Global Brassicaceae phylogeny based on filtering of 1,000-gene dataset

Hendriks, Kasper P.; Al-Shehbaz, Ihsan A.; Bailey, C. Donovan; Hooft van Huysduynen, Alex; Nauheimer, Lars; Zuntini, Alexandre R.; Franzke, Andreas; Schranz, M. Eric; Ly, Elfy; More Authors

DOI
10.1016/j.cub.2023.08.026

Publication date
2023

Document Version
Final published version

Published in
Current Biology

Article

Current Biology

Global Brassicaceae phylogeny based on filtering of 1,000-gene dataset

Highlights

- An unparalleled Brassicaceae phylogeny covering nearly all 349 genera is presented
- Cytonuclear discordance is omnipresent and a likely sign of rampant hybridization
- The family originated during the late Eocene to late Oligocene
- Results are used to come up with a new family classification

Authors

Kasper P. Hendriks, Christiane Kiefer, Ihsan A. Al-Shehbaz, ..., Félix Forest, Klaus Mummenhoff, Frederic Lens

Correspondence

kasper.hendriks@naturalis.nl (K.P.H.), kmummenh@uni-osnabrueck.de (K.M.), frederic.lens@naturalis.nl (F.L.)

In brief

The Brassicaceae family contains the model plant *Arabidopsis thaliana* and many important crop species. Surprisingly, relationships within the family were poorly known. Hendriks et al. present an unprecedented family phylogeny with representatives from nearly all 349 genera. These results will boost fundamental and applied plant biological research.

Hendriks et al., 2023, Current Biology *33*, 4052–4068
October 9, 2023
https://doi.org/10.1016/j.cub.2023.08.026

CellPress

CellPress
OPEN ACCESS

Current Biology

Article

Global Brassicaceae phylogeny based on filtering of 1,000-gene dataset

Kasper P. Hendriks,[1,2,38,39,*] Christiane Kiefer,[3] Ihsan A. Al-Shehbaz,[4] C. Donovan Bailey,[5] Alex Hooft van Huysduynen,[2,6] Lachezar A. Nikolov,[7] Lars Nauheimer,[8] Alexandre R. Zuntini,[9] Dmitry A. German,[10] Andreas Franzke,[11] Marcus A. Koch,[3] Martin A. Lysak,[12] Óscar Toro-Núñez,[13] Barış Özüdoğru,[14] Vanessa R. Invernón,[15] Nora Walden,[3] Olivier Maurin,[9] Nikolai M. Hay,[16] Philip Shushkov,[17] Terezie Mandáková,[12] M. Eric Schranz,[18] Mats Thulin,[19] Michael D. Windham,[16] Ivana Rešetnik,[20] Stanislav Španiel,[21] Elfy Ly,[2,22,23] J. Chris Pires,[24] Alex Harkess,[25] Barbara Neuffer,[1] Robert Vogt,[26]

(Author list continued on next page)

[1]Department of Biology, Botany, University of Osnabrück, Barbarastraße 11, 49076 Osnabrück, Germany
[2]Functional Traits Group, Naturalis Biodiversity Center, Darwinweg 2, 2333 CR Leiden, the Netherlands
[3]Centre for Organismal Studies (COS), Heidelberg University, Im Neuenheimer Feld 345, 69120 Heidelberg, Germany
[4]Missouri Botanical Garden, 4344 Shaw Blvd, St. Louis, MO 63110, USA
[5]Department of Biology, New Mexico State University, PO Box 30001, MSC 3AF, Las Cruces, NM 88003, USA
[6]Department of Biology, University of Antwerp, Groenenborgerlaan 171, 2020 Antwerp, Belgium
[7]Department of Molecular, Cell and Developmental Biology, University of California, 610 Charles E. Young Dr. S., Los Angeles, CA 90095, USA
[8]Australian Tropical Herbarium, James Cook University, PO Box 6811, Cairns, QLD 4870, Australia
[9]Royal Botanic Gardens, Kew, Richmond, Surrey TW9 3AE, UK
[10]South-Siberian Botanical Garden, Altai State University, Barnaul, Lesosechnaya Ulitsa, 25, Barnaul, Altai Krai, Russia
[11]Heidelberg Botanic Garden, Heidelberg University, Im Neuenheimer Feld 361, 69120 Heidelberg, Germany
[12]CEITEC–Central European Institute of Technology, Masaryk University, Kamenice 5, Brno 625 00, Czech Republic
[13]Departamento de Botánica, Universidad de Concepción, Barrio Universitario, Concepción, Chile
[14]Department of Biology, Hacettepe University, Beytepe, Ankara 06800, Türkiye
[15]Sorbonne Université, Muséum National d'Histoire Naturelle, Institut de Systématique, Évolution, Biodiversité (ISYEB), CP 39, 57 rue Cuvier, 75231 Paris Cedex 05, France

(Affiliations continued on next page)

SUMMARY

The mustard family (Brassicaceae) is a scientifically and economically important family, containing the model plant *Arabidopsis thaliana* and numerous crop species that feed billions worldwide. Despite its relevance, most phylogenetic trees of the family are incompletely sampled and often contain poorly supported branches. Here, we present the most complete Brassicaceae genus-level family phylogenies to date (Brassicaceae Tree of Life or BrassiToL) based on nuclear (1,081 genes, 319 of the 349 genera; 57 of the 58 tribes) and plastome (60 genes, 265 genera; all tribes) data. We found cytonuclear discordance between the two, which is likely a result of rampant hybridization among closely and more distantly related lineages. To evaluate the impact of such hybridization on the nuclear phylogeny reconstruction, we performed five different gene sampling routines, which increasingly removed putatively paralog genes. Our cleaned subset of 297 genes revealed high support for the tribes, whereas support for the main lineages (supertribes) was moderate. Calibration based on the 20 most clock-like nuclear genes suggests a late Eocene to late Oligocene origin of the family. Finally, our results strongly support a recently published new family classification, dividing the family into two subfamilies (one with five supertribes), together representing 58 tribes. This includes five recently described or re-established tribes, including Arabidopsideae, a monogeneric tribe accommodating *Arabidopsis* without any close relatives. With a worldwide community of thousands of researchers working on Brassicaceae and its diverse members, our new genus-level family phylogeny will be an indispensable tool for studies on biodiversity and plant biology.

INTRODUCTION

The mustard family (Brassicaceae) is a globally distributed, medium-sized plant family (~4,000 species) with huge economic and scientific impact, characterized by high morphological diversity[1–5] (Figure 1). The family contains numerous species grown for food and biofuel (cabbage, rapeseed) as well as many model species (*Arabidopsis thaliana*, *Arabis alpina*,

Christian Bräuchler,[27] Heimo Rainer,[27] Steven B. Janssens,[28,29] Michaela Schmull,[30] Alan Forrest,[31] Alessia Guggisberg,[32] Sue Zmarzty,[9] Brendan J. Lepschi,[33] Neville Scarlett,[34] Fred W. Stauffer,[35] Ines Schönberger,[36] Peter Heenan,[36] William J. Baker,[9] Félix Forest,[9] Klaus Mummenhoff,[1,*] and Frederic Lens[2,37,39,*]

[16]Department of Biology, Duke University, Durham, NC 27708, USA
[17]Department of Chemistry, Indiana University, 800 E. Kirkwood Ave., Bloomington, IN 47405, USA
[18]Biosystematics Group, Wageningen University, Droevendaalsesteeg 1, 6708 PB Wageningen, the Netherlands
[19]Department of Organismal Biology, Uppsala University, Norbyvägen 18, 752 36 Uppsala, Sweden
[20]Department of Biology, University of Zagreb, Marulićev trg 20/II, 10000 Zagreb, Croatia
[21]Institute of Botany, Slovak Academy of Sciences, Plant Science and Biodiversity Centre, Dúbravská cesta 9, 845 23 Bratislava, Slovakia
[22]Wetsus, European Centre of Excellence for Sustainable Water Technology, Oostergoweg 9, 8911 MA Leeuwarden, the Netherlands
[23]Department of Biotechnology, Delft University of Technology, Van der Maasweg 9, 2629 HZ Delft, the Netherlands
[24]Soil and Crop Sciences, Colorado State University, 307 University Ave., Fort Collins, CO 80523-1170, USA
[25]HudsonAlpha Institute for Biotechnology, 601 Genome Way Northwest, Huntsville, AL 35806, USA
[26]Botanischer Garten und Botanisches Museum, Freie Universität Berlin, Königin-Luise-Straße 6-8, 14195 Berlin, Germany
[27]Department of Botany, Natural History Museum Vienna, Burgring 7, 1010 Vienna, Austria
[28]Department of Biology, KU Leuven, Kasteelpark Arenberg 31 - box 2435, 3001 Leuven, Belgium
[29]Meise Botanic Garden, Nieuwelaan 38, 1860 Meise, Belgium
[30]Harvard University Herbaria, 22 Divinity Ave., Cambridge, MA 02138, USA
[31]Centre for Middle Eastern Plants, Royal Botanic Garden Edinburgh, 20A Inverleith Row, Edinburgh EH3 5LR, UK
[32]ETH Zürich, Institut für Integrative Biologie, Universitätstrasse 16, 8092 Zürich, Switzerland
[33]Australian National Herbarium, Centre for Australian National Biodiversity Research, Clunies Ross St, Acton, ACT 2601, Australia
[34]La Trobe University, Plenty Road and Kingsbury Dr., Bundoora, VIC 3086, Australia
[35]Conservatory and Botanic Gardens of Geneva, CP 60, Chambésy, 1292 Geneva, Switzerland
[36]Manaaki Whenua Landcare Research, Allan Herbarium, PO Box 69040, Lincoln, New Zealand
[37]Institute of Biology Leiden, Plant Sciences, Leiden University, Sylviusweg 72, 2333 BE Leiden, the Netherlands
[38]Twitter: @FunctionalTrai3
[39]Lead contact
*Correspondence: kasper.hendriks@naturalis.nl (K.P.H.), kmummenh@uni-osnabrueck.de (K.M.), frederic.lens@naturalis.nl (F.L.)
https://doi.org/10.1016/j.cub.2023.08.026

Brassica spp., and *Capsella* spp.).[1,6] Moreover, the overwhelming availability of genetic tools and resources in these species make Brassicaceae an ideal model family in flowering plants. These tools have facilitated, among other things, plant developmental studies that (1) disentangle genotype-phenotype interactions,[7,8] (2) investigate impacts of whole-genome duplications (WGDs) at different timescales,[9,10] and (3) unravel the evolutionary variation in metabolic pathways, leading to a huge diversity in natural products.[11] In many of these comparative studies that go beyond model species, a robust evolutionary framework is required, ideally encompassing the ~4,000 currently accepted species of Brassicaceae that are divided among 349 genera and 50–60 tribes.[12] This is also true for studies on crop wild relatives aiming to introgress desirable traits (e.g., drought tolerance and disease resistance) into crops using plant breeding.[13–15] Therefore, the evolution of Brassicaceae has been the subject of study for a long time.[2,3,16–23] However, a robust, densely sampled Brassicaceae Tree of Life (hereafter named BrassiToL) remains lacking.

Early phylogenetic inferences relying on flower and fruit morphology[25–27] were often misled by rampant convergent evolution.[28–30] Subsequently, the use of few molecular markers generated poorly supported phylogenies,[17,31] traditionally attributed to an early rapid radiation[20] and multiple WGD and hybridization events.[1] More recent family-wide phylogenetic studies based on multiple molecular markers offered support for three,[32] four,[1] or five[3,22,33–35] main lineages in addition to tribe Aethionemeae, which is sister to all other Brassicaceae. However, several evolutionary events in the family keep challenging the reconstruction of the BrassiToL backbone,[10,36–40] including incomplete lineage sorting and introgression,[41] and (ancient) inter-tribal and inter-lineage hybridization, commonly followed by post-polyploid diploidization.[42,43]

Here, we present results from the largest global Brassicaceae phylogenetic study to date including nearly all genera, with the following objectives: (1) reconstruct the most complete and robust family-wide BrassiToL based on nuclear genomic data and plastome data, (2) investigate the influence of paralogous genes and polyploid species on phylogeny reconstruction, (3) re-evaluate the temporal origin of the Brassicaceae, its main lineages (subfamilies and supertribes), and tribes, and (4) provide an updated taxonomic delimitation of these lineages and tribes. Our improved phylogenetic framework will further develop Brassicaceae as the prime model family in flowering plants.

RESULTS

Unparalleled family-wide sampling

Our ingroup sampling for the nuclear analyses included 380 samples, covering 375 species, 319 of the 349 currently accepted genera (91%) and 57 of the 58 accepted tribes following the latest taxonomic revision of German et al.[24] (excluding Hillielleae; Data S1A and S1B). We added 23 outgroup species representing all Brassicales families and 14 former Brassicaceae genera now synonymized with accepted genera. Three additional genera (*Hilliella*, *Onuris*, and *Thelypodium*) were included in the plastome dataset (Data S1C).

We recovered 1,081 nuclear genes (Data S1D; Methods S1A) across 397 samples (Data S1E) from target capture sequencing of 764 Brassicaceae-specific genes[3] (hereafter B764) and 353 Angiosperm-wide genes[44,45] (hereafter A353), following a "mixed

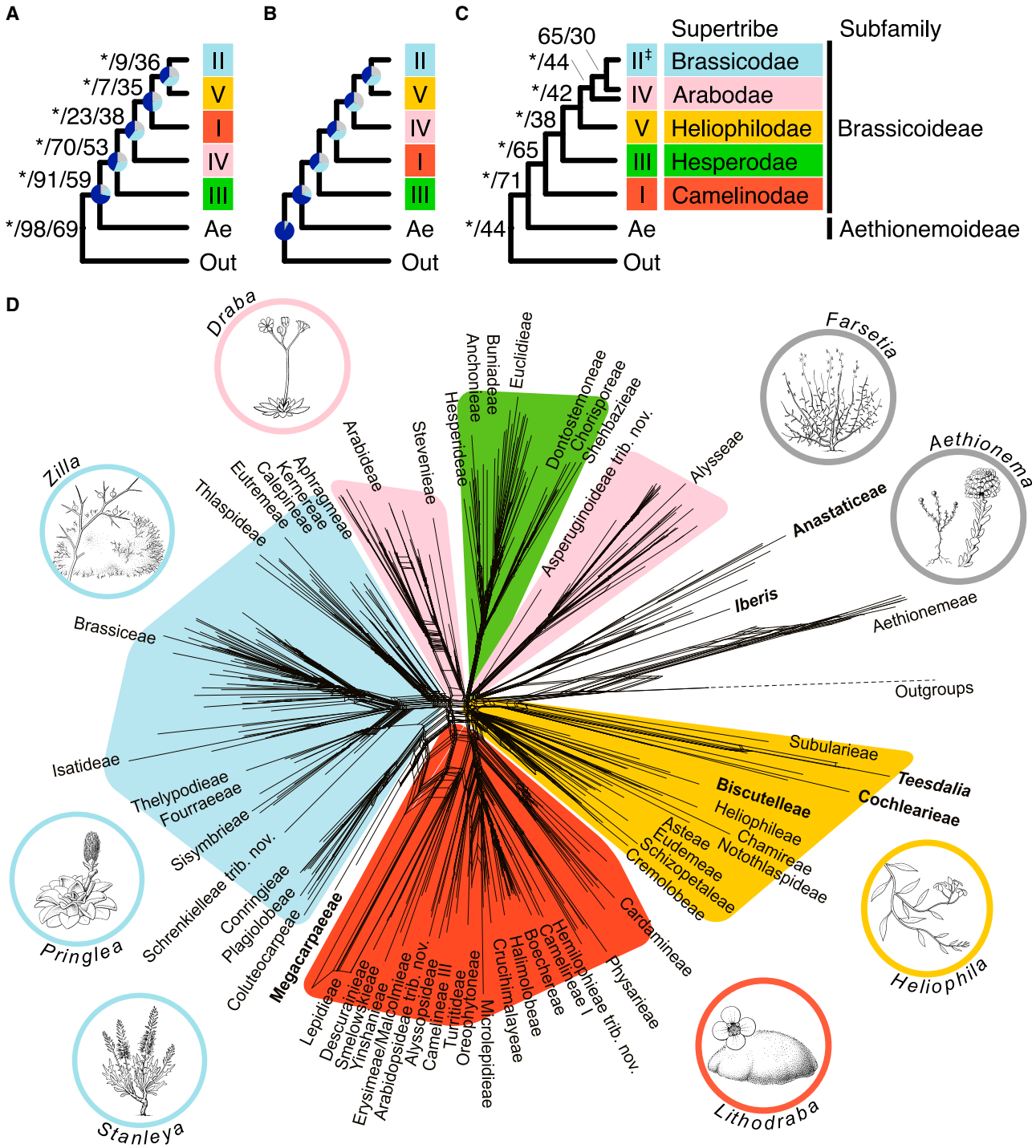

Figure 1. Overview of the main mustard family relationships from different phylogenetic reconstruction approaches and gene sampling routines

Color group is the main core Brassicaceae lineages (I–V) described by Nikolov et al.,[3] with recently proposed names for subfamilies and supertribes following German et al.[24] See also Figure S3.

(A) Cladogram of main lineages as recovered from nuclear stricter routines (strict, superstrict, superstrict by tribe, and superstrict excluding hybrids) using either supermatrix ML or coalescent-based approaches (ASTRAL-III), with node support from the supermatrix ML approach on the superstrict excluding hybrids routine (BS/gCF/sCF; * indicates 100%) and the ASTRAL-III approach on the superstrict routine (pie charts represent quartet scores: dark blue for the first quartet, light blue for the first alternative, and gray for the second alternative).

(B) Cladogram of main lineages as recovered from nuclear inclusive routine using coalescent-based approach (using both ASTRAL-III and ASTRAL-Pro) with node support from the ASTRAL-III approach on the inclusive routine (colors as in previous panel).

(legend continued on next page)

baits" methodology.[46] For B764, we recovered 91% of the total target length of 888,392 bp (mean across all samples), and 84% of the 728 unique genes targeted (36 genes in common with the A353 kit were first removed). For A353, mean values were 83% of the total target length of 294,516 bp and 80% of 353 genes targeted. Interestingly, gene recovery hardly decreased with sample age, and we successfully obtained hundreds of nuclear genes even from pre-1900 samples (Methods S1B).

Unparalleled global nuclear Brassicaceae phylogeny

We used HybPhaser v2.0[47] to define five "sampling routines" ("inclusive": 1,018 genes, "strict": 1,013 genes, "superstrict": 297 genes, "superstrict by tribe": 1,049 genes, and "superstrict excluding hybrids": 303 genes; Table 1). In the superstrict by tribe routine, putative paralogous genes were removed by tribe, meaning that a relatively high number of genes could be kept for further analyses (see STAR Methods for details and Methods S1C and S1D). In addition, we used both supermatrix maximum likelihood (ML) and coalescent-based approaches in ASTRAL-III[48] and ASTRAL-Pro[49] to infer the species phylogeny.

Node support was high for most tribes, regardless of the sampling routine or phylogenetic approach (Figure 2; Data S1E and S2), providing strong support for the monophyly of many (but not all) of the tribes. For example, from the supermatrix approach of the superstrict routine (concatenation of 297 target genes; total length 700,445 bp; 82.7% complete), the median support values across tribal nodes were 100% for bootstrap (BS), 50% for gene concordance factor (gCF), and 64% for site concordance factors (sCFs). From the coalescent-based approach (strict routine), the median local posterior probability (LPP) across tribal nodes was 100%, and support for the first quartet (Q1) was 78%. Tribes Camelineae and Iberideae were polyphyletic in all nuclear phylogenies (Data S2).

Based on the nuclear dataset, we consistently found tribe Aethionemeae sister to the rest of the Brassicaceae (hereafter defined as "core Brassicaceae"), main lineage III sister to a clade of lineages I–V, and lineages II + V sharing a common ancestor (Figures 1A and 1B).

Based on the superstrict routine (361 samples; Table 1), support for the split between Aethionemeae and the core Brassicaceae was high (BS: 100%; gCF: 70%; sCF: 63%; LPP: 100%; Q1: 70%; Figure 2; Data S2A), whereas support for the five remaining main lineages varied. Although main lineage III received good support (100%, 40%, 58%, 100%, and 91%, respectively), support for the remaining main lineages was relatively poor in terms of gCF (ranging from 1% to 5%) and sCF (ranging from 36% to 58%) values (Data S1F). Remarkably, BS and LPP values for these main lineages were (nearly) always 100%. The generally low support for the phylogenetic backbone (focusing on gCF, sCF, and Q1) was reflected in the topological differences among the family phylogenies based on the different sampling routines (Figures 1A–1C) and was visualized as a large, complex reticulate core in our split network analysis (Figure 1D). Our approach to increase backbone support—and thereby confidence in the family's main lineage relationships—with our superstrict excluding hybrids routine indeed resulted in higher support for all main lineages (gCF now ranging from 16% to 76%; range of sCF unchanged, but support for all main lineages the same or slightly higher; Data S1F). However, this came at the cost of removing many samples (and therefore species, genera, and tribes), with now only 138 samples left (Table 1).

Our inclusive routine resulted in placement of main lineage I as sister to a clade of main lineages II + IV + V (Figure 1A), whereas the stricter routines of the nuclear dataset showed a consistent placement of lineage IV as sister to a clade of lineages I + II + V (Figure 1B) in both the ML and coalescent-based approaches. This showed that removing a first set of most variable genes (strict routine) considerably impacts the backbone of the topology, whereas an additional removal of less variable genes (superstrict, superstrict by tribe, and superstrict excluding hybrids) has little effect.

We used Townsend's phylogenetic informativeness[50] to quantify the impact of each of the 1,081 nuclear genes on the final species tree (Figure S1). Gene informativeness varied markedly, with genes from the B764 bait set on average more informative than those from the A353 bait set, which likely reflects the family-specific design of the B764 bait set.

Extended global plastome Brassicaceae phylogeny

Our plastome family ingroup sampling included 502 samples, covering 438 species, 266 genera (76%), and all 58 tribes (following German et al.[24]), with raw input data from newly sequenced samples (Data S1A) and previously published data from Nikolov et al.[3] (Data S1B) and Walden et al.[2] (Data S1G).

Similar to the nuclear phylogeny, tribal support was generally high, with median nodal sCF of 81% and BS of 100% (Figure 3; Data S1F). As in the nuclear phylogeny, tribe Aethionemeae was sister to all remaining Brassicaceae. Tribes within lineage II were similar to those in the nuclear phylogeny. However, cytonuclear discordance manifested by among others the following topological differences (Figure 4): (1) in the plastome phylogeny, lineage I (not III) was sister to all other lineages (II–V), with lineage III sister to a clade formed by II + IV + V; (2) lineage II was polyphyletic in the plastome phylogeny, with several tribes (Aphragmeae, Coluteocarpeae, Conringieae, Kernereae, and Plagiolobeae) recovered within lineage V; (3) tribe Stevenieae, recovered within lineage IV in the nuclear phylogeny, was recovered within lineage I in the plastome phylogeny; and (4) new tribe Asperuginoideae (see below), assigned to lineage IV in the nuclear phylogeny, formed a clade with Biscutelleae in the plastome phylogeny, and these tribes together were sister to all other remaining lineages (II + IV + V).

(C) Cladogram of main lineages as recovered from plastome supermatrix ML approach with its node support (BS/sCF). The ‡ indicates a major polyphyly in main lineage II (supertribe Brassicodae).

(D) Split network of the mustard family tribes computed from uncorrected p-distances on a supermatrix of the nuclear genes covered by the superstrict routine, covering 317 genera of 56 tribes. The network highlights the complex reticulate evolution both in the ancestors of extant main lineages, as well as within some of the main lineages. Names in bold highlight rogue taxa described in the main text. Inset drawings show the high morphological diversity within the family, which contains growth forms such as herbaceous, frutescent, and woody species, the latter comprising lianas, shrubs, and cushion plants. Drawings by Esmée Winkel, Naturalis Biodiversity Center.

Table 1. Sampling routine definitions

Sampling routine	Inclusive	Strict	Superstrict	Superstrict by tribe	Superstrict excl. hybrids
Gene and sample selection criteria					
Gene minimum proportion of samples recovered	0.1	0.2	0.2	0.2	0.2
Gene minimum proportion of target length recovered	0.1	0.2	0.2	0.2	0.2
Sample minimum proportion of total target length recovered	0.0	0.4	0.4	0.4	0.4
Sample minimum proportion of genes recovered	0.0	0.2	0.2	0.2	0.2
Gene SNPs proportion threshold for all samples	none	outliers	0.02	0.02	0.02
Remove outlier genes per sample?	yes	yes	yes	yes	yes
Remove putative hybrids and rogue taxa?	no	no	no	no	yes[a]
Dataset summary					
Number of genes that passed HybPhaser criteria					
All genes	1,081 (100)	1,057 (97.8)	306 (28.3)	1,076 (99.5)	306 (28.3)
B764 genes	728[b] (100)	721 (99)	194 (26.6)	728 (100)	194 (26.6)
A353 genes	353 (100)	336 (95.2)	112 (31.7)	348 (98.6)	112 (31.7)
Number of genes that passed de-noising loop and used in phylogenetic analysis					
All genes	1,018 (94.2)	1,013 (93.7)	297 (27.5)	1,049 (97) (1,031[c])	303 (28)
B764 genes	684 (94)	688 (94.5)	189 (26.0)	709 (97.4)	193 (26.5)
A353 genes	334 (94.6)	325 (92.0)	108 (30.6)	340 (96.3)	110 (31.2)
Number of samples in final dataset	380	361	361	361	138
Number of species	375	356	356	356	138
Number of genera as currently accepted	319 (332)	305 (317)	305 (317)	305 (317)	124
Number of tribes[d]	57	56	56	56	36
Coalescent-based results from ASTRAL-III					
LPP (mean of all nodes)	0.94	0.95	0.92	0.88	–
LPP (median of all nodes)	1	1	1	1	–
First quartile proportion (mean of all nodes)	0.55	0.56	0.57	0.56	–
First quartile proportion (median of all nodes)	0.48	0.50	0.50	0.49	–

Definition of sampling routines and criteria used in HybPhaser to select nuclear genes for downstream analyses, with dataset summary and coalescent-based results from ASTRAL-III. Values in parentheses are percentages relative to the total number of genes in each bait set (728 for B764, i.e., 764 minus 36 overlapping genes; 353 for A353). See also Methods S1.

[a]Putative hybrids were selected based on locus heterozygosity and allele divergence, with all samples within the upper 50% removed; rogue tribes as discussed in the main text.

[b]This corresponds to the 764 genes from the bait set minus 36 genes overlapping with the A353 bait set.

[c]Mean number across samples, with genes retained differently in different tribes (i.e., tribe-specific). As a result, more genes could be retained in total as compared with the inclusive routine, but at a lower mean sample occupancy across all genes, thus leading to a relatively sparse matrix.

[d]Tribe definitions following latest taxonomic insights presented in German at al.,[24] i.e., 58 tribes in total.

Topological incongruences within tribes between the nuclear and plastome phylogenies could be calculated for 23 tribes for which the number of shared species was >3 (Data S1H). Results varied from full concordance (tribes Biscutelleae, Chorisporeae, Cremolobeae, Descurainieae, Eudemeae, Heliophileae, Isatideae, and Physarieae; generalized Robinson-Foulds distance 0) to full discordance (tribe Lepidieae only; generalized Robinson-Foulds distance 1), with a median generalized Robinson-Foulds distance across all tribes of 0.31.

Allelic variability highlights polyploid origins

We assessed allelic variation using HybPhaser v2.0[47] by calculating the proportions of loci with divergent alleles (% locus heterozygosity [LH]) and average divergence between alleles (% allele divergence [AD]), using the 1,013 nuclear genes included in the strict routine (ignoring 38 samples with coverage too low to calculate LH and AD; Data S1E). Although most samples fell into the expected range of "normal" diploid species (roughly LH < 90% and AD < 1), a number of samples showed signs of polyploidization (Figures 5 and S2). Mean values for both LH and AD were high at 75.1% and 2.25%, respectively, with no clear difference between genes from the B764 and A353 bait sets (Data S1E and S1I).

Based on the list of LH and AD values for all species, we tentatively distinguished four family-specific classes: "hybrid," "old polyploid," "highly polyploid," and "old and highly polyploid" (Figure 5; see STAR Methods for details and class circumscription). Nearly all representatives of tribe Thelypodieae were highly polyploid, suggesting that the tribe experienced one or more recent polyploidization events. Similarly, all representatives of tribe Lepidieae fell within the hybrid or highly polyploid classes, and also, tribes Alysseae, Brassiceae, Cardamineae,

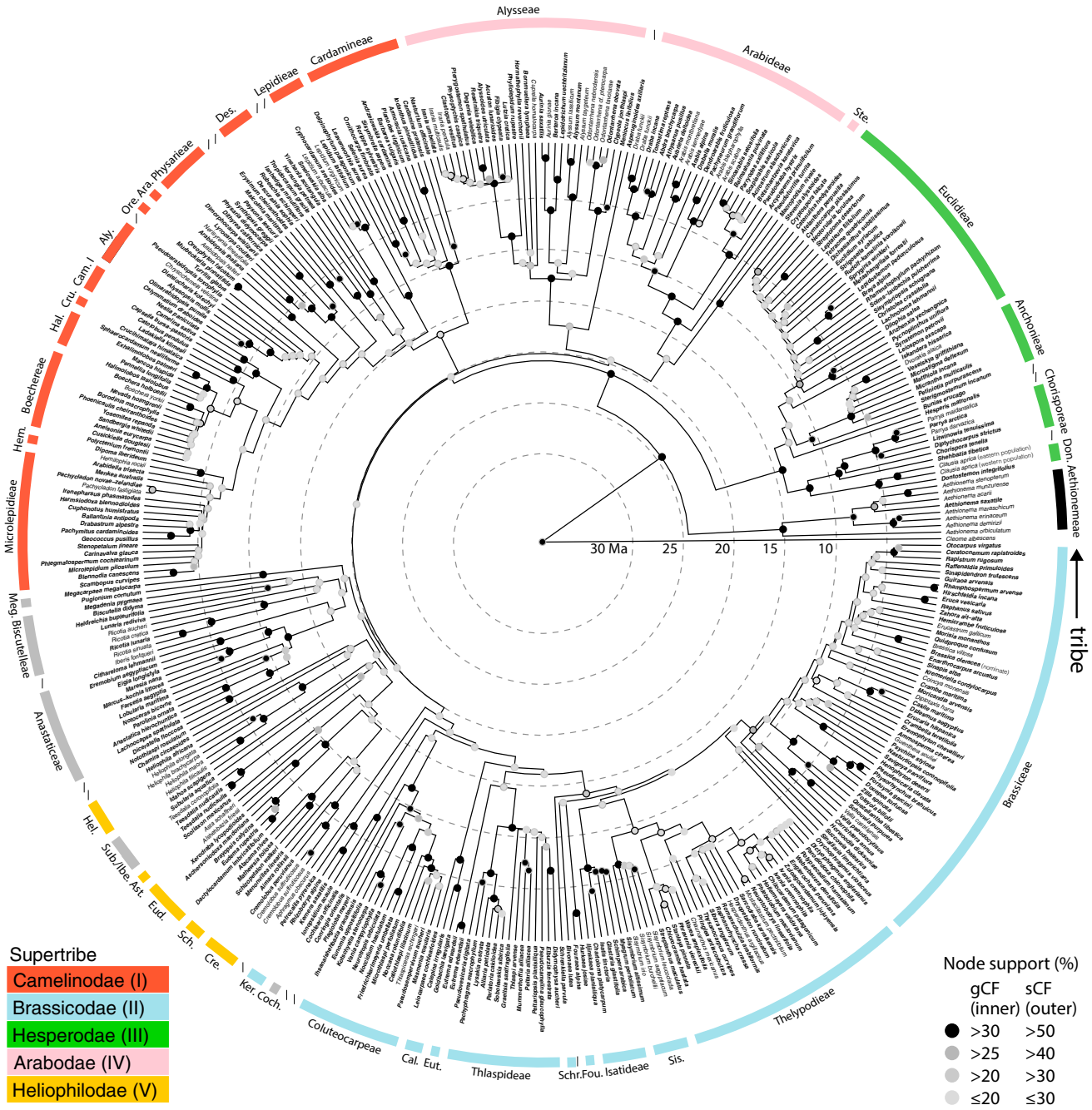

Figure 2. Time-calibrated genus-level Brassicaceae Tree of Life (BrassiToL) from a maximum likelihood analysis of a 297 nuclear genes supermatrix (superstrict routine)

Genus type species are highlighted in bold. All tribes with more than a single representative are listed. Abbreviations of tribes are as follows: Aly., Alyssopsideae; Ara., Arabidopsideae trib. nov.; Ast., Asteae; Cal., Calepineae; Cam. I, Camelineae I; Coch., Cochlearieae; Cre., Cremolobeae; Cru., Crucihimalayeae; Des., Descurainieae; Don., Dontostemoneae; Eud., Eudemeae; Eut., Eutremeae; Fou., Fourraeeae; Hal., Halimolobeae; Hel., Heliophileae; Hem., Hemilophieae; Ibe., Iberideae; Ker., Kernereae; Meg., Megacarpaeeae; Ore., Oreophytoneae; Sch., Schizopetaleae; Schr., Schrenkielleae trib. nov.; Sis., Sisymbrieae; Ste., Stevenieae; Sub., Subularieae. See also Data S2A for a fully annotated version of this phylogeny (including bootstrap, gCF, sCF and node age 95% HPD intervals) with outgroups representing all families within the order Brassicales and calibration nodes. See also Data S2A.

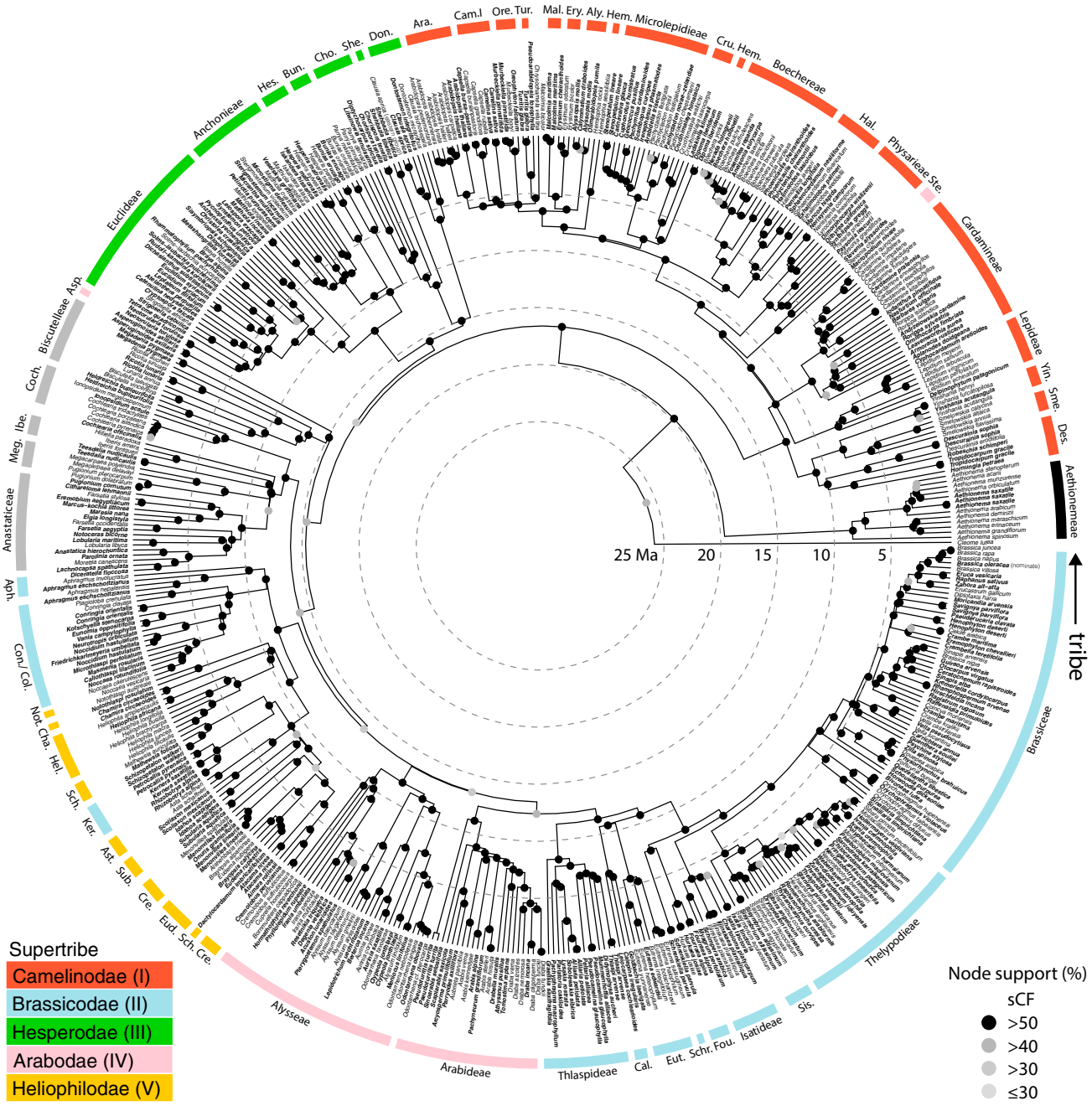

Figure 3. Time-calibrated genus-level Brassicaceae Tree of Life from a maximum likelihood analysis of a 60 plastome genes supermatrix

Genus type species are highlighted in bold. All tribes with more than a single representative are listed. Abbreviations of tribes follow those of Figure 2, with additionally: Aph., Aphragmeae; Asp., Asperuginoideae trib. nov.; Bun., Buniadeae; Cha., Chamireae; Cho., Chorisporeae; Col., Coluteocarpeae; Con., Conringieae; Don., Dontostemoneae; Ery., Erysimeae; Hes., Hesperideae; Mal., Malcolmieae; Not., Notothlaspideae; Ore., Oreophytoneae; She., Shehbazieae; Sme., Smelowskieae; Tur., Turritideae; Yin., Yinshanieae. See also Data S2B for a fully annotated version of this phylogeny (including bootstrap, sCF and node age 95% HPD intervals) with outgroups representing all families within the order Brassicales and calibration nodes. See also Data S1.

and Microlepidieae had the bulk of their representatives in one of the polyploid classes (Figure 5).

Our phylogenetic results highlighted several jumpy or rogue taxa (mostly tribes) that were recovered in different positions in the phylogenies from different sampling routines. Most importantly, in the nuclear supermatrix ML approach and the stricter coalescent-based approaches (Figures S3A and S3D–S3F), Anastaticeae, Biscutelleae, and Megacarpaeeae formed a clade sister to lineage V, whereas in the inclusive coalescent-based approaches (ASTRAL-III and ASTRAL-Pro; Figures S3B and S3C), Megacarpaeeae was sister to Anastaticeae, with the two tribes sister to lineage II. The position of tribe Biscutelleae

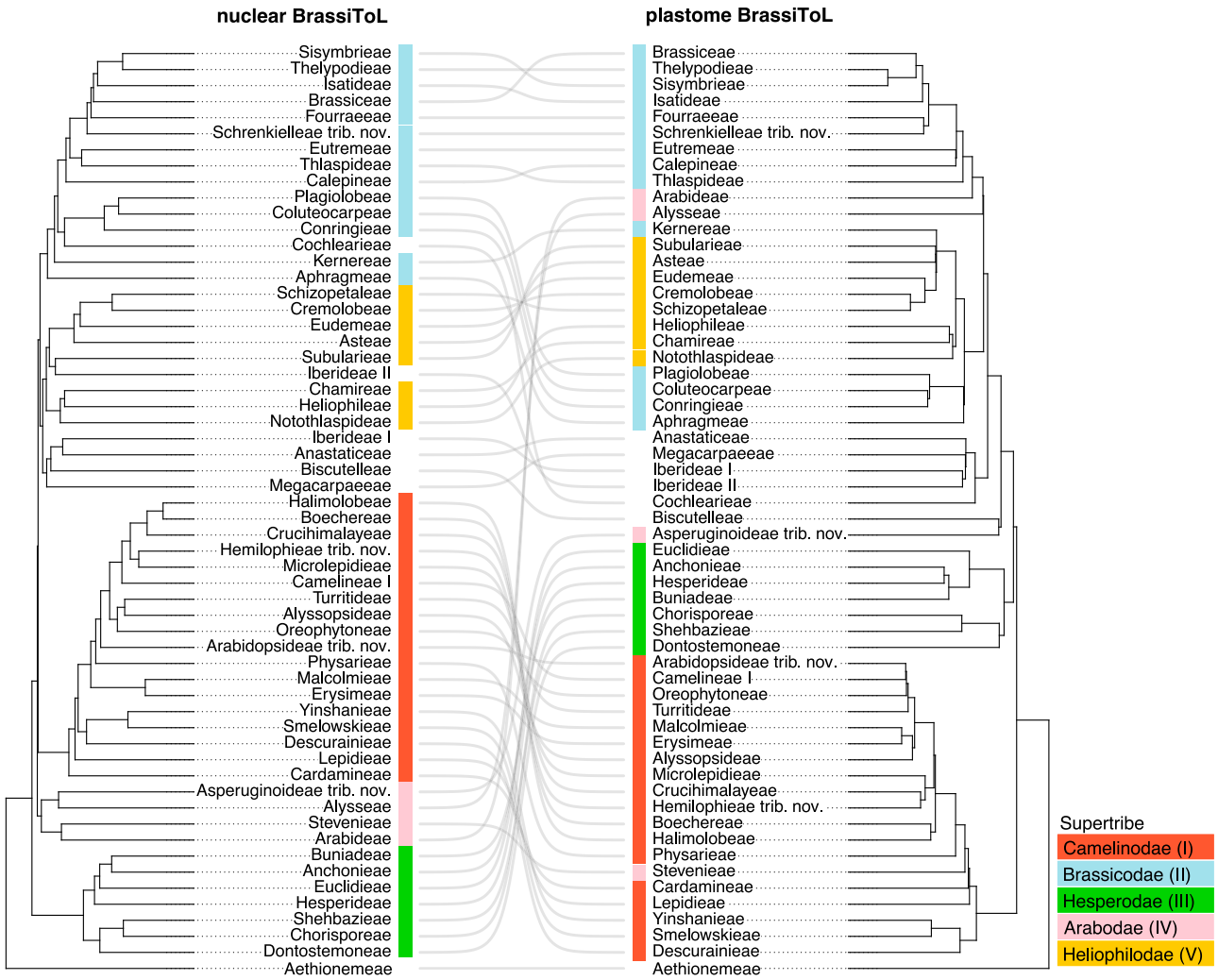

Figure 4. Cytonuclear discordance at tribe level in newly derived nuclear and plastome Brassicaceae Trees of Life

Curved lines between tip labels from the two phylogenies link the same tribes. Tribes are represented by a randomly chosen sample from each tribe. Rogue tribes have not yet been assigned to a supertribe due to their changing topological position from different sampling routines and phylogenetic approaches.

changed depending on the coalescent-based approach used (ASTRAL-III vs. ASTRAL-Pro). Contrary to our plastome phylogeny, we recovered genus *Iberis* (tribe Iberideae) as sister to tribe Anastaticeae in our nuclear phylogenies, and not as sister to *Teesdalia*, the other genus assigned to tribe Iberideae. Species that belong to one of the rogue taxa generally also show relatively high LH and AD values (17 of the 29 samples fall within the "realm of polyploids," with another 6 samples bordering it; Figure 5), suggesting a polyploid origin.

Fossil calibration shows an Eocene origin of the mustard family

We used the Turonian *Dressiantha bicarpellata* fossil (93.6–89.3 Ma[51]) to calibrate the stem node of order Brassicales (Data S2A and S2B). We validated our dating estimates with nine other fossils and biogeographical events (Data S1J), as suggested by Franzke et al.[52] In general, results from our nuclear phylogeny supported the expected dates (based on literature) or were younger. In the plastome phylogeny, our age estimates were significantly younger. We found good corroboration with age estimates for the Miocene *Capparidoxylon holleisii* fossil (16.3 Ma[53,54]) in the nuclear phylogeny (poor in the plastome phylogeny), the maximum crown age of the Mediterranean genus *Ricotia* (11.3–9.2 Ma[55]) in both phylogenies, the estimated age of the most recent common ancestor of the *Arabis alpina* clade (3.27–2.65 Ma[56]) in the nuclear phylogeny (poor in the plastome phylogeny), and the maximum age of the New Zealand alpine genus *Pachycladon* (max. 1.9 Ma[57,58]) in both phylogenies. Support was medium in the nuclear phylogeny (and poor in the plastome phylogeny) for the Paleocene fossil *Akania* sp. (~61 Ma[59,60]) and the early Tertiary *Palaeocleome lakensis* fossil (55.8–48.6 Ma[61]). Support was poor for the vicariance event in *Clausia aprica*[62] and the maximum age of Hawaiian *Lepidium*,[63] with recovered ages in both the nuclear and plastome phylogenies actually older than expected. Corroboration with the Brassicaceae *Thlaspi primaevum* fossil (~32 Ma[64–66]) was poor

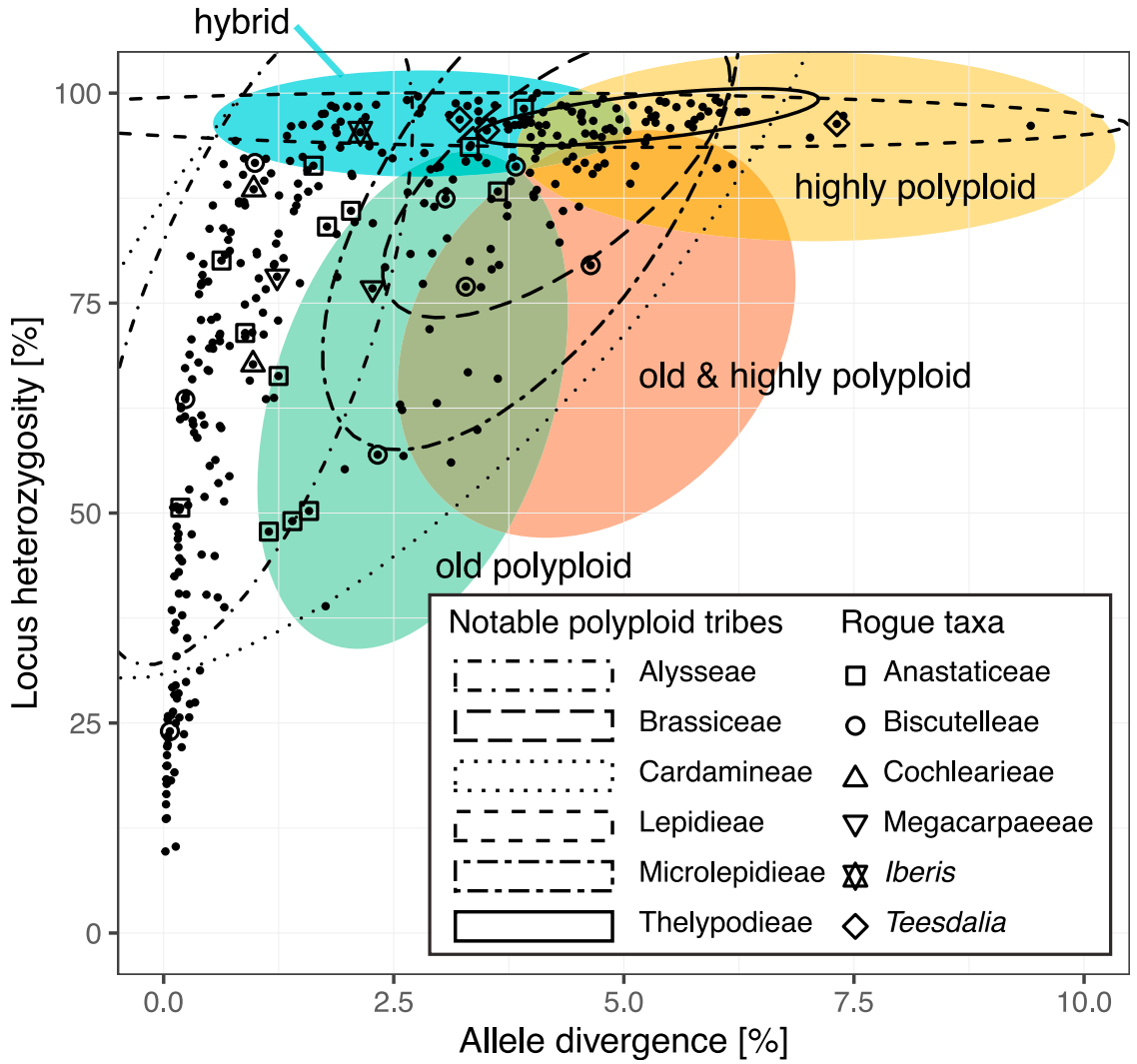

Figure 5. Scatterplot displaying the locus heterozygosity and allele divergence of all samples included in the strict routine

Colored ovals indicate four classes in the expected realm of polyploids, rough estimates of what could be considered likely hybrids, polyploids, and their ages (see main text for the rationale behind assigning these different classes). The overlap between these classes highlights the uncertainty. The six largest tribes (represented by >10 samples) for which (most) of their samples fall within this realm of polyploids are annotated using ellipses showing the 95% confidence level for a multivariate t-distribution of their data. Samples from rogue taxa are highlighted using different shapes and colors. For more details, see Data S1E and S1I and Figure S2.

(node estimated at 10.8 Ma) in both phylogenies, which is unsurprising, given the ambiguous taxonomic identity of this fossil.[52]

Based on our nuclear dataset, we found that the core Brassicales (Brassicaceae + Cleomaceae) and Capparaceae split around 43.2 Ma (95% highest posterior density, or HPD: 45.0–41.1) in the middle Eocene (Data S2A). Subsequently, the Brassicaceae split from sister family Cleomaceae around 38.8 Ma (95% HPD: 40.5–36.9) in the middle to late Eocene (Data S1J). The core Brassicaceae split from tribe Aethionemeae around 24.5 Ma (95% HPD: 25.7–23.1), in the late Oligocene. All five main lineages (supertribes) in the core Brassicaceae originated during rapid radiation in the early to middle Miocene (median stem ages 21.2–19.8 Ma; median crown ages 19.9–14.4 Ma; Figure S4; Data S1F). Mean median stem age across tribes was 12.1 Ma (max. 18.9 Ma, Subularieae; min. 4.0 Ma, Boechereae; n = 52), whereas mean median crown age was 8.0 Ma (max. 18.1 Ma, Subularieae; min. 0.2 Ma, Oreophytoneae; n = 37). Results from our plastome dataset were much younger than those from our nuclear dataset, with a family crown age of 20.2 Ma (95% HPD: 29.0–13.0) and a core Brassicaceae crown age of 16.9 Ma (95% HPD: 24.3–10.2; Data S1F).

DISCUSSION

Unprecedented genus-level BrassiToL

We present here the most complete nuclear and plastome-derived BrassiToL to date, together based on more than 1,000 genes and covering 92% of the 349 genera, representing all 58 tribes (Figures 1, 2, and 3; Data S2). This global genus-level BrassiToL will be an indispensable tool for family experts

interested in understanding the evolutionary processes that shaped current biodiversity patterns. In addition, many scientists working on *Arabidopsis* or other Brassicaceae model species or crops will benefit from this improved phylogenetic framework when investigating gene regulatory mechanisms across the family or exploring specific traits related to stress/disease in crops and their wild relatives.

The backbone of our nuclear BrassiToL largely agrees with that of Nikolov et al.[3] (based solely on the Brassicaceae-specific bait set targeting 764 genes), which included 63 species covering most tribes. Specifically, our nuclear BrassiToL confirms (1) the recognition of tribe Aethionemeae as sister to the rest of the Brassicaceae (i.e., core Brassicaceae) and (2) main lineage III (supertribe Hesperodae, Figures 1A and 1B) as sister to the remaining core Brassicaceae.[3,33,34] The relationships among the other main lineages varied depending on the applied sampling routine and/or phylogenetic approach, although the sister relationship between main lineages II (Brassicodae) and V (Heliophilodae) was consistently recovered as well (Figures 1A and 1B). We are confident that our expanded dataset and more diverse analyses of putative single-copy markers within Brassicaceae offer the most trustworthy estimation of the family relationships to date. We acknowledge, however, that—even with this massive dataset and careful analyses—we have not fully resolved the well-known problem of recovering deeper nodes in the Brassicaceae phylogeny (see below).

When comparing our plastome phylogeny (representing 265 currently accepted genera, all 58 tribes) with the nuclear phylogeny (representing 319 currently accepted genera, 57 tribes), we found strong cytonuclear discordance among the main lineages, as recently also demonstrated by Nikolov et al.[3] and Mabry et al.[33] (Figure 4). At a lower taxonomic level, such cytonuclear discordance was studied in more detail for *Arabidopsis* and its putative close relatives,[41] as well as tribe Biscutelleae,[43] where it was hypothesized that complex hybridization and introgression events caused topological incongruences. For example, tribe Biscutelleae harbors four genus-specific WGDs resulting from hybridization between parental genomes belonging to the same lineages, closely related lineages, and even two different supertribes[43] (Camelinodae and Brassicodae).

Toward resolving the family phylogeny backbone

Relationships at the shallower nodes are generally well-resolved in our phylogenies, leading to strongly supported relationships among genera within most of the tribes (e.g., Q1 node support was >50% in 34 of the 42 tribes for which we could calculate node support, i.e., tribes represented by >1 species; Figure S4; Data S1F). Interestingly, most of the deepest nodes (connecting the Brassicales families) are generally also well supported (Figure S4). However, the deeper nodes *within* the Brassicaceae family—reflecting the relative positions of the five supertribes in the subfamily Brassicoideae—have been notoriously hard to recover in the past and remain incompletely resolved even with more than a thousand genes to study (Figure 1). Importantly, we show that high BS support (used in the past to claim a solid family backbone[35]) can be recovered nonetheless, even for conflicting topologies from different sampling routines, thereby challenging the value of BSs in phylogenomics. Instead, metrics on node support that take underlying gene variation into account (gCF; Q1, first quartet from coalescent-based analysis) and single-nucleotide polymorphisms (SNPs) (sCF) are more informative when assessing branch support from genomic datasets.[67]

Low backbone support in a phylogenetic tree can have several causes: (1) biological processes within the family, such as gene duplication and loss, hybridization, incomplete lineage sorting, a rapid radiation, and gene saturation, (2) one or more artifacts (e.g., deficient or erroneous data such as paralogs interpreted as orthologs), or (3) a combination thereof.[68] As gene saturation is a process that increases over time, we believe that it is not influencing our dataset to a large extent because node support among Brassicales families is generally higher than support for main lineages (subfamilies and supertribes) within Brassicaceae (Figure S4). Our dataset clearly indicates, however, that hybridization events have occurred frequently throughout Brassicaceae evolution. We found LH and AD to be very high in many species, with mean values of 75.1% and 2.25%, respectively (Data S1E and S1I). In comparison, these values were only 52.3% and 0.21%, respectively, across all non-hybrid natural accessions of pitcher plants (*Nepenthes* spp.) and 89.4% and 0.86%, respectively, in their known and suspected hybrids.[47] This highlights the abundant presence of hybrids in Brassicaceae, such as in tribes Alysseae, Brassiceae, Cardamineae, Lepideae, Microlepideae, and Thelypodieae, and rogue tribes Anastaticeae, Biscutelleae, Cochlearieae, Iberideae, and Megacarpaeeae (Figure 5; see next section).

Brassicaceae are notoriously known for their rampant hybridization (see below), and it is challenging to include all such evolutionary oddballs into a single evolutionary model to recover the family's "true" systematic relationships. A network approach could potentially represent these relationships more faithfully[69] (Figure 1D). Aware of the issues with data deficiency and paralogs, we specifically designed our analyses to disentangle some of the deeper nodes by stepwise removal of putative paralogous genes and samples of a putative hybrid origin ("inclusive," "strict," "superstrict," "superstrict by tribe," and "superstrict by tribe excluding hybrids" routines; Table 1; see STAR Methods for details). A recent alternative approach to exclude paralogs using synteny between whole-genome sequences[70] showed similar differences between tree topologies at the deeper nodes as those among our inclusive and stricter datasets (Figures 1A and 1B). Specifically, the branching order of clades I–IV based exclusively on strict syntenic orthologs was identical to the branching order recovered in our stricter datasets (Figure 1A), lending independent support for the importance of our substantial paralog filtering routines.

Interestingly, we found that going from the strict routine (removing only outlier genes and keeping 97.8% of all genes; Table 1) toward the superstrict routine (removing all genes with a mean SNP proportion >0.02, keeping only 28.3% of all genes), the topology of our BrassiToL from the coalescent-based approach hardly changed (Figure S3). We therefore believe that the superstrict routine (Figure 2) provides the best estimate of the phylogeny, as it includes most samples and at the same time excludes any "unnecessary" genes.

Rogue taxa share a mesopolyploid origin

Our results corroborate with previous studies that highlighted several difficult-to-place or rogue taxa,[3] including tribes

Anastaticeae, Biscutelleae, Cochlearieae, and Megacarpaeeae, and the genera *Iberis*, *Idahoa*, and *Subularia*. Contrary to the bulk of the tribes, which were consistently placed within a main lineage in our different nuclear phylogenies, these taxa "jumped" positions across the different phylogenies (Figure S3). This has previously been interpreted as the result of inter-tribal or inter-lineage hybridizations between distantly related parental species (genus- or tribe-specific meso-polyploidizations), as explained in the next paragraph.

The genus *Brassica* and the tribe Brassiceae were the first Brassicaceae taxa reported with a WGD postdating the family-specific palaeotetraploidization At-α event.[71–73] Since this pioneering work, more than a dozen genus- and tribe-specific mesopolyploid WGDs have been discovered throughout the family.[9,39,43,74] Because most mesopolyploid taxa have an allopolyploid origin arising from hybridization between distantly related parental species,[42,43,75] inferring their phylogenetic position within the family tree is often a challenging endeavor. For all the rogue taxa we identified, a mesopolyploid origin has indeed been demonstrated or claimed. For instance, for tribe Cochlearieae, whole-genome triplication has been demonstrated, whereas Anastaticeae and *Iberis* have a mesotetraploid origin.[39] Biscutelleae harbor four different WGDs specific to *Biscutella*, *Heldreichia*, *Lunaria*, and *Ricotia*.[43,76] Both genera of Megacarpaeeae (*Megacarpaea* and *Pugionium*) have been shown to be formed by independent WGDs.[77,78] *Teesdalia*, based on available chromosome numbers (2n = 36 in *T. coronopifolia* and *T. nudicaulis*, 2n = 20 in *T. conferta*), most likely has a polyploid origin, but (cyto)genomic data are lacking. Interestingly, based on our results, not all mesopolyploids act as rogue taxa, including tribe Brassiceae.

Brassicaceae originated during Earth's icehouse era

Our results suggest a somewhat younger age for the Brassicaceae than previously published (Figure 2; Table 2). We recovered a middle to late Eocene stem age for the mustard family (38.8 Ma; 95% HPD: 40.5–36.9), with a late Oligocene crown age (24.5 Ma; 95% HPD: 25.7–23.1). Although the family's crown node age estimates ranged from 54.3[19] to 15.0 Ma[18] in the earlier studies, results from more recent Brassicaceae studies converged to a crown age of 37.1–32.4 Ma, which appears insensitive to data type (nuclear/plastome), methods, and fossils used.[23]

Genomic datasets like ours are notoriously difficult to use in time calibration due to the sheer amount of data that needs to be analyzed simultaneously.[85] There is no consensus about the best approach to retrieve reliable dating estimates based on big genomic datasets, but our experience is that including a subset of only clock-like genes yields a more reliable time-calibrated tree compared with a more inclusive dataset (including, e.g., all genes), which commonly results in a forward shift in time.

Based on our time-calibrated analyses using the 20 most clock-like genes, the family's origin and the onset of its diversification coincide with the cooling of the Earth during the Eocene-Oligocene transition (so-called greenhouse to icehouse transition[86,87]). This period was characterized by a worldwide replacement of tropical forests with temperate forests, open vegetation, and deserts, which are all typical habitats of extant Brassicaceae. Tribe Aethionemeae and the five supertribes in subfamily Brassicoideae originated quickly after in the early Miocene (median stem ages range from 21.2 to 19.8 Ma; median crown ages range from 19.9 to 14.4 Ma; Data S1F). A combination of short branch lengths and low support in the BrassiToL around these events (Figure S4) supports the idea of an early rapid radiation of the family.

Results from our plastome phylogeny generally show a ~5 Ma forward shift in time relative to our nuclear phylogeny, possibly the result of including all 60 plastome genes in our study, compared with a subset of 20 clock-like genes in our nuclear analysis (Figure 3; Table 2). Median crown age of the family was 20.2 Ma (vs. 24.5 Ma in nuclear analysis), and median crown age of core Brassicaceae 16.9 Ma (vs. 21.1 Ma). Results from our plastome study are also much younger (nearly 10 Ma) than previously found by Walden et al.[2] based on partly the same dataset (but a different set of fossils used for calibration; Table 2).

Improved Brassicaceae phylogeny warrants updated family classification

Our study provides new systematic insights and consolidates results from recent phylogenetic studies in Brassicaceae, fully supporting the new family classification of German et al.[24] This includes the formal definition of two subfamilies (Aethionemoideae, including only *Aethionema*, and Brassicoideae), five supertribes (Camelinodae, previously lineage I; Brassicodae, lineage II; Hesperodae, lineage III; Arabodae, lineage IV; and Heliophilodae, lineage V), and 58 tribes.

At the tribal level, our results confirm that the scientifically important genus *Arabidopsis* is not closely related to any other genera traditionally placed in tribe Camelineae,[3,22] supporting the movement of *Arabidopsis* into a new monogeneric tribe, Arabidopsideae (supertribe Camelinodae[24]). Our results also support the following taxonomic changes as recently published by German et al.[24]: (1) erecting the monospecific and distinct genus *Asperuginoides* to its own new tribe, Asperuginoideae, (2) combining the genera *Dipoma* and *Hemilophia* in the new tribe Hemilophieae, (3) combining the genera *Idahoa* and *Subularia* in the re-established tribe Subularieae (plastome phylogeny only), and (4) raising the monospecific and distinct genus *Schrenkiella* into its own tribe, Schrenkielleae. We found all tribes to be monophyletic and highly supported in both nuclear and plastome phylogenies, except for Camelineae (polyphyletic in both nuclear and plastome phylogenies), Iberideae (polyphyletic in the nuclear phylogeny), and Subularieae (polyphyletic in the nuclear phylogeny). These tribes, in addition to the aforementioned rogue and hybrid tribes, require further systematic studies to uncover their exact evolutionary history and determine their taxonomic status.

Conclusions

We provide the first global, calibrated, nuclear, and plastome BrassiToL, offering an important step forward in untangling the notoriously difficult phylogenetic relationships across Brassicaceae. We applied the latest bioinformatic tools to select the most reliable genes to construct the species tree, thereby increasing topological accuracy and highlighting likely polyploid taxa. Our improved phylogenetic framework supports the reinstatement of the new family classification by German et al.,[24] including two subfamilies (Aethionemoideae and Brassicoideae),

Table 2. Brassicaceae divergence time estimates

Studies	Crown Brassicaceae (Ma)	Crown core Brassicaceae[a] (Ma)	Method	Dataset	Calibration
Koch et al.[79]	–	25.9–23.1	synonymous substitution rate	*Adh* and *Chs*	synonymous substitution rate
Franzke et al.[18]	35.0–15.0–1.0	28.0–11.0–1.0	BEAST	*nad4*	one secondary calibration
Beilstein et al.[19]	64.2–54.3–45.2	54.3–46.9–39.4	BEAST	*ndhF* and *PHYA*	four fossils
Couvreur et al.[20]	49.4–37.6–24.2	43.8–32.3–20.9	BEAST	eight genes from nuclei, chloroplast, and mitochondria	one fossil
Kagale et al.[80]	–	26.6	synonymous substitution rate	213 nuclear orthologs	synonymous substitution rate
Edger et al.[11]	45.9–31.8–16.8	–	BEAST	1,115 single-copy nuclear genes	two fossils
Hohmann et al.[9]	38.6–32.4–27.1	27.3–23.4–19.9	BEAST	plastomes	four fossils
Huang et al.[22]	37.8–37.1–36.3	30.3–29.7–29.1	r8s	113 low-copy nuclear orthologs	18 fossils[b]
Cardinal-McTeague et al.[54]	44.1–37.7–31.4	–	BEAST	Chloroplast DNA (ndhF, matK, rbcL) and mitochondrial DNA (matR, rps3)	three fossils[b]
Mohammadin et al.[81]	58.9–48.0–37.5	35.4	BEAST	plastomes	one secondary calibration
Guo et al.[82]	41.8–34.9–29.0	29.8–25.1–21.3	MCMCTree	plastomes	14 fossils[c]
Mandáková et al.[39]	54.7–40.1–29.4	30.6	BEAST	plastomes	four fossils
Huang et al.[23]	33.2–29.9–26.8	22.9–21.3–19.6	BEAST	plastomes	four fossils
Ramírez-Barahona et al.[83]	52.7–41.9–30.5	–	BEAST	rbcL, atpB, matK, ndhF, 18S, 26S, 5.8S	238 fossils, angiosperm-wide
Walden et al.[2]	35.7–29.9–24.3	29.6–25.1–20.9	BEAST	plastomes	four fossils
Legalov et al.[84]	–	>36.4	indirect calibration via phytophagous beetles	–	*Ceutorhynchus* beetle fossils
This study (nuclear dataset)	25.7–24.5–23.1	22.4–21.1–19.9	treePL	phylogeny using superstrict routine; calibration using subset of 20 most clock-like genes	one fossil validated by nine secondary calibration points
This study (plastome dataset)	29.0–20.2–13.0	24.3–16.9–10.2	treePL	plastomes	one fossil validated by eight secondary calibration points

Comparison of estimates from past studies and the current study (taken and updated from Huang et al.[23]).
[a]Core Brassicaceae are all Brassicaceae, excluding basal tribe Aethionemeae.
[b]Dating results excluding *Thlaspi primaevum*.
[c]Only results that exclude Brassicales fossils.

with five supertribes in the latter (Arabodae, Brassicodae, Camelinodae, Heliophilodae, and Hesperodae), and a total of 58 tribes of which five newly described or re-established.

The ultimate goal of our Brassicaceae consortium is to build a complete Brassicaceae family phylogeny including all ~4,000 species. The methods that we applied (including our "mixed baits" target capture sequencing from—sometimes over 200 years old—herbarium material[46]) have proven successful and are becoming more affordable (now roughly € 40 per sample) and can easily be scaled up. Such a new species-level phylogeny will further boost the significance of this model plant family—for both fundamental and applied research programs covering all fields of plant biology.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead contact
  - Materials availability
  - Data and code availability
- EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS
- METHOD DETAILS

- Taxon sampling
- Library preparation, target capture, and sequencing
- Sequence assembly of target capture data
- Sequence assembly of plastome
- Taxonomic verification
- Allelic variation and paralog detection
- Nuclear phylogenomics
- Phylogenetic informativeness
- Plastome phylogenetics
- Fossil calibration

● QUANTIFICATION AND STATISTICAL ANALYSIS

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.cub.2023.08.026.

ACKNOWLEDGMENTS

We thank Helen Barnes, Roxali Bijmoer, Manuel Benito Crespo, Ivalu Cacho, Cyrille Chatelain, Suzanne Cubey, Gabi Droege, Catherine Gallagher, Mary Korver, Alicia Marticorena, Pina Milne, Hamid Moazzeni, Peter Sack, Yasaman Salmaki, Marnel Scherrenberg, Jan Wieringa, Hasan Yıldırım, and Shahin Zarre for assistance in sampling herbarium material. Elza Duijm, Izai Sabino Kikuchi, and Marina Ventayol Garcia kindly provided fruitful discussions on optimization of laboratory techniques. David Alejandro Duchene Garzon provided valuable ideas on calculations of clock-likeness of gene trees. Muséum national d'Histoire naturelle kindly provided V.R.I. access to the collections in the framework of the RECOLNAT national Research Infrastructure. This work was supported by the German Research Foundation (DFG; grant numbers MU1137/17-1 to K.M. and KO2302/23-2 to M.A.K.), the Czech Science Foundation (grant numbers 21-03909S to M.A.L. and 21-06839S to T.M.), grants from the Calleva Foundation to the Plant and Fungal Trees of Life project at the Royal Botanic Gardens, Kew, and CONICYT PAI Subvención, la Instalación, en La Academia Convocatoria 2019 (grant number n°77190055 to Ó.T.-N.). This work used the Extreme Science and Engineering Discovery Environment (XSEDE, project BIO210079), which is supported by the National Science Foundation (grant number ACI-1548562).

AUTHOR CONTRIBUTIONS

F.L., K.M., and K.P.H. designed the study. F.L., K.M., K.P.H., I.A.A.-S., C.D.B., C.K., A.H.v.H., D.A.G., M.A.K., Ó.T.-N., B.Ö., V.R.I., O.M., N.M.H., M.T., M.D.W., I.R., S.Š., B.N., R.V., C.B., H.R., S.B.J., M.S., A.F., A.G., S.Z., B.J.L., N.S., F.W.S., I.S., P.H., W.J.B., and F.F. contributed to the sampling process. K.P.H., A.H.v.H., C.K., N.W., C.D.B., and A.R.Z. performed laboratory work. K.P.H., C.K., L.N., N.M.H., A.R.Z., and E.L. performed the various analyses. K.P.H., F.L., C.K., K.M., W.J.B., I.A.A.-S., C.D.B., A.H.v.H., L.N., D.A.G., M.A.K., M.A.L., and N.W. took the lead in writing the manuscript, with subsequent input from all co-authors.

DECLARATION OF INTERESTS

The authors declare no competing interests.

INCLUSION AND DIVERSITY

We support inclusive, diverse, and equitable conduct of research.

Received: February 3, 2023
Revised: June 22, 2023
Accepted: August 8, 2023
Published: September 1, 2023

REFERENCES

1. Franzke, A., Lysak, M.A., Al-Shehbaz, I.A., Koch, M.A., and Mummenhoff, K. (2011). Cabbage family affairs: the evolutionary history of Brassicaceae. Trends Plant Sci. *16*, 108–116. https://doi.org/10.1016/j.tplants.2010.11.005.
2. Walden, N., German, D.A., Wolf, E.M., Kiefer, M., Rigault, P., Huang, X.C., Kiefer, C., Schmickl, R., Franzke, A., Neuffer, B., et al. (2020). Nested whole-genome duplications coincide with diversification and high morphological disparity in Brassicaceae. Nat. Commun. *11*, 3795, https://doi.org/10.1038/s41467-020-17605-7.
3. Nikolov, L.A., Shushkov, P., Nevado, B., Gan, X., Al-Shehbaz, I.A., Filatov, D., Bailey, C.D., and Tsiantis, M. (2019). Resolving the backbone of the Brassicaceae phylogeny for investigating trait diversity. New Phytol. *222*, 1638–1651. https://doi.org/10.1111/nph.15732.
4. Nikolov, L.A. (2019). Brassicaceae Flowers: diversity amid Uniformity. J. Exp. Bot. *70*, 2623–2635. https://doi.org/10.1093/jxb/erz079.
5. Mummenhoff, K., Al-Shehbaz, I.A., Bakker, F.T., Linder, H.P., and Mühlhausen, A. (2005). Phylogeny, morphological evolution, and speciation of endemic Brassicaceae genera in the Cape Flora of Southern Africa. Ann. Mo. Bot. Gard. *92*, 400–424.
6. Warwick, S.I. (2011). Brassicaceae in Agriculture. In Genetics and Genomics of the Brassicaceae Plant Genetics and Genomics: Crops and Models, R. Schmidt, and I. Bancroft, eds. (Springer), pp. 33–65. https://doi.org/10.1007/978-1-4419-7118-0_2.
7. Nikolov, L.A., and Tsiantis, M. (2017). Using mustard genomes to explore the genetic basis of evolutionary change. Curr. Opin. Plant Biol. *36*, 119–128. https://doi.org/10.1016/j.pbi.2017.02.005.
8. Rahimi, A., Karami, O., Lestari, A.D., de Werk, T., Amakorová, P., Shi, D., Novák, O., Greb, T., and Offringa, R. (2022). Control of cambium initiation and activity in Arabidopsis by the transcriptional regulator AHL15. Curr. Biol. *32*, 1764–1775.e3. https://doi.org/10.1016/j.cub.2022.02.060.
9. Hohmann, N., Wolf, E.M., Lysak, M.A., and Koch, M.A. (2015). A time-calibrated road map of Brassicaceae species radiation and evolutionary history. Plant Cell *27*, 2770–2784. https://doi.org/10.1105/tpc.15.00482.
10. Mandáková, T., and Lysak, M.A. (2018). Post-polyploid diploidization and diversification through Dysploid changes. Curr. Opin. Plant Biol. *42*, 55–65. https://doi.org/10.1016/j.pbi.2018.03.001.
11. Edger, P.P., Heidel-Fischer, H.M., Bekaert, M., Rota, J., Glöckner, G., Platts, A.E., Heckel, D.G., Der, J.P., Wafula, E.K., Tang, M., et al. (2015). The butterfly plant arms-race escalated by gene and genome duplications. Proc. Natl. Acad. Sci. USA *112*, 8362–8366. https://doi.org/10.1073/pnas.1503926112.
12. Koch, M.A., German, D.A., Kiefer, M., and Franzke, A. (2018). Database Taxonomics as Key to modern plant biology. Trends Plant Sci. *23*, 4–6. https://doi.org/10.1016/j.tplants.2017.10.005.
13. Castañeda-Álvarez, N.P., Khoury, C.K., Achicanoy, H.A., Bernau, V., Dempewolf, H., Eastwood, R.J., Guarino, L., Harker, R.H., Jarvis, A., Maxted, N., et al. (2016). Global conservation priorities for crop wild relatives. Nat. Plants *2*, 16022, https://doi.org/10.1038/nplants.2016.22.
14. Castillo-Lorenzo, E., Finch-Savage, W.E., Seal, C.E., and Pritchard, H.W. (2019). Adaptive significance of functional germination traits in crop wild relatives of *Brassica*. Agric. For. Meteorol. *264*, 343–350. https://doi.org/10.1016/j.agrformet.2018.10.014.
15. Quezada-Martinez, D., Addo Nyarko, C.P., Schiessl, S.V., and Mason, A.S. (2021). Using wild relatives and related species to build climate resilience in *Brassica* Crops. Theor. Appl. Genet. *134*, 1711–1728. https://doi.org/10.1007/s00122-021-03793-3.
16. Al-Shehbaz, I.A., Beilstein, M.A., and Kellogg, E.A. (2006). Systematics and phylogeny of the Brassicaceae (Cruciferae): an overview. Plant Syst. Evol. *259*, 89–120. https://doi.org/10.1007/s00606-006-0415-z.
17. Bailey, C.D., Koch, M.A., Mayer, M., Mummenhoff, K., O'Kane, S.L., Jr., Warwick, S.I., Windham, M.D., and Al-Shehbaz, I.A. (2006). Toward a

global phylogeny of the Brassicaceae. Mol. Biol. Evol. *23*, 2142–2160. https://doi.org/10.1093/molbev/msl087.

18. Franzke, A., German, D., Al-Shehbaz, I.A., and Mummenhoff, K. (2009). *Arabidopsis* Family ties: molecular phylogeny and age estimates in Brassicaceae. Taxon *58*, 425–437. https://doi.org/10.1002/tax.582009.
19. Beilstein, M.A., Nagalingum, N.S., Clements, M.D., Manchester, S.R., and Mathews, S. (2010). Dated molecular phylogenies indicate a Miocene origin for *Arabidopsis thaliana*. Proc. Natl. Acad. Sci. USA *107*, 18724–18728. https://doi.org/10.1073/pnas.0909766107.
20. Couvreur, T.L.P., Franzke, A., Al-Shehbaz, I.A., Bakker, F.T., Koch, M.A., and Mummenhoff, K. (2010). Molecular phylogenetics, temporal diversification, and principles of evolution in the mustard family (Brassicaceae). Mol. Biol. Evol. *27*, 55–71. https://doi.org/10.1093/molbev/msp202.
21. Warwick, S.I., Mummenhoff, K., Sauder, C.A., Koch, M.A., and Al-Shehbaz, I.A. (2010). Closing the gaps: phylogenetic relationships in the Brassicaceae Based on DNA sequence data of nuclear ribosomal ITS Region. Plant Syst. Evol. *285*, 209–232. https://doi.org/10.1007/s00606-010-0271-8.
22. Huang, C.-H., Sun, R., Hu, Y., Zeng, L., Zhang, N., Cai, L., Zhang, Q., Koch, M.A., Al-Shehbaz, I., Edger, P.P., et al. (2016). Resolution of Brassicaceae phylogeny using nuclear genes uncovers nested radiations and supports convergent morphological evolution. Mol. Biol. Evol. *33*, 394–412. https://doi.org/10.1093/molbev/msv226.
23. Huang, X.-C., German, D.A., and Koch, M.A. (2020). Temporal patterns of diversification in Brassicaceae demonstrate decoupling of rate shifts and Mesopolyploidization events. Ann. Bot. *125*, 29–47. https://doi.org/10.1093/aob/mcz123.
24. German, D.A., Hendriks, K.P., Koch, M.A., Lens, F., Lysak, M.A., Bailey, C.D., Mummenhoff, K., and Al-Shehbaz, I.A. (2023). An updated classification of the Brassicaceae (Cruciferae). PhytoKeys *220*, 127–144. https://doi.org/10.3897/phytokeys.220.97724.
25. von Hayek, A. (1911). Entwurf eines Cruciferen-Systems auf Phylogenetischer Grundlage (Heinrich).
26. Schulz, O.E. (1919). Cruciferae-Brassiceae. Part 1. In Pflanzenreich IV. 105 (Heft 70), A. Engler, ed. (Verlag von Wilhelm Engelmann), pp. 1–290.
27. Janchen, E. (1942). Das System der Cruciferen. Österr. Bot. Z. *91*, 1–28.
28. Koch, M., Al-Shehbaz, I.A., and Mummenhoff, K. (2003). Molecular systematics, evolution, and population biology in the mustard family (Brassicaceae). Ann. Mo. Bot. Gard. *90*, 151–171. https://doi.org/10.2307/3298580.
29. Koch, M.A., and Mummenhoff, K. (2006). Editorial: Evolution and phylogeny of the Brassicaceae. Editorial. Plant Syst. Evol. *259*, 81–83.
30. Mitchell-Olds, T., Al-Shehbaz, I.A., Koch, M., and Sharbel, T.F. (2005). Crucifer evolution in the post-genomic era. In Plant Diversity and Evolution: Genotypic and Phenotypic Variation in Higher Plants (CAB International), pp. 119–137.
31. Koch, M., Haubold, B., and Mitchell-Olds, T. (2001). Molecular systematics of the Brassicaceae: evidence from coding plastidic matK and nuclear Chs sequences. Am. J. Bot. *88*, 534–544. https://doi.org/10.2307/2657117.
32. Koch, M.A., and Al-Shehbaz, I.A. (2009). Molecular systematics and evolution of "wild" crucifers (Brassicaceae or Cruciferae). Biology and breeding of crucifers. In Biology and Breeding of Crucifers, S.K. Gupta, ed. (Taylor and Francis Group).
33. Mabry, M.E., Brose, J.M., Blischak, P.D., Sutherland, B., Dismukes, W.T., Bottoms, C.A., Edger, P.P., Washburn, J.D., An, H., Hall, J.C., et al. (2020). Phylogeny and multiple independent whole-genome duplication events in the Brassicales. Am. J. Bot. *107*, 1148–1164. https://doi.org/10.1002/ajb2.1514.
34. Beric, A., Mabry, M.E., Harkess, A.E., Brose, J., Schranz, M.E., Conant, G.C., Edger, P.P., Meyers, B.C., and Pires, J.C. (2021). Comparative phylogenetics of repetitive elements in a diverse order of flowering plants (Brassicales). G3 (Bethesda) *11*, jkab140, https://doi.org/10.1093/g3journal/jkab140.
35. Liu, L.-M., Du, X.-Y., Guo, C., and Li, D.-Z. (2021). Resolving robust phylogenetic relationships of core Brassicaceae using genome skimming data. J. Syst. Evol. *59*, 442–453.
36. Maddison, W.P. (1997). Gene trees in species trees. Syst. Biol. *46*, 523–536. https://doi.org/10.1093/sysbio/46.3.523.
37. Yue, J.-P., Sun, H., Baum, D.A., Li, J.-H., Al-Shehbaz, I.A., and Ree, R. (2009). Molecular phylogeny of *Solms-laubachia* (Brassicaceae) s.l., based on multiple nuclear and plastid DNA sequences, and its biogeographic implications. J. Syst. Evol. *47*, 402–415. https://doi.org/10.1111/j.1759-6831.2009.00041.x.
38. German, D.A., Grant, J.R., Lysak, M.A., and Al-Shehbaz, I.A. (2011). Molecular phylogeny and systematics of the tribe Chorisporeae (Brassicaceae). Plant Syst. Evol. *294*, 65–86. https://doi.org/10.1007/s00606-011-0452-0.
39. Mandáková, T., Li, Z., Barker, M.S., and Lysak, M.A. (2017). Diverse genome organization following 13 independent mesopolyploid events in Brassicaceae contrasts with convergent patterns of gene retention. Plant J. *91*, 3–21. https://doi.org/10.1111/tpj.13553.
40. Dogan, M., Pouch, M., Mandáková, T., Hloušková, P., Guo, X., Winter, P., Chumová, Z., Van Niekerk, A., Mummenhoff, K., Al-Shehbaz, I.A., et al. (2020). Evolution of tandem repeats is mirroring post-polyploid cladogenesis in *Heliophila* (Brassicaceae). Front. Plant Sci. *11*, 607893.
41. Forsythe, E.S., Nelson, A.D.L., and Beilstein, M.A. (2020). Biased gene retention in the face of introgression obscures species relationships. Genome Biol. Evol. *12*, 1646–1663. https://doi.org/10.1093/gbe/evaa149.
42. Mandáková, T., Pouch, M., Harmanová, K., Zhan, S.H., Mayrose, I., and Lysak, M.A. (2017). Multispeed genome diploidization and diversification after an ancient allopolyploidization. Mol. Ecol. *26*, 6445–6462. https://doi.org/10.1111/mec.14379.
43. Guo, X., Mandáková, T., Trachtová, K., Özüdoğru, B., Liu, J., and Lysak, M.A. (2021). Linked by ancestral bonds: multiple whole-genome duplications and reticulate evolution in a Brassicaceae Tribe. Mol. Biol. Evol. *38*, 1695–1714. https://doi.org/10.1093/molbev/msaa327.
44. Baker, W.J., Dodsworth, S., Forest, F., Graham, S.W., Johnson, M.G., McDonnell, A., Pokorny, L., Tate, J.A., Wicke, S., and Wickett, N.J. (2021). Exploring Angiosperms353: an open, community toolkit for collaborative phylogenomic research on flowering plants. Am. J. Bot. *108*, 1059–1065. https://doi.org/10.1002/ajb2.1703.
45. Johnson, M.G., Pokorny, L., Dodsworth, S., Botigué, L.R., Cowan, R.S., Devault, A., Eiserhardt, W.L., Epitawalage, N., Forest, F., Kim, J.T., et al. (2019). A universal probe set for targeted sequencing of 353 nuclear genes from any flowering plant designed using k-Medoids clustering. Syst. Biol. *68*, 594–606. https://doi.org/10.1093/sysbio/syy086.
46. Hendriks, K.P., Mandáková, T., Hay, N.M., Ly, E., Hooft van Huysduynen, A.H. van, Tamrakar, R., Thomas, S.K., Toro-Núñez, O., Pires, J.C., Nikolov, L.A., et al. (2021). The best of both worlds: combining lineage-specific and universal bait sets in target-enrichment hybridization reactions. Appl. Plant Sci. *9*, e11438, https://doi.org/10.1002/aps3.11438.
47. Nauheimer, L., Weigner, N., Joyce, E., Crayn, D., Clarke, C., and Nargar, K. (2021). HybPhaser: A workflow for the detection and phasing of hybrids in target capture data sets. Appl. Plant Sci. *9*, https://doi.org/10.1002/aps3.11441.
48. Zhang, C., Rabiee, M., Sayyari, E., and Mirarab, S. (2018). ASTRAL-III: polynomial time species tree reconstruction from partially resolved gene trees. BMC Bioinformatics *19*, 153, https://doi.org/10.1186/s12859-018-2129-y.
49. Zhang, C., Scornavacca, C., Molloy, E.K., and Mirarab, S. (2020). ASTRAL-Pro: quartet-based species-tree inference despite paralogy. Mol. Biol. Evol. *37*, 3292–3307. https://doi.org/10.1093/molbev/msaa139.
50. Townsend, J.P. (2007). Profiling phylogenetic informativeness. Syst. Biol. *56*, 222–231. https://doi.org/10.1080/10635150701311362.

51. Gandolfo, M.A., Nixon, K.C., and Crepet, W.L. (1998). A new fossil flower from the Turonian of New Jersey: *Dressiantha bicarpellata* gen. et sp. nov. (Capparales). Am. J. Bot. *85*, 964, https://doi.org/10.2307/2446363.

52. Franzke, A., Koch, M.A., and Mummenhoff, K. (2016). Turnip time travels: age estimates in Brassicaceae. Trends Plant Sci. *21*, 554–561. https://doi.org/10.1016/j.tplants.2016.01.024.

53. Selmeier, A. (2005). *Capparidoxylon holleisii* nov. spec. a Silicified Capparis (Capparaceae) Wood with Insect Coprolites from the Neogene of Southern Germany. Zitteliana, 199–209.

54. Cardinal-McTeague, W.M., Sytsma, K.J., and Hall, J.C. (2016). Biogeography and diversification of Brassicales: A 103 million year tale. Mol. Phylogenet. Evol. *99*, 204–224. https://doi.org/10.1016/j.ympev.2016.02.021.

55. Özüdoğru, B., Akaydın, G., Erik, S., Al-Shehbaz, I.A., and Mummenhoff, K. (2015). Phylogeny, diversification and biogeographic implications of the eastern Mediterranean endemic genus *Ricotia* (Brassicaceae). Taxon *64*, 727–740. https://doi.org/10.12705/644.5.

56. Karl, R., and Koch, M.A. (2013). A world-wide perspective on crucifer speciation and evolution: phylogenetics, biogeography and trait evolution in tribe Arabideae. Ann. Bot. *112*, 983–1001. https://doi.org/10.1093/aob/mct165.

57. Heenan, P.B., and Mitchell, A.D. (2003). Phylogeny, biogeography and adaptive radiation of *Pachycladon* (Brassicaceae) in the mountains of South Island, New Zealand. J. Biogeogr. *30*, 1737–1749. https://doi.org/10.1046/j.1365-2699.2003.00941.x.

58. Heenan, P.B., and McGlone, M.S. (2013). Evolution of New Zealand alpine and open-habitat plant species during the Late Cenozoic. N. Z. J. Ecol. *37*, 105–113.

59. Romero, E.J., and Hickey, L.J. (1976). A fossil leaf of Akaniaceae from Paleocene beds in Argentina. Bull. Torrey Bot. Club *103*, 126–131.

60. Iglesias, A., Wilf, P., Johnson, K.R., Zamuner, A.B., Cúneo, N.R., Matheos, S.D., and Singer, B.S. (2007). A Paleocene lowland macroflora from Patagonia reveals significantly greater richness than North American analogs. Geology *35*, 947–950. https://doi.org/10.1130/G23889A.1.

61. Chandler, M. (1962). Flora of the Pipe-Clay Series of Dorset (lower Bagshot). The Lower Tertiary floras of Southern England, II (British Museum (Natural History)), pp. 1–176.

62. Franzke, A., Hurka, H., Janssen, D., Neuffer, B., Friesen, N., Markov, M., and Mummenhoff, K. (2004). Molecular signals for Late Tertiary/Early Quaternary range splits of an Eurasian steppe plant: *Clausia aprica* (Brassicaceae). Mol. Ecol. *13*, 2789–2795. https://doi.org/10.1111/j.1365-294X.2004.02272.x.

63. Mummenhoff, K., Brüggemann, H., and Bowman, J.L. (2001). Chloroplast DNA phylogeny and biogeography of *Lepidium* (Brassicaceae). Am. J. Bot. *88*, 2051–2063. https://doi.org/10.2307/3558431.

64. Becker, H.F. (1961). Oligocene plants from the upper ruby river basin, southwestern Montana*82*, pp. 1–122.

65. Lielke, K., Manchester, S., and Meyer, H. (2012). Reconstructing the environment of the northern Rocky Mountains during the Eocene/Oligocene transition: constraints from the palaeobotany and geology of south-western Montana, USA. Acta Palaeobot. *52*, 317–358.

66. Esmailbegi, S., Al-Shehbaz, I.A., Pouch, M., Mandáková, T., Mummenhoff, K., Rahiminejad, M.R., Mirtadzadini, M., and Lysak, M.A. (2018). Phylogeny and systematics of the tribe Thlaspideae (Brassicaceae) and the recognition of two new genera. Taxon *67*, 324–340. https://doi.org/10.12705/672.4.

67. Minh, B.Q., Hahn, M.W., and Lanfear, R. (2020). New methods to calculate concordance factors for phylogenomic datasets. Mol. Biol. Evol. *37*, 2727–2733. https://doi.org/10.1093/molbev/msaa106.

68. Boussau, B., and Scornavacca, C. (2020). Reconciling gene trees with species trees. In Phylogenetics in the Genomic Era, C. Scornavacca, F. Delsuc, and N. Galtier, eds. (No commercial publisher|Authors open access book), pp. 3.2:1–3.2:23.

69. Huson, D.H., and Bryant, D. (2006). Application of phylogenetic networks in evolutionary studies. Mol. Biol. Evol. *23*, 254–267. https://doi.org/10.1093/molbev/msj030.

70. Walden, N., and Schranz, M.E. (2023). Synteny identifies reliable orthologs for phylogenomics and comparative genomics of the Brassicaceae. Genome Biol. Evol. *15*, evad034, https://doi.org/10.1093/gbe/evad034.

71. Lysak, M.A., Koch, M.A., Pecinka, A., and Schubert, I. (2005). Chromosome triplication found across the tribe Brassiceae. Genome Res. *15*, 516–525. https://doi.org/10.1101/gr.3531105.

72. Parkin, I.A.P., Gulden, S.M., Sharpe, A.G., Lukens, L., Trick, M., Osborn, T.C., and Lydiate, D.J. (2005). Segmental structure of the *Brassica napus* genome based on comparative analysis with *Arabidopsis thaliana*. Genetics *171*, 765–781. https://doi.org/10.1534/genetics.105.042093.

73. Lysak, M.A., Cheung, K., Kitschke, M., and Bureš, P. (2007). Ancestral chromosomal blocks are triplicated in Brassiceae species with varying chromosome number and genome size. Plant Physiol. *145*, 402–410. https://doi.org/10.1104/pp.107.104380.

74. Dogan, M., Mandáková, T., Guo, X., and Lysak, M.A. (2022). *Idahoa* and *Subularia*: hidden Polyploid Origins of Two Enigmatic Genera of Crucifers. Am. J. Bot. *109*, 1273–1289. https://doi.org/10.1002/ajb2.16042.

75. German, D.A., and Friesen, N.W. (2014). *Shehbazia* (Shehbazieae, Cruciferae), a new monotypic genus and tribe of hybrid origin from Tibet. Turczaninowia *17*, 17–23.

76. Mandáková, T., Guo, X., Özüdoğru, B., Mummenhoff, K., and Lysak, M.A. (2018). Hybridization-facilitated genome merger and repeated chromosome fusion after 8 million years. Plant J. *96*, 748–760. https://doi.org/10.1111/tpj.14065.

77. Yang, Q., Bi, H., Yang, W., Li, T., Jiang, J., Zhang, L., Liu, J., and Hu, Q. (2020). The genome sequence of alpine *Megacarpaea delavayi* identifies species-specific whole-genome duplication. Front. Genet. *11*, 812.

78. Hu, Q., Ma, Y., Mandáková, T., Shi, S., Chen, C., Sun, P., Zhang, L., Feng, L., Zheng, Y., Feng, X., et al. (2021). Genome evolution of the psammophyte *Pugionium* for desert adaptation and further speciation. Proc. Natl. Acad. Sci. USA *118*, e2025711118, https://doi.org/10.1073/pnas.2025711118.

79. Koch, M.A., Haubold, B., and Mitchell-Olds, T. (2000). Comparative evolutionary analysis of chalcone synthase and alcohol dehydrogenase loci in *Arabidopsis*, *Arabis*, and related genera (Brassicaceae). Mol. Biol. Evol. *17*, 1483–1498. https://doi.org/10.1093/oxfordjournals.molbev.a026248.

80. Kagale, S., Robinson, S.J., Nixon, J., Xiao, R., Huebert, T., Condie, J., Kessler, D., Clarke, W.E., Edger, P.P., Links, M.G., et al. (2014). Polyploid evolution of the Brassicaceae during the Cenozoic era. Plant Cell *26*, 2777–2791. https://doi.org/10.1105/tpc.114.126391.

81. Mohammadin, S., Peterse, K., Kerke, S.J. van de, Chatrou, L.W., Dönmez, A.A., Mummenhoff, K., Pires, J.C., Edger, P.P., Al-Shehbaz, I.A., and Schranz, M.E. (2017). Anatolian origins and diversification of Aethionema, the sister lineage of the core Brassicaceae. Am. J. Bot. *104*, 1042–1054. https://doi.org/10.3732/ajb.1700091.

82. Guo, X., Liu, J., Hao, G., Zhang, L., Mao, K., Wang, X., Zhang, D., Ma, T., Hu, Q., Al-Shehbaz, I.A., et al. (2017). Plastome phylogeny and early diversification of Brassicaceae. BMC Genomics *18*, 176, https://doi.org/10.1186/s12864-017-3555-3.

83. Ramírez-Barahona, S., Sauquet, H., and Magallón, S. (2020). The delayed and geographically heterogeneous diversification of flowering plant families. Nat. Ecol. Evol. *4*, 1232–1238. https://doi.org/10.1038/s41559-020-1241-3.

84. Legalov, A.A., Nazarenko, V.Y., Vasilenko, D.V., and Perkovsky, E.E. (2022). *Ceutorhynchus* Germar (Coleoptera, Curculionidae) as proxy for Eocene core Brassicaceae: first record of the genus from rovno amber. J. Paleontol. *96*, 379–386. https://doi.org/10.1017/jpa.2021.82.

85. Smith, S.A., Brown, J.W., and Walker, J.F. (2018). So many genes, so Little Time: A practical approach to divergence-time estimation in the genomic era. PLoS One *13*, e0197433, https://doi.org/10.1371/journal.pone.0197433.

86. Zanazzi, A., Kohn, M.J., MacFadden, B.J., and Terry, D.O. (2007). Large temperature drop across the Eocene–Oligocene transition in central North America. Nature *445*, 639–642. https://doi.org/10.1038/nature05551.

87. Sun, J., Ni, X., Bi, S., Wu, W., Ye, J., Meng, J., and Windley, B.F. (2014). Synchronous turnover of Flora, fauna and climate at the Eocene–Oligocene boundary in Asia. Sci. Rep. *4*, 7463, https://doi.org/10.1038/srep07463.

88. Bolger, A.M., Lohse, M., and Usadel, B. (2014). Trimmomatic: A flexible trimmer for Illumina sequence data. Bioinformatics *30*, 2114–2120. https://doi.org/10.1093/bioinformatics/btu170.

89. Johnson, M.G., Gardner, E.M., Liu, Y., Medina, R., Goffinet, B., Shaw, A.J., Zerega, N.J.C., and Wickett, N.J. (2016). HybPiper: extracting coding sequence and introns for phylogenetics from high-throughput sequencing reads using target enrichment. Appl. Plant Sci. *4*, 1600016, https://doi.org/10.3732/apps.1600016.

90. Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with burrows–Wheeler transform. Bioinformatics *25*, 1754–1760. https://doi.org/10.1093/bioinformatics/btp324.

91. Bankevich, A., Nurk, S., Antipov, D., Gurevich, A.A., Dvorkin, M., Kulikov, A.S., Lesin, V.M., Nikolenko, S.I., Pham, S., Prjibelski, A.D., et al. (2012). SPAdes: A new genome assembly algorithm and its applications to single-cell sequencing. J. Comput. Biol. *19*, 455–477. https://doi.org/10.1089/cmb.2012.0021.

92. Tange, O. (2011). GNU parallel: the command-line power tool. USENIX Mag. *36*, 42–47.

93. McLay, T.G., Gunn, B.F., Ning, W., Tate, J.A., Nauheimer, L., Joyce, E.M., Simpson, L., Schmidt-Lebuhn, A.N., Baker, W.J., Forest, F., et al. (2020). New targets acquired: improving locus recovery from the Angiosperms353 probe set. https://doi.org/10.1101/2020.10.04.325571.

94. Danecek, P., Bonfield, J.K., Liddle, J., Marshall, J., Ohan, V., Pollard, M.O., Whitwham, A., Keane, T., McCarthy, S.A., Davies, R.M., et al. (2021). Twelve years of SAMtools and BCFtools. GigaScience *10*, giab008, https://doi.org/10.1093/gigascience/giab008.

95. McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimella, K., Altshuler, D., Gabriel, S., Daly, M., et al. (2010). The genome analysis toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res. *20*, 1297–1303. https://doi.org/10.1101/gr.107524.110.

96. Van der Auwera, G.A., and O'Connor, B.D. (2020). Genomics in the Cloud: Using Docker, GATK, and WDL in Terra (O'Reilly Media).

97. Quinlan, A.R., and Hall, I.M. (2010). BEDTools: A flexible suite of utilities for comparing genomic features. Bioinformatics *26*, 841–842. https://doi.org/10.1093/bioinformatics/btq033.

98. Katoh, K., and Standley, D.M. (2013). MAFFT Multiple Sequence Alignment, software version 7: improvements in performance and usability. Mol. Biol. Evol. *30*, 772–780.

99. Price, M.N., Dehal, P.S., and Arkin, A.P. (2010). FastTree 2 – Approximately maximum-likelihood trees for large alignments. PLoS One *5*, e9490, https://doi.org/10.1371/journal.pone.0009490.

100. Mirarab, S., Nguyen, N., Guo, S., Wang, L.S., Kim, J., and Warnow, T. (2015). PASTA: ultra-large multiple sequence alignment for nucleotide and amino-acid sequences. J. Comput. Biol. *22*, 377–386. https://doi.org/10.1089/cmb.2014.0156.

101. Minh, B.Q., Schmidt, H.A., Chernomor, O., Schrempf, D., Woodhams, M.D., von Haeseler, A., and Lanfear, R. (2020). IQ-TREE 2: New models and efficient methods for phylogenetic inference in the genomic era. Mol. Biol. Evol. *37*, 1530–1534. https://doi.org/10.1093/molbev/msaa015.

102. Capella-Gutiérrez, S., Silla-Martínez, J.M., and Gabaldón, T. (2009). trimal: A Tool for Automated Alignment Trimming in Large-Scale phylogenetic Analyses. Bioinformatics *25*, 1972–1973. https://doi.org/10.1093/bioinformatics/btp348.

103. Zhang, C., Zhao, Y., Braun, E.L., and Mirarab, S. (2021). TAPER: pinpointing errors in multiple sequence alignments despite varying rates of evolution. Methods Ecol. Evol. *12*, 2145–2158. https://doi.org/10.1111/2041-210X.13696.

104. Mai, U., and Mirarab, S. (2018). TreeShrink: fast and accurate detection of outlier long branches in collections of phylogenetic trees. BMC Genomics *19*, 272, https://doi.org/10.1186/s12864-018-4620-2.

105. Xu, S., Li, L., Luo, X., Chen, M., Tang, W., Zhan, L., Dai, Z., Lam, T.T., Guan, Y., and Yu, G. (2022). GGTREE: A serialized data object for visualization of a phylogenetic tree and annotation data. iMeta *n/a*, e56. iMeta *1*, https://doi.org/10.1002/imt2.56.

106. Mayrose, I., Graur, D., Ben-Tal, N., and Pupko, T. (2004). Comparison of site-specific rate-inference methods for protein sequences: empirical bayesian methods are superior. Mol. Biol. Evol. *21*, 1781–1791. https://doi.org/10.1093/molbev/msh194.

107. Dornburg, A., Fisk, J.N., Tamagnan, J., and Townsend, J.P. (2016). PhyInformR: phylogenetic experimental design and phylogenomic data exploration in R. BMC Evol. Biol. *16*, 262, https://doi.org/10.1186/s12862-016-0837-3.

108. Nguyen, L.-T., Schmidt, H.A., von Haeseler, A., and Minh, B.Q. (2015). IQ-TREE: A fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. Mol. Biol. Evol. *32*, 268–274. https://doi.org/10.1093/molbev/msu300.

109. Smith, S.A., and O'Meara, B.C. (2012). treePL: divergence time estimation using penalized likelihood for large phylogenies. Bioinformatics *28*, 2689–2690. https://doi.org/10.1093/bioinformatics/bts492.

110. Bouckaert, R., Heled, J., Kühnert, D., Vaughan, T., Wu, C.-H., Xie, D., Suchard, M.A., Rambaut, A., and Drummond, A.J. (2014). Beast2: A software platform for bayesian evolutionary analysis. PLoS Comput. Biol. *10*, e1003537.

111. Bouckaert, R., Vaughan, T.G., Barido-Sottani, J., Duchêne, S., Fourment, M., Gavryushkina, A., Heled, J., Jones, G., Kühnert, D., De Maio, N.D., et al. (2019). BEAST 2.5: an advanced software platform for bayesian evolutionary analysis. PLoS Comput. Biol. *15*, e1006650, https://doi.org/10.1371/journal.pcbi.1006650.

112. Kiefer, M., Schmickl, R., German, D.A., Mandáková, T., Lysak, M.A., Al-Shehbaz, I.A., Franzke, A., Mummenhoff, K., Stamatakis, A., and Koch, M.A. (2014). BrassiBase: introduction to a novel knowledge database on Brassicaceae evolution. Plant Cell Physiol. *55*, e3. https://doi.org/10.1093/pcp/pct158.

113. Weitemier, K., Straub, S.C.K., Cronn, R.C., Fishbein, M., Schmickl, R., McDonnell, A., and Liston, A. (2014). Hyb-Seq: combining target enrichment and genome skimming for plant phylogenomics. Appl. Plant Sci. *2*, 1400042, https://doi.org/10.3732/apps.1400042.

114. Baker, W.J., Bailey, P., Barber, V., Barker, A., Bellot, S., Bishop, D., Botigué, L.R., Brewer, G., Carruthers, T., Clarkson, J.J., et al. (2022). A comprehensive phylogenomic platform for exploring the angiosperm tree of Life. Syst. Biol. *71*, 301–319. https://doi.org/10.1093/sysbio/syab035.

115. Towns, J., Cockerill, T., Dahan, M., Foster, I., Gaither, K., Grimshaw, A., Hazlewood, V., Lathrop, S., Lifka, D., and Peterson, G.D. (2014). XSEDE: accelerating scientific discovery. Comput. Sci. Eng. *16*, 62–74.

116. Li, H. (2013). Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. https://doi.org/10.48550/arXiv.1303.3997.

117. Hoang, D.T., Chernomor, O., von Haeseler, A., Minh, B.Q., and Vinh, L.S. (2018). UFBoot2: improving the ultrafast bootstrap approximation. Mol. Biol. Evol. *35*, 518–522. https://doi.org/10.1093/molbev/msx281.

118. Revell, L.J. (2012). phytools: an R package for phylogenetic comparative biology (and other things). Methods Ecol. Evol. *3*, 217–223.

119. Robinson, D.F., and Foulds, L.R. (1981). Comparison of phylogenetic trees. Math. Biosci. *53*, 131–147. https://doi.org/10.1016/0025-5564(81)90043-2.

120. Smith, M.R. (2020). Information theoretic generalized Robinson–Foulds metrics for comparing phylogenetic trees. Bioinformatics *36*, 5007–5013.

121. Sanderson, M.J. (2002). Estimating absolute rates of molecular evolution and divergence times: A penalized likelihood approach. Mol. Biol. Evol. *19*, 101–109. https://doi.org/10.1093/oxfordjournals.molbev.a003974.

122. Vankan, M., Ho, S.Y.W., and Duchêne, D.A. (2022). Evolutionary Rate Variation among Lineages in Gene Trees has a Negative Impact on Species-Tree Inference. Syst. Biol. *71*, 490–500. https://doi.org/10.1093/sysbio/syab051.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Biological Samples		
See Data S1A, S1B, and S1G	This study	N/A
Critical Commercial Assays		
DNeasy PowerPlant Pro Kit	Qiagen, Hilden, Germany	Cat#13400
DNeasy PowerClean Pro Cleanup Kit	Qiagen, Hilden, Germany	Cat#12997
NEBNext Ultra II FS Kit	New England Biolabs, Ipswich, Massachusetts, USA	Cat#*E7645L*
IDT 10 index	Integrated DNA Technologies, Coralville, Iowa, USA	Cat#10008052
NovaSeq 6000 sequencer	Illumina, San Diego, California, USA	N/A
Deposited Data		
Raw target capture sequence data	This study	NCBI SRA PRJNA806513
Raw target capture sequence data	Nikolov et al.[3]	NCBI SRA PRJNA518905
Raw target capture sequence data	Hendriks et al.[46]	NCBI SRA PRJNA678873
Raw chloroplast sequence data	Walden et al.[2]	ENA/GenBank PRJEB38700
Multiple sequence alignments	This study	https://doi.org/10.5281/zenodo.8214354
Oligonucleotides		
myBaits Custom 20-40K	Arbor Biosciences, Ann Arbor, Michigan, USA	Cat#300296R
myBaits Angiosperms-353	Arbor Biosciences, Ann Arbor, Michigan, USA	Cat#308196
Software and Algorithms		
Various scripts to map raw data, create multiple sequence alignments, gene trees, and species trees	This study	https://doi.org/10.5281/zenodo.8214354
Trimmomatic v0.38	Bolger et al.[88]	https://github.com/usadellab/Trimmomatic
HybPiper v1.3.1	Johnson et al.[89]	https://github.com/mossmatters/HybPiper/wiki/HybPiper-Legacy-Wiki
BWA v0.7.16a	Li and Durbin[90]	https://github.com/lh3/bwa
BWA v0.7.17	Li and Durbin[90]	https://github.com/lh3/bwa
SPAdes v3.14.1	Bankevich et al.[91]	https://github.com/ablab/spades
GNU Parallel	Tange[92]	https://github.com/martinda/gnu-parallel
Python script 'filter_megatarget.py'	McLay et al.[93]	See publication
SAMtools v1.3.1	Danecek et al.[94]	https://github.com/samtools/samtools
Picard tools	N/A	http://broadinstitute.github.io/picard/
GATK3	McKenna et al.[95]	https://github.com/broadinstitute/gatk
GATK4	Van der Auwera and O'Connor[96]	https://github.com/broadinstitute/gatk
Shell script 'masker.sh'	This study	https://doi.org/10.5281/zenodo.8214354
Shell script 'cpanno.py'	This study	https://doi.org/10.5281/zenodo.8214354
BEDTools v2.27.1	Quinlan and Hall[97]	https://github.com/arq5x/bedtools2
MAFFT v7.273	Katoh and Standley[98]	https://mafft.cbrc.jp/alignment/software/
FastTree v2.1	Price et al.[99]	http://www.microbesonline.org/fasttree/
PASTA v1.8.6	Mirarab et al.[100]	https://github.com/smirarab/pasta
ASTRAL-III v5.7.8	Zhang et al.[48]	https://github.com/smirarab/ASTRAL
HybPhaser v2.0	Nauheimer et al.[47]	https://github.com/LarsNauheimer/HybPhaser
SplitsTree4 v4.17.1	Huson and Bryant[69]	https://github.com/husonlab/splitstree4

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
IQ-TREE2 v2.1.3	Minh et al.[101]	https://github.com/iqtree/iqtree2
trimAl v1.2	Capella-Gutiérrez et al.[102]	https://github.com/inab/trimal
TAPER v1.0.0	Zhang et al.[103]	https://github.com/chaoszhang/TAPER
TreeShrink v1.3.9	Mai and Mirarab[104]	https://github.com/uym2/TreeShrink
ASTRAL-Pro v1.1.6	Zhang et al.[49]	https://github.com/chaoszhang/A-pro
R package 'ggtree'	Xu et al.[105]	https://github.com/YuLab-SMU/ggtree
Rate4Site v3.2	Mayrose et al.[106]	N/A
R package 'PhyInformR v1.0'	Dornburg et al.[107]	https://github.com/carolinafishes/PhyInformR
Perl script 'catfasta2phyml.pl'	N/A	https://github.com/nylander/catfasta2phyml
IQ-TREE v1.6.12	Nguyen et al.[108]	https://github.com/Cibiv/IQ-TREE
IQ-TREE v2.2.0	Minh et al.[101]	https://github.com/Cibiv/IQ-TREE
treePL v1.0	Smith and O'Meara[109]	https://github.com/blackrim/treePL
TreeAnnotator v2.4.7	Bouckaert et al.[110]	https://beast.community/treeannotator
TreeAnnotator v2.6.7	Bouckaert et al.[111]	https://beast.community/treeannotator

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact, Kasper P. Hendriks (kasper.hendriks@naturalis.nl).

Materials availability

This study did not generate new unique reagents.

Data and code availability

- Target capture data have been deposited to the NCBI Sequence Read Archive (SRA): BioProject PRJNA678873 and PRJNA806513; all data are publicly available as of the date of publication. Sample accession numbers are listed in Table S1.
- All original code has been deposited to Zenodo and accessible through https://doi.org/10.5281/zenodo.8214354.
- Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS

This study did not use any experimental model or study participants.

METHOD DETAILS

Taxon sampling

To reconstruct the nuclear Brassicaceae phylogeny, we aimed to include at least one species from each of the 349 currently accepted genera—preferably each genus' type species—along with one species of every non-Brassicaceae Brassicales families and all species needed for fossil calibration checks of vicariance and colonisation within lineages of Brassicaceae (Data S1J). We followed BrassiBase[112] for taxonomic delimitation of species, genera, and tribes, added with more recent insights from taxonomic experts. Where possible, we sampled type specimens to support future taxonomic judgements (Table S1).

We generated new nuclear sequence data for 365 samples (Table S1) and added available sequences from 38 samples from Nikolov et al.[3] (Table S2). All new data were sequenced from dried herbarium specimens or silica dried tissue from 29 different herbarium collections across the world, with plants collected between 1807 and 2020 (including 35 pre-1900 and 29 samples collected between 1900 and 1950; Methods S1B). Old samples were included either because we did not have access to younger material, or to show the possibilities of our methodology with regard to natural history collections. We used the original type material of 24 species (Table S1).

New plastome data were generated from genome spiking (see below) for 237 samples (Table S1). We used additional data for 196 samples from Walden et al.,[2] 60 plastid genomes downloaded from GenBank and used in the same study (Table S7), and 31 samples from Nikolov et al.[3] (Table S2).

Library preparation, target capture, and sequencing

Wet lab methods followed Hendriks et al.[46] Briefly, we extracted genomic DNA from 25 mg of dried leaf tissue (or less if insufficient material was available; in case no leaf tissue was available from any herbarium voucher available to us, we used branches and/or flowers) using the DNeasy PowerPlant Pro Kit (Qiagen, Hilden, Germany), following the manufacturer's protocol (but with a final elution time of 1 h). DNA extracts with visible impurities (green or brown colour; ~25% of samples) were subsequently purified using the DNeasy PowerClean Pro Cleanup Kit (Qiagen, Hilden, Germany). Genomic DNA was stored in the DNA bank of Naturalis Biodiversity Center, Leiden, the Netherlands (Table S1). Genomic libraries were generated using the NEBNext Ultra II FS kit (New England Biolabs, Ipswich, Massachusetts, USA) with sonication in an M220 Focused-ultrasonicator (Covaris, Woburn, Massachusetts, USA; only for libraries with fragment peak length > 400 bp). Indexing was performed with 384 unique combinations from IDT10 primers (Integrated DNA Technologies, Coralville, Iowa, USA), with protocol adjustments described by Hendriks et al.[46] Target sequence capture was carried out on pools of 10–30 libraries each, using the 'mixed baits' approach described by Hendriks et al.[46] This method targets putatively single-copy nuclear genes from two different bait sets in a single capture reaction: a Brassicaceae-specific set targeting 1,827 exons from 764 genes, using 40k probes[3] (hereafter B764), and the now widely used Angiosperms353 v1 universal bait set targeting 353 genes, using 80k probes[44,45] (hereafter A353; both kits available from Arbor Biosciences, Ann Arbor, Michigan, USA). To maintain the ratio of probes among the bait sets during target capture, we used a B764: A353 = 1 : 2 (v/v) mixture. To aid skimming of chloroplast gene reads during sequencing,[113] we used genome spiking of each enriched library with its unenriched library at a ratio of 1 : 1 (M/M). Sequencing was performed on an Illumina NovaSeq 6000 sequencer (Illumina, San Diego, California, USA) at BaseClear, the Netherlands, producing 150-bp paired-end reads, at a targeted 100× technical coverage. Raw sequence data files were uploaded to the National Center for Biotechnology Information (NCBI) Sequence Read Archive (SRA) under BioProjects PRJNA806513 and PRJNA678873. Target capture and sequencing for a sample were repeated if results from sequence assembly (described below) were poor (< 500 genes recovered).

New 'mixed baits' data for 12 outgroup species were generated by the Plant and Fungal Trees of Life project at the Royal Botanic Gardens, Kew, UK, following methods described by Baker et al.[114] A single sample was processed by the Bailey Lab, New Mexico State University, USA, and three more at Heidelberg University, Germany. To further increase sampling, we added previously published raw sequence data for 37 samples from Nikolov et al.,[3] available from NCBI SRA as BioProject PRJNA518905 (target capture for B764 only; Table S2).

Sequence assembly of target capture data

Raw sequence data (from our study, as well as from other sources; see taxon sampling) were quality controlled and trimmed with Trimmomatic v0.38[88] using parameters from Baker, et al.[114] Trimmed reads were mapped against two reference files (i.e., for B764 and A353 bait sets) using HybPiper v1.3.1[89] with BWA v0.7.16a[90] and SPAdes v3.14.1,[91] using GNU Parallel[92] to manage parallel computing of samples on the XSEDE Stampede2 HPC.[115] We built a gene reference file for the B764 dataset from the 1,827 exon reference file of Nikolov et al.[3] by concatenation of same-gene exons, resulting in a total target length of 919,712 bp. For the A353 dataset, we used the 'mega353′ target file with the script 'filter_megatarget.py' to create a mustard family-specific reference[93] with a total target length of 294,516 bp. We identified a total of 36 genes with overlap among the two bait sets, which was expected because the two bait sets have been developed independently (Data S1K). To avoid studying the same genetic marker twice, we discarded these genes from the B764 dataset (generally the shorter targets), leaving a final nuclear dataset of 1,081 (i.e., 764 + 353 - 36) genes with a total target length of 888,392 bp.

Sequence assembly of plastome

We used off-target reads from genome spiking (samples with new raw data) and genome skimming (raw data from previous studies listed above) to reconstruct 60 plastid genes for 237 new samples and 31 samples from Nikolov et al.[3] New data were integrated into an already existing plastid dataset containing 231 Brassicaceae species.[2]

Trimmed sequencing reads were mapped using BWA v0.7.17 using option 'BWA-MEM'[116] and the *Arabidopsis thaliana* plastid genome (NCBI GenBank accession number NC_000932) as reference. Prior to mapping, the second copy of the inverted repeat region was removed, as identical regions lead to secondary alignments which are omitted by tools used in subsequent analysis. SAMtools v1.3.1[94] was used to enhance mapping quality and to sort and index the .bam files. Duplicates were removed using Picard tools (http://broadinstitute.github.io/picard/). Variant calling was performed using the GATK4[96] function 'HaplotypeCaller' setting ploidy to 1 and pcr-indel-model to none. The GATK3[95] function 'FastaAlternateReferenceMaker' was used to generate sequences including the detected SNPs and indels. Regions of low coverage (< 5) and low mapping quality (< 30) were detected using GATK3 function 'CallableLoci'. After having adjusted the positions of the regions to be masked using the inhouse script 'masker.sh' (Markus Kiefer, Heidelberg University; see Key Resources Table), BEDTools[97] function 'maskfasta' was used to mask regions of bad mapping quality (< 30) and low coverage (< 5). The annotation of genes was transferred by alignment to above mentioned *A. thaliana* plastid reference using the inhouse script 'cpanno.py' (Markus Kiefer, Heidelberg University; see Key Resources Table). After removal of gap columns, the final data matrix had a total target length of 29,120 bp and was 96.2% complete.

Taxonomic verification

We performed 'taxonomic verification' on a preliminary species phylogeny. To reconstruct this phylogeny, multiple sequence alignments were created for each gene using MAFFT v7.273,[98] with a quick gene tree inference using FastTree v2.1,[99] both within the

pipeline PASTA v1.8.6.[100] Unfiltered gene trees were used as input to ASTRAL-III v5.7.8.[48] Data from samples marked as possible or likely errors (found in highly unlikely positions in the preliminary species tree) were removed, followed by either repetition of library preparations or resampling of the species. This routine was repeated once more, such that the phylogenetic position of each sample was verified by taxonomic experts. Subsequently, any data resulting from multiple DNA extractions and/or library preparations from the same voucher were merged after trimming and again mapped following the same routine.

Allelic variation and paralog detection

We used HybPhaser v2.0,[47] an extension to HybPiper, to assess allelic variation and to detect possible paralogs in our nuclear dataset. In short, HybPhaser performs a re-mapping of raw sequence data, using the contig of each sample (created by HybPiper) as a new reference. Whereas HybPiper by default constructs the most likely allele for a—supposedly single-copy—gene based on the relative nucleotide frequency of each heterozygous site, HybPhaser instead takes SNP variation into account using nucleotide ambiguity codes, and uses these to quantify divergence between gene variants to detect paralogy and hybridisation. Single genes with high SNP count are considered likely paralogs, while samples with high SNP count across all genes are considered likely hybrids or polyploids.[47] Putative paralogs were removed from the dataset. However, since there is no single criterion to define a paralog, we used five 'routines' to do this (Table 1). In the 'inclusive' routine (1,018 genes, 332 genera, 375 species, after running a de-noising loop; see below), we retained as much data as possible (including all samples, and thus all genera for which we had any data) and only discarded poorly recovered genes (gene recovered for < 10% of the samples and/or proportion of gene target length recovered < 10% on average across all samples). We discarded putative paralogs in the 'strict' routine (1,013 genes, 317 genera, 356 species) by removing all 'outlier' genes, defined as loci that have more than 1.5*IQR (interquartile range) above the 3rd quartile of mean SNPs. In a 'superstrict' routine (297 genes, 317 genera, 356 species), we removed all genes with a mean proportion of SNPs across the dataset of > 0.02. After noticing large differences in mean SNP proportions among tribes within the mustard family (Methods S1C), we added the 'superstrict by tribe' routine (mean 1,031 genes, with gene selection varying by tribe, 317 genera, 356 species) in which we assessed and removed mean SNP proportions by tribe, leading to a very sparse sample-gene matrix (Methods S1D). Finally, we aimed to improve phylogenetic backbone support further by using a 'superstrict excluding hybrids' routine (303 genes, 124 genera, 138 species). This dataset was the same as in the 'superstrict' routine, but all samples belonging to rogue taxa and/or having a locus heterozygosity and/or allelic divergence (see next) in the upper 50% of detected values from all samples were removed. This led to a highly reduced (in terms of both genes and samples/species/genera/tribes) dataset, but with the anticipated advantage of having removed as much noise from the dataset as possible, while at the same time including enough representatives from all main lineages.

We used HybPhaser to detect possible hybrids by calculating each sample's allele divergence (AD; percentage of SNPs across all genes) and locus heterozygosity (LH; percentage of genes with SNPs), two metrics that are useful in the detection of hybrids.[47] Because hybrids are expected to inherit multiple alleles from their different parent species, they are expected to show relatively high levels of LH and an AD that corresponds to the divergence of the parental lineages. Very high values for AD are expected in lineages with multiple polyploidizations.[47] With time, polyploid lineages are expected to lose duplicated genes leading to a decrease in LH. Therefore, high AD combined with intermediate LH can indicate that samples are more ancient polyploids. While there is no universal circumscription of what values correspond to hybrids or other types of polyploids, these values can give a good indication on the history of hybridisation in samples. Here we broadly distinguish four classes: hybrid (high LH, medium AD), highly polyploid (high LH and high AD), old polyploid (medium LH, medium AD), and old and highly polyploid (medium LH, high AD).

Nuclear phylogenomics

We applied four different phylogenomic approaches to analyse our nuclear dataset. We used default settings and parameters for all tools, unless specified.

First, we applied a network approach to visualise possible evolutionary reticulations, inferring a splits graph (based on uncorrected p-distances) with SplitsTree4 v4.17.1.[69] We used a nuclear supermatrix from the 297 gene alignments from the 'superstrict' routine as input.

Second, we used a ML supermatrix approach with IQ-TREE2 v2.1.3[101] with again the nuclear supermatrix approach with the 297 genes from the 'superstrict' routine (with outgroup sample S1321, *Synsepalum afzelii*), and separately with a supermatrix of the 303 genes from the 'superstrict excluding hybrids' routine (with outgroup sample MYZV, *Tropaeolum peregrinum*). We used 1,000 ultrafast bootstraps[117] (saving bootstrap replicates) with a GTR+F+R model. In this first analysis we specifically kept all outgroup samples needed for fossil calibration, even if these would have been removed based on our sampling routine criteria. Contrary to our coalescent-based approach (see below), these analyses generated phylogenies in which branch lengths were representative of evolutionary change (number of mutations), which was needed for the subsequent divergence time estimation. As an input for fossil calibration (see below) we again used a nuclear supermatrix approach with the 297 genes from the 'superstrict' routine. We repeated the IQ-TREE2 analysis with the resultant species tree and all 297 gene trees (see below) to calculate gene (gCF) and site concordance factors (sCF; parameter –scf 1,000) for all nodes.[67]

Third, we applied a coalescent-based approach with ASTRAL-III.[48] As input, we took the consensus sequences from the four paralog detection routines in HybPhaser (see above) to infer gene trees that served as input for a coalescent-based analysis. For each of the routines, we started by running a de-noising loop: sequences were aligned using MAFFT v7.273[98] and trimmed using trimAl v1.2[102] with parameters resoverlap 0.75, seqoverlap 0.90, and gt 0.90. Any remaining likely sequencing errors were masked using

TAPER v1.0.0[103] with default parameters. Gene trees were inferred using IQ-TREE2 v2.1.3[101] inferring branch support using ultrafast bootstrapping,[117] with other parameters following Baker, et al.[114] We used TreeShrink v1.3.9[104] with default parameters on the complete set of gene trees to detect and remove outlier branches and update gene trees and alignments. Discordance among gene trees was scored using all gene trees associated with each routine and calculated using normalised quartet scores for the main topology, along with first and second alternatives.

Fourth, we applied another coalescent-based approach with ASTRAL-Pro v1.1.6.[49] Contrary to ASTRAL-III, this version allows the input of multiple alleles (including possible paralogs) for each individual, acknowledging that this may actually be informative in species tree inference (e.g., no a priori choice of a definitive homologous gene copy needs to be made). Gene trees were now collected from mapping done by HybPiper using the script 'paralog_investigator.py', which saves all possible alleles.[89] Genes were again aligned with MAFFT v7.273 and trimmed using trimAl v1.2 with parameters resoverlap 0.75, seqoverlap 0.90, and gt 0.90, with subsequent gene tree inference with IQ-TREE2 v2.1.3. All phylogenetic trees from the approaches 2–4 were plotted using the R package ggtree.[105]

Phylogenetic informativeness

To study any differences in support from different nuclear genes in inferring the nuclear species tree, we calculated Townsend's phylogenetic informativeness[50] for all genes included in the 'strict' routine (Figure S1). First, we calculated relative evolutionary rates for all sites in each gene's multiple sequence alignment, constrained on the ML supermatrix approach species tree, using Rate4Site v3.2.[106] Second, we used the R package PhyInformR v1.0[107] to calculate phylogenetic informativeness profiles for each gene, making a distinction between genes obtained from either the B764 and A353 bait sets, and genes with a mean SNP proportion of <= 0.02 and > 0.02 (i.e., the threshold applied for paralog detection in HybPhaser).

Plastome phylogenetics

To generate a plastome-based phylogeny, coding sequences and sequences encoding tRNAs and rRNAs were extracted using BEDTools v2.27.1 function 'getfasta[97]'. We used a reduced gene set of 60 loci as previously used by Walden et al.[2] New sequences were aligned, together with the corresponding sequences from Walden et al.,[2] using MAFFT v7.45.3.[98] In a last step, gap columns were deleted and alignments were concatenated using the script 'catfasta2phyml.pl' (https://github.com/nylander/catfasta2phyml).

Phylogenetic reconstruction was performed using IQ-TREE v1.6.12[108] with partition information from the alignment, defining the outgroup, and running 1,000 ultrafast bootstrap replicates.[117] A second phylogenetic reconstruction was performed using IQ-TREE v2.2.0[101] to calculate site concordance factors (sCF) as a measure of support for the splits in the tree,[67] with the former species tree topology and gene alignment as input. Note that in the plastome phylogeny, gene concordance factors (gCF) cannot be calculated as in the nuclear phylogeny, because the 60 plastome genes studied were considered to be a single heritable unit (supermatrix approach), and consequently separate gene trees (needed to calculate gCF) were not inferred.

Nuclear versus plastome BrassiToL species tree incongruences were visualised using a cophylogeny plot created in R package phytools.[118] Incongruences within tribs were quantified using the classical Robinson-Foulds metric[119] and the generalised Robinson-Foulds metric following Smith[120] using R package TreeDist.[120] Species trees were first pruned to include only species present in both trees, and any duplicates were removed (after random selection per species).

Fossil calibration

For both the nuclear and the plastome dataset, we estimated divergence times within the family using the penalized likelihood approach[121] as implemented in treePL v1.0[109] and the Turonian fossil *Dressiantha bicarpellata*, estimated at 93.6–89.3 Ma[51] (which we took as maximum and minimum ages, respectively), as a single calibration point at the stem node of order Brassicales (cf. Couvreur et al.[20]). For the nuclear dataset, we used the topology from the ML supermatrix approach as input species tree, and reran IQ-TREE using the gene alignments from the 20 most clock-like genes only to infer relative branch lengths, acknowledging that inclusion of too many genes can easily result in an artificial pushback in time of internal nodes. To do so, we first calculated the clock-likeness of all genes following Vankan et al.,[122] who defined clock-likeness as the coefficient of variation of all root-to-tip distances in the gene tree. When running the priming analysis in treePL, the value for 'opt' was set to 2, and 'optad' set to 1. The 'moredetail' and 'moredetailad' options were in effect and 'optcvad' was set to 1. Cross validation analysis indicated 10 as the best smoothing value. We assessed node age estimates by repeating the treePL (using the above optimised settings) analysis for 1,000 bootstrap trees generated with IQ-TREE, this time fixing the topology of the species tree (but not branch lengths) and summarising with TreeAnnotator v2.4.7[110] to obtain 95% HPD confidence intervals.

For the plastid dataset, when running the priming analysis and later adjustments, the values for 'opt' and 'optad' were both set to 3. The 'moredetail' and 'moredetailad' options were in effect and 'optcvad' was set to 4. Cross validation analysis indicated 0.000001 as best smoothing value. Again, we assessed node age estimates by repeating the treePL analysis (using the above optimised settings) for the 1,000 bootstrap replicates generated in IQ-TREE. Calibrated gene trees were summarised using TreeAnnotator v2.6.7[111] to obtain 95% HPD confidence intervals.

We performed multi-evidence validation of our new results against four other fossils and five biogeographical dating events as suggested by Franzke et al.[52] by comparing expected and recovered node ages (Data S1J).

QUANTIFICATION AND STATISTICAL ANALYSIS

Support for nodes in phylogenetic trees was calculated using multiple methods (LPP, bootstrap, quartet score, gCF, and sCF), as detailed under method details.