Received: 8 November 2021 | Accepted: 14 April 2022

DOI: 10.1002/tpg2.20223

The Plant Genome

ORIGINAL RESEARCH

Genome sequence for the blue-flowered Andean shrub *Iochroma cyaneum* reveals extensive discordance across the berry clade of Solanaceae

Adrian F. Powell[1,†] | **Jing Zhang**[1,†] | **Duncan Hauser**[1] | **Julianne A. Vilela**[4] | **Alice Hu**[1] | **Daniel J. Gates**[2,3] | **Lukas A. Mueller**[1] | **Fay-Wei Li**[1,5] | **Susan R. Strickler**[1] | **Stacey D. Smith**[6]

[1]Boyce Thompson Institute, Ithaca, NY, USA

[2]School of Biological Sciences, Univ. of Nebraska, Lincoln, NE, USA

[3]Current address: Checkerspot, Inc., Alameda, CA, USA

[4]Philippine Genome Center, Program for Agriculture, Livestock, Forestry and Fisheries, Univ. of the Phillipines Los Baños, Laguna, Phillipines

[5]Plant Biology Section, Cornell Univ., Ithaca, NY, USA

[6]Dep. of Ecology and Evolutionary Biology, Univ. of Colorado, Boulder, CO, USA

Correspondence
Stacey D. Smith, Dep. of Ecology and Evolutionary Biology, Univ. of Colorado, Boulder, CO 80309, USA.
Email: stacey.d.smith@colorado.edu

Assigned to Associate Editor Allen Van Deynze.

[†]These authors contributed equally to this work.

Abstract

The tomato (*Solanum lycopersicum* L.) family, Solanaceae, is a model clade for a wide range of applied and basic research questions. Currently, reference-quality genomes are available for over 30 species from seven genera, and these include numerous crops as well as wild species [e.g., *Jaltomata sinuosa* (Miers) Mione and *Nicotiana attenuata* Torr. ex S. Watson]. Here we present the genome of the showy-flowered Andean shrub *Iochroma cyaneum* (Lindl.) M. L. Green, a woody lineage from the tomatillo (*Physalis philadelphica* Lam.) subfamily Physalideae. The assembled size of the genome (2.7 Gb) is more similar in size to pepper (*Capsicum annuum* L.) (2.6 Gb) than to other sequenced diploid members of the berry clade of Solanaceae [e.g., potato (*Solanum tuberosum* L.), tomato, and *Jaltomata*]. Our assembly recovers 92% of the conserved orthologous set, suggesting a nearly complete genome for this species. Most of the genomic content is repetitive (69%), with *Gypsy* elements alone accounting for 52% of the genome. Despite the large amount of repetitive content, most of the 12 *I. cyaneum* chromosomes are highly syntenic with tomato. Bayesian concordance analysis provides strong support for the berry clade, including *I. cyaneum*, but reveals extensive discordance along the backbone, with placement of chili pepper and *Jaltomata* being highly variable across gene trees. The *I. cyaneum* genome contributes to a growing wealth of genomic resources in Solanaceae and underscores the need for expanded sampling of diverse berry genomes to dissect major morphological transitions.

1 | INTRODUCTION

Advances in comparative genomics rely on moving from assembling high-quality genomes from single model species to building model clades (Rogers, 2018). Model clades, as described by Donoghue and Edwards (2019), are lineages in

Abbreviations: BUSCO, benchmarking universal single-copy ortholog; CF, concordance factor; GO, gene ontology; LTR, long-terminal repeat.

which we sample densely across species to identify evolutionary transitions and build multilayered datasets to understand the mechanisms and drivers of those transitions. The genomic layer of clade biology has been quickly accumulated in taxa with small genomes (Feng et al., 2020; Kim et al., 2021; Miyauchi et al., 2020), but more slowly in plants, where genomes can be as large as 149 Gb (Pellicer et al., 2010). Still, clusters of genomes have been built around plant model species and crops where comparative evolutionary studies can result in direct applications (Ma et al., 2021; Saad et al., 2021).

One such emerging model clade is the tomato (*Solanum lycopersicum* L.) family, Solanaceae. This family comprises nearly 3,000 species, roughly 40 of which have been domesticated, particularly in the fleshy-fruited subclade Solanoideae (Pickersgill, 2007; Samuels, 2015). The first published genome from this clade was potato (*Solanum tuberosum* L.) (The Potato Genome Sequencing Consortium, 2011) closely followed by tomato (The Tomato Genome Consortium, 2012). More recently sequenced economically important species include tobacco (*Nicotiana tabacum* L.) (Sierro et al., 2014), pepper (*Capsicum annuum* L.) (Kim et al., 2014), eggplant (*Solanum melongena* L.) (Barchi et al., 2019; Hirakawa et al., 2014), and Chinese wolfberry (*Lycium barbarum* L.) (Cao et al., 2021). In addition to these crops and model organisms, many wild species have recently been sequenced, for example, for members of the genera *Nicotiana* (Xu et al., 2017), *Petunia* (Bombarely et al., 2016), *Solanum* (Aversano et al., 2015; Razali et al., 2018; Schmidt et al., 2017), *Capsicum* (Qin et al., 2014), and *Jaltomata* (Wu et al., 2018). These taxa capture wide trait variation, from fleshy to dry fruits, self-incompatible to self-compatible, and annuals to perennials. Accordingly, comparative analyses have provided insights into the genomic basis for a range of key traits. Studies in this family have been particularly informative with respect to developmental processes (Kim et al., 2014), such as fruit ripening, and the evolution of specialized metabolites such as the defensive alkaloids and the colorful flavonoids and carotenoids (Cardenas et al., 2015; Gebhardt, 2016).

Here we present a de novo assembly of the genome of *Iochroma cyaneum* (Lindl.) M. L. Green, a blue-flowered shrub native to the Andes. The genus *Iochroma* falls in the large fleshy-fruited subfamily (Solanoideae) (Särkinen et al., 2013) and is related to the tomatillo (*Physalis philadelphica* Lam.) and pepper (Deanna et al., 2019). Unlike species in these genera, *Iochroma* species are woody shrubs or treelets with some reaching up to 15 m (Shaw, 1998). Moreover, while its close relatives in the tomatillo tribe Physalideae are largely insect-pollinated (Knapp, 2010), most species of *Iochroma* are specialized for hummingbird pollination (Smith et al., 2008). Their colorful tubular flowers are arranged in large inflorescences, and with the ease of hybridization among species of different colors (Smith & Baum, 2007), they have become increasingly popular in the horticultural trade (Meerow et al., 2004). Given their wide range of flower colors and sizes, *I. cyaneum* has served as a model for understanding the ecological factors and genetic mechanisms that drive floral evolution (Muchhala et al., 2014; Smith, Ane, et al., 2008; Smith & Rausher, 2011).

Comparative genomic analyses of *I. cyaneum* and related taxa have the potential to provide new insights into the evolutionary history of Solanaceae broadly as well as the changes unique to this hummingbird-pollinated lineage. For example, phylogenomic analyses may reveal discordant gene histories, even in parts of the tree that were well supported in previous phylogenetic analyses with fewer markers (Gagnon et al., 2021). Moreover, the expansion of sequenced genomes will allow us to isolate major genomic events, such as the amplification of repetitive content, rearrangements, and the gain and loss of coding genes, which may be tied to particular morphological or ecological transitions. In particular, the addition of the *I. cyaneum* genome will likely divide the branch between the Solanaeae (*Solanum* + *Jaltomata*) and Capsiceae (*Capsicum* + *Lycianthes*) clades, helping us to distinguish genomic variation unique to those lineages with variation that is shared because of common ancestry. In order to explore these evolutionary questions, we assembled and annotated a de novo genome for *I. cyaneum* and applied phylogenetic and comparative analyses to estimate its relationship to other Solanaceae along with historical changes in genome content.

Core Ideas

- Expanding genome sequences beyond crop species is important for understanding their evolution.
- The tomato family is an emerging model clade, with many genomes for crops and wild species.
- We assembled a reference-quality genome for a wild shrub in the tomatillo clade.
- Phylogenetic analyses including this new member of the berry clade shows deep discordance.
- This discordance will challenge efforts to connect genomic changes to morphological transitions.

2 | MATERIALS AND METHODS

2.1 | Genome sequencing and assembly

Genomic DNA was prepared from fresh leaf material of *I. cyaneum* (voucher: Smith 265 [WIS]) using the 2XCTAB protocol (Doyle & Doyle, 1987). We chose *I. cyaneum*

because it is the type of the genus and exhibits the deep violet flowers for which the genus is named (Bentham, 1845). Although native to the northern Andes, this species is widely cultivated as an ornamental with several commercial cultivars (Meerow et al., 2004; Shaw, 1998). The sequenced accession was grown from seed from cultivated material at the Missouri Botanical Garden and originally collected from the wild by W. G. D'Arcy.

Paired-end libraries with an insert size of 400 bp were sequenced on four lanes of an Illumina Hi-Seq 2000 flow cell. Mate pair libraries of 2 and 5 kb were sequenced on two lanes. Additionally, we sequenced a Hi-C library (Phase Genomics) on one lane of a Hi-Seq 4000 with 100× paired-end reads to assemble the contigs into larger scaffolds. All Illumina sequencing was completed at the Cornell Weill Genome Sequencing Facility and the numbers of reads are provided in Supplemental Table s1. Nanopore sequencing was performed on six flow cells of an Oxford Nanopore Minion device to provide an additional 5,809,839 reads. Nanopore and Illumina reads were assembled with MaSurca v3.3.2 (Zimin et al., 2013) and polished with three rounds of Pilon v1.23 (Walker et al., 2014) using Illumina reads. The Hi-C data was processed using the 3D-DNA v180922 pipeline (Dudchenko et al., 2017), and the scaffolds were manually edited in Juicebox (Dudchenko et al., 2018). Gaps were filled with LR_gapcloser (Xu et al., 2018), and Pilon was used to correct errors.

2.2 | Analysis of repeat content

We examined repetitive DNA in *I. cyaneum* and additional Solanaceae genomes for comparison. For this purpose, we downloaded assemblies for *C. annuum* cv. CM334 v.1.55 (Kim et al., 2014), *S. lycopersicum* v.4.0 (The Tomato Genome Consortium, 2012), the large white petunia [*Petunia axillaris* (Lam.) Britton et al.] v.1.6.2 (Bombarely et al., 2016), and *Nicotiana attenuata* Torr. ex S. Watson r.2.0 (Xu et al., 2017) from solgenomics.net and peppergenome.snu.ac.kr. We used LTRHarvest (Ellinghaus et al., 2008) and LTR_finder (Xu & Wang, 2007) to identify de novo putative long-terminal repeat (LTR) retrotransposons and LTR_retriever with default settings to filter the results and reduce false positives (Ou & Jiang, 2018). We then masked each genome using RepeatMasker v4.0.7 (Smit et al., 2013) with the resulting LTR library and used RepeatModeler v2.0.1 (Flynn et al., 2020) to identify additional repeats in the remaining unmasked regions of the genome. Known protein-coding sequences were excluded from the RepeatModeler library using the ProtExcluder.pl script (Campbell et al., 2014). For each genome, the LTR_retriever and RepeatModeler libraries were joined to generate a final library, which was used to mask the genome. We obtained coverage values from the RepeatMasker output, by using the fam_coverage.pl and fam_summary.pl scripts included with LTR_retriever and inputting the estimated sizes of each genome.

2.3 | Annotation

To aid in annotation, we conducted RNA sequencing on four pools of tissues: developing corollas, vegetative tissue (shoot plus root), reproductive tissue (stamen plus pistil), and seedlings from the same accession of *I. cyaneum*. Total RNA was extracted using the Spectrum Kit (Sigma-Aldrich) with on-column DNAse digestion (Qiagen). The corolla RNA was prepared with a TruSeq kit (Illumina) and sequenced with half of a lane of Hi-Seq2000 with 100-bp paired-end reads. We also carried out 454 GS-FLX Titanium sequencing (half of plate) on normalized libraries for the corolla RNA IU at Indiana University's Center for Genomics and Bioinformatics. The remaining RNAs for the other tissues were prepared with the TruSeq kit and sequenced on a single lane of HiSeq 2500 with 100-bp single reads. The 454 reads were collapsed using cd-hit v4.6.8 (Li & Godzik, 2006). Illumina and 454 reads were mapped to the genome assembly using Hisat2 v2.1.0 (Kim et al., 2015). The bam files containing mapped reads were provided as input to the BRAKER2.−2.1.5-2 pipeline (Bruna et al., 2021), which makes use of both GeneMark-ET (Lomsadze et al., 2014) and AUGUSTUS (Hoff & Stanke, 2019) for gene prediction.

Functional annotation of predicted coding genes was performed by BLASTp v2.2.31+ (Altschul et al., 1990) to the UniProt (Boutet et al., 2016) and TrEMBL (Boeckmann et al., 2003) databases using an e-value cut off of 1×10^{-20}. We also removed any predicted proteins both with few to no mapped reads (FPKM < 0.01) and which had no hits within the NCBI NR, tomato, or pepper databases. Protein domains were predicted with InterProScan v5.46-81.0 (Jones et al., 2014) and genes labeled as transposons were discarded. BUSCO v3 analysis (Simão et al., 2015), with the Embryophyta dataset, was used to quantify genome and annotation content and examine the completeness of the genome assembly and annotation in comparison with other published genomes. We used OrthoFinder v2.5.2 (Emms & Kelly, 2015) to identify groups of orthologous genes shared between *I. cyaneum*, pepper, tomato, and robusta coffee (*Coffea canephora* Pierre ex A. Froehner). For pepper and tomato, we used the same genome assemblies as cited above and for coffee, we used *Coffea canephora* v.1.0 (Denoeud et al., 2014). These results were used to create a Venn diagram depicting shared and unique gene clusters across taxa.

Finally, we used maximum likelihood methods to identify significantly expanded and contracted gene families in *I. cyaneum*. For these analyses, we expanded our sampling to include all the tips that were present in the phylogenetic

analysis (see below). Again, we used OrthoFinder to identify groups of orthologous genes found in one or more of the species. We input these gene families from Orthofinder and the species tree (see below) into CAFE v.3.0 (Han et al., 2013). Before inputting, the tree was ultrametricized with penalized likelihood using the chronopl() function in the R package APE (Paradis et al., 2004). For the gene families showing significant expansion and contraction ($p < .05$) in *I. cyaneum*, we conducted BLAST searches to examine their possible functions. We extracted the two longest sequences from each expanded or contracted orthogroup in *I. cyaneum* and ran BLAST searches using DIAMOND BLASTp v0.9.30.131 (Buchfink et al., 2015). We kept the top hits for each of those sequences and retrieved the list of gene ontology (GO) terms for them with InterProScan. The resulting list of expanded or contracted *I. cyaneum* orthogroups and their associated GO terms was input to topGO (Alexa & Rahnenfuhrer, 2021) for enrichment analyses. We searched for enrichment in GO terms associated with biological functions and used Fisher's exact test to determine significance.

2.4 | Phylogeny estimation

We investigated the phylogenetic relationship of *I. cyaneum* to other Solanaceae using Bayesian concordance analysis (Ane et al., 2007; Baum, 2007). This approach estimates the population or species tree with branch lengths in coalescent units using quartet methods along with the proportion of the genome that supports each clade in this tree (Larget et al., 2010). We included seven other species of Solanaceae (large white petunia [*P. axillaris* (Lam.) Britton, Sterns & Poggenb.], *N. attenuata*, potato, tomato, eggplant, and pepper) plus little-bell (*Ipomoea triloba* L.) (Convolvulaceae) (Wu et al., 2018) and robusta coffee (Rubiaceae) as outgroups. We chose these taxa based on the availability of reference-quality annotated genomes at the time of dataset assembly. For the Solanaceae genomes, we used the same assembly versions and sources as listed above for gene family analyses. For species tree estimation, we first generated posterior distributions of gene trees for the 1355 single-copy genes from the Orthofinder analysis that were present in all genomes (zero missing data). Each protein alignment was run in MrBayes v3.2.7a (Ronquist & Huelsenbeck, 2003) for two million generations, sampling every 100 generations, with a mixed prior on amino acid models, an exponential prior on branch lengths with mean set to 0.001, and a gamma distribution for rate heterogeneity across sites with an estimated proportion of invariant sites. Convergence was assessed with the potential scale reduction factor, which was near 1.0 for all model parameters for all genes. We removed the first 5,000 trees as burn-in and summarized the remaining sample from the posterior with the mbsum program in BUCKy 1.4.4 (Larget et al., 2010). We estimated the population tree and the concordance factors (CFs) in BUCKy with four Markov chain Monte Carlo chains, each of one million steps and the initial value for the discordance parameter, alpha, set to 1. The results of the concordance analysis were summarized as a population tree with branch lengths in coalescent units, rooted on the outgroup taxa, and CFs with credibility intervals for each clade.

2.5 | Synteny analysis

In order to assess patterns of synteny between *I. cyaneum* and closely related crop genomes, we first created whole-genome alignments with NUCmer v3.1, part of the MUMmer software (Kurtz et al., 2004). For visualization, the alignments were filtered to select one-to-one aligned segments with a minimum length of maximal exact matches of 2,000, as well as either a minimum alignment identity of 88, in the case of *I. cyaneum* compared with pepper, or of 85, in the case of *I. cyaneum* to tomato and pepper to tomato. The coordinates of the filtered alignments were then input as links to generate plots using Circos v0.69-6 (Krzywinski et al., 2009). We used tomato as a benchmark for numbering and orienting the *I. cyaneum* pseudomolecules.

3 | RESULTS

3.1 | Genome assembly and annotation

The length of our de novo sequence assembly for *I. cyaneum* is 2.7 Gb, making it very similar to pepper (Table 1). This assembled size for *I. cyaneum* is slightly smaller than the size previously estimated from flow cytometry, ~ 3.2 Gb (Gates et al., 2016). Our chromosome-level assembly (Supplemental Figure s1) was quite similar to *C. annuum*, with 84% of the assembly anchored, and our sequencing strategy resulted in a lower percentage of N bases and gaps (Table 1). Although the genomes of *I. cyaneum* and pepper are over three times the size of those in sequenced *Solanum* species (Bolger et al., 2014; Hirakawa et al., 2014; The Tomato Genome Consortium, 2012), we recovered similar numbers of annotated genes (Table 1). Our annotation for *I. cyaneum* includes 92% of the highly conserved benchmarking universal single-copy orthologs (BUSCOs). Overall, the BUSCO analysis showed few fragmented or missing BUSCOs (Figure 1), suggesting that the quality of the genome is on par with those of related economically important plants. In addition to these highly conserved orthologous genes, we found a large number of unique gene clusters in *I. cyaneum*, nearly twice those found in tomato or pepper (Figure 1).

Our CAFE analyses revealed a strong bias toward gene family expansion in *I. cyaneum*. A total of 1,959 gene families

TABLE 1 Summary statistics for *Iochroma cyaneum* genome assembly compared with closely related Solanaceae

Summary statistic	*Iochroma cyaneum*	*Capsicum annuum*	*Solanum lycopersicum*
Genome assembly total length, Mb	2,716.02	2,633.68	782.52
Percentage of assembly assigned to chromosomes	84.13	86.00	98.77
No. of contigs	37,881	117,244	448
Contig N50, kb	212.94	55.87	6,007.83
Longest contig, kb	3,996.25	608.96	26,291.69
No. of N bases, Mb	0.64	78.12	0.04
No. of gaps	19,176	217,286	435
No. of genes	38,625	34,903	34,075
Repeat percentage of genome, %	69.35	72.26	58.30

Note. Values for assembly length, number of N bases, and number of gaps based on currently available assemblies on SolGenomics.net (SL4.0 for tomato and v.1.55 for pepper) calculated with assembly-stats 0.1.4 (Trizna, 2020). Contig statistics were calculated with the same tool after splitting the assemblies at Ns. Remaining values estimated during the comparative repeat analyses (Figure 3) or, for annotation information, gathered from the literature (Hosmani et al., 2019; Kim et al., 2014).

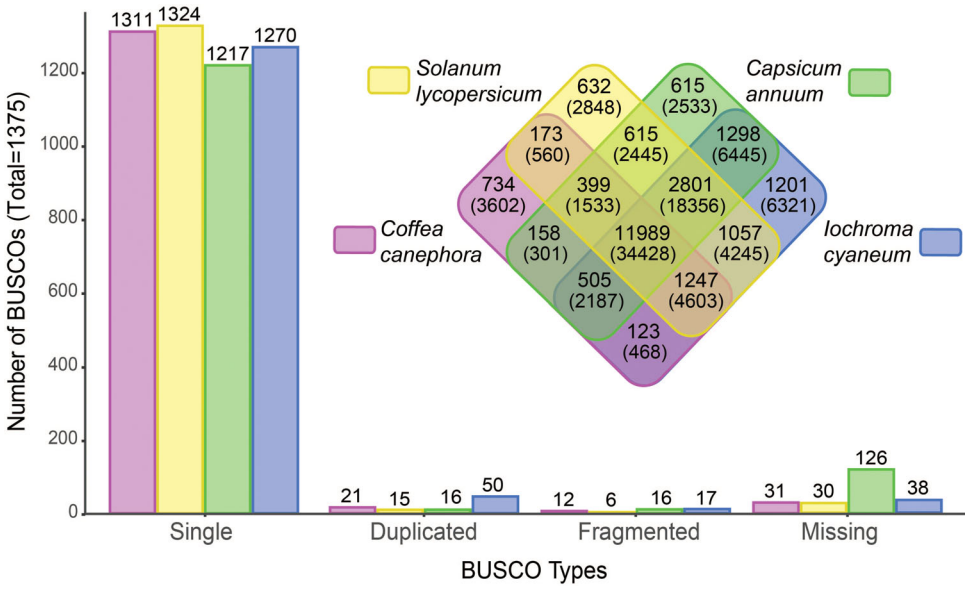

FIGURE 1 Comparison of *Iochroma cyaneum* annotation to related crop genomes. Bar graph shows the results of the BUSCO analysis with coffee, tomato, pepper, and *I. cyaneum*, left to right, for each BUSCO type. The numbers of genes in each category are shown at the top of each bar. Inset is a Venn diagram showing the results of the orthogroup analysis with unique and shared clusters shown for each species. The total numbers of genes in each orthogroup are shown in parentheses

had a significant change in size along the *I. cyaneum* branch ($p < .05$) with 654 contracted and 1,305 expanded (Supplemental Table s2). The contracted families were spread across a range of biological processes with the most significant enrichment in ribonucleoprotein complex assembly ($p = .0043$; Supplemental Figure s2). By contrast, the most highly enriched GO terms for the expanded gene families were all related to pollen recognition ($p = .00037$; Supplemental Figure s3). We used BLAST searches to determine the identity of the nine expanded families with this GO term, and all appear to be G-type lectin S-receptor-like serine/threonine-protein kinases (Supplemental Table s3).

3.2 | Phylogeny

Our phylogenetic analysis recovered the core relationships among lineages of Solanaceae that have been estimated in previous studies (Olmstead et al., 2008; Särkinen et al., 2013). *Nicotiana* is sister to the large fleshy-fruited clade containing tomato, potato, eggplant, *Jaltomata*, pepper, and *I. cyaneum* (CF = 0.72; Figure 2a). Together, they form the x=12 clade, united by the base chromosome number of 12 (Olmstead & Palmer, 1992). We find strong agreement across the 1,355 genes for all the relationships within *Solanum* (CF = 0.91–0.99), but less so among the other fleshy-fruited species. For example, the estimated proportion of the genome for which

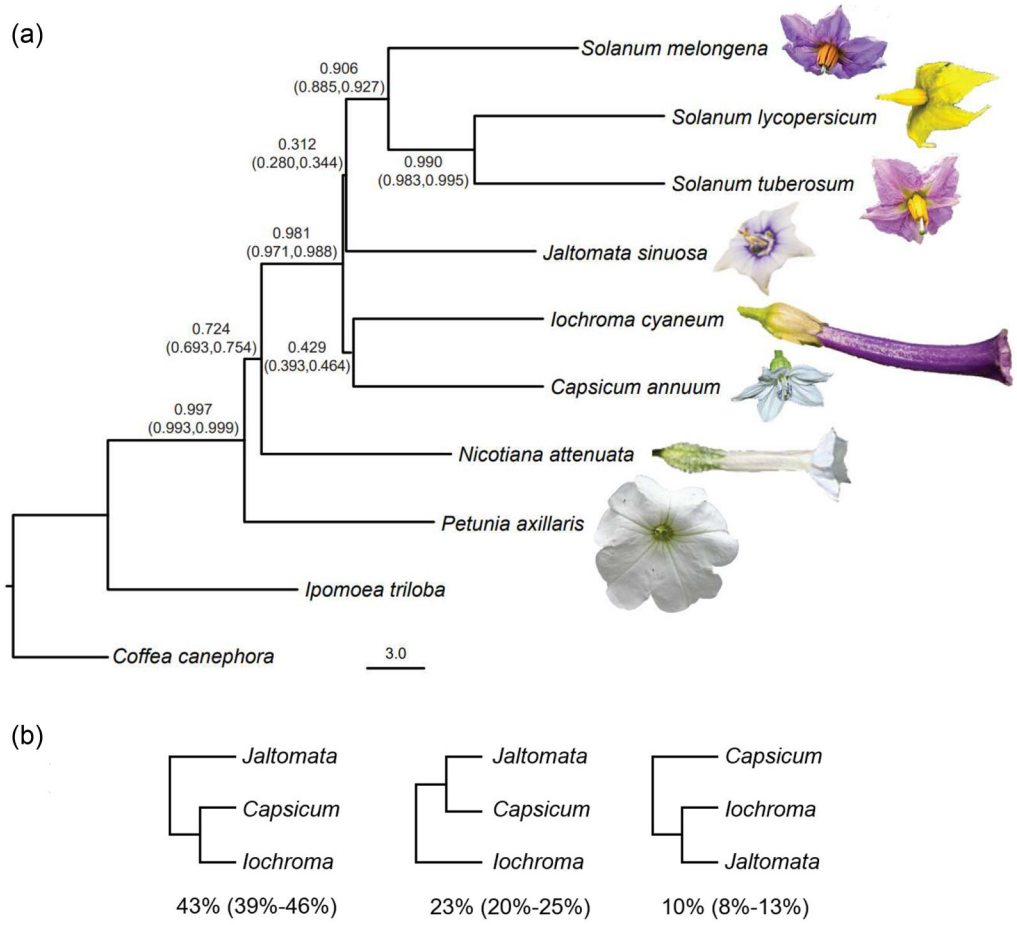

FIGURE 2 Phylogenetic position of *Iochroma cyaneum*. (a) Population tree for Solanaceae estimated with BUCKy. Branch lengths are in coalescent units, and branches are annotated with the estimated genome-wide concordance factors (with credibility intervals in parentheses). Each concordance factor corresponds to the proportion of the genome estimated to have the clade in its history. Photos are from Wikimedia commons with the exception of *Jaltomata sinuosa* (image from Thomas Mione, Central Connecticut State University). (b) Genome-wide variation in the relationships among *Jaltomata*, *Iochroma*, and *Capsicum*. Concordance factors (and their credibility intervals) are shown as percentages

the true tree places *Capsicum* sister to *I. cyaneum* is 0.43 and there is even less agreement regarding the placement of *Jaltomata*. Indeed, the population tree shown in Figure 2 varies from the primary concordance tree in *Jaltomata*'s position, putting it instead sister to *Capsicum* + *Iochroma* with a CF of 0.32 with an overlapping credibility interval (0.287–0.353) (Supplemental Table s4). We also estimate a sizeable proportion (23%) of the genome supporting a *Jaltomata* + *Capsicum* relationship (Figure 2b) and 19% placing *Capsicum* closer to *Solanum* than to *Iochroma* (Supplemental Table s4). Overall, these analyses point to significant discordance along the backbone of the berry clade, with large numbers of loci supporting alternate relationships to those in the population tree.

3.3 | Repetitive content in *Iochroma cyaneum*

Our analyses show that the *I. cyaneum* genome comprises largely repetitive content as in other Solanaceae and indeed in most plant genomes (Feschotte et al., 2002). Only 31% of the *I. cyaneum* genome is nonrepetitive, which is slightly more than *Capsicum* and *Jaltomata* but less than the other genomes analyzed (Figure 3a; Supplemental Table s5). Despite being closely related and sharing similar percentages of repetitive DNA, the composition of the repeats varies markedly between *I. cyaneum* and *Capsicum*. In *I. cyaneum*, *Gypsy* elements account for the majority of the repetitive content (75%) and over half (52%) of the entire genome. The other types of elements have contracted in *I. cyaneum*, which has a smaller proportion of *Copia* elements among its LTR repeats than any other Solanaceae examined (Supplemental Table s5). In this context, all the lineages have a significant fraction of repetitive elements that cannot be classified either within interspersed repeats or as a type of LTR specifically. Nonetheless, as the same pipeline was applied to all taxa, the estimated variation in the fraction of each element in the genome points to substantial macroevolutionary shifts in the composition of the repetitive DNA.

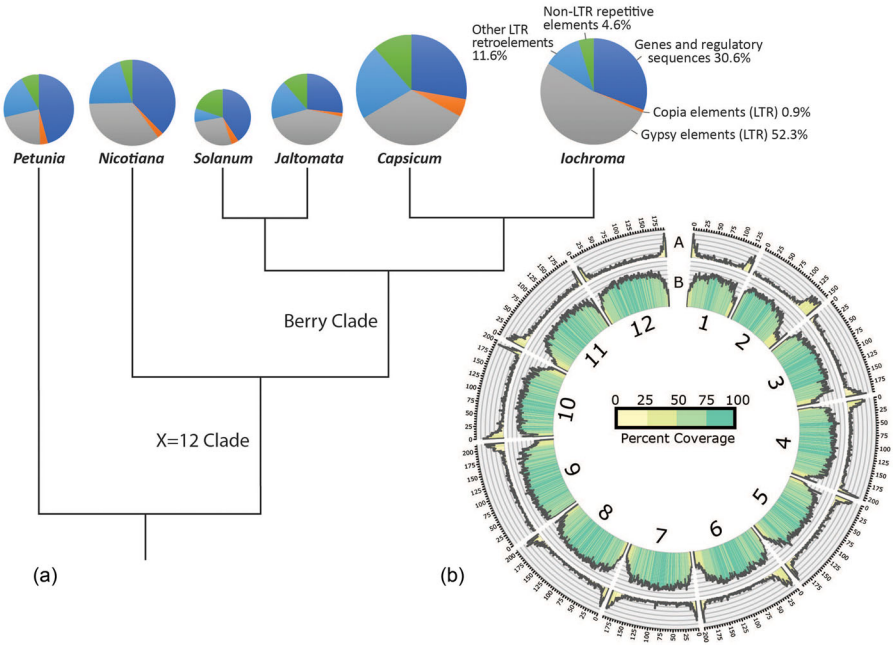

FIGURE 3 Repetitive content in *Iochroma cyaneum* and related Solanaceae. (a) Phylogenetic relationships from Figure 2. The pie charts for each species are proportional to genome size. The other long-terminal repeat (LTR) retroelements category includes caulimovirus, ERK, and unknown retroelements and the non-LTR elements category includes long interspersed nuclear elements, DNA elements, simple and low complexity repeats, and other unclassified repetitive elements (see Supplemental Table s2). (b) The distribution of repetitive content across *Iochroma* chromosomes. The inside ring shows the percentage of repetitive content in each 1-Mb window and the outside ring shows the percentage of annotated genic content in that window, where the gray lines denote 10% increments. Each genomic window is colored to show the percent coverage of repetitive content in that window as indicated by the legend in the center. Chromosomes are numbered and ordered following patterns of synteny with tomato (Figure 4). The length of each chromosome is shown with the outermost ring in units of mega bases

We also examined how this repetitive content was distributed along chromosomes within the *I. cyaneum* genome. We found that the nonrepetitive genic regions are clustered at the very ends of the chromosomes while the centromeric regions tended to be less gene rich and more repetitive (Figure 3b). While most chromosomes have genic regions at either end, two of them (chromosomes 2 and 9) have only a single cluster at one end. This chromosomal organization (with repetitive DNA most dense at the center and coding regions at the distal ends) is common for plant genomes and was also observed in *Capsicum* (Kim et al., 2014).

3.4 | Collinearity between *Iochroma cyaneum* and other Solanaceae

Despite the large difference in genome size between *I. cyaneum* and tomato, we found strong synteny for much of the genome. Most *I. cyaneum* chromosomes (1, 2, 4, and 6–10) were easily aligned to tomato, having only small structural arrangements between the two taxa. For example, the content of *I. cyaneum* chromosome 9 closely matches that of tomato chromosome 9, although a few areas that match more highly to sectors of tomato chromosomes 1 and 11 (Figure 4). We did, however, observe some connections that indicate major rearrangements between the two taxa. In one clear case, the roughly 20 Mb at 3′ end of tomato chromosome 4 is highly syntenic with the 5′ end of *I. cyaneum* chromosome 11, suggesting a translocation event (Figure 4). This relationship between chromosomes 4 and 11 is apparent in our synteny analysis of *I. cyaneum* and pepper (Supplemental Figure s4) but not pepper and tomato (Supplemental Figure s5), which is consistent with a translocation event specific to the *I. cyaneum* branch of the phylogeny. In fact, visual comparison of the two synteny maps (tomato vs. *I. cyaneum*; Figure 4, and tomato vs. pepper, Supplemental Figure s5) points to no major shared rearrangements in *I. cyaneum* and *Capsicum*, suggesting that instead, most of the translocations and inversions are lineage specific. This result is consistent with the relatively short internal branch uniting these two genera (Figure 2).

4 | DISCUSSION

The family Solanaceae has witnessed an explosion in whole-genome sequencing accompanied by efforts to expand beyond crop species into wild relatives (Bolger et al., 2014; Cao et al.,

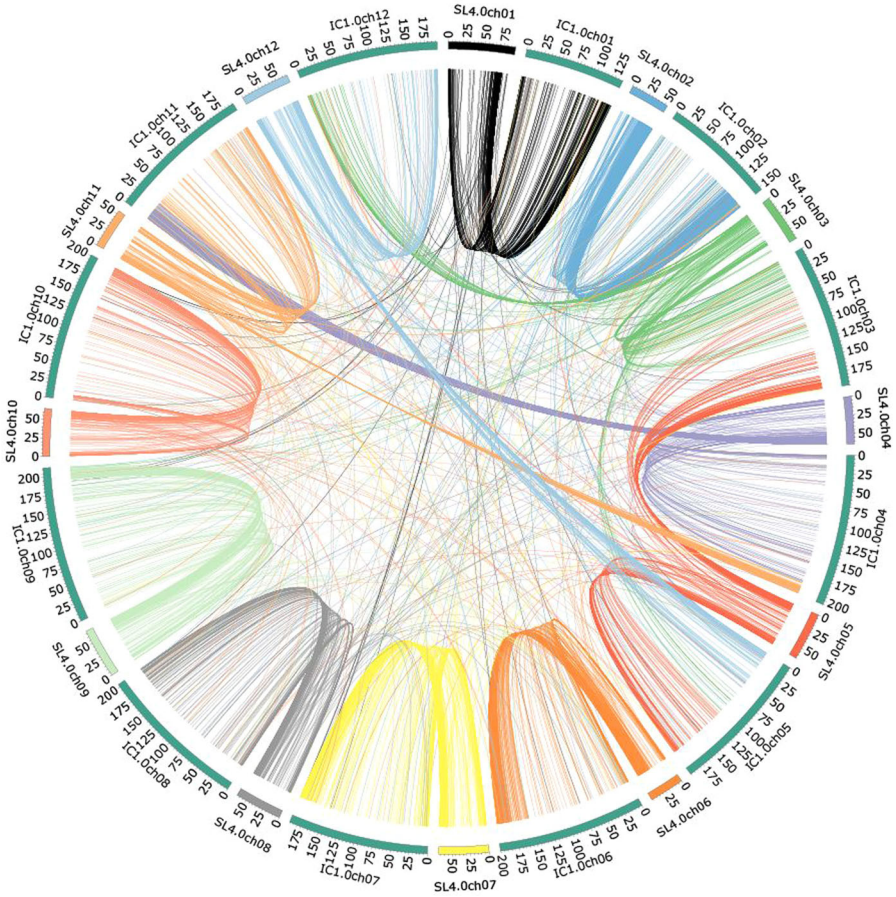

FIGURE 4 Patterns of synteny between tomato and *Iochroma cyaneum*. Tomato and *I. cyaneum* chromosomes are shown with lines connecting syntenic segments. Line coloring follows tomato. The length of each chromosome is marked in 25-Mb increments

2021; Wu et al., 2018). Analyses of these new genomes have solidified aspects of the family's evolutionary history, such as the whole-genome triplication at the base of the family (Bombarely et al., 2016; Cao et al., 2021; The Tomato Genome Consortium, 2012) while also revealing the complexities of the phylogenetic relationships and genomic rearrangements (Barchi et al., 2019). As the first member of the tomatillo subfamily (Physalideae) with a chromosome-level assembly, our analysis of the *I. cyaneum* genome brings new insights regarding the radiation of the berry clade and the accompanying changes in genome size, content, and organization.

4.1 | Discordance along the berry clade backbone

Phylogenetic analyses, including *I. cyaneum* together with seven other Solanaceae, point to significant discordance within the berry-fruited clade Solanoideae. This clade includes pepper and its allies (Capsiceae), tomatillo and its allies (Physalideae), and the large genus *Solanum* and its sister genus *Jaltomata* (Solaneae). Recent family-level analyses with plastid and nuclear markers have shown strong support for the dominant relationship, with *Capsicum* more closely related to *Physalis* and *Jaltomata* sister to *Solanum* (Olmstead et al., 2008; Särkinen et al., 2013). Nevertheless, alternative relationships have often appeared in phylogenetic analyses (Bohs & Olmstead, 1997; Olmstead et al., 1999; Smith & Baum, 2006), and previous phylogenomic analyses suggest extensive discordance involving *Capsicum* and *Jaltomata* (Wu et al., 2018; Wheeler et al., 2022). Our Bayesian concordance analysis expands the scope of this discordance, as the relationship of *I. cyaneum* to these two taxa is also highly variable across the genome. Following previous family-level studies (Olmstead et al., 2008; Särkinen et al., 2013), we expected *I. cyaneum* to be most closely related to *Capsicum*, and indeed, 43% of the genes in the genome are estimated to follow this dominant history (Figure 2a). However, many genes show alternative resolutions, that is, with *Capsicum* sister to *Jaltomata* (22%) or *Jaltomata* sister to *Iochroma* (10%) (Figure 2b). Meanwhile, the position of *Jaltomata* is nearly evenly split across gene trees between appearing as sister to *Solanum* (31%) vs. sister to *Capsicum* + *Iochroma* (32%). These patterns contrast with other nodes in the tree (e.g.,

the common ancestor of *Solanum*, the common ancestor of Solanaceae), where nearly all genes share the same underlying history. The high discordance along the backbone of the berry clade may reflect a range of evolutionary processes including hybridization and introgression or incomplete lineage sorting because of rapid radiation or large population sizes (Maddison, 1997). In the case of *I. cyaneum*, the large difference between the dominant history (43% for *Capsicum* sister) and the minor histories (22 and 10%) is most consistent with incomplete lineage sorting (Baum, 2007). Expanding the phylogenomic analysis to include other major lineages of the large and diverse berry clade (~2,000 species) would be valuable to distinguish among these possible causes.

4.2 | Gene family evolution in *Iochroma cyaneum*

Although quite similar in total genome size, our annotation pipeline retrieved more gene models in *I. cyaneum* than were estimated in pepper (38.6 vs. 34.9 K, Table 1), and we estimate a slightly higher proportion of nonrepetitive (including genic) content in *I. cyaneum* (30.6 vs. 27.7%). Consistent with the possibility of gene family expansion along the *I. cyaneum* lineage, the orthogroup analysis recovered a larger number of unique orthogroups compared with pepper and more genes in those orthogroups (Figure 1). Using maximum likelihood birth–death models, we estimated significant expansions in 1,305 gene families (Supplemental Table s2), and we found that these families were enriched for function in pollen recognition (Supplemental Figure s3). The BLAST searches suggest that these orthogroups, which are significantly expanded in *I. cyaneum* and involved in pollen recognition, are G-type lectin S-receptor-like serine/threonine-protein kinases. Receptor kinases are known to play an important role in sporophytic self-incompatibility in the Brassiceae, but they have not been documented to be involved in pollen recognition in species with gametophytic self-incompatibility like Solanaceae (Kachroo et al., 2001; McCubbin & Kao, 2000). Beyond pollination, these G-type lectin receptor-like kinases are known to be involved in other aspects of signaling, in particular, mediating responses to insect attacks (Gilardoni et al., 2011). Plant–insect interactions have emerged as major drivers of genome evolution, especially in Solanaceae (De-la-Cruz et al., 2021; Fan et al., 2020), and our findings from *I. cyaneum* suggest that lectin receptor-like kinases merit additional investigation as mediators of these interactions (Sun et al., 2020).

4.3 | Diversity and distribution of repetitive DNA

With a genome estimated at 3.2 Gb with flow cytometry (Gates et al., 2016) and 2.7 Gb in our reference assembly (Table 1), *I. cyaneum* presents the largest diploid genome sequenced in the Solanaceae thus far and is most similar in size to pepper. The large size of the pepper genome compared with tomato was attributed to the expansion of repetitive content and, in particular, to LTR retroelements (Kim et al., 2014). Using a single pipeline for six Solanaceae species, we estimated that the proportion of the genome occupied by LTRs in *I. cyaneum* is even higher than in pepper and roughly 1.5 times that in tomato (Supplemental Table s5). We also uncovered a high turnover in the type of LTR retrotransposon in *I. cyaneum*, which has much higher proportion of *Gypsy* elements compared with pepper (81 vs. 55%) and a correspondingly smaller proportion of the other classes of retroelements (Figure 3; Supplemental Table s5). Thus, while the genomes of these species are both composed of over 60% LTR retrotransposons, the individual classes of element have shifted dramatically in frequency possibly because of repeated rounds of transposable element expansion and contraction (i.e., the genomic 'accordion'; Kapusta et al., 2017). Although LTR retrotransposons, like other transposable elements, seem to be largely inactive (Feschotte et al., 2002), lineage-specific amplification and contractions are often uncovered in comparative genomic analyses in plants (e.g., Lee et al., 2017; Zhang et al., 2019). Whole-genome duplications and hybridization events are hypothesized to trigger transposable element proliferation (Wendel et al., 2016), offering an intriguing area for future research given the apparent frequency of hybridization in Iochrominae (Smith & Baum, 2006) and possibly more broadly in the tomatillo clade (Zamora-Tavares et al., 2016).

As in many plant genomes, we also found that the repetitive content in the *I. cyaneum* genome occurs in the centers of the chromosomes with genic regions clustered at the tips (Figure 3b). This organization is common to plants with metacentric chromosomes, and the repetitive content plays a key role in coordinating chromosome movement during meiosis and mitosis (Nagaki et al., 2003; Zhong et al., 2002). All 12 chromosomes of *I. cyaneum* are metacentric, and such highly symmetric karyotypes are typical in the genus (Deanna et al., 2018). Tomato and pepper share this karyotype (mostly or all metacentric; Chiarini et al., 2018) and, in turn, this chromosomal organization, with an expanse of repetitive content at the center and gene-rich content only near the ends (Jouffroy et al., 2016; Kim et al., 2014).

Despite their similarity in genome organization, patterns of synteny between these three taxa suggest several major rearrangements. The comparison of tomato and *I. cyaneum* revealed regions of up to 50 Mb with disrupted synteny, likely because of translocations toward the ends of chromosomes 4, 5, 11, and 12 (Figure 4). Given that *I. cyaneum* is more closely related to pepper than to tomato, we expected fewer rearrangements between them, but instead observed less synteny than with tomato (Supplemental Figure s4). These results suggest that genomic events, such as large translocations, inversions, and deletions, are frequent at this ~20-million-yr

intergeneric-scale (Barchi et al., 2019) and that a much denser taxon sampling will be needed to infer the order and timing of any particular event. The addition of a high-quality reference genome for *Physalis* (Lemmon et al., 2018) will aid in determining which of the rearrangements that appear distinct to *I. cyaneum* are in fact shared more widely across the tomatillo clade. Karyotypic analyses across Physalideae point to several shifts in chromosome size, symmetry, and number that can help to guide taxon sampling (Deanna et al., 2018; Rodriguez et al., 2020). With more targeted sampling across the berry clade, together with the development of new comparative genomic tools (e.g., GENESPACE; Lovell et al., 2018), we may look toward building a berry core genome that captures the shared elements in the fleshy-fruited common ancestor as well as a pangenome that spans the genomic diversity of the clade.

5 | CONCLUSIONS

With clusters of genomes emerging around crop species of Solanaceae, our challenge now is to expand in terms of phylogenetic diversity using wild species to span the connections between these clusters. The berry clade of Solanaceae comprises roughly 50 genera (Hunziker, 2001), but the 20 genomes sequenced thus far include only five of these. As a member of the tomatillo clade, the addition of *I. cyaneum* splits the evolutionary path between pepper and tomato with slightly closer affinity to pepper. Nevertheless, our phylogenetic analyses reinforce and expand the findings of Wu et al. (2018), namely that the relationships among berry clade genera are highly discordant across the genome. This discordance has important implications for downstream applications of these comparative genomics resources. For example, the genes that underlie traits of interest, such as fruit characteristics or secondary metabolites, may not follow the inferred species tree, potentially leading to incorrect inferences about the number and timing of evolutionary transitions (Hahn & Nakhleh, 2016). Moreover, the disagreement about relationships means there is no clear sister group for genomic comparison with crop-containing genera (*Solanum*, *Capsicum*). Instead, functional comparative studies will need to make use of the suite of sequenced berry clade genomes to reconstruct gene histories and dissect the origins of mutations with functional consequences (Martin & Orgogozo, 2013). Adding genomic resources for other genera is unlikely to resolve the deeply discordant backbone of the berry species tree but will allow us to build a more complete picture of the evolutionary diversification of this economically important clade of plants.

ACKNOWLEDGMENTS

Funding for this work was provided by the National Science Foundation (NSF-DEB 1413855 and 1355518 to SDS) and start-up funds from the University of Nebraska–Lincoln to SDS. SDS also thanks Lynn Bohs for providing seed of this accession of *I. cyaneum*. Research conducted for this manuscript complies with the ethical rules applicable for this journal, as stated in the Instructors for Authors.

AUTHOR CONTRIBUTIONS

Adrian F. Powell: Formal analysis; Investigation; Methodology; Visualization; Writing – review & editing. Jing Zhang: Data curation; Formal analysis; Methodology. Duncan Hauser: Methodology; Resources. Julianne A. Vilela: Formal analysis; Investigation. Alice Hu: Formal analysis; Investigation. Daniel J. Gates: Conceptualization; Data curation. Lukas A. Mueller: Conceptualization. Fay-Wei Li: Resources; Supervision. Susan R. Strickler: Conceptualization; Data curation; Formal analysis; Methodology; Project administration; Supervision; Writing – review & editing. Stacey D. Smith: Conceptualization; Formal analysis; Funding acquisition; Investigation; Methodology; Project administration; Resources; Supervision; Visualization; Writing – original draft; Writing – review & editing.

CONFLICT OF INTEREST

The authors declare that they have no conflict of interest.

DATA AVAILABILITY STATEMENT

Raw sequencing reads used in the assembly of the genome are available from the NCBI database under BioProject PRJNA777841. The completed genome assembly and annotation files are available on the Sol Genomics Network website (https://solgenomics.net).

REFERENCES

Alexa, A., & Rahnenfuhrer, J. (2021). topGO: Enrichment Analysis for Gene Ontology. R package v2.45.0.

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., & Lipman, D. J. (1990). Basic Local Alignment Search Tool. *Journal of Molecular Biology*, *215*, 403–410. https://doi.org/10.1016/S0022-2836(05)80360-2

Ane, C., Larget, B., Baum, D. A., Smith, S. D., & Rokas, A. (2007). Bayesian estimation of concordance among gene trees. *Molecular Biology and Evolution*, *24*, 412–426. https://doi.org/10.1093/molbev/msl170

Aversano, R., Contaldi, F., Ercolano, M. R., Grosso, V., Iorizzo, M., Tatino, F., Xumerle, L., Dal Molin, A., Avanzato, C., Ferrarini, A., Delledonne, M., Sanseverino, W., Cigliano, R. A., Capella-Gutierrez, S., Gabaldón, T., Frusciante, L., Bradeen, J. M., & Carputo, D. (2015). The *Solanum commersonii* genome sequence provides insights into adaptation to stress conditions and genome evolution of wild potato relatives. *Plant Cell*, *27*, 954–968. https://doi.org/10.1105/tpc.114.135954

Barchi, L., Pietrella, M., Venturini, L., Minio, A., Toppino, L., Acquadro, A., Andolfo, G., Aprea, G., Avanzato, C., Bassolino, L., Comino, C., Molin, A. D., Ferrarini, A., Maor, L. C., Portis, E., Reyes-Chin-Wo, S., Rinaldi, R., Sala, T., Scaglione, D., … Rotino, G. L. (2019). A chromosome-anchored eggplant genome sequence reveals key events

in Solanaceae evolution. *Scientific Reports*, *9*, 11769. https://doi.org/10.1038/s41598-019-47985-w

Baum, D. A. (2007). Concordance trees, concordance factors, and the exploration of reticulate genealogy. *Taxon*, *56*, 417–426. https://doi.org/10.1002/tax.562013

Bentham, G. (1845). *Iochroma tubulosa*. *Edwards's Botanical Register*, *31*, t. 20.

Boeckmann, B. (2003). The SWISS-PROT protein knowledgebase and its supplement TrEMBL in 2003. *Nucleic Acids Research*, *31*, 365–370. https://doi.org/10.1093/nar/gkg095

Bohs, L., & Olmstead, R. G. (1997). Phylogenetic relationships in *Solanum* (Solanaceae) based on *ndhF* sequences. *Systematic Botany*, *22*, 5–17. https://doi.org/10.2307/2419674

Bolger, A., Scossa, F., Bolger, M. E., Lanz, C., Maumus, F., Tohge, T., Quesneville, H., Alseekh, S., Sørensen, I., Lichtenstein, G., Fich, E. A., Conte, M., Keller, H., Schneeberger, K., Schwacke, R., Ofner, I., Vrebalov, J., Xu, Y., Osorio, S., … Fernie, A. R. (2014). The genome of the stress-tolerant wild tomato species *Solanum pennellii*. *Nature Genetics*, *46*, 1034–1038. https://doi.org/10.1038/ng.3046

Bombarely, A., Moser, M., Amrad, A., Bao, M., Bapaume, L., Barry, C. S., Bliek, M., Boersma, M. R., Borghi, L., Bruggmann, R., Bucher, M., D'agostino, N., Davies, K., Druege, U., Dudareva, N., Egea-Cortines, M., Delledonne, M., Fernandez-Pozo, N., Franken, P., … Kuhlemeier, C. (2016). Insight into the evolution of the Solanaceae from the parental genomes of *Petunia hybrida*. *Nature Plants*, *2*, 16074. https://doi.org/10.1038/nplants.2016.74

Boutet, E., Lieberherr, D., Tognolli, M., Schneider, M., Bansal, P., Bridge, A. J., et al., & Xenarios, I. (2016). UniProtKB/Swiss-Prot, the manually annotated section of the UniProt KnowledgeBase: How to use the entry view. *High-Throughput Next Generation Sequencing: Methods and Application*, *1374*, 23–54.

Brůna, T., Hoff, K. J., Lomsadze, A., Stanke, M., & Borodovsky, M. (2021). BRAKER2: Automatic eukaryotic genome annotation with GeneMark-EP+ and AUGUSTUS supported by a protein database. *NAR Genomics and Bioinformatics*, *3*, lqaa108. https://doi.org/10.1093/nargab/lqaa108

Buchfink, B., Xie, C., & Huson, D. H. (2015). Fast and sensitive protein alignment using DIAMOND. *Nature Methods*, *12*, 59–60. https://doi.org/10.1038/nmeth.3176

Campbell, M. S., Law, M., Holt, C., Stein, J. C., Moghe, G. D., Hufnagel, D. E., Lei, J., Achawanantakun, R., Jiao, D., Lawrence, C. J., Ware, D., Shiu, S.-H., Childs, K. L., Sun, Y., Jiang, N., & Yandell, M. (2014). MAKER-P: A tool kit for the rapid creation, management, and quality control of plant genome annotations. *Plant Physiology*, *164*, 513–524. https://doi.org/10.1104/pp.113.230144

Cao, Y.-L., Li, Y.-L., Fan, Y.-F., Li, Z., Yoshida, K., Wang, J.-Y., Ma, X.-K., Wang, N., Mitsuda, N., Kotake, T., Ishimizu, T., Tsai, K.-C., Niu, S.-C., Zhang, D., Sun, W.-H., Luo, Q., Zhao, J.-H., Yin, Y., Zhang, B., … Liu, Z.-J. (2021). Wolfberry genomes and the evolution of *Lycium* (Solanaceae). *Communications Biology*, *4*, 671. https://doi.org/10.1038/s42003-021-02152-8

Cárdenas, P. D., Sonawane, P. D., Heinig, U., Bocobza, S. E., Burdman, S., & Aharoni, A. (2015). The bitter side of the nightshades: Genomics drives discovery in Solanaceae steroidal alkaloid metabolism. *Phytochemistry*, *113*, 24–32. https://doi.org/10.1016/j.phytochem.2014.12.010

Chiarini, F., Sazatornil, F., & Bernardello, G. (2018). Data reassessment in a phylogenetic context gives insight into chromosome evolution in the giant genus *Solanum* (Solanaceae). *Systematics and Biodiversity*, *16*, 397–416. https://doi.org/10.1080/14772000.2018.1431320

Deanna, R., Larter, M. D., Barboza, G. E., & Smith, S. D. (2019). Repeated evolution of a morphological novelty: A phylogenetic analysis of the inflated fruiting calyx in the Physalideae tribe (Solanaceae). *American Journal of Botany*, *106*, 270–279. https://doi.org/10.1002/ajb2.1242

Deanna, R., Smith, S. D., Särkinen, T., & Chiarini, F. (2018). Patterns of chromosomal evolution in the florally diverse Andean clade Iochrominae (Solanaceae). *Perspectives in Plant Ecology Evolution and Systematics*, *35*, 31–43. https://doi.org/10.1016/j.ppees.2018.09.004

De-la-Cruz, I. M., Hallab, A., Olivares-Pinto, U., Tapia-López, R., Velázquez-Márquez, S., Piñero, D., Oyama, K., Usadel, B., & Núñez-Farfán, J. (2021). Genomic signatures of the evolution of defence against its natural enemies in the poisonous and medicinal plant *Datura stramonium* (Solanaceae). *Science Reports*, *11*, 882. https://doi.org/10.1038/s41598-020-79194-1

Denoeud, F., Carretero-Paulet, L., Dereeper, A., Droc, G., Guyot, R., Pietrella, M., Zheng, C., Alberti, A., Anthony, F., Aprea, G., Aury, J.-M., Bento, P., Bernard, M., Bocs, S., Campa, C., Cenci, A., Combes, M.-C., Crouzillat, D., Da Silva, C., … Lashermes, P. (2014). The coffee genome provides insight into the convergent evolution of caffeine biosynthesis. *Science*, *345*, 1181–1184. https://doi.org/10.1126/science.1255274

DeWitt Smith, S., & Baum, D. A. (2007). Systematics of Iochrominae (Solanaceae): Patterns in floral diversity and interspecific crossability. *Acta Horticulturae*, *745*, 241–254. https://doi.org/10.17660/ActaHortic.2007.745.10

Donoghue, M. J., & Edwards, E. J. (2019). Model clades are vital for comparative biology, and ascertainment bias is not a problem in practice: A response to Beaulieu and O'Meara (2018). *American Journal of Botany*, *106*, 327–330. https://doi.org/10.1002/ajb2.1255

Doyle, J. J., & Doyle, J. L. (1987). A rapid DNA isolation procedure from small quantities of fresh leaf tissues. *Phytochemical Bulletin*, *19*, 11–15.

Dudchenko, O., Batra, S. S., Omer, A. D., Nyquist, S. K., Hoeger, M., Durand, N. C., Shamim, M. S., Machol, I., Lander, E. S., Aiden, A. P., & Aiden, E. L. (2017). De novo assembly of the *Aedes aegypti* genome using Hi-C yields chromosome-length scaffolds. *Science*, *356*, 92–95. https://doi.org/10.1126/science.aal3327

Dudchenko, O., Shamim, M. S., Batra, S. S., Durand, N. C., Musial, N. T., Mostofa, R., & Aiden, E. L. (2018). The Juicebox Assembly Tools module facilitates *de novo* assembly of mammalian genomes with chromosome-length scaffolds for under $1000. *bioRxiv*, *254797*, https://doi.org/10.1101/254797

Ellinghaus, D., Kurtz, S., & Willhoeft, U. (2008). LTRharvest, an efficient and flexible software for de novo detection of LTR retrotransposons. *BMC Bioinformatics*, *9*, 18. https://doi.org/10.1186/1471-2105-9-18

Emms, D. M., & Kelly, S. (2015). OrthoFinder: Solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy. *Genome Biology*, *16*, 157. https://doi.org/10.1186/s13059-015-0721-2

Fan, P., Wang, P., Lou, Y.-R., Leong, B. J., Moore, B. M., Schenck, C. A., Combs, R., Cao, P., Brandizzi, F., Shiu, S.-H., & Last, R. L. (2020). Evolution of a plant gene cluster in Solanaceae and emergence of metabolic diversity. *elife*, *9*, e56717. https://doi.org/10.7554/eLife.56717

Feng, S., Stiller, J., Deng, Y., Armstrong, J., Fang, Q., Reeve, A. H., Xie, D., Chen, G., Guo, C., Faircloth, B. C., Petersen, B., Wang, Z., Zhou, Q., Diekhans, M., Chen, W., Andreu-Sánchez, S., Margaryan, A., Howard, J. T., Parent, C., … Zhang, G. (2020). Dense sampling of bird diversity increases power of comparative genomics. *Nature*, *587*, 252–257. https://doi.org/10.1038/s41586-020-2873-9

Feschotte, C., Jiang, N., & Wessler, S. R. (2002). Plant transposable elements: Where genetics meets genomics. *Nature Reviews Genetics*, *3*, 329–341. https://doi.org/10.1038/nrg793

Flynn, J. M., Hubley, R., Goubert, C., Rosen, J., Clark, A. G., Feschotte, C., & Smit, A. F. (2020). RepeatModeler2 for automated genomic discovery of transposable element families. *Proceedings of the National Academy of Sciences*, *117*, 9451–9457. https://doi.org/10.1073/pnas.1921046117

Gagnon, E., Hilgenhof, R., Orejuela, A., McDonnell, A., Sablok, G., Aubriot, X., et al., & Sarkinen, T. (2021). Phylogenomic data reveal hard polytomies across the backbone of the large genus *Solanum* (Solanaceae). *American Journal of Botany*, *109*, 580601. https://doi.org/10.1002/ajb2.1827

Gates, D. J., Strickler, S. R., Mueller, L. A., Olson, B. J. S. C., & Smith, S. D. (2016). Diversification of R2R3-MYB transcription factors in the tomato family Solanaceae. *Journal of Molecular Evolution*, *83*, 26–37. https://doi.org/10.1007/s00239-016-9750-z

Gebhardt, C. (2016). The historical role of species from the Solanaceae plant family in genetic research. *Theoretical and Applied Genetics*, *129*, 2281–2294. https://doi.org/10.1007/s00122-016-2804-1

Gilardoni, P. A., Hettenhausen, C., Baldwin, I. T., & Bonaventure, G. (2011). *Nicotiana attenuata* LECTIN RECEPTOR KINASE1 suppresses the insect-mediated inhibition of induced defense responses during *Manduca sexta* herbivory. *Plant Cell*, *23*, 3512–3532. https://doi.org/10.1105/tpc.111.088229

Hahn, M. W., & Nakhleh, L. (2016). Irrational exuberance for resolved species trees. *Evolution*, *70*, 7–17. https://doi.org/10.1111/evo.12832

Han, M. V., Thomas, G. W. C., Lugo-Martinez, J., & Hahn, M. W. (2013). Estimating gene gain and loss rates in the presence of error in genome assembly and annotation using CAFE 3. *Molecular Biology and Evolution*, *30*, 1987–1997. https://doi.org/10.1093/molbev/mst100

Hirakawa, H., Shirasawa, K., Miyatake, K., Nunome, T., Negoro, S., Ohyama, A., Yamaguchi, H., Sato, S., Isobe, S., Tabata, S., & Fukuoka, H. (2014). Draft genome sequence of eggplant (*Solanum melongena* L.): The representative *Solanum* species indigenous to the Old World. *DNA Research*, *21*, 649–660. https://doi.org/10.1093/dnares/dsu027

Hoff, K. J., & Stanke, M. (2019). Predicting genes in single genomes with AUGUSTUS. *Current Protocols in Bioinformatics*, *65*, e57. https://doi.org/10.1002/cpbi.57

Hosmani, P. S., Flores-Gonzalez, M., Van de Geest, H., Maumus, F., Bakker, L. V., Schijlen, E., & Saha, S. (2019). An improved de novo assembly and annotation of the tomato reference genome using single-molecule sequencing, Hi-C proximity ligation and optical maps. *bioRxiv.*, *767764*, https://doi.org/10.1101/767764

Hunziker, A. T. (2001). *The genera of Solanaceae*. A. R. G. Ganter Verlag.

Jones, P., Binns, D., Chang, H. - Y., Fraser, M., Li, W., McAnulla, C., McWilliam, H., Maslen, J., Mitchell, A., Nuka, G., Pesseat, S., Quinn, A. F., Sangrador-Vegas, A., Scheremetjew, M., Yong, S.-Y., Lopez, R., & Hunter, S. (2014). InterProScan 5: Genome-scale protein function classification. *Bioinformatics*, *30*, 1236–1240. https://doi.org/10.1093/bioinformatics/btu031

Jouffroy, O., Saha, S., Mueller, L., Quesneville, H., & Maumus, F. (2016). Comprehensive repeatome annotation reveals strong potential impact of repetitive elements on tomato ripening. *BMC Genomics*, *17*, https://doi.org/10.1186/s12864-016-2980-z

Kachroo, A., Schopfer, C. R., Nasrallah, M. E., & Nasrallah, J. B. (2001). Allele-specific receptor-ligand interactions in *Brassica* self-incompatibility. *Science*, *293*, 1824–1826. https://doi.org/10.1126/science.1062509

Kapusta, A., Suh, A., & Feschotte, C. (2017). Dynamics of genome size evolution in birds and mammals. *Proceedings of the National Academy of Sciences*, *114*, E1460–E1469. https://doi.org/10.1073/pnas.1616702114

Kim, B. Y., Wang, J. R., Miller, D. E., Barmina, O., Delaney, E., Thompson, A., Comeault, A. A., Peede, D., D'agostino, E. R., Pelaez, J., Aguilar, J. M., Haji, D., Matsunaga, T., Armstrong, E. E., Zych, M., Ogawa, Y., Stamenković-Radak, M., Jelić, M., Veselinović, M. S., … Petrov, D. A. (2021). Highly contiguous assemblies of 101 drosophilid genomes. *elife*, *10*, 66405. https://doi.org/10.7554/eLife.66405

Kim, D., Langmead, B., & Salzberg, S. L. (2015). HISAT: A fast spliced aligner with low memory requirements. *Nature Methods*, *12*, 357–360. https://doi.org/10.1038/nmeth.3317

Kim, S., Park, M., Yeom, S.-I., Kim, Y.-M., Lee, J. M., Lee, H. A., Seo, E., Choi, J., Cheong, K., Kim, K.-T., Jung, K., Lee, G.-W., Oh, S.-K., Bae, C., Kim, S.-B., Lee, H.-Y., Kim, S.-Y., Kim, M.-S., Kang, B.-C., … Choi, D. (2014). Genome sequence of the hot pepper provides insights into the evolution of pungency in *Capsicum* species. *Nature Genetics*, *46*, 270–278. https://doi.org/10.1038/ng.2877

Knapp, S. (2010). On 'various contrivances': Pollination, phylogeny and flower form in the Solanaceae. *Philosophical Transactions of the Royal Society B-Biological Sciences*, *365*, 449–460. https://doi.org/10.1098/rstb.2009.0236

Krzywinski, M., Schein, J., Birol, I., Connors, J., Gascoyne, R., Horsman, D., Jones, S. J., & Marra, M. A. (2009). Circos: An information aesthetic for comparative genomics. *Genome Research*, *19*, 1639–1645. https://doi.org/10.1101/gr.092759.109

Kurtz, S., Phillippy, A., Delcher, A. L., Smoot, M., Shumway, M., Antonescu, C., & Salzberg, S. L. (2004). Versatile and open software for comparing large genomes. *Genome Biology*, *5*, R12. https://doi.org/10.1186/gb-2004-5-2-r12

Larget, B. R., Kotha, S. K., Dewey, C. N., & Ané, C. (2010). BUCKy: Gene tree/species tree reconciliation with Bayesian concordance analysis. *Bioinformatics*, *26*, 2910–2911. https://doi.org/10.1093/bioinformatics/btq539

Lee, J., Waminal, N. E., Choi, H.-I., Perumal, S., Lee, S.-C., Nguyen, V. B., Jang, W., Kim, N.-H., Gao, L.-Z., & Yang, T.-J. (2017). Rapid amplification of four retrotransposon families promoted speciation and genome size expansion in the genus *Panax*. *Science Reports*, *7*, 9045. https://doi.org/10.1038/s41598-017-08194-5

Lemmon, Z. H., Reem, N. T., Dalrymple, J., Soyk, S., Swartwood, K. E., Rodriguez-Leal, D., Van Eck, J., & Lippman, Z. B. (2018). Rapid improvement of domestication traits in an orphan crop by genome editing. *Nature Plants*, *4*, 766–770. https://doi.org/10.1038/s41477-018-0259-x

Li, W., & Godzik, A. (2006). Cd-hit: A fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics*, *22*, 1658–1659. https://doi.org/10.1093/bioinformatics/btl158

Lomsadze, A., Burns, P. D., & Borodovsky, M. (2014). Integration of mapped RNA-Seq reads into automatic training of eukaryotic gene

finding algorithm. *Nucleic Acids Research*, *42*, e119. https://doi.org/10.1093/nar/gku557

Lovell, J. T., Jenkins, J., Lowry, D. B., Mamidi, S., Sreedasyam, A., Weng, X., Barry, K., Bonnette, J., Campitelli, B., Daum, C., Gordon, S. P., Gould, B. A., Khasanova, A., Lipzen, A., Macqueen, A., Palacio-Mejía, J. D., Plott, C., Shakirov, E. V., Shu, S., … Juenger, T. E. (2018). The genomic landscape of molecular responses to natural drought stress in *Panicum hallii*. *Nature Communications*, *9*, 5213. https://doi.org/10.1038/s41467-018-07669-x

Ma, Z., Zhang, Y., Wu, L., Zhang, G., Sun, Z., Li, Z., Jiang, Y., Ke, H., Chen, B., Liu, Z., Gu, Q., Wang, Z., Wang, G., Yang, J., Wu, J., Yan, Y., Meng, C., Li, L., Li, X., … Wang, X. (2021). High-quality genome assembly and resequencing of modern cotton cultivars provide resources for crop improvement. *Nature Genetics*, *53*, 1385–1391. https://doi.org/10.1038/s41588-021-00910-2

Maddison, W. P. (1997). Gene trees in species trees. *Systematic Biology*, *46*, 523–536. https://doi.org/10.1093/sysbio/46.3.523

Martin, A., & Orgogozo, V. (2013). The loci of repeated evolution: A catalog of genetic hotspots of phenotypic variation. *Evolution*, *67*, 1235–1250. https://doi.org/10.1111/evo.12081

McCubbin, A. G., & Kao, T.-H. (2000). Molecular recognition and response in pollen and pistil interactions. *Annual Review of Cell and Developmental Biology*, *16*, 333–364. https://doi.org/10.1146/annurev.cellbio.16.1.333

Meerow, A. W., Schoellhorn, R. J., & Kartuz, M. (2004). Four cultivars of *Iochroma*. *Hortscience*, *39*, 194–197. https://doi.org/10.21273/HORTSCI.39.1.194

Miyauchi, S., Kiss, E., Kuo, A., Drula, E., Kohler, A., Sánchez-García, M., Morin, E., Andreopoulos, B., Barry, K. W., Bonito, G., Buée, M., Carver, A., Chen, C., Cichocki, N., Clum, A., Culley, D., Crous, P. W., Fauchery, L., Girlanda, M., … Martin, F. M. (2020). Large-scale genome sequencing of mycorrhizal fungi provides insights into the early evolution of symbiotic traits. *Nature Communications*, *11*, 5125. https://doi.org/10.1038/s41467-020-18795-w

Mohd Saad, N. S., Severn-Ellis, A. A., Pradhan, A., Edwards, D., & Batley, J. (2021). Genomics armed with diversity leads the way in Brassica improvement in a changing global environment. *Frontiers in Genetics*, *12*, 600789. https://doi.org/10.3389/fgene.2021.600789

Muchhala, N., Johnsen, S., & Smith, S. D. (2014). Competition for hummingbird pollination shapes flower color variation in Andean Solanaceae. *Evolution*, *68*, 2275–2286. https://doi.org/10.1111/evo.12441

Nagaki, K., Talbert, P. B., Zhong, C. X., Dawe, R. K., Henikoff, S., & Jiang, J. (2003). Chromatin immunoprecipitation reveals that the 180-bp satellite repeat is the key functional DNA element of *Arabidopsis thaliana* centromeres. *Genetics*, *163*, 1221–1225. https://doi.org/10.1093/genetics/163.3.1221

Olmstead, R. G., Bohs, L., Migid, H. A., Santiago-Valentin, E., Garcia, V. F., & Collier, S. M. (2008). A molecular phylogeny of the Solanaceae. *Taxon*, *57*, 1159–1181. https://doi.org/10.1002/tax.574010

Olmstead, R. G., & Palmer, J. D. (1992). A chloroplast DNA phylogeny of the Solanaceae: Subfamilial relationships and character evolution. *Annals of the Missouri Botanical Garden*, *79*, 346–360. https://doi.org/10.2307/2399773

Olmstead, R. G., Sweere, J. A., Spangler, R. E., Bohs, L., & Palmer, J. D. (1999). Phylogeny and provisional classification of the Solanaceae based on chloroplast DNA. In M. Nee, D. E. Symon, R. N. Lester, & J. P. Jessop (Eds.), *Solanaceae IV* (pp. 111–137). Royal Botanical Gardens.

Ou, S., & Jiang, N. (2018). LTR_retriever: A highly accurate and sensitive program for identification of long terminal repeat retrotransposons. *Plant Physiology*, *176*, 1410–1422. https://doi.org/10.1104/pp.17.01310

Paradis, E., Claude, J., & Strimmer, K. (2004). APE: Analyses of phylogenetics and evolution in R language. *Bioinformatics*, *20*, 289–290. https://doi.org/10.1093/bioinformatics/btg412

Pellicer, J., Garcia, S., Canela, M. Á., Garnatje, T., Korobkov, A. A., Twibell, J. D., & Vallès, J. (2010). Genome size dynamics in *Artemisia* L. (Asteraceae): Following the track of polyploidy. *Plant Biology*, *12*, 820–830. https://doi.org/10.1111/j.1438-8677.2009.00268.x

Pickersgill, B. (2007). Domestication of plants in the Americas: Insights from Mendelian and molecular genetics. *Annals of Botany*, *100*, 925–940. https://doi.org/10.1093/aob/mcm193

The Potato Genome Sequencing Consortium. (2011). Genome sequence and analysis of the tuber crop potato. *Nature*, *475*, 189–194. https://doi.org/10.1038/nature10158

Qin, C., Yu, C., Shen, Y., Fang, X., Chen, L., Min, J., Cheng, J., Zhao, S., Xu, M., Luo, Y., Yang, Y., Wu, Z., Mao, L., Wu, H., Ling-Hu, C., Zhou, H., Lin, H., González-Morales, S., Trejo-Saavedra, D. L., … Zhang, Z. (2014). Whole-genome sequencing of cultivated and wild peppers provides insights into Capsicum domestication and specialization. *Proceedings of the National Academy of Sciences*, *111*, 5135–5140. https://doi.org/10.1073/pnas.1400975111

Razali, R., Bougouffa, S., Morton, M. J. L., Lightfoot, D. J., Alam, I., Essack, M., Arold, S. T., Kamau, A. A., Schmöckel, S. M., Pailles, Y., Shahid, M., Michell, C. T., Al-Babili, S., Ho, Y. S., Tester, M., Bajic, V. B., & Negrão, S. (2018). The genome sequence of the wild tomato *Solanum pimpinellifolium* provides insights into salinity tolerance. *Frontiers in Plant Science*, *9*, 1402. https://doi.org/10.3389/fpls.2018.01402

Rodríguez, J., Deanna, R., & Chiarini, F. (2020). Karyotype asymmetry shapes diversity within the physaloids (Physalidinae, Physalideae, Solanaceae). *Systematics and Biodiversity*, *19*, 168–185. https://doi.org/10.1080/14772000.2020.1832156

Rogers, J. (2018). Adding resolution and dimensionality to comparative genomics: Moving from reference genomes to clade genomics. *Genome Biology*, *19*, 115. https://doi.org/10.1186/s13059-018-1500-7

Ronquist, F., & Huelsenbeck, J. P. (2003). MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics*, *19*, 1572–1574. https://doi.org/10.1093/bioinformatics/btg180

Samuels, J. (2015). Biodiversity of food species of the Solanaceae family: A preliminary taxonomic inventory of the subfamily Solanoideae. *Resources*, *4*, 277–322. https://doi.org/10.3390/resources4020277

Särkinen, T., Bohs, L., Olmstead, R. G., & Knapp, S. (2013). A phylogenetic framework for evolutionary study of the nightshades (Solanaceae): A dated 1000-tip tree. *BMC Evolutionary Biology*, *13*, 214. https://doi.org/10.1186/1471-2148-13-214

The Tomato Genome Consortium. (2012). The tomato genome sequence provides insights into fleshy fruit evolution. *Nature*, *485*, 635–641. https://doi.org/10.1038/nature11119

Schmidt, M. H.-W., Vogel, A., Denton, A. K., Istace, B., Wormit, A., Van De Geest, H., Bolger, M. E., Alseekh, S., Maß, J., Pfaff, C., Schurr, U., Chetelat, R., Maumus, F., Aury, J.-M., Koren, S., Fernie, A. R., Zamir, D., Bolger, A. M., & Usadel, B. (2017). De novo assembly of a new *Solanum pennellii* accession using nanopore sequencing. *Plant Cell*, *29*, 2336–2348. https://doi.org/10.1105/tpc.17.00521

Shaw, J. M. H. (1998). *Iochroma*—A review. *New Plantsman*, *5*, 153–191.

Sierro, N., Battey, J. N. D., Ouadi, S., Bakaher, N., Bovet, L., Willig, A., Goepfert, S., Peitsch, M. C., & Ivanov, N. V. (2014). The tobacco genome sequence and its comparison with those of tomato and potato. *Nature Communications*, *5*, 3833. https://doi.org/10.1038/ncomms4833

Simão, F. A., Waterhouse, R. M., Ioannidis, P., Kriventseva, E. V., & Zdobnov, E. M. (2015). BUSCO: Assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics*, *31*, 3210–3212. https://doi.org/10.1093/bioinformatics/btv351

Smit, A. F. A., Hubley, R., & Green, P. (2013). RepeatMasker Open-4.0. http://www.repeatmasker.org

Smith, S. D., Ané, C., & Baum, D. A. (2008). The role of pollinator shifts in the floral diversification of *Iochroma* (Solanaceae). *Evolution*, *62*, 793–806. https://doi.org/10.1111/j.1558-5646.2008.00327.x

Smith, S. D., & Baum, D. A. (2006). Phylogenetics of the florally diverse Andean clade Iochrominae (Solanaceae). *American Journal of Botany*, *93*, 1140–1153. https://doi.org/10.3732/ajb.93.8.1140

Smith, S. D., Hall, S. J., Izquierdo, P. R., & Baum, D. A. (2008). Comparative pollination biology of sympatric and allopatric Andean *Iochroma* (Solanaceae). *Annals of the Missouri Botanical Garden*, *95*, 600–617. https://doi.org/10.3417/2007037

Smith, S. D., & Rausher, M. D. (2011). Gene loss and parallel evolution contribute to species difference in flower color. *Molecular Biology and Evolution*, *28*, 2799–2810. https://doi.org/10.1093/molbev/msr109

Sun, Y., Qiao, Z., Muchero, W., & Chen, J.-G. (2020). Lectin receptor-like kinases: The sensor and mediator at the plant cell surface. *Frontiers in Plant Science*, *11*, 596301. https://doi.org/10.3389/fpls.2020.596301

Trizna, M. (2020). assembly-stats 0.1.4 https://doi.org/10.5281/zenodo.3968774

Walker, B. J., Abeel, T., Shea, T., Priest, M., Abouelliel, A., Sakthikumar, S., Cuomo, C. A., Zeng, Q., Wortman, J., Young, S. K., & Earl, A. M. (2014). Pilon: An integrated tool for comprehensive microbial variant detection and genome assembly improvement. *Plos One*, *9*, e112963. https://doi.org/10.1371/journal.pone.0112963

Wendel, J. F., Jackson, S. A., Meyers, B. C., & Wing, R. A. (2016). Evolution of plant genome architecture. *Genome Biology*, *17*, 37. https://doi.org/10.1186/s13059-016-0908-1

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., De Freitas, L. B., Barboza, G. E., Moyroud, E., & Smith, S. D. (2022). Transcription factors evolve faster than their structural gene targets in the flavonoid pigment pathway. *Molecular Biology and Evolution*, *39*, msac044. https://doi.org/10.1093/molbev/msac044

Wu, M., Kostyun, J. L., & Moyle, L. C. (2018). Genome sequence of *Jaltomata* addresses rapid reproductive trait evolution and enhances comparative genomics in the hyper-diverse Solanaceae. *Genome Biology and Evolution*, *11*, 335. https://doi.org/10.1093/gbe/evy274

Wu, S., Lau, K. H., Cao, Q., Hamilton, J. P., Sun, H., Zhou, C., Eserman, L., Gemenet, D. C., Olukolu, B. A., Wang, H., Crisovan, E., Godden, G. T., Jiao, C., Wang, X., Kitavi, M., Manrique-Carpintero, N., Vaillancourt, B., Wiegert-Rininger, K., Yang, X., … Fei, Z. (2018). Genome sequences of two diploid wild relatives of cultivated sweetpotato reveal targets for genetic improvement. *Nature Communications*, *9*, 4580. https://doi.org/10.1038/s41467-018-06983-8

Xu, G. C., Xu, T. J., Zhu, R., Zhang, Y., Li, S. Q., Wang, H. W., & Li, J. T. (2018). LR_Gapcloser: A tiling path-based gap closer that uses long reads to complete genome assembly. *GigaScience*, *8*, giy157. https://doi.org/10.1093/gigascience/giy157

Xu, S., Brockmöller, T., Navarro-Quezada, A., Kuhl, H., Gase, K., Ling, Z., Zhou, W., Kreitzer, C., Stanke, M., Tang, H., Lyons, E., Pandey, P., Pandey, S. P., Timmermann, B., Gaquerel, E., & Baldwin, I. T. (2017). Wild tobacco genomes reveal the evolution of nicotine biosynthesis. *Proceedings of the National Academy of Science*, *114*, 6133–6138. https://doi.org/10.1073/pnas.1700073114

Xu, Z., & Wang, H. (2007). LTR_FINDER: An efficient tool for the prediction of full-length LTR retrotransposons. *Nucleic Acids Research*, *35*, W265–W268. https://doi.org/10.1093/nar/gkm286

Zamora-Tavares, M. D. P., Martínez, M., Magallón, S., Guzmán-Dávalos, L., & Vargas-Ponce, O. (2016). *Physalis* and physaloids: A recent and complex evolutionary history. *Molecular Phylogenetics and Evolution*, *100*, 41–50. https://doi.org/10.1016/j.ympev.2016.03.032

Zhang, T., Qiao, Q., Novikova, P. Y., Wang, Q., Yue, J., Guan, Y., Ming, S., Liu, T., De, J., Liu, Y., Al-Shehbaz, I. A., Sun, H., Van Montagu, M., Huang, J., Van De Peer, Y., & Qiong, L. (2019). Genome of *Crucihimalaya himalaica*, a close relative of *Arabidopsis*, shows ecological adaptation to high altitude. *Proceedings of the National Academy of Sciences*, *116*, 7137–7146. https://doi.org/10.1073/pnas.1817580116

Zhong, C. X., Marshall, J. B., Topp, C., Mroczek, R., Kato, A., Nagaki, K., Birchler, J. A., Jiang, J., & Dawe, R. K. (2002). Centromeric retroelements and satellites interact with maize kinetochore protein CENH3. *Plant Cell*, *14*, 2825–2836. https://doi.org/10.1105/tpc.006106

Zimin, A. V., Marçais, G., Puiu, D., Roberts, M., Salzberg, S. L., & Yorke, J. A. (2013). The MaSuRCA genome assembler. *Bioinformatics*, *29*, 2669–2677. https://doi.org/10.1093/bioinformatics/btt476

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Powell, A. F., Zhang, J., Hauser, D., Vilela, J. A., Hu, A., Gates, D. J., Mueller, L. A., Li, F.-W., Strickler, S. R., & Smith, S. (2022). Genome sequence for the blue-flowered Andean shrub *Iochroma cyaneum* reveals extensive discordance across the berry clade of Solanaceae. *The Plant Genome*, *15,* e20223. https://doi.org/10.1002/tpg2.20223