Mesopolyploidy as a taxonomic clade marker for Brassica and relatives (tribe Brassiceae)

Shawn K. Thomas, R. Shawn Abrahams, Daniel Robert Kick, Nora Walden, Gavin Conant, Michael R. McKain, Hong An, Tatiana Arias, Patrick P. Edger, Alex Harkess, Kasper P. Hendriks, Marcus A. Koch, Fredric Lens, Martin A. Lysak, Alex McAlvay, Klaus Mummenhoff, Ihsan A. Al-Shehbaz, Jacob D. Washburn, J. Chris Pires

Annals of Botany **00**: 1–15, 2025
https://doi.org/10.1093/aob/mcaf170, available online at www.academic.oup.com/aob

ANNALS OF BOTANY
Nonprofit since 1887

ORIGINAL ARTICLE

Mesopolyploidy as a taxonomic clade marker for *Brassica* and relatives (tribe Brassiceae)

Shawn K. Thomas[1,2,3,*,†], R. Shawn Abrahams[4,†], Daniel Robert Kick[5,6], Nora Walden[7], Gavin Conant[8,9], Michael R. McKain[10], Hong An[2], Tatiana Arias[11], Patrick P. Edger[12], Alex Harkess[13], Kasper P. Hendriks[14,15], Marcus A. Koch[7], Fredric Lens[14,16], Martin A. Lysak[17], Alex McAlvay[18], Klaus Mummenhoff[19], Ihsan A. Al-Shehbaz[20], Jacob D. Washburn[21,*] and J. Chris Pires[22,*]

[1]Division of Biological Sciences, University of Missouri, Columbia, MO 65211, USA, [2]Bioinformatics and Analytics Core, University of Missouri, Columbia, MO 65211, USA, [3]Living Earth Collaborative, Washington University in St. Louis, St. Louis, MO 63130, USA, [4]Department of Plant Biology, University of Illinois, Urbana-Champaign, IL 61801, USA, [5]United States Department of Agriculture, Agricultural Research Service, Columbia, MO 65211, USA, [6]Division of Plant Sciences, University of Missouri, Columbia, MO 65211, USA, [7]Centre for Organismal Studies, University of Heidelberg, Heidelberg 69120, Germany, [8]Department of Biological Sciences, North Carolina State University, Raleigh, NC 27695, USA, [9]Bioinformatics Research Center, North Carolina State University, Raleigh, NC 27695, USA, [10]Department of Biological Sciences, The University of Alabama, Tuscaloosa, AL 35487, USA, [11]Grupo de Investigación Biodiversidad, Medio Ambiente y Salud (BIOMAS), Universidad de Las Américas (UDLA), Quito 170513, Ecuador, [12]Department of Horticulture, Michigan State University, East Lansing, MI 48824, USA, [13]HudsonAlpha Institute for Biotechnology, Huntsville, AL 35806, USA, [14]Naturalis Biodiversity Center, Leiden 2300 RA, The Netherlands, [15]Department of Biology, Botany, University of Osnabrück, Osnabrück D-49076, Germany, [16]Institute of Biology Leiden, Plant Sciences, Leiden University, Leiden 2333 BE, The Netherlands, [17]Central European Institute of Technology and Department of Experimental Plant Biology, Faculty of Science, Masaryk University, Brno 625 00, Czech Republic, [18]Center for Plants, People, and Culture, New York Botanical Garden, The Bronx, NY 10458, USA, [19]Department of Biology/Chemistry, University of Osnabrück, Osnabrück D-49076, Germany, [20]Missouri Botanical Garden, St Louis, MO 63110, USA, [21]Plant Genetics Research Unit, USDA-ARS, Columbia, MO 65211, USA and [22]Soil and Crop Sciences, Colorado State University, Fort Collins, CO 80523-1170, USA

[*]For correspondence. E-mail shawnkt4@gmail.com, jacob.washburn@usda.gov or chris.pires@colostate.edu
[†]Co-first authorship.

Received: 31 January 2025 Returned for revision: 20 May 2025 Editorial decision: 15 July 2025 Accepted: 23 July 2025

• **Background and Aims** Whole-genome duplications (WGDs) are rampant in flowering plant genomes. Within Brassicaceae, the genus *Brassica* (including crop mustards) and relatives (tribe Brassiceae) are hypothesized to share an ancient mesohexaploidy, or whole-genome triplication (WGT), resulting from two WGD events (Br-α WGT). However, the phylogenetic boundaries of the Br-α WGT remain unknown.
• **Methods** We use phylogenomic assessments and divergence time analyses to place and date the Br-α WGT. We see conflicting topologies among the plastid and nuclear trees putatively due to polyploidy, hybridization and reticulate evolution. Despite this, we find tribe Brassiceae to be monophyletic in both trees.
• **Key Results** As currently circumscribed, tribe Brassiceae does not share the Br-α WGT. The sister clade to the rest of the tribe, containing the genera *Orychophragmus* and *Sinalliaria*, show no evidence of the Br-α WGT. Based on this placement, divergence time analyses indicate that the Br-α WGT would have occurred between 12.1 and 10.7 million years ago.
• **Conclusions** We propose a new taxonomic revision for the tribe Brassiceae based on the shared characteristics of the Br-α WGT. This presents a stable characteristic for the tribe, which was not the case in previous taxonomies based on morphological characters. These findings help clarify the history of the mustard crops and their relatives and resolve long-standing issues with the circumscription of the tribe Brassiceae.

Key words: Whole-genome duplication, phylogenomics, mustard crops, *Brassica*, Brassicaceae, Cruciferae

INTRODUCTION

Many plant taxonomic groups established before the advent of comparative genomics exhibit a history of whole-genome duplication (WGD) or polyploidy (Soltis *et al*., 2009; Schranz *et al*., 2012; One Thousand Plant Transcriptomes Initiative, 2019). This pattern likely stems from the influence of WGD events on phenotypic characteristics commonly used in taxonomic classification, such as the development of novel synapomorphic traits (Soltis and Soltis, 2016; Porturas *et al*., 2019), accelerated rates of diversification (Landis *et al*., 2018; Huang *et al*., 2020) and an increased capacity for hybridization (Koch, 2012; Nieto Feliner *et al*., 2020; Tseng *et al*., 2024). These clades have been critical to understanding how gene duplicates lead to the key innovations we associate with many plant lineages today (Soltis and Soltis, 2016; Clark and Donoghue, 2018; Hämälä *et al*., 2024). Taxonomic groups that share a history of WGD or other rare genomic characters have also emphasized the usefulness of natural taxonomic classification for comparative biology (Rokas and Holland, 2000; Renner, 2016). Natural taxonomic classification requires the group to represent all members of a shared ancestor and provide insight into the evolutionary processes that have defined them (Doolittle, 1999). However, determining which evolutionary events are important to the character of a clade and describing events like WGD is non-trivial. For example, distinguishing WGD event signals and placing those events on a phylogeny remains challenging for families that boast several WGD events at various time scales, including palaeo-, meso- and neopolyploid events (Mandáková *et al*., 2017; Mandáková and Lysak, 2018). Of these event types, mesopolyploid lineages are especially difficult to resolve phylogenetically as they are defined by an ongoing process of extensive fractionation in extant species, where mixed retention of duplicate genes and complex histories of hybridization can lead to rampant gene tree incongruence (Kuzmin *et al*., 2020; Booker *et al*., 2022; Hoang *et al*., 2024). Despite these challenges, describing and placing WGD events remains important to understanding evolution through comparative biology.

The Brassicaceae includes numerous polyploid lineages (Mandáková *et al*., 2017 ; Walden *et al*., 2020; Guo *et al*., 2021). Among these, the ancestral polyploidy in the tribe Brassiceae (Br-α whole-genome triplication (WGT)) is the most studied mesopolyploidy event (Lysak *et al*., 2005, 2007; Wang *et al*., 2011; Hao *et al*., 2021). The duplicated genes arising from this ancestral polyploidy are believed to be a key driver of the remarkable diversity and adaptability observed in crops belonging to this tribe. This diversity manifests in various ways, including enhanced resilience to environmental stresses, altered nutrient content, the evolution of self-incompatibility systems, and innovations in glucosinolate profiles, which are compounds with important roles in plant defence and flavour (Abrahams *et al*., 2020; Azibi *et al*., 2020; Thomas *et al*., 2024). This event was initially identified through research on economically significant mustard crops within the genus *Brassica*. Even before the advent of whole-genome sequencing technologies, comparative genetic maps constructed between three diploid *Brassica* species (*B. nigra*, *B. oleracea* and *B. rapa*) and the model plant species *Arabidopsis thaliana* offered the first clues that this ancient polyploidy was a WGT or hexaploidy (Lagercrantz and Lydiatet, 1996; Lagercrantz, 1998; Babula *et al*., 2003; Parkin *et al*., 2005). These maps revealed that the diploid *Brassica* genomes harbour three copies of an ancestral genome from a shared hexaploid ancestor. Subsequent comparative cytogenetic studies extended this finding beyond the *Brassica* genus. Researchers found that members of various Brassiceae subtribes, including Brassicinae, Cakilinae, Moricandiinae, Raphaniae, Vellinae and potentially Zillinae, all retained three copies of ancestral chromosomal segments (Lysak *et al*., 2005, 2007; Ziolkowski *et al*., 2006) (Fig. 1, Supplementary Data Table S1), indicating a basal placement of the Br-α WGT in the evolutionary history of the tribe. Consistent with the expectations for a WGT event, comparative genomic investigations in *Brassica* species suggest that the ancestral hexaploid genome arose through a two-step process (Wang *et al*., 2011; Tang *et al*., 2012; Tang and Lyons, 2012; Cheng *et al*., 2013, 2014; He *et al*., 2021).

Accurately mapping multi-step WGD events onto a phylogenetic tree is challenging due to factors including the ages of overlapping duplication events, incomplete lineage sampling, and whether a lineage that experienced only part of the multi-step event still exists (Bayat *et al*., 2018; Mabry *et al*., 2020; Parma *et al*., 2021; Hoang *et al*., 2023). The current understanding of the Br-α WGT is that there was an initial tetraploidization event followed by a secondary allopolyploidy after a period of diploidization (Wang *et al*., 2011). This has resulted in the species of the Brassiceae having two fractionated subgenomes from the initial tetraploidy (MF and IF, also referred to as MF2 and MF1, respectively) and a less fractionated subgenome (LF) from the secondary event, providing the ancestral mesohexaploid genome with three subgenomes (Wang *et al*., 2011; Hao *et al*., 2021). The genera *Orychophragmus* (up to six species; Hu *et al*. (2016)) and *Sinalliaria* (two species; Zhang *et al*. (2018)) are the earliest-diverging lineage of the Brassiceae, forming a sister clade (henceforth referred to as the Orychophragmus lineage) to the remaining Brassiceae genera (Walden *et al*., 2020; Hendriks *et al*., 2023). Previous studies using chloroplast markers had identified new lineages within the Brassiceae but did not sample *Orychophragmus* spp. (Arias and Pires, 2012; Arias *et al*., 2014). Chromosome painting in *Orychophragmus violaceus* revealed an ancient tetraploidy event (Lysak *et al*., 2007), and recent genome sequencing of *O. violaceus* suggests that the mesotetraploid *O. violaceus* genome originated independently and later than the Br-α mesohexaploidy (Huang *et al*., 2023; Zhang *et al*., 2023).

These findings raise two key questions: (1) where to place the ancestral Br-α WGT in the tribe's evolutionary history, both in dated time and along a specific branch or branches of the phylogeny, and (2) whether the Orychophragmus lineage presents a distinct enough evolutionary history to be excluded from the Brassiceae based on a system of natural classification. In this study, we leverage transcriptome data across tribe Brassiceae and out-groups to assess which subtribal lineages share the Br-α WGT by analysing gene tree topologies, genomic synteny and K_s plots. We combine plastid data across the rosids (sampling focused on tribe Brassiceae) with fossil evidence to provide a robust date for the clade sharing the Br-α WGT. Further, based on these results, we explore incongruence in nuclear and plastid phylogenies and taxonomic implications for the group. Our findings support a robust placement for the Br-α WGT between 12.1 and 10.7 million years ago (Ma)

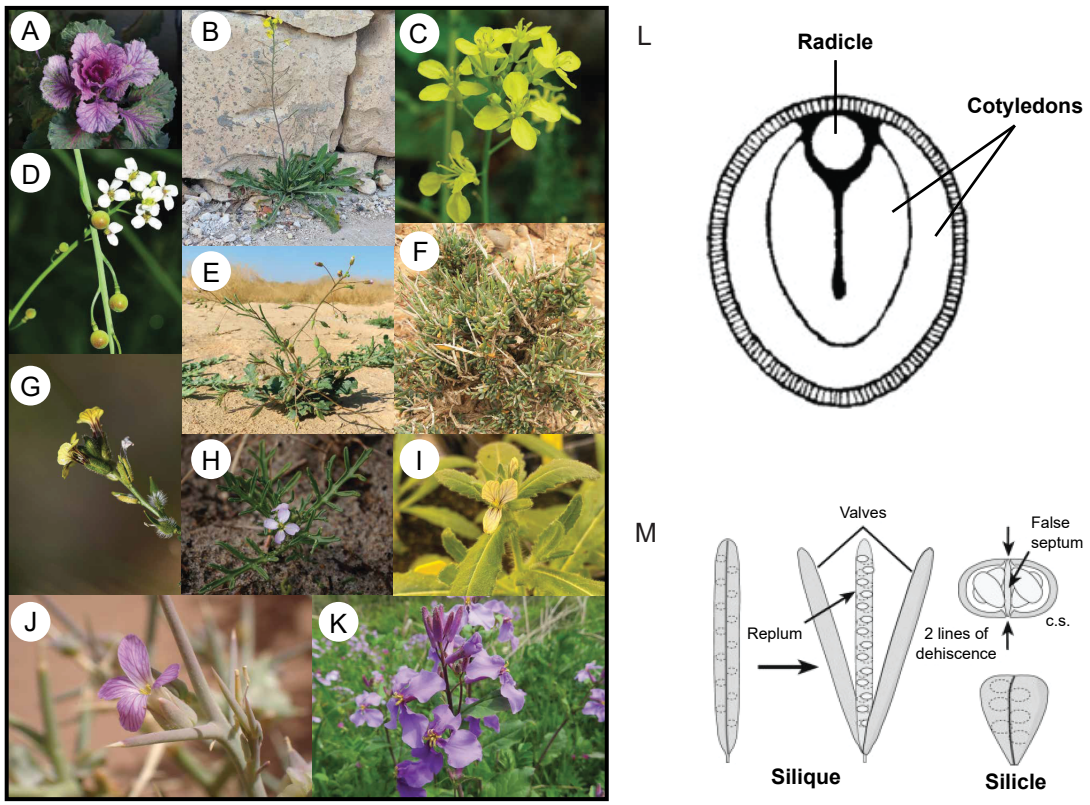

FIG. 1. Morphological diversity in the tribe Brassiceae. Panels A–J indicate species in a lineage that share the Br-α WGT, and panel K indicates species in the tribe Orychophragmeae. (A) *Brassica oleracea* (Brassica clade): https://www.inaturalist.org/observations/207775994. Observed in New Zealand by Agnes Trekker (licensed under https://creativecommons.org/publicdomain/zero/1.0/). (B) *Diplotaxis tenuifolia* (Brassica clade): https://www.inaturalist.org/observations/200285504. Observed in Malta by Daniel Cahen (licensed under https://creativecommons.org/licenses/by/4.0/). (C) *Erucastrum nasturtiifolium* (Brassica clade): https://www.inaturalist.org/observations/44368621. Observed in Austria by Patrick Hacker (licensed under https://creativecommons.org/licenses/by/4.0/). (D) *Crambe hispanica* (Crambe clade): https://www.inaturalist.org/observations/74446084. Observed in Spain by faluke (licensed under https://creativecommons.org/licenses/by-nc/4.0/). (E) *Savignya parviflora* (Savignya clade): https://www.inaturalist.org/observations/241829952. Observed in Israel by Theodore (licensed under https://creativecommons.org/licenses/by-nc/4.0/). (F) *Henophyton zygarrhenum* (Henophyton clade): https://www.inaturalist.org/observations/180244532. Observed in Morocco by missour (licensed under https://creativecommons.org/licenses/by-nc/4.0/). (G) *Vella aspera* (Vella clade): https://www.inaturalist.org/observations/256012267. Observed in Spain by Francisco Javier Carela Quilez (licensed under https://creativecommons.org/licenses/by-nc/4.0/). (H) *Cakile maritima* (Cakile clade): https://www.inaturalist.org/observations/168123356. Observed in Denmark by Margit Kildevang (licensed under https://creativecommons.org/publicdomain/zero/1.0/). (I) *Psychine stylosa* (Psychine clade): Observed in Missouri (greenhouse setting) by Shawn K. Thomas. (J) *Zilla spinosa* (Zilla clade): https://www.inaturalist.org/observations/206145185. Observed in Morocco by frederikl (licensed under https://creativecommons.org/licenses/by-nc/4.0/). (K) *Orychophragmus violaceus* (Orychophragmus clade): https://www.inaturalist.org/observations/153751462. Observed in Japan by belvedere04 (licensed under https://creativecommons.org/licenses/by/4.0/). (L) Cartoon image depicting conduplicate cotyledons, a morphological character shared by many members of the tribe Brassiceae (modified from Iltis *et al.* (2011)). (M) Cartoon image depicting silique and silicle fruit types (from Simpson (2020)). Fruits in the tribe Brassiceae are often segmented by the replum (in siliques) or the false septum (in silicles).

including the majority of the tribe Brassiceae, but excluding its earliest-diverging clade, the Orychophragmus lineage. We propose that a circumscription of the tribe Brassiceae based on the placement of the Br-α WGT presents a more accurate representation of the evolutionary history of the lineage under natural taxonomic classification and a better model clade for comparative biology.

MATERIALS AND METHODS

Taxon sampling

In order to sample both transcriptome and plastid DNA sequences, seeds or living plant collections from across the tribe were required. Seeds were obtained from various sources (see accession information in Supplementary Data Tables S2 and S3) and grown in greenhouse conditions. A limitation of this approach was that seed was not available for some taxa; notably, living tissue could not be obtained for *Henophyton*, precluding the generation of transcriptomes for this taxon. Plastid sequence data for *Henophyton* were available (Walden *et al.*, 2020), enabling plastid phylogenetics. However, these missing taxa do not hinder estimating the phylogeny and chronogram of the remaining taxa nor the placement and age of the Br-α WGT. All taxa sampled and their sources can be found in Supplementary Data Tables S2 and S3.

DNA and RNA isolation and sequencing

Plants were grown in a sterile growth chamber or greenhouse setting at the University of Missouri as described in Qi *et al.* (2017), and samples were taken from whole leaves or shoots. For RNA, samples were taken as close to noon as possible for consistency. However, since the aim of this sampling was to

extract sequence data from RNA samples rather than perform expression analysis, the growth stage was not controlled. Different samples were taken for DNA and RNA. In some cases, the two samples were from the same plant and in other cases they were siblings or progeny of inbred lines, depending on the availability of tissue and the health of the plants.

DNA was isolated using the DNeasy Plant Kit (Qiagen, Germantown, MD, USA) or, in some cases when the kit failed to provide sufficient DNA, a standard CTAB protocol was used (Doyle and Doyle, 1987). Libraries were prepared using TruSeq DNA Library Preparation Kits (Illumina, San Diego, CA, USA) or Nextera DNA Library Preparation Kits (Illumina, San Diego, CA, USA) and sequenced using the HiSeq (Illumina, San Diego, CA, USA) with paired end reads of 100-bp length or the NextSeq 500 (Illumina, San Diego, CA, USA) with paired end reads of 150-bp length at the University of Missouri (Supplementary Data Table S2). Total RNA was isolated from flash-frozen tissue using the Invitrogen PureLink RNA Mini Kit (Thermo Fisher Scientific, Carlsbad, CA) and libraries were prepared using the TruSeq mRNA Library Preparation Kit (Illumina, San Diego, CA, USA) with some libraries being stranded and others not stranded due to kit availability and timing (Supplementary Data Table S3). mRNA libraries were sequenced on an Illumina HiSeq with paired end reads of 100-bp lengths.

Plastid gene and transcriptome assembly

DNA sequencing raw reads were cleaned using fastp v. 0.21.0 (Chen *et al.*, 2018) to automatically detect and remove adapter sequences. All other cleaning options in fastp were kept as default and only cleaned reads in which both read pairs survived were kept for downstream analyses. The *Brassica rapa* plastid RefSeq proteins (NCBI Reference Sequence NC_040849.1) were selected as target genes for assembly in the HybPiper pipeline v. 2.1.6 (Johnson *et al.*, 2016). In this pipeline, reads were mapped to the *B. rapa* targets using the Burrows–Wheeler alignment (BWA) tool (Li and Durbin, 2009). Mapped reads for each target were assembled *de novo* into contigs using the best *k*-mer detected using SPAdes v. 3.13.0 (Prjibelski *et al.*, 2020). The exon contigs were then aligned to the reference to extract coding sequence and then scaffolded using exonerate (Slater and Birney, 2005). Finally, CDS sequences were included from existing data on NCBI for *Brassica oleracea* (NCBI Reference Sequence NC_041167.1) and *Brassica nigra* (NCBI Reference Sequence NC_030450.1). These data were combined with existing data from Walden *et al.* (2020) for downstream analyses.

RNA sequencing (RNA-seq) raw reads were cleaned using Trimmomatic v. 0.39 (Bolger *et al.*, 2014) to remove adapter sequences, remove leading and trailing low-quality or N bases, and keep reads with a minimum length of 36 bp (LEADING:3, TRAILING:3, MINLEN:36). Transcriptomes were assembled using Trinity v. 2.11.0 (Grabherr *et al.*, 2011) using default parameters. Each transcriptome was clustered to produce representative sequences using CD-HIT-EST (Fu *et al.*, 2012) and a threshold of 90 % sequence identity and word size of 8 or 9 (-c 0.90 -n 8,9). These representative sequences were filtered to only keep open reading frames of at least 100 amino acids in length and to predict likely coding regions using Transdecoder v. 5.5.0 (B. J. Haas, https://github.com/TransDecoder/TransDecoder). These assemblies were then validated for quality using Transrate v. 1.0.3 (Smith-Unna *et al.*, 2016) and completeness using BUSCO v. 5.4.7 (Manni *et al.*, 2021).

Finally, to increase sampling, existing sequence data from Phytozome, the Brassicales Map Alignment Project (BMAP), the Brassicaceae Database (BRAD) and past publications were used. RNA-seq data from An *et al.* (2019) for *Brassica cretica*, *B. hilarionis*, *B. montana* and *B. villosa* and from Wang *et al.* (2019) for *Sinalliaria limprichtiana* and *S. limprichtiana* var. *grandifolia* were used to generate transcriptomes (as described above). In addition, primary CDS and peptide sequences were pulled for the following species: *Arabidopsis thaliana* (Lamesch *et al.*, 2012), *Brassica nigra* (Perumal *et al.*, 2020), *B. oleracea* (Liu *et al.*, 2014), *B. rapa* (Wang *et al.*, 2011; Lou *et al.*, 2020), *Cakile maritima*, *Caulanthus amplexicaulis*, *Eutrema salsugineum* (Yang *et al.*, 2013), *Iberis amara*, *Isatis tinctoria*, *Lunaria annua*, *Myagrum perfoliatum*, *Schrenkiella parvula* (Dassanayake *et al.*, 2011; Oh *et al.*, 2014), *Sisymbrium irio* (Haudry *et al.*, 2013), *Stanleya pinnata* and *Thlaspi arvense* (data sources can be found in Supplementary Data Table S3). These sequence data were used in conjunction with the transcriptomes generated for this study in downstream analyses.

Coalescent tree estimation

A multilocus species tree estimation approach was used to reconstruct the phylogeny from transcriptomic data. Orthology was inferred in OrthoFinder v. 2.5.4 (Emms and Kelly, 2019) using parameters -t 20 and -a 20 to use 20 parallel analysis threads for sequence search, multiple sequence alignment (MSA), and tree inference (-t 20), and internal, RAM-intensive tasks (-a 20). All other analysis parameters were left as default. Then orthogroups were filtered by taxon occupancy for a minimum of 40 (out of 54) samples with at least one sequence each, resulting in 13 506 orthogroups. These orthogroups were aligned using MAFFT v. 7.453 (Katoh and Standley, 2013) and gene tree inference was conducted using IQ-TREE v. 2.3.6 (Minh *et al.*, 2020). Finally, ASTRAL-Pro 3 v. 1.19.3.6 (Zhang *et al.*, 2020; Zhang and Mirarab, 2022; Tabatabaee *et al.*, 2023) was used to construct the multilocus species tree from all gene trees, including trees with duplicate/multicopy genes per species.

Mesohexaploidy placement

One line of evidence for ancient polyploidy events is to compare patterns of putative paralogues across gene trees with a single species tree. The multilocus species tree inferred by ASTRAL-Pro 3 (Zhang *et al.*, 2020; Zhang and Mirarab, 2022; Tabatabaee *et al.*, 2023) and all of the gene trees used to infer that multilocus species tree were used as inputs for the program PUG v. 2.1.1 (McKain *et al.*, 2016). This program uses an algorithm that queries putative paralogues in a gene tree and identifies species congruence relative to the species tree as described in McKain *et al.* (2016). Parameters were set to estimate paralogues from the gene trees by identifying all potential combinations of multicopy genes from each taxon. The R script packaged with PUG was used to generate the phylogenetic tree with duplicate genes indicated on the branches. Only duplicate pairs that had a bootstrap value of ≥80 leading to their coalescing branch were considered. To aid in the visualization of the data, branches with a minimum of 25 % of the highest total number

of unique duplicate genes found on any branch (here, the one leading to the two *Orychophragmus violaceus* samples) were marked on the phylogeny.

The distribution of the synonymous divergence (K_s) between paralogues in each transcriptome is another line of evidence for the placement of an ancient polyploidy event. As gene copies diverge from each other over evolutionary time, their K_s will increase and the position of the peak along the K_s axis provides an estimate of when the event occurred. Large-scale duplication events like polyploidy provide a burst of many new duplicate genes and therefore are expected to create a peak corresponding to when the event occurred. Thus, we can use these peaks as a line of evidence of when an ancient polyploidy occurred in a species, and further compare them across the phylogeny to get a sense of which lineages share ancient polyploidy events. To obtain K_s values, translated CDS and peptide sequences from Transdecoder or Phytozome were used as input for the FASTKs pipeline (McKain *et al.*, 2016). In this pipeline, CDS sequences for each sample were aligned against themselves with an e-value cut-off of 1e−40 in BLAST (Altschul *et al.*, 1990). Putative pairs were identified as BLAST hits when they met the following criteria: (1) >40 % identity in alignment, (2) not an exact match (<100 % identity), and (3) >300 bp of alignment length. Peptide sequences for the putative pairs were aligned with MUSCLE (Edgar, 2004), and back-translated to CDS using PAL2NAL (Suyama *et al.*, 2006). K_s, non-synonymous divergence (K_a) and K_a/K_s were estimated for the aligned pairs using codeml in PAML v.4.8 (Yang, 2007) using the same parameters used by McKain *et al.* (2012).

To determine if the K_s distributions could be statistically differentiated, the distribution's empirical cumulative distribution functions were compared using a Kolmogorov–Smirnov (K–S) test. This test was performed using R's built-in implementation, ks.test(). In cases where distributions contained an unequal number of samples, the distribution with more samples was systematically downsampled, retaining samples evenly spaced through the observed distribution.

Simulated K_s distributions were generated for the following clades: the Br-α in-group (samples that share Br-α WGT), the non-Br-α out-group (samples that do not share Br-α WGT) and the subtribal lineages of species sharing the Br-α WGT. Distributions were simulated by drawing samples from each species' substitution distribution for those species in a given set (e.g. the members of a clade). The samples drawn from each species were dictated by the total number of samples times the desired proportion in the simulated distribution rounded to the nearest integer. The total number of samples was defined as the average number of samples among the species to be used, rounded to the nearest integer. The desired proportion for each species was determined using the relatedness in the phylogenetic tree. For each species the path length to every parent node was calculated and used to find the distance from each node to its child species. The representation of each species' distribution at a given node was assumed to be inversely proportional to the distance between that node and the species. In cases where one or more species was removed from consideration (e.g. simulating the distribution of the out-group) the expected fractions of the disallowed species were set to zero and then the fractions were normalized to sum to one. Although this approach resulted in stochastic results, samples were randomly drawn and the sheer number of samples being drawn (17 223–167 646 samples) reduced the expected variation between simulations. Custom R functions were created for simulating substitution distributions based on the phylogenetic tree and distributions for each species and for comparison and visualization of empirical cumulative distribution functions. For details, please refer to the associated GitHub repository (http://www.github.com/danielkick/brassica_ks).

Modelling the proposed WGD events with POInT

We also modelled the polyploidy events in *O. violaceus* and Br-α using the software package POInT v. 1.61 (the Polyploidy Orthology Inference Tool) (Hao and Conant, 2022) to assess the evidence for *O. violaceus* sharing the first event in Br-α. To identify duplicated genes in the genome of *O. violaceus*, we conducted a homology search between it and the genome of *A. thaliana* Col-0 (v. 10.29; CoGe genome id 20342) using GenomeHistory v. 2.1 (Conant and Wagner, 2002) and BLASTP v. 2.7.1 (Altschul *et al.*, 1997). We required a BLAST e-value 10^{-9} and ≥60 % amino acid identity to retain a homologue pair. We then used simulated annealing to search for blocks of double-conserved synteny (DCS) in *O. violaceus* relative to *A. thaliana* (Hao and Conant, 2022). Homologue pairs with K_a >0.5 were omitted from this analysis.

To assess whether the genome duplication in *O. violaceus* corresponded to a shared polyploidy with any of the pairs of subgenomes of the Br-α WGT, we next extracted all three pairs of subgenomes from the four Brassiceae genomes we had previously analysed (Hao *et al.*, 2021). These correspond to the least and intermediately fractionated subgenomes (LF/IF), least and most fractionated subgenomes (LF/MF) and intermediately and most fractionated subgenomes (IF/MF). In each case, we extracted pillars from our dataset where all genes were assigned to their respective subgenomes with ≥95 % confidence (Hao *et al.*, 2021). We then merged these pillars with our inferences from *O. violaceus* to produce three hypothesized shared genome duplications: LF/IF with 7093 pillars, LF/MF with 6472 pillars and IF/MF with 4535 pillars. For each dataset, we sought an optimal syntenic pillar order using simulated annealing as previously described (Hao and Conant, 2022).

Chronogram of clades and mesohexaploidy timing

Plastome data from 45 samples that were newly sequenced or acquired from publicly available data were combined with an existing dataset comprising 60 plastid-encoded genes from all Brassicaceae tribes and rosid out-groups (Walden *et al.*, 2020) to estimate divergence times in the Brassiceae. From the existing dataset, all samples from non-Brassicaceae out-groups, tribe Brassiceae and closely related tribes Thelypodieae and Sisymbrieae were included. Finally, for all other Brassicaceae tribes outside Brassiceae, Thelypodieae and Sisymbrieae a few representative species from each were used, resulting in a combined number of 216 samples (Supplementary Data Table S2). Newly generated coding sequences were manually aligned to existing gene alignments in AliView v. 2021 (Larsson, 2014), and alignments were trimmed to the same size as the existing dataset.

Divergence time estimation was conducted in BEAST v. 2.6.7 (Bouckaert *et al.*, 2019) following Walden *et al.* (2020). This analysis used a single partition with substitution model GTR with estimated frequencies, four γ categories and

estimated substitution rates, shape parameter and proportion of invariant sites. The log-normal relaxed clock (Drummond *et al.*, 2006) with estimated clock rate was used with a birth–death tree model (Gernhard, 2008). Four fossil calibration points from across the rosids (Supplementary Data Table S4; Knobloch and Mai, 1986; Sims *et al.*, 1999; Takahashi *et al.*, 1999; Bell *et al.*, 2010; Moore *et al.*, 2010; Benedict *et al.*, 2011; Njuguna *et al.*, 2013) were implemented as minimum ages for calibration with a uniform distribution, and the root height was constrained using a uniform prior with minimum age 92 Ma; 125 Ma was set as upper bound for all five calibration points following methods from Hohmann *et al.* (2015) and Walden *et al.* (2020). A starting tree was generated using RAxML v. 8.2.12 (Stamatakis, 2014) with substitution model GTRGAMMA and 1000 rapid bootstrap replicates followed by thorough ML search (option -f a). Node ages in the starting tree were adjusted to match fossil calibration time using chronos in R package ape v. 5.6-2 (Paradis and Schliep, 2019) in R v. 4.2.0. Eight independent runs with 200 000 000 generations each were started, logging every 20 000 generations.

Convergence of the runs was checked in Tracer v. 1.7.2 (Rambaut *et al.*, 2018), then trees from all eight runs were combined using LogCombiner v. 2.6.7 (Bouckaert *et al.*, 2019) while discarding the first 10 % of generations as burn-in. The maximum clade credibility tree was obtained using TreeAnnotator v. 2.6.7 (Bouckaert *et al.*, 2019).

RESULTS

Sequence matrices

DNA read pools after trimming ranged in size from 245 347 to 95 765 353 read pairs per sample with an average of 5 750 314 read pairs. RNA read pools after trimming ranged in size from 18 534 657 to 51 625 028 reads per sample, with an average of 28 931 113 reads per sample. BUSCO scores for transcriptomes assembled for this study had an average of 86.01 % complete genes (ranging from 74.80 to 93.30 %), 6.95 % fragmented genes (ranging from 4.50 to 12.80 %) and 7.04 % missing genes (ranging from 2.20 to 13.40 %). The publicly available transcriptomes had BUSCO scores with an average of 94.51 % complete genes (ranging from 70.70 to 99.30 %), 1.85 % fragmented genes (ranging from 0.20 to 8.30 %) and 3.65 % missing genes (ranging from 0.50 to 21.00 %). All BUSCO and Transrate metrics for all samples can be found in Supplementary Data Tables S5–S8. From all the transcriptomes, 106 109 orthogroups (homologous loci across samples) were identified with OrthoFinder 2.5.4 using default settings. Orthogroups were then filtered by taxon occupancy for a minimum of 40 (out of 52) samples with at least one sequence each, resulting in 13 506 orthogroups.

Plastid phylogenomics

The divergence time analysis resulted in a plastid gene phylogeny that is congruent with the family-level phylogeny presented in Walden *et al.* (2020). Within tribe Brassiceae all branches except five were recovered with 0.7 local posterior probability (LPP) support or better in the plastid phylogeny (Supplementary Data Fig. S1). The backbone of the plastid tree is well supported, with the exception of the branch leading to the clade of the subtribal lineages (as classified by Arias and Pires (2012)) [Nigra + Crambe + Rapa/Oleracea + Savignya + Henophyton], which had an LPP of 0.56. The branch leading to [Brassica-juncea + Brassica-rapa-pekinensis + Brassica-rapa Brassica-oleracea + Brassica-napus + Raphanus-raphanistrum + Raphanus-sativus] had an LPP of 0.68 and the remaining three branches with <0.7 LPP were branches leading to closely related species or species complexes. While species relationships within subtribal lineages are fairly consistent, the relationships across these lineages have some differences. Notably, we see differences in the branching patterns of the tribe Brassiceae. The lineage Henophyton, which was sister to the Rapa/Oleracea + Savignya + Nigra + Cakile + Crambe lineages in Arias and Pires (2012), is placed sister to Rapa/Oleracea + Savignya. Nigra + Crambe lineages form a clade sister to Rapa/Oleracea + Savignya + Henophyton lineages, and the Cakile lineage is sister to the rest of this group. The early-diverging lineages Vella and Zilla are shown to form a grade with the Zilla lineage diverging first and then the Vella lineage, whereas in Arias and Pires (2012) the Vella and Zilla lineages formed a clade. In addition, we find *Psychine stylosa* (henceforth referred to as the Psychine lineage) sister to the Vella lineage in the plastid phylogeny. In previous studies, the Psychine lineage has been placed within either the Vella (Warwick and Black, 1994) or the Savignya (Crespo, 2000) lineage. Finally, the Orychophragmus lineage is found to be sister to the rest of the tribe Brassiceae.

Transcriptome phylogenomics

The coalescent tree estimation resulted in a well-supported nuclear tree (Supplementary Data Fig. S2). All branches were recovered with full support (1 LPP). The topology of the tree is inconsistent with the plastid gene phylogeny and relationships between major clades in the tribe are shown in Fig. 2. Quartet frequencies between major clades in the tribe are shown in Supplementary Data Fig. S3. The Nigra and Rapa/Oleracea lineages are nested within each other and form a monophyletic clade (henceforth referred to as the Brassica lineage). The Crambe lineage is sister to this clade, the Cakile lineage is sister to the Brassica + Crambe lineages, and the Savignya lineage forms a clade sister to all of these. Finally, we see the lineages Psychine, Zilla and Vella respectively forming a grade with the Vella lineage sister to the rest. In contrast, the plastid tree has the lineages Psychine and Vella in a clade with the Zilla lineage sister to the rest. Despite the inconsistent placement of lineages in both trees, the Orychophragmus lineage is found to be sister to the rest of tribe Brassiceae in both trees, preserving the monophyly of the tribe Brassiceae.

Using signatures from retained paralogues post-mesohexaploidy, we were able to place the Br-α WGT. These signatures include the analysis of (1) gene tree topologies and (2) K_s plots. Using PUG (github.com/mrmckain/PUG), we queried gene trees to identify branches in the species tree with patterns of shared duplications in the gene trees (Supplementary Data Fig. S4). We see evidence for 1425 unique shared duplications in tribe Brassiceae on the branch sharing most of the subtribal lineages, but excluding the sister Orychophragmus lineage (Fig. 3, Supplementary Data Fig. S4). The two samples of *O. violaceus* have 7283 unique duplications. In addition, we see evidence for

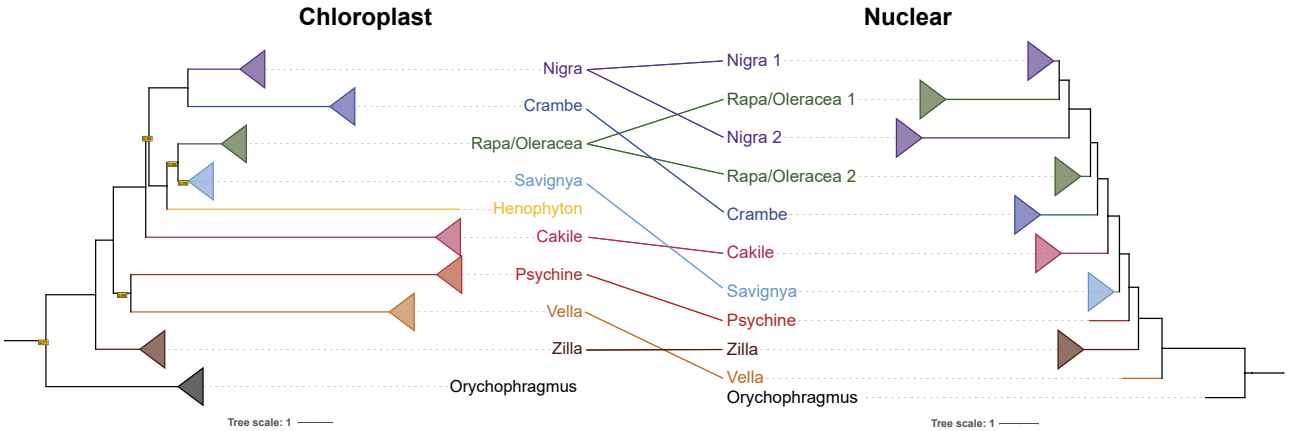

FIG. 2. Plastid and nuclear tree incongruence. Comparison of maximum likelihood 60 plastid gene phylogeny (Chloroplast) with coalescent-based species phylogeny (Nuclear) of tribe Brassiceae. Major clades of the tribe are indicated. Support values are indicated wherever the local posterior probability is <0.99.

shared duplications in the backbone of the tree on the branches leading to Brassiceae + Isatideae + Sisymbrieae + Thelypodieae (3082 duplications) and the next branch out including Brassiceae + Isatideae + Sisymbrieae + Thelypodieae + Schrenkielleae (3581 duplications) (Supplementary Data Fig. S4). Furthermore, we see shared duplications within the tribe at shallower branches. The branch leading to the clade with *B. rapa*, *B. oleracea* and wild relatives of *B. oleracea* has 1247 shared duplications (Supplementary Data Fig. S4). In addition, *B. oleracea* and its wild relatives (excluding *B. rapa*) share 2749 duplications and 739 duplications on the branch excluding *B. villosa* (Supplementary Data Fig. S4). The branch leading to *Diplotaxis harra* and *Coincya longirostra* has 2250 shared duplications, and the branch leading to *Zilla macroptera* and *Foleyola billotii* has 1394 shared duplications (Supplementary Data Fig. S4). The branches leading to the genera *Eruca* and *Cakile* have 1584 and 2201 shared duplications, respectively (Supplementary Data Fig. S4). The branch leading to *Crambe filiformis*, *C. kralikii* and *C. hispanica* shares 850 duplications (Supplementary Data Fig. S4). Finally, outside of the tribe Brassiceae there are 1270 shared duplications on the branch leading to tribe Isatideae (represented in the transcriptome phylogeny by *Myagrum perfoliatum* and *Isatis tinctoria*) (Supplementary Data Fig. S4). All counts of unique shared duplications per node can be found in Supplementary Data Table S9.

The distribution of the synonymous divergence (K_s) between paralogues in each transcriptome is one method that can provide evidence for the placement of a polyploid event. Our data show that across the tribe Brassiceae all samples except for ones in the Orychophragmus lineage share a peak close to $K_s \sim 0$ and another one around $K_s \sim 0.3$ (Fig. 3, Supplementary Data Figs S5 and S6). This result also held when assessing peaks from the simulated clade K_s distributions (Fig. 3). Typically peaks around $K_s \sim 0$ correspond to recent tandem duplications and are not relevant to ancient polyploidy events, so the peak around $K_s \sim 0.3$ gives evidence that the Brassiceae mesohexaploidy (Br-α WGT) is only shared with tribe Brassiceae excluding the Orychophragmus lineage. In the Orychophragmus lineage and other lineages outside of tribe Brassiceae, we see that the distributions in the K_s plots are widely variable, and these could be indicative of independent ancient polyploidy events in those lineages. There is some signal for a peak around $K_s \sim 0.6$–0.7, which may be from the family level WGD (At-α WGD).

Although the peaks can be distinguished by eye, we used the K–S test to statistically differentiate the shape of simulated clade distributions. Simulated clade distributions were used to reduce the number of pairwise comparisons. When assessing the simulated in-group (tribe Brassiceae excluding the Orychophragmus lineage) distribution against the simulated out-group and the *O. violaceus* sample, we get $P = 0$ and $P < 2.2e{-}16$, respectively (Fig. 3D, E). In addition, K–S tests between the in-group and *Sinalliaria* species result in $P = 0$ (Supplementary Data Fig. S7). We conducted pairwise K–S tests comparing simulated clades within the in-group (tribe Brassiceae excluding the Orychophragmus lineage) with each other and with the simulated in-group and could not statistically differentiate them (P values ranging from 0.20 to 0.99) (Supplementary Data Figs S8 and S9). Finally, we compared each simulated clade with the simulated out-group and got $P = 0$ for each comparison (Supplementary Data Fig. S10). Given both the visual and statistical assessment of the K_s distributions, the results indicate that the Br-α WGT is shared only with tribe Brassiceae excluding the Orychophragmus lineage (henceforth referred to as the Br-α WGT clade).

Test for a single, shared polyploidy event

For all three datasets, we modelled the loss of ohnologous genes (gene duplicates originated from WGDs) using a model of duplicated gene loss (Fig. 4A) that allowed for duplicate fixation and biased fractionation using POInT. In each case, we then used POInT to search for the maximum likelihood phylogenetic topology using an exhaustive tree search, which yielded the two topologies seen in Fig. 4B, C.

To test whether there was evidence for a shared WGD between *O. violaceus* and any of the pairs of subgenomes from the Brassiceae mesohexaploidy, we used a simulation approach as previously described (Conant and Wolfe, 2008). The null hypothesis of no shared event corresponds to a tree like that in Fig. 4B, C with a zero length root branch, corresponding to no shared losses between the *O. violaceus* polyploidy and the subgenomes of the Brassiceae hexaploidy. We hence fitted a model with a fixed zero-length root branch to each of the three datasets. Using the model parameters found when doing so, we

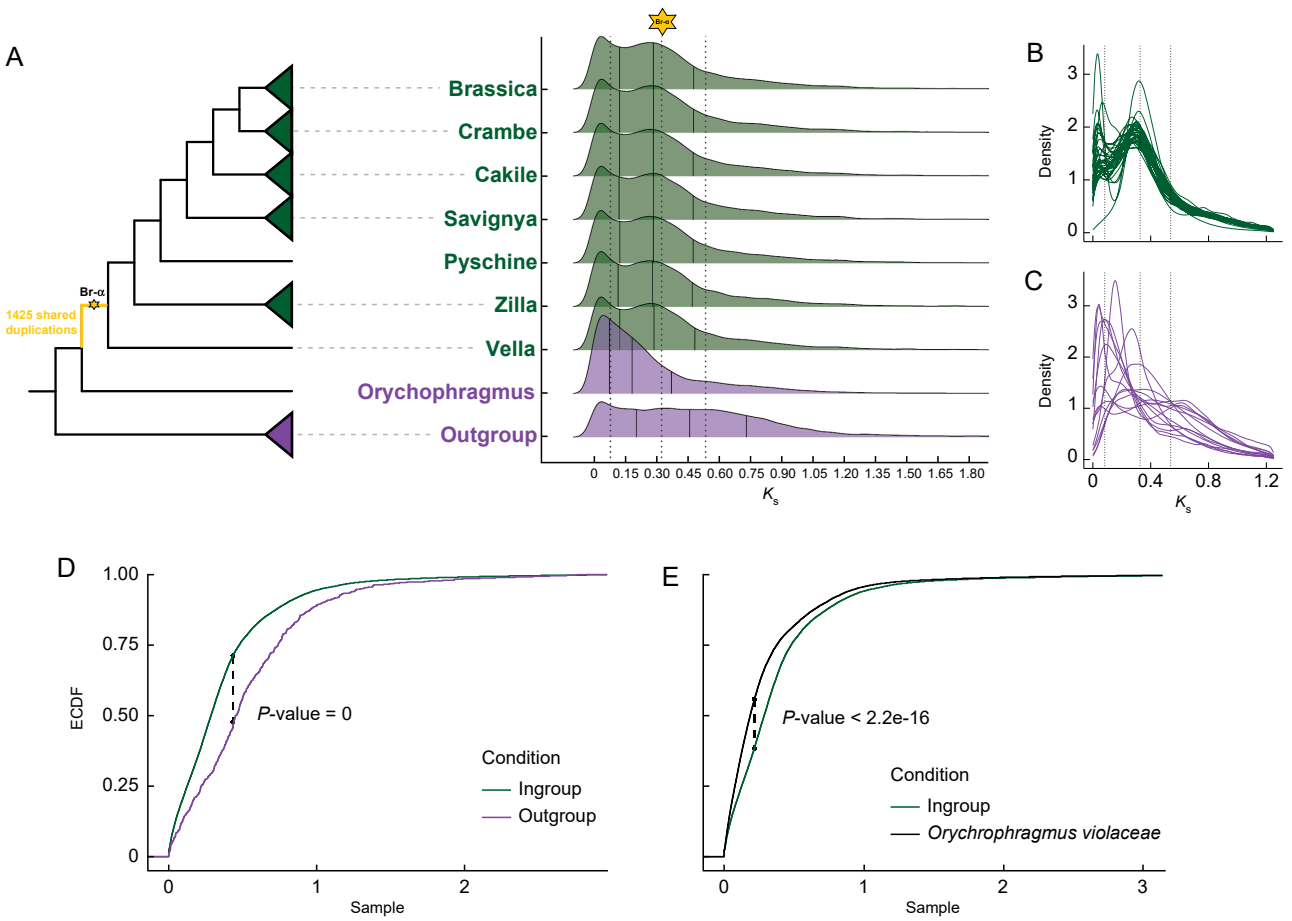

FIG. 3. Placement of the Br-α WGT. (A) Coalescent-based species phylogeny and simulated clade K_s distributions. Branch colour denotes the number of unique gene duplications as determined by PUG (github.com/mrmckain/PUG). Simulated clade K_s distributions are shown as density plots corresponding to tips on the phylogeny. Dotted lines, based on Qi *et al.* (2021), show the K_s range of the Br-α WGT inferred from *Brassica rapa*. Solid black lines separate distributions into quartiles. Yellow star indicates the Br-α WGT. (B) All K_s distributions from samples inferred to share the Br-α WGT. (C) All K_s distributions from samples inferred to not share the Br-α WGT. (D) Comparison of the simulated in-group (green) and out-group (purple) ECDFs with the K–S test ($P = 0$). (E) Comparison of the simulated in-group (green) and sister lineage Orychophragmus (black) ECDFs with the K–S test ($P < 2.2e{-}16$).

then simulated 100 sets of five genomes from this topology that assumes two independent events. To each simulated dataset, we fitted a model with a zero-length root branch (Model 1) and a model where the root branch was allowed to have an arbitrary length (Model 2). We then compared the distribution of the length of the root branch in the simulations analysed under Model 2 with the observed length of the root branch in the actual datasets (Fig. 4D–F for LF/IF, LF/MF and IF/MF, respectively). In every case, the real dataset showed a shorter root branch than seen in the simulations, corresponding to a failure to reject the null hypothesis of no shared polyploidy between *O. violaceus* and the Brassiceae species.

Chronogram and mesohexaploidy timing

Combining plastid sequences focused on the tribe Brassiceae (but spanning the rosids) and robust fossil calibrations in the rosids, we were able to estimate divergence times for clades of interest. Overall, the divergence time estimates were consistent with the results presented in Walden *et al.* (2020). With the additional sampling included in this study we see slightly earlier age estimates for tribe Brassiceae. This study estimates tribe Brassiceae with a crown age of 12.14 Ma, stem age of 12.29 Ma and crown 95 % HPD of 9.7–14.7 Ma. The Walden *et al.* (2020) study estimated tribe Brassiceae with a crown age of 12.29 Ma, stem age of 12.57 Ma and crown 95 % HPD of 9.81–15.16 Ma. Both the K_s and phylogenomic analyses presented above place the Br-α WGT on the branch leading to tribe Brassiceae excluding the Orychophragmus lineage (Br-α WGT clade), so we used the age estimates at that branch to represent the age of Br-α WGT. Therefore the age estimates for the branch where Br-α WGT occurred have a crown age of 10.68 Ma, stem age of 12.14 Ma and crown 95 % HPD of 8.8–13.0 Ma. Age estimates across tribe Brassiceae are presented in Fig. 5, and clade age estimates are summarized in Table 1 based on the lineages used in this study (modified classifications from Arias and Pires (2012)).

DISCUSSION

Polyploidy as a clade marker

The tension between the ideal of natural taxonomic classification and the practical challenges of reconstructing complex evolutionary histories is exemplified in ancestral polyploid lineages

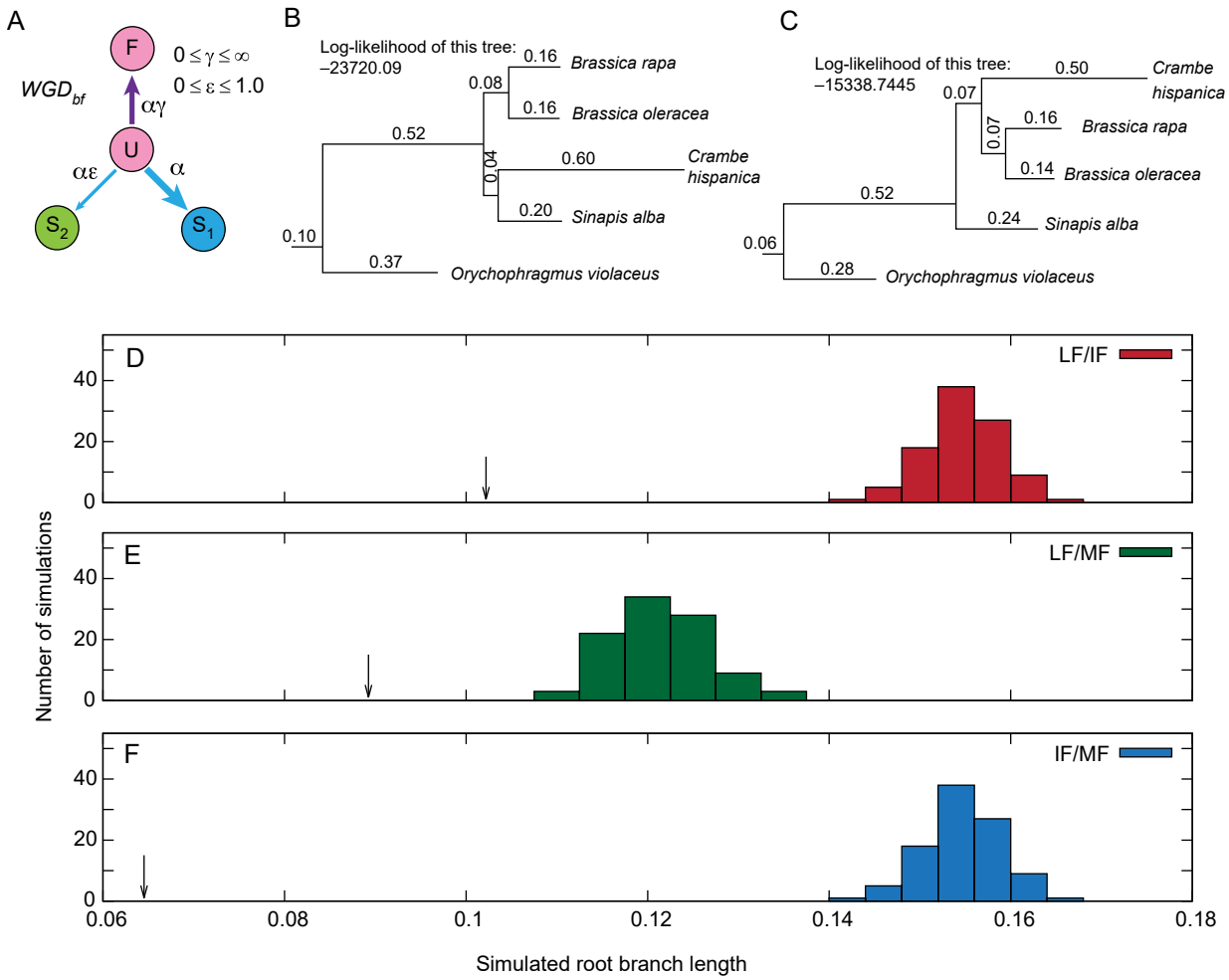

FIG. 4. No evidence for a shared WGD between *O. violaceus* and species with the Brassiceae hexaploidy. (A) A model of duplicate loss after WGD. Ohnolog pairs all start in state U (undifferentiated duplicates) immediately post-WGD. Loss of the copy from subgenome 2 moves the pair to state S_1 (only copy 1 retained) or symmetrically to S_2, which occurs less often due to biased fractionation (rate ϵ). Duplicates can move in to a fixed duplicate state F at rate γ. (B) The maximum likelihood topology fix to the pillars combining LF/IF subgenomes from the four Brassiceae species and *O. violaceus*. (C) The topology for LF/MF and IF/MF analyses (IF/MF pictured). (D–F) Distribution of the estimated root branch length under Model 2 (see Materials and methods) for the LF/IF, LF/MF and IF/MF comparisons, respectively. In each case, we simulated datasets under the assumption of no shared losses between the Brassiceae and *O. violaceus* and then fitted a model allowing such shared losses. The lengths of this branch seen in simulation always exceeded that seen for the real datasets (arrows), failing to reject the null hypothesis of independent events in the two groups.

like the tribe Brassiceae. In our study, we addressed the tribe's complex history as a step towards developing the Brassiceae as a model clade for comparative biology. Determining the placement of the ancestral mesohexaploidy, Br-α WGT, is critical for contextualizing the effects of polyploidy in the group, such as our understanding of how fractionation has shaped the A, B and C subgenome lineages of the Triangle of U (Nagaharu and Nagaharu, 1935) and other distinct chromosomal lineages in the tribe (Hoang *et al.*, 2024). By clarifying whether or not the *O. violaceus* WGD is shared as a first step of Br-α WGT or independent, we lay the groundwork for future out-group comparisons. By utilizing ancient polyploidy as a rare genomic character informing the tribal taxonomic boundary, we enable the tribal designation of Brassiceae to represent a distinct evolutionary history. Our findings will thus provide a phylogenetic definition for the Brassiceae, grounded in both evolutionary history and the unique genomic signature of the Br-α WGT polyploidization event.

Developing a robust phylogeny is the first step in placing Br-α. To achieve this, our sampling included species from the most recognized tribal lineages to represent overall diversity in the plastid and nuclear phylogenies (Fig. 1, Supplementary Data Table S1). We followed a modified version of clade boundaries previously defined by Arias and Pires (2012) and used the Orychophragmus clade (including genera *Orychophragmus* and *Sinalliaria*) as the earliest diverging lineage, totalling all nine clades in the plastid phylogeny and eight clades in the nuclear phylogeny. Nuclear and plastid trees align with previous phylogenetic studies of the group, showing evident incongruence (Fig. 2). The Crambe, Savignya, Cakile, Psychine, Vella, Zilla and Orychophragmus lineages showed incongruence between the plastid and nuclear trees, though they were found to be monophyletic. The Rapa/Oleracea and Nigra clades (as defined by Arias and Pires (2012)) were not monophyletic in the nuclear data and demonstrated a nested relationship and in this study are referred to as the Brassica clade/

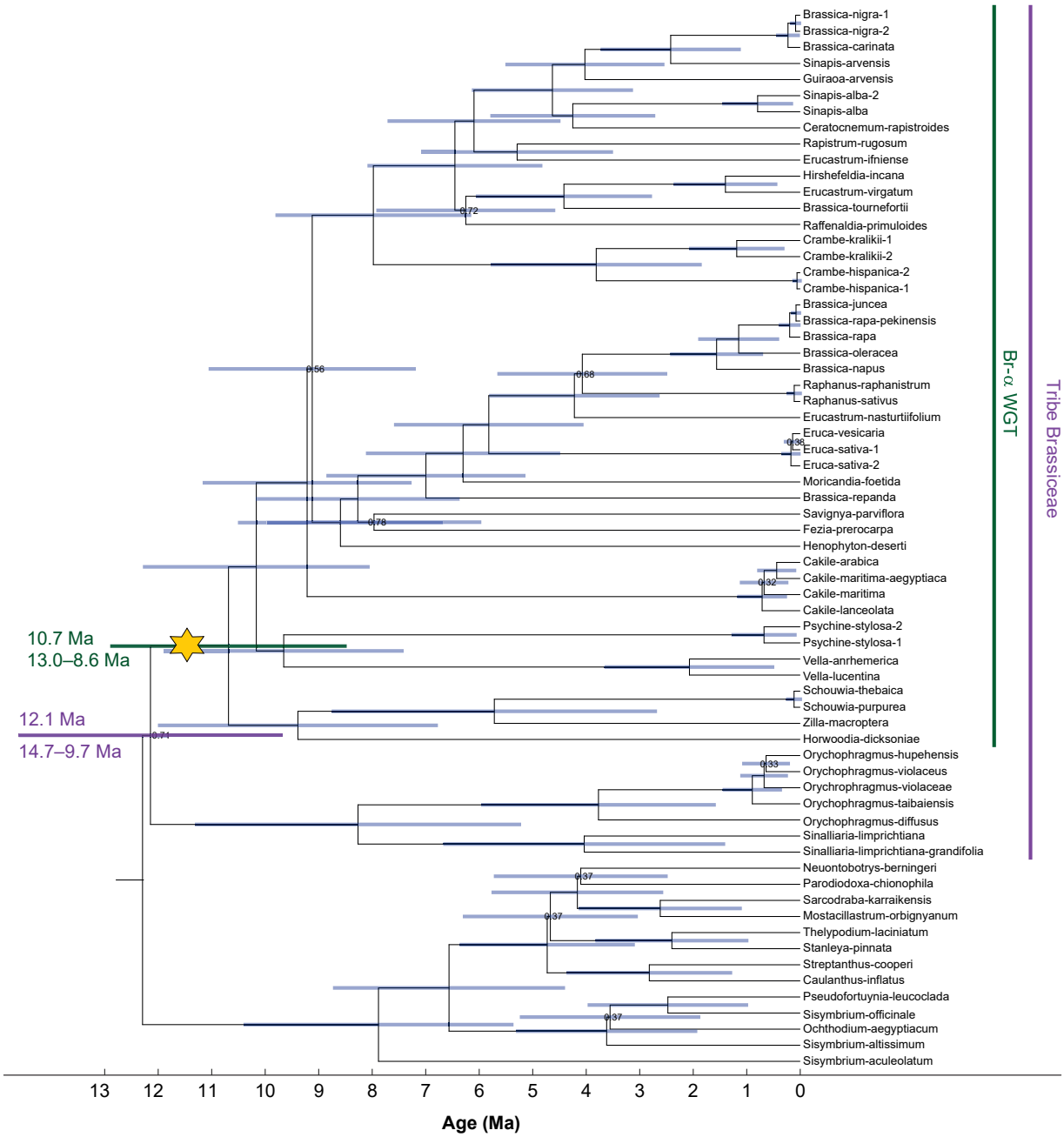

FIG. 5. Timing of the Br-α WGT. BEAST divergence time analysis of 60 plastid gene phylogeny. Phylogeny in this figure is pruned to only include one sister lineage of tribe Brassiceae. Tribe Brassiceae is indicated in purple and species within tribe Brassiceae sharing the Br-α WGT are indicated in green. Yellow star indicates the Br-α WGT. Blue bars next to nodes indicate their 95 % HPD interval for the node age. Purple and green intervals indicate the tribe Brassiceae and Br-α nodes, respectively, and the text at intervals indicates the mean age and 95 % HPD interval. Scale below the tree indicates time in million years (Ma). Support values are indicated if local posterior probabilities are <0.8.

lineage. These differences suggest that ancient gene flow and group hybridization were likely rampant, and in fact Hoang *et al*. (2024) provide strong evidence that the genomes of *Hirschfeldia incana* and *Raphanus sativus* (Brassica lineage) show signatures of hybridization or introgression from within the Brassiceae. Gene tree discordance measurements support this assumption, as calculated from the nuclear tree (Supplementary Data Fig. S3). Equal quartet frequencies across clades would be expected if incomplete lineage sorting were the only source of gene tree discordance (Pollard *et al*., 2006; Yu *et al*., 2011; Morales-Briones *et al*., 2021; Thomas *et al*., 2021). However, across the Br-α WGT clade backbone, unequal alternative quartet frequencies indicate that ancient gene flow and hybridization significantly impacted the evolutionary history of its subclades.

The Br-α WGT occurred in two steps: an initial tetraploidy followed by a later addition of a third subgenome, with the two events potentially separated by millions of years of evolution. WGT events, when localized, may appear across multiple branches of a phylogeny, so that some extant taxa have

Table 1. Divergence time estimates. Mean crown/stem ages and crown 95 % HPD interval for major clades in tribe Brassiceae.

Clade	Crown age (Ma)	Stem age (Ma)	Crown 95 % HPD (Ma)
Brassiceae	12.1395	12.2876	9.7–14.7
Br-α WGT	10.681	12.1395	8.8–13.0
Orychophragmus	8.2604	12.1395	5.2–11.3
Zilla	9.3837	10.681	6.7–11.9
Psychine	0.6717	9.6488	0.15–1.4
Vella	2.0657	9.6488	0.7–3.8
Cakile	0.7071	9.2113	0.2–1.2
Savignya	7.9614	8.2639	6.0–10.0
Rapa/Oleracea	6.9909	8.2639	5.3–8.9
Henophyton	NA	8.5887	NA
Crambe	3.808	7.9737	1.9–5.9
Nigra	6.4492	7.9737	4.9–8.1

experienced only the first tetraploidy event; for example, *Gynandropsis gynandra* (Cleomaceae) experienced only part of the Th-α WGT (Hoang *et al.*, 2023). Using a syntenic comparison of tribal genomes (Fig. 4), we tested whether the *O. violaceus* tetraploidy was one step in the Br-α WGT, but we found that it was likely a separate event, similar to the findings of Zhang *et al.* (2023). Visual assessments of the K_s distributions supported this conclusion, showing that the Br-α WGT clade shares a peak at ∼0.3, which is indicative of the known mesohexaploidy in crop *Brassica* species (Qi *et al.*, 2021). Statistical assessments show that simulated K_s distributions of lineages within the Br-α WGT clade cannot be differentiated, but the simulated clade distribution of the Br-α WGT clade and lineages within it are statistically different from both the Orychophragmus lineage and the out-group simulated distributions. These analyses provide strong evidence that the entire Br-α WGT is phylogenetically localized to the branch leading to the Br-α WGT clade. (Fig. 2, Supplementary Data Figs S4 and S6). If two polyploid events occurred close enough in time, then distinguishing them may not be possible because the K_s values between pairwise paralogues would be too close to distinguish multiple peaks. Patterns of unique shared duplications on branches in the nuclear tree from the PUG analyses also support the placement of the Br-α WGT (Fig. 3, Supplementary Data Fig. S4), but shared duplications in the backbone of the tree (Supplementary Data Fig. S4) may suggest complex patterns of hybridization in ancestral lineages. Given our sampling and the monophyletic nature of most Brassiceae clades, it is unlikely that an extant lineage sharing only part of the Br-α WGT exists, though *Succowia balearica* and *Bivonaea lutea* have been shown in recent studies (Hendriks *et al.*, 2023) to occupy key phylogenetic positions between the Orychophragmus lineage and the Br-α WGT clade that make them potential candidates for an intermediate lineage.

Timing of the Br-α WGT

The length of time between events and the lack of an extant intermediate lineage could be a product of the environmental context during the geological timing of the events as much as the randomness of evolutionary processes. To date the Br-α WGT, we performed divergence time analyses and provided evidence for the date of clades based on plastid sequence data and fossil records. To avoid using contentious fossil calibrations within the Brassicales and Brassicaceae, this study utilized robust fossil calibrations across the rosids (Bell *et al.*, 2010; Njuguna *et al.*, 2013; Walden *et al.*, 2020). Although the WGT is shared by other tribal lineages (Lysak *et al.*, 2005, 2007), most calculated dates for the Br-α WGT have been inferred genomically from only Brassica lineage genomes. To infer a well-supported and dated phylogeny (Fig. 5, Supplementary Data Fig. S1) across the rosids, with an emphasis on sampling of tribe Brassiceae taxa, this study used fossil data, along with both publicly available and newly generated plastid sequence data. The crown age at which tribe Brassiceae diverged from the rest of the family Brassicaceae was around 12.1 Ma (95 % HPD 9.7–14.7 Ma), and the Br-α WGT clade diverged from the Orychophragmus clade around 10.7 Ma (95 % HPD 8.8–13.0 Ma) (Fig. 5). Given this range, the two-step hexaploidy of the Br-α WGT would have likely occurred between 12.1 and 10.7 Ma during the mid-to-late Miocene.

Although previous divergence estimates for the tribe Brassiceae ranged from 6 to 24 Ma (Table 1), recent estimates have converged around ∼12 Ma (Walden *et al.*, 2020; Hendriks *et al.*, 2023). Hendriks *et al.* (2023) produced the most recent time-calibrated phylogeny of the mustard family, estimating ages using both nuclear and plastid data. While their nuclear data estimated a crown age of 12.8 Ma for the tribe Brassiceae, aligning with our estimate of 12.1 Ma, their plastid data yielded a much younger age of 6.5 Ma. Our methodological choices likely explain the differences between our plastid tree estimates. Studies comparing treePL and BEAST have demonstrated that treePL produces both younger and older age estimates, in contrast to BEAST (Magallón *et al.*, 2015 ; Cai *et al.*, 2016). However, the agreement between our BEAST-generated plastid tree and the Hendriks *et al.* (2023) treePL nuclear tree estimate emphasizes the growing consensus on the age of the tribe.

Polyploidy events occur more often during periods of environmental stress and are hypothesized to enhance species' adaptability during such periods (Van de Peer *et al.*, 2021; Tossi *et al.*, 2022; Thomas *et al.*, 2023). This study's findings indicate that the Br-α WGT occurred concurrently with the long-term cooling and aridification of Eurasia during the Mid-Late Miocene epoch, specifically spanning the end of the Serravallian age (13.82–11.63 Ma) and the beginning of the Tortonian age (11.63–7.246 Ma) (Hamon *et al.*, 2013; Quan *et al.*, 2014). Globally, decreasing carbon dioxide levels during this period caused the major extinction event known as the Middle Miocene disruption (Pearson and Palmer, 2000; Torfstein and Steinberg, 2020). Aridification and decreasing carbon dioxide levels during this period drove the large expansion of C4 grasses and rapid species radiations in CAM lineages (Sage *et al.*, 2018, 2023), and likely influenced the evolutionary trajectory of the Brassiceae lineages. The Brassiceae tribe is notable for maintaining several xeric-adapted lineages, its unique inclusion of several C3–C4 intermediate species within the Brassicaceae (Guerreiro *et al.*, 2023), and for serving as a model for the recurrent evolution of phylogenetically derived woodiness as a drought adaptation (Zizka *et al.*, 2022; Mabry

et al., 2024). Post-polyploidy changes in gene networks may have enabled this lineage to survive extinction during shifting climatic conditions and promoted the exaptations that support the recurrent evolution of arid-adapted traits in the tribe today.

Taxonomic implications

The tribal placement of *Orychophragmus* (seven species) and its sister genus *Sinalliaria* (two species) within the Brassiceae remains controversial. Schulz (1923) first included the genus in the tribe, mainly because of its conduplicate cotyledons (Fig. 5), a character shared with most species sharing the Br-α WGT and not found anywhere else in the Brassicaceae family. Some authors subsequently maintained its inclusion (e.g. Gomez-Campo, 1999; Warwick and Sauder, 2005; Al-Shehbaz, 2012), while others excluded it (Gomez-Campo, 1980). Unlike the rest of the tribe Brassiceae, which shares the Br-α WGT, (Lysak *et al.*, 2005, 2007; Zhang *et al.*, 2023), this paper shows that *O. violaceus* has an independent WGD. This difference provides solid evidence supporting the exclusion of the genus from the tribe. Moreover, the Orychophragmus/Sinalliaria clade is geographically restricted to eastern China, with *O. violaceus* also native to Korea; no other native Brassiceae species occur in eastern Asia. The genera are morphologically distinguished by the absence of segmented fruits seen in many other members of the Brassiceae (Supplementary Data Table S1). Additionally, most species in these genera feature long petiolate stem leaves, characterized by either a cordate blade or a terminal lobe on a divided leaf. Therefore, morphological, geographical and molecular evidence supports placing *Orychophragmus* and *Sinalliaria* in a new tribe, Orychophragmeae, which is formally validated below. In fact, Khosravi *et al.* (2009) first suggested placing *Orychophragmus* and *Conringia planisiliqua* (currently in the Isatideae) as one tribe, though no nomenclatural adjustments were proposed.

The genera *Orychophragmus* (seven species) and *Sinalliaria* (two species) are herein formally recognized as the tribe Orychophragmeae. This adds to the recently updated taxonomic framework of the Brassicaceae family (German *et al.*, 2023), in total now with 59 recognized tribes in the family.

Orychophragmeae Shawn K. Thomas, R. Shawn Abrahams, J.C. Pires & Al-Shehbaz, trib. nov.

Type: *Orychophragmus* Bunge

Description: Herbs annuals or biennials. Trichomes simple or absent. Multicellular glands absent. Cauline leaves petiolate to subsessile, cordate to pinnatisect with subcordate terminal lobe, auriculate or not. Racemes ebracteate, elongated in fruit. Flowers actinomorphic; sepals ascending to spreading, base of lateral pair not saccate; petals purple, lavender, or white; claw differentiated from blade or not; filaments dilated at base; pollen 3-colpate; ovules 20–70 per ovary. Fruits dehiscent siliques, linear, terete, unsegmented; septum complete; stigma capitate, 2-lobed. Seeds uniseriate; cotyledons conduplicate.

Distribution: *Orychophragmus* is widespread in China (Anhui, Beijing, Gansu, Hebei, Henan, Hubei, Hunan, Jiangsu, Nei Mongol, Shandong, Shanxi, Sichuan, Zhejiang) and Korea; it is naturalized in Japan. However, *Sinalliaria* is restricted to the Chinese provinces Anhui and Zhejiang.

Notes: *Sinalliaria* is easily separated from *Orychophragmus* by being rhizomatous perennial, with non-rosulate basal leaves, non-auriculate cauline leaves, distinctly 3-veined fruit valves, and 20–30 verrucose seeds. By contrast, *Orychophragmus* is non-rhizomatous biennial, with rosulate basal leaves, auriculate or non-auriculate cauline leaves, obscurely to distinctly 1-veined fruit valves, and 20–70 alveolate-reticulate seeds.

Conclusions

This study phylogenetically localizes and dates the Br-α mesohexaploidy. An independent polyploidy event in *O. violaceus*, differences in morphology and biogeography, and the results from this study support placing the Orychophragmus lineage (including both the genera *Orychophragmus* and *Sinalliaria*) in its own tribe called the Orychophragmeae. The emergence of the Br-α WGT, dated around 12.1–10.7 Ma, coincides with a major extinction event and the rise of C4 grasses in the middle Miocene, which may indicate that the Br-α WGT provided an evolutionary advantage to ancestral Br-α taxa during this stressful time. This study has illuminated aspects of the Br-α WGT, but several key questions remain to be answered. Because this study did not sample *Succowia balearica* and *Bivonaea lutea*, whether they share the Br-α WGT remains unclear. Further research incorporating these taxa and leveraging whole genomes are necessary to better resolve the evolutionary history of this economically significant group.

SUPPLEMENTARY DATA

Supplementary data are available at *Annals of Botany* online and consist of the following. Figure S1: full plastid phylogeny. Figure S2: full nuclear phylogeny. Figure S3: gene tree discordance across tribe Brassiceae and out-groups. Figure S4: full nuclear phylogeny with unique gene duplications mapped to branches. Figure S5: K_s frequency plots for all transcriptome data. Figure S6: modified version of Fig. 3 with the full nuclear tree and all K_s plots. Figure S7: *Sinalliaria* vs Br-α WGT clade. Figure S8: comparison of the simulated lineages based on the modified Arias and Pires (2012) clades used in this study (black) and the simulated in-group (green) ECDFs with the K–S test. Figure S9: comparison of the simulated lineages based on the modified Arias and Pires (2012) Brassiceae clades used in this study against each other (red or black) ECDFs with the K–S test. Figure S10: comparison of the simulated lineages based on the modified Arias and Pires (2012) clades used in this study (black) and out-group (purple) ECDFs with the K–S test. Table S1: overview of Brassiceae genera, morphology and WGT presence/absence. Table S2: plastid tree sampling list. Table S3: transcriptome sampling list. Table S4: fossil calibrations list. Table S5: BUSCO scores for transcriptomes assembled in this study. Table S6: BUSCO scores for pre-existing, publicly available transcriptomes. Table S7: Transrate software statistics for transcriptomes assembled in this study. Table S8: Transrate software transcriptomes pre-existing, publicly available transcriptomes. Table S9: PUG duplication counts per node at 80 and 50 % bootstrap support thresholds.

FUNDING

We thank our funding sources, including the National Science Foundation (NSF) grant DBI-1339156 to J.C.P. and G.C., and USDA-ARS. The computational aspect of this work was done using the high-performance computing infrastructure provided by Research Computing Support Services and funded in part by the National Science Foundation under grant number CNS-1429294 at the University of Missouri, Columbia MO (MU).

ACKNOWLEDGEMENTS

We thank the Department of Energy Joint Genome Institute and collaborators for prepublication access to the Brassicales Map Alignment Project (BMAP) sequence data. The work (proposal 10.46936/10.25585/60000980) conducted by the US Department of Energy Joint Genome Institute (https://ror.org/04xm1d337), a DOE Office of Science User Facility, is supported by the Office of Science of the US Department of Energy operated under Contract No. DE-AC02-05CH11231.

AUTHOR CONTRIBUTIONS

S.K.T., J.C.P. and J.D.W. designed the project. R.S.A. and J.D.W. generated new sequence data for this study. S.K.T. and N.W. acquired, organized and curated publicly available sequence data. S.K.T., D.R.K., N.W., G.C. and M.R.M. performed data analyses and created visualizations. The first draft of the manuscript was written by S.K.T., R.S.A., D.R.K., N.W., G.C., I.A.A. and J.C.P. All authors read, revised and approved the final version of the manuscript.

DATA AVAILABILITY

Raw sequence reads are available in the NCBI SRA BioProject PRJNA1254719 (https://www.ncbi.nlm.nih.gov/bioproject/PRJNA1254719). Scripts, intermediate data and trees are available at https://github.com/danielkick/brassica_ks and/or https://github.com/shawnkt/BrassiceaeMesohexaploidy. Specific data sources for publicly available data can be found in Supplementary Data Tables S2 and S3.

REFERENCES

Abrahams RS, **Pires JC**, **Schranz ME. 2020**. Genomic origin and diversification of the glucosinolate MAM locus. *Frontiers in Plant Science* **11**: 711.

Al-Shehbaz IA. 2012. A generic and tribal synopsis of the Brassicaceae (Cruciferae). *Taxon* **61**: 931–954.

Altschul SF, **Gish W**, **Miller W**, **Myers EW**, **Lipman DJ. 1990**. Basic local alignment search tool. *Journal of Molecular Biology* **215**: 403–410.

Altschul SF, **Madden TL**, **Schäffer AA**, ***et al.* 1997**. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Research* **25**: 3389–3402.

An H, **Qi X**, **Gaynor ML**, ***et al.* 2019**. Transcriptome and organellar sequencing highlights the complex origin and diversification of allotetraploid *Brassica napus*. *Nature Communications* **10**: 2878.

Arias T, **Beilstein MA**, **Tang M**, **McKain MR**, **Pires JC. 2014**. Diversification times among *Brassica* (Brassicaceae) crops suggest hybrid formation after 20 million years of divergence. *American Journal of Botany* **101**: 86–91.

Arias T, **Pires JC. 2012**. A fully resolved chloroplast phylogeny of the brassica crops and wild relatives (Brassicaceae: Brassiceae): novel clades and potential taxonomic implications. *Taxon* **61**: 980–988.

Azibi T, **Hadj-Arab H**, **Lodé M**, ***et al.* 2020**. Impact of whole genome triplication on the evolutionary history and the functional dynamics of regulatory genes involved in *Brassica* self-incompatibility signalling pathway. *Plant Reproduction* **33**: 43–58.

Babula D, **Kaczmarek M**, **Barakat A**, **Delseny M**, **Quiros CF**, **Sadowski J. 2003**. Chromosomal mapping of *Brassica oleracea* based on ESTs from *Arabidopsis thaliana*: complexity of the comparative map. *Molecular Genetics and Genomics* **268**: 656–665.

Bayat S, **Schranz ME**, **Roalson EH**, **Hall JC. 2018**. Lessons from Cleomaceae, the sister of crucifers. *Trends in Plant Science* **23**: 808–821.

Bell CD, **Soltis DE**, **Soltis PS. 2010**. The age and diversification of the angiosperms re-revisited. *American Journal of Botany* **97**: 1296–1303.

Benedict JC, **DeVore ML**, **Pigg KB. 2011**. *Prunus* and *Oemleria* (Rosaceae) flowers from the late early Eocene republic flora of northeastern Washington state, U.S.A. *International Journal of Plant Sciences* **172**: 948–958.

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

Booker WW, **Gerhardt HC**, **Lemmon AR**, ***et al.* 2022**. The complex history of genome duplication and hybridization in North American gray treefrogs. *Molecular Biology and Evolution* **39**: msab316.

Bouckaert R, **Vaughan TG**, **Barido-Sottani J**, ***et al.* 2019**. BEAST 2.5: an advanced software platform for Bayesian evolutionary analysis. *PLoS Computational Biology* **15**: e1006650.

Cai L, **Xi Z**, **Peterson K**, **Rushworth C**, **Beaulieu J**, **Davis CC. 2016**. Phylogeny of Elatinaceae and the tropical Gondwanan origin of the Centroplacaceae (Malpighiaceae, Elatinaceae) clade. *PLoS One* **11**: e0161881.

Chen S, **Zhou Y**, **Chen Y**, **Gu J. 2018**. fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* **34**: i884–i890.

Cheng F, **Mandáková T**, **Wu J**, **Xie Q**, **Lysak MA**, **Wang X. 2013**. Deciphering the diploid ancestral genome of the mesohexaploid *Brassica rapa*. *Plant Cell* **25**: 1541–1554.

Cheng F, **Wu J**, **Wang X. 2014**. Genome triplication drove the diversification of *Brassica* plants. *Horticulture Research* **1**: 14024.

Clark JW, **Donoghue PCJ. 2018**. Whole-genome duplication and plant macroevolution. *Trends in Plant Science* **23**: 933–945.

Conant GC, **Wagner A. 2002**. GenomeHistory: a software tool and its application to fully sequenced genomes. *Nucleic Acids Research* **30**: 3378–3386.

Conant GC, **Wolfe KH. 2008**. Probabilistic cross-species inference of orthologous genomic regions created by whole-genome duplication in yeast. *Genetics* **179**: 1681–1692.

Crespo M. 2000. Subtribe Vellinae (Brassiceae, Brassicaceae): a combined analysis of ITS nrDNA sequences and morphological data. *Annals of Botany* **86**: 53–62.

Dassanayake M, **Oh D-H**, **Haas JS**, ***et al.* 2011**. The genome of the extremophile crucifer *Thellungiella parvula*. *Nature Genetics* **43**: 913–918.

Doolittle WF. 1999. Phylogenetic classification and the universal tree. *Science* **284**: 2124–2129.

Doyle JJ, **Doyle JL. 1987**. A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochemical Bulletin* **19**: 11–15.

Drummond AJ, **Ho SYW**, **Phillips MJ**, **Rambaut A. 2006**. Relaxed phylogenetics and dating with confidence. *PLoS Biology* **4**: e88.

Edgar RC. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research* **32**: 1792–1797.

Emms DM, **Kelly S. 2019**. OrthoFinder: phylogenetic orthology inference for comparative genomics. *Genome Biology* **20**: 238.

Fu L, **Niu B**, **Zhu Z**, **Wu S**, **Li W. 2012**. CD-HIT: accelerated for clustering the next-generation sequencing data. *Bioinformatics* **28**: 3150–3152.

German DA, **Hendriks KP**, **Koch MA**, ***et al.* 2023**. An updated classification of the Brassicaceae (Cruciferae). *PhytoKeys* **220**: 127–144.

Gernhard T. 2008. The conditioned reconstructed process. *Journal of Theoretical Biology* **253**: 769–778.

Gomez-Campo C. 1980. Morphology and morpho-taxonomy of the tribe Brassiceae. In: **Tsunoda S**, **Hinata K**, **Gómez-Campo C.** eds. *Brassica crops and wild allies: biology and breeding*. Tokyo: Japan Scientific Societies Press, 3–31.

Gomez-Campo C. ed. 1999. *Biology of Brassica coenospecies. Vol. 4*. London: Elsevier Science.

Grabherr MG, **Haas BJ**, **Yassour M**, ***et al.*** **2011**. Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nature Biotechnology* **29**: 644–652.

Guerreiro R, **Bonthala VS**, **Schlüter U**, ***et al.*** **2023**. A genomic panel for studying C3-C4 intermediate photosynthesis in the Brassiceae tribe. *Plant, Cell & Environment* **46**: 3611–3627.

Guo X, **Mandáková T**, **Trachtová K**, **Özüdo?ru B**, **Liu J**, **Lysak MA. 2021**. Linked by ancestral bonds: multiple whole-genome duplications and reticulate evolution in a Brassicaceae tribe. *Molecular Biology and Evolution* **38**: 1695–1714.

Hämälä T, **Moore C**, **Cowan L**, ***et al.*** **2024**. Impact of whole-genome duplications on structural variant evolution in *Cochlearia*. *Nature Communications* **15**: 5377.

Hamon N, **Sepulchre P**, **Lefebvre V**, **Ramstein G. 2013**. The role of eastern Tethys seaway closure in the Middle Miocene Climatic Transition (ca. 14 Ma). *Climate of the Past* **9**: 2687–2702.

Hao Y, **Conant GC. 2022**. POInT: a tool for modeling ancient polyploidies using multiple polyploid genomes. *Methods in Molecular Biology* **2512**: 81–91.

Hao Y, **Mabry ME**, **Edger PP**, ***et al.*** **2021**. The contributions of the allopolyploid parents of the mesopolyploid Brassiceae are evolutionarily distinct but functionally compatible. *Genome Research* **31**: 799–810.

Haudry A, **Platts AE**, **Vello E**, ***et al.*** **2013**. An atlas of over 90,000 conserved noncoding sequences provides insight into crucifer regulatory regions. *Nature Genetics* **45**: 891–898.

He Z, **Ji R**, **Havlickova L**, ***et al.*** **2021**. Genome structural evolution in *Brassica* crops. *Nature Plants* **7**: 757–765.

Hendriks KP, **Kiefer C**, **Al-Shehbaz IA**, ***et al.*** **2023**. Global Brassicaceae phylogeny based on filtering of 1,000-gene dataset. *Current Biology* **33**: 4052–4068.e6.

Hoang NV, **Sogbohossou EOD**, **Xiong W**, ***et al.*** **2023**. The *Gynandropsis gynandra* genome provides insights into whole-genome duplications and the evolution of C4 photosynthesis in Cleomaceae. *Plant Cell* **35**: 1334–1359.

Hoang NV, **Walden N**, **Caracciolo L**, ***et al.*** **2024**. Expanding the triangle of U: comparative analysis of the *Hirschfeldia incana* genome provides insights into chromosomal evolution, phylogenomics and high photosynthesis-related traits. *Annals of Botany*. mcae179. doi:10.1093/aob/mcae179.

Hohmann N, **Wolf EM**, **Lysak MA**, **Koch MA. 2015**. A time-calibrated road map of Brassicaceae species radiation and evolutionary history. *The Plant Cell*. 27: 2770–2784. doi:. 10.1105/tpc.15.00482

Hu H, **Hu Q**, **Al-Shehbaz IA**, ***et al.*** **2016**. Species delimitation and interspecific relationships of the genus *Orychophragmus* (Brassicaceae) inferred from whole chloroplast genomes. *Frontiers in Plant Science* **7**: 1826.

Huang F, **Chen P**, **Tang X**, ***et al.*** **2023**. Genome assembly of the Brassicaceae diploid *Orychophragmus violaceus* reveals complex whole-genome duplication and evolution of dihydroxy fatty acid metabolism. *Plant Communications* **4**: 100432.

Huang X-C, **German DA**, **Koch MA. 2020**. Temporal patterns of diversification in Brassicaceae demonstrate decoupling of rate shifts and mesopolyploidization events. *Annals of Botany* **125**: 29–47.

Iltis HH, **Hall JC**, **Cochrane TS**, **Sytsma KJ. 2011**. Studies in the Cleomaceae I. On the separate recognition of Capparaceae, Cleomaceae, and Brassicaceae. *Annals of the Missouri Botanical Garden* **98**: 28–36.

Johnson MG, **Gardner EM**, **Liu Y**, ***et al.*** **2016**. HybPiper: extracting coding sequence and introns for phylogenetics from high-throughput sequencing reads using target enrichment. *Applications in Plant Sciences* **4**: 1600016.

Katoh K, **Standley DM. 2013**. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Molecular Biology and Evolution* **30**: 772–780.

Khosravi AR, **Mohsenzadeh S**, **Mummenhoff K. 2009**. Phylogenetic relationships of Old World Brassicaceae from Iran based on nuclear ribosomal DNA sequences. *Biochemical Systematics and Ecology* **37**: 106–115.

Knobloch E, **Mai D. 1986**. Monograph of the fruits and seeds in the Cretaceous of Central Europe. *Rozpravy Ústreõdního Ústavu Geologickéh* **47**: 1–219.

Koch MA. 2012. Mid-Miocene divergence of *Ionopsidium* and *Cochlearia* and its impact on the systematics and biogeography of the tribe Cochlearieae (Brassicaceae). *Taxon* **61**: 76–92.

Kuzmin E, **VanderSluis B**, **Nguyen Ba AN**, ***et al.*** **2020**. Exploring whole-genome duplicate gene retention with complex genetic interaction analysis. *Science* **368**: eaaz5667.

Lagercrantz U. 1998. Comparative mapping between *Arabidopsis thaliana* and *Brassica nigra* indicates that *Brassica* genomes have evolved through extensive genome replication accompanied by chromosome fusions and frequent rearrangements. *Genetics* **150**: 1217–1228.

Lagercrantz U, **Lydiatet DJ. 1996**. Comparative genome mapping in *Brassica*. *Genetics* **144**: 1903–1910.

Lamesch P, **Berardini TZ**, **Li D**, ***et al.*** **2012**. The Arabidopsis Information Resource (TAIR): improved gene annotation and new tools. *Nucleic Acids Research* **40**: D1202–D1210.

Landis JB, **Soltis DE**, **Li Z**, ***et al.*** **2018**. Impact of whole-genome duplication events on diversification rates in angiosperms. *American Journal of Botany* **105**: 348–363.

Larsson A. 2014. AliView: a fast and lightweight alignment viewer and editor for large datasets. *Bioinformatics* **30**: 3276–3278.

Li H, **Durbin R. 2009**. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* **25**: 1754–1760.

Liu S, **Liu Y**, **Yang X**, ***et al.*** **2014**. The *Brassica oleracea* genome reveals the asymmetrical evolution of polyploid genomes. *Nature Communications* **5**: 3930.

Lou P, **Woody S**, **Greenham K**, ***et al.*** **2020**. Genetic and genomic resources to study natural variation in *Brassica rapa*. *Plant Direct* **4**: e00285. doi:10.1002/pld3.285

Lysak MA, **Cheung K**, **Kitschke M**, **Bures P. 2007**. Ancestral chromosomal blocks are triplicated in Brassiceae species with varying chromosome number and genome size. *Plant Physiology* **145**: 402–410.

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

Mabry ME, **Abrahams RS**, **Al-Shehbaz IA**, ***et al.*** **2024**. Complementing model species with model clades. *Plant Cell* **36**: 1205–1226.

Mabry ME, **Brose JM**, **Blischak PD**, ***et al.*** **2020**. Phylogeny and multiple independent whole-genome duplication events in the Brassicales. *American Journal of Botany* **107**: 1148–1164.

Magallón S, **Gómez-Acevedo S**, **Sánchez-Reyes LL**, **Hernández-Hernández T. 2015**. A metacalibrated time-tree documents the early rise of flowering plant phylogenetic diversity. *New Phytologist* **207**: 437–453.

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

Mandáková T, **Lysak MA. 2018**. Post-polyploid diploidization and diversification through dysploid changes. *Current Opinion in Plant Biology* **42**: 55–65.

Manni M, **Berkeley MR**, **Seppey M**, **Simão FA**, **Zdobnov EM. 2021**. BUSCO update: novel and streamlined workflows along with broader and deeper phylogenetic coverage for scoring of eukaryotic, prokaryotic, and viral genomes. *Molecular Biology and Evolution* **38**: 4647–4654.

McKain MR, **Tang H**, **McNeal JR**, ***et al.*** **2016**. A phylogenomic assessment of ancient polyploidy and genome evolution across the Poales. *Genome Biology and Evolution* **8**: 1150–1164.

McKain MR, **Wickett N**, **Zhang Y**, ***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.

Minh BQ, **Schmidt HA**, **Chernomor O**, ***et al.*** **2020**. IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era. *Molecular Biology and Evolution* **37**: 1530–1534.

Moore MJ, **Soltis PS**, **Bell CD**, **Burleigh JG**, **Soltis DE. 2010**. Phylogenetic analysis of 83 plastid genes further resolves the early diversification of eudicots. *Proceedings of the National Academy of Sciences of the United States of America* **107**: 4623–4628.

Morales-Briones DF, **Kadereit G**, **Tefarikis DT**, ***et al.*** **2021**. Disentangling sources of gene tree discordance in phylogenomic data sets: testing ancient hybridizations in Amaranthaceae s.l. *Systematic Biology* **70**: 219–235.

Nagaharu U, **Nagaharu N. 1935**. Genome analysis in *Brassica* with special reference to the experimental formation of *B. napus* and peculiar mode of fertilization. *Journal of Japanese Botany* **7**: 389–452.

Nieto Feliner G, **Casacuberta J**, **Wendel JF. 2020**. Genomics of evolutionary novelty in hybrids and polyploids. *Frontiers in Genetics* **11**: 792.

Njuguna W, **Liston A**, **Cronn R**, **Ashman T-L**, **Bassil N. 2013**. Insights into phylogeny, sex function and age of *Fragaria* based on whole chloroplast genome sequencing. *Molecular Phylogenetics and Evolution* **66**: 17–29.

Oh D-H, **Hong H**, **Lee SY**, **Yun D-J**, **Bohnert HJ**, **Dassanayake M. 2014**. Genome structures and transcriptomes signify niche adaptation for the multiple-ion-tolerant extremophyte *Schrenkiella parvula*. *Plant Physiology* **164**: 2123–2138.

One Thousand Plant Transcriptomes Initiative. 2019. One thousand plant transcriptomes and the phylogenomics of green plants. *Nature* **574**: 679–685.

Paradis E, **Schliep K. 2019**. Ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. *Bioinformatics* **35**: 526–528.
Parkin IAP, **Gulden SM**, **Sharpe AG**, ***et al.* 2005**. Segmental structure of the *Brassica napus* genome based on comparative analysis with *Arabidopsis thaliana*. *Genetics* **171**: 765–781.
Parma DF, **Vaz MGMV**, **Falquetto P**, ***et al.* 2021**. New insights into the evolution of C4 photosynthesis offered by the *Tarenaya* cluster of Cleomaceae. *Frontiers in Plant Science* **12**: 756505.
Pearson PN, **Palmer MR. 2000**. Atmospheric carbon dioxide concentrations over the past 60 million years. *Nature* **406**: 695–699.
Perumal S, **Koh CS**, **Jin L**, ***et al.* 2020**. A high-contiguity *Brassica nigra* genome localizes active centromeres and defines the ancestral *Brassica* genome. *Nature Plants* **6**: 929–941.
Pollard DA, **Iyer VN**, **Moses AM**, **Eisen MB. 2006**. Widespread discordance of gene trees with species tree in *Drosophila*: evidence for incomplete lineage sorting. *PLoS Genetics* **2**: e173.
Porturas LD, **Anneberg TJ**, **Curé AE**, **Wang S**, **Althoff DM**, **Segraves KA. 2019**. A meta-analysis of whole genome duplication and the effects on flowering traits in plants. *American Journal of Botany* **106**: 469–476.
Prjibelski A, **Antipov D**, **Meleshko D**, **Lapidus A**, **Korobeynikov A. 2020**. Using SPAdes De Novo assembler. *Current Protocols in Bioinformatics* **70**: e102.
Qi X, **An H**, **Hall TE**, ***et al.* 2021**. Genes derived from ancient polyploidy have higher genetic diversity and are associated with domestication in *Brassica rapa*. *New Phytologist* **230**: 372–386.
Qi X, **An H**, **Ragsdale AP**, ***et al.* 2017**. Genomic inferences of domestication events are corroborated by written records in *Brassica rapa*. *Molecular Ecology* **26**: 3373–3388.
Quan C, **Liu Y-SC**, **Tang H**, **Utescher T. 2014**. Miocene shift of European atmospheric circulation from trade wind to westerlies. *Scientific Reports* **4**: 5660.
Rambaut A, **Drummond AJ**, **Xie D**, **Baele G**, **Suchard MA. 2018**. Posterior summarization in Bayesian phylogenetics using tracer 1.7. *Systematic Biology* **67**: 901–904.
Renner SS. 2016. A return to Linnaeus's focus on diagnosis, not description: the use of DNA characters in the formal naming of species. *Systematic Biology* **65**: 1085–1095.
Rokas A, **Holland PW. 2000**. Rare genomic changes as a tool for phylogenetics. *Trends in Ecology & Evolution* **15**: 454–459.
Sage RF, **Gilman IS**, **Smith JAC**, **Silvera K**, **Edwards EJ. 2023**. Atmospheric CO_2 decline and the timing of CAM plant evolution. *Annals of Botany* **132**: 753–770.
Sage RF, **Monson RK**, **Ehleringer JR**, **Adachi S**, **Pearcy RW. 2018**. Some like it hot: the physiological ecology of C4 plant evolution. *Oecologia* **187**: 941–966.
Schranz ME, **Mohammadin S**, **Edger PP. 2012**. Ancient whole genome duplications, novelty and diversification: the WGD Radiation Lag-Time Model. *Current Opinion in Plant Biology* **15**: 147–153.
Schulz OE. 1923. Cruciferae–Brassiceae. In: **Engler A.** ed. *Pflanzenreich IV. 105 (Heft 84)*. Leipzig: von Wilhelm Engelmann, 1–100.
Simpson MG. 2020. *Plant systematics*. San Diego: Academic Press.
Sims HJ, **Herendeen PS**, **Lupia R**, **Christopher RA**, **Crane PR. 1999**. Fossil flowers with Normapolles pollen from the Upper Cretaceous of southeastern North America. *Review of Palaeobotany and Palynology* **106**: 131–151.
Slater GSC, **Birney E. 2005**. Automated generation of heuristics for biological sequence comparison. *BMC Bioinformatics* **6**: 31.
Smith-Unna R, **Boursnell C**, **Patro R**, **Hibberd JM**, **Kelly S. 2016**. TransRate: reference-free quality assessment of de novo transcriptome assemblies. *Genome Research* **26**: 1134–1144.
Soltis DE, **Albert VA**, **Leebens-Mack J**, ***et al.* 2009**. Polyploidy and angiosperm diversification. *American Journal of Botany* **96**: 336–348.
Soltis PS, **Soltis DE. 2016**. Ancient WGD events as drivers of key innovations in angiosperms. *Current Opinion in Plant Biology* **30**: 159–165.
Stamatakis A. 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30**: 1312–1313.
Suyama M, **Torrents D**, **Bork P. 2006**. PAL2NAL: robust conversion of protein sequence alignments into the corresponding codon alignments. *Nucleic Acids Research* **34**: W609–W612.
Tabatabaee Y, **Zhang C**, **Warnow T**, **Mirarab S. 2023**. Phylogenomic branch length estimation using quartets. *Bioinformatics* **39**: i185–i193.
Takahashi M, **Crane PR**, **Ando H. 1999**. Fossil flowers and associated plant fossils from the Kamikitaba locality (Ashizawa Formation, Futaba Group, Lower Coniacian, Upper Cretaceous) of northeast Japan. *Journal of Plant Research* **112**: 187–206.
Tang H, **Lyons E. 2012**. Unleashing the genome of *Brassica rapa*. *Frontiers in Plant Science* **3**: 172.
Tang H, **Woodhouse MR**, **Cheng F**, ***et al.* 2012**. Altered patterns of fractionation and exon deletions in *Brassica rapa* support a two-step model of paleohexaploidy. *Genetics* **190**: 1563–1574.
Thomas SK, **An H**, **Pires JC. 2023**. Mangroves and multiplications: influence of genome duplications on salt tolerance. *Molecular Ecology* **32**: 275–277.
Thomas SK, **Hoek KV**, **Ogoti T**, ***et al.* 2024**. Halophytes and heavy metals: a multi-omics approach to understand the role of gene and genome duplication in the abiotic stress tolerance of *Cakile maritima*. *American Journal of Botany* **111**: e16310.
Thomas SK, **Liu X**, **Du Z-Y**, ***et al.* 2021**. Comprehending Cornales: phylogenetic reconstruction of the order using the Angiosperms353 probe set. *American Journal of Botany* **108**: 1112–1121.
Torfstein A, **Steinberg J. 2020**. The Oligo-Miocene closure of the Tethys Ocean and evolution of the proto-Mediterranean sea. *Scientific Reports* **10**: 13817.
Tossi VE, **Martínez Tosar LJ**, **Laino LE**, ***et al.* 2022**. Impact of polyploidy on plant tolerance to abiotic and biotic stresses. *Frontiers in Plant Science* **13**: 869423.
Tseng Y-H, **Kuo L-Y**, **Borokini I**, **Fawcett S. 2024**. The role of deep hybridization in fern speciation: examples from the Thelypteridaceae. *American Journal of Botany* **111**: e16388.
Van de Peer Y, **Ashman T-L**, **Soltis PS**, **Soltis DE. 2021**. Erratum to: Polyploidy: an evolutionary and ecological force in stressful times. *Plant Cell* **33**: 2899.
Walden N, **German DA**, **Wolf EM**, ***et al.* 2020**. Nested whole-genome duplications coincide with diversification and high morphological disparity in Brassicaceae. *Nature Communications* **11**: 3795.
Wang T-J, **Ru D-F**, **Zhang D**, **Hu Q-J. 2019**. Analyses of genome-scale variation reveal divergence of two *Sinalliaria* species (Brassicaceae) with continuous but limited gene flow. *Journal of Systematics and Evolution*. **57**:268–277. doi:10.1111/jse.12461
Wang X, **Wang H**, **Wang J**, ***et al.* 2011**. The genome of the mesopolyploid crop species *Brassica rapa*. *Nature Genetics* **43**: 1035–1039.
Warwick SI, **Black LD. 1994**. Evaluation of the subtribes Moricandiinae, Savignyinae, Vellinae, and Zillinae (Brassicaceae, tribe Brassiceae) using chloroplast DNA restriction site variation. *Canadian Journal of Botany* **72**: 1692–1701.
Warwick SI, **Sauder CA. 2005**. Phylogeny of tribe Brassiceae (Brassicaceae) based on chloroplast restriction site polymorphisms and nuclear ribosomal internal transcribed spacer and chloroplast *trn*L intron sequences. *Canadian Journal of Botany* **83**: 467–483.
Yang Z. 2007. PAML 4: phylogenetic analysis by maximum likelihood. *Molecular Biology and Evolution* **24**: 1586–1591.
Yang R, **Jarvis DE**, **Chen H**, ***et al.* 2013**. The reference genome of the halophytic plant *Eutrema salsugineum*. *Frontiers in Plant Science* **4**: 46.
Yu Y, **Than C**, **Degnan JH**, **Nakhleh L. 2011**. Coalescent histories on phylogenetic networks and detection of hybridization despite incomplete lineage sorting. *Systematic Biology* **60**: 138–149.
Zhang C, **Mirarab S. 2022**. ASTRAL-Pro 2: ultrafast species tree reconstruction from multi-copy gene family trees. *Bioinformatics* **38**: 4949–4950.
Zhang C, **Scornavacca C**, **Molloy EK**, **Mirarab S. 2020**. ASTRAL-Pro: quartet-based species-tree inference despite paralogy. *Molecular Biology and Evolution* **37**: 3292–3307.
Zhang K, **Yang Y**, **Zhang X**, ***et al.* 2023**. The genome of *Orychophragmus violaceus* provides genomic insights into the evolution of Brassicaceae polyploidization and its distinct traits. *Plant Communications* **4**: 100431.
Zhang L, **Zeng T**, **Hu H**, **Fan L**, **Zheng H**, **Hu Q. 2018**. Interspecific divergence of two *Sinalliaria* (Brassicaceae) species in eastern China. *Frontiers in Plant Science* **9**: 77.
Ziolkowski PA, **Kaczmarek M**, **Babula D**, **Sadowski J. 2006**. Genome evolution in *Arabidopsis*/*Brassica*: conservation and divergence of ancient rearranged segments and their breakpoints. *Plant Journal* **47**: 63–74.
Zizka A, **Onstein RE**, **Rozzi R**, ***et al.* 2022**. The evolution of insular woodiness. *Proceedings of the National Academy of Sciences of the United States of America* **119**: e2208629119.