

**INVITED SPECIAL ARTICLE** For the Special Issue: Using and Navigating the Plant Tree of Life

# Brassicales phylogeny inferred from 72 plastid genes: A reanalysis of the phylogenetic localization of two paleopolyploid events and origin of novel chemical defenses

Patrick P. Edger<sup>1,2,12</sup>, Jocelyn C. Hall<sup>3</sup>, Alex Harkess<sup>4,5</sup>, Michelle Tang<sup>6</sup>, Jill Coombs<sup>6</sup>, Setareh Mohammadin<sup>7</sup>, M. Eric Schranz<sup>7</sup>, Zhiyong Xiong<sup>8</sup>, James Leebens-Mack<sup>4</sup>, Blake C. Meyers<sup>5</sup>, Kenneth J. Sytsma<sup>9</sup>, Marcus A. Koch<sup>10</sup>, Ihsan A. Al-Shehbaz<sup>11</sup>, and J. Chris Pires<sup>6</sup>

Manuscript received 21 August 2017; revision accepted 6 November 2017.

<sup>1</sup> Department of Horticulture, Michigan State University, East Lansing, Michigan 48864, USA

<sup>2</sup> Ecology, Evolutionary Biology and Behavior, Michigan State University, East Lansing, MI 48864, USA

<sup>3</sup> Department of Biological Sciences, University of Alberta, Edmonton, Alberta T6G 2E9, Canada

<sup>4</sup> Department of Plant Biology, University of Georgia, Athens, GA 30602, USA

<sup>5</sup> Donald Danforth Plant Science Center, 975 North Warson Road, St. Louis, MO 63132, USA

<sup>6</sup> Division of Biological Sciences, University of Missouri, Columbia, MO 65211, USA

 $^7$ Biosystematics, Plant Science Group, Wageningen University and Research, Wageningen, Netherlands

<sup>8</sup> Potato Engineering & Technology Research Center, Inner Mongolia University, Hohhot, China

<sup>9</sup> Department of Botany, University of Wisconsin, Madison, WI 53706, USA

<sup>10</sup> Department of Biodiversity and Plant Systematics, Centre for Organismal Studies, Heidelberg University, Heidelberg, Germany

<sup>11</sup> Missouri Botanical Garden, St. Louis, MO 63166-0299, USA

 $^{\rm 12}$  Author for correspondence (e-mail: edgerpat@msu.edu)

Citation: 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(3): 463–469.

doi:10.1002/ajb2.1040

**PREMISE OF THE STUDY**: Previous phylogenetic studies employing molecular markers have yielded various insights into the evolutionary history across Brassicales, but many relationships between families remain poorly supported or unresolved. A recent phylotranscriptomic approach utilizing 1155 nuclear markers obtained robust estimates for relationships among 14 of 17 families. Here we report a complete family-level phylogeny estimated using the plastid genome.

**METHODS:** We conducted phylogenetic analyses on a concatenated data set comprising 44,926 bp from 72 plastid genes for species distributed across all 17 families. Our analysis includes three additional families, Tovariaceae, Salvadoraceae, and Setchellanthaceae, that were omitted in the previous phylotranscriptomic study.

**KEY RESULTS**: Our phylogenetic analyses obtained fully resolved and strongly supported estimates for all nodes across Brassicales. Importantly, these findings are congruent with the topology reported in the phylotranscriptomic study. This consistency suggests that future studies could utilize plastid genomes as markers for resolving relationships within some notoriously difficult clades across Brassicales. We used this new phylogenetic framework to verify the placement of the At- $\alpha$  event near the origin of Brassicaceae, with median date estimates of 31.8 to 42.8 million years ago and restrict the At- $\beta$  event to one of two nodes with median date estimates between 85 to 92.2 million years ago. These events ultimately gave rise to novel chemical defenses and are associated with subsequent shifts in net diversification rates.

**CONCLUSIONS**: We anticipate that these findings will aid future comparative evolutionary studies across Brassicales, including selecting candidates for whole-genome sequencing projects.

**KEY WORDS** phylogenomics; glucosinolates; secondary metabolites; evolutionary novelties.

American Journal of Botany 105(3): 463–469, 2018; http://www.wileyonlinelibrary.com/journal/AJB © 2018 The Authors. American Journal of Botany is published by Wiley Periodicals, Inc. on behalf of the Botanical Society of America. This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.
463

The order Brassicales, which contains ~4700 species or ~2.2% of eudicot diversity (Magallón et al., 1999; Kiefer et al., 2014; Cardinal-McTeague et al., 2016), is a monophyletic group consisting of 17 families including several model families for evolutionary biology (Rodman et al., 1996, 1998; Hall et al., 2004; Soltis et al., 2011; Chase et al., 2016). The primary families for such studies include Brassicaceae, Cleomaceae, and Caricaceae with dozens of sequenced reference genomes (Koenig and Weigel, 2015) and numerous genome projects that are currently underway. Brassicales are known for their rich diversity of morphological, physiological, developmental, and chemical traits, and a range of projects are investigating the genetic basis of this diversity, including transitioning from C<sub>3</sub> to C<sub>4</sub> photosynthesis [e.g., Gynandropsis gynandra (L.) Briq.; Külahoglu et al., 2014; van den Bergh et al., 2014], abiotic stress tolerances [e.g., Eutrema salsugineum (Pall.) Al-Shehbaz & S.I.Warwick; Yang et al., 2013], and evolution of sex chromosomes (e.g., Carica papaya L.; Ming et al., 2008).

Despite this incredible diversity in form, members of the order are united by chemical defenses called glucosinolates (i.e., mustard oils) (Rodman et al., 1998; Schranz et al., 2011). Glucosinolates are highly variable across the order (Fig. 1), including over 120 unique compounds that are grouped into distinct classes based on their amino acid precursor (Fahey et al., 2001; Kliebenstein et al., 2001; Mithen et al., 2010). Some glucosinolates have been described outside the Brassicales, including Euphorbiaceae (Rodman et al., 1998). However, methionine-derived and tryptophan-derived (indolic) glucosinolates are novel to only certain families within Brassicales (Fahey et al., 2001; Schranz et al., 2011; van den Bergh et al., 2016). These novel classes of glucosinolates evolved, at least in part, due to an ancient arms race with insect herbivores, including butterflies of the subfamily Pierinae (Pieridae, Lepidoptera) (Ehrlich and Raven, 1964; Edger et al., 2015). Gene and genome duplication events followed by neofunctionalization played a central role driving the escalation of novel glucosinolate diversity across Brassicales (Kliebenstein, 2008; Hofberger et al., 2013; Edger et al., 2015). For example, the regulatory and core biosynthetic pathways for these novel glucosinolate classes arose from retained duplicates from both single-gene (tandem) and an ancient whole-genome duplication  $(At-\beta)$  shared by most families of Brassicales (Edger et al., 2015). The duplicated pathways from the At- $\beta$  event evolved to synthesize a novel class of chemical defense compounds from a new amino acid substrate. Similarly, new biosynthetic steps involved in novel structural elaborations unique to Brassicaceae arose due to retained duplicate genes from the At-a whole-genome and tandem duplications (Hofberger et al., 2013; Edger et al., 2015). A more recent study investigated the origin of certain specialized metabolites across Solanaceae, revealing that gene duplications and functional diversification also played a central role in generating novelty (Moghe et al., 2017). These studies, among others, support Susumu Ohno's hypothesis (Ohno, 1970) that gene duplications have been instrumental in driving the origin of new features and increases in organismal complexity and species diversity over evolutionary time (Freeling and Thomas, 2006; Edger and Pires, 2009).

A robust and well-resolved phylogeny is needed to interpret the wealth of data available from Brassicales species in an evolutionary context, including glucosinolate diversity. As with many diverse plant lineages, resolving the Brassicales phylogeny has been a combination of robust relationships being evident in earliest analyses with other groups proving more challenging to place robustly. For example, early studies identified a species-rich core group of families, referred to as the core Brassicales (Rodman et al., 1994, 1996), based on synapomorphy of an extended *rbcL* sequence found in Brassicaceae, Capparaceae, Cleomaceae, Gyrostemonaceae, Reseduaceae, and Tovariaceae. Additional studies expanded this well-supported clade to eight families that also have an aspartic acid rather than the usual stop codon at the end of the gene: Brassicaceae, Cleomaceae, Capparaceae, Emblingiaceae, Gyrostemonaceae, Pentadiplandraceae, Resedaceae, and Tovariaceae (Hall et al., 2004; Su et al., 2012; Cardinal-McTeague et al., 2016). Within the core Brassicales, many relationships are strongly supported based on multiple analyses. A Brassicaceae-Cleomaceae clade is strongly supported as sister to Capparaceae (Hall et al., 2002, 2004; Hall, 2008; Edger et al., 2015). Gyrostemonaceae and Resedaceae are a strongly supported clade (Hall et al., 2004; Ronse De Craene and Haston, 2006; Edger et al., 2015; Cardinal-McTeague et al., 2016) that are sister to Pentadiplandraceae (Edger et al., 2015; Cardinal-McTeague et al., 2016). The Brassicaceae-Cleomaceae-Capparaceae and Gyrostemonaceae-Resedaceae-Pentadiplandraceae clades are strongly supported as sister groups, with this clade being sister to Emblingaceae (Edger et al., 2015). In contrast, the phylogenetic placement of Tovariaceae within the core Brassicales remains unresolved despite broad taxonomic sampling and five gene regions (e.g., in polytomy in Cardinal-McTeague et al., 2016). Notably, Edger et al. (2015) were able to robustly estimated relationships within core Brassicales using 1155 nuclear markers, but they did not include Tovariaceae. Thus, understanding relationships of Tovariaceae may be dependent on a substantial increase in data.

The theme of variable resolution of phylogenetic relationships extends beyond the core Brassicales, which is compounded by the omission of key taxa in some analyses. There are a number of family-level phylogenetic relationships that have been established outside the core Brassicales. Studies support the following sisterfamily pairs: Koeberliniaceae-Bataceae, Tropaeolaceae-Akaniaceae, and Caricaceae-Moringaceae (Olson, 2002a, b; Hall et al., 2004; Cardinal-McTeague et al., 2016). However, an alternate relationship for Bataceae has been estimated, a Bataceae-Salvadoraceae clade (Ronse De Craene and Haston, 2006), which is congruent with findings of separate Bataceae and Koeberliniaceae lineages (Edger et al., 2015). Salvadoraceae was also not included in our previous phylotranscriptomic study (Edger et al., 2015). A clade that includes the core Brassicales, Koeberlinaceae, Bataceae, Salvadoraceae, Limnanthaceae, and Setchellanthaceae is strongly supported (Hall et al., 2004), with Setchellanthaceae as the earliest-diverging lineage of this group (Rodman et al., 1996, 1998; Karol et al., 1999; Chandler and Bayer, 2000; Ronse De Craene and Haston, 2006). Notably, our previous phylotranscriptomic study (Edger et al., 2015) did not sample either Salvadoraceae or Setchellanthaceae, of which the latter family was previously hypothesized as an important family to phylogenetically localize the At- $\beta$  event (Schranz et al., 2011). The Caricaceae-Moringaceae and Akaniaceae-Tropaeolaceae clades are strongly supported clades and are the earliest-diverging families of the order Brassicales (Ronse De Craene and Haston, 2006; Edger et al., 2015; Cardinal-McTeague et al., 2016). Last, the nodes along the backbone of the phylogeny are strongly supported in recent studies (Edger et al., 2015; Cardinal-McTeague et al., 2016), but neither study included every Brassicales family.

Phylogenomic analyses utilizing coding sequences from the plastid genome have obtained strongly supported phylogenetic estimates for some notoriously difficult plant clades at various taxonomic levels, including angiosperms (Jansen et al., 2007),



**FIGURE 1.** Phylogenetic relationships across Brassicales: ancient whole-genome duplications, glucosinolate diversity, and shifts in net diversification rates. The Brassicales topology of the 17 families, estimated using 72 genes from the plastid genome, was utilized to phylogenetically localize the At- $\alpha$  (red star) and At- $\beta$  (purple star) whole-genome duplications using data from Edger et al. (2015). Bootstrap values for nodes less than 100% are shown below adjacent branches. *Arabidopsis, Brassica*, and *Aethionema* are Brassicaceae (orange box). The core Brassicales are highlighted within the green box. Species richness for every family is provided in the column to the right of the phylogeny, with Brassicaceae split into the Aethionema lineage (48 total) (Mohammadin et al., 2017) and all other species (3925 total) (Al-Shehbaz, 2012; Koch et al., 2012). The two possible phylogenetic positions for the At- $\beta$  whole-genome duplication are depicted with purple stars. The origins of indolic glucosinolates, Met-derived glucosinolates, and novel structural elaborations are shown with blue arrows; see Edger et al. (2015) for additional details and timing of these events. Date estimates for nodes are shown above adjacent branches. Three significant shifts in net diversification rates (black circles) were previously reported, including at the origin of Met-glucosinolates, base of the Gyrostemonaceae-Resedaceae clade, and origin of novel structural elaborations (Edger et al., 2015; Cardinal-McTeague et al., 2016). The whole-genome duplication unique to Cleomaceae is shown with a green star (Barker et al., 2009).

basal angiosperms (Leebens-Mack et al., 2005), eudicots (Moore et al., 2010), grasses (Cotton et al., 2015; Saarela et al., 2015), and Brassicaceae (Hohmann et al., 2015; Mandáková et al., 2017). Here, we report a phylogenetic estimate of the relationships among all 17 Brassicales families based on 72 plastid coding regions assembled from Illumina next-generation sequencing (NGS) data. We had three major goals for this study: (1) obtain a strongly supported phylogenetic estimate for all family-level relationships across Brassicales, (2) phylogenetically localize the At- $\alpha$  and At- $\beta$  events, and (3) re-evaluate the phylogenetic distribution of novel glucosinolates that are associated with significant shifts in net diversification rates. Last, we hope these results will help guide the selection of species for future genome sequencing projects for evolutionary studies across Brassicales.

# MATERIALS AND METHODS

### Data sets

Total DNA was extracted from either fresh or silica-dried leaf material using the Qiagen DNeasy Plant Mini Kit (Qiagen, Germantown, MD, USA) (Invitrogen, Carlsbad, CA, USA). A complete list of taxa, voucher information, and access to sequence data is given in Appendix S1 (see the Supplemental Data with this article). The barcoded DNA Illumina libraries were constructed using the NEB prep E600L kit (New England Biolabs, Ipswich, MA, USA) and sequenced single end with 100-bp reads on both the Illumina (San Diego, CA, USA) GAII and HiSeq-2000 instrument at the University of Missouri DNA core. For Gyrostemon racemiger H.Walter, total RNA was extracted, instead of DNA, from fresh young leaf tissue using the Invitrogen PureLink RNA Mini Kit, converted into an Illumina library using the TruSeq RNA kit (Illumina), and pairedend 100-bp reads were sequenced on the HiSeq-2000 instrument at the University of Missouri DNA core. Sequence data for Arabidopsis thaliana (L.) Heynh. (NC\_001284.2 and NC\_000932.1), Carica papaya (NC\_012116.1 and NC\_010323.1), and Brassica oleracea L. (SRP010908) were obtained from NCBI-GenBank and NCBI-SRA. Raw illumina data generated in this study have been deposited in either the Dryad Digital Repository (https://doi.org/10.5061/dryad. kc5b8) or the NCBI database under BioProject PRJNA283303 (https://www.ncbi.nlm.nih.gov/bioproject/?term=PRJNA283303).

#### Assembly and phylogenetic analyses

The NextGENe V2.17 (SoftGenetics, State College, PA, USA) software package was used to remove low-quality Illumina data. The quality-filtered data was assembled *de novo* using both the Maximum Overlap assembler in NextGENe and CAP3 (Huang and Madan, 1999), and assembly errors were manually screened utilizing NextGENe-Viewer. We utilized BLASTn (Altschul et al., 1990) combined with published reference sequences from *Arabidopsis thaliana* (NC\_001284.2 and NC\_000932.1) to parse plastid contigs larger than 300 bp. The *Gyrostemon* transcriptome library was assembled using Trinity (Grabherr et al., 2011). Target coding sequences (CDS) were identified using *Arabidopsis thaliana* organellar CDS (GenBank Y08501.2 and AP000423.1). We only utilized loci shared by all species. All alignments were generated using MUSCLE (Edgar, 2004), and columns were removed (i.e., across all taxa) if a base-pair position was missing in one species. We

obtained a complete 44,926-bp data matrix derived from 72 different coding regions of the plastid genome. All phylogenetic analyses were conducted using the CIPRES V3.1 portal (Miller et al., 2010). The program RAxML version 7.3.1 (Stamatakis, 2006) was utilized to search for the optimal maximum likelihood (ML) tree using the GTR+GAMMA substitution model, combined with 1000 bootstrap replicates. Consensus trees were summarized with the program Consense (Felsenstein, 1989). The final alignment of assembled plastid sequences used in phylogenetic analyses available from the Dryad Digital Repository: https://doi.org/10.5061/dryad.kc5b8. Glucosinolate profiles and Ks age distribution data for species was extracted from the data of Edger et al. (2015).

# RESULTS

Previous phylogenetic analyses have provided important insights into the Brassicales phylogeny, including delineating all ~4700 species into 17 monophyletic families. However, these studies have been unable to resolve many familial relationships, specifically along the entire backbone of the tree. Here, we have employed a 72-marker approach to resolve the major relationships among all families. We obtained a strongly supported topology with 100% bootstrap support for the majority of nodes and two nodes with over 87% bootstrap support. Our phylogenetic estimates are a significant improvement in the overall phylogeny (e.g., core Brassicales, backbone, earlier-diverging lineages), permitting us to more accurately pinpoint two ancient whole-genome duplications and the origin of two novel classes of glucosinolates.

### **Core Brassicales**

The core Brassicales is the most speciose monophyletic clade in the order, comprising of the Brassicaceae (3973 species), Cleomaceae (300 species), Capparaceae (480 species), Emblingiaceae (1 species), Gyrostemonaceae (18 species), Pentadiplandraceae (1 species), Resedaceae (75 species), and Tovariaceae (2 species) (Hall et al., 2004). For the most up-to-date species numbers, see BrassiBase (Koch et al., 2012; Kiefer et al., 2014). Our analyses using markers from the plastid genome yielded a fully resolved phylogenetic estimate for the entire core Brassicales (Fig. 1) and is identical to the tree estimated from 1155 nuclear genes (Edger et al., 2015) with the sole exception of the inclusion of Tovariaceae. The previously reported Brassicaceae-Cleomaceae clade is still strongly supported (bootstrap 100%), which is sister to the Capparaceae (bootstrap 100%). Our phylogenetic estimates suggest that Tovariaceae is sister to the Brassicaceae-Cleomaceae-Capparaceae clade (bootstrap 89.5%). In addition, our analyses also recovered a Gyrostemonaceae-Resedaceae clade (bootstrap 100%), with Pentadiplandraceae as the immediate sister lineage (bootstrap 100%). Last, Emblingia calceoliflora F.Meull., the sole species in the family Emblingiaceae, is strongly supported as being sister to the remainder of the core Brassicales (bootstrap 100%). Thus, we present the first robust and well-resolved phylogeny that includes all families in the core Brassicales.

#### **Earlier-diverging lineages**

Our estimates for the relationships among the nine families distributed across the earliest-diverging lineages are also congruent with our previous phylotranscriptomic study (Edger et al., 2015) with the exception of the inclusion of two new families (Salvadoraceae and Setchellanthaceae). These nine families, with 166 species (~3.5% Brassicales) that are distributed across six major lineages, includes Koeberliniaceae (1 species), Bataceae (2 species), Salvadoraceae (11 species), Setchellanthaceae (1 species), Limnanthaceae (8 species), Caricaceae (35 species), Moringaceae (12 species), Tropaeolaceae (95 species), and Akaniaceae (2 species). The monotypic Koeberliniaceae, which consists of Koeberlinia spinosa Zucc., is the lineage sister to the core Brassicales (bootstrap 100%) with the Bataceae-Salvadoraceae clade sister to that clade (bootstrap 100%). Limnanthaceae (bootstrap 87.4%) is sister to the aforementioned clades with Setchellanthaceae (bootstrap 100%) sister to this clade in turn. Caricaceae-Moringaceae (bootstrap 100%) are sister to all Brassicales except for the Tropaeolaceae-Akaniaceae clade (100%), which represents the earliest-diverging lineage in the order (100%). These relationships are consistent with previously published phylogenies, including the Bataceae-Salvadoraceae clade (Ronse De Craene and Haston, 2006) and placement of Setchellanthaceae (Cardinal-McTeague et al., 2016).

# DISCUSSION

Our phylogenetic analysis of coding regions from the plastid genome obtained fully resolved and strongly supported estimates for all family-level relationships across the Brassicales. However, this analysis reflects only the maternal history of the order due to the inheritance of organellar genomes. Importantly, these estimates are congruent with the topology reported in our previous phylotranscriptomic study (Edger et al., 2015) with the exception of the addition of the three new families. This consistency suggests that future studies could utilize plastid genomes as markers for resolving relationships within some notoriously difficult clades across Brassicales, as was previously demonstrated within Brassicaceae (Hohmann et al., 2015; Mandáková et al., 2017). The sequencing and assembly of plastid genomes has remained an easier and more cost-effective approach compared to phylotranscriptomics (Wysocki et al., 2014; Honaas et al., 2016). We encourage future studies to assess the robustness of our phylogenetic estimates by sampling more species across various families and to resolve within family relationships (Al-Shehbaz, 2012; Huang et al., 2016).

# Phylogenetic placement of two ancient whole-genome duplications

The precise phylogenetic placement of the At- $\alpha$  and At- $\beta$  wholegenome duplications in the Brassicales has been unknown due to two major reasons: (1) lack of a robust, well-resolved phylogeny and (2) lack of additional genomic resources from all families. We previously determined the history of these two events in 14 of the 17 families by analyzing transcriptome data, which included constructing Ks age distributions of duplicated genes (Edger et al., 2015). Based on the phylogenetic placement of the three new families, the At- $\alpha$  event is confirmed to have occurred at the origin of the Brassicaceae (Fig. 1). The At- $\alpha$  event occurred before the divergence of *Arabidopsis* and *Aethionema* (Haudry et al., 2013) but following the divergence from the sister family Cleomaceae (Edger et al., 2015) with an independent polyploid event (Barker et al., 2009). We were also able to phylogenetically restrict the At- $\beta$  event to one of two nodes. The At- $\beta$  event occurred either after the divergence of the Caricaceae-Moringaceae clade or after the divergence of Setchellanthaceae (Fig. 1). We previously determined that the At- $\beta$ event was not shared by Akaniaceae, Caricaceae, Moringaceae, and Tropaeolaceae, but present in the evolutionary history of all other surveyed families (Edger et al., 2015). Based on the phylogenetic placement of Salvadoraceae and Tovariaceae in this study, we are able to determine that these families must also share the At- $\beta$  event. However, we are unable to determine whether Setchellanthaceae shares the At- $\beta$  event. Thus, we are able to restrict the phylogenetic placement of this ancient whole-genome duplication to the entire order with the exception of a single species (*Setchellanthus caeruleus*; Setchellanthaceae). The exact phylogenetic placement of the At- $\beta$  could be determined from future analyses of either transcriptome or a reference genome of *Setchellanthus caeruleus*.

### Phylogenetic distribution of novel glucosinolate compounds

Methionine-derived and tryptophan-derived (indolic) glucosinolates, which are unique to the order Brassicales, are a set of novel chemical defenses (Rodman et al., 1998; Kliebenstein et al., 2002; Windsor et al., 2005; Bekaert et al., 2012). It is these compounds that give mustard and horseradish that distinctive taste and unique aroma. Our previous studies showed that these novel classes of glucosinolates are synthesized from new amino acid substrates via functionally diverged, duplicated pathways that arose due to the At-α and At-β whole-genome duplications (Hofberger et al., 2013; Edger et al., 2015). For example, in Arabidopsis thaliana, two paralogous groups of Myb transcription factors regulate different glucosinolate pathways, with Myb28/29/76 regulating the biosynthesis of Met-derived glucosinolates and Myb34/51/122 regulating the biosynthesis of indolic glucosinolates. In Carica papaya (Caricaceae), which does not share the At- $\beta$  event in its evolutionary history, the ancestral glucosinolate pathway is regulated by a single Myb transcription factor gene (Schranz et al., 2011; Bekaert et al., 2012).

A reanalysis of the phylogenetic distribution of these novel glucosinolate classes, using the new family-level phylogeny, placed the origin of indolic glucosinolates, methionine-derived glucosinolates, and novel structural variants at nodes consistent with our previous analysis (Edger et al., 2015) (Fig. 1). In our previous work, we also constructed and compared glucosinolate pathways for the majority of families, revealing that Limnanthaceae did not have the genes required for indolic glucosinolate biosynthesis. This analysis further supports the placement for indolic glucosinolates shown in Fig. 1. A reference genome for any Limnanthaceae (e.g., Limnanthes douglasii R.Br.), combined with additional targeted glucosinolate profiling, would be tremendously helpful to verify these findings. Furthermore, we are confident in the placement for the origin of methionine-derived glucosinolates since these are unique to the Brassicaceae-Cleomaceae-Capparaceae clade. Similarly, novelstructural elaborations unique to the core Brassicaceae are synthesized from neofunctionalized duplicated genes retained from the At-α event (Kliebenstein, 2008; Hofberger et al., 2013).

The origin of these novel glucosinolates are also associated with significant shifts in net diversification rates (Fig. 1) (Edger et al., 2015; Cardinal-McTeague et al., 2016), supporting predicted evolutionary arms-race dynamics with certain butterflies (Ehrlich and Raven, 1964). However, a shift in net diversification rates is not observed for the origin of indolic glucosinolates but this is likely due to the age of this event (~85–92.2 million years ago; Edger et al.,

2015) pre-dating the Cretaceous-Tertiary (KT) mass extinction event (Fawcett et al., 2009). Regarding the diversification rate shift detected at the base of the Reseaceae-Gyrostemnon clade (Cardinal-McTeague et al., 2016), the causes of this radiation remain unknown. Last, our findings are consistent with the predictions of the wholegenome duplication (WGD) radiation-lag time model (Schranz et al., 2012) in which the whole-genome duplication and the evolution of a novel adaptive trait (e.g., glucosinolates) derived from neofunctionalized duplicate genes should be separated by long time periods (e.g., millions of years). Neofunctionalization, developing novel gene functions from background mutations, requires long time periods especially without perturbing the established network stoichiometry (Birchler and Veitia, 2007; Edger and Pires, 2009). In conclusion, we anticipate that this new Brassicales phylogeny will aid future evolutionary studies across the order and guide the selection of future candidates for genome projects.

#### ACKNOWLEDGEMENTS

We thank J. Birchler, G. Conant, K. Newton, and J. Taylor for valuable comments on the manuscript; R. Schnabel and K. Wells for computational assistance and the National Science Foundation (DEB 1146603 and NPGI 1202793) and USDA-HATCH (1009804) for support. We also thank the editors and reviewers for their helpful comments during the review of this manuscript.

## SUPPORTING INFORMATION

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

# LITERATURE CITED

- Al-Shehbaz, I. A. 2012. A generic and tribal synopsis of the Brassicaceae (Cruciferae). *Taxon* 61: 931–954.
- Altschul, S. F., W. Gish, W. Miller, E. W. Myers, and D. J. Lipman. 1990. Basic local alignment search tool. *Journal of Molecular Biology* 215: 403–410.
- 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.
- Bekaert, M., P. P. Edger, C. M. Hudson, J. C. Pires, and G. C. Conant. 2012. Metabolic and evolutionary costs of herbivory defense: systems biology of glucosinolate synthesis. *New Phytologist* 196: 596–605.
- Birchler, J. A., and R. A. Veitia. 2007. The gene balance hypothesis: from classical genetics to modern genomics. *Plant Cell* 19: 395–402.
- Cardinal-McTeague, W. M., 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.
- Chandler, G. T., and R. J. Bayer. 2000. Phylogenetic placement of the enigmatic Western Australian genus *Emblingia* based on *rbcL* sequences. *Plant Species Biology* 15: 67–72.
- Chase, M. W., M. J. M. Christenhusz, M. F. Fay, J. W. Byng, W. S. Judd, D. E. Soltis, D. J. Mabberley, et al. 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.
- Cotton, J. L., W. P. Wysocki, L. G. Clark, S. A. Kelchner, J. C. Pires, P. P. Edger, D. Mayfield-Jones, and M. R. Duvall. 2015. Resolving deep relationships of PACMAD grasses: a phylogenomic approach. *BMC plant biology* 15: 178.

- Edgar, R. C. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research* 32: 1792–1797.
- 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, USA* 112: 8362–8366.
- Edger, P. P., and J. C. Pires. 2009. Gene and genome duplications: the impact of dosage-sensitivity on the fate of nuclear genes. *Chromosome Research* 17: 699–717.
- Ehrlich, P. R., and P. H. Raven. 1964. Butterflies and plants: a study in coevolution. *Evolution* 18: 586–608.
- Fahey, J. W., A. T. Zalcmann, and P. Talalay. 2001. The chemical diversity and distribution of glucosinolates and isothiocyanates among plants. *Phytochemistry* 56: 5–51.
- Fawcett, J. A., S. Maere, and Y. Van de Peer. 2009. Plants with double genomes might have had a better chance to survive the Cretaceous-Tertiary extinction event. *Proceedings of the National Academy of Sciences, USA* 106: 5737–5742.
- Felsenstein, J. 1989. PHYLIP (version 3.6) Phylogeny Inference Package. Cladistics 5: 164–166.
- Freeling, M., and B. C. Thomas. 2006. Gene-balanced duplications, like tetraploidy, provide predictable drive to increase morphological complexity. *Genome Research* 16: 805–814.
- 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: 644–652.
- 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: 891–898.
- Hofberger, J. A., E. Lyons, P. P. Edger, J. C. Pires, and M. E. Schranz. 2013. Whole genome and tandem duplicate retention facilitated glucosinolate pathway diversification in the mustard family. *Genome Biology and Evolution* 5: 2155–2173.
- Hohmann, N., E. M. Wolf, M. A. Lysak, and M. A. Koch. 2015. A time-calibrated road map of Brassicaceae species radiation and evolutionary history. *Plant Cell* 27: 2770–2784.
- Honaas, L. A., E. K. Wafula, N. J. Wickett, J. P. Der, Y. Zhang, P. P. Edger, N. S. Altman, et al. 2016. Selecting superior de novo transcriptome assemblies: lessons learned by leveraging the best plant genome. *Plos One* 11: E0146062.
- 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.
- Huang, X., and A. Madan. 1999. CAP3: A DNA sequence assembly program. *Genome Research* 9: 868–877.
- Jansen, R. K., Z. Cai, L. A. Raubeson, H. Daniell, C. W. Depamphilis, J. Leebens-Mack, K. F. Müller, et al. 2007. Analysis of 81 genes from 64 plastid genomes resolves relationships in angiosperms and identifies genome-scale evolutionary patterns. *Proceedings of the National Academy of Sciences, USA* 104: 19369–19374.
- Karol, K. G., J. E. Rodman, E. Conti, and K. J. Sytsma. 1999. Nucleotide sequence of *rbcL* and phylogenetic relationships of *Setchellanthus caeruleus* (Setchellanthaceae). *Taxon* 48: 303–315.
- Kiefer, M., R. Schmickl, D. A. German, T. Mandáková, M. A. Lysak, I. A. Al-Shehbaz, A. Franzke, et al. 2014. BrassiBase: introduction to a novel knowledge database on Brassicaceae evolution. *Plant & Cell Physiology* 55: e3.

- Kliebenstein, D., D. Pedersen, B. Barker, and T. Mitchell-Olds. 2002. Comparative analysis of quantitative trait loci controlling glucosinolates, myrosinase and insect resistance in *Arabidopsis thaliana*. *Genetics* 161: 325–332.
- Kliebenstein, D. J. 2008. A role for gene duplication and natural variation of gene expression in the evolution of metabolism. *PloS One* 3: e1838.
- Kliebenstein, D. J., J. Kroymann, P. Brown, A. Figuth, D. Pedersen, J. Gershenzon, and T. Mitchell-Olds. 2001. Genetic control of natural variation in Arabidopsis glucosinolate accumulation. *Plant Physiology* 126: 811–825.
- Koch, M. A., M. Kiefer, D. A. German, I. A. Al-Shehbaz, A. Franzke, K. Mummenhoff, and R. Schmickl. 2012. BrassiBase: Tools and biological resources to study characters and traits in the Brassicaceae—version 1.1. *Taxon* 61: 1001–1009.
- Koenig, D., and D. Weigel. 2015. Beyond the thale: comparative genomics and genetics of Arabidopsis relatives. *Nature Reviews. Genetics* 16: 285–298.
- Külahoglu, C., A. K. Denton, M. Sommer, J. Maß, S. Schliesky, T. J. Wrobel, B. Berckmans, et al. 2014. Comparative transcriptome atlases reveal altered gene expression modules between two Cleomaceae C<sub>3</sub> and C<sub>4</sub> plant species. *Plant Cell* 26: 3243–3260.
- Leebens-Mack, J., L. A. Raubeson, L. Cui, J. V. Kuehl, M. H. Fourcade, T. W. Chumley, J. L. Boore, et al. 2005. Identifying the basal angiosperm node in chloroplast genome phylogenies: sampling one's way out of the Felsenstein zone. *Molecular Biology and Evolution* 22: 1948–1963.
- Magallón, 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., P. Hloušková, D. A. German, and M. A. Lysak. 2017. Monophyletic origin and evolution of the largest crucifer genomes. *Plant Physiology* 174: 2062–2071.
- Miller, M. A., W. Pfeiffer, and T. Schwartz. 2010. Creating the CIPRES Science Gateway for inference of large phylogenetic trees. *Gateway Computing Environments Workshop (GCE)* 2010: 1–8.
- 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: 991–996.
- Mithen, R., R. Bennett, and J. Marquez. 2010. Glucosinolate biochemical diversity and innovation in the Brassicales. *Phytochemistry* 71: 2074–2086.
- Moghe, G. D., B. J. Leong, S. M. Hurney, A. Daniel Jones, and R. L. Last. 2017. Evolutionary routes to biochemical innovation revealed by integrative analysis of a plant-defense related specialized metabolic pathway. *eLife* 6: e28468.
- Mohammadin, S., K. Peterse, devan Kerke S. J., L.W. Chatrou, A. A. Dönmez, K. Mummenhoff, J. C. Pires, et al. 2017. Anatolian origins and diversification of Aethionema, the sister lineage of the core Brassicaceae. *American Journal of Botany* 104: 1042–1054.
- Moore, M. J., P. S. Soltis, C. D. Bell, J. G. Burleigh, and D. E. Soltis. 2010. Phylogenetic analysis of 83 plastid genes further resolves the early diversification of eudicots. *Proceedings of the National Academy of Sciences, USA* 107: 4623–4628.
- Ohno, S. 1970. Evolution by gene duplication. Spinger-Verlag, New York, NY, USA. Olson, M. E. 2002a. Combining data from DNA sequences and morphology for
- a phylogeny of Moringaceae (Brassicales). *Systematic Botany* 27: 55–73. Olson, M. E. 2002b. Intergeneric relationships within the Caricaceae-Moringaceae clade (Brassicales) and potential morphological synapomorphies of the clade and its families. *International Journal of Plant Sciences* 163: 51–65.

- Rodman, J. E., K. G. Karol, R. A. Price, E. Conti, and K. J. Systma. 1994. Nucleotide sequences of *rbcL* confirm the capparalean affinity of the Australian endemis Gyrostemonaceae. *Australian Systematic Botany* 7: 57–69.
- Rodman, J. E., K. G. Karol, R. A. Price, and K. J. Sytsma. 1996. Molecules, morphology, and Dahlgren's expanded order Capparales. *Systematic Botany* 21: 289–307.
- Rodman, J., P. Soltis, D. Soltis, K. Sytsma, and K. Karol. 1998. Parallel evolution of glucosinolate biosynthesis inferred from congruent nuclear and plastid gene phylogenies. *American Journal of Botany* 85: 997–1006.
- Ronse De Craene, L. P., and E. Haston. 2006. The systematic relationships of glucosinolate-producing plants and related families: a cladistic investigation based on morphological and molecular characters. *Botanical Journal of the Linnean Society* 151: 453–494.
- Saarela, J. M., W. P. Wysocki, C. F. Barrett, R. J. Soreng, J. I. Davis, L. G. Clark, and S. A. Kelchner, et al. 2015. Plastid phylogenomics of the cool-season grass subfamily: clarification of relationships among early-diverging tribes. *AoB PLANTS* 7: plv046.
- Schranz, M. E., P. P. Edger, J. C. Pires, N. M. van Dam, and C. W. Wheat. 2011. Comparative genomics in the Brassicales: ancient genome duplications, glucosinolate diversification and Pierinae herbivore radiation. In: D. Edwards, J. Batley, I. Parkin, and C. Kole (eds), Genetics, Genomics and Breeding of Oilseed Brassicas 206, 206–217. CRC Press, Boca Raton, FL, USA.
- Schranz, M. E., S. Mohammadin, and P. P. Edger. 2012. Ancient whole genome duplications, novelty and diversification: the WGD radiation lag-time model. *Current Opinion in Plant Biology* 15: 147–153.
- Soltis, D. E., S. A. Smith, N. Cellinese, K. J. Wurdack, D. C. Tank, S. F. Brockington, N. F. Refulio-Rodriguez, et al. 2011. Angiosperm phylogeny: 17 genes, 640 taxa. American Journal of Botany 98: 704–730.
- Stamatakis, A. 2006. RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics* 22: 2688–2690.
- Su, J.-X., W. Wang, L.-B. Zhang, and Z.-D. Chen. 2012. Phylogenetic placement of two enigmatic genera, *Borthwickia* and *Stixis*, based on molecular and pollen data, and the description of a new family of Brassicales, Borthwickiaceae. *Taxon* 61: 601–611.
- van den Bergh, E., J. A. Hofberger, and M. E. Schranz. 2016. Flower power and the mustard bomb: Comparative analysis of gene and genome duplications in glucosinolate biosynthetic pathway evolution in Cleomaceae and Brassicaceae. *American Journal of Botany* 103: 1212–1222.
- 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 C 4 photosynthesis: birth of a trait in the Cleomaceae. *Current Plant Biology* 1: 2–9.
- Windsor, A. J., M. Reichelt, A. Figuth, A. Svatos, J. Kroymann, D. J. Kliebenstein, J. Gershenzon, and T. Mitchell-Olds. 2005. Geographic and evolutionary diversification of glucosinolates among near relatives of *Arabidopsis thaliana* (Brassicaceae). *Phytochemistry* 66: 1321–1333.
- Wysocki, W. P., L. G. Clark, S. A. Kelchner, S. V. Burke, J. C. Pires, P. P. Edger, D. R. Mayfield, et al. 2014. A multi-step comparison of short-read full plastome sequence assembly methods in grasses. *Taxon* 63: 899–910.
- Yang, R., D. E. Jarvis, H. Chen, M. A. Beilstein, J. Grimwood, J. Jenkins, S. Shu, et al. 2013. The reference genome of the halophytic plant *Eutrema salsugineum*. *Frontiers in Plant Science* 4: 46.