephora*, *E. guttata*, and *A. thaliana*) were included in the phylogenetic analyses of gene families (Table S1). The selected Solanaceae species included 15 with publicly available sequenced genomes (*P. axillaris*, *P. inflata*, *N. attenuata*, *N. obtusifolia*, *N. tabacum*, *N. sylvestris*, *C. annuum*, *C. baccatum*, *C. chinense*, *J. sinuosa*, *Solanum chilense*, *S. pennellii*, *S. pimpinellifolium*, *S. lycopersicum*, and *S. tuberosum*). In addition, 26 representative transcriptomes were selected from different clades (especially the early-branching clades) with relatively high quality for their respective clades (average gene number, 45 005; average N50, 1028). These transcriptome datasets were from Schizanthoideae (*Schizanthus litoralis* and *S. pinnatus*), Goetzeoideae (*Goetzea ekmanii*), Schwenckioideae (*Schwenckia americana*), Cestroideae (*Browallia americana*, *Streptosolen jamesonii*, *Cestrum diurnum*, *C. parqui*, *Vestia foetida*, *Reyesia chilensis*, *R. juniperoides*, *Salpiglossis sinuata*), Petunieae (*Nierembergia linariifolia* var. *linariifolia*, *Brunfelsia americana*, *B. grandiflora*, *Fabiana imbricata*), Hyoscyameae (*Physochlaina praealta*), Nolaneae (*N. crassulifolia*, *N. paradoxa*), Lyceae (*L. barbarum*), Mandragoreae (*Mandragora caulescens*), Solandreae (*Solandra maxima*), Datureae (*Datura stramonium*), Nicandreae (*Nicandra physalodes*), Capsiceae (*L. rantonnetii*), and Physalideae (*Physalis angulata*).

Protein homologs of query sequences of the genes were used for BLASTP with an e-value cutoff of 10^{-10} . MADS-box homologs were also retrieved with HMMER (Mistry et al., 2013) using the SRF-TF entry (pfam00319) from Pfam (El-Gebali et al., 2019). Protein sequences for each gene set were aligned in MAFFT (Kato and Standley, 2013), and poorly aligned regions were trimmed using trimAl v1.4 (Capella-Gutierrez et al., 2009); the trimmed alignments were then used for gene tree reconstruction with FastTree (Price et al., 2009). The subclade containing the query sequence was extracted, and nucleotide alignments were generated according to the realigned, trimmed, and manually examined protein alignments. Gene trees of each subclade were then inferred with nucleotide alignments using IQ-TREE (Nguyen et al., 2014) under the tested optimal model. Incomplete or suspicious MADS-box family sequences obtained from public genomes were examined and re-annotated using the genomic sequences and their location information with exonerate (Slater and Birney, 2005). The corrected sequences were given an “r” suffix after the original sequence IDs. The duplicated gene pairs of five model plants (*C. annuum*, *N. tabacum*, *P. axillaris*, *P. inflata*, and/or *S. lycopersicum*) were further examined to detect collinearity support. The reference gene IDs of *A. thaliana* and *S. lycopersicum* used in phylogenetic analyses are presented in Table S11.

DATA AVAILABILITY

Raw reads generated in this study were deposited in the NCBI (<https://www.ncbi.nlm.nih.gov>) SRA database under bioproject PRJNA827705 and in the National Genomics Data Center (<https://ngdc.cnbc.ac.cn>) as GSA of CRA010127.

SUPPLEMENTAL INFORMATION

Supplemental information is available at *Plant Communications Online*.

FUNDING

This work was supported by funds from the National Natural Science Foundation of China (grant nos. 31770242, 31970224, and 32270232), the Key Laboratory of Biodiversity Science and Ecological Engineering and State Key Laboratory of Genetic Engineering at Fudan University, and Eberly College of Science and the Huck Institutes of the Life Sciences at the Pennsylvania State University.

AUTHOR CONTRIBUTIONS

H.M. and C.-H.H. designed this study. J.H., H.M., J.Z., W.X., C.M., and C.Z. collected plant materials. J.H. and Y.H. extracted RNA and/or DNA. J.H. performed all the analyses with the help of C.-H.H., J.G., L.Z., and Y.Z. J.H., C.-H.H., and H.M. wrote the manuscript. All authors read and approved the final manuscript.

ACKNOWLEDGMENTS

We thank Prof. Gang Wang and Drs. Xinxin Zhu, Jing Liu, Zhaoen Lu, Holly Forbes, Clare Loughran, Peter Brownless, Suzanne Cubey, Andrew Stephenson, Andrew Wyatt, Jocelyn Frazelle, Haldre Rogers, Brittany Cavazos, Pablo Bolaños-Villegas, and Sergio Castro-Pacheco for assistance with plant sampling. We also thank Profs. Ji Yang, Wenju Zhang, Zhiping Song, Yuguo Wang, Ji Qi, and Qiang Zhang for discussion. We thank Xinxin Zhu for providing beautiful plant photos. We thank Drs. Taikui Zhang and Duoyuan Chen for their technical assistance in the analyses. We thank Missouri Botanical Garden, Royal Botanic Garden Edinburgh, the New York Botanical Garden, University of California Botanical Garden at Berkeley, the Royal Botanic Gardens Kew, the Germplasm Bank of Wild Species at Kunming Institute of Botany, and the Herbarium of Institute of Botany, Xishuangbanna Tropical Botanical Garden, for providing materials. No conflict of interest is declared.

Received: September 27, 2022

Revised: March 2, 2023

Accepted: March 22, 2023

Published: March 25, 2023

REFERENCES

- Alaria, A., Chau, J.H., Olmstead, R.G., and Peralta, I.E. (2022). Relationships among *Calibrachoa*, *Fabiana* and *petunia* (Petunieae tribe, Solanaceae) and a new generic placement of Argentinean endemic *Petunia patagonica*. *PhytoKeys* **194**:75–93.
- Alvarez, J., and Smyth, D.R. (2002). *CRABS CLAW* and *SPATULA* genes regulate growth and pattern formation during gynoecium development in *Arabidopsis thaliana*. *Int. J. Plant Sci.* **163**:17–41.
- Ampomah-Dwamena, C., Morris, B.A., Sutherland, P., Veit, B., and Yao, J.-L. (2002). Down-regulation of *TM29*, a tomato *SEPALLATA* homolog, causes parthenocarpic fruit development and floral reversion. *Plant Physiol.* **130**:605–617.
- Arrigo, N., and Barker, M.S. (2012). Rarely successful polyploids and their legacy in plant genomes. *Curr. Opin. Plant Biol.* **15**:140–146.
- Azzi, L., Deluche, C., Gévaudant, F., Frangne, N., Delmas, F., Hernoould, M., and Chevalier, C. (2015). Fruit growth-related genes in tomato. *J. Exp. Bot.* **66**:1075–1086.
- Barboza, G.E., Hunziker, A.T., Bernardello, G., Cocucci, A.A., Moscone, A.E., Carrizo García, C., Fuentes, V., Dillon, M.O., Bittrich, V., Cosa, M.T., et al. (2016). Solanaceae. In *The Families and Genera of Vascular Plants: Aquifoliales, Boraginales, Bruniales, Dipsacales, Escalloniales, Garryales, Paracryphiales, Solanales (Except Convolvulaceae), Icacinaceae, Metteniusaceae, K.K. Vahliaaceae*, ed. (Berlin: Springer), pp. 295–357.
- Bemer, M., Karlova, R., Ballester, A.R., Tikunov, Y.M., Bovy, A.G., Wolters-Arts, M., Rossetto, P.d.B., Angenent, G.C., and de Maagd, R.A. (2012). The tomato FRUITFULL homologs TDR4/FUL1 and MBP7/FUL2 regulate ethylene-independent aspects of fruit ripening. *Plant Cell* **24**:4437–4451.
- Blanc, G., and Wolfe, K.H. (2004). Widespread paleopolyploidy in model plant species inferred from age distributions of duplicate genes. *Plant Cell* **16**:1667–1678.
- Bohs, L. (2005). Major clades in *Solanum* based on *ndhF* sequence data. In *A Festschrift for William G. D'Arcy: The Legacy of a Taxonomist. Monographs in Systematic Botany of the Missouri Botanical Garden* 104, R.C. Keating, V.C. Hollowell, and T.B. Croat, eds. (St. Louis, Missouri: Missouri Botanical Garden Press), pp. 27–49.
- Bombarely, A., Moser, M., Amrad, A., Bao, M., Bapaume, L., Barry, C.S., Blied, M., Boersma, M.R., Borghi, L., Bruggmann, R., et al. (2016). Insight into the evolution of the Solanaceae from the parental genomes of *Petunia hybrida*. *Nat. Plants* **2**:16074.
- Bouckaert, R., Heled, J., Kühnert, D., Vaughan, T., Wu, C.-H., Xie, D., Suchard, M.A., Rambaut, A., and Drummond, A.J. (2014). BEAST 2: a software platform for Bayesian evolutionary analysis. *PLoS Comput. Biol.* **10**:e1003537.
- Buchfink, B., Xie, C., and Huson, D.H. (2015). Fast and sensitive protein alignment using DIAMOND. *Nat. Methods* **12**:59–60.
- Busi, M.V., Bustamante, C., D'angelo, C., Hidalgo-Cuevas, M., Boggio, S.B., Valle, E.M., and Zabaleta, E. (2003). MADS-box genes expressed during tomato seed and fruit development. *Plant Mol. Biol.* **52**:801–815.
- Cai, D., Rodríguez, F., Teng, Y., Ané, C., Bonierbale, M., Mueller, L.A., and Spooner, D.M. (2012). Single copy nuclear gene analysis of polyploidy in wild potatoes (*Solanum* section *Petota*). *BMC Evol. Biol.* **12**:70.
- Capella-Gutiérrez, S., Silla-Martínez, J.M., and Gabaldón, T. (2009). TrimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* **25**:1972–1973.
- The Angiosperm Phylogeny Group, Christenhusz, M.J.M., Fay, M.F., Byng, J.W., Judd, W.S., Mabberley, D.J., Sennikov, A.N., Soltis, D.E., Soltis, P.S., Stevens, P.F., et al. (2016). An update of the Angiosperm Phylogeny Group classification for the orders and families of flowering plants: APG IV. *Bot. J. Linn. Soc.* **181**:1–20.
- Chowański, S., Adamski, Z., Marciniak, P., Rosiński, G., Büyükgüzel, E., Büyükgüzel, K., Falabella, P., Scranò, L., Ventrella, E., Lelario, F., and Bufo, S.A. (2016). A review of bioinsecticidal activity of Solanaceae alkaloids. *Toxins* **8**:60.
- Christenhusz, M.J., and Byng, J.W. (2016). The number of known plants species in the world and its annual increase. *Phytotaxa* **261**:201–217.
- Cong, B., Barrero, L.S., and Tanksley, S.D. (2008). Regulatory change in YABBY-like transcription factor led to evolution of extreme fruit size during tomato domestication. *Nat. Genet.* **40**:800–804.
- De Bie, T., Cristianini, N., Demuth, J.P., and Hahn, M.W. (2006). CAFE: a computational tool for the study of gene family evolution. *Bioinformatics* **22**:1269–1271.
- Deanna, R., Wilf, P., and Gandolfo, M.A. (2020). New physaloid fruit-fossil species from early Eocene South America. *Am. J. Bot.* **107**:1749–1762.
- Deanna, R., Larter, M.D., Barboza, G.E., and Smith, S.D. (2019). Repeated evolution of a morphological novelty: a phylogenetic

Plant Communications

analysis of the inflated fruiting calyx in the Physalideae tribe (Solanaceae). *Am. J. Bot.* **106**:270–279.

- Dillon, M.O.** (2005). The Solanaceae of the *Iomas* formations of coastal Peru and Chile. In *Monographs in Systematic Botany from the Missouri Botanical Garden*, R.C. Keating, V.C. Hollowell, and T.B. Croat, eds. (St Louis: Missouri Botanical Garden Press), pp. 131–156.
- Dillon, M.O., Tu, T., Xie, L., Quipuscoa Silvestre, V., and Wen, J.** (2009). Biogeographic diversification in *Nolana* (Solanaceae), a ubiquitous member of the Atacama and Peruvian deserts along the western coast of south America. *J. Syst. Evol.* **47**:457–476.
- Dillon, M.O., Tu, T., Soejima, A., Yi, T., Nie, Z., Tye, A., and Wen, J.** (2007). Phylogeny of *Nolana* (Nolaneae, Solanoideae, Solanaceae) as inferred from granule-bound starch synthase I (GBSSI) sequences. *Taxon* **56**:1000–1011.
- Ditta, G., Pinyopich, A., Robles, P., Pelaz, S., and Yanofsky, M.F.** (2004). The *SEP4* gene of *Arabidopsis thaliana* functions in floral organ and meristem identity. *Curr. Biol.* **14**:1935–1940.
- Doyle, J.J., and Doyle, J.L.** (1987). A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochem. Bull.* **19**:11–15.
- Drummond, A.J., Suchard, M.A., Xie, D., and Rambaut, A.** (2012). Bayesian phylogenetics with BEAUti and the BEAST 1.7. *Mol. Biol. Evol.* **29**:1969–1973.
- Dupin, J., and Smith, S.D.** (2018). Phylogenetics of Datureae (Solanaceae), including description of the new genus *Trompettia* and re-circumscription of the tribe. *Taxon* **67**:359–375.
- Dupin, J., Matzke, N.J., Särkinen, T., Knapp, S., Olmstead, R.G., Bohs, L., and Smith, S.D.** (2017). Bayesian estimation of the global biogeographical history of the Solanaceae. *J. Biogeogr.* **44**:887–899.
- Ebersberger, I., Strauss, S., and von Haeseler, A.** (2009). HaMStR: profile hidden markov model based search for orthologs in ESTs. *BMC Evol. Biol.* **9**:157.
- Edgar, R.C.** (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* **32**:1792–1797.
- Edwards, K.D., Fernandez-Pozo, N., Drake-Stowe, K., Humphry, M., Evans, A.D., Bombarely, A., Allen, F., Hurst, R., White, B., Kernodle, S.P., et al.** (2017). A reference genome for *Nicotiana tabacum* enables map-based cloning of homeologous loci implicated in nitrogen utilization efficiency. *BMC Genom.* **18**:448.
- El-Gebali, S., Mistry, J., Bateman, A., Eddy, S.R., Luciani, A., Potter, S.C., Qureshi, M., Richardson, L.J., Salazar, G.A., Smart, A., et al.** (2019). The Pfam protein families database in 2019. *Nucleic Acids Res.* **47**:D427–D432.
- Emms, D.M., and Kelly, S.** (2019). OrthoFinder: phylogenetic orthology inference for comparative genomics. *Genome Biol.* **20**:238.
- Enright, A.J., Van Dongen, S., and Ouzounis, C.A.** (2002). An efficient algorithm for large-scale detection of protein families. *Nucleic Acids Res.* **30**:1575–1584.
- Ezekiel, R., Singh, N., Sharma, S., and Kaur, A.** (2013). Beneficial phytochemicals in potato - a review. *Food Res. Int.* **50**:487–496.
- FAO.** (2020). *World Food and Agriculture - Statistical Yearbook 2020*.
- Favaro, R., Pinyopich, A., Battaglia, R., Kooiker, M., Borghi, L., Ditta, G., Yanofsky, M.F., Kater, M.M., and Colombo, L.** (2003). MADS-box protein complexes control carpel and ovule development in *Arabidopsis*. *Plant Cell* **15**:2603–2611.
- Fawcett, J.A., Van de Peer, Y., and Maere, S.** (2013). Significance and biological consequences of polyploidization in land plant evolution. In *Plant Genome Diversity*, I.J. Leitch, ed. (Vienna: Springer), pp. 277–293.
- Ferrández, C., Liljegen, S.J., and Yanofsky, M.F.** (2000). Negative regulation of the SHATTERPROOF genes by FRUITFULL during *Arabidopsis* fruit development. *Science* **289**:436–438.
- ## Genome duplications and development in Solanaceae
- Finot, V.L., Marticorena, C., and Marticorena, A.** (2018). Pollen grain morphology of *Nolana* L. (Solanaceae: Nolanoideae: Nolaneae) and related genera of southern south American Solanaceae. *Grana* **57**:415–455.
- Fourquin, C., and Ferrández, C.** (2014). The essential role of NGATHA genes in style and stigma specification is widely conserved across eudicots. *New Phytol.* **202**:1001–1013.
- Franks, R.G., Liu, Z., and Fischer, R.L.** (2006). *SEUSS* and *LEUNIG* regulate cell proliferation, vascular development and organ polarity in *Arabidopsis* petals. *Planta* **224**:801–811.
- Fu, L., Niu, B., Zhu, Z., Wu, S., and Li, W.** (2012). CD-HIT: accelerated for clustering the next-generation sequencing data. *Bioinformatics* **28**:3150–3152.
- Fujisawa, M., Nakano, T., and Ito, Y.** (2011). Identification of potential target genes for the tomato fruit-ripening regulator RIN by chromatin immunoprecipitation. *BMC Plant Biol.* **11**:26.
- Fujisawa, M., Nakano, T., Shima, Y., and Ito, Y.** (2013). A large-scale identification of direct targets of the tomato MADS box transcription factor RIPENING INHIBITOR reveals the regulation of fruit ripening. *Plant Cell* **25**:371–386.
- Gagnon, E., Hilgenhof, R., Orejuela, A., McDonnell, A., Sablok, G., Aubriot, X., Giacomini, L., Gouvêa, Y., Bragionis, T., Stehmann, J.R., et al.** (2022). Phylogenomic discordance suggests polytomies along the backbone of the large genus *Solanum*. *Am. J. Bot.* **109**:580–601.
- Grabherr, M.G., Haas, B.J., Yassour, M., Levin, J.Z., Thompson, D.A., Amit, I., Adiconis, X., Fan, L., Raychowdhury, R., Zeng, Q., et al.** (2011). Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat. Biotechnol.* **29**:644–652.
- Guan, H., Huang, B., Chen, M., Wang, X., Song, S., Liu, H., Chen, R., and Hao, Y.** (2018). Genome-wide identification, phylogeny analysis, expression profiling, and determination of protein-protein interactions of the LEUNIG gene family members in tomato. *Gene* **679**:1–10.
- Guo, J., Xu, W., Hu, Y., Huang, J., Zhao, Y., Zhang, L., Huang, C.H., and Ma, H.** (2020). Phylotranscriptomics in Cucurbitaceae reveal multiple whole-genome duplications and key morphological and molecular innovations. *Mol. Plant* **13**:1117–1133.
- Haas, B.J., Papanicolaou, A., Yassour, M., Grabherr, M., Blood, P.D., Bowden, J., Couger, M.B., Eccles, D., Li, B., Lieber, M., et al.** (2013). *De novo* transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nat. Protoc.* **8**:1494–1512.
- Hahn, M.W.** (2009). Distinguishing among evolutionary models for the maintenance of gene duplicates. *J. Hered.* **100**:605–617.
- Hawkes, J.G.** (1990). *The Potato: Evolution, Biodiversity and Genetic Resources* (Washington, DC: Belhaven Press).
- Herrera, C.M.** (1982). Defense of ripe fruit from pests: its significance in relation to plant-disperser interactions. *Am. Nat.* **120**:218–241.
- Hijmans, R.J., Gavrilenko, T., Stephenson, S., Bamberg, J., Salas, A., and Spooner, D.M.** (2007). Geographical and environmental range expansion through polyploidy in wild potatoes (*Solanum* section Petota). *Glob. Ecol. Biogeogr.* **16**:485–495.
- Hoare, A.L., and Knapp, S.** (1997). Phylogenetic conspectus of the tribe Hyoscyameae (Solanaceae). *Bull. Nat. Hist. Mus. Bot. ser.* **27**:11.
- Huang, C.H., Qi, X., Chen, D., Qi, J., and Ma, H.** (2020). Recurrent genome duplication events likely contributed to both the ancient and recent rise of ferns. *J. Integr. Plant Biol.* **62**:433–455.
- Huang, C.H., Zhang, C., Liu, M., Hu, Y., Gao, T., Qi, J., and Ma, H.** (2016a). Multiple polyploidization events across Asteraceae with two nested events in the early history revealed by nuclear phylogenomics. *Mol. Biol. Evol.* **33**:2820–2835.

- Huang, C.H., Sun, R., Hu, Y., Zeng, L., Zhang, N., Cai, L., Zhang, Q., Koch, M.A., Al-Shehbaz, I., Edger, P.P., et al. (2016b). Resolution of Brassicaceae phylogeny using nuclear genes uncovers nested radiations and supports convergent morphological evolution. *Mol. Biol. Evol.* **33**:394–412.
- Huang, W., Zhang, L., Columbus, J.T., Hu, Y., Zhao, Y., Tang, L., Guo, Z., Chen, W., McKain, M., Bartlett, M., et al. (2022). A well-supported nuclear phylogeny of Poaceae and implications for the evolution of C4 photosynthesis. *Mol. Plant* **15**:755–777.
- Huang, Z., Van Houten, J., Gonzalez, G., Xiao, H., and van der Knaap, E. (2013). Genome-wide identification, phylogeny and expression analysis of SUN, OFP and YABBY gene family in tomato. *Mol. Genet. Genomics*. **288**:111–129.
- Hunziker, A.T. (2001). Genera *Solanacearum*: The Genera of Solanaceae Illustrated, Arranged According to a New System (Gantner: Ruggell, Liechtenstein).
- Ireland, H.S., Yao, J.L., Tomes, S., Sutherland, P.W., Nieuwenhuizen, N., Gunaseelan, K., Winz, R.A., David, K.M., and Schaffer, R.J. (2013). Apple SEPALLATA1/2-like genes control fruit flesh development and ripening. *Plant J.* **73**:1044–1056.
- Janssens, S.B., Knox, E.B., Huysmans, S., Smets, E.F., and Merckx, V.S.F.T. (2009). Rapid radiation of *Impatiens* (Balsaminaceae) during Pliocene and Pleistocene: result of a global climate change. *Mol. Phylogenet. Evol.* **52**:806–824.
- Jiao, Y., Wickett, N.J., Ayyampalayam, S., Chanderbali, A.S., Landherr, L., Ralph, P.E., Tomsho, L.P., Hu, Y., Liang, H., Soltis, P.S., et al. (2011). Ancestral polyploidy in seed plants and angiosperms. *Nature* **473**:97–100.
- Katoh, K., and Standley, D.M. (2013). MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol. Biol. Evol.* **30**:772–780.
- Kim, S., Park, J., Yeom, S.-I., Kim, Y.-M., Seo, E., Kim, K.-T., Kim, M.-S., Lee, J.M., Cheong, K., Shin, H.-S., et al. (2017). New reference genome sequences of hot pepper reveal the massive evolution of plant disease-resistance genes by retroduplication. *Genome Biol.* **18**:210.
- Knapp, S. (2002a). Tobacco to tomatoes: a phylogenetic perspective on fruit diversity in the Solanaceae. *J. Exp. Bot.* **53**:2001–2022.
- Knapp, S. (2002b). Floral diversity and evolution in the Solanaceae. In *Developmental Genetics and Plant Evolution*, Q.C.B. Cronk, R.M. Bateman, and J.A. Hawkins, eds. (London: Taylor and Francis), pp. 267–297.
- Knapp, S., and Peralta, I.E. (2016). The tomato (*Solanum lycopersicum* L., Solanaceae) and its botanical relatives. In *The Tomato Genome*, M. Causse, J. Giovannoni, M. Bouzayen, and M. Zouine, eds. (Berlin, Heidelberg: Springer), pp. 7–21.
- Landis, J.B., Soltis, D.E., Li, Z., Marx, H.E., Barker, M.S., Tank, D.C., and Soltis, P.S. (2018). Impact of whole-genome duplication events on diversification rates in angiosperms. *Am. J. Bot.* **105**:348–363.
- One Thousand Plant Transcriptomes Initiative, Barker, M.S., Carpenter, E.J., Deyholos, M.K., Gitzendanner, M.A., Graham, S.W., Grosse, I., Li, Z., Melkonian, M., and Mirarab, S. (2019). One thousand plant transcriptomes and the phylogenomics of green plants. *Nature* **574**:679–685.
- Leitch, I.J., Hanson, L., Lim, K.Y., Kovarik, A., Chase, M.W., Clarkson, J.J., and Leitch, A.R. (2008). The ups and downs of genome size evolution in polyploid species of *Nicotiana* (Solanaceae). *Ann. Bot.* **101**:805–814.
- Levin, D.A. (1983). Polyploidy and novelty in flowering plants. *Am. Nat.* **122**:1–25.
- Li, H.-Q., Gui, P., Xiong, S.-Z., and Averett, J.E. (2013). The generic position of two species of tribe Physaleae (Solanaceae) inferred from three DNA sequences: a case study on *Physaliastrum* and *Archiphysalis*. *Biochem. Syst. Ecol.* **50**:82–89.
- Li, L., Stoeckert, C.J., and Roos, D.S. (2003). OrthoMCL: identification of ortholog groups for eukaryotic genomes. *Genome Res.* **13**:2178–2189.
- Li, Z., Tiley, G.P., Galuska, S.R., Reardon, C.R., Kidder, T.I., Rundell, R.J., and Barker, M.S. (2018). Multiple large-scale gene and genome duplications during the evolution of hexapods. *Proc. Natl. Acad. Sci. USA* **115**:4713–4718.
- Liu, Z., Ma, H., Jung, S., Main, D., and Guo, L. (2020). Developmental mechanisms of fleshy fruit diversity in Rosaceae. *Annu. Rev. Plant Biol.* **71**:547–573.
- Luo, R., Liu, B., Xie, Y., Li, Z., Huang, W., Yuan, J., He, G., Chen, Y., Pan, Q., Liu, Y., et al. (2012). SOAPdenovo2: an empirically improved memory-efficient short-read *de novo* assembler. *GigaScience* **1**:18.
- Maere, S., and Van de Peer, Y. (2010). Duplicate retention after small- and large-scale duplications. In *Evolution after Gene Duplication*, K. Dittmar and D.A. Liberles, eds. (Hoboken, NJ: Wiley), pp. 31–56.
- Magallón, S., Gómez-Acevedo, S., Sánchez-Reyes, L.L., and Hernández-Hernández, T. (2015). A metacalibrated time-tree documents the early rise of flowering plant phylogenetic diversity. *New Phytol.* **207**:437–453.
- Malcomber, S.T., and Kellogg, E.A. (2005). *SEPALLATA* gene diversification: brave new whorls. *Trends Plant Sci.* **10**:427–435.
- Martins, T.R., and Barkman, T.J. (2005). Reconstruction of Solanaceae phylogeny using the nuclear gene *SAMT*. *Syst. Bot.* **30**:435–447.
- Mayrose, I., Zhan, S.H., Rothfels, C.J., Magnuson-Ford, K., Barker, M.S., Rieseberg, L.H., and Otto, S.P. (2011). Recently formed polyploid plants diversify at lower rates. *Science* **333**:1257.
- McKey, D. (1979). The distribution of secondary compounds within plants. In *Herbivores: Their Interaction with Secondary Plant Metabolites*, G.A. Rosenthal and D.H. Janzen, eds. (New York: Academic), pp. 55–134.
- Milner, S.E., Brunton, N.P., Jones, P.W., O'Brien, N.M., Collins, S.G., and Maguire, A.R. (2011). Bioactivities of glycoalkaloids and their aglycones from *Solanum* species. *J. Agric. Food Chem.* **59**:3454–3484.
- Mirarab, S., Reaz, R., Bayzid, M.S., Zimmermann, T., Swenson, M.S., and Warnow, T. (2014). ASTRAL: genome-scale coalescent-based species tree estimation. *Bioinformatics* **30**:i541–i548.
- Mistry, J., Finn, R.D., Eddy, S.R., Bateman, A., and Punta, M. (2013). Challenges in homology search: HMMER3 and convergent evolution of coiled-coil regions. *Nucleic Acids Res.* **41**:e121.
- Ng, J., and Smith, S.D. (2016). Widespread flower color convergence in Solanaceae via alternate biochemical pathways. *New Phytol.* **209**:407–417.
- Nguyen, L.-T., Schmidt, H.A., von Haeseler, A., and Minh, B.Q. (2015). IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol. Biol. Evol.* **32**:268–274.
- Ocares, N., and Mejía, N. (2016). Suppression of the D-class MADS-box *AGL11* gene triggers seedlessness in fleshy fruits. *Plant Cell Rep.* **35**:239–254.
- Ohno, S. (1970). *Evolution by Gene Duplication* (New York: Springer).
- Olmstead, R.G., and Palmer, J.D. (1992). A chloroplast DNA phylogeny of the Solanaceae: subfamilial relationships and character evolution. *Ann. Mo. Bot. Gard.* **79**:346–360.
- Olmstead, R.G., and Bohs, L. (2007). A summary of molecular systematic research in Solanaceae: 1982–2006. *Acta Hort.* **745**:255–268.
- Olmstead, R.G., Sweere, J.A., Spangler, R.E., Bohs, L., and Palmer, J.D. (1999). Phylogeny and provisional classification of the Solanaceae based on chloroplast DNA. *Solanaceae IV* **1**:1–137.

Plant Communications

- Olmstead, R.G., Bohs, L., Migid, H.A., Santiago-Valentin, E., Garcia, V.F., and Collier, S.M. (2008). A molecular phylogeny of the Solanaceae. *Taxon* **57**:1159–1181.
- Orejuela, A., Wahlert, G., Orozco, C.I., Barboza, G., and Bohs, L. (2017). Phylogeny of the tribes Juanulloae and Solandreae (Solanaceae). *Taxon* **66**:379–392.
- Ortiz-Ramírez, C.I., Plata-Arboleda, S., and Pabón-Mora, N. (2018). Evolution of genes associated with gynoecium patterning and fruit development in Solanaceae. *Ann. Bot.* **121**:1211–1230.
- Otto, S.P. (2007). The evolutionary consequences of polyploidy. *Cell* **131**:452–462.
- Pelaz, S., Ditta, G.S., Baumann, E., Wisman, E., and Yanofsky, M.F. (2000). B and C floral organ identity functions require *SEPALLATA* MADS-box genes. *Nature* **405**:200–203.
- Pfannebecker, K.C., Lange, M., Rupp, O., and Becker, A. (2017). An evolutionary framework for carpel developmental control genes. *Mol. Biol. Evol.* **34**:330–348.
- Pfannebecker, K.C., Lange, M., Rupp, O., and Becker, A. (2017). Seed plant-specific gene lineages involved in carpel development. *Mol. Biol. Evol.* **34**:925–942.
- Pinyopich, A., Ditta, G.S., Savidge, B., Liljegren, S.J., Baumann, E., Wisman, E., and Yanofsky, M.F. (2003). Assessing the redundancy of MADS-box genes during carpel and ovule development. *Nature* **424**:85–88.
- Price, M.N., Dehal, P.S., and Arkin, A.P. (2009). FastTree: computing large minimum evolution trees with profiles instead of a distance matrix. *Mol. Biol. Evol.* **26**:1641–1650.
- Qi, X., Kuo, L.Y., Guo, C., Li, H., Li, Z., Qi, J., Wang, L., Hu, Y., Xiang, J., Zhang, C., et al. (2018). A well-resolved fern nuclear phylogeny reveals the evolution history of numerous transcription factor families. *Mol. Phylogenet. Evol.* **127**:961–977.
- Qin, C., Yu, C., Shen, Y., Fang, X., Chen, L., Min, J., Cheng, J., Zhao, S., Xu, M., Luo, Y., et al. (2014). Whole-genome sequencing of cultivated and wild peppers provides insights into *Capsicum* domestication and specialization. *Proc. Natl. Acad. Sci. USA* **111**:5135–5140.
- Raj, S.P., Solomon, P.R., and Thangaraj, B. (2022). Solanaceae. In *Biodiesel from Flowering Plants* (Springer), pp. 543–549.
- Rambaut, A., Drummond, A.J., Xie, D., Baele, G., and Suchard, M.A. (2018). Posterior summarization in Bayesian phylogenetics using Tracer 1.7. *Syst. Biol.* **67**:901–904.
- Ren, R., Wang, H., Guo, C., Zhang, N., Zeng, L., Chen, Y., Ma, H., and Qi, J. (2018). Widespread whole genome duplications contribute to genome complexity and species diversity in angiosperms. *Mol. Plant* **11**:414–428.
- Rieseberg, L.H., and Willis, J.H. (2007). Plant speciation. *Science* **317**:910–914.
- Rodríguez, G.R., Muñoz, S., Anderson, C., Sim, S.C., Michel, A., Causse, M., Gardener, B.B.M., Francis, D., and van der Knaap, E. (2011). Distribution of SUN, OVATE, LC, and FAS in the tomato germplasm and the relationship to fruit shape diversity. *Plant Physiol.* **156**:275–285.
- Särkinen, T., Bohs, L., Olmstead, R.G., and Knapp, S. (2013). A phylogenetic framework for evolutionary study of the nightshades (Solanaceae): a dated 1000-tip tree. *BMC Evol. Biol.* **13**:214.
- Tomato Genome Consortium, Tabata, S., Hirakawa, H., Asamizu, E., Shirasawa, K., Isobe, S., Kaneko, T., Nakamura, Y., Shibata, D., Aoki, K., et al. (2012). The tomato genome sequence provides insights into fleshy fruit evolution. *Nature* **485**:635–641.
- Schiavinato, M., Marcet-Houben, M., Dohm, J.C., Gabaldón, T., and Himmelbauer, H. (2020). Parental origin of the allotetraploid tobacco *Nicotiana benthamiana*. *Plant J.* **102**:541–554.

Genome duplications and development in Solanaceae

- Schlueter, J.A., Dixon, P., Granger, C., Grant, D., Clark, L., Doyle, J.J., and Shoemaker, R.C. (2004). Mining EST databases to resolve evolutionary events in major crop species. *Genome* **47**:868–876.
- Schranz, M.E., Mohammadin, S., and Edger, P.P. (2012). Ancient whole genome duplications, novelty and diversification: the WGD Radiation Lag-Time Model. *Curr. Opin. Plant Biol.* **15**:147–153.
- Schwarz-Sommer, Z., Huijser, P., Nacken, W., Saedler, H., and Sommer, H. (1990). Genetic control of flower development by homeotic genes in *Antirrhinum majus*. *Science* **250**:931–936.
- Seymour, G.B., Østergaard, L., Chapman, N.H., Knapp, S., and Martin, C. (2013). Fruit development and ripening. *Annu. Rev. Plant Biol.* **64**:219–241.
- Seymour, G.B., Ryder, C.D., Cevik, V., Hammond, J.P., Popovich, A., King, G.J., Vrebalov, J., Giovannoni, J.J., and Manning, K. (2011). A *SEPALLATA* gene is involved in the development and ripening of strawberry (*Fragaria × ananassa* Duch.) fruit, a non-climacteric tissue. *J. Exp. Bot.* **62**:1179–1188.
- Shan, H., Zhang, N., Liu, C., Xu, G., Zhang, J., Chen, Z., and Kong, H. (2007). Patterns of gene duplication and functional diversification during the evolution of the AP1/SQUA subfamily of plant MADS-box genes. *Mol. Phylogenet. Evol.* **44**:26–41.
- Shore, P., and Sharrocks, A.D. (1995). The MADS-box family of transcription factors. *Eur. J. Biochem.* **229**:1–13.
- Shrestha, B., Guragain, B., and Sridhar, V.V. (2014). Involvement of corepressor LUH and the adapter proteins SLK1 and SLK2 in the regulation of abiotic stress response genes in *Arabidopsis*. *BMC Plant Biol.* **14**:54.
- Sierro, N., Battey, J.N.D., Ouali, S., Bovet, L., Goepfert, S., Bakaher, N., Peitsch, M.C., and Ivanov, N.V. (2013). Reference genomes and transcriptomes of *Nicotiana glauca* and *Nicotiana glauca*. *Genome Biol.* **14**:R60.
- Slater, G.S.C., and Birney, E. (2005). Automated generation of heuristics for biological sequence comparison. *BMC Bioinf.* **6**:31.
- Smith, S.A., and O'Meara, B.C. (2012). treePL: divergence time estimation using penalized likelihood for large phylogenies. *Bioinformatics* **28**:2689–2690.
- Smith, S.A., Beaulieu, J.M., and Donoghue, M.J. (2010). An uncorrelated relaxed-clock analysis suggests an earlier origin for flowering plants. *Proc. Natl. Acad. Sci. USA* **107**:5897–5902.
- Smith, S.A., Brown, J.W., and Walker, J.F. (2018). So many genes, so little time: a practical approach to divergence-time estimation in the genomic era. *PLoS One* **13**:e0197433.
- Smith, S.A., Moore, M.J., Brown, J.W., and Yang, Y. (2015). Analysis of phylogenomic datasets reveals conflict, concordance, and gene duplications with examples from animals and plants. *BMC Evol. Biol.* **15**:150.
- Soltis, D.E., Soltis, P.S., and Tate, J.A. (2004). Advances in the study of polyploidy since plant speciation. *New Phytol.* **161**:173–191.
- Soltis, D.E., Albert, V.A., Leebens-Mack, J., Bell, C.D., Paterson, A.H., Zheng, C., Sankoff, D., Depamphilis, C.W., Wall, P.K., and Soltis, P.S. (2009). Polyploidy and angiosperm diversification. *Am. J. Bot.* **96**:336–348.
- Spalink, D., Stoffel, K., Walden, G.K., Hulse-Kemp, A.M., Hill, T.A., Van Deynze, A., and Bohs, L. (2018). Comparative transcriptomics and genomic patterns of discordance in Capsiceae (Solanaceae). *Mol. Phylogenet. Evol.* **126**:293–302.
- Stahle, M.I., Kuehlich, J., Staron, L., von Arnim, A.G., and Golz, J.F. (2009). YABBYs and the transcriptional corepressors LEUNIG and LEUNIG_HOMOLOG maintain leaf polarity and meristem activity in *Arabidopsis*. *Plant Cell* **21**:3105–3118.

- Stamatakis, A.** (2006). RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics* **22**:2688–2690.
- Stebbins, G.L.** (1940). The significance of polyploidy in plant evolution. *Am. Nat.* **74**:54–66.
- Suyama, M., Torrents, D., and Bork, P.** (2006). PAL2NAL: robust conversion of protein sequence alignments into the corresponding codon alignments. *Nucleic Acids Res.* **34**:W609–W612.
- Szymański, J., Bocobza, S., Panda, S., Sonawane, P., Cárdenas, P.D., Lashbrooke, J., Kamble, A., Shahaf, N., Meir, S., Bovy, A., et al.** (2020). Analysis of wild tomato introgression lines elucidates the genetic basis of transcriptome and metabolome variation underlying fruit traits and pathogen response. *Nat. Genet.* **52**:1111–1121.
- Tang, D., Jia, Y., Zhang, J., Li, H., Cheng, L., Wang, P., Bao, Z., Liu, Z., Feng, S., Zhu, X., et al.** (2022). Genome evolution and diversity of wild and cultivated potatoes. *Nature* **606**:535–541.
- Tank, D.C., Eastman, J.M., Pennell, M.W., Soltis, P.S., Soltis, D.E., Hinchliff, C.E., Brown, J.W., Sessa, E.B., and Harmon, L.J.** (2015). Nested radiations and the pulse of angiosperm diversification: increased diversification rates often follow whole genome duplications. *New Phytol.* **207**:454–467.
- Tepe, E.J., Anderson, G.J., Spooner, D.M., and Bohs, L.** (2016). Relationships among wild relatives of the tomato, potato, and pepino. *Taxon* **65**:262–276.
- Tu, T., Volis, S., Dillon, M.O., Sun, H., and Wen, J.** (2010). Dispersals of Hyoscyameae and Mandragoreae (Solanaceae) from the new world to Eurasia in the early Miocene and their biogeographic diversification within Eurasia. *Mol. Phylogenet. Evol.* **57**:1226–1237.
- Tu, T.Y., Sun, H., Gu, Z.J., and Yue, J.P.** (2005). Cytological studies on the Sino-Himalayan endemic *Anisodus* and four related genera from the tribe Hyoscyameae (Solanaceae) and their systematic and evolutionary implications. *Bot. J. Linn. Soc.* **147**:457–468.
- Valenta, K., Burke, R.J., Styler, S.A., Jackson, D.A., Melin, A.D., and Lehman, S.M.** (2013). Colour and odour drive fruit selection and seed dispersal by mouse lemurs. *Sci. Rep.* **3**:2424.
- Velichkevich, F.Y., and Zastawniak, E.** (2003). The Pliocene flora of Kholmech, south-eastern Belarus and its correlation with other Pliocene floras of Europe. *Acta Palaeobot.* **43**:137–259.
- Busi, M.V., Bustamante, C., D'Angelo, C., Hidalgo-Cuevas, M., Boggio, S.B., Valle, E.M., and Zabaleta, E.** (2003). MADS-box genes expressed during tomato seed and fruit development. *Plant Mol. Biol.* **52**:801–815.
- Volis, S., Fogel, K., Tu, T., Sun, H., and Zaretsky, M.** (2018). Evolutionary history and biogeography of *Mandragora* L. (Solanaceae). *Mol. Phylogenet. Evol.* **129**:85–95.
- Vrebalov, J., Ruezinsky, D., Padmanabhan, V., White, R., Medrano, D., Drake, R., Schuch, W., and Giovannoni, J.** (2002). A MADS-box gene necessary for fruit ripening at the tomato *ripening-inhibitor (rin)* locus. *Science* **296**:343–346.
- Wang, L., Li, J., Zhao, J., and He, C.** (2015). Evolutionary developmental genetics of fruit morphological variation within the Solanaceae. *Front. Plant Sci.* **6**:248.
- Wang, Y., Zhang, J., Hu, Z., Guo, X., Tian, S., and Chen, G.** (2019). Genome-Wide analysis of the MADS-Box transcription factor family in *Solanum lycopersicum*. *Int. J. Mol. Sci.* **20**:2961.
- Wang, Y., Tang, H., DeBarry, J.D., Tan, X., Li, J., Wang, X., Lee, T.H., Jin, H., Marler, B., Guo, H., et al.** (2012). MCScanX: a toolkit for detection and evolutionary analysis of gene synteny and collinearity. *Nucleic Acids Res.* **40**:e49.
- Weese, T.L., and Bohs, L.** (2007). A three-gene phylogeny of the genus *Solanum* (Solanaceae). *Syst. Bot.* **32**:445–463.
- Wheeler, L.C., Walker, J.F., Ng, J., Deanna, R., Dunbar-Wallis, A., Backes, A., Pezzi, P.H., Palchetti, M.V., Robertson, H.M., Monaghan, A., et al.** (2022). Transcription factors evolve faster than their structural gene targets in the flavonoid pigment pathway. *Mol. Biol. Evol.* **39**:msac044.
- Whitson, M., and Manos, P.S.** (2005). Untangling *Physalis* (Solanaceae) from the physaloids: a two-gene phylogeny of the Physalinae. *Syst. Bot.* **30**:216–230.
- Wilf, P., Carvalho, M.R., Gandolfo, M.A., and Cúneo, N.R.** (2017). Eocene lantern fruits from Gondwanan Patagonia and the early origins of Solanaceae. *Science* **355**:71–75.
- Wu, M., Kostyun, J.L., and Moyle, L.C.** (2019). Genome sequence of *Jaltomata* addresses rapid reproductive trait evolution and enhances comparative genomics in the hyper-diverse Solanaceae. *Genome Biol. Evol.* **11**:335–349.
- Wu, M., Kostyun, J.L., Hahn, M.W., and Moyle, L.C.** (2018). Dissecting the basis of novel trait evolution in a radiation with widespread phylogenetic discordance. *Mol. Ecol.* **27**:3301–3316.
- Xiang, Y., Huang, C.H., Hu, Y., Wen, J., Li, S., Yi, T., Chen, H., Xiang, J., and Ma, H.** (2017). Evolution of Rosaceae fruit types based on nuclear phylogeny in the context of geological times and genome duplication. *Mol. Biol. Evol.* **34**:262–281.
- Xu, S., Brockmüller, T., Navarro-Quezada, A., Kuhl, H., Gase, K., Ling, Z., Zhou, W., Kreitzer, C., Stanke, M., Tang, H., et al.** (2017). Wild tobacco genomes reveal the evolution of nicotine biosynthesis. *Proc. Natl. Acad. Sci. USA* **114**:6133–6138.
- Potato Genome Sequencing Consortium, Xu, X., Pan, S., Cheng, S., Zhang, B., Mu, D., Ni, P., Zhang, G., Yang, S., Li, R., et al.** (2011). Genome sequence and analysis of the tuber crop potato. *Nature* **475**:189–195.
- Yang, Y., and Smith, S.A.** (2014). Orthology inference in nonmodel organisms using transcriptomes and low-coverage genomes: improving accuracy and matrix occupancy for phylogenomics. *Mol. Biol. Evol.* **31**:3081–3092.
- Yang, Y., Moore, M.J., Brockington, S.F., Mikenas, J., Olivieri, J., Walker, J.F., and Smith, S.A.** (2018). Improved transcriptome sampling pinpoints 26 ancient and more recent polyploidy events in Caryophyllales, including two allopolyploidy events. *New Phytol.* **217**:855–870.
- Yang, Y., Moore, M.J., Brockington, S.F., Soltis, D.E., Wong, G.K.S., Carpenter, E.J., Zhang, Y., Chen, L., Yan, Z., Xie, Y., et al.** (2015). Dissecting molecular evolution in the highly diverse plant clade Caryophyllales using transcriptome sequencing. *Mol. Biol. Evol.* **32**:2001–2014.
- Yang, Z.** (2007). PAML 4: phylogenetic analysis by maximum likelihood. *Mol. Biol. Evol.* **24**:1586–1591.
- Yu, G., Wang, L.G., Han, Y., and He, Q.Y.** (2012). ClusterProfiler: an R package for comparing biological themes among gene clusters. *OMICS A J. Integr. Biol.* **16**:284–287.
- Zachos, J.C., Dickens, G.R., and Zeebe, R.E.** (2008). An early Cenozoic perspective on greenhouse warming and carbon-cycle dynamics. *Nature* **451**:279–283.
- Zahn, L.M., Feng, B., and Ma, H.** (2006). Beyond the ABC-model: regulation of floral homeotic genes. *Adv. Bot. Res.* **44**:163–207.
- Zahn, L.M., Kong, H., Leebens-Mack, J.H., Kim, S., Soltis, P.S., Landherr, L.L., Soltis, D.E., Depamphilis, C.W., and Ma, H.** (2005). The evolution of the SEPALLATA subfamily of MADS-box genes: a preangiosperm origin with multiple duplications throughout angiosperm history. *Genetics* **169**:2209–2223.
- Zamora-Tavares, M.D.P., Martínez, M., Magallón, S., Guzmán-Dávalos, L., and Vargas-Ponce, O.** (2016). *Physalis* and physaloids: a recent and complex evolutionary history. *Mol. Phylogenet. Evol.* **100**:41–50.

Plant Communications

- Zhang, C., Scornavacca, C., Molloy, E.K., and Mirarab, S. (2020a). ASTRAL-Pro: quartet-based species-tree inference despite paralogy. *Mol. Biol. Evol.* **37**:3292–3307.
- Zhang, C., Huang, C.H., Liu, M., Hu, Y., Panero, J.L., Luebert, F., Gao, T., and Ma, H. (2021). Phylotranscriptomic insights into Asteraceae diversity, polyploidy, and morphological innovation. *J. Integr. Plant Biol.* **63**:1273–1293.
- Zhang, C., Zhang, T., Luebert, F., Xiang, Y., Huang, C.H., Hu, Y., Rees, M., Frohlich, M.W., Qi, J., Weigend, M., and Ma, H. (2020b). Asterid phylogenomics/phylotranscriptomics uncover morphological evolutionary histories and support phylogenetic placement for numerous whole genome duplications. *Mol. Biol. Evol.* **37**:3188–3210.
- Zhang, J., Hu, Z., Wang, Y., Yu, X., Liao, C., Zhu, M., and Chen, G. (2018). Suppression of a tomato *SEPALLATA* MADS-box gene, *SICMB1*, generates altered inflorescence architecture and enlarged sepals. *Plant Sci.* **272**:75–87.

Genome duplications and development in Solanaceae

- Zhang, L., Zhu, X., Zhao, Y., Guo, J., Zhang, T., Huang, W., Huang, J., Hu, Y., Huang, C.H., and Ma, H. (2022). Phylotranscriptomics resolves the phylogeny of Pooideae and uncovers factors for their adaptive evolution. *Mol. Biol. Evol.* **39**:msac026.
- Zhang, N., Zeng, L., Shan, H., and Ma, H. (2012). Highly conserved low-copy nuclear genes as effective markers for phylogenetic analyses in angiosperms. *New Phytol.* **195**:923–937.
- Zhao, Y., Zhang, R., Jiang, K.W., Qi, J., Hu, Y., Guo, J., Zhu, R., Zhang, T., Egan, A.N., Yi, T.S., et al. (2021). Nuclear phylotranscriptomics and phylogenomics support numerous polyploidization events and hypotheses for the evolution of rhizobial nitrogen-fixing symbiosis in Fabaceae. *Mol. Plant* **14**:748–773.
- Zhou, Y., Zhang, Z., Bao, Z., Li, H., Lyu, Y., Zan, Y., Wu, Y., Cheng, L., Fang, Y., Wu, K., et al. (2022). Graph pangenome captures missing heritability and empowers tomato breeding. *Nature* **606**:527–534.