

ORIGINAL ARTICLE

Dissecting the basis of novel trait evolution in a radiation with widespread phylogenetic discordance

Meng Wu¹ | Jamie L. Kostyun^{1,2} | Matthew W. Hahn^{1,3} | Leonie C. Moyle¹ ¹Department of Biology, Indiana University, Bloomington, Indiana²Department of Plant Biology, University of Vermont, Burlington, Vermont³Department of Computer Science, Indiana University, Bloomington, Indiana**Correspondence**Leonie C. Moyle, Department of Biology, Indiana University, Bloomington, IN 47405.
Email: lmoyle@indiana.edu**Funding information**

National Science Foundation, Grant/Award Number: DEB-1135707

Abstract

Phylogenetic analyses of trait evolution can provide insight into the evolutionary processes that initiate and drive phenotypic diversification. However, recent phylogenomic studies have revealed extensive gene tree–species tree discordance, which can lead to incorrect inferences of trait evolution if only a single species tree is used for analysis. This phenomenon—dubbed “hemiplasy”—is particularly important to consider during analyses of character evolution in rapidly radiating groups, where discordance is widespread. Here, we generate whole-transcriptome data for a phylogenetic analysis of 14 species in the plant genus *Jaltomata* (the sister clade to *Solanum*), which has experienced rapid, recent trait evolution, including in fruit and nectar colour, and flower size and shape. Consistent with other radiations, we find evidence for rampant gene tree discordance due to incomplete lineage sorting (ILS) and to introgression events among the well-supported subclades. As both ILS and introgression increase the probability of hemiplasy, we perform several analyses that take discordance into account while identifying genes that might contribute to phenotypic evolution. Despite discordance, the history of fruit colour evolution in *Jaltomata* can be inferred with high confidence, and we find evidence of de novo adaptive evolution at individual genes associated with fruit colour variation. In contrast, hemiplasy appears to strongly affect inferences about floral character transitions in *Jaltomata*, and we identify candidate loci that could arise either from multiple lineage-specific substitutions or standing ancestral polymorphisms. Our analysis provides a generalizable example of how to manage discordance when identifying loci associated with trait evolution in a radiating lineage.

KEYWORDSconvergence, hemiplasy, *Jaltomata*, phylogenomics, rapid radiation, *Solanum*

1 | INTRODUCTION

Phylogenies contribute to our understanding of the evolutionary history of traits (Felsenstein, 1985). When the patterns of relationship among species are known, robust inferences about character state evolution can be made, including the number of times a character evolved, the direction of character evolution and the most likely ancestral character state. Phylogenies can also reveal whether lineages with similar phenotypic traits have evolved these via

independent evolution (convergence or parallelism) or whether a single origin is more likely (Wake, Wake, & Specht, 2011). The recent use of whole genomes or transcriptomes to make phylogenetic inferences from thousands to millions of sites (“phylogenomics”) has succeeded in its aim of generating species trees with high levels of statistical support. However, genome-wide analyses have also begun to reveal unexpected complexities in the evolutionary history of rapidly radiating lineages—including widespread gene tree discordance due to incomplete lineage sorting (ILS) and/or introgression

(Degnan & Rosenberg, 2009). This frequent discordance among individual gene trees can amplify incorrect inferences of trait evolution on even well-supported species trees. In particular, when a trait is determined by genes whose topologies do not match the species topology, incorrect inferences of homoplasy (independent evolution of the same character state) are substantially elevated—a phenomenon known as “hemiplasy” (Avisé & Robinson, 2008; Hahn & Nakhleh, 2016; Storz, 2016). Because understanding trait evolution—including the underlying genetic changes—is of particular interest in species radiations, extra care must be taken to consider and account for the influence of hemiplasy in these cases.

The fraction of the genome affected by hemiplasy will depend upon the amount and sources of gene tree discordance in a clade. In rapidly radiating species groups, widespread discordance has been attributed to the effects of both ILS and introgression between lineages (Degnan & Rosenberg, 2009). ILS can affect gene tree topologies when two lineages do not coalesce in their common ancestor, but instead in an ancestral population shared with a third lineage (Madison, 1997). Because the effect of ILS is proportional to ancestral population size, and inversely proportional to the time between speciation events (Hudson, 1983; Pamilo & Nei, 1988), ILS is expected to be particularly exaggerated in radiations where a diverse ancestral population undergoes rapid speciation. Indeed, gene tree discordance has been noted for a substantial fraction of the genome in rapidly radiating groups, including the *Drosophila simulans* subclade (Garrigan et al., 2012), African cichlid fishes (Brawand et al., 2014), wild tomatoes (Pease, Haak, Hahn, & Moyle, 2016) and the genus *Arabidopsis* (Novikova et al., 2016). When there is introgression, discordance emerges because genes that are introgressed among lineages will show historical patterns of relatedness that differ from the loci in the genome into which they are introduced. Substantial introgression has also been identified among rapidly radiating lineages through genome-wide analysis, including *Xiphophorus* fishes (Cui et al., 2013), *Heliconius* butterflies (Martin et al., 2013), Darwin's finches (Lamichhane et al., 2015) and *Anopheles* mosquitoes (Fontaine et al., 2015).

Both ILS and introgression contribute to hemiplasy because they cause a proportion of gene trees to disagree with the species tree (Avisé & Robinson, 2008; Hahn & Nakhleh, 2016; Storz, 2016). In particular, the probability of hemiplasy is expected to be (a) proportional to the fraction of gene trees that are discordant with the species tree; and (b) negatively correlated with the branch length leading to clades with similar phenotypes (Hahn & Nakhleh, 2016). A higher proportion of discordant gene trees increase the probability that a character of interest is underpinned by genes that have a tree topology that differs from the species tree; shorter branch lengths increase the chance of incorrectly inferring homoplasy, as they leave relatively little time for convergent evolution to happen (Hahn & Nakhleh, 2016). Both conditions are expected to be exaggerated in rapidly diversifying groups. Therefore, in these cases mapping characters onto a single species tree has an elevated risk of incorrectly inferring the number of times a trait has evolved and the timing of trait changes (Avisé & Robinson, 2008; Hahn & Nakhleh, 2016; Storz, 2016). Hemiplasy also affects inferences about the specific loci

inferred to underlie trait transitions because the substitutions underlying trait transitions may occur on gene trees that are discordant with the species tree, when ILS or introgression are common (Mendes, Hahn, & Hahn, 2016). Therefore, genome-wide analyses must take into account the extent and distribution of ILS and introgression if they are to accurately infer the number and timing of evolutionary changes in specific traits, and the genes underlying these changes.

In this study, we used genome-wide data to investigate the morphologically and ecologically diverse plant genus *Jaltomata*, in which several key trait transitions appear to have occurred in parallel (Miller, Mione, Phan, & Olmstead, 2011), and have previously been inferred to be independent convergent responses to similar selective pressures. However, because trait diversification has occurred in a relatively short period in this group, the probability of hemiplasy is also expected to be elevated. Our main goals were to assess the timing of lineage and trait diversification in the group, and to identify sources of genetic variation that potentially contribute to rapid trait diversification in *Jaltomata*, while taking into account the potential for hemiplasy. To do so, we generated a whole-transcriptome data set from representative species across the clade and explicitly evaluated alternative scenarios to explain trait evolution by (a) reconstructing phylogenetic relationships among target species and evaluating the extent of discordance with the resulting inferred species tree; (b) evaluating patterns of trait variation and evolution in key reproductive (flower and fruit) characters, in the context of best and least supported nodes in this tree; and (c) evaluating specific scenarios of the genetic changes associated with this trait evolution, to identify candidate loci that might be causally responsible. Our results imply two different scenarios of trait evolution for fruit colour vs. floral traits, reflecting the different amounts of hemiplasy associated with the two traits. Fruit colour evolution in *Jaltomata* could be confidently inferred—along with potential de novo molecular changes on the relevant evolutionary branches—whereas inferring the history of floral trait evolution and the potential contributing loci required more careful treatment that considered a high probability of hemiplasy.

2 | MATERIALS AND METHODS

2.1 | Study system

The plant genus *Jaltomata* includes approximately 60–80 species, distributed from the south-western United States through to the Andes of South America (Moine, 1992; Mione, Leiva González, & Yacher, 2015; Supporting Information Figure S1). It is the sister genus to *Solanum*, the largest and most economically important genus in the family Solanaceae (Olmstead et al., 2008; Särkinen, Bohs, Olmstead, & Knapp, 2013). Species of *Jaltomata* live in a wide range of habitats and are phenotypically diverse in vegetative, floral and other reproductive traits (Kostyun & Moyle, 2017; Mione, 1992). Floral diversity is particularly pronounced in *Jaltomata*. In comparison with closely related clades (including *Solanum*, *Capsicum* and *Lycianthes*) that predominantly have “flattened” rotate corollas (petals) (Knapp, 2010), *Jaltomata* species exhibit a variety of corolla

shapes, including rotate, campanulate (bell shaped) and tubular (Miller et al., 2011). All *Jaltomata* species also produce at least some nectar, including noticeably red- or orange-coloured nectar in some lineages, while nectar is not produced by species in *Solanum*. *Jaltomata* species also differ in mature fruit colour, and this variation appears to characterize major subgroups within the genus as separate dark purple-, red- and orange-fruited clades (Miller et al., 2011; Särkinen et al., 2013). Several species also have green fruit at maturity; however, these lineages appear to be distributed across the three major *Jaltomata* clades, suggesting multiple convergent losses of fruit pigment (Miller et al., 2011). The first molecular phylogeny of this genus (Mione, Olmstead, Jansen, & Anderson, 1994), inferred from 15 AFLP markers, recovered two major sister clades (purple-fruited and red-/orange-fruited). A subsequent study with more species (Miller et al., 2011) and using a single gene (*waxy*) indicated that the lineage of species with red fruits is sister to the rest of the genus. Most recent, an analysis of seven genes (five plastid and two nuclear, including *waxy*) (Särkinen et al., 2013) showed yet another conflicting topology, with purple-fruited lineages sister to the remaining groups and red-fruited lineages more closely related to lineages with orange fruits. The inconsistency between these studies might be the result of using few loci or of reconstructions performed with loci that have different evolutionary histories.

2.2 | RNA preparation and sequencing

We chose 14 target *Jaltomata* species that are distributed across the three previously identified major clades (Miller et al., 2011) and that span representative floral diversity within the genus (Figure 2a, Supporting Information Table S1). Tissues for RNA extraction included seven reproductive tissues (ranging from early bud, to mature pollinated flower, to early fruit) and four vegetative tissues (roots, early leaf buds and young and mature leaves), from a single representative individual of each target species (see Supporting Information). All sampled individuals were housed at the Indiana University research greenhouse, under standardized temperature (15–20°C), watering (twice daily) and lighting (14-hr days) conditions.

Tissue collection and RNA extraction followed Pease et al. (2016): In brief, tissue was collected into prechilled tubes under liquid nitrogen, each sample was individually ground under liquid nitrogen, and RNA was extracted from <100 mg ground tissue using the Qiagen Plant RNeasy kit. RNA quality/quantity was checked via Nanodrop (Thermo Fisher Scientific); qualified samples of >50 ng/μl with 260/280 and 260/230 between 1.8 and 2.0 were brought to the IU Center for Genomics and Bioinformatics (CGB) for library preparation. Separate reproductive and vegetative libraries for RNA-seq were prepared by pooling equimolar RNA samples from all reproductive tissues, and all vegetative tissues, respectively, for each species. Both reproductive and vegetative libraries were prepared for all species except for *Jaltomata grandibaccata*, for which only vegetative RNA could be obtained.

Libraries were sequenced using 100-bp paired-end reads in a single lane of Illumina Hi-seq 2000 (San Diego, CA, USA). Raw paired-

end reads were filtered for quality using the program SHEAR (<https://github.com/jbpease/shear>) by removing low-quality reads and ambiguous bases, and trimming adapter ends (see Supporting Information).

2.3 | Estimating the amount of nucleotide variation

To quantify the amount of variation within species, and among species and subclades, in *Jaltomata* (Figure 1a), the trimmed reads from all 14 species were mapped to the reference tomato genome (The Tomato Genome Consortium, 2012) using STAR v2.5.2 (Dobin et al., 2013). SAM files generated were converted to sorted BAM files using SAMTOOLS v. 0.1.19, with the flag “-q 255” (requiring mapping score equal to 255) to extract only uniquely mapped reads (Li et al., 2009). SAMTOOLS *mpileup* was then used to call alleles from the BAM files for all lineages. VCF files were processed into Multisample Variant Format (MVF) files using VCF2MVF from the MVFTOOLS package (Pease & Rosenzweig, 2015), requiring nonreference allele calls to have quality scores ≥ 30 and mapped read coverage ≥ 10 . Based on the MVF files, the numbers of variable sites within species and shared between different subclades of *Jaltomata* species were counted.

2.4 | Transcriptome assembly

For transcriptome assembly, reads retained from filtering raw paired-end reads (length >50 bp) from both vegetative and reproductive transcriptomes of each species were combined prior to assembly using Trinity with the default settings (Grabherr et al., 2011) (Figure 1b). The open reading frame of each assembled transcript was predicted using TRANSDCODER v.2.0.1 with default settings (Haas et al., 2013). All the predicted protein-coding sequences within each *Jaltomata* species transcriptome were reduced using CD-HIT v4.6 with -c 0.99 -n 10 (Fu, Niu, Zhu, Wu, & Li, 2012). At any heterozygous site, we randomly chose an allele to generate a representative haplotype for each transcript within each species, for all tree estimation and downstream analyses. To include domestic tomato (*Solanum lycopersicum*) as the outgroup in the following analyses, we also downloaded the annotated tomato protein-coding sequences from SOLGENOMICS (<ftp://ftp.solgenomic.net>).

2.5 | Protein-coding gene ortholog identification

To infer orthologous gene clusters, we followed a pipeline designed for transcriptome data in nonmodel species, that begins with an all-by-all BLAST search followed by several steps that iteratively split subclusters of homologs at long internal branches, until the subtree with the highest number of nonrepeating/nonredundant taxa is obtained (Yang & Smith, 2014; Yang et al., 2015) (see Supporting Information, Figure 1). For the primary analyses, our homologous clusters were required to include a *S. lycopersicum* (tomato) homolog in each cluster. For one of our downstream analyses (molecular evolution on the basal branch leading to *Jaltomata*; see below), we also used *Capsicum annuum* (pepper) sequence data. To do so, we added the *C. annuum*

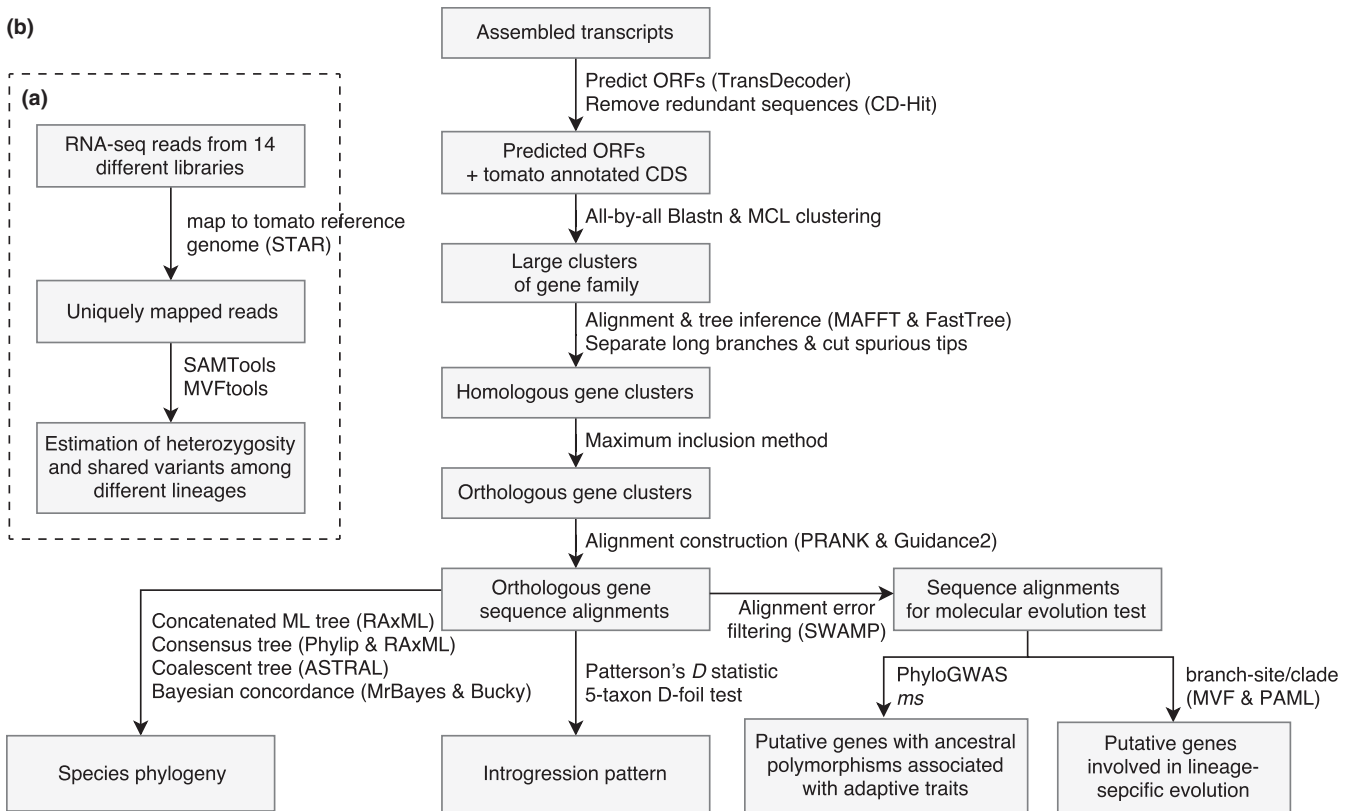


FIGURE 1 Workflow of *Jaltomata* phylogenomics study. (a, inset) The steps to estimate heterozygosity and shared variants among *Jaltomata* lineages. (b) The steps to conduct comparative phylogenetic analyses including phylogeny reconstruction, introgression analyses and downstream tests

sequences if the tomato sequence in the orthologous cluster had an identified 1-to-1 ortholog in a pepper gene model (<http://peppersequence.genomics.cn>).

We prepared multiple sequence alignments of orthologous genes using the program GUIDANCE v.2.0 (Sela, Ashkenazy, Katoh, & Pupko, 2015) with PRANK v.150803 (Löytynoja & Goldman, 2005) as the alignment algorithm, with codons enforced and ten bootstrap replicates. As a final quality check, we further removed poorly aligned regions using a sliding window approach that masked any 15-bp window from alignment if it had more than three mismatches (indels/gaps were not counted) between ingroup sequences or had more than five/seven mismatches when tomato/pepper sequences were included. After this process, any alignment with more than 20% of its sequence masked was removed from the analysis. The resulting sequence alignments were converted to the MVF files, and then, genetic distances were computed in all possible pairs of species using MVFTOOLS (Pease & Rosenzweig, 2015).

2.6 | Phylogenetic analysis

We used four different, but complementary, inference approaches to perform phylogenetic reconstruction: (a) maximum likelihood applied to concatenated alignments; (b) consensus of gene trees; (c) quartet-based gene tree reconciliation; and (d) Bayesian concordance of gene trees. Because these four approaches use different methods to

generate a phylogeny, we applied all four to evaluate the extent to which they generated phylogenies that disagreed, as well as to identify the specific nodes and branches that were robust to all methods of phylogenetic reconstruction. For the concatenation approach, we first aligned all orthologous genes ($n = 6,431$) and then used those alignments to build a supermatrix of sequences (6,223,350 sites in total). The species tree was then inferred by maximum likelihood using the GTRGAMMA model in RAXML v8.23 with 100 bootstraps (Stamatakis, 2006). We also inferred chromosome-concatenated phylogenies with this method (assuming synteny with the *S. lycopersicum* genome and concatenating the data for each chromosome prior to reconstruction). The other three methods (i.e., consensus, quartet-based and Bayesian concordance) infer species relationships based on gene trees. The majority rule consensus tree was inferred with internode certainty (IC) and internode certainty all (ICA) support scores using RAXML with the option for majority rule extended (Salichos & Rokas, 2013). A quartet-based estimation of the species tree was inferred using the program ASTRAL v.4.10.9 with 100 bootstraps (Mirarab & Warnow, 2015), based on the RAXML-inferred gene trees. The Bayesian primary concordance tree and associated concordance factors (CFs; indicative of the posterior probability of gene trees supporting a node) at each internode of the primary concordance tree were computed in the program BUCKY v1.4.4 (Larget, Kotha, Dewey, & Ané, 2010). The input of a posterior distribution of gene trees was generated from an analysis with MRBAYES v3.2 (Huelsenbeck &

Ronquist, 2001). We ran MRBAYES for one million Markov chain Monte Carlo generations, and every 1,000th tree was sampled. After discarding the first half of the 1,000 resulting trees from MRBAYES as burnin, BUCKY was run for one million generations with the default prior probability that two randomly sampled gene trees share the same tree topology set to 50% ($\alpha = 1$) (Larget et al., 2010).

For concatenated, consensus and quartet-based methods, our primary analysis used all orthologous genes in our data set ($n = 6,431$). However, because BUCKY is computationally intensive for a large number of input gene trees, we only used orthologous gene sets that were potentially informative for resolving gene trees in these analyses. In particular, we used 1,190 genes that showed average bootstrap values >50 and minimum bootstrap values >10 at each node, across the RAXML-inferred gene trees. (These are generally the loci with sufficient genetic variation across the tree to provide information about branch support.) To directly compare results from all four phylogeny reconstruction methods, we also applied the other three methods to the same data set of 1,190 gene trees. At last, to examine whether a set of increasingly more highly resolved gene trees affected the resolution of individual nodes in the species tree, we reconstructed the consensus tree and estimated the IC/ICA values using four different gene tree data sets with increasingly higher power (i.e., average node bootstrap values of 50%, 60%, 70% and 80%; see Supporting Information).

All inferred species trees were plotted using the R package "PHYTOOLS" (Revell, 2012). To estimate dates of divergence, we used the function "chronos" in the R package "APE" (Paradis, Claude, & Strimmer, 2004) to fit a chronogram to the RAXML genome-wide concatenated phylogeny using penalized likelihood and maximum-likelihood methods implemented in chronos. Times were calibrated using a previous estimate of the divergence time between *Solanum* and *Jaltomata* at c. 17 Ma (Särkinen et al., 2013). To visualize gene tree discordance, a "cloudogram" of 207 gene trees with average node bootstrap values >70 was prepared using DENSITREE v 2.2.1 (Bouckaert, 2010).

2.7 | Ancestral state reconstruction

The number of sampled species (14) is small compared to the size of this clade (60–80 species), and sparse taxon sampling is known to affect the reconstruction of ancestral character states (Heath, Hedtke, & Hillis, 2008). Nonetheless, to assess whether our confidence in the number and placement of transitions generally differs among different traits in our clade, we reconstructed ancestral states for fruit colour, nectar colour, nectar volume and corolla shape. We used the ultrametric species tree inferred from the RAXML genome-wide concatenated phylogeny and the distribution of traits at the tips of phylogeny (Figure 2a) as input. While nectar volume is a quantitative trait, the other three traits are categorical (fruit colour: purple/red/orange/green; nectar colour: red/clear; corolla shape: rotate/campanulate/tubular). Ancestral character states were inferred using the standard maximum-likelihood method within the PHYTOOLS package (Revell, 2012), which models the evolution of continuous-

valued traits using Brownian motion, and the evolution of discrete-valued traits using a Markov chain. For the latter, we performed ancestral state reconstruction using the equal rates model "ER" and two additional models (symmetric model and all rates different model) and compared model fits with the likelihood ratio test (LRT) in GEIGER (Harmon, Weir, Brock, Glor, & Challenger, 2007). We chose the ER model as there was no significant improvement ($p > 0.05$) using more heavily parameterized models (data not shown).

2.8 | Testing for introgression

We searched for evidence of postspeciation gene flow, or introgression, using the ABBA – BABA test (Durand, Patterson, Reich, & Slatkin, 2011; Green et al., 2010) on the concatenated orthologous sequence alignment ($n = 6,431$, and 6,223,350 sites in total). The ABBA – BABA test detects introgression by comparing the frequency of alternate ancestral ("A") and derived ("B") allele patterns among four taxa. In the absence of gene flow, the alternate patterns ABBA and BABA should be approximately equally frequent, given the equal chance of either underlying discordant topology under ILS. An excess of either ABBA or BABA patterns is indicative of gene flow. Because of the low resolution of many recent branches within major clades of the phylogenetic tree (see Results), evidence for introgression was only evaluated between four major subclades with low discordance in our specific data set (i.e., the purple-, red- and orange-fruited major clades, and a two-species clade of green-fruited taxa; see Results). Patterson's *D*-statistic was calculated for all four-taxon combinations including one taxon from the green-fruited lineage, one from red or orange-fruit lineage, one from purple-fruit lineages and with tomato as the outgroup. Patterson's *D*-statistic is calculated as $(ABBA - BABA)/(ABBA + BABA)$ for biallelic sites in the multiple sequence alignment (Durand et al., 2011; Green et al., 2010).

2.9 | Identifying genetic variation associated with trait evolution

We used two general strategies to identify loci that might contribute to important phenotypic trait transitions (e.g., fruit and floral) within *Jaltomata*. First, to identify loci that have experienced lineage-specific de novo adaptive molecular evolution, we evaluated loci for patterns of molecular evolution indicative of positive selection on specific phylogenetic branches (i.e., $d_N/d_S > 1$). Second, to identify variants that might have been selected from segregating ancestral variation, we identified genetic variants that had polyphyletic topologies that grouped lineages according to shared trait variation rather than phylogenetic relationships ("PHYLOGWAS"; Pease et al., 2016).

2.9.1 | Lineage-specific de novo evolution associated with trait variation

We identified loci with signatures of de novo adaptive molecular evolution (i.e., significantly elevated rates of nonsynonymous substitution) across each available locus in our transcriptome, as well as in

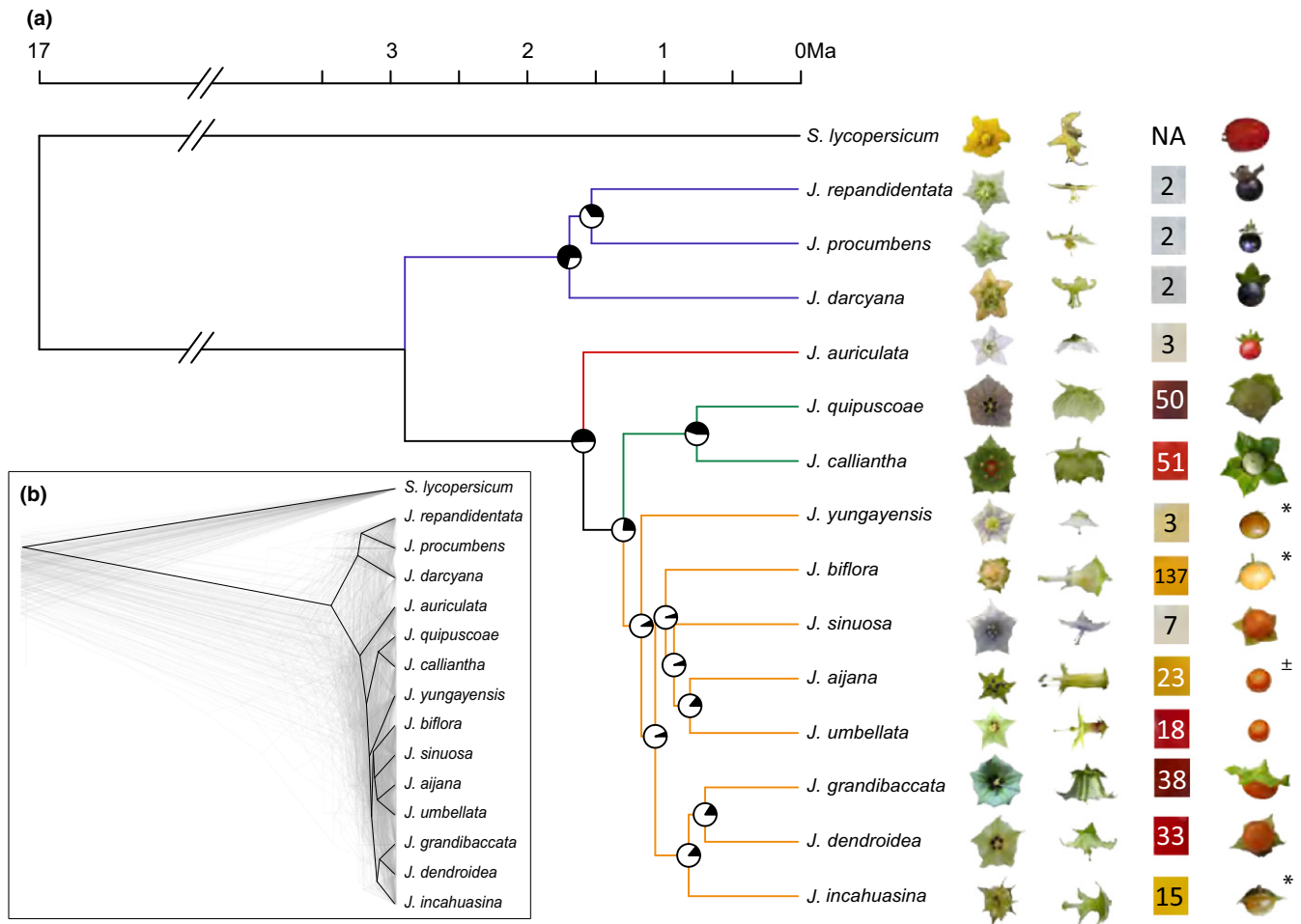


FIGURE 2 The phylogeny of investigated *Jaltomata* species. (a) A whole-transcriptome concatenated phylogeny of *Jaltomata* species with *Solanum lycopersicum* as outgroup. The optimization likelihood for the concatenated ML tree is $-12,312,815.35$. Pie charts on each internode show the proportion of ($n = 6,431$) gene trees that support this specific bipartition. The axis above indicates the estimated divergence times among lineages calibrated using a divergence time between *Solanum* and *Jaltomata* previously estimated at c. 17 Ma (Särkinen et al., 2013). Representative flower and fruit images to the right of species names: front view of flower, lateral view of flower, nectar colour and volume (μl) per flower and ripe fruit. Image with \pm indicates that fruit from a similar species is shown, and * indicates contributed by Dr. Thomas Mione at Central Connecticut State University. (b) A “cloudogram” of 207 gene trees whose average bootstrap values are larger than 70 across the nodes. For contrast, the concatenated tree (a) is shown in black [Colour figure can be viewed at wileyonlinelibrary.com]

a set of a priori candidate loci identified based on known or putative functional roles associated with floral or fruit trait variation (Krizek & Anderson, 2013; Rausher, 2008; Specht & Howarth, 2015) (see Supporting Information). Tests were only performed on the four highest concordance branches within the phylogeny (see Results). For each locus (group of orthologs), we inferred putative adaptive evolution (i.e., $d_N/d_S > 1$) using PAML v4.4 branch-site model (model = 2 and NS sites = 2) on the target branches (Yang, 2007). In each analysis, a LRT was used to determine whether the alternative test model (fixed_omega = 0) was significantly better than the null model (fixed_omega = 1). In addition, because PAML uses a tree-based d_N/d_S model to reconstruct ancestral states and lineage-specific substitutions, and because high levels of incongruence of gene trees caused by ILS and introgression can produce misleading results when gene trees do not match the assumed species tree (Mendes et al., 2016; Pease et al., 2016), we limited our tests of molecular evolution to the subset of

genes for which (a) the RAXML gene tree contained the target ancestral branch (that is, the target branch was supported by the genealogy of the tested/target locus); and (b) there was at least one nonsynonymous substitution that could be unambiguously assigned to this branch. Prior to testing individual loci, we further filtered our data (using SWAMP v1.0; Harrison, Jordan, & Montgomery, 2014) to ensure that poorly aligned and/or error-rich regions were excluded from our alignments (as these tests are particularly sensitive to alignment errors that generate spurious nonsynonymous changes; see Supporting Information). Putative genes showing positive selection were identified at two levels of significance: first, using the uncorrected p -value < 0.01 as a cut-off; and, second, at a false discovery rate (FDR; Benjamini & Hochberg, 1995) < 0.1 , as calculated for the PAML p -values in each branch-specific test.

For the PAML analyses of a priori candidates, we used a slightly less stringent statistical cut-off for consideration and required only

that each cluster of orthologous sequences had a minimum representation of species from each of the major clades (see Supporting Information); this allowed us to evaluate more of these loci while still testing molecular evolution only on the four high-concordance branches. For a priori genes showing a significant signature of positive selection ($p < 0.05$) and for genes identified by the genome-wide unbiased analyses ($FDR < 0.1$), we manually checked the sequence alignments to examine whether they contain putative multinucleotide mutations (MNM), which can cause false inferences of positive selection in the PAML branch-site test (Venkat, Hahn, & Thornton, 2018). Here, we assigned an MNM in cases where we observed that a single codon had two or three substitutions on the selected branch. To determine whether genes with elevated per site nonsynonymous substitution rates were enriched for particular functional categories, we also ran a gene ontology (GO) enrichment analysis on all genes with p -value < 0.05 from the transcriptome-wide analysis (see Supporting Information).

2.9.2 | Ancestral genetic variation associated with trait variation

We used a “PHYLOGWAS” approach (Pease et al., 2016) to search for variable sites that were shared by current accessions that share the same character state, regardless of their phylogenetic relatedness. This approach is only informative in cases where trait variation is not confounded with phylogenetic relationships, which in *Jaltomata* applies to floral shape transitions from the ancestral “rotate” form to the two derived forms (see Results). For our analysis, we treated both campanulate and tubular corolla as the derived state and rotate corolla as the ancestral state (Supporting Information Figure S2). These categories of floral shape are also perfectly associated with nectar colour variation; species with rotate corollas have small amounts of clear/very lightly coloured nectar (ancestral), whereas species with campanulate or tubular corollas have larger amounts of darkly coloured red or orange nectar (derived). To assess whether the number of nonsynonymous variants found to be associated with our defined groups of floral traits (see Results) was greater than expected by chance, we generated a null distribution due to ILS alone by simulating data sets over the species tree (Figure 2a) using the program ms (Hudson, 2002). An associated p -value was determined by the proportion of simulated data sets that have a greater number of genes perfectly associated with the floral trait distribution than our observed value (see Supporting Information).

3 | RESULTS

3.1 | Transcriptome assembly and ortholog inference identified >6,000 orthologs

In assembled transcriptomes from both reproductive and vegetative tissues for each of 14 *Jaltomata* species (except for *J. grandibaccata*, which only included vegetative tissues), the number of transcripts per lineage ranged from 46,841 to 132,050, and mean transcript

length ranged from 736 to 925 bp (Supporting Information Table S2). Based on our criteria for ortholog identification (see Methods, Figure 1), we ultimately identified 6,431 one-to-one orthologous genes for which we had sequences from all 14 investigated *Jaltomata* species and a unique annotated tomato coding sequence. All of these 6,431 genes were used in the concatenation, majority rule and quartet-based phylogeny reconstructions. From this data set, we also extracted 1,190 high-resolution genes to directly compare our four different inference methods (i.e., concatenation, majority rule, quartet-based and BUCKY) (see Methods). As we did not sample RNA from the reproductive tissues of *J. grandibaccata*, we excluded this species from analyses of locus-specific adaptive evolution; this resulted in a slightly larger data set of 6,765 alignments of orthologous coding sequences, each containing sequences from the remaining 13 *Jaltomata* species (with *J. grandibaccata* excluded) and tomato. Among them, 4,248 genes also had *C. annuum* orthologs, thus could also be used to test for positive selection on the ancestral branch leading to *Jaltomata*.

3.2 | Phylogenomic reconstruction of *Jaltomata* lineages supports several major clades

All four phylogenetic inference methods (concatenation, majority rule, quartet-based and Bayesian concordance) generated a species tree topology identical with respect to the placement of major subclades (Figure 2a, Supporting Information Figures S3 and S4). In all trees, the first split in the species tree produces a clade of the three north- and central-American species (*J. procumbens*, *J. repandidentata* and *J. darciana*) that all share floral traits (rotate corollas and light nectar) and produce dark purple/purple fruit. The remaining 11 species, that are found exclusively in South America and vary in floral traits and fruit colours, form a single clade. Within this non-purple-fruited group, our reconstruction indicates that the red-fruited species *J. auriculata* is sister to the remaining species, which are split into a clade of two species (*J. calliantha* and *J. quipuscoae*) that share floral traits and green fruits, and a group consisting of the remaining eight species that vary extensively in floral traits but all produce orange fruit (Figure 2a). Based on sequence divergence at synonymous sites (Supporting Information Table S3), lineages within the non-purple-fruited clade have pairwise distance of 0.26%–0.53% and differ from the purple-fruited lineages by 0.95%–1.29%.

3.3 | Segregating variation is broadly shared among species in different subclades

To quantify how much variation is shared among present subclades—presumably because of either shared ancestral variation or ongoing introgression—we mapped RNA-seq reads from each species to the tomato reference genome and called high-quality variants from ~8 million sites with more than 10× sequencing depth for all investigated species. We identified a large number of sites that are sorting the same allele among different subclades. Among them, 572 variant sites are sorting in all four subclades (Figure 3a). We also quantified

how many sites that are heterozygous in one lineage (accession) have the same two alleles sorting in other subclades. Within each lineage, the proportion of heterozygous sites range from 0.02% to 0.17% (Supporting Information Table S4), which is comparable to the level of heterozygosity observed in self-compatible tomato species (Pease et al., 2016). For 14.04%–65.54% of heterozygous sites in one species (Figure 3a, Supporting Information Table S4), both alleles could also be identified in other subclades, again indicative of a large amount of shared allelic variation. Because we have only one accession per species, we expect that additional sampling could reveal more shared variation among species and subclades.

3.4 | Phylogenomic discordance accompanies rapid diversification

As expected given the large number of genes ($n = 6,431$ and $6,223,350$ sites in total) used for phylogenetic inference, our species trees had very high statistical support in terms of bootstrap values at almost all nodes (Supporting Information Figure S3A,C). Despite this, reconstructions also revealed evidence of extensive gene tree discordance (Figures 2b, Supporting Information Figures S3B, and S4B,D) including Bayesian CFs < 50 at numerous nodes. High discordance was particularly elevated on short internal branches, and within the most rapidly diversifying groups, especially the orange-fruited clade. For example, ICA values around 0 at several nodes within this subclade indicate that the proportion of gene trees supporting the representative bipartition at this internode is almost equivalent to the proportion of conflicting bipartitions (Supporting Information Figures S3B, S4B and S5). Similar short branches and extensive discordance were detected for trees built individually for

each of the 12 chromosomes (Supporting Information Figure S7). Overall, the 6,431 genes inferred 6,431 different topologies, none of which matched the topology of the inferred species tree (Figure 2a).

Several lines of evidence indicate that this phylogenetic discordance is primarily due to biological processes that accompany rapid consecutive lineage-splitting events—especially ILS—rather than low phylogenetic signal or other nonbiological effects. First, based on the reduced data set of 1,190 genes (in which the average bootstrap value on the gene tree is $> 50\%$), we detected strong correlations between the internal branch length and both levels of discordance (as measured by CFs; $p = 4.5 \times 10^{-5}$, Supporting Information Figure S6A) and ICA ($p = 6.3 \times 10^{-8}$, Supporting Information Figure S6B), consistent with either ILS or introgression. Second, smaller data sets of higher bootstrap gene trees continued to show high levels of discordance: As the average power (bootstrap value) of each gene used in the analysis increased, levels of concordance were only marginally improved at some internodes, notably those that were already relatively highly concordant; most internodes that were previously highly discordant still retained high discordance (Supporting Information Figure S8). Finally, the number and distribution of site counts reflecting alternative topologies (i.e., ABBA – BABA) at specific nodes provide strong evidence that discordance is not due to low power: Even on very short branches, we have hundreds to thousands of informative variants that support the two alternative minority topologies. For example, when we count the proportion of informative sites supporting discordant topologies (i.e., $(ABBA + BABA)/(BBAA + ABBA + BABA)$) among the orange-fruited lineages on a genome-wide scale, a large number (39%–68% of $\sim 9,000$) of informative sites disagree with the representative bipartitions (Figure 3a; Supporting Information Table S23). As these counts are not

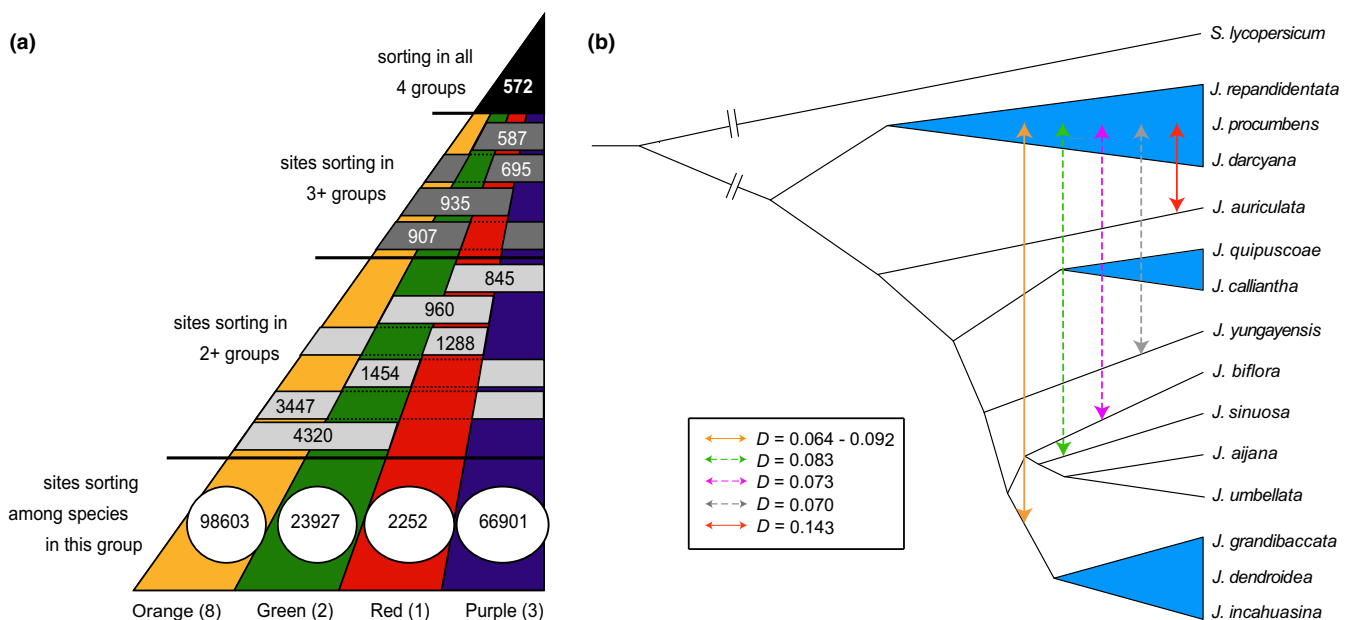


FIGURE 3 (a) Allele sorting, with the number of genetic variants within or shared between each *Jaltomata* subclade. (b) The introgression pattern among *Jaltomata* lineages. The solid lines indicate strong evidence of introgression between two lineages or subclades, while the dashed lines indicate putative introgression. The corresponding Patterson's D -statistic value is labelled for each putative introgression event [Colour figure can be viewed at wileyonlinelibrary.com]

dependent on tree reconstruction (and can be accurately estimated for such recent splits), they provide further evidence of ILS.

Only three branches within *Jaltomata* had relatively little discordance, for example Bayesian CFs >50 (Supporting Information Figure S4D): the branch leading to the purple-fruited clade, the branch uniting all non-purple-fruit *Jaltomata* lineages and the branch leading to the two green-fruited lineages (Figure 2a). Along with the ancestral *Jaltomata* branch, these were the four branches on which most of our subsequent analyses were performed.

3.5 | Introgression after speciation detected among major clades of *Jaltomata* lineages

Given the apparent high level of phylogenetic discordance among our examined species, we also tested for evidence of introgression across the most high-concordance splits within *Jaltomata* (between the purple subclade and each of the red-, green- or orange-fruited subclades), on the background of inferred ILS. We found several such cases (Figure 3b). For example, a significant excess of sites supporting a minority topology that groups the red-fruited lineage (*J. auriculata*) with a purple-fruited lineage (relative to sites grouping the green- and purple-fruited lineages) indicates gene flow between the red- and purple-fruited lineages since their split (Figure 3b and Supporting Information Table S5). We also inferred putative introgression, in at least two separate events, involving six species in the orange-fruited clade with the purple-fruited clade (Figure 3b and Supporting Information Table S5). First, we inferred a shared introgression event between the purple-fruited group and three of the orange-fruited species (*J. grandibaccata*, *J. dendroidea* and *J. incahuasina*); this excess includes shared specific sites that support the same alternative tree topology for each of the three ingroup species, consistent with it involving the common ancestor of all three contemporary orange-fruited species (Supporting Information Table S6). Second, we detected evidence for gene flow between the remaining orange-fruited species (*J. yungayensis*, *J. biflora* and *J. sinuosa*) and the purple-fruited lineage, in the form of significant genome-wide *D*-statistics (Supporting Information Table S5). A lack of shared specific sites supporting the same alternative tree topology among these three orange-fruited species (Supporting Information Table S6) suggests three putative independent introgression events; however, given very low resolution of patterns of relatedness among orange-fruited species, the specific timing of these events is hard to resolve.

3.6 | Ancestral state reconstruction suggests different histories for fruit colour and floral trait evolution

Based on the inferred species tree (Figure 2a), we reconstructed the ancestral states of fruit and floral traits (Figure 4 and Supporting Information Figure S9). The four subclades of *Jaltomata* were inferred to have evolved different fruit colours at their corresponding common ancestors (Figure 4a). Our reconstruction suggests that the derived nectar traits (orange/red nectar colour and increased

nectar volume) probably evolved at the common ancestor of the green/orange-fruited clade (Supporting Information Figure S6A,B), with two subsequent reversions to ancestral conditions within this clade. The evolution of the two derived corolla shapes in *Jaltomata* (campanulate and tubular) appears to be more complex (Figure 4b). At the majority of internodes within the non-purple-fruited lineages, all three corolla shapes (i.e., rotate (ancestral), campanulate and tubular) show $\geq 10\%$ probability of being the ancestral state, making specific inferences about corolla shape evolution within this clade uncertain. These patterns are consistent with very low CFs at almost all internodes within the radiating subgroup that displays the derived floral traits (i.e., the non-purple-fruited lineages) (Figures 2a and Supporting Information Figure S5D), but considerably higher CFs on branches associated with fruit colour evolution (including the branch uniting the two green-fruited species analysed).

These analyses suggest alternative evolutionary and genetic histories for our traits of interest. In particular, strong associations between fruit colour transitions and specific branches/clades within *Jaltomata* suggests that the underlying genetic changes are more likely due to conventional lineage-specific de novo evolution along the relevant branches. In contrast, the ambiguous reconstruction of floral shape trait transitions (Figure 4b) does not exclude de novo evolution (with the same trait evolving independently multiple times), but another alternative is that current transitions drew upon shared variation segregating in the ancestor of these lineages. Therefore, in the next sections, we use lineage-specific de novo evolution analyses to identify potential candidates for fruit colour evolution, whereas both lineage-specific de novo evolution and selection from standing ancestral variation are evaluated when searching for genetic variants that might have contributed to floral trait evolution.

3.7 | Loci with patterns of positive selection associated with lineage-specific trait evolution

We performed tests of molecular evolution for all orthologous clusters that contained a sequence from every *Jaltomata* accession and an ortholog from the tomato outgroup. Depending upon the specific branch, between 1,531 and 3,556 loci in our data set were testable; we detected evidence for positive selection in $\sim 1\%$ – 2% of these loci, based on whether the locus had d_N/d_S ratios significantly >1 ($p < 0.01$) (Supporting Information Tables S7–S11). Many of these genes appear to have general molecular functions (e.g., transcription, protein synthesis or signalling), including stress responses, such as heavy metal tolerance, sugar starvation response, UV and temperature protection, and herbivore and pathogen resistance (Supporting Information Table S7–S11); they were significantly enriched for functions associated with photosynthesis, fatty/lipid biosynthesis and transportation, and sugar signal transduction (in the GO enrichment analysis; Supporting Information Tables S12–S16), as well as loci with unknown functions.

After controlling for multiple tests using an FDR < 0.1 on each branch tested, only three genes on the *Jaltomata* ancestral branch, one gene on the purple-fruited ancestral branch and four genes on

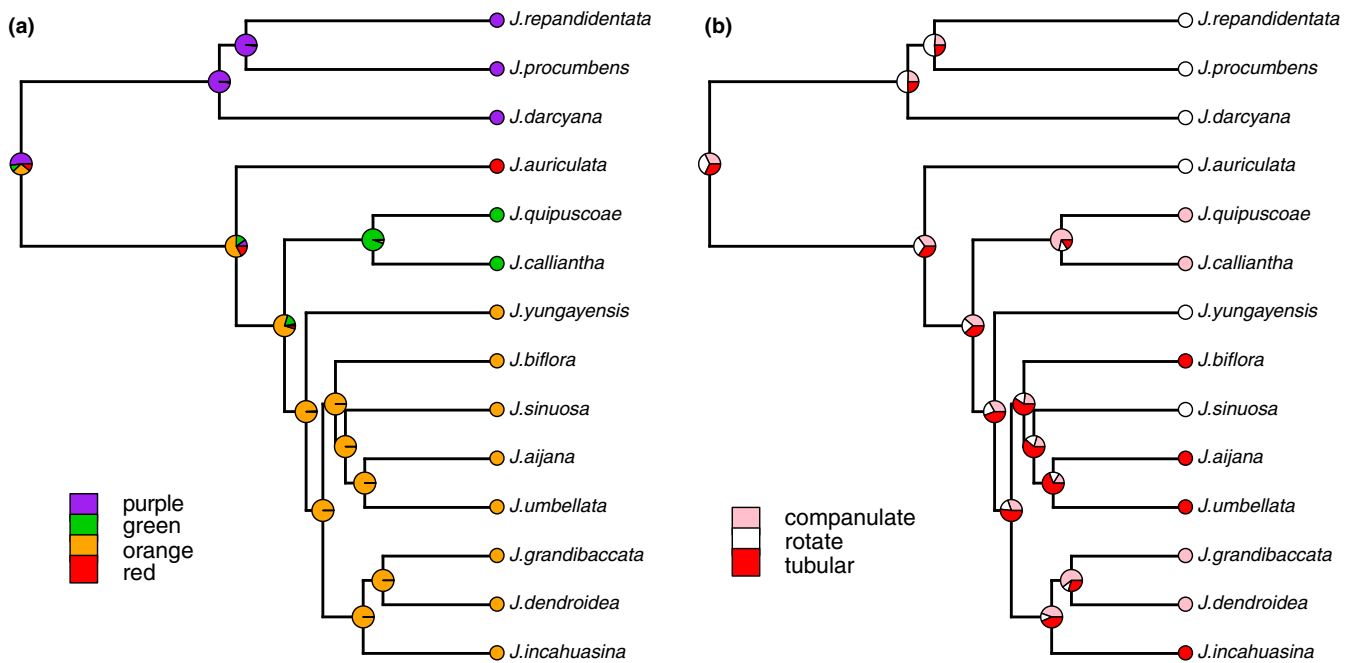


FIGURE 4 Ancestral character state reconstruction of (a) fruit colour, (b) corolla shape in investigated *Jaltomata* using maximum likelihood [Colour figure can be viewed at wileyonlinelibrary.com]

the red ancestral branch remained significant for $d_N/d_S > 1$. Notably, for seven of these eight loci, the inference of positive selection appears to be due to the presence of a multinucleotide mutation (MNM) specifically on the target branch, a mutational pattern known to produce spurious inferences of positive selection in PAML's branch-site test (Venkat et al., 2018). In addition, given the young age of the *Jaltomata* clade, there is a strong possibility that some of the nucleotide changes included in the analysis of d_N/d_S are still polymorphic within populations. Because deleterious mutations are more likely to be sampled as polymorphisms than as fixed differences between species, analyses of this type may be misleading (e.g., Kryazhimskiy & Plotkin, 2008; Li, Costello, Holloway, & Hahn, 2008; Peterson & Masel, 2009). Nevertheless, this is unlikely to be a major influence on our data set, as we do not observe any association between the age of a branch and the number of significant genes detected.

In our a priori candidate gene set, we also detected several instances where slightly less stringent criteria ($d_N/d_S > 1$; $p < 0.05$) revealed lineage-specific adaptive evolution occurring on a branch that is also inferred to be associated with the evolution of derived traits (Supporting Information Table S17). Most notable, we found evidence of positive selection on candidate loci that are likely to be involved in fruit colour, including two genes significant on the ancestral branch of the green-fruited lineages encoding carotenoid cleavage enzyme 1A (*CCD1A*; ortholog to *Solyc01g087250*) and zeaxanthin epoxidase (*ZEP*; ortholog to *Solyc02g090890*) (Figure 5, see Discussion). We also detected positive selection on a gene encoding ζ -carotene isomerase (*ZISO*; ortholog to *Solyc12g098710*) a key enzyme in the production of red-coloured lycopene in the carotenoid biosynthetic pathway (Chen, Li, & Wurtzel, 2010) on the red-

fruited lineage—the same locus found to show adaptive evolution specifically on the branch leading to the red-fruited (Esculentum) group in wild tomatoes (Pease et al., 2016)—however, this locus also contains an MNM on the target branch. We detected signatures of positive selection on fewer of the genes involved in floral development, mostly notably in the MADS-box gene *APETALA3* (*AP3/DEF*, ortholog to *Solyc04g081000*) on the ancestral branch to the purple-fruited lineage. Overall, however, many of our loci (including a priori candidates) did not meet the requirements to be tested for positive selection (Supporting Information Tables S7–S11 and S17); in particular, gene trees for many loci lacked the required support for a specific internal branch, either because of incongruence or an insufficient number of substitutions, especially within the orange-fruited clade (Supporting Information Table S17).

3.8 | Loci potentially associated with trait evolution from standing ancestral variation

To investigate whether ancestral variants are potentially associated with floral trait diversification, we performed a “PHYLOGWAS” analysis (Pease et al., 2016) and found 31 genes with nonsynonymous variants perfectly associated with ancestral vs. derived floral forms (Supporting Information Table S18), this is significantly more than the number of loci expected by chance to have segregation patterns that exactly match the tip states ($p < 9.3 \times 10^{-5}$). Most of these genes are characterized by only one or few nucleotide differences, which is an expected pattern for variants recently selected from standing ancestral variation (Pease et al., 2016). These results suggest that one or few molecular variants present in ancestral populations could contribute to the multiple apparent transitions to derived floral

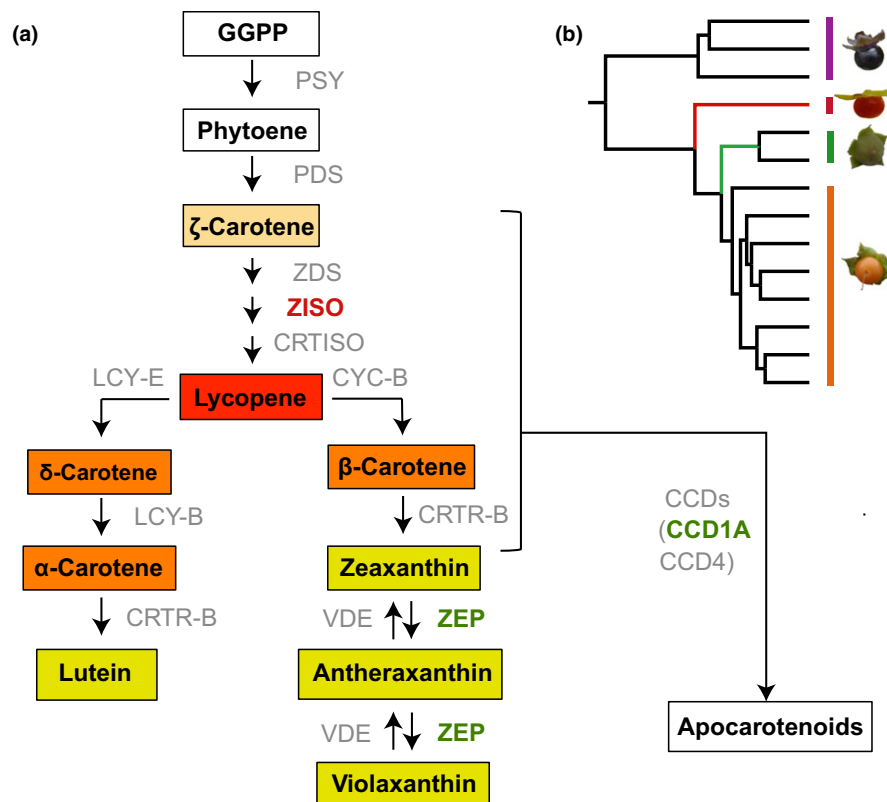


FIGURE 5 Genes under adaptive evolution in the carotenoid biosynthesis pathway. (a) Simplified carotenoid biosynthesis pathway modified from Yuan et al. (2015). Genes putatively under adaptive evolution are indicated by their names highlighted in colours corresponding to particular branches (see panel b). CCD, carotenoid cleavage dioxygenase; CRTISO, carotenoid isomerase; CRTR-B, β -ring hydroxylase; CYC-B, chromoplast specific lycopene β -cyclase; LCY-E, lycopene ϵ -cyclase; LCY-B, lycopene β -cyclase; PDS, phytoene desaturase; PSY, phytoene synthase; VDE, violaxanthin de-epoxidase; ZDS, ζ -carotene desaturase; ZEP, zeaxanthin epoxidase; ZISO, ζ -carotene isomerase. Metabolites are boxed and coloured according to their compound colours, whereas white boxes indicate no colour. (b) Positive selection signatures of genes on different branches are indicated by different colours: red-fruited lineages (Red), and green-fruited lineages (Green). Note that ZISO also has a multinucleotide mutation (MNM) on the target branch (see main text) [Colour figure can be viewed at wileyonlinelibrary.com]

shapes in *Jaltomata*. Among the loci identified by our approach, some genes are potentially functionally related to petal development, including *ARGONAUTE1* (*AGO1*) and *xyloglucan endotransglucosylase/hydrolase 2* (*DcXTH2*) (see Discussion).

4 | DISCUSSION

Within rapidly radiating groups, the patterns of genetic relatedness among lineages provide essential data for determining the pace and timing of important trait transitions and their underlying genes. Both are critical for understanding the drivers of rapid diversification and speciation. Our phylogenomic analyses of the 14 investigated *Jaltomata* species revealed genome-wide gene tree discordance and a highly complex history of genetic relatedness among contemporary lineages. We inferred that substantial ILS, together with putative introgressions among the major subclades of *Jaltomata*, are the sources of this observed complex genome-wide history. This complexity was also reflected in inferences about the evolution of major trait transitions within the group. We found differences in the patterns of fruit vs. floral character evolution and in our inferred

confidence in the reconstruction of these patterns, including their likely risk of hemiplasy. Given this, to identify functionally relevant candidate genes for our target trait transitions, we examined lineage-specific de novo adaptive evolution only along highly concordant branches for both classes of trait, but were also able to identify variants that might have been selected from standing ancestral variation for floral trait transitions. Overall, combining evidence from molecular evolution with data on trait variation across a clade—and a more direct accounting for the risk of hemiplasy—generates more conservative but credible inferences of candidate genes responsible for the evolution of ecologically important phenotypic traits; these draw from different potential sources of adaptive evolution depending upon this risk of hemiplasy.

4.1 | ILS and introgression are emerging as common signals during rapid diversification

An emerging inference from genome-wide studies is that the evolutionary history of rapidly radiating lineages is frequently complex (Brawand et al., 2014; Garrigan et al., 2012; Novikova et al., 2016; Pease et al., 2016). Our analysis of *Jaltomata* fits squarely within this

emerging literature. Our reconstruction of phylogenetic relationships based on a transcriptome-wide data set indicates several well-supported and relatively concordant subclades (primarily distinguished from each other by their fruit colours and also recovered in previous analyses (Mione et al., 1994; Miller et al., 2011; Särkinen et al., 2013)), and both concatenation and quartet-based approaches generated a species tree with high statistical support in terms of bootstrap values (commonly observed when inferring trees from large amounts of data (Salichos & Rokas, 2013)). Nonetheless, gene tree discordance was rampant, with individual gene trees showing highly variable support for the specific placement of individual species (Figure 2b) especially at short branches (Figure 2a). Along with site count data supporting alternative topologies, we infer this discordance is due to extensive genome-wide ILS in *Jaltomata*. Our finding agrees with other recent studies on recent (Novikova et al., 2016; Pease et al., 2016) and relatively ancient adaptive radiations (Wickett et al., 2014; Yang et al., 2015), and suggests that ILS is emerging as a universal signal of rapid radiations, as is expected from theory (Hudson, 1983; Pamilo & Nei, 1988).

A second increasingly common inference from radiating groups is the unexpected prevalence of postspeciation introgression. In our analysis, resolution of phylogenetic relationships among the species within each *Jaltomata* subclade was insufficiently clear to investigate introgression within subgroups. However, across major subclades, we identified at least two clear introgression events that involved either orange or red-fruited lineages with the purple-fruited lineages. Interestingly, in one case an excess of sites supported a shared introgression pattern between three orange-fruited species (i.e., *J. grandibaccata*, *J. dendroidea* and *J. incahuasina*) and the purple-fruited clade, consistent with a scenario in which introgression involved the recent common ancestor of these three orange-fruited species. Moreover, in this case, the inference of a single shared introgression event itself provided more confidence in this specific ancestral branch within the orange-fruited clade. Regardless, as with ILS, postspeciation introgression is another inference increasingly emerging from contemporary phylogenomic studies of radiations, whether these are in plants (Novikova et al., 2016; Owens, Baute, & Rieseberg, 2016; Pease et al., 2016) or in animal groups (Cui et al., 2013; Fontaine et al., 2015; Lamichaney et al., 2015; Martin et al., 2013).

4.2 | Inferring the history of trait evolution and the contributing loci in the presence of rampant discordance

The complex history of genomic divergence in *Jaltomata* and other rapidly diversifying groups has clear consequences for inferences of trait and gene evolution. When the relevant branches and resulting relationships are associated with higher levels of concordance, the evolutionary transitions of traits can be more confidently inferred (Hahn & Nakhleh, 2016). However, when ILS or introgression can plausibly explain the discordant distribution of traits, it might be impossible to infer trait evolution with any certainty in the absence

of additional independent information about target traits, such as their genetic basis (Hahn & Nakhleh, 2016). One of the main goals in this study was to better understand the evolutionary history of distinctive trait diversity within *Jaltomata* in fruit colour, corolla shape, and nectar volume and colour (Figure 2a) (Miller et al., 2011), including the genetic basis of the associated trait transitions. A previous phylogenetic study based on a single locus suggested that floral traits (including corolla shape and nectar colour) might have evolved multiple times independently in *Jaltomata* species (Miller et al., 2011). However, rampant discordance makes inferring the history of trait transitions and their genetic basis especially challenging in this group. Indeed, our analyses indicated that different classes of trait transition—most notably fruit colour vs. floral shape variation—were differently susceptible to hemiplasy.

Accounting for the potential influence of hemiplasy is also critical when generating hypotheses about the loci that could have contributed to trait transitions. In general, incorrect reconstructions of trait history will suggest incorrect candidates involved in the evolution of those traits. Moreover, tests of molecular evolution can be specifically misled if trait transitions occur on discordant gene trees (Mendes et al., 2016). For traits evolving on branches where discordance is low, confidence is high, and hemiplasy is unlikely, it is reasonable to expect that lineage-specific *de novo* substitutions are a substantial contributor to relevant trait evolution. However, when there is a high risk of hemiplasy, genetic variation underpinning trait evolution could potentially come from additional sources, including recruitment of ancestral polymorphisms and/or introgression. These differences are exemplified in our study by the alternative histories, and different genetic hypotheses, generated for fruit colour vs. floral shape traits.

4.2.1 | Floral shape evolution

Multiple lines of evidence indicate that the probability of hemiplasy is high for floral shape traits in *Jaltomata*: Floral shape is distributed paraphyletically, the branch lengths leading to lineages with derived character states are uniformly short with high levels of gene tree discordance, and the three alternative corolla morphs were inferred to be almost equally likely at the common ancestor of non-purple-fruited lineages. It is possible that introgression contributes to these patterns. For example, the ancestral floral character states (i.e., rotate corolla shapes and clear nectar) found in *J. yungayensis* and *J. sinuosa* within the orange-fruited clade could be due to alleles introgressed from purple-fruited species, as we identified putative introgression events between those lineages (Figure 3b). However, because we lack a reference genome for *Jaltomata*, we were precluded from more directly investigating evidence (e.g., locus-specific patterns of introgression) for these scenarios. Instead, we evaluated the two other potential sources of trait variation in this case.

First, based on the rationale that when the paraphyletic distribution of derived traits is due to hemiplasy among species, the relevant nucleotide differences should be at the same sites in all lineages that share derived traits (Hahn & Nakhleh, 2016; Pease et al., 2016), we

identified 31 candidate genes with paraphyletic nonsynonymous variants that were perfectly associated with the distribution of floral trait variation (derived vs. ancestral; Supporting Information Figure S2). Among them, the gene *AGO1* (ortholog to *Solyc02g069260*) is known to be necessary in *Arabidopsis* floral stem cell termination and might act through *CUC1* and *CUC2* (a priori candidate genes), which redundantly specify floral meristem boundaries (Ji et al., 2011; Kidner & Martienssen, 2005). Another gene *DcEXPA2* (ortholog to *Solyc02g091920*) is known to be markedly upregulated in the petals of carnation (*Dianthus caryophyllus*) and is potentially associated with the petal growth and development (Harada et al., 2010). Our sampling of only a fraction of species within the genus, and only a single individual per species, means that we have not captured all of the variation present in *Jaltomata*. While additional sampling will certainly detect more molecular variation, each of the traits tested here is fixed within species. Therefore, although future studies may uncover that some trait-associated alleles found here are also in individuals who do not share these traits (cf. Thomas, Hahn, & Hahn, 2017), they will almost certainly also have more power to distinguish among candidate causative loci. In addition, while the shared hemiplasious variants detected here could be due to postspeciation introgression or sorting from ancestral variation, because the major subclades are primarily allopatric (Supporting Information Figure S1), our interim conclusion is that these variants were more likely sorted from ancestral variation.

Second, because our reconstruction of floral trait evolution could not exclude a role for them, we also examined our data set for genes showing lineage-specific de novo mutations associated with the derived floral traits. Of the small number of testable genes, none of our a priori candidate floral development genes showed suggestive patterns of molecular evolution on internal branches of *Jaltomata* (Supporting Information Table S7), nor did we find any other functionally suggestive (floral development-related) genes adaptively evolving on branches leading to subclades within *Jaltomata*. The one exception was *APETALA3* (*AP3*)—a gene associated with the formation of petals and stamens in flowering plants, including *Arabidopsis* (Wuest et al., 2012)—although this was evolving adaptively on the branch leading to the purple-fruited clade, within which all species retain the ancestral rotate corolla form.

4.2.2 | Fruit colour evolution

In contrast to floral traits, our data indicate that fruit colour transitions are much less susceptible to hemiplasy: Ancestral states of fruit colours at most relevant internodes could be inferred with high confidence (Figure 4a) with fruit colour transitions generally following phylogenetic relationships, and the internal branches leading to fruit trait transitions are highly concordant (Figure 2a). Accordingly, we identified a set of loci showing adaptive evolution specifically on these branches and found multiple adaptively evolving candidate loci with clear functional relevance to these trait transitions, including several a priori candidate genes involved in the carotenoid pathway (Yuan, Zhang, Nageswaran, & Li, 2015; Figure 5). In particular, on

the branch leading to our two green-fruited species we found significantly elevated d_N/d_S ratios for both *CCD1A* (ortholog to *Solyc01g087250*), a gene whose product participates in the conversion of carotenoid pigments to isoprenoid volatiles (Ilg, Bruno, Beyer, & Al-Babili, 2014), and for *ZEP* (ortholog to *Solyc02g090890*), which converts zeaxanthin to violaxanthin (Marin et al., 1996). *CCD1* has previously been identified in tomato fruits as responsible for generating flavour volatiles (Auldrige, McCarty, & Klee, 2006; Simkin, Schwartz, Auldrige, Taylor, & Klee, 2004). This functional observation from a closely related group is intriguing because fruits of our two green-fruited species (*J. quipuscoae* and *J. calliantha*) appear to produce the strongest scent within the 14 *Jaltomata* species analysed here (*J. Kostyun*, unpublished data). This apparent increase in fragrance—presumably due to changed volatile organic compounds—might play a role in attracting vertebrate frugivores for seed dispersal.

In addition to carotenoids, among a priori candidate genes involved in the biosynthesis pathway of water-soluble vacuolar anthocyanin pigments, we detected *BANYULS* (ortholog to *Solyc03g031470*) selected on the red-fruited branch. (PAML also indicates adaptive evolution of *BANYULS* on the purple-fruited branch, but this appears to be due to the presence of a MNM; see Results). In addition to a priori candidates, genome-wide analyses also identified multiple genes belonging to *R2R3MYB*, *BHLH* and *WD40*-repeats classes of loci under positive selection in purple-fruited lineages and the red-fruited lineage. The *MYB-BHLH-WD40* TF complexes are known to regulate cellular differentiation pathways, including of the epidermis, as well as transcription of anthocyanin structural genes (Gonzalez, Zhao, Leavitt, & Lloyd, 2008; Jaakola, 2013; Ramsay & Glover, 2005).

4.3 | Implications for the inference of phenotypic trait evolution and causal genetic variation in rapid radiating lineages

If both ILS and introgression frequently contribute to the history of diversification within radiating clades, evolution in these groups will be more complex than can be represented by a simple bifurcating species tree (Eaton & Ree, 2013; Novikova et al., 2016; Owens et al., 2016; Pease et al., 2016). This complexity clearly has important implications for empirical inferences about historical relationships and trait evolution, because assuming resolved relationships without taking into account incongruence can fundamentally mislead inferences in both these cases (Hahn & Nakhleh, 2016). In general, when the branch lengths leading to lineages with derived character states are uniformly short with high levels of gene tree discordance, the probability of hemiplasy is expected to be very high (Hahn & Nakhleh, 2016). Similarly, to identify loci that might be responsible for any particular trait transitions, different approaches will be appropriate depending upon the confidence with which hemiplasy can be excluded or not. When a trait has several possible alternative evolutionary histories, the range of alternative sources of genetic variation—including de novo lineage-

specific evolution and selection from ancestral variation—that could underpin this trait evolution should be examined. Even cases where trait transitions are associated with relatively low-discordance branches require some caution, for example to avoid potential inference errors from examining genes that have topologies that do not support the branch being tested (Mendes et al., 2016).

Here, we provided a genome-wide analysis of the recently diversified plant genus *Jaltomata* in which we consider the relative risk of hemiplasy while identifying candidates for the specific loci underlying trait evolution. Our analysis highlights a growing appreciation that rapid radiations can and likely do draw on multiple sources of genetic variation (Hedrick, 2013; Pease et al., 2016; Richards & Martin, 2017). Indeed, while independently originating variants could explain the recurrent evolution of phenotypic similarity—a frequent observation in adaptive radiations—it is clear that shared ancestral genetic variation, or alleles introgressed from other lineages, has also made substantial contributions (Elmer & Meyer, 2011; Stern, 2013). Going forward, it will be important to generate data and implement approaches that are able to distinguish between these alternative scenarios, if we are to understand how different evolutionary paths contribute to phenotypic convergence and differentiation (Martin & Orgogozo, 2013; Stern, 2013) and to identify the specific variants responsible.

ACKNOWLEDGEMENTS

The authors thank James Pease, Rafael Guerrero and Fabio Mendes for advice on performing comparative phylogenomic and molecular evolution analyses. This research was funded by National Science Foundation grant DEB-1135707 to LCM and MWH.

DATA AVAILABILITY

Raw reads (FASTQ files) for generating the 14 species transcripts are deposited in the NCBI SRA (BioProject: PRJNA380644). Multiple sequence alignments used for phylogenetic tree reconstruction and molecular evolution analyses are available at Dryad (<https://doi.org/10.5061/dryad.cv270>). All commands and scripts used for analyses in this study can be found in our project directory on GitHub (<https://github.com/wum5/JaltPhylo>).

AUTHOR CONTRIBUTIONS

L.C.M., M.W., M.W.H. and J.L.K. designed the experiments; J.L.K. generated the experimental materials; M.W. conducted the bioinformatics analyses; and M.W. and L.C.M. wrote the paper with contributions from M.W.H. and J.L.K.

ORCID

Leonie C. Moyle  <http://orcid.org/0000-0003-4960-8001>

REFERENCES

- Auldridge, M. E., McCarty, D. R., & Klee, H. J. (2006). Plant carotenoid cleavage oxygenases and their apocarotenoid products. *Current Opinion in Plant Biology*, 9, 315–321. <https://doi.org/10.1016/j.pbi.2006.03.005>
- Avise, J. C., & Robinson, T. J. (2008). Hemiplasy: A new term in the lexicon of phylogenetics. *Systematic Biology*, 57, 503–507. <https://doi.org/10.1080/10635150802164587>
- Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society. Series B (Methodological)*, 57, 289–300.
- Bouckaert, R. R. (2010). DensiTree: Making sense of sets of phylogenetic trees. *Bioinformatics*, 26, 1372–1373. <https://doi.org/10.1093/bioinformatics/btq110>
- Brawand, D., Wagner, C. E., Li, Y. I., Malinsky, M., Keller, I., Fan, S., ... Bezaul, E. (2014). The genomic substrate for adaptive radiation in African cichlid fish. *Nature*, 513, 375–381. <https://doi.org/10.1038/nature13726>
- Chen, Y., Li, F., & Wurtzel, E. T. (2010). Isolation and characterization of the Z-ISO gene encoding a missing component of carotenoid biosynthesis in plants. *Plant Physiology*, 153, 66–79. <https://doi.org/10.1104/pp.110.153916>
- Cui, R., Schumer, M., Kruesi, K., Walter, R., Andolfatto, P., & Rosenthal, G. G. (2013). Phylogenomics reveals extensive reticulate evolution in Xiphophorus fishes. *Evolution*, 67, 2166–2179. <https://doi.org/10.1111/evo.12099>
- Degnan, J. H., & Rosenberg, N. A. (2009). Gene tree discordance, phylogenetic inference and the multispecies coalescent. *Trends in Ecology & Evolution*, 24, 332–340. <https://doi.org/10.1016/j.tree.2009.01.009>
- Dobin, A., Davis, C. A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., ... Gingeras, T. R. (2013). STAR: Ultrafast universal RNA-seq aligner. *Bioinformatics*, 29, 15–21. <https://doi.org/10.1093/bioinformatics/bts635>
- Durand, E. Y., Patterson, N., Reich, D., & Slatkin, M. (2011). Testing for ancient admixture between closely related populations. *Molecular Biology and Evolution*, 28, 2239–2252. <https://doi.org/10.1093/molbev/msr048>
- Eaton, D. A., & Ree, R. H. (2013). Inferring phylogeny and introgression using RADseq data: An example from flowering plants (Pedicularis: Orobanchaceae). *Systematic Biology*, 62, 689–706. <https://doi.org/10.1093/sysbio/syt032>
- Elmer, K. R., & Meyer, A. (2011). Adaptation in the age of ecological genomics: Insights from parallelism and convergence. *Trends in Ecology & Evolution*, 26, 298–306. <https://doi.org/10.1016/j.tree.2011.02.008>
- Felsenstein, J. (1985). Phylogenies and the comparative method. *The American Naturalist*, 125, 1–15. <https://doi.org/10.1086/284325>
- Fontaine, M. C., Pease, J. B., Steele, A., Waterhouse, R. M., Neafsey, D. E., Sharakhov, I. V., ... Kakani, E. (2015). Extensive introgression in a malaria vector species complex revealed by phylogenomics. *Science*, 347, 1258524. <https://doi.org/10.1126/science.1258524>
- Fu, L., Niu, B., Zhu, Z., Wu, S., & Li, W. (2012). CD-HIT: Accelerated for clustering the next-generation sequencing data. *Bioinformatics*, 28, 3150–3152. <https://doi.org/10.1093/bioinformatics/bts565>
- Garrigan, D., Kingan, S. B., Geneva, A. J., Andolfatto, P., Clark, A. G., Thornton, K. R., & Presgraves, D. C. (2012). Genome sequencing reveals complex speciation in the *Drosophila simulans* clade. *Genome Research*, 22, 1499–1511. <https://doi.org/10.1101/gr.130922.111>
- Gonzalez, A., Zhao, M., Leavitt, J. M., & Lloyd, A. M. (2008). Regulation of the anthocyanin biosynthetic pathway by the TTG1/bHLH/Myb transcriptional complex in *Arabidopsis* seedlings. *The Plant Journal*, 53, 814–827. <https://doi.org/10.1111/j.1365-313X.2007.03373.x>

- Grabherr, M. G., Haas, B. J., Yassour, M., Levin, J. Z., Thompson, D. A., Amit, I., ... Zeng, Q. (2011). Trinity: Reconstructing a full-length transcriptome without a genome from RNA-Seq data. *Nature Biotechnology*, *29*, 644. <https://doi.org/10.1038/nbt.1883>
- Green, R. E., Krause, J., Briggs, A. W., Maricic, T., Stenzel, U., Kircher, M., ... Fritz, M. H.-Y. (2010). A draft sequence of the Neandertal genome. *Science*, *328*, 710–722. <https://doi.org/10.1126/science.1188021>
- Haas, B. J., Papanicolaou, A., Yassour, M., Grabherr, M., Blood, P. D., Bowden, J., ... Lieber, M. (2013). De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nature Protocols*, *8*, 1494–1512. <https://doi.org/10.1038/nprot.2013.084>
- Hahn, M. W., & Nakhleh, L. (2016). Irrational exuberance for resolved species trees. *Evolution*, *70*, 7–17. <https://doi.org/10.1111/evo.12832>
- Harada, T., Torii, Y., Morita, S., Onodera, R., Hara, Y., Yokoyama, R., ... Satoh, S. (2010). Cloning, characterization, and expression of xyloglucan endotransglucosylase/hydrolase and expansin genes associated with petal growth and development during carnation flower opening. *Journal of Experimental Botany*, *62*, 815–823.
- Harmon, L. J., Weir, J. T., Brock, C. D., Glor, R. E., & Challenger, W. (2007). GEIGER: Investigating evolutionary radiations. *Bioinformatics*, *24*, 129–131.
- Harrison, P. W., Jordan, G. E., & Montgomery, S. H. (2014). SWAMP: Sliding window alignment masker for PAML. *Evolutionary Bioinformatics*, *10*, 197.
- Heath, T. A., Hedtke, S. M., & Hillis, D. M. (2008). Taxon sampling and the accuracy of phylogenetic analyses. *Journal of Systematics and Evolution*, *46*, 239–257.
- Hedrick, P. W. (2013). Adaptive introgression in animals: Examples and comparison to new mutation and standing variation as sources of adaptive variation. *Molecular Ecology*, *22*, 4606–4618. <https://doi.org/10.1111/mec.12415>
- Hudson, R. R. (1983). Testing the constant-rate neutral allele model with protein-sequence data. *Evolution*, *37*, 203–217. <https://doi.org/10.1111/j.1558-5646.1983.tb05528.x>
- Hudson, R. R. (2002). Generating samples under a Wright-Fisher neutral model of genetic variation. *Bioinformatics*, *18*, 337–338. <https://doi.org/10.1093/bioinformatics/18.2.337>
- Huelsenbeck, J. P., & Ronquist, F. (2001). MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics*, *17*, 754–755. <https://doi.org/10.1093/bioinformatics/17.8.754>
- Ilg, A., Bruno, M., Beyer, P., & Al-Babili, S. (2014). Tomato carotenoid cleavage dioxygenases 1A and 1B: Relaxed double bond specificity leads to a plenitude of dialdehydes, mono-apocarotenoids and isoprenoid volatiles. *FEBS Open Bio*, *4*, 584–593. <https://doi.org/10.1016/j.fob.2014.06.005>
- Jaakola, L. (2013). New insights into the regulation of anthocyanin biosynthesis in fruits. *Trends in Plant Science*, *18*, 477–483. <https://doi.org/10.1016/j.tplants.2013.06.003>
- Ji, L., Liu, X., Yan, J., Wang, W., Yumul, R. E., Kim, Y. J., ... Zheng, B. (2011). ARGONAUTE10 and ARGONAUTE1 regulate the termination of floral stem cells through two microRNAs in Arabidopsis. *PLoS Genetics*, *7*, e1001358. <https://doi.org/10.1371/journal.pgen.1001358>
- Kidner, C. A., & Martienssen, R. A. (2005). The role of ARGONAUTE1 (AGO1) in meristem formation and identity. *Developmental Biology*, *280*, 504–517. <https://doi.org/10.1016/j.ydbio.2005.01.031>
- Knapp, S. (2010). On 'various contrivances': Pollination, phylogeny and flower form in the Solanaceae. *Philosophical Transactions of the Royal Society B*, *365*, 449–460. <https://doi.org/10.1098/rstb.2009.0236>
- Kostyun, J. L., & Moyle, L. C. (2017). Multiple strong postmating and intrinsic postzygotic reproductive barriers isolate florally diverse species of *Jaltomata* (Solanaceae). *Evolution*, *71*, 1556–1571. <https://doi.org/10.1111/evo.13253>
- Krizek, B. A., & Anderson, J. T. (2013). Control of flower size. *Journal of Experimental Botany*, *64*, 1427–1437. <https://doi.org/10.1093/jxb/ert025>
- Kryazhimskiy, S., & Plotkin, J. B. (2008). The population genetics of dN/dS. *PLoS Genetics*, *4*, e1000304.
- Lamichhaney, S., Berglund, J., Almén, M. S., Maqbool, K., Grabherr, M., Martínez-Barrio, A., ... Zamani, N. (2015). Evolution of Darwin's finches and their beaks revealed by genome sequencing. *Nature*, *518*, 371–375. <https://doi.org/10.1038/nature14181>
- Large, 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>
- Li, Y. F., Costello, J. C., Holloway, A. K., & Hahn, M. W. (2008). "Reverse ecology" and the power of population genomics. *Evolution*, *62*, 2984–2994. <https://doi.org/10.1111/j.1558-5646.2008.00486.x>
- Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., ... Durbin, R. (2009). The sequence alignment/map format and SAMtools. *Bioinformatics*, *25*, 2078–2079. <https://doi.org/10.1093/bioinformatics/btp352>
- Löytynoja, A., & Goldman, N. (2005). An algorithm for progressive multiple alignment of sequences with insertions. *Proceedings of the National Academy of Sciences of the United States of America*, *102*, 10557–10562. <https://doi.org/10.1073/pnas.0409137102>
- Maddison, W. P. (1997). Gene trees in species trees. *Systematic Biology*, *46*, 523–536. <https://doi.org/10.1093/sysbio/46.3.523>
- Marin, E., Nussaume, L., Quesada, A., Gonneau, M., Sotta, B., Huguency, P., ... Marion-Poll, A. (1996). Molecular identification of zeaxanthin epoxidase of *Nicotiana plumbaginifolia*, a gene involved in abscisic acid biosynthesis and corresponding to the ABA locus of *Arabidopsis thaliana*. *The EMBO Journal*, *15*, 2331.
- Martin, S. H., Dasmahapatra, K. K., Nadeau, N. J., Salazar, C., Walters, J. R., Simpson, F., ... Jiggins, C. D. (2013). Genome-wide evidence for speciation with gene flow in *Heliconius* butterflies. *Genome Research*, *23*, 1817–1828. <https://doi.org/10.1101/gr.159426.113>
- Martin, A., & Orgogozo, V. (2013). The loci of repeated evolution: A catalog of genetic hotspots of phenotypic variation. *Evolution*, *67*, 1235–1250.
- Mendes, F. K., Hahn, Y., & Hahn, M. W. (2016). Gene tree discordance can generate patterns of diminishing convergence over time. *Molecular Biology and Evolution*, *33*, 3299–3307. <https://doi.org/10.1093/molbev/msw197>
- Miller, R. J., Mione, T., Phan, H.-L., & Olmstead, R. G. (2011). Color by numbers: Nuclear gene phylogeny of *Jaltomata* (Solanaceae), sister genus to *Solanum*, supports three clades differing in fruit color. *Systematic Botany*, *36*, 153–162. <https://doi.org/10.1600/036364411X553243>
- Moine, T. (1992). *Systematics and evolution of Jaltomata (Solanaceae)* (Ph.D. Dissertation). University of Connecticut, Storrs, CT.
- Mione, T., Leiva González, S., & Yacher, L. (2015). Two new Peruvian species of *Jaltomata* (Solanaceae, Solaneae) with red floral nectar. *Brittonia*, *67*, 105–112. <https://doi.org/10.1007/s12228-014-9360-2>
- Mione, T., Olmstead, R. C., Jansen, R. K., & Anderson, G. J. (1994). Systematic Implications of chloroplast DNA variation in *Jaltomata* and selected physaloid genera (Solanaceae). *American Journal of Botany*, *81*(7), 912–918. <https://doi.org/10.1002/j.1537-2197.1994.tb15572.x>
- Mirarab, S., & Warnow, T. (2015). ASTRAL-II: Coalescent-based species tree estimation with many hundreds of taxa and thousands of genes. *Bioinformatics*, *31*, i44–i52. <https://doi.org/10.1093/bioinformatics/btv234>
- Novikova, P. Y., Hohmann, N., Nizhynska, V., Tsuchimatsu, T., Ali, J., Muir, G., ... Fedorenko, O. M. (2016). Sequencing of the genus *Arabidopsis* identifies a complex history of nonbifurcating speciation and

- abundant trans-specific polymorphism. *Nature Genetics*, 48, 1077–1082. <https://doi.org/10.1038/ng.3617>
- 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.
- Owens, G. L., Baute, G. J., & Rieseberg, L. H. (2016). Revisiting a classic case of introgression: Hybridization and gene flow in Californian sunflowers. *Molecular Ecology*, 25, 2630–2643. <https://doi.org/10.1111/mec.13569>
- Pamilo, P., & Nei, M. (1988). Relationships between gene trees and species trees. *Molecular Biology and Evolution*, 5, 568–583.
- 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>
- Pease, J. B., Haak, D. C., Hahn, M. W., & Moyle, L. C. (2016). Phylogenomics reveals three sources of adaptive variation during a rapid radiation. *PLoS Biology*, 14, e1002379. <https://doi.org/10.1371/journal.pbio.1002379>
- Pease, J., & Rosenzweig, B. (2015). Encoding data using biological principles: The Multisample Variant Format for phylogenomics and population genomics. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*. <https://doi.org/10.1109/tccb.2015.2509997>
- Peterson, G. I., & Masel, J. (2009). Quantitative prediction of molecular clock and Ka/Ks at short timescales. *Molecular Biology and Evolution*, 26, 2595–2603. <https://doi.org/10.1093/molbev/msp175>
- Ramsay, N. A., & Glover, B. J. (2005). MYB–bHLH–WD40 protein complex and the evolution of cellular diversity. *Trends in Plant Science*, 10, 63–70. <https://doi.org/10.1016/j.tplants.2004.12.011>
- Rauscher, M. D. (2008). Evolutionary transitions in floral color. *International Journal of Plant Sciences*, 169, 7–21. <https://doi.org/10.1086/523358>
- Revell, L. J. (2012). phytools: An R package for phylogenetic comparative biology (and other things). *Methods in Ecology and Evolution*, 3, 217–223. <https://doi.org/10.1111/j.2041-210X.2011.00169.x>
- Richards, E. J., & Martin, C. H. (2017). Adaptive introgression from distant Caribbean islands contributed to the diversification of a microendemic adaptive radiation of trophic specialist pupfishes. *PLoS Genetics*, 13, e1006919. <https://doi.org/10.1371/journal.pgen.1006919>
- Salichos, L., & Rokas, A. (2013). Inferring ancient divergences requires genes with strong phylogenetic signals. *Nature*, 497, 327–331. <https://doi.org/10.1038/nature12130>
- 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>
- Sela, I., Ashkenazy, H., Katoh, K., & Pupko, T. (2015). GUIDANCE2: Accurate detection of unreliable alignment regions accounting for the uncertainty of multiple parameters. *Nucleic Acids Research*, 43, W7–W14. <https://doi.org/10.1093/nar/gkv318>
- Simkin, A. J., Schwartz, S. H., Auldridge, M., Taylor, M. G., & Klee, H. J. (2004). The tomato carotenoid cleavage dioxygenase 1 genes contribute to the formation of the flavor volatiles β -ionone, pseudoionone, and geranylacetone. *The Plant Journal*, 40, 882–892. <https://doi.org/10.1111/j.1365-313X.2004.02263.x>
- Specht, C. D., & Howarth, D. G. (2015). Adaptation in flower form: A comparative evo devo approach. *New Phytologist*, 206, 74–90. <https://doi.org/10.1111/nph.13198>
- Stamatakis, A. (2006). RAxML-VI-HPC: Maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics*, 22, 2688–2690. <https://doi.org/10.1093/bioinformatics/btl446>
- Stern, D. L. (2013). The genetic causes of convergent evolution. *Nature Reviews Genetics*, 14, 751–764. <https://doi.org/10.1038/nrg3483>
- Storz, J. F. (2016). Causes of molecular convergence and parallelism in protein evolution. *Nature Reviews Genetics*, 17, 239–250. <https://doi.org/10.1038/nrg.2016.11>
- The Tomato Genome Consortium (2012). The tomato genome sequence provides insights into fleshy fruit evolution. *Nature*, 485, 635–641.
- Thomas, G. W. C., Hahn, M. W., & Hahn, Y. (2017). The effects of increasingly the number of taxa on inferences of molecular convergence. *Genome Biology and Evolution*, 9, 213–221.
- Venkat, A., Hahn, M. W., & Thornton, J. W. (2018). Multinucleotide mutations cause false inferences of positive selection. *Nature Ecology and Evolution*. <https://doi.org/10.1038/s41559-018-0584-5>
- Wake, D. B., Wake, M. H., & Specht, C. D. (2011). Homoplasy: From detecting pattern to determining process and mechanism of evolution. *Science*, 331, 1032–1035. <https://doi.org/10.1126/science.1188545>
- Wickett, N. J., Mirarab, S., Nguyen, N., Warnow, T., Carpenter, E., Matasci, N., ... Gitzendanner, M. A. (2014). Phylotranscriptomic analysis of the origin and early diversification of land plants. *Proceedings of the National Academy of Sciences of the United States of America*, 111, E4859–E4868. <https://doi.org/10.1073/pnas.1323926111>
- Wuest, S. E., O'Maoidigh, D. S., Rae, L., Kwasniewska, K., Raganelli, A., Hanczaryk, K., ... Wellmer, F. (2012). Molecular basis for the specification of floral organs by APETALA3 and PISTILLATA. *Proceedings of the National Academy of Sciences*, 109, 13452–13457. <https://doi.org/10.1073/pnas.1207075109>
- Yang, Z. (2007). PAML 4: Phylogenetic analysis by maximum likelihood. *Molecular Biology and Evolution*, 24, 1586–1591. <https://doi.org/10.1093/molbev/msm088>
- Yang, Y., Moore, M. J., Brockington, S. F., Soltis, D. E., Wong, G. K.-S., Carpenter, E. J., & Xie, Y. (2015). Dissecting molecular evolution in the highly diverse plant clade Caryophyllales using transcriptome sequencing. *Molecular Biology and Evolution*, 32, 2001–2014. <https://doi.org/10.1093/molbev/msv081>
- Yang, Y., & Smith, S. A. (2014). Orthology inference in nonmodel organisms using transcriptomes and low-coverage genomes: Improving accuracy and matrix occupancy for phylogenomics. *Molecular Biology and Evolution*, 31, 3081–3092. <https://doi.org/10.1093/molbev/msu245>
- Yuan, H., Zhang, J., Nageswaran, D., & Li, L. (2015). Carotenoid metabolism and regulation in horticultural crops. *Horticulture Research*, 2, 15036. <https://doi.org/10.1038/hortres.2015.36>

SUPPORTING INFORMATION

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

How to cite this article: Wu M, Kostyun JL, Hahn MW, Moyle LC. Dissecting the basis of novel trait evolution in a radiation with widespread phylogenetic discordance. *Mol Ecol*. 2018;27:3301–3316. <https://doi.org/10.1111/mec.14780>