RESEARCH ARTICLE

American Journal of Botany

Phylogeny and multiple independent whole-genome duplication events in the Brassicales

Makenzie E. Mabry[1,11], Julia M. Brose[1], Paul D. Blischak[2], Brittany Sutherland[2], Wade T. Dismukes[1], Christopher A. Bottoms[3], Patrick P. Edger[4], Jacob D. Washburn[5], Hong An[1], Jocelyn C. Hall[6], Michael R. McKain[7], Ihsan Al-Shehbaz[8], Michael S. Barker[2], M. Eric Schranz[9], Gavin C. Conant[10], and J. Chris Pires[1,11]

Manuscript received 10 December 2019; revision accepted 5 May 2020.

[1] Division of Biological Sciences and Christopher S. Bond Life Sciences Center, University of Missouri, Columbia, Missouri 65211, USA

[2] Department of Ecology and Evolutionary Biology, University of Arizona, Tucson, Arizona 85719, USA

[3] Informatics Research Core Facility and Christopher S. Bond Life Sciences Center, University of Missouri, Columbia, Missouri 65211, USA

[4] Department of Horticulture, Michigan State University, East Lansing, Michigan 48824, USA

[5] Plant Genetics Research Unit, USDA-ARS, Columbia, Missouri 65211, USA

[6] Department of Biological Sciences, University of Alberta, Edmonton T6G 2E9, Canada

[7] Department of Biological Sciences, The University of Alabama, Tuscaloosa, Alabama 35401, USA

[8] Missouri Botanical Garden, St. Louis, Missouri 63110, USA

[9] Wageningen University and Research, Wageningen, The Netherlands

[10] Bioinformatics Research Center, Program in Genetics and Department of Biological Sciences, North Carolina State University, Raleigh, North Carolina 27695, USA

[11] Authors for correspondence (e-mail: mmabry44@gmail.com; piresjc@missouri.edu)

Citation: Mabry, M. E., J. M. Brose, P. D. Blischak, B. Sutherland, W. T. Dismukes, C. A. Bottoms, P. P. Edger, J. D. Washburn, et al. 2020. Phylogeny and multiple independent whole-genome duplication events in the Brassicales. *American Journal of Botany* 107(8): 1148–1164.

doi:10.1002/ajb2.1514

PREMISE: Whole-genome duplications (WGDs) are prevalent throughout the evolutionary history of plants. For example, dozens of WGDs have been phylogenetically localized across the order Brassicales, specifically, within the family Brassicaceae. A WGD event has also been identified in the Cleomaceae, the sister family to Brassicaceae, yet its placement, as well as that of WGDs in other families in the order, remains unclear.

METHODS: Phylo-transcriptomic data were generated and used to infer a nuclear phylogeny for 74 Brassicales taxa. Genome survey sequencing was also performed on 66 of those taxa to infer a chloroplast phylogeny. These phylogenies were used to assess and confirm relationships among the major families of the Brassicales and within Brassicaceae. Multiple WGD inference methods were then used to assess the placement of WGDs on the nuclear phylogeny.

RESULTS: Well-supported chloroplast and nuclear phylogenies for the Brassicales and the putative placement of the Cleomaceae-specific WGD event Th-α are presented. This work also provides evidence for previously hypothesized WGDs, including a well-supported event shared by at least two members of the Resedaceae family, and a possible event within the Capparaceae.

CONCLUSIONS: Phylogenetics and the placement of WGDs within highly polyploid lineages continues to be a major challenge. This study adds to the conversation on WGD inference difficulties by demonstrating that sampling is especially important for WGD identification and phylogenetic placement. Given its economic importance and genomic resources, the Brassicales continues to be an ideal group for assessing WGD inference methods.

KEY WORDS Brassicales; Capparaceae; Cleomaceae; phylogeny; phylo-transcriptomics; Resedaceae; whole-genome duplication.

The Brassicales is an economically important order of flowering plants, home to crop species (kale, broccoli, cabbage, cauliflower, papaya, capers, and canola) as well as several model species, including *Arabidopsis* spp. There are currently 17 accepted families within the order (APG IV, 2016), with the Brassicaceae being the most well-studied family due to its many crop and model plant species. Together, the Brassicaceae and the closely related Cleomaceae and Capparaceae contain 94% of the species diversity in the order (Edger et al., 2015). Sister to these three families is an unresolved clade of four families: Tovariaceae, Gyrostemonaceae, Resedaceae, and Pentadiplandraceae. This clade is followed by Emblingiaceae, [Salvadoraceae + Bataceae], Koeberliniaceae], Limnanthaceae, Setchellanthaceae, [Caricaceae + Morginaceae], and [Tropaeolaceae + Akaniaceaeae] (Appendix S1; APG IV, 2016). The order diverged from other rosids ~103 mya and represents 2.2% of the total extant core eudicot diversity (Magallon et al., 1999; Cardinal-McTeague et al., 2016). Previous research has identified multiple whole-genome duplication (WGD) events across the order using a variety of comparative methods, including genomics, transcriptomics, and molecular cytogenetics (Vision et al., 2000; Schranz and Mitchell-Olds, 2006; Barker et al., 2009; Cheng et al., 2013; Kagale et al., 2014; Edger et al., 2015, 2018; Mandáková et al., 2017; Lysak, 2018; One Thousand Plant Transcriptomes Initiative, 2019; Appendix S1). Four of the most studied events include one near the base of the order (At-β; Edger et al., 2015, 2018), an event at the base of the Brassicaceae (At-α; Vision et al., 2000; Haudry et al., 2013; Edger et al., 2015), a triplication at the base of the tribe Brassiceae in the Brassicaceae (Lysak et al., 2005; Tang et al., 2012), and an unplaced event within the Cleomaceae (Th-α; Schranz and Mitchell-Olds, 2006; Barker et al., 2009).

Within the Brassicales, the Brassicaceae has the largest number of accepted species (>4000; BrassiBase 2019). It contains the model plant *Arabidopsis thaliana* (Arabidopsis Genome Initiative, 2000) as well as important crops of the *Brassica* and *Raphanus* groups. Its clades have been placed into three major lineages (Lineage I, Lineage II, and Lineage III; Beilstein et al., 2006), with notable named clades acknowledged more recently (i.e., Clade C; Huang et al., 2016; Nikolov et al., 2019). The relationships among these lineages and clades are unclear. Besides elucidating the relationships within the Brassicaceae, another major area of research has focused on the considerable glucosinolate diversity within the family (Kliebenstein et al., 2001; Ratzka et al., 2002; Züst et al., 2018; Blažević et al., 2019), including the impact of WGD events on the glucosinolate chemical structures (Edger et al., 2015; Barco and Clay, 2019).

Sister to the Brassicaceae is the Cleomaceae. A mostly herbaceous family of ~270 species of pantropical plants, it diverged from the Brassicaceae ~40 mya (Edger et al., 2015). The Cleomaceae displays a much wider range of floral morphologies than its sister family, a characteristic that has been the focus of several recent studies (Bhide et al., 2014; Brock, 2014; Bayat et al., 2018). This family is unique among the Brassicales for containing species that utilize C_4 photosynthesis (*Gynandropsis gynandra* and *Coalisina angustifolia*, formerly *Cleome angustifolia*; Feodorova et al., 2010) as well as, though not unique to Cleomaceae (Schlüter et al., 2016), a C_3-C_4 intermediate (*Coalisina paradoxa*, formerly *C. paradoxa*; van den Bergh et al., 2014). Cleomaceae is known to have undergone at least one independent polyploidy event that occurred after the split from the Brassicaceae, named Th-α (after *Tarenaya hassleriana*), and has been dated to ~13.7 mya (Schranz and Mitchell-Olds, 2006; Barker et al., 2009; Cheng et al., 2013). The analyses for this identification and placement used only partial genomic fragments, ESTs, or a single genome. It was subsequently determined that the Th-α event is shared with the species *G. gynandra*, a C_4 species (van den Bergh et al., 2014), but not with *Cleome violacea*, *Arivela viscosa*, or *Polanisia trachysperma* (Emery et al., 2018; One Thousand Plant Transcriptomes Initiative, 2019). The precise phylogenetic location of Th-α remains a mystery (van den Bergh et al., 2014; Bayat et al., 2018).

The Capparaceae—a mostly woody tropical family of 450 species—is less studied than either of its two sister families, Brassicaceae and Cleomaceae. Like the Cleomaceae, the Capparaceae is very diverse in its floral morphology (Endress, 1992), and, like other members of the order (with the exception of *Koeberlinia spinosa*; Tobe and Raven, 2008), it produces glucosinolates but shares the production of unique methyl-glucosinolates with only the Cleomaceae (Hall et al., 2002; Mithen et al., 2010). In this group is the economically important species *Capparis spinosa*, or capers. Recent work using chromosome counts hypothesized that the Capparaceae and a more distant family, the Resedaceae, may possess unique WGD events (Lysak, 2018). Members of the Resedaceae, a relatively small clade of ~85 species, are mostly distributed across Europe, the Middle East, and Africa, with one taxon occurring in North America (*Oligomeris linifolia*) due to a long-distance dispersal event (Martín-Bravo et al., 2007, 2009; Cardinal-McTeague et al., 2016).

To infer phylogenetic relationships within the Brassicales, we use phylo-transcriptomics, a quickly evolving subdiscipline of phylogenomics that uses RNA-seq data as the basis of its inferences (Dunn et al., 2008; McKain et al., 2012; Yang et al., 2015; Washburn et al., 2017; Unruh et al., 2018; Godden et al., 2019). Transcriptomics gives access to many more nuclear genes than traditional PCR-based approaches and is less expensive than sequencing an entire genome. RNA-seq data also allows for assessing gene and genome duplication events (Barker et al., 2009; McKain et al., 2012). One difficulty with using transcriptomes for phylogenetic inference is determining orthology (Dunn et al., 2013; Yang and Smith 2014; Washburn et al., 2017; Emms and Kelly, 2019). Several methods have been developed to address this problem, including those that aim to identify orthogroups, or sets of genes descended from a single gene in the last common ancestor of the group or species of interest (Duarte et al., 2010; Emms and Kelly, 2015). OrthoFinder version 2 (Emms and Kelly, 2019) offers improvements in both orthogroup inference accuracy and in computational speed, especially when using Diamond (an alternative to BLAST; Buchfink et al., 2015). Together, these methods have enabled phylo-transcriptomics to be extremely useful for inferring species relationships, understanding gene evolution, and elucidating WGD events.

The Brassicales are an intriguing group for the study of polyploidy. With well-established WGD events across the Brassicaceae, including At-α at the base (Vision et al., 2000; Edger et al., 2015) and the identification of a unique and more recent, albeit unplaced, event in the Cleomaceae (Th-α; Schranz and Mitchell-Olds, 2006; Barker et al., 2009; Cheng et al., 2013), one wonders what processes drive the abundance of WGD in this group of plants. Here, we aim to answer the remaining questions on the placement of WGD events, including Th-α, using phylo-transcriptomics, with a focus on sampling the Brassicaceae and Cleomaceae and additional sampling of the Capparaceae, Resedaceae, Bataceae, Caricaceae, and Moringaceae families. We ask if Th-α is shared across the Cleomaceae or if the family, like the Brassicaceae, is characterized by multiple events. We also test the recent hypothesis that the Resedaceae and the Capparaceae possess independent WGD events (Lysak, 2018). We demonstrate that the Brassicales are a powerful resource for the study of WGD and, in the future, will be an important group to test how WGD correlates with variation in floral morphology, photosynthesis types, metabolism (especially glucosinolates), and other traits of interest.

MATERIALS AND METHODS

Taxon sampling

Sampling of 74 taxa from 57 genera across the Brassicales spanned seven families (Brassicaceae, Cleomaceae, Capparaceae, Resedaceae, Bataceae, Moringaceae, Caricaceae), with a focus on the Brassicaceae (48 taxa) and Cleomaceae (17 taxa) (Appendix S2). Seeds were grown at the University of Missouri or the University of Alberta in a sterile growth chamber environment. At maturity, but before flowering, leaf tissue was collected for both RNA and DNA extraction.

DNA and RNA isolation and sequencing

DNA was extracted from leaf tissue for 69 of the 74 taxa using a DNeasy Plant Kit (Qiagen, Germantown, Maryland, USA). To increase yield, slight modifications to the manufacturer's protocol included increasing lysis buffer incubation time to 1 h and using 25 µL of buffer to elute the final sample. TruSeq library preparation (Illumina, San Diego, California, USA) and genome survey sequencing (GSS, also known as skim sequencing) on a NextSeq instrument were carried out at the University of Missouri, resulting in 2 × 150 bp reads.

RNA from leaf tissue was collected and immediately flash frozen using liquid nitrogen. For 38 samples, RNA was isolated using an Invitrogen PureLink RNA Mini Kit (Thermo Fisher Scientific, Carlsbad, California) followed by TruSeq library preparation and sequencing on the NextSeq, resulting in 2 × 75 bp reads (Appendix S3). For 16 samples, RNA was isolated using an Invitrogen PureLink RNA Mini Kit then sequenced on an Illumina HiSeq instrument, resulting in 2 × 100 bp reads (Appendix S3). For 17 samples, RNA was sequenced on a HiSeq for 2 × 100 bp reads but using a Qiagen RNeasy Plant Kit for RNA isolation (Appendix S3). Two samples were isolated using a ThermoFisher Invitrogen PureLink RNA Mini Kit) and sequenced on a HiSeq for 2 × 250 bp reads (Appendix S3). All sequencing and library preparation for the above samples was performed by the University of Missouri DNA Core Facility.

At the University of Alberta, the sample *Cleomella serrulata* had tissue pooled from leaves, apical meristematic tissue, and floral tissue of different developmental stages including small, medium, and large buds and open flowers from two plants. Total RNA was extracted using a Qiagen RNeasy Plant Mini Kit following the manufacturer's protocol, then treated with DNAse I (New England Biolabs, Ipswich, Massachusetts, USA) for 30 min at 37°C to remove residual DNA from the total RNA. Sequencing was conducted by Plateforme d'analyses génomique (l'Université Laval, Quebec City, Quebec, Canada) with Illumina TruSeq RNASeq for library preparation and Illumina for sequencing of paired-end 2 × 100 bp reads.

Chloroplast assembly, alignment, and phylogenomics

To verify identification of taxa, an analysis was performed with two previously published chloroplast genes, *matK* and *ndhF*, for 91 taxa (Hall, 2008) plus 66 samples from this study. The two chloroplast genes were annotated and extracted from de novo whole-chloroplast sequences (discussed below) using Geneious version 8.1.9 (Kearse et al., 2012). We were unable to annotate and extract *ndhF* for the taxon *Batis maritima*. Alignment of resulting genes was performed in MAFFT version 7 (Katoh et al., 2002) and cleaned with Phyutility version 2.7.1 (Smith and Dunn, 2008) using the parameter *-clean 0.5*. For maximum likelihood (ML) phylogenetic inference, RAxML version 8 (Stamatakis, 2014) was run with a separate partition for each gene, GTRGAMMA as the model, and 1000 bootstrap replicates.

To assemble the de novo chloroplasts sequences from the GSS data, we used Fast-Plast version 1.2.8 (McKain and Wilson, 2017). This method utilizes Trimmomatic version 0.35 (Bolger et al., 2014) to clean the reads of adaptors using a Phred score of 33, Bowtie2 version 2.3.4.3 (Langmead and Salzberg, 2012) to separate chloroplast reads by mapping them to a reference database of angiosperm chloroplasts, followed by both SPAdes version 3.13.0 (Bankevich et al., 2012) and "afin" to assemble reads (https://github.com/mrmckain/Fast-Plast/tree/master/afin). For 13 samples that would not assemble with the default options, the *--subsample* option yielded successful assemblies (Appendix S2). Since we obtained only partial regions of the chloroplast genomes for *Polanisia dodecandra*, *Farsetia aegyptia*, and *Cardamine hirsuta*, these samples were excluded from the downstream analyses. Following assembly, MAFFT was used to align the large single copy (LSC), the small single copy (SSC), and one copy of the inverted repeat (IR). Alignments were cleaned with Phyutility using the parameter *-clean 0.5*. Maximum likelihood phylogenomic inference was performed in RAxML with partitions for each region, GTRGAMMA as the model, and 1000 bootstrap replicates.

Transcriptome assembly, alignment, and phylogenomics

For transcriptome analyses, reads were trimmed with Trimmomatic using the parameters *SLIDINGWINDOW:4:5*, *LEADING:5*, *TRAILING:5*, and *MINLEN:25*, followed by assembly with Trinity version 2.2 (Grabherr et al., 2011). The resulting de novo transcriptomes were checked for completeness in BUSCO version 3 (Simão et al., 2015; Waterhouse et al., 2017) and compared to the Embryophyta database. Transcriptomes were translated to protein sequences by extracting the longest open reading frame, and coding regions were predicted using TransDecoder version 3.0 (github.com/TransDecoder/TransDecoder). Finally, orthology was inferred in OrthoFinder version 2.2.6, first using the parameter *-S diamond* (Buchfink et al., 2015) and then the parameter *-M msa -ot* for multiple sequence alignments and only trees. Using custom scripts, alignments were filtered for 80% taxon occupancy (github.com/MU-IRCF/filter_by_ortho_group) and alignment quality, allowing for only 40% gaps (github.com/MU-IRCF/filter_by_gap_fraction). To estimate gene trees using ML inference, RAxML was used (Stamatakis, 2014) under the PROTCATWAG model and 100 bootstrap replicates. Gene trees were analyzed in PhyloTreePruner version 1.0 (Kocot et al., 2013) to remove any paralogous genes by using a cutoff of 10 for the minimum number of taxa required to keep a group. Resulting alignments were then used to estimate final gene trees with RAxML under the PROTCATWAG model and 100 bootstrap replicates. Species tree estimation for the Brassicales was performed with ASTRAL-III version 5.6.1 (Zhang et al., 2018) and included the parameter *-t 2* to assess discordance among gene trees. Species tree analyses were also performed at the family level (Brassicaceae, Cleomaceae, Capparaceae, [Resedaceae + Bataceae + Moringaceae + Cariacacae]; Appendix S4).

Whole-genome duplication

To estimate the phylogenetic placement of whole-genome duplications, we used PUG version 2.1 (github.com/mrmckain/PUG) to query putative paralogs over multiple gene trees with the estimated ASTRAL-III tree as the input species tree. For each analysis, we used the original ML gene trees before running them through PhyloTreePruner (i.e., gene trees with all duplicates retained), the ASTRAL-III tree (rooted and with node labels removed), and parameters *--estimate_paralogs* and *--outgroups Carica_papaya, Moringa_oleifera* as input. Output duplicate gene counts were used for nodes with bootstrap values ≥80%.

As another confirmation of duplication events, we constructed histograms giving the distribution of the synonymous divergence (K_s) between paralogs in each transcriptome. This method allows for the potential identification of peaks in the distribution that may be indicative of a WGD event. The position of the peak along the K_s axis provides an estimate of when the event occurred. Typically, the peak closest to time zero (or $K_s \sim 0$) corresponds to recent tandem duplicates, not relevant to WGD events. Plots of K_s distributions were made for all taxa in FASTKs version 1.1 (github.com/mrmckain/FASTKs), as described in McKain et al. (2016), and in DupPipe, following Barker et al. (2010). R version 3.5.1 (R Core Team, 2018) was used to estimate normal mixture models for K_s values using the "mclust" package version 5.0.2 (Fraley and Raftery, 2002; Fraley et al., 2012). To assess for the best number of peaks to explain the data, we tested one to four components for each mixture model. We chose the component with the lowest Bayesian information criterion (BIC) score as the best fit (Appendix S5).

Ortholog divergence was estimated using OrthoPipe, as described in Barker et al. (2010). Using the estimated ortholog divergence and the DupPipe K_s estimates, we bookended the positions of potential events by comparing when species diverged to the mean paralog divergence of an estimated WGD event: if the ortholog divergence between pairs of species is greater (larger K_s value) than the paralog divergence of a WGD event, then the species do not share the event; if the ortholog divergence between species is less than the WGD paralog divergence, then the species share the proposed event.

RESULTS

Sequence matrices

DNA read pools range in size from 6,637,717 to 13,335,392 reads per sample. After assembly of complete chloroplasts, the inferred genomes for the 66 taxa range in length from 137,110 to 160,272 bp. The LSC, SSC, and IR regions were isolated and aligned separately, with total alignment lengths of 84,350 bp, 17,931 bp, and 26,500 bp, respectively. The analysis of two previously published copies of *matK* and *ndhF*, in combination with our own data, resulted in alignment lengths of 1521 and 985 bp for each gene, respectively. Both chloroplast analyses had 100% occupancy for taxa included.

RNA read pools range in size from 5,555,024 to 59,723,745 reads per sample, with an average of 22,520,865 reads per sample. To check the completeness of transcriptomes, the assemblies were run though BUSCO. All assemblies had >66% complete genes, with <12% of genes missing or fragmented (Appendix S6). OrthoFinder version 2.2.6 recovered 47,600 orthogroups across the Brassicales. Filtering for 80% taxon occupancy (59/74 taxa) yielded 10,968 orthogroups. After filtering for alignment quality by allowing for only 40% gaps, we recovered 2663 orthogroups. Pruning trees for any remaining paralogs, by using a minimum of 10 taxa as a cutoff, resulted in 1284 orthogroups, which were then used for species tree inference. Following the steps above for each family (Brassicaceae, Capparaceae, Cleomaceae, and Resedaceae + Bataceae + Moringaceae + Caricaceae), we recovered 2100, 10,214, 3626, and 8476 orthogroups, respectively (Appendix S4).

Chloroplast phylogenomics of the Brassicales

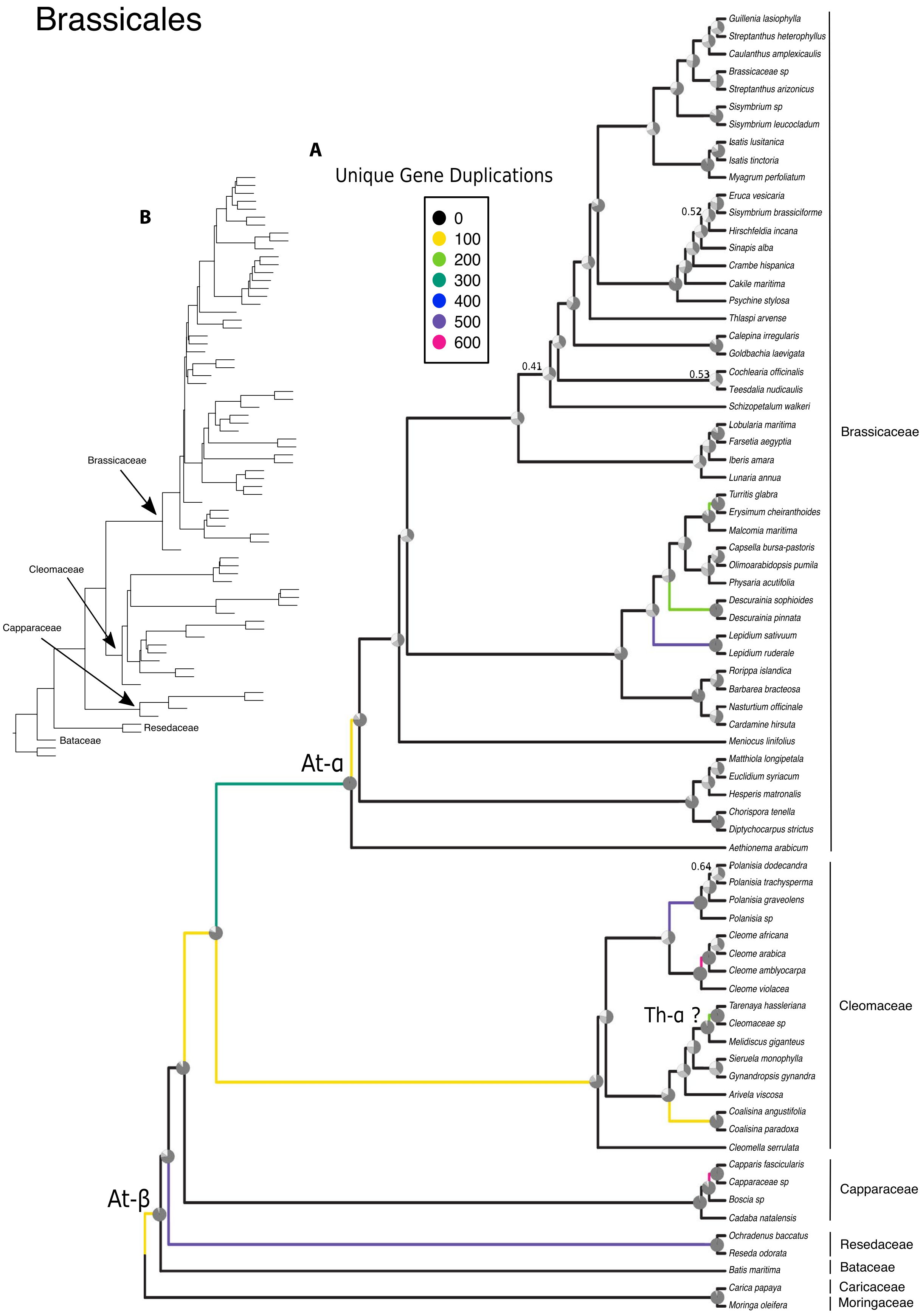
The analysis of the chloroplast genes *matK* and *ndhF*, the 91 taxa from the study by Hall (2008), and our 66 samples identified some inconsistencies of species placement but recovered the same overall relationships as published for other chloroplast phylogenies of the Brassicales (Hall, 2008; Cardinal-McTeague et al., 2016; Edger et al., 2018; Appendix S7). Species sampled in Hall (2008) and in the present study that are not recovered as sister to one another include *Stanleya pinnata*, *Cleomella lutea*, *Andinocleome pilosa*, and *Capparis tomentosa*. The lack of congruence for species placement may be due to species being mislabeled (e.g., *Cleomella lutea*) or species being more genetically diverse than previously thought. Due to this uncertainty in taxon identification, we refer to these samples as Brassicaceae sp., *Polanisia* sp., Cleomaceae sp., and Capparaceae sp., respectively.

For the whole-chloroplast analyses, using just one copy of the IR, all nodes except four were recovered with 70% bootstrap support or better and with a topology largely congruent with previous studies (Hall, 2008; Cardinal-McTeague et al., 2016; Edger et al., 2018). This agreement includes a clade of *Moringa oleifera* and *Carica papaya* sister to a clade of [Bataceae + Resedaceae + Capparaceae + Cleomaceae + Brassicaceae], followed by Bataceae sister to [Resedaceae + Capparaceae + Cleomaceae + Brassicaceae], Resedaceae sister to [Capparaceae + Cleomaceae + Brassicaceae], and Capparaceae sister to [Cleomaceae + Brassicaceae] (Appendix S8). The relationships among the major lineages within the Brassicaceae were also in agreement with previous studies (Guo et al., 2017). We recovered *Aethionema arabicum* as sister to the rest of the family, followed by Lineage I sister to [Lineage III + Clade C + Lineage II and Expanded Lineage II] and Lineage III sister to [Clade C + Lineage II and expanded Lineage II]. Relationships within the Cleomaceae were congruent with previous studies (Hall, 2008; Patchell et al., 2014), with *Cleome* sensu stricto (after Patchell et al., 2014) sister to *Polanisia* plus the rest of the family. Most likely due to sampling, our relationships among the Capparaceae were not congruent with previous studies (Hall, 2008; Tamboli et al., 2018). Previous studies with more sampling recovered *Boscia* sp. sister to *Cadaba*, while we recovered *Boscia* as sister to *Capparis*.

Phylo-transcriptiomics of the Brassicales

Analysis of nuclear data from the transcriptome with ASTRAL-III recovers a well-resolved tree with all nodes but four having a local posterior probability ≥0.7 (Fig. 1). The overall relationships of the

FIGURE 1. ASTRAL-III species phylogeny and whole-genome duplication events of the Brassicales. (A) Species tree with branch lengths proportional and discordance plotted as pie charts at nodes. Dark gray denotes proportion of gene trees agreeing with main topology; lighter gray denotes proportion of gene trees agreeing with first alternative topology; lightest gray denotes proportion of gene trees agreeing with second alternative topology. Previously inferred events (At-α and At-β) and possible placement of Th-α are indicated. Branch color denotes number of unique gene duplications as determined by PUG (github.com/mrmckain/PUG). Support values are indicated if <0.7 local posterior probability. (B) Coalescent-based species tree with branch lengths.


Brassicales
A
Unique Gene Duplications
● 0
● 100
● 200
● 300
● 400
● 500
● 600
B
Brassicaceae
Cleomaceae
Capparaceae
Resedaceae
Bataceae
At-α
At-β
Th-α ?
0.52
0.41
0.53
0.64
Guillenia lasiophylla
Streptanthus heterophyllus
Caulanthus amplexicaulis
Brassicaceae sp
Streptanthus arizonicus
Sisymbrium sp
Sisymbrium leucocladum
Isatis lusitanica
Isatis tinctoria
Myagrum perfoliatum
Eruca vesicaria
Sisymbrium brassiciforme
Hirschfeldia incana
Sinapis alba
Crambe hispanica
Cakile maritima
Psychine stylosa
Thlaspi arvense
Calepina irregularis
Goldbachia laevigata
Cochlearia officinalis
Teesdalia nudicaulis
Schizopetalum walkeri
Lobularia maritima
Farsetia aegyptia
Iberis amara
Lunaria annua
Turritis glabra
Erysimum cheiranthoides
Malcomia maritima
Capsella bursa-pastoris
Olimoarabidopsis pumila
Physaria acutifolia
Descurainia sophioides
Descurainia pinnata
Lepidium sativuum
Lepidium ruderale
Rorippa islandica
Barbarea bracteosa
Nasturtium officinale
Cardamine hirsuta
Meniocus linifolius
Matthiola longipetala
Euclidium syriacum
Hesperis matronalis
Chorispora tenella
Diptychocarpus strictus
Aethionema arabicum
Polanisia dodecandra
Polanisia trachysperma
Polanisia graveolens
Polanisia sp
Cleome africana
Cleome arabica
Cleome amblyocarpa
Cleome violacea
Tarenaya hassleriana
Cleomaceae sp
Melidiscus giganteus
Sieruela monophylla
Gynandropsis gynandra
Arivela viscosa
Coalisina angustifolia
Coalisina paradoxa
Cleomella serrulata
Capparis fascicularis
Capparaceae sp
Boscia sp
Cadaba natalensis
Ochradenus baccatus
Reseda odorata
Batis maritima
Carica papaya
Moringa oleifera
Brassicaceae
Cleomaceae
Capparaceae
Resedaceae
Bataceae
Caricaceae
Moringaceae

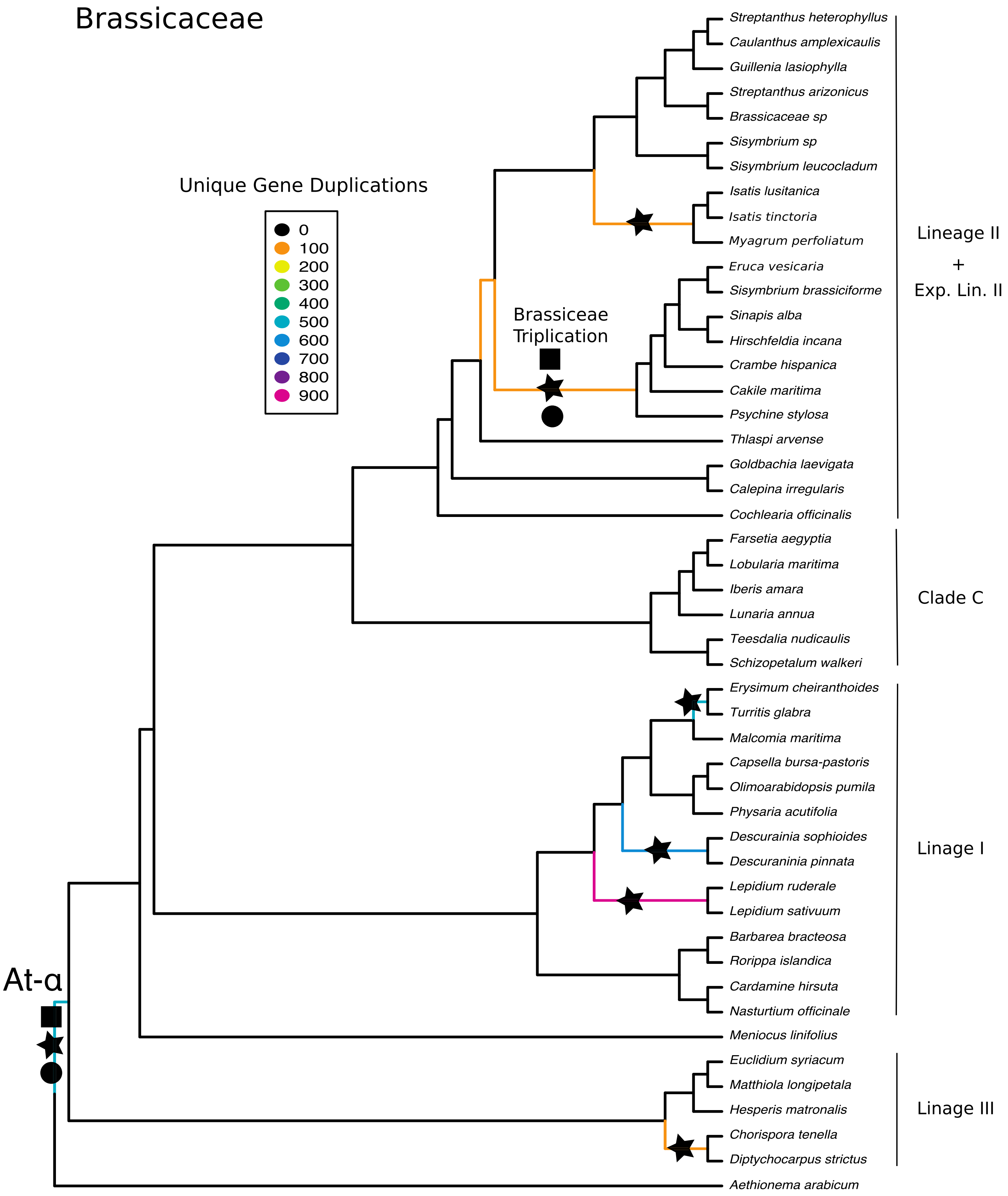


FIGURE 2. Coalescent-based species phylogeny and whole-genome duplication events of the Brassicaceae. Branch color denotes number of unique gene duplications as determined by PUG (github.com/mrmckain/PUG). WGD events identified by PUG (black stars) FASTKs (black squares; McKain et al., 2016); and by DupPipe (black circles; Barker et al., 2010) are indicated. Support values are all >0.7 local posterior probabilities.

families and major lineages were congruent with previous studies using transcriptomics (Edger et al., 2015). As with the whole-chloroplast phylogeny, we recover a clade of *Moringa oleifera* and *Carica papaya* sister to a clade of [Bataceae + Resedaceae + Capparaceae + Cleomaceae + Brassicaceae], Bataceae sister to [Resedaceae + Capparaceae + Cleomaceae + Brassicaceae], Resedaceae sister to [Capparaceae + Cleomaceae + Brassicaceae], and Capparaceae sister to [Cleomaceae + Brassicaceae]. Within Brassicaceae, the major lineages were recovered, as supported by previous literature (Huang et al., 2016; Nikolov et al., 2019), with *Aethionema arabicum* as sister to the rest of the family, followed by Lineage III sister to [Lineage I + Clade C + Lineage II and expanded Lineage II], and Lineage I sister to [Clade C + Lineage II and expanded Lineage II]. Within Cleomaceae, the relationships were mostly congruent with previous nuclear phylogenies (Patchell et al., 2014) with *Polanisia* sister to *Cleome* sensu stricto, but differing in the placement of *Gynandropsi*s (unsupported in Patchell et al., 2014). Additionally, the sampling of only four Capparaceae limited our ability to say much about the relationships within the family; however, to date, there is no phylogeny based solely on nuclear data for this group of plants. Discordance analyses of the Brassicales showed agreement among gene trees for nodes along the backbone, except within and between Clade C, Lineage II, and expanded Lineage II in the Brassicaceae.

WGD events across the Brassicales

Two popular (and most cost-effective) methods used to detect WGDs across a phylogeny include the analysis of (1) gene tree topologies and (2) K_s plots, which allowed for the identification of signatures left behind in paralogs after WGD. We used a combination of these approaches to test hypotheses of proposed WGD across the Brassicales. Using PUG (github.com/mrmckain/PUG), a gene tree topology WGD estimation method, we recovered some previously inferred events with high support (e.g., At-α and At-β) but did not find strong support for other WGD events, such as the more recent Brassiceae triplication (Fig. 1). Notably, PUG identified only 65 unique gene duplications at the Brassiceae node when only gene trees with >80% bootstrap support were considered. This number is surprisingly low when compared to At-α and At-β with counts >300 and >150, respectively. To increase the number of orthogroups used to infer species trees as well as to increase the number of gene trees to query putative paralogs against, we independently analyzed phylogenies of each family for evidence of WGD. By evaluating the Brassicaceae, Capparaceae, Cleomaceae, and [Resedaceae + Bataceae + Moriagaceae + Caricaceae] families separately, we increased gene tree counts in the analyses and improved WGD detection of previously inferred events (Fig. 2).

WGD events in Brassicaceae—Analysis of the Brassicaceae identified At-α at the base of the family and also successfully identified the Brassiceae whole-genome triplication (Lysak et al., 2005; Tang et al., 2012; Fig. 2). We also recovered neopolyploid (relatively young) events shared between (1) *Chorispora tenella* and *Diptychocarpus strictus*, (2) *Lepidium ruderale* and *L. sativum*, (3) *Descurainia sophioides* and *D. pinnata*, (4) *Turritis glabra* and *Erysimum cheiranthoides*, and (5) a clade of *Isatis lusitanica*, *I. tinctoria*, and *Myagrum perfoliatum*. Using both FASTKs to estimate pairwise K_s values (github.com/mrmckain/FASTKs; McKain et al., 2016) and DupPipe to estimate K_s values using duplications in gene trees (Barker et al., 2010), the K_s plots mostly showed agreement with the WGD events inferred by the phylogenetic method, PUG. For example, K_s plots from both analyses recovered the Brassiceae triplication (K_s ~ 0.3; Appendix S9). However, for the neopolyploid events and At-α, the K_s plots showed differing results between FASTKs and DupPipe, some with and others without evidence for WGD events (Appendix S9).

Independent WGD events in Cleomaceae—When running PUG using the Cleomaceae family, we placed Th-α as potentially shared between *T. hassleriana* and Cleomaceae sp. We also identified additional events between (1) *Coalisina paradoxa* and *C. angustifolia* (2) four species of *Polanisia*, and (3) *Cleome amblyocarpa*, *C. africana*, and *C. arabica* (Fig. 3).

Both methods of K_s estimation provided support for the placement of Th-α with peaks at ~0.4 for *T. hassleriana* and Cleomaceae sp. and also for *Melidiscus giganteus*, *G. gynandra*, and *Sieruela monophylla*, suggesting that Th-α is shared across more than just *T. hassleriana* and *Cleomaceae* sp. (Fig. 3). We did not see evidence for this peak in *A. viscosa*, which is sister to the species above. As for the other three events, the story is more complicated. When compared to *C. violacea* (which lacks evidence for Th-α; Emery et al., 2018; One Thousand Plant Transcriptomes Initiative, 2019), we concluded that there was no evidence for two of these novel events: the event shared by the four species of *Polanisia* and the event shared by *Cleome amblyocarpa*, *C. africana*, and *C. arabica*. However, the event between *Coalisina paradoxa* and *C. angustifolia* does have a signal for a WGD in the K_s plots (Fig. 3).

Due to incongruence of results for the placement of Th-α, we divided potential placements into four hypotheses, H1–H4, to test the age of ortholog divergence between taxa to the age of Th-α (K_s ~ 0.4). We found evidence that Th-α is shared with at least *T. hassleriana*, Cleomaceae sp., and *M. giganteus* and that Th-α occurred before the divergence between *M. giganteus* and *T. hassleriana* and around the same time as the divergence of *G. gynandra* and *T. hassleriana* (Th-α H2; Fig. 4A). When we compare the divergence between *A. viscosa* and *G. gynandra* to the K_s values of these three species along with *S. monophylla*, *A. viscosa*, and *G. gynandra*, we found that *A. viscosa* and *G. gynandra* diverged more recently than Th-α and that, as in earlier K_s plots, *A. viscosa* lacks evidence for Th-α (Th-α H3; Fig. 4B). This perplexing result could indicate that the data from *A. viscosa* is of poor quality or that the genome has lost such a large fraction of the duplicates that the signal for this event is not detected. To further test for the placement of Th-α, we expanded our comparisons to include the ortholog divergence of *Coalisina angustifolia* and *T. hassleriana* as well as the divergence between *C. violacea* and *T. hassleriana* to test if the proposed independent WGD events between these two clades may be a single event (Th-α H4; Fig. 4C). Surprisingly, the ortholog divergence for both pairs of taxa is about the same age as Th-α. Based on these results, Th-α either is shared across the whole clade (Th-α H4) or is two separate events that happened at approximately the same time. A comparison of ortholog divergence to K_s peaks for the two other identified WGD events using phylogenomics suggests that there is no other WGD event in the Cleomaceae (Appendix S10).

WGD in Capparaceae—In agreement with Lysak (2018), our PUG analysis recovered evidence for an independent WGD event in the Capparaceae that is shared between a species of *Capparis* and another species of Capparaceae included in our analyses (Fig. 5A). This event was supported by K_s plots using

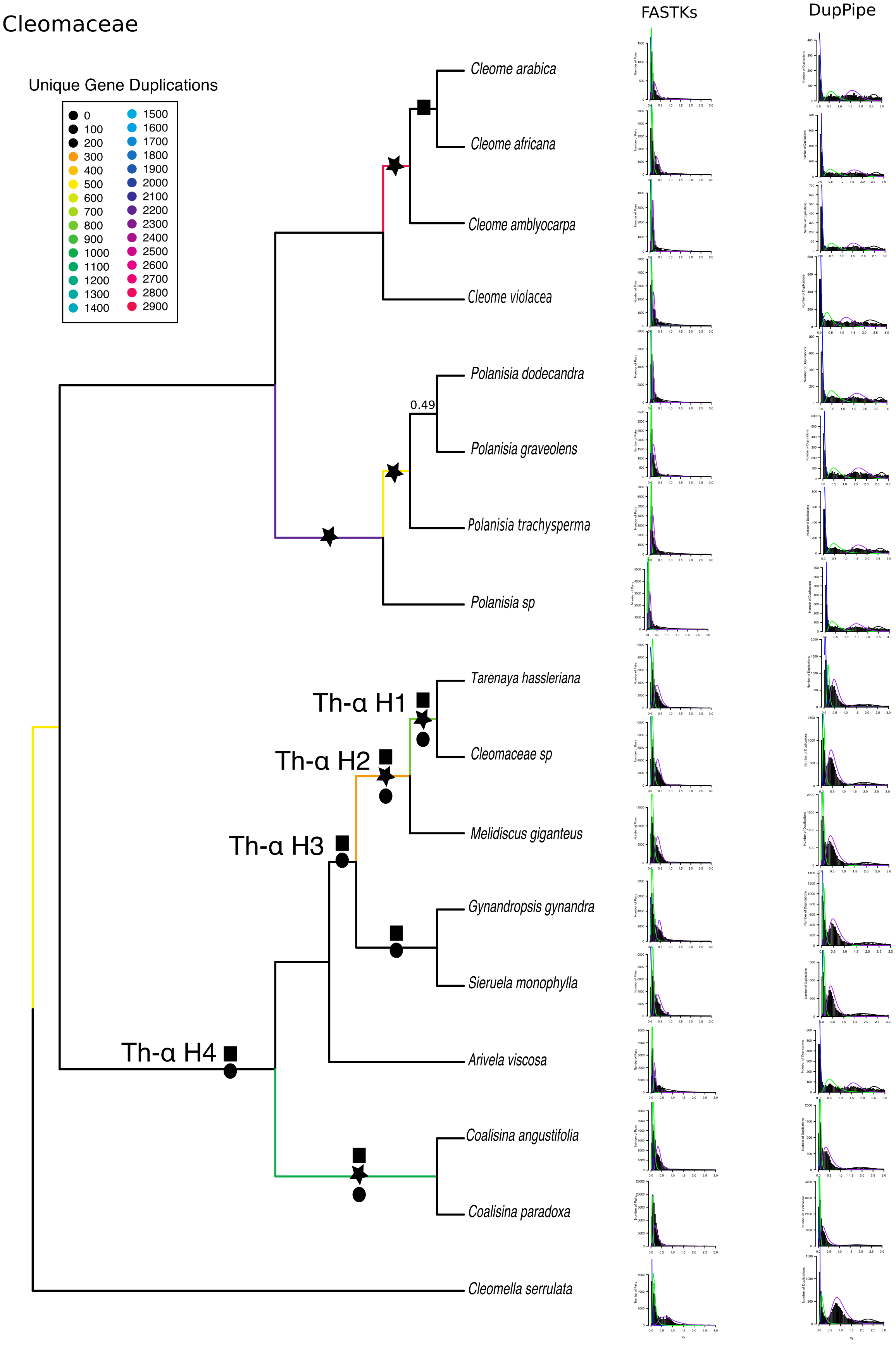

Cleomaceae
Unique Gene Duplications
0
100
200
300
400
500
600
700
800
900
1000
1100
1200
1300
1400
1500
1600
1700
1800
1900
2000
2100
2200
2300
2400
2500
2600
2700
2800
2900
FASTKs
DupPipe
Cleome arabica
Cleome africana
Cleome amblyocarpa
Cleome violacea
Polanisia dodecandra
0.49
Polanisia graveolens
Polanisia trachysperma
Polanisia sp
Tarenaya hassleriana
Th-α H1
Cleomaceae sp
Th-α H2
Melidiscus giganteus
Th-α H3
Gynandropsis gynandra
Sieruela monophylla
Arivela viscosa
Th-α H4
Coalisina angustifolia
Coalisina paradoxa
Cleomella serrulata

FIGURE 3. Coalescent-based species phylogeny and whole-genome duplication events of the Cleomaceae. Branch color denotes number of unique gene duplications as determined by PUG (github.com/mrmckain/PUG). WGD events identified by PUG (black stars), by FASTKs (black squares; McKain et al., 2016), and by DupPipe (black circles; Barker et al., 2010) are indicated. K_s plots using both FASTKs and DupPipe are placed next to their corresponding branch. *Y*-axes of K_s plots are not congruent: FASTKs measures number of pairs, and DupPipe measures numbers of duplications. Support values are indicated if <0.7 local posterior probability.

FastKs, but not DupPipe, with a peak centered at K_s ~ 0.3 (Fig. 5A). Ortholog divergences between members of the Capparaceae showed conflicting patterns. When comparing K_s values of *Boscia* sp., *Capparis fascicularis*, Capparaceae sp., and *Cadaba natalensis* from DupPipe to the ortholog divergence time between *Boscia* sp. and *Capparis fascicularis*, we found that the divergence between these two species occurs before the possible WGD event, agreeing with the PUG analysis. All four taxa shared a peak in their K_s distribution, although their K_s plots from both analyses were not in agreement, providing conflicting results for the identification of this event. The divergences tested between *Boscia* sp. and *Cadaba natalensis* and between *Capparis fascicularis* and *Cadaba natalensis* also occur before the proposed event. However, the divergence between *Capparis fascicularis* and a misidentified species of Capparaceae seems to occur at the same time as the peak in K_s values (Appendix S11).

Resedaceae specific WGD—When combining the Resedaceae (*O. barcardis* and *R. odorata*), Bataceae, Moringaceae, and Caricaceae families together, we found strong evidence for a Resedaceae-specific WGD event in all three analyses, with K_s plots indicating a peak at ~0.4 (Fig. 5B). The ortholog divergences seemed to support the proposal of this WGD as shared between the samples of Resedaceae. Both samples (*R. odorata* and *O. barcardis*) shared a K_s peak of ~0.4, which occurred before the divergence between these two samples and after the divergence between Resedaceae from *B. maritima* (Appendix S11). In addition, we recovered evidence for At-β using both PUG and DupPipe (K_s ~ 1.7; Fig. 5B).

DISCUSSION

Studies of the relationships within the Brassicales have included either many taxa but few genes (Hall et al., 2004; Hall, 2008; Cardinal-McTeague et al., 2016), a few taxa and few genes (Rodman et al., 1998), or few taxa and many genes (Edger et al., 2015, 2018). In this study, we balance taxa and genes to present a well-supported chloroplast and nuclear phylogeny for the Brassicales, both in overall agreement with previous studies at the interfamilial and intrafamilial levels (Edger et al., 2015, 2018; Cardinal-McTeague et al., 2016; Huang et al., 2016; Guo et al., 2017). Using the nuclear phylogeny, we highlight several potential placements of the Th-α WGD event and identify other possible novel events in the Cleomaceae, Capparaceae, and Resedaceae.

Incongruences between the chloroplast and nuclear trees across the Brassicaceae

Although relationships in our nuclear and chloroplast phylogenies are congruent with previous analyses, we highlight incongruence *between* the nuclear and chloroplast trees among the major lineages of the Brassicaceae, a well-documented pattern between these genomes (Beilstein et al., 2008; Huang et al., 2016; Nikolov et al., 2019; summarized in Fig. 6). We find Lineage I sister to [Lineage III + Lineage II + Expanded Lineage II + Clade C] in the chloroplast tree and Lineage III sister to [Lineage I + Lineage II + Expanded Lineage II + Clade C] in the nuclear tree. Huang et al. (2016), using 113 low-copy nuclear genes from 55 Brassicaceae species, recovered a tree congruent with our nuclear phylogeny, as did Nikolov et al. (2019) in their study using 79 species and 1421 exons. Additionally, Guo et al. (2017), using 77 chloroplast genes from 53 samples, recovered a phylogeny in agreement with our chloroplast tree. With additional taxon sampling, an increase in data, and using the same samples across analyses, we too recover incongruent relationships between nuclear and chloroplast trees, leaving us to conclude that the trees from these genomes are in disagreement due to different evolutionary histories, which could include ancient hybridization or introgression events (Forsythe et al., 2018 [Preprint]). For future users of these phylogenies, the differences between these two trees are important to consider when using the phylogeny to assess character evolution and divergence dating, as node ordering depends on which tree is used.

Putative placement of Th-α in the Cleomaceae

Previous studies have identified a WGD event unique to Cleomaceae (Th-α) using a variety of sources from syntenic regions to ESTs (Schranz and Mitchell-Olds, 2006; Barker et al., 2009; reviewed in Bayat et al., 2018; One Thousand Plant Transcriptomes Initiative, 2019). Placement of Th-α within the Cleomaceae had yet to be confirmed. By using K_s plots to assess for the signatures of WGD left behind in the paralogs they created, phylogenetics using individual gene tree topologies, gene counts, a known species tree, and ortholog divergences, we are able to putatively place Th-α as shared between *T. hassleriana*, Cleomaceae sp., *Melidiscus giganteus*, *G. gynandra*, and *S. monophyla*, *A. viscosa*, *Coalisina angustifolia*, and *Coalisina paradoxa* (Th-α H4; Fig. 3). We include these last three species due to both the evidence from ortholog divergences and the signatures in K_s plots that strongly suggest this event is shared with all species (Figs. 3 and 4). It is possible that two separate events occurred independently and that *A. viscosa* does indeed lack a WGD. Although this hypothesis (Th-α H3) is in agreement with the most recent analysis of this group by the One Thousand Plant Transcriptomes Initiative (2019), their sampling of the Cleomaceae was limited and did not include *Coalisina angustifolia* and *C. paradoxa*. K_s plots of all samples, other than *A. viscosa*, identify a peak at K_s ~ 0.4, agreeing with previous studies (Barker et al., 2009; van den Bergh et al., 2014) that first identified this peak in *T. hassleriana* followed by *G. gynandra*. PUG, however, supports two separate events. The difficulty in placing this event may be due to the gene-tree discordance found within this clade (Fig. 1) or to the possibility that, like others in the order, it is a whole-genome triplication, in which case it will be hard to tease apart due to having to resolve two serial duplication events that occurred in a short period (as with the tribe Brassiceae; Tang et al., 2012).

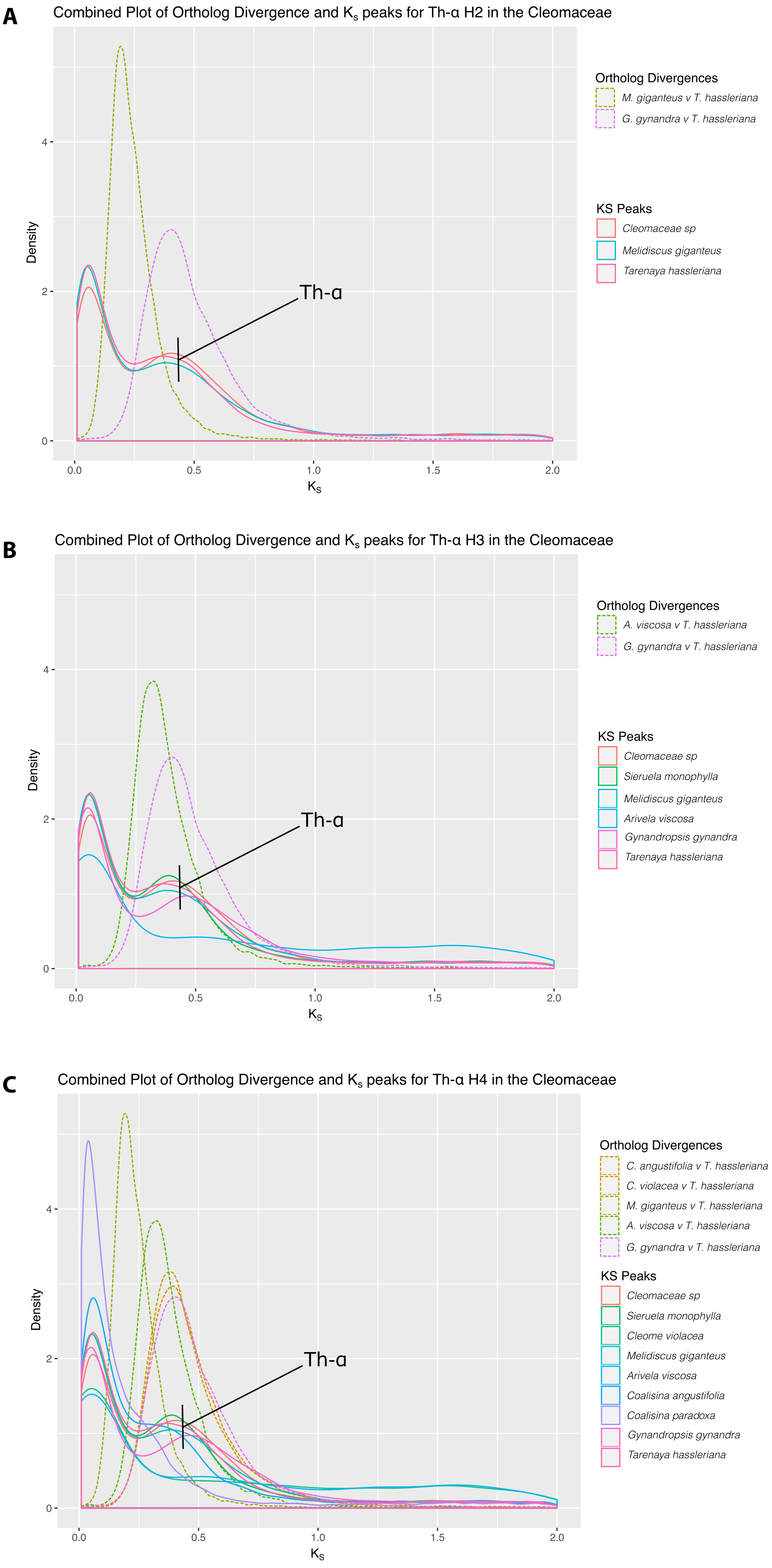

A
Combined Plot of Ortholog Divergence and K_s peaks for Th-α H2 in the Cleomaceae
Density
0
2
4
0.0
0.5
1.0
1.5
2.0
K_S
Th-α
Ortholog Divergences
M. giganteus v T. hassleriana
G. gynandra v T. hassleriana
KS Peaks
Cleomaceae sp
Melidiscus giganteus
Tarenaya hassleriana
B
Combined Plot of Ortholog Divergence and K_s peaks for Th-α H3 in the Cleomaceae
Density
0
2
4
0.0
0.5
1.0
1.5
2.0
K_S
Th-α
Ortholog Divergences
A. viscosa v T. hassleriana
G. gynandra v T. hassleriana
KS Peaks
Cleomaceae sp
Sieruela monophylla
Melidiscus giganteus
Arivela viscosa
Gynandropsis gynandra
Tarenaya hassleriana
C
Combined Plot of Ortholog Divergence and K_s peaks for Th-α H4 in the Cleomaceae
Density
0
2
4
0.0
0.5
1.0
1.5
2.0
K_S
Th-α
Ortholog Divergences
C. angustifolia v T. hassleriana
C. violacea v T. hassleriana
M. giganteus v T. hassleriana
A. viscosa v T. hassleriana
G. gynandra v T. hassleriana
KS Peaks
Cleomaceae sp
Sieruela monophylla
Cleome violacea
Melidiscus giganteus
Arivela viscosa
Coalisina angustifolia
Coalisina paradoxa
Gynandropsis gynandra
Tarenaya hassleriana

FIGURE 4. Comparison of ortholog divergences and K_s peaks of the Cleomaceae to test hypotheses of placement for Th-α. (A) Testing H2 by comparison of ortholog divergences of *Melidiscus giganteus* and *Gynandropsis gynandra* to *Tarenaya hassleriana* compared with K_s peaks of Cleomaceae sp., *M. giganteus*, and *T. hassleriana*. (B) Testing of H3 by comparison of ortholog divergence between *Arivela viscosa* and *T. hassleriana*, and *G. gynandra* to *T. hassleriana* with K_s values of Cleomaceae sp., *Sieruela monophylla*, *M. giganteus*, *A. viscosa*, *G. gynandra*, and *T. hassleriana*. (C) Testing the H4 hypothesis for placement of Th-α.

Multiple WGD events in the Cleomaceae?

In addition to placing Th-α, we report two possible additional WGD events in the Cleomaceae. Both events are identified in the Brassicales and the Cleomaceae analyses, but with much greater support in the analysis of the Cleomaceae species. These WGD events are placed at common ancestors shared between (1) *Cleome amblyocarpa*, *C. africana*, and *C. arabica* and (2) four species of *Polanisia* (Fig. 3). K_s plots provide contrasting support for these events. K_s plots from FASTKs of *C. africana* and *C. arabica* show a small peak of duplicates at K_s ~ 0.3. Yet there is no evidence of a WGD event when the same data are run through DupPipe. The K_s plots of *C. amblyocarpa* also give conflicting evidence for this event. The K_s plot from FASTKs looks much more similar to that of *C. violacea*, which shows no evidence of a recent WGD event (based on genome sequencing; Emery et al., 2018). The second event shared between the four species of *Polanisia* is supported by a large number (2200) of unique gene duplications using PUG but not by K_s plots from either FASTKs or DupPipe; the resulting plots look more similar to *C. violacea*. Analyses of ortholog divergence between *C. amblyocarpa*, *C. africana*, and *C. arabica* also lack support for a WGD (Appendix S10), as do analyses between the four species of *Polanisia* (Appendix S10). To further test how WGD and C_4 photosynthesis have evolved in this family, we suggest a study focusing primarily on Cleomaceae sampling. C_4 photosynthesis has evolved at least three times independently in Cleomaceae, specifically in (of the taxa sampled) *G. gynandra* and *Coalisina angustifolia* with *C. paradoxa* as a C_3-C_4 intermediate in anatomy and physiology (Bhide et al., 2014). If our putative placement of Th-α is correct, then all of these samples share this event. It will be interesting to investigate the role of polyploidy and, more specifically, Th-α, in character evolution in this group.

WGD events in the Capparaceae and Resedaceae

Although our analysis included only two samples, we recover some support for an event between at least one species of *Capparis* and the misidentified species of Capparaceae (Fig. 5A). Given this ambiguous species identification, inconclusiveness from K_s plots, and no support in comparison between ortholog divergence and K_s peaks, this event, although supported by many unique duplicates in the PUG analysis, should be interpreted carefully. That said, Lysak (2018), using chromosome counts, also proposed that Capparaceae experienced a unique event; however, chromosome counts alone may be misleading in concluding that a WGD event has occurred (Evans et al., 2017). There is also a lack of agreement between K_s plots derived using FASTKs and DupPipe, which estimate K_s values differently (i.e., pairwise K_s estimates in FASTKs vs. estimates of K_s at nodes in gene trees in DupPipe), further confounding evidence for either presence or absence of a Capparaceae-specific event. Between information presented by Lysak (2018) and the evidence presented here, this possible event certainly warrants additional study.

We find good evidence to support the presence of a separate WGD event in the Resedaceae, also hypothesized by Lysak (2018). This polyploidy is one of the few events recovered with consensus between K_s plots (from both FASTKs and DupPipe), phylogenetics, and ortholog divergences (Fig. 4B and Appendix S11). Therefore, we are confident in the identification of this event (named Rs-α here). This event was recently identified using a single species (*Reseda odorata*; One Thousand Plant Transcriptomes Initiative, 2019); however, we determine that it is shared with at least one other species in the family, *Ochradenus barcardis*. The sister families, Caricaceae and Moringaceae, show no evidence of unique WGD events, which is in agreement with the recent whole-genome sequencing of *Moringa oleifera* (Chang et al., 2018). When Tian et al. (2015) compared the papaya genome, which shows no evidence of a (recent) WGD (Ming et al., 2008), to the newly sequenced genome of *M. oleifera*, they too concluded that Moringaceae did not experience a family-specific genome duplication. The identified and well-supported Resedaceae event warrants additional sampling and investigation to test if it is shared across the whole family.

Methodological challenges with placing WGD events: sampling matters

Currently, three types of methods are used to detect WGD events: (1) K_s plots to assess for signatures left behind in paralogs after WGD, (2) identification of retained duplicate blocks in a genome, and (3) phylogenetics using individual gene-tree topologies. Since all three methods have limitations in identifying WGD events, we use a combination of approaches to test hypotheses, reduce proposing events that may not exist, and simultaneously provide multiple lines of evidence for recovered events.

Recently, an abundance of papers have highlighted the difficulties and complexities of determining WGD events across the tree of life (Conover et al., 2018; Tiley et al., 2018; Li and Barker, 2019; Li et al., 2019; Nakatani and McLysaght, 2019; Zwaenepoel and van de Peer, 2019; Zwaenepoel et al., 2019). We add another dimension to this conversation by demonstrating that the different taxonomic levels from which we sample, such as the order or family, make a difference in support of previously identified WGDs (i.e., the Brassiceae triplication). Recent research has demonstrated that differences in taxonomic sampling and taxon occupancy in data matrices can influence the inference of WGDs, particularly if adding taxa decreases taxon occupancy in gene families (Yang et al., 2015; Li et al., 2018; Li and Barker, 2019; Zwaenepoel and van de Peer, 2019). Testing for WGD events across the Brassicales phylogeny led to less certain topologies; therefore, signals of WGD are missed when filtering for nodes with only high bootstrap support to count duplicates. To account for this and to increase taxon and gene-family occupancy in our data sets, we reduce sampling to just the family level to infer WGD. However, at each level of analysis, we choose an arbitrary cutoff for the number of duplicates that we feel is sufficient to infer a WGD event, a

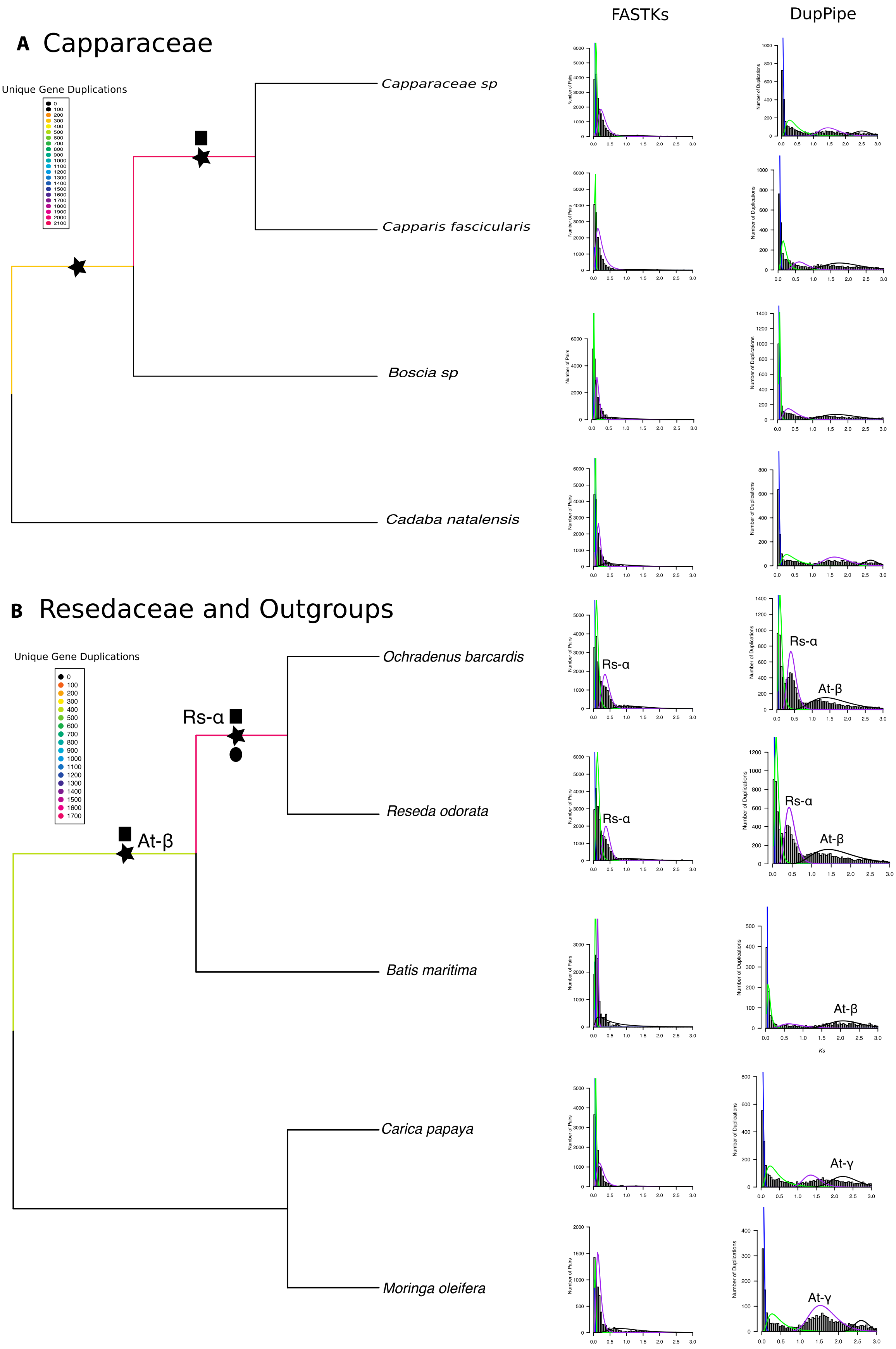

FASTKs
DupPipe
A Capparaceae
Unique Gene Duplications
0
100
200
300
400
500
600
700
800
900
1000
1100
1200
1300
1400
1500
1600
1700
1800
1900
2000
2100
Capparaceae sp
Capparis fascicularis
Boscia sp
Cadaba natalensis
B Resedaceae and Outgroups
Unique Gene Duplications
0
100
200
300
400
500
600
700
800
900
1000
1100
1200
1300
1400
1500
1600
1700
Rs-α
At-β
Ochradenus barcardis
Reseda odorata
Batis maritima
Carica papaya
Moringa oleifera
Rs-α
At-β
At-γ
Number of Pairs
Number of Duplications
Ks
0.0
0.5
1.0
1.5
2.0
2.5
3.0

FIGURE 5. Coalescent-based species phylogenies and whole-genome duplication events of the (A) Capparaceae and (B) Resedaceae + Outgroups. Branch color denotes number of unique gene duplications as determined by PUG (github.com/mrmckain/PUG). WGD events identified by PUG (black stars), by FASTKs (black squares; McKain et al., 2016), and by DupPipe (black circles; Barker et al., 2010) are indicated. K_s plots using both FASTKs and DupPipe are placed next to their corresponding branch. *Y*-axes of K_s plots are not congruent: FASTKs measures number of pairs, and DupPipe measures numbers of duplications. At-β and At-γ events are noted above corresponding peaks in the K_s plots. Support values are all >0.7 local posterior probabilities.

documented criticism of these types of methods (Zwaenepoel and van de Peer, 2019). Others note that it is also important to consider heterogeneity in substitution rates (Cui et al., 2006; Barker et al., 2009; Yang et al., 2015), but phylogenomic methods and denser sampling should obviate the need for rate corrections. Variation in the duplication and loss rate across the species tree may also impact tests for WGD events (Li et al., 2018; Zwaenepoel and van de Peer, 2019).

Although our K_s-based inferences of WGDs are largely consistent with phylogenomic inferences, there are some differences among the approaches. FASTKs and DupPipe use different estimates of K_s that likely produce the observable differences in the respective K_s plots. FASTKs uses a pairwise approach for estimating K_s values (github.com/mrmckain/FASTKs; McKain et al., 2016), whereas DupPipe estimates K_s values from nodes of gene trees (Barker et al., 2010). Tiley et al. (2018) explored the difference in K_s estimates from these types

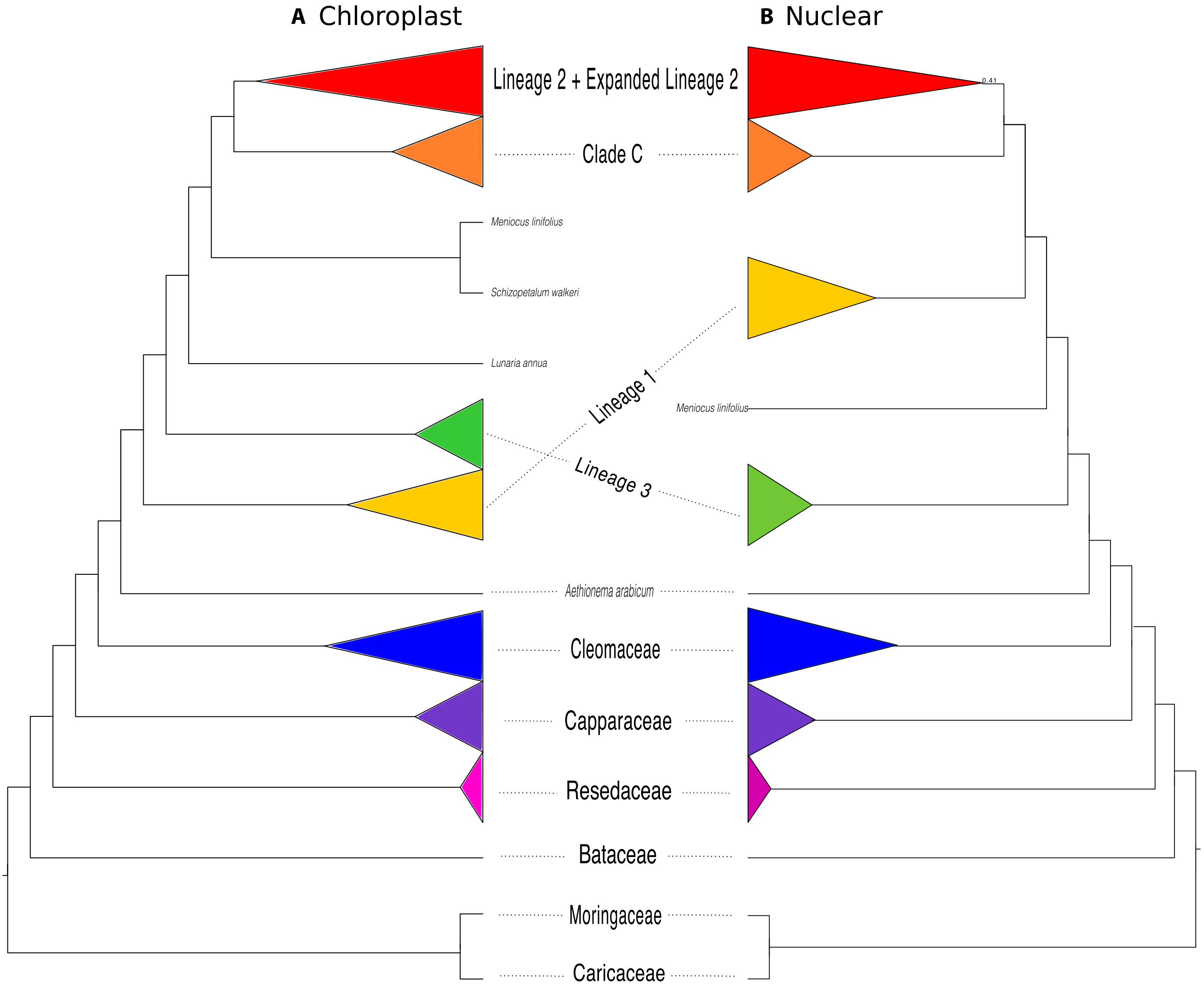


FIGURE 6. Comparison of (A) maximum likelihood whole-chloroplast phylogeny to (B) coalescent-based species phylogeny of the Brassicales. Major lineages and clades of the Brassicaceae are indicated. Support values are indicated if <0.7 local posterior probabilities or <70% bootstrap support.

of approaches and found that the observed differences in peaks of duplications between the two different methods is consistent with simulations. The node-based estimates of K_s from DupPipe often yield sharper peaks in putative WGDs, with overall lower numbers of duplications because of the difference in number of nodes vs. pairwise comparisons. However, the results of both approaches are largely consistent after close inspection. Perhaps more confounding for K_s analyses is the interpretation of mixture models to identify putative peaks associated with a WGD. Mixture models, which are typically fit to the distribution of duplicates, tend to overestimate the number of true peaks (Naik et al., 2007; Tiley et al., 2018; Zwaenepoel et al., 2019). Using two different methods across multiple species allowed us to evaluate and compare putative peaks from different analyses to identify the expected signatures of WGDs. Furthermore, because paralogs from WGDs tend to be more highly expressed than those resulting from tandem duplications (Casneuf et al., 2006), transcriptomes may yield a data set that is more enriched for WGD duplicates than (fragmented) genomic data. Therefore, transcriptome data, as shown by Tiley et al. (2018), may actually improve success in detecting WGD events.

CONCLUSIONS

The Brassicales is an excellent group for comparing methods of WGD identification because of the wealth of genomic data and the previously inferred WGDs that are available. With many chromosome-level genomes available, analyses based on synteny, which seem to be regarded as most reliable in detecting these events (Nakatani and McLysaght, 2019), can be used as controls for comparing WGD methods. Sequenced genomes, which are placed throughout the Brassicales, provide strong evidence for taxa that we know do not have recent WGD events (i.e., *Cleome violacea* and *Carica papaya*) and for taxa that do show evidence for recent WGD events (i.e., *A. thaliana* and many *Brassica* crops). These resources provide calibration points that can be used to verify results when testing for novel events. This group of plants, combined with recent insights on difficulties in placing WGD events, can help further the development of innovative methods in describing and identifying WGDs.

ACKNOWLEDGMENTS

We thank our funding sources, including National Science Foundation (NSF) grants IOS-1339156 and EF-1550838 to M.S.B., NSF-DEB 1737898 to P.P.E., NSF IOS-1811784 to P.D.B., NIH PERT K12GM000708 to B.S., and NSF DEB-1146603 to J.C.P. We thank the staff from the University of Missouri, specifically N. Bivens, M.-Y. Zhou, and K. Bromert with the DNA Core, for their patience and assistance in getting successful libraries for all samples; J. Gotburg with the Research Computing Support Services for assistance in running large jobs across the Lewis cluster; and B. Sonderman with the Division of Biological Sciences for assistance in keeping many of these plants alive. Finally, we acknowledge S. Unruh and M. Kroll and our anonymous reviewers for providing feedback that improved the manuscript.

AUTHOR CONTRIBUTIONS

M.E.M., J.C.P., G.C.C., J.C.H., P.P.E., and M.E.S. designed the project. M.E.M., J.M.B., H.A., J.D.W., and W.T.D. grew plants, collected tissue, and isolated RNA. M.E.M. and J.M.B. isolated DNA. J.C.H. contributed *Cleomella serrulate*. M.E.M. and J.M.B. analyzed the data. P.D.B. and J.D.W. assisted with processing and analyzing the data. C.A.B. wrote alignment filtering scripts. M.S.B. and B.S. analyzed data using DupPipe and OrthoPipe. M.R.M. helped with processing data using FastPlast. I.A. and J.C.H. updated nomenclature and helped correct plant identifications. M.E.M. wrote the original manuscript draft, which was read and approved by all authors.

DATA AVAILABILITY

Sequence data from this article can be found in the NCBI SRA data libraries under BioProject accession number PRJNA542714. Individual BioSample accession numbers can be found in Appendix S2. Seeds corresponding to samples are available upon request.

SUPPORTING INFORMATION

Additional Supporting Information may be found online in the supporting information tab for this article.

APPENDIX S1. Current understanding of the phylogenetic relationships between the 17 families of the Brassicales and whole-genome duplication events.

APPENDIX S2. Taxon sampling, accessions, and additional analysis information.

APPENDIX S3. RNA and DNA extraction method, library preparation method, sequencing method, read size, and raw read numbers.

APPENDIX S4. Orthogroups retained for each analysis.

APPENDIX S5. BIC scores for 1–4 components for both FASTKs (McKain et al., 2016) and DupPipe (Barker et al., 2010).

APPENDIX S6. BUSCO analysis of de novo transcriptomes.

APPENDIX S7. Maximum likelihood phylogeny of the Brassicales using two chloroplast genes, *MatK* and *NdhF*.

APPENDIX S8. Maximum likelihood whole chloroplast phylogeny of the Brassicales.

APPENDIX S9. Brassicaceae K_s plots using both FASTKs (McKain et al., 2016) and DupPipe (Barker et al., 2010).

APPENDIX S10. Additional ortholog divergences and K_s peaks of the Cleomaceae.

APPENDIX S11. Ortholog divergences and K_s peaks of the (A) Capparaceae and (B) Resedaceae + Outgroups.

LITERATURE CITED

APG IV. 2016. An update of the Angiosperm Phylogeny Group classification for the orders and families of flowering plants: APG IV. *Botanical Journal of the Linnean Society* 181: 1–20.

Arabidopsis Genome Initiative. 2000. Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature* 408: 796.

Bankevich, A., S. Nurk, D. Antipov, A. A. Gurevich, M. Dvorkin, A. S. Kulikov, V. M. Lesin, et al. 2012. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *Journal of Computational Biology* 19: 455–477.

Barco, B., and N. K. Clay. 2019. Evolution of glucosinolate diversity via whole-genome duplications, gene rearrangements, and substrate promiscuity. *Annual Review of Plant Biology* 70: 585–604.

Barker, M. S., K. M. Dlugosch, L. Dinh, R. S. Challa, N. C. Kane, M. G. King, and L. H. Rieseberg. 2010. EvoPipes.net: bioinformatic tools for ecological and evolutionary genomics. *Evolutionary Bioinformatics* 6: EBO-S5861.

Barker, M. S., H. Vogel, and M. E. Schranz. 2009. Paleopolyploidy in the Brassicales: analyses of the Cleome transcriptome elucidate the history of genome duplications in *Arabidopsis* and other Brassicales. *Genome Biology and Evolution* 1: 391–399.

Bayat, S., M. E. Schranz, E. H. Roalson, and J. C. Hall. 2018. Lessons from Cleomaceae, the sister of crucifers. *Trends in Plant Science* 23: 808–821.

Beilstein, M. A., I. A. Al-Shehbaz, and E. A. Kellogg. 2006. Brassicaceae phylogeny and trichome evolution. *American Journal of Botany* 93: 607–619.

Beilstein, M. A., I. A. Al-Shehbaz, S. Mathews, and E. A. Kellogg. 2008. Brassicaceae phylogeny inferred from phytochrome A and ndhF sequence data: tribes and trichomes revisited. *American Journal of Botany* 95: 1307–1327.

Bhide, A., S. Schliesky, M. Reich, A. P. Weber, and A. Becker. 2014. Analysis of the floral transcriptome of *Tarenaya hassleriana* (Cleomaceae), a member of the sister group to the Brassicaceae: towards understanding the base of morphological diversity in Brassicales. *BMC Genomics* 15: 140.

Blažević, I., S. Montaut, F. Burčul, C. E. Olsen, M. Burow, P. Rollin, and N. Agerbirk. 2019. Glucosinolate structural diversity, identification, chemical synthesis and metabolism in plants. *Phytochemistry* 112100.

Bolger, A. M., M. Lohse, and B. Usadel. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30: 2114–2120.

BrassiBase. Tools and biological resources for Brassicaceae character and trait studies. Website: https://brassibase.cos.uni-heidelberg.de. Accessed 24 March 2019.

Brock, K. C. 2014. Tracking the Evolutionary History of Development Genes: Implications for the Diversification of Fruits and Flowers in the Brassicaceae and Cleomaceae. *Master's thesis*. University of Alberta, Edmonton, Alberta, Canada.

Buchfink, B., C. Xie, and D. H. Huson. 2015. Fast and sensitive protein alignment using DIAMOND. *Nature Methods* 12: 59.

Cardinal-McTeague, W., K. J. Sytsma, and J. C. Hall. 2016. Biogeography and diversification of Brassicales: a 103 million year tale. *Molecular Phylogenetics and Evolution* 99: 204–224.

Casneuf, T., S. De Bodt, J. Raes, S. Maere, and Y. Van de Peer. 2006. Nonrandom divergence of gene expression following gene and genome duplications in the flowering plant *Arabidopsis thaliana*. *Genome Biology* 7: R13.

Chang, Y., H. Liu, M. Liu, X. Liao, S. K. Sahu, Y. Fu, B. Song, et al. 2018. The draft genomes of five agriculturally important African orphan crops. *GigaScience* 8: giy152.

Cheng, S., E. van den Bergh, P. Zeng, X. Zhong, J. Xu, X. Liu, J. Hofberger, et al. 2013. The *Tarenaya hassleriana* genome provides insight into reproductive trait and genome evolution of crucifers. *The Plant Cell* 25: 2813–2830.

Conover, J. L., N. Karimi, N. Stenz, C. Ané, C. E. Grover, C. Skema, J. A. Tate, et al. 2018. A Malvaceae mystery: a mallow maelstrom of genome multiplications and maybe misleading methods? *Journal of Integrative Plant Biology* 61: 12–31.

Cui, L., P. K. Wall, J. H. Leebens-Mack, B. G. Lindsay, D. E. Soltis, J. J. Doyle, P. S. Soltis, et al. 2006. Widespread genome duplications throughout the history of flowering plants. *Genome Research* 16: 738–749.

Duarte, J. M., P. K. Wall, P. P. Edger, L. L. Landherr, H. Ma, J. C. Pires, and J. Leebens-Mack. 2010. Identification of shared single copy nuclear genes in *Arabidopsis*, *Populus*, *Vitis* and *Oryza* and their phylogenetic utility across various taxonomic levels. *BMC Evolutionary Biology* 10: 61.

Dunn, C. W., A. Hejnol, D. Q. Matus, K. Pang, W. E. Browne, S. A. Smith, E. Seaver, et al. 2008. Broad phylogenomic sampling improves resolution of the animal tree of life. *Nature* 452: 745.

Dunn, C., M. Howison, and F. Zapata. 2013. Agalma: an automated phylogenomics workflow. *BMC Bioinformatics* 14: 330.

Edger, P. P., J. C. Hall, A. Harkess, M. Tang, J. Coombs, S. Mohammadin, M. E. Schranz, et al. 2018. Brassicales phylogeny inferred from 72 plastid genes: a reanalysis of the phylogenetic localization of two paleopolyploid events and origin of novel chemical defenses. *American Journal of Botany* 105: 463–469.

Edger, P. P., H. M. Heidel-Fischer, M. Bekaert, J. Rota, G. Glöckner, A. E. Platts, D. G. Heckel, et al. 2015. The butterfly plant arms-race escalated by gene and genome duplications. *Proceedings of the National Academy of Sciences* 112: 8362–8366.

Emery, M., M. M. S. Willis, Y. Hao, K. Barry, K. Oakgrove, Y. Peng, J. Schmutz, et al. 2018. Preferential retention of genes from one parental genome after polyploidy illustrates the nature and scope of the genomic conflicts induced by hybridization. *PLoS Genetics* 14: e1007267.

Emms, D. M., and S. Kelly. 2015. OrthoFinder: solving fundamental biases in whole-genome comparisons dramatically improves orthogroup inference accuracy. *Genome Biology* 16: 157.

Emms, D. M., and S. Kelly. 2019. OrthoFinder: phylogenetic orthology inference for comparative genomics. *Genome Biology* 20: 1–14.

Endress, P. K. 1992. Evolution and floral diversity: the phylogenetic surroundings of *Arabidopsis* and *Antirrhinum*. *International Journal of Plant Sciences* 153: S106–S122.

Evans, B. J., N. S. Upham, G. B. Golding, R. A. Ojeda, and A. A. Ojeda. 2017. Evolution of the largest mammalian genome. *Genome Biology and Evolution* 9: 1711–1724.

Feodorova, T. A., E. V. Voznesenskaya, G. E. Edwards, and E. H. Roalson. 2010. Biogeographic patterns of diversification and the origins of C4 in *Cleome* (Cleomaceae). *Systematic Botany* 35: 811–826.

Forsythe, E. S., A. D. Nelson, and M. A. Beilstein. 2018. Biased gene retention in the face of massive nuclear introgression obscures species relationships. *bioRxiv* 197087 [Preprint].

Fraley, C., and A. E. Raftery. 2002. Model-based clustering, discriminant analysis, and density estimation. *Journal of the American Statistical Association* 97: 611–631.

Fraley, C., A. E. Raftery, and L. Scrucca. 2012. Normal mixture modeling for model-based clustering, classification, and density estimation. *Department of Statistics University of Washington* 23: 2012.

Godden, G. T., T. J. Kinser, P. S. Soltis, and D. E. Soltis. 2019. Phylotranscriptomic analyses reveal asymmetrical gene duplication dynamics and signatures of ancient polyploidy in mints. *Genome Biology and Evolution* 11: 3393–3408.

Grabherr, M. G., B. J. Haas, M. Yassour, J. Z. Levin, D. A. Thompson, I. Amit, X. Adiconis, et al. 2011. Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nature Biotechnology* 29: 644.

Guo, X., J. Liu, G. Hao, L. Zhang, K. Mao, X. Wang, D. Zhang, et al. 2017. Plastome phylogeny and early diversification of Brassicaceae. *BMC Genomics* 18: 176.

Hall, J. C. 2008. Systematics of Capparaceae and Cleomaceae: an evaluation of the generic delimitations of Capparis and Cleome using plastid DNA sequence data. *Botany* 86: 682–696.

Hall, J. C., H. H. Iltis, and K. J. Sytsma. 2004. Molecular phylogenetics of core Brassicales, placement of orphan genera Emblingia, Forchhammeria, Tirania, and character evolution. *Systematic Botany* 29: 654–669.

Hall, J. C., K. J. Sytsma, and H. H. Iltis. 2002. Phylogeny of Capparaceae and Brassicaceae based on chloroplast sequence data. *American Journal of Botany* 89: 1826–1842.

Haudry, A., A. E. Platts, E. Vello, D. R. Hoen, M. Leclercq, R. J. Williamson, E. Forczek, et al. 2013. An atlas of over 90,000 conserved noncoding sequences provides insight into crucifer regulatory regions. *Nature Genetics* 45: 891.

Huang, C. H., R. Sun, Y. Hu, L. Zeng, N. Zhang, L. Cai, Q. Zhang, et al. 2016. Resolution of Brassicaceae phylogeny using nuclear genes uncovers nested radiations and supports convergent morphological evolution. *Molecular Biology and Evolution* 33: 394–412.

Kagale, S., S. J. Robinson, J. Nixon, R. Xiao, T. Huebert, J. Condie, D. Kessler, et al. 2014. Polyploid evolution of the Brassicaceae during the Cenozoic era. *The Plant Cell* 26: 2777–2791.

Katoh, K., K. Misawa, K. I. Kuma, and T. Miyata. 2002. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Research* 30: 3059–3066.

Kearse, M., R. Moir, A. Wilson, S. Stones-Havas, M. Cheung, S. Sturrock, S. Buxton, et al. 2012. Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* 28: 1647–1649.

Kliebenstein, D. J., V. M. Lambrix, M. Reichelt, J. Gershenzon, and T. Mitchell-Olds. 2001. Gene duplication in the diversification of secondary metabolism: tandem 2-oxoglutarate–dependent dioxygenases control glucosinolate biosynthesis in Arabidopsis. *The Plant Cell* 13: 681–693.

Kocot, K. M., M. R. Citarella, L. L. Moroz, and K. M. Halanych. 2013. PhyloTreePruner: a phylogenetic tree-based approach for selection of orthologous sequences for phylogenomics. *Evolutionary Bioinformatics* 9: EBO-S12813.

Langmead, B., and S. L. Salzberg. 2012. Fast gapped-read alignment with Bowtie2. *Nature Methods* 9: 357.

Li, Z., and M. S. Barker. 2019. Inferring putative ancient whole genome duplications in the 1000 Plants (1KP) initiative: access to gene family phylogenies and age distributions. *GigaScience* 9: giaa004.

Li, Z., G. P. Tiley, S. R. Galuska, C. R. Reardon, T. I. Kidder, R. J. Rundell, and M. S. Barker. 2018. Multiple large-scale gene and genome duplications during the evolution of hexapods. *Proceedings of the National Academy of Sciences* 115: 4713–4718.

Li, Z., G. P. Tiley, R. J. Rundell, and M. S. Barker. 2019. Reply to Nakatani and McLysaght: Analyzing deep duplication events. *Proceedings of the National Academy of Sciences* 116: 1819–1820.

Lysak, M. A. 2018. Brassicales: an update on chromosomal evolution and ancient polyploidy. *Plant Systematics and Evolution* 304: 757–762.

Lysak, M. A., M. A. Koch, A. Pecinka, and I. Schubert. 2005. Chromosome triplication found across the tribe Brassiceae. *Genome Research* 15: 516–525.

Magallon, S., P. R. Crane, and P. S. Herendeen. 1999. Phylogenetic pattern, diversity, and diversification of eudicots. *Annals of the Missouri Botanical Garden* 86: 297–372.

Mandáková, T., Z. Li, M. S. Barker, and M. A. Lysak. 2017. Diverse genome organization following 13 independent mesopolyploid events in Brassicaceae contrasts with convergent patterns of gene retention. *The Plant Journal* 91: 3–21.

Martín-Bravo, S., H. Meimberg, M. Luceño, W. Märkl, V. Valcárcel, C. Bräuchler, P. Vargas, and G. Heubl. 2007. Molecular systematics and biogeography of Resedaceae based on ITS and trnL-F sequences. *Molecular Phylogenetics and Evolution* 44: 1105–1120.

Martín-Bravo, S., P. Vargas, and M. Luceño. 2009. Is Oligomeris (Resedaceae) indigenous to North America? Molecular evidence for a natural colonization from the Old World. *American Journal of Botany* 96: 507–518.

McKain, M. R., H. Tang, J. R. McNeal, S. Ayyampalayam, J. I. Davis, C. W. dePamphilis, T. J. Givnish, et al. 2016. A phylogenomic assessment of ancient polyploidy and genome evolution across the Poales. *Genome Biology and Evolution* 8: 1150–1164.

McKain, M. R., N. Wickett, Y. Zhang, S. Ayyampalayam, W. R. McCombie, M. W. Chase, J. C. Pires, et al. 2012. Phylogenomic analysis of transcriptome data elucidates co-occurrence of a paleopolyploid event and the origin of bimodal karyotypes in Agavoideae (Asparagaceae). *American Journal of Botany* 99: 397–406.

McKain, M. R., and M. Wilson. 2017. Fast-Plast: rapid *de novo* assembly and finishing for whole chloroplast genomes. Version 1.2.6. Website: https://github.com/mrmckain/Fast-Plast.

Ming, R., S. Hou, Y. Feng, Q. Yu, A. Dionne-Laporte, J. H. Saw, P. Senin, et al. 2008. The draft genome of the transgenic tropical fruit tree papaya (*Carica papaya* Linnaeus). *Nature* 452: 991.

Mithen, R., R. Bennett, and J. Marquez. 2010. Glucosinolate biochemical diversity and innovation in the Brassicales. *Phytochemistry* 71: 2074–2086.

Naik, P. A., P. Shi, and C. L. Tsai. 2007. Extending the Akaike information criterion to mixture regression models. *Journal of the American Statistical Association* 102: 244–254.

Nakatani, Y., and A. McLysaght. 2019. Macrosynteny analysis shows the absence of ancient whole-genome duplication in lepidopteran insects. *Proceedings of the National Academy of Sciences* 116: 1816–1818.

Nikolov, L. A., P. Shushkov, B. Nevado, X. Gan, I. A. Al-Shehbaz, D. Filatov, C. D. Bailey, and M. Tsiantis. 2019. Resolving the backbone of the Brassicaceae phylogeny for investigating trait diversity. *New Phytologist* 222: 1638–1651.

One Thousand Plant Transcriptomes Initiative. 2019. One thousand plant transcriptomes and the phylogenomics of green plants. *Nature* 574: 679.

Patchell, M. J., E. H. Roalson, and J. C. Hall. 2014. Resolved phylogeny of Cleomaceae based on all three genomes. *Taxon* 63: 315–328.

Ratzka, A., H. Vogel, D. J. Kliebenstein, T. Mitchell-Olds, and J. Kroymann. 2002. Disarming the mustard oil bomb. *Proceedings of the National Academy of Sciences* 99: 11223–11228.

R Core Team. 2018. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. Website: https://www.R-project.org/.

Rodman, J. E., P. S. Soltis, D. E. Soltis, K. J. Sytsma, and K. G. Karol. 1998. Parallel evolution of glucosinolate biosynthesis inferred from congruent nuclear and plastid gene phylogenies. *American Journal of Botany* 85: 997–1006.

Schlüter, U., A. Bräutigam, U. Gowik, M. Melzer, P. A. Christin, S. Kurz, T. Mettler-Altmann, and A. P. Weber. 2016. Photosynthesis in C3–C4 intermediate *Moricandia* species. *Journal of Experimental Botany* 68: 191–206.

Schranz, M. E., and T. Mitchell-Olds. 2006. Independent ancient polyploidy events in the sister families Brassicaceae and Cleomaceae. *The Plant Cell* 18: 1152–1165.

Simão, F. A., R. M. Waterhouse, P. Ioannidis, E. V. Kriventseva, and E. M. Zdobnov. 2015. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* 31: 3210–3212.

Smith, S. A., and C. W. Dunn. 2008. Phyutility: a phyloinformatics tool for trees, alignments, and molecular data. *Bioinformatics* 24: 715–716.

Stamatakis, A. 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 30: 1312–1313.

Tamboli, A. S., P. B. Yadav, A. A. Gothe, S. R. Yadav, and S. P. Govindwar. 2018. Molecular phylogeny and genetic diversity of genus *Capparis* (Capparaceae) based on plastid DNA sequences and ISSR markers. *Plant Systematics and Evolution* 304: 205–217.

Tang, H., M. R. Woodhouse, F. Cheng, J. C. Schnable, B. S. Pedersen, G. Conant, X. Wang, M. Freeling, and J. C. Pires. 2012. Altered patterns of fractionation and exon deletions in *Brassica* rapa support a two-step model of paleohexaploidy. *Genetics* 190: 1563–1574.

Tian, Y., Y. Zeng, J. Zhang, C. Yang, L. Yan, X. Wang, C. Shi, et al. 2015. High quality reference genome of drumstick tree (*Moringa oleifera* Lam.), a potential perennial crop. *Science China Life Sciences* 58: 627–638.

Tiley, G. P., M. S. Barker, and J. G. Burleigh. 2018. Assessing the performance of K_s plots for detecting ancient whole-genome duplications. *Genome Biology and Evolution* 10: 2882–2898.

Tobe, H., and P. H. Raven. 2008. Embryology of *Koeberlinia* (Koeberliniaceae): Evidence for core-Brassicalean affinities. *American Journal of Botany* 95: 1475–1486.

Unruh, S. A., M. R. McKain, Y. I. Lee, T. Yukawa, M. K. McCormick, R. P. Shefferson, A. Smithson, et al. 2018. Phylotranscriptomic analysis and genome evolution of the Cypripedioideae (Orchidaceae). *American Journal of Botany* 105: 631–640.

van den Bergh, E., C. Külahoglu, A. Bräutigam, J. M. Hibberd, A. P. M. Weber, X. G. Zhu, and M. E. Schranz. 2014. Gene and genome duplications and the origin of C4 photosynthesis: birth of a trait in the Cleomaceae. *Current Plant Biology* 1: 2–9.

Vision, T. J., D. G. Brown, and S. D. Tanksley. 2000. The origins of genomic duplications in *Arabidopsis*. *Science* 290: 2114–2117.

Washburn, J. D., J. C. Schnable, G. C. Conant, T. P. Brutnell, Y. Shao, Y. Zhang, M. Ludwig, et al. 2017. Genome-guided phylo-transcriptomic methods and the nuclear phylogenetic tree of the Paniceae grasses. *Scientific Reports* 7: 13528.

Waterhouse, R. M., M. Seppey, F. A. Simão, M. Manni, P. Ioannidis, G. Klioutchnikov, E. V. Kriventseva, and E. M. Zdobnov. 2017. BUSCO applications from quality assessments to gene prediction and phylogenomics. *Molecular Biology and Evolution* 35: 543–548.

Yang, Y., M. J. Moore, S. F. Brockington, D. E. Soltis, G. K. S. Wong, E. J. Carpenter, Y. Zhang, et al. 2015. Dissecting molecular evolution in the highly diverse plant clade Caryophyllales using transcriptome sequencing. *Molecular Biology and Evolution* 32: 2001–2014.

Yang, Y., and S. A. Smith. 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.

Zhang, C., M. Rabiee, E. Sayyari, and S. Mirarab. 2018. ASTRAL-III: polynomial time species tree reconstruction from partially resolved gene trees. *BMC Bioinformatics* 19: 153.

Züst, T., M. Mirzaei, and G. Jander. 2018. *Erysimum cheiranthoides*, an ecological research system with potential as a genetic and genomic model for studying cardiac glycoside biosynthesis. *Phytochemistry Reviews* 17: 1239–1251.

Zwaenepoel, A., Z. Li, R. Lohaus, and Y. van de Peer. 2019. Finding evidence for whole-genome duplications: a reappraisal. *Molecular Plant* 12: 133–136.

Zwaenepoel, A., and Y. van de Peer. 2019. Ancient whole-genome duplications and the evolution of the gene duplication and loss rate. *Molecular Biology and Evolution* 36: 1384–1404.