Plant Communications
Research article

CellPress
Partner Journal

Genomes of *Meniocus linifolius* and *Tetracme quadricornis* reveal the ancestral karyotype and genomic features of core Brassicaceae

Jie Liu[1,4,5], Shi-Zhao Zhou[1,4,5], Yun-Long Liu[2,5], Bin-Yan Zhao[1,4,5], Dongmei Yu[1], Mi-Cai Zhong[1], Xiao-Dong Jiang[1], Wei-Hua Cui[1], Jiu-Xia Zhao[1,4], Juan Qiu[3], Liang-Min Liu[2,4], Zhen-Hua Guo[2], Hong-Tao Li[2], Dun-Yan Tan[3], Jin-Yong Hu[1,*] and De-Zhu Li[2,*]

[1]CAS Key Laboratory for Plant Diversity and Biogeography of East Asia & Yunnan Key Laboratory of Crop Wild Relatives Omics, Kunming Institute of Botany, Chinese Academy of Sciences, Kunming 650201, China

[2]Germplasm Bank of Wild Species & Yunnan Key Laboratory for Crop Wild Relatives Omics, Kunming Institute of Botany, Chinese Academy of Sciences, Kunming 650201, China

[3]College of Life Sciences, Xinjiang Agricultural University, Ürümqi 830052, China

[4]University of Chinese Academy of Sciences, Beijing 100049, China

[5]These authors contributed equally to this article.

***Correspondence: Jin-Yong Hu (hujinyong@mail.kib.ac.cn), De-Zhu Li (dzl@mail.kib.ac.cn)**

https://doi.org/10.1016/j.xplc.2024.100878

ABSTRACT

Brassicaceae represents an important plant family from both a scientific and economic perspective. However, genomic features related to the early diversification of this family have not been fully characterized, especially upon the uplift of the Tibetan Plateau, which was followed by increasing aridity in the Asian interior, intensifying monsoons in Eastern Asia, and significantly fluctuating daily temperatures. Here, we reveal the genomic architecture that accompanied early Brassicaceae diversification by analyzing two high-quality chromosome-level genomes for *Meniocus linifolius* (Arabodae; clade D) and *Tetracme quadricornis* (Hesperodae; clade E), together with genomes representing all major Brassicaceae clades and the basal Aethionemeae. We reconstructed an ancestral core Brassicaceae karyotype (CBK) containing 9 pseudochromosomes with 65 conserved syntenic genomic blocks and identified 9702 conserved genes in Brassicaceae. We detected pervasive conflicting phylogenomic signals accompanied by widespread ancient hybridization events, which correlate well with the early divergence of core Brassicaceae. We identified a successive Brassicaceae-specific expansion of the class I *TREHALOSE-6-PHOSPHATE SYNTHASE 1* (*TPS1*) gene family, which encodes enzymes with essential regulatory roles in flowering time and embryo development. The *TPS1*s were mainly randomly amplified, followed by expression divergence. Our results provide fresh insights into historical genomic features coupled with Brassicaceae evolution and offer a potential model for broad-scale studies of adaptive radiation under an ever-changing environment.

Key words: Cruciferae, genomic features, ancient hybridization, core Brassicaceae karyotype, CBK, *TREHALOSE-6-PHOSPHATE SYNTHASE 1* genes, *TPS1*s

Liu J., Zhou S.-Z., Liu Y.-L., Zhao B.-Y., Yu D., Zhong M.-C., Jiang X.-D., Cui W.-H., Zhao J.-X., Qiu J., Liu L.-M., Guo Z.-H., Li H.-T., Tan D.-Y., Hu J.-Y., and Li D.-Z. (2024). Genomes of *Meniocus linifolius* and *Tetracme quadricornis* reveal the ancestral karyotype and genomic features of core Brassicaceae. Plant Comm. **5**, 100878.

INTRODUCTION

Understanding the molecular and genetic mechanisms that underlie species radiations remains an important, albeit challenging, topic in evolutionary biology. Radiation can result from increased speciation, decreased extinction rates, or both (Naciri and Linder, 2020). This phenomenon is prevalent in nature, with numerous examples, such as butterflies (Edelman et al., 2019), Malawi cichlids (Malinsky et al., 2018), lizards (Garcia-Porta et al., 2019), Darwin's giant daisies (Fernandez-Mazuecos et al., 2020), and rhododendrons (Ma et al., 2022;

Published by the Plant Communications Shanghai Editorial Office in association with Cell Press, an imprint of Elsevier Inc., on behalf of CSPB and CEMPS, CAS.

Xia et al., 2022). However, rapid radiations make phylogenetic reconstruction challenging owing to the limited accumulation of substitutions within a short time. In addition, large population sizes and close evolutionary relationships can result in incomplete lineage sorting (ILS) and hybridization, leading to conflicting gene and species trees (Cai et al., 2021; Guo et al., 2023).

Brassicaceae (Cruciferae) contains approximately 4140 species, many of which are important crops (e.g., cabbage, rapeseed, mustard, and broccoli, among others) and/or model plants in *Arabidopsis*, *Capsella*, *Brassica*, and *Arabis* (German et al., 2023). Except for the basal Aethionemeae, core Brassicaceae comprises approximately 98.6% of species in five supertribes/clades, i.e., Camelinodae/A, Brassicodae/B, Hesperodae/E, Arabodae/D, and Heliophilodae/C, whose phylogenetic relationships lack resolution because of the likelihood of an early rapid radiation, as implied by many studies (Al-Shehbaz et al., 2006; Bailey et al., 2006; Beilstein et al., 2006, 2008; German et al., 2009; Couvreur et al., 2010; Warwick et al., 2010; Franzke et al., 2011; Hohmann et al., 2015; Huang et al., 2016; Guo et al., 2017; Nikolov et al., 2019; Walden et al., 2020b; Liu et al., 2020, 2024; Hendriks et al., 2023). In addition, a whole-genome duplication (WGD) event (i.e., At-α WGD) appears to have been essential for novel traits and species diversification in Brassicaceae (Franzke et al., 2011; Edger et al., 2018). This pattern has been observed in several other large and economically important families, including Poaceae, Asteraceae, Fabaceae, and Solanaceae, and has inspired the development of the WGD radiation lag-time model (Schranz et al., 2012). Indeed, WGD or polyploidization followed by rediploidization has been a major driving force in plant diversification (Soltis et al., 2015; Qiao et al., 2019). Hybridization, which is thought to be the first and most important step toward allopolyploidy, also contributes to species radiation by increasing allelic diversity, thus enhancing adaptability to new and challenging environments (Guo et al., 2021; Slovak et al., 2023).

Recent research suggests that Brassicaceae originated during the middle to late Eocene (40.5–36.9 million years ago [mya]), a period in Earth's history known as the "icehouse era". The crown age of Brassicaceae, marked by the divergence of the basal Aethionemeae and core Brassicaceae, is estimated to be between 23.1 and 25.7 mya (Hendriks et al., 2023). This period corresponds to the transition from the Oligocene to the Miocene and the rapid uplift of the Tibetan Plateau, which was followed by increasing aridity in the Asian interior, intensifying monsoons in Eastern Asia, and markedly fluctuating daily temperatures (Zachos et al., 2001; Kagale et al., 2014; Ding et al., 2020; Miao et al., 2022). However, molecular and genomic features related to this early diversification remain poorly characterized.

Inferring ancestral genomes is one of the central aspects of comparative genomic analyses (Murat et al., 2017; Anselmetti et al., 2018; Gao et al., 2022). The ancestral crucifer karyotype (ACK) for Brassicaceae has been proposed to feature 8 pseudochromosomes with 22 conserved ancestral genomic blocks (GBs) (Schranz et al., 2006; Lysak et al., 2016). The ACK has been widely used to study the phylogenomic and karyotypic evolutionary patterns of Brassicaceae (Mandakova and Lysak, 2008; Willing et al., 2015; Geiser et al., 2016; Mandakova et al., 2017, 2018, 2020; Walden et al., 2020a; Bayat et al., 2021; Guo et al., 2021). However, the ACK was essentially based on genome sequences and genetic maps of *A. thaliana*, *A. lyrata*, *C. rubella*, and *B. rapa* (Schranz et al., 2006; Lysak et al., 2016). Later efforts that incorporated the *Thellungiella parvula* genome did not markedly alter the ACK (Murat et al., 2015). More recently, inclusion of additional genomes from earlier-diverging lineages resulted in the reconstruction of a new haploid ancestral genome (n = 9) for core Brassicaceae (Walden and Schranz, 2023). However, this new karyotype contains only 3392 strictly conserved syntenic genes, likely owing to use of the non-chromosome-level *Eutrema parvulum* genome to represent Hesperodae/clade E. With the generation of more high-quality chromosome-level genomes and better phylogenomic coverage, an improved ancestral karyotype may provide a better understanding of the Brassicaceae genome and chromosomal evolution.

Species radiation is often accompanied by adaptation to a fluctuating or otherwise challenging environment, in which the timing of the vegetative to reproductive growth transition (i.e., flowering time) plays a pivotal role (Alonso-Blanco et al., 2009; Franzke et al., 2011). Flowering time is tightly regulated by a highly wired gene-regulatory network that links endogenous developmental signals and exogenous environmental signals such as light, temperature, and water availability (Andres and Coupland, 2012; Hyun et al., 2017; Gaudinier and Blackman, 2020). As a signal of sugar or energy availability, the level of trehalose-6-phosphate (T6P), which is synthesized mainly by T6P synthase 1 (TPS1) and T6P phosphatase, determines flowering time, embryogenesis, and other developmental processes, as well as various stress responses (Iordachescu and Imai, 2008; Fernandez et al., 2010; Fichtner et al., 2021; Fichtner and Lunn, 2021). The TPS1-mediated T6P pathway is highly conserved even in prokaryotes; however, the evolutionary patterns of *TPS1* gene expansion and/or contraction have not been fully characterized in plants, including Brassicaceae.

Here, we systematically explore the genomic features associated with the early Brassicaceae radiation by analyzing a set of high-quality chromosome-level genomes, including two newly assembled genomes, from species representing each supertribe and the basal Aethionemeae (Huang et al., 2016; German et al., 2023; Hendriks et al., 2023). We propose an ancestral core Brassicaceae karyotype (CBK) that includes 9 pseudochromosomes and 65 conserved GBs. We detect strong phylogenomic conflicts associated with ancient hybridization and identify a highly dynamic pattern of Brassicaceae-specific expansion of class I *TPS1* genes. Notably, the expression of these genes responds differentially to fluctuating temperature. Our efforts provide new genomic resources and improve our understanding of Brassicaceae diversification during historical environmental changes.

RESULTS

High-quality chromosome-level reference genomes for Arabodae and Hesperodae

We reconstructed high-quality chromosome-level genome assemblies for two species, *Meniocus linifolius* (*Mli*), representing

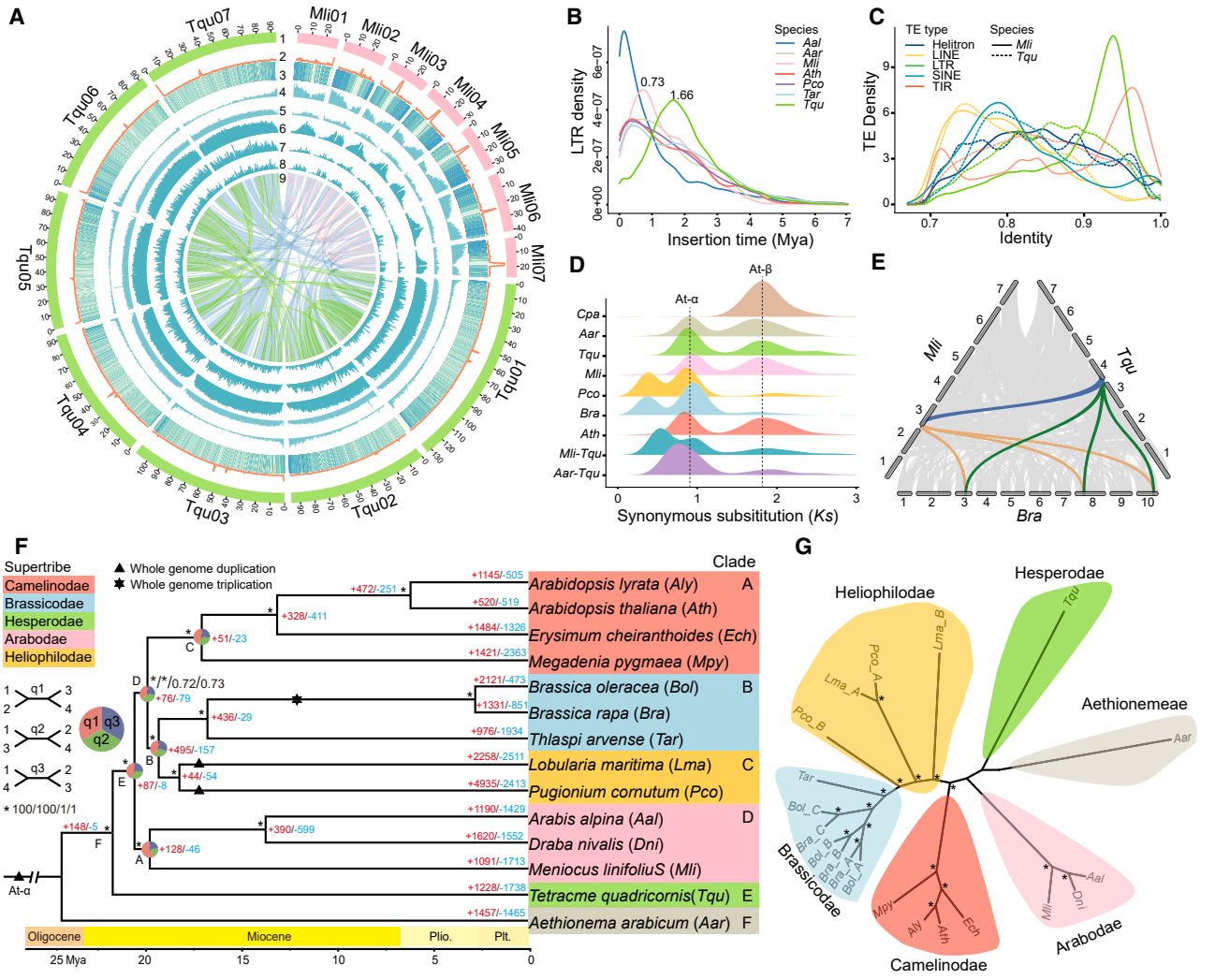


Figure 1. Phylogenomic relationships among Brassicaceae.

(A) Genomic landscapes for *Tetracme quadricornis* (*Tqu*; left) and *Meniocus linifolius* (*Mli*; right). (1) Pseudo-chromosomes, (2) tandem repeats, (3) gene expression profiles of mixed leaf, root, stem, and flower samples, (4) GC content, (5) gene density, (6) repetitive sequences along chromosomes, (7) *Copia* density in 500-kb sliding window, (8) *Gypsy* density in 500-kb sliding window, (9) intra- (green and pink lines) and interspecies synteny (blue lines).

(B) Density and insertion time (mya) distribution for intact LTRs in seven Brassicaceae species. *Aal*, *Arabis alpina*; *Aar*, *Aethionema arabicum*; *Ath*, *Arabidopsis thaliana*; *Pco*, *Pugionium cornutum*; *Tar*, *Thlaspi arvense*.

(C) Distribution of sequence identity values between genomic copies and consensus repeats for different types of transposable elements (TEs) in *Tqu* and *Mli* assemblies.

(D) Synonymous substitutions per synonymous site (*Ks*) distributions of intra- and intergenomic syntenic blocks. At-α and At-β indicate the Brassicaceae- and Brassicales-specific WGD events, respectively. *Cpa*, *Carica papaya*; *Bra*, *Brassica rapa*. The younger peak observed in the *Ks* distribution for *Mli-Tqu* suggests a diversification corresponding to the split time between *Mli* and *Tqu*.

(E) Collinearity patterns among genomes of *Tqu*, *Mli*, and *Bra*. The green and orange wedges highlight an example of triplication of the *Tqu* (Chr4) and *Mli* (Chr3) genome segments in *Bra* (Chr3/8/10), and the blue wedge indicates the syntenic block between *Tqu* and *Mli*.

(F) A supertribe-level phylogeny for Brassicaceae. A maximum likelihood (ML) tree based on the concatenation of 1463 single-copy orthologous genes (SCOGs) identified with *Orthofinder* is shown, with estimated divergence times at the bottom. Bootstrap support values and posterior probabilities are marked with an * indicating 100% support, and numbers indicate the real values for four datasets: concatenation-based with *Orthofinder*, concatenation-based with *SonicParanoid*, coalescence-based with *Orthofinder*, and coalescence-based with *SonicParanoid*. The black triangles and hexagon indicate the WGDs in Brassicaceae (At-α), *Lobularia maritima* (*Lma*), and *Pco* and the WGT (whole-genome triplication) in the most recent common ancestor of *Bra* and *Bol*, respectively. Numbers on the branches represent the numbers of expanded gene families (+, red) or contracted gene families (–, blue) among lineages. Pie charts at five internal nodes **(A–E)** show the frequency of three topologies (q1–q3, sector representations in different colors).

(G) A coalescence-based phylogeny based on 4434 syntenic genes. The A/B/C marks indicate the sub-genomes of *Pco*, *Lma*, *Bra*, and *Bol*. *Bol*, *Brassica oleracea*; *Dni*, *Draba nivalis; Ech*, *Erysimum cheiranthoides*.

Arabodae/clade D, and *Tetracme quadricornis* (*Tqu*), representing Hesperodae/clade E (Figure 1A). A combination of flow cytometry and genomic surveys using ~50× Illumina short reads revealed that genome sizes were between 246 and 288 Mb for *Mli* and 721 and 768 Mb for *Tqu* (supplemental Figures 1 and 2; supplemental Table 1). Cytological analysis indicated that the chromosome

number for *Mli* was 2*n* = 14 (supplemental Figure 3), although both 2*n* = 16 and 2*n* = 14 have been reported for this species (Spaniel et al., 2015). Both species exhibited a low level of heterozygosity with 0.0586% for *Mli* and 0.0344% for *Tqu* (supplemental Figure 2). Both genomes were sequenced and assembled using PacBio HiFi long reads and high-throughput chromosome conformation capture (Hi-C) sequencing (supplemental Figure 4; supplemental Tables 2 and 3). The final nuclear genome assemblies for *Mli* and *Tqu* were approximately 244.24 and 724.93 Mb in length, with contig N50s of approximately 15.43 and 45.06 Mb and maximum scaffold lengths of 40.91 and 133.35 Mb, respectively (supplemental Table 4). Both species have seven pseudochromosomes, to which more than 93% (*Mli*) and 99% (*Tqu*) of the contigs were anchored via Hi-C data (supplemental Figures 5 and 6; supplemental Table 5). Both assemblies exhibited high completeness, with >98.5% of the eudicot Benchmarking Universal Single-Copy Orthologs (BUSCOs) identified (supplemental Figure 7; supplemental Table 6) (Simao et al., 2015). In addition, >99% of the Illumina short reads were properly mapped to the genome assemblies (supplemental Table 7). The long terminal repeat (LTR) assembly index was 11.07 for *Mli* and 23.46 for *Tqu*, meeting the "Reference" and "Gold" standards, respectively (supplemental Figure 8; supplemental Table 4) (Ou et al., 2018).

Using a combination of homology-, transcriptome-, and *ab initio*-based predictive approaches, we annotated 33 038 and 27 225 protein-coding genes for *Mli* and *Tqu* (supplemental Figure 9; supplemental Table 8). For both species, >97% of the predicted protein-coding genes were functionally annotated using eight databases (supplemental Table 9). The proteomes were estimated to be >97.5% complete for both species on the basis of BUSCO assessments (supplemental Figure 10; supplemental Table 10). We also annotated 169 pseudogenes, 6074 rRNAs, 3751 tRNAs, 83 miRNAs, 98 snRNAs, and 185 snoRNAs in *Mli* and 194 pseudogenes, 4132 rRNAs, 1538 tRNAs, 109 miRNAs, 182 snRNAs, and 203 snoRNAs in *Tqu* (supplemental Tables 11 and 12).

The *Mli* and *Tqu* assemblies were composed of 38.74% (85.22 Mb) and 73.4% (524.22 Mb) transposable elements (TEs), respectively (supplemental Table 13). The most abundant repetitive elements were retrotransposons, which accounted for 51.19% and 60.9% of all repetitive sequences in *Mli* and *Tqu*. Retrotransposons accounted for 19.83% and 44.7% of the *Mli and Tqu* genome assemblies, and DNA transposons accounted for 17.31% and 26.63%. *Tqu* contained the most ancient LTR insertion at ~1.66 mya, whereas *Mli* and the other five species contained insertions dating between 0.7 and 0.8 mya (Figure 1B). Corroborating the very recent burst history, the LTRs and terminal inverted repeats in *Mli* featured sequence identities between 90% and 100% (Figure 1C). An additional *de novo* annotation of two previously published Arabodae genomes, *Arabis alpina* (*Aal*) (Jiao et al., 2017) and *Draba nivalis* (*Dni*) (Nowak et al., 2021), revealed that all three genomes had similar TE compositions but varied in tandem repeat content (ranging from 4.78% in *Dni* to 9.13% in *Mli*) (supplemental Tables 13 and 14; supplemental Figure 11).

Genome sizes varied significantly among Hesperodae. Tetraploid *Hesperis matronalis* had the largest genome (8117 Mb) and diploid *E. syriacum* (*Esy*) had the smallest (256 Mb), which was approximately one-third the size of diploid *Tqu* (supplemental Table 13) (Kiefer et al., 2014; Mandakova et al., 2017; Hlouskova et al., 2019). Compared with *Esy*, *Tqu* contained approximately 2.2-fold more LTRs, 3.47-fold more helitrons, and longer introns (Wilcoxon test, P = 2e−10; supplemental Tables 13 and 15; supplemental Figures 12 and 13). These results suggest that variation in TE content and gene structure may contribute to genome size diversity in Hesperodae.

Two genomes free of extra WGD events

Two well-known features of Brassicaceae are the family-specific At-α WGD and the At-β WGD shared within Brassicales (Huang et al., 2020b; Walden et al., 2020b; Guo et al., 2021). To explore the genomic history of *Mli* and *Tqu*, we compared their genomes with those of five representative Brassicaceae species and the basal Brassicales species *Carica papaya* (*Cpa*) (Figure 1D). These species exhibited several WGD patterns. For example, *A. thaliana* (*Ath*; Camelinodae) and *Aethionema arabicum* (*Aar*; Aethionemeae) shared both the At-α and At-β WGDs; *B. rapa* (*Bra*; Brassicodae) and *Pugionlum comutum* (*Pco*; Heliophilodae) featured more recent polyploidization events (Cheng et al., 2014; Hu et al., 2021); and *Cpa* experienced only the At-β WGD (Edger et al., 2018). Both *Mli* and *Tqu* featured high levels of intragenomic synteny coincident with the At-α and At-β WGDs (Figure 1A; supplemental Figures 14 and 15). In addition, synonymous substitutions per synonymous site (*Ks*) distributions of homologous pairs from intra- and intergenomic syntenic blocks demonstrated that both genomes were free of extra WGDs (Figure 1D; supplemental Figure 16). Finally, analysis of intergenomic synteny revealed that both *Mli* and *Tqu* exhibited syntenic depth ratios of 1:3 with *B. rapa*, whereas a 1:1 ratio persisted for *Mli* relative to *Aal* and *Dni* in Arabodae and *Tqu* relative to *Esy* in Hesperodae (Figures 1E; supplemental Figures 17 and 18).

A consistent phylogeny for core Brassicaceae supertribes

Recent research suggests that orthologous or single-copy nuclear sequences from *de novo* assemblies tend to be more accurate than those from either transcriptomic or targeted sequence-capture data (Hu et al., 2023). To characterize the phylogenomic relationships among supertribes of the core Brassicaceae, we selected chromosome-level genomes from 14 species representing all supertribes and the basal Aethionemeae. We identified 1463 shared single-copy orthologous genes (SCOGs) (supplemental Table 16). Highly supported (100% bootstrap values for all nodes) maximum-likelihood (ML) species trees were produced from the concatenated sequences using both nucleotide (all three codons or with third position removed) and amino acid sequences (Figures 1F; supplemental Figures 19 and 20). A coalescent-based analysis with individual gene trees after removal of nodes with bootstrap values <60% resulted in a topology identical to that of the concatenation-based analysis (supplemental Figure 21 and 22). Our results revealed that Hesperodae (*Tqu*) was sister to the remaining supertribes within the core Brassicaceae. Arabodae (*Aal*, *Dni*, and *Mli*) was a successive sister to Camelinodae, Brassicodae, and Heliophilodae, consistent with previous reports (Huang et al.,

2016; Kiefer et al., 2019; Nikolov et al., 2019; Hendriks et al., 2023). Within Arabodae, *Mli* clustered with the two Arabideae species *Aal* and *Dni*. The rogue species *Megadenia pygmaea* (*Mpy*) of the tribe Biscutelleae, which is well known for having a contentious phylogenetic placement (Guo et al., 2021), was grouped within Camelinodae (*Ath*, *A. lyrata*, and *Erysimum cheiranthoides*) in our analysis. This result is in agreement with previous transcriptome- and genome-based phylogenies (Kiefer et al., 2019; Guo et al., 2021) but in conflict with other studies that place the species within Heliophilodae (Huang et al., 2016).

To minimize orthology inference errors, we next extracted 2546 SCOGs using *SonicParanoid* (Cosentino and Iwasaki, 2019). A concatenation-based phylogenomic reconstruction resulted in the same tree topology described above (Figures 1F; supplemental Figures 19, 20, and 23). However, a coalescent-based analysis using amino acid sequences resulted in Camelinodae being sister to Brassicodae, Arabodae, and Heliophilodae (supplemental Figure 24). We next performed a synteny-based phylogenomic reconstruction using 4344 collinear genes identified with the *WGDI* pipeline (Sun et al., 2022) using *C. rubella* as a reference (Slotte et al., 2013) (supplemental Figure 25). Again, the same topology was obtained, with the polyploidization histories clearly reflected for *Bra*, *Bol*, *Lma*, and *Pco* (Figure 1G; supplemental Figure 26) (Cheng et al., 2014; Huang et al., 2020a; Hu et al., 2021).

To minimize phylogenetic errors resulting from poor taxon sampling, we included up to 27 Brassicaceae species and used *Cleome violacea* (Cleomaceae) as the outgroup. With 1092 SCOGs, both concatenation- and coalescent-based methods produced consistent topologies (supplemental Figure 27). An additional expansion to 55 Brassicales genomes with 5217 low-copy orthologous genes generated a nearly identical topology, except that *Mpy* and *Lunaria annua* (Biscutelleae) moved from Camelinodae to Heliophilodae (supplemental Figure 28). However, the majority of nodes, especially deeper nodes of the core Brassicaceae, exhibited low support values, suggesting the presence of high levels of genomic complexity (Mandakova et al., 2018; Guo et al., 2021; Hendriks et al., 2023).

Despite such consistent core Brassicaceae tree topologies, phylogenetic discordance was frequently observed among nuclear gene trees (Figure 2), reflecting the complex evolutionary history of Brassicaceae. Indeed, analyses of internode certainty all (ICA) values and numbers of conflicting and concordant bipartitions revealed strong conflicts, especially in the deepest nodes of the core Brassicaceae (supplemental Figure 29). This pattern was easily observed in gene trees visualized with cloudogram through *DensiTree* (Bouckaert, 2010). These conflicts occurred mainly in the period between 10 and 15 mya (Figure 2A). The deeper nodes (A to E) of the core Brassicaceae were supported by around 50% of gene trees for Arabodae (node A, q1) or less for other nodes (q1; Figure 1F). In addition, the sister relationships among different supertribes were weakly, or sometimes strongly, rejected by >60% of relevant gene trees (Figure 2C). Thus, the core Brassicaceae supertribes featured a high level of phylogenomic complexity. We also identified pervasive incongruence between plastome- and nuclear-genome-based phylogenies (Figure 2B; supplemental Figures 30–32), a pattern previously reported in core Brassicaceae (Walden et al., 2020a, 2020b; Liu et al., 2020, 2024; Hendriks et al., 2023). This indicated that hybridization and/or ILS may have occurred during the early diversification of Brassicaceae.

Frequent hybridization and introgression during Brassicaceae evolution

At least four historical hybridization scenarios were identified in a network analysis (Figure 2B; supplemental Figure 33). The strongest signal of gene flow occurred between *Mpy* and the most recent common ancestor (MRCA) of supertribe Camelinodae and *Pco*. Other hybridizations were identified between the MRCA of Brassicodae and Heliophilodae. An ABBA-BABA-derived test (*D* statistic) (Malinsky et al., 2021) revealed significant introgressions in 86 of the 120 tested species triplets (three-taxon subtrees) (Z score > 3 and $P < 0.002$). The maximum pairwise *D* and *f4*-ratio statistics were observed for *Mpy-Pco* and *Mpy-Tar* (*Thlaspi arvense*) (Figure 2D; supplemental Figure 34; supplemental Table 17). A derived *Fbranch* (f_b) analysis identified a strong hybridization signal between *Mpy* and all other supertribes (Brassicodae, Hesperodae, Arabodae, and Heliophilodae), and other introgression events were identified between the MRCA of (Brassicodae + Heliophilodae) and Hesperodae/Arabodae (Figure 2E). Quantification of tree branch lengths with *QuIBL* (Edelman et al., 2019) revealed that 15.8% of tested triplets featured significant hybridization signals (19 of 120, $\Delta BIC < -10$), with the introgression gene trees having an average ratio of 17.4% (ranging from 9.91% for *Tqu-Aly* to 26.52% for *Mpy-Pco*; Figure 2D; supplemental Tables 18 and 19). A final calculation of the introgression intensity showed that deeper nodes had high reticulation indices, especially in supertribes Hesperodae, Arabodae, and Heliophilodae (supplemental Figure 35) (Cai et al., 2021). Collectively, these data suggest a complex history of hybridization and introgression during the early radiation of Brassicaceae, with *Mpy* likely being of hybrid origin.

Even so, we could not overlook the contribution of ILS. Estimation of theta, a parameter that reflects ILS levels (Cai et al., 2021), revealed high levels of ILS in node D and ancestor of Heliophilodae (supplemental Figure 36). A comparison between the real distribution of tree-to-tree distances and the simulated distribution of Robinson and Foulds tree-to-tree distances revealed a largely overlapping pattern (supplemental Figure 37) (Bogdanowicz et al., 2012). In addition, a strong positive correlation was observed between branch lengths and ICA values (Pearson's correlation coefficient $R = 0.96$, $P = 2.2e{-}16$; supplemental Figure 38) (Zhou et al., 2022). These results suggest that ILS has at least partially contributed to the phylogenetic conflicts described above.

Ancestral core Brassicaceae karyotype (CBK)

We next reconstructed the ancestral karyotype of the MRCA for the core Brassicaceae using a synteny-based gene-family-clustering approach (Wu et al., 2023). Specifically, we used our reliable phylogeny based on high-quality genomes of nine diploid species representing major supertribes and the basal Aethionemeae, which have not undergone additional WGDs. We excluded species in Heliophilodae from reconstruction of

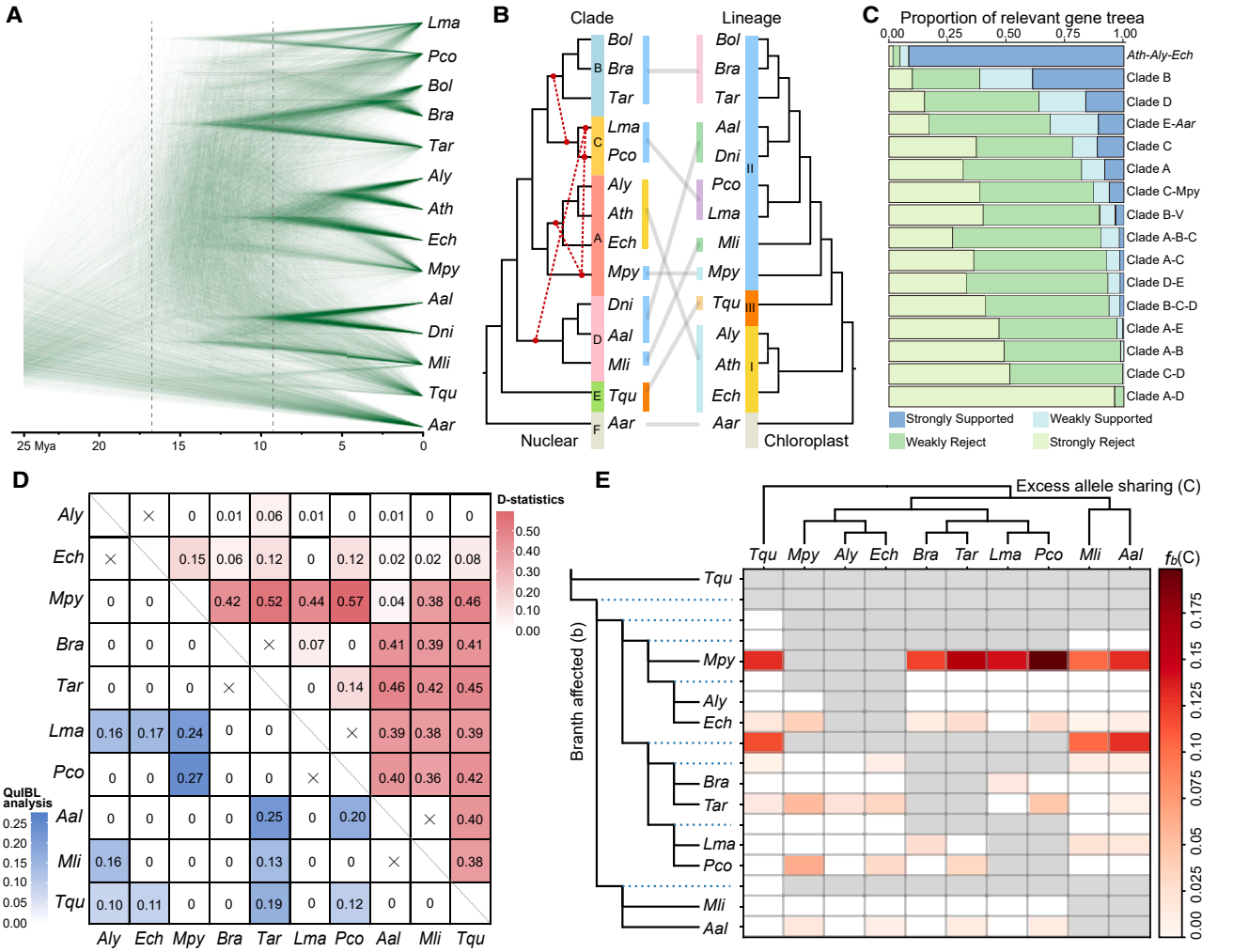


Figure 2. Pervasive topology discordance, prevalent hybridization, and introgression in Brassicaceae.

(A) Cloudogram inferred from 1463 SCOGs (*Orthofinder*). Scale in mya.

(B) Extensive conflicts between plastome-based (right panel) and nuclear-genome-based (left panel) species ML trees using concatenated data. Introgression events are shown as broken red lines on the nuclear tree.

(C) Gene-tree compatibility as revealed by the portion of gene trees that are highly (weakly) supported or rejected. Weakly rejected refers to those not in the tree but compatible if low support branches (<75%) are contracted.

(D) Tests for introgression with *D* statistics (upper right panel) and *QuIBL* analysis (lower left panel). Heatmaps of mean pairwise *D* per species pair and the mean total proportion of introgressed loci per species pair inferred with *QuIBL*.

(E) Test for introgression, with identification of excess sharing of derived alleles via the branch-specific statistic f_b(C) approach. The branch-specific statistic f_b(C) value indicates excess sharing of derived alleles between a given branch **(b)** on the y-axis, relative to its sister branch, and species C on the x-axis.

the CBK because of their complex evolutionary history and potentially hybrid origin (Hendriks et al., 2023).

We first detected syntenic gene pairs between each of the 9 species. A minimum of 13 982 pairs were identified between *Dni* and *Aar* and a maximum of 20 472 pairs were found between *Ath* and *Cru* (supplemental Table 20). Next, we identified a total of 118 980 non-redundant syntenic groups, among which 9702 were conserved genes (putative protogenes, or pPGs) in all 9 species and represented a gene pool for the MRCA of the Brassicaceae. A total of 11 682 pPGs were present in 8 core Brassicaceae species, and 15 778 pPGs were found in 4 species representing supertribes Camelinodae and Brassicodae (supplemental Table 21). By analyzing the synteny between the 9 extant species and the previously predicted common ancestor for Brassicaceae (Schranz et al., 2006; Lysak et al., 2016), we refined the GB boundaries using the 9702 pPGs and identified 43 additional breakpoints dividing the 22 conserved GBs (Lysak et al., 2016) into 65 GBs, which were present in all 9 species (supplemental Figure 39; supplemental Table 22). Finally, with consideration of the phylogenetic topology and these 65 GBs, we built the karyotypes of extant species and traced the evolutionary scenario in Brassicaceae by reconstructing the ancestral karyotypes for all seven internal nodes (1–7) using *Aar* as the outgroup (Figure 3; Table 1). The predicted chromosomal pattern at node 7, i.e., the CBK, represents the ancestral karyotype of core Brassicaceae.

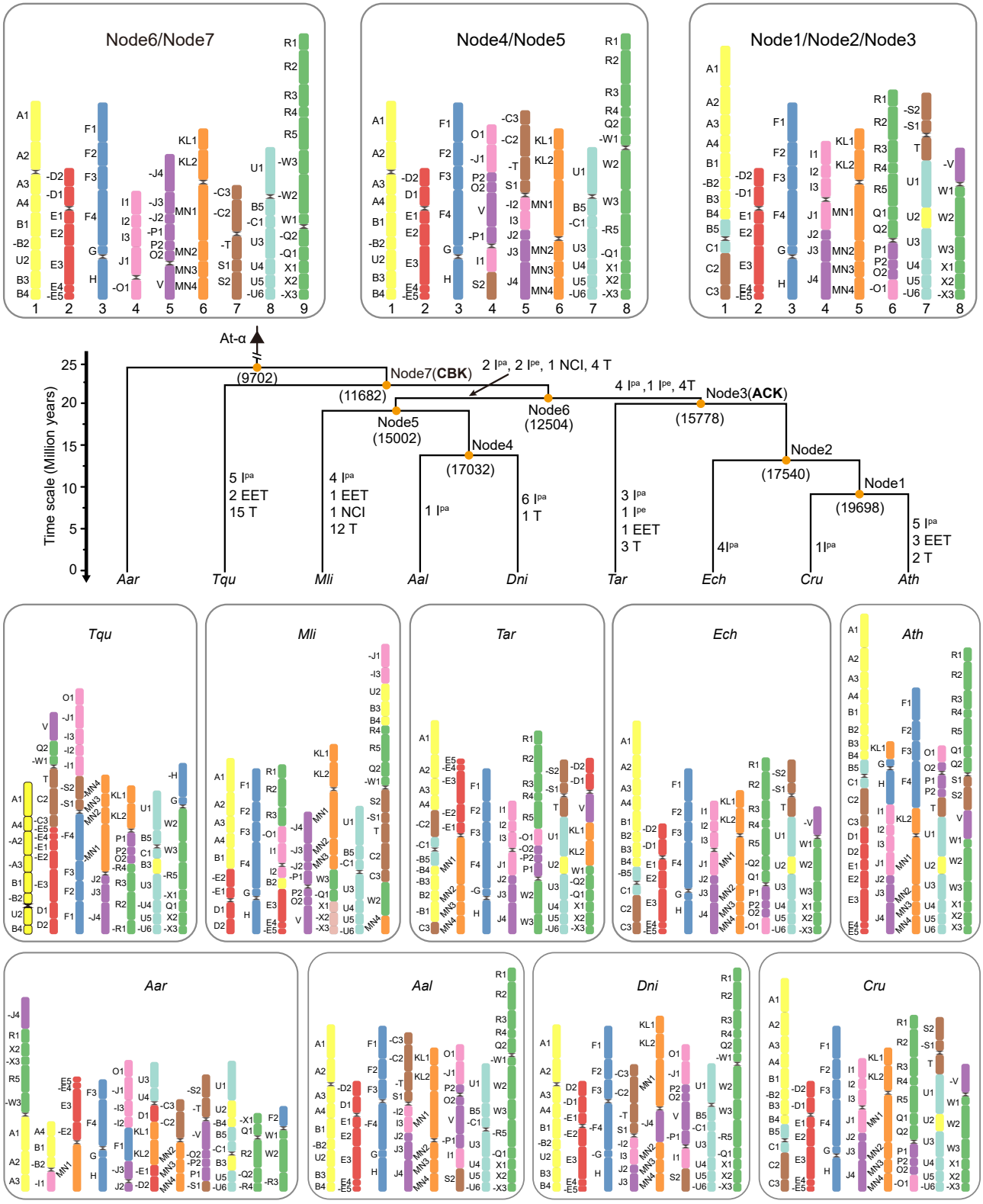


Figure 3. The ancestral core Brassicaceae karyotype (CBK) and the evolutionary history of Brassicaceae karyotypes.
Numbers in brackets indicate the conserved pPGs at each node (see supplementary Table 21). Rearrangement processes include nested chromosome insertions (NCI), end-to-end translocations (EET), and translocations (T), as well as paracentric (I^{pa}) and pericentric (I^{pe}) inversions. Black triangles indicate WGD events in Brassicaceae (At-α), and black sandglass-like symbols in karyotypes represent centromeres.

CBK	GB	GB start	GB end	pPG start	pPG end
CBK1	A1	AT1G01010	AT1G08100	AT1G01010	AT1G07650
	A2	AT1G08110	AT1G12960	AT1G08540	AT1G12950
	A3	AT1G12970	AT1G16610	AT1G12970	AT1G16610
	A4	AT1G16630	AT1G19840	AT1G16650	AT1G19840
	B1	AT1G19850	AT1G24256	AT1G19850	AT1G24140
	B2	AT1G24260	AT1G27280	AT1G24310	AT1G27210
	U2	AT4G24160	AT4G27730	AT4G24740	AT4G27540
	B3	AT1G27290	AT1G30755	AT1G27320	AT1G30755
	B4	AT1G30757	AT1G32750	AT1G30760	AT1G32750
CBK2	D2	AT1G61210	AT1G56210	AT1G61210	AT1G56230
	D1	AT1G64670	AT1G61215	AT1G64670	AT1G61240
	E1	AT1G64960	AT1G67270	AT1G65020	AT1G67260
	E2	AT1G67280	AT1G71100	AT1G67530	AT1G71100
	E3	AT1G71110	AT1G78310	AT1G71110	AT1G78310
	E4	AT1G78320	AT1G79720	AT1G78380	AT1G79720
	E5	AT1G79730	AT1G80950	AT1G79730	AT1G80950
CBK3	F1	AT3G01015	AT3G07530	AT3G01015	AT3G07490
	F2	AT3G07540	AT3G12180	AT3G07540	AT3G12180
	F3	AT3G12190	AT3G16010	AT3G12200	AT3G16010
	F4	AT3G16020	AT3G25520	AT3G16050	AT3G25470
	G	AT2G05170	AT2G07690	/	/
	H	AT2G10940	AT2G20900	AT2G13810	AT2G20890
CBK4	I1	AT2G20920	AT2G25260	AT2G20930	AT2G25050
	I2	AT2G25270	AT2G27540	AT2G25800	AT2G27170
	I3	AT2G27550	AT2G31035	AT2G27900	AT2G31010
	J1	AT2G31040	AT2G35850	AT2G31060	AT2G35610
	O1	AT4G00026	AT4G03190	AT4G00026	AT4G03190
CBK5	J4	AT2G41420	AT2G48150	AT2G41420	AT2G48080
	J3	AT2G37670	AT2G41417	AT2G37670	AT2G41290
	J2	AT2G35860	AT2G37660	AT2G35880	AT2G37650
	P1	AT4G12620	AT4G09680	AT4G12620	AT4G09830
	P2	AT4G09670	AT4G07390	AT4G09610	AT4G08280
	O2	AT4G03200	AT4G05450	AT4G03200	AT4G05430
	V	AT5G47810	AT5G42130	AT5G47540	AT5G42340
CBK6	KL1	AT2G01060	AT2G05160	AT2G01060	AT2G04410
	KL2	AT3G25540	AT3G32960	AT3G25540	AT3G30340
	MN1	AT3G42180	AT3G52970	AT3G42880	AT3G52960
	MN2	AT3G52980	AT3G56550	AT3G52990	AT3G56550
	MN3	AT3G56560	AT3G59550	AT3G56570	AT3G59520
	MN4	AT3G59570	AT3G63530	AT3G59600	AT3G63530

Table 1. The ancestral core Brassicaceae karyotype

(Continued on next page)

CBK	GB	GB start	GB end	pPG start	pPG end
CBK7	C3	AT1G53720	AT1G56190	AT1G53730	AT1G56140
	C2	AT1G47960	AT1G53710	AT1G47980	AT1G53710
	T	AT4G12700	AT4G16240	AT4G12840	AT4G16230
	S1	AT5G42110	AT5G39890	AT5G42080	AT5G39900
	S2	AT5G39880	AT5G32470	AT5G39860	AT5G36210
CBK8	U1	AT4G16250	AT4G24150	AT4G16250	AT4G24150
	B5	AT1G32760	AT1G37130	AT1G32760	AT1G36310
	C1	AT1G43020	AT1G47940	AT1G43190	AT1G47720
	U3	AT4G27740	AT4G32990	AT4G27745	AT4G32780
	U4	AT4G33000	AT4G35730	AT4G33560	AT4G35730
	U5	AT4G35733	AT4G38100	AT4G35740	AT4G38100
	U6	AT4G38120	AT4G40100	AT4G38120	AT4G40100
CBK9	R1	AT5G23000	AT5G19350	AT5G22940	AT5G19350
	R2	AT5G19340	AT5G13390	AT5G19330	AT5G13390
	R3	AT5G13380	AT5G08540	AT5G13360	AT5G08540
	R4	AT5G08535	AT5G06740	AT5G08535	AT5G06750
	R5	AT5G06730	AT5G01010	AT5G06440	AT5G01030
	W3	AT5G56550	AT5G60800	AT5G56550	AT5G60800
	W2	AT5G49620	AT5G56540	AT5G49920	AT5G56530
	W1	AT5G47820	AT5G49610	AT5G47820	AT5G49580
	Q2	AT5G26220	AT5G23010	AT5G26220	AT5G23010
	Q1	AT5G30510	AT5G26230	AT5G28910	AT5G26230
	X1	AT5G60805	AT5G63090	AT5G60820	AT5G63090
	X2	AT5G63100	AT5G65925	AT5G63120	AT5G65910
	X3	AT5G65930	AT5G67640	AT5G65950	AT5G67640

Table 1. *Continued*

CBK1–9, pseudochromosomes of the CBK; GB, genomic block, named corresponding to the 22 ACK blocks with numbers indicating the breakdown of each ACK block; pPGs, putative protogenes conserved in all nine extant species.

In contrast to the 8 pseudochromosomes of the ACK, the CBK is characterized by 9 haploid chromosomes. This finding is consistent with a recent study that included only 3392 strictly conserved syntenic genes (supplemental Figure 40) (Walden and Schranz, 2023). The karyotypes for node 1 to node 3 are identical, and so also for node 4 and node 5. Notably, node 6 has the same pattern as the CBK (Figure 3). These results imply relatively short divergence times between node 1 and node 3, node 4 and node 5, and the CBK and node 6. It is noteworthy that the karyotype of node 3, the common ancestor of Camelinodae and Brassicodae, is nearly identical to that of the ACK, with only a few GBs having undergone inversion (Figures 3 and 4). Overall, we observed considerable consistency between the CBK and the ACK. Specifically, three pseudochromosomes of node 3 (_2/_3/_5) correspond to the three ancestral chromosomes of the CBK (CBK2/3/6). By contrast, node 3_1 and node 3_7 originated from CBK1/8/7, and node 3_4/_6/_8 were derived from CBK4/5/9, respectively. This transformation occurred through a series of translocations, including both paracentric and pericentric inversions (Figure 4A). Similarly, five pseudochromosomes of node 5 (_1/_2/_3/_6/_7) appear to have remained in their ancestral states, whereas the remaining three (_4/_5/_8) are derived from four translocations, one nested chromosome insertion, two paracentric inversions, and one pericentric inversions (Figure 4B). Because there is only one high-quality chromosome-level genome assembly available (*Tqu*), we were unable to infer the ancestral genome for Hesperodae. However, a comparison between the *Tqu* genome and the reconstructed ancestral karyotype for Hesperodae (CEK, $n = 7$) with a comparative cytogenetic approach (Mandakova et al., 2017) revealed extensive chromosomal rearrangements, with all *Tqu* chromosomes having undergone batches of rearrangements (supplemental Figure 41). Finally, frequent centromere repositioning may have occurred (supplemental Table 23), consistent with previous reports (Mandakova et al., 2020).

Previous research has reported chromosome counts of both $n = 7$ and $n = 8$ for *Mli*, with $n = 8$ being more frequent in Alysseae (Spaniel et al., 2015). This phenomenon has also been observed in *Camelina microcarpa*, likely owing to variation among different genetic populations (Brock et al., 2022). Interestingly, the ancestors of node 4 (Arabideae) and node 5 (Arabodae) had chromosome counts of $n = 8$ (Figure 3).

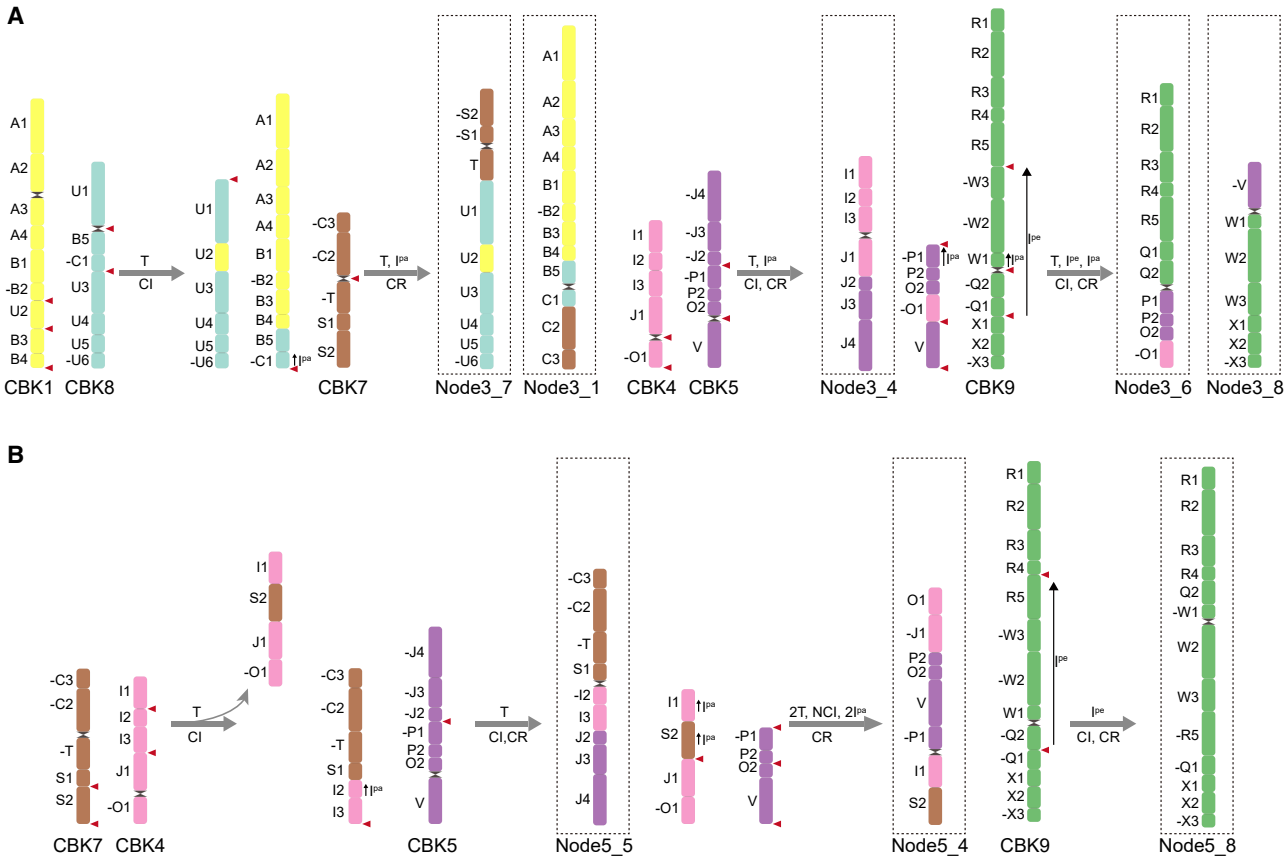


Figure 4. Karyotype evolutionary scenarios from the CBK to node 3 (A) and node 5 (B).
Rearrangement processes include nested chromosome insertions (NCI), end-to-end translocations (EET), and translocations (T), as well as paracentric (I^{pa}) and pericentric (I^{pe}) inversions. Black sandglass-like symbols represent centromere locations. Red triangles denote the positions of genomic fission or fusion.

Compared with *Aal* and *Dni*, two species also in Arabodae, *Mli* seems to have experienced more intense chromosomal recombination, resulting in a decrease in chromosome number to seven from the ancestral state of eight. More specifically, the ancestral node 5_2 chromosome in *Mli* is divided into two segments, each of which is fused to the Mli_1 and Mli_3 chromosomes, respectively (Figure 3; supplemental Figure 42). Similarly, extensive chromosomal rearrangements were identified in *Mpy* (supplemental Figure 43). Considering its ancient hybridization history, jumped phylogenomic position, and shared PCK-specific chromosomes with *Pco*, the diploid *Mpy* appears to have originated from homoploid hybridization between node 3 (ACK; n = 8) and ancPCK (n = 8) (Guo et al., 2021; Hu et al., 2021).

Expansion and expression diversification of class I *TPS* genes

We next evaluated whether any genes or gene families were associated with the early Brassicaceae radiation by searching for node-specific gene family expansion and contraction patterns (nodes A–F; Figure 1F; supplemental Figures 44–48; supplemental Tables 24–29). Among these nodes, only a small number of gene families were found to be expanded or contracted. The most significant expansion (495 families) and contraction (157 families) occurred at node B, the MRCA of supertribes Brassicodae and Heliophilodae (Figure 1F). Gene ontology analysis revealed that defense-related genes were significantly enriched at nodes B, C, and E (supplemental Figures 45, 46, and 48). Node F (the MRCA of core Brassicaceae) was enriched for genes encoding TPS enzymes, as well as genes related to several other pathways (Figure 5A).

TPS genes encode enzymes responsible for the biosynthesis of T6P, a signaling molecule involved in the regulation of abiotic stress tolerance and developmental processes such as flowering time and embryo development (Fichtner and Lunn, 2021). *A. thaliana* has 11 *TPS* genes that form 2 distinct classes: I (*TPS1* to *TPS4*) and II (*TPS5* to *TPS11*) (Leyman et al., 2001; Avonce et al., 2006; Lunn, 2007). Phylogenetic and synteny analyses revealed that *TPS1* is highly conserved and exhibits good collinearity between Brassicaceae and *C. violacea* (Cleomaceae), whereas *TPS2* to *TPS4* are present only in Brassicaceae (Figure 5B; supplemental Figure 49; supplemental Table 30). As collinearity was not identified between *TPS1*s and *TPS4*s (supplemental Figure 50), *TPS4* was likely generated by random duplication of *TPS1* in the MRCA of Brassicaceae and stably inherited with very good collinearity in core Brassicaceae. Notably, *Pco* contains two copies (*PcoTPS4.1* and *PcoTPS4.2*), which are tandemly duplicated on

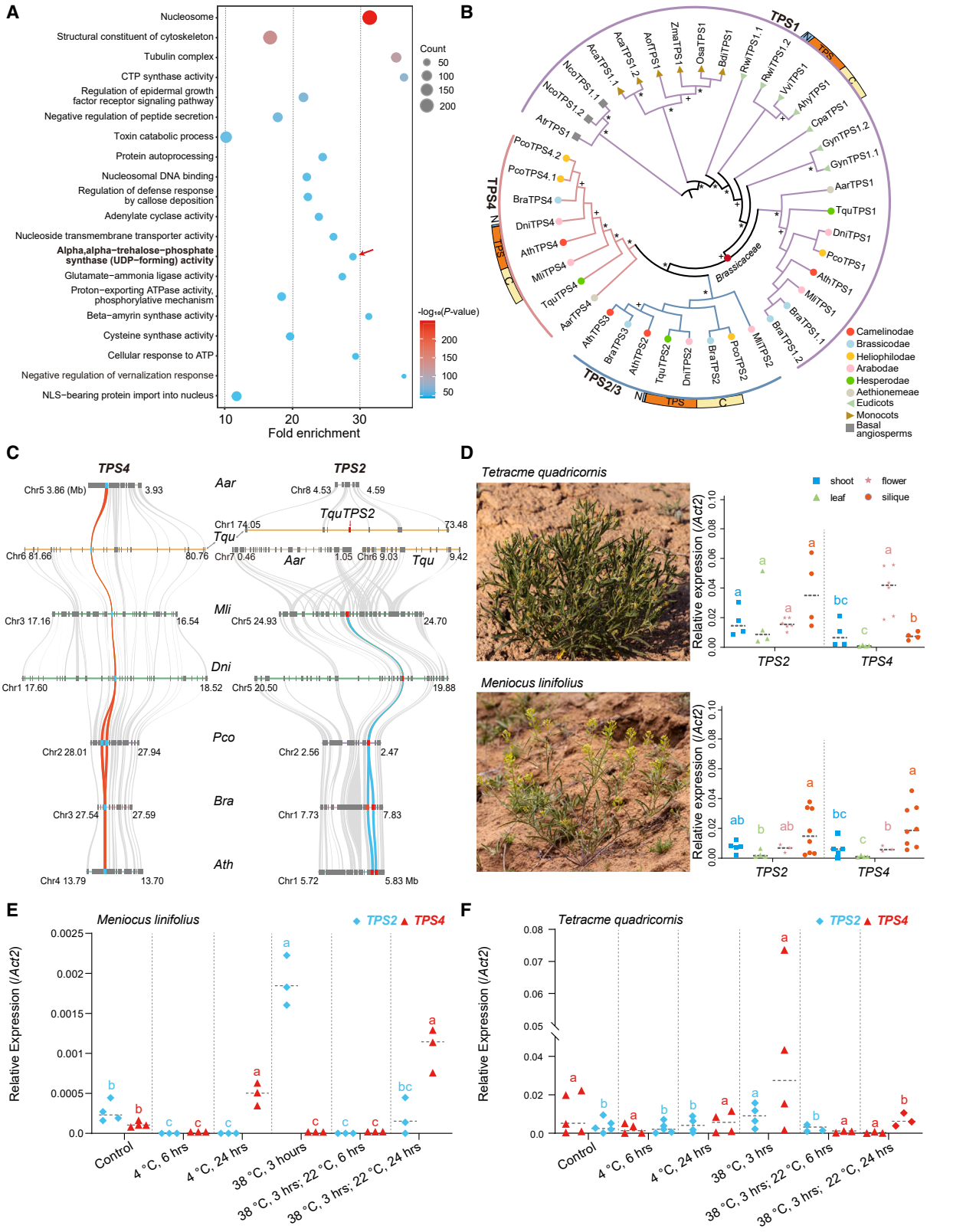


Figure 5. Brassicaceae-specific expansion of *Trehalose-6-Phosphate (T6P) synthase 1* (*TPS1*) genes.

(A) Gene ontology (GO) enrichment of molecular functions for expanded gene families on node F (Figure 1F) corresponding to the split between core Brassicaceae and Aethionemeae. GO terms in bold and marked with red arrows highlight the expansion of trehalose-6-phosphate (T6P) synthase-related genes.

(legend continued on next page)

chromosome 2 (Figure 5B and 5C). *TPS2-like* genes are expanded in Hesperodae (*Tqu*) and the remaining core Brassicaceae supertribes via two independent random duplication events. The *TquTPS2* locus on chromosome 1 exhibits good synteny with *Aar* chromosome 4, whereas no *TPS2-like* sequence was identified in the blocks harboring *TPS2-like* genes in supertribes Camelinodae, Brassicodae, Arabodae, and Heliophilodae. Tandemly duplicated *TPS3s* were detected only in *Arabidopsis* and *Brassica* (Figure 5B). All the Brassicaceae TPS2 and TPS4 proteins lacked the N-terminal domain, which harbors a nuclear localization signal that targets the AthTPS1 protein primarily to the nucleus (Figure 5B; supplemental Figure 51) (Fichtner et al., 2020). By contrast, nearly all Brassicaceae TPS1s contain the N-terminal domain. These results suggest that the Brassicaceae-specific *TPS2*- and *TPS4-like* genes may have different evolutionary histories.

In *Arabidopsis*, class I TPSs have catalytic activity (Van Dijck et al., 2002). *TPS1* is broadly expressed, *TPS2* and *TPS4* are expressed at low levels and exclusively in developing seeds, and *TPS3*, a potential pseudogene, is only minimally expressed (supplemental Figures 52 and 53) (Vandesteene et al., 2010). Both *TquTPS1* and *MliTPS1* were highly and nearly universally expressed in all tissues sampled from wild plants harvested in Xinjiang, China (supplemental Figure 54). A similar pattern was observed for *AthTPS1* (supplemental Figure 52). Despite their extremely low levels of expression under laboratory conditions, both *MliTPS2* and *MliTPS4* were expressed at relatively high levels in siliques and seeds sampled from plants growing under fluctuating natural conditions (Figure 5D and 5E). These results suggest that *TPS2* and *TPS4* genes from supertribes Camelinodae, Brassicodae, Arabodae, and Heliophilodae may exhibit similar expression patterns. *TquTPS4* is expressed at relatively high levels in flowers (Figure 5D), indicating that this gene may play a role in the regulation of flower development. Notably, these *TPS* genes exhibit diverse expression patterns under fluctuating temperatures (Figure 5F). The lineage-specific expansion of class I *TPSs* was thus followed by diversification of expression, and likely function, during the early evolution of Brassicaceae.

DISCUSSION

Because of the economic and scientific importance of Brassicaceae, its phylogeny and genome evolution remain at the forefront of plant evolutionary biology. Here, we generated high-quality chromosome-level genome assemblies for two species representing supertribes Hesperodae and Arabodae and identified the complex genomic features that accompanied the early evolution of Brassicaceae.

Using available high-quality genomes representing all supertribes and the basal Aethionemeae, we produced a phylogeny consistent with recent reports (Huang et al., 2016; Kiefer et al., 2019; Nikolov et al., 2019; Hendriks et al., 2023). Overall, Hesperodae (*Tqu*) appears to have diverged successive to the basal Aethionemeae (*Aar*) followed by supertribes Camelinodae, Brassicodae, Arabodae, and Heliophilodae in the late Oligocene to early Miocene (19.3–24.4 mya; Figure 1). Intriguingly, this was coincident with the accelerated uplift of the Tibetan Plateau, which was followed by significant aridity in the Asian interior and monsoon intensification in Eastern Asia (Kagale et al., 2014; Ding et al., 2020; Miao et al., 2022). These results thus link the diversification of Brassicaceae to global environmental change.

The observed phylogenetic discordances correspond well with the complex evolutionary history of Brassicaceae, in which frequent and ancient inter-supertribe hybridizations have been identified (Figure 2). Such patterns, marked by extensive gene-tree heterogeneity, have been documented previously and can result from rampant hybridization events between members of closely or distantly related groups (Nikolov et al., 2019; Hendriks et al., 2023). Furthermore, the influence of ILS during early radiation, leading to low resolution of deeper nodes within Brassicaceae, cannot be overlooked. Therefore, reducing such groups exclusively to existing models that strictly adhere to bifurcating trees significantly oversimplifies the reality and hinders our ability to accurately describe the evolutionary process (Cai et al., 2021; Guo et al., 2023; Hendriks et al., 2023).

The availability of high-quality diploid genomes without additional WGD events for species representing nearly all major supertribes as well as the basal Aethionemeae makes it possible to reconstruct the CBK (Table 1; Figures 3 and 4). The previously proposed ACK is based mainly on genomes representing Camelinodae and Brassicodae (Schranz et al., 2006; Lysak et al., 2016). Using the ACK, phylogenomic variation and chromosomal evolution have been characterized in Brassicodae (PCK, $n = 7$) (Mandakova and Lysak, 2008; Cheng et al., 2013) and Hesperodae (CEK, $n = 7$) (Mandakova et al., 2017) and tribes Thlaspideae (Bayat et al., 2021), Biscutelleae (Geiser et al., 2016; Mandakova et al., 2018; Guo et al., 2021), and Arabideae (Willing et al., 2015; Mandakova et al., 2020), as well as basal Aethionemeae (Walden et al., 2020a). Very recently, Walden and Schranz (2023) reported their effort in reconstruction of the ancestral karyotype covering the major clades (Walden and Schranz, 2023). Their ancestral karyotype reconstruction, which also features 9 haploid chromosomes, largely aligns with our CBK, albeit with less resolution and significant structural differences (supplemental Figure 40). This is very likely due to the identification of fewer

(B) An ML tree shows the phylogenetic clustering of class I TPS proteins in 19 angiosperms, including 7 Brassicaceae species representing all supertribes and the basal Aethionemeae. The * and + indicate bootstrap supports of >90% and 50%–90%, respectively. The summarized protein structure for each subgroup is shown, with N, TPS, and C indicating the N-terminal domain, TPS domain, and C-terminal domain, respectively. Supertribes or outgroups are indicated with colored circles, triangles, and squares, as shown at the bottom right.
(C) The collinearity relationships of subgroups *TPS2* and *TPS4* in Brassicaceae. The red and blue curves show the positions and syntenic relationships of *TPS2* (red squares) and *TPS4* (blue squares) between selected species. Gray curves show the surrounding syntenic genes between species.
(D) Relative expression patterns of *TPS2*s and *TPS4*s in different tissues (right panels) of plants grown under natural conditions in Fukang County, Xinjiang Province, China. A minimum of four biological replicates of leaves, shoots, flowers, and siliques were collected and pooled from at least five individuals of each species (left panels).
(E and F) Relative expression patterns of *TPS2*s and *TPS4*s under fluctuating temperature conditions for *Mli* **(E)** and *Tqu* **(F)**. A minimum of three biological replicates were used for each assay. Lowercase letters indicate statistically significant differences (ANOVA, $P < 0.05$).

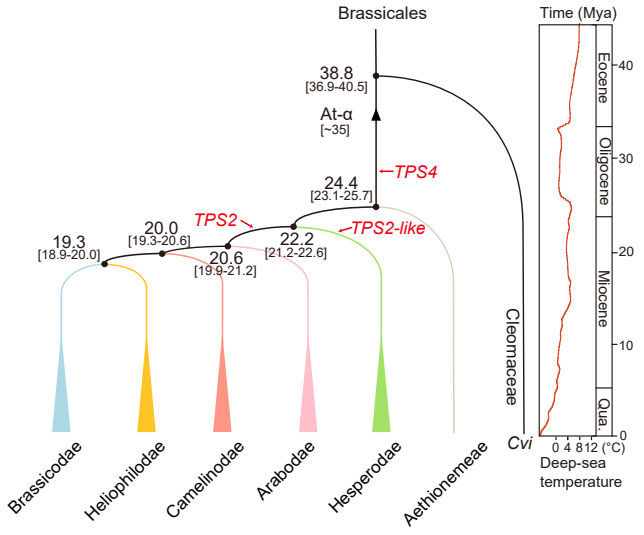


Figure 6. A simplified model of evolutionary events in Brassicaceae.

After splitting from Aethionemeae, the progenitors of core Brassicaceae (colored triangles) diversified in a short period between 19.3 and 24.4 mya during the late Oligocene to the early Miocene. The first random duplication of *TPS4*s occurred after the At-α WGD. Two independent random duplication events doubled the *TPS2-like*s and *TPS2*s in Hesperodae and the MRCA of other supertribes (Brassicodae, Heliophilodae, Camelinodae, and Arabodae), respectively. The deep-sea temperatures are modified from Zachos et al. (2001).

strictly conserved syntenic genes (3392 compared with our 9702) caused by the segmental genome of *Esy* (Jiao et al., 2017). Alternatively, inclusion of the rogue *Mpy* genome, which features significant hybridization and has a contentious phylogenetic position, may have lowered the accuracy of their effort.

Although the ACK represents essentially just the ancestors of Camelinodae and Brassicodae (Schranz et al., 2006; Lysak et al., 2016), its 22 GBs are well conserved throughout Brassicaceae (supplemental Figure 39; supplemental Table 22). Compared with the 8 chromosomes of ACK, the CBK exhibits a more ancestral state with 9 chromosomes and 65 GBs (Figures 3 and 4; supplemental Figure 39; Table 1). It is noteworthy that the reconstructed ancestral karyotype for node 3 (the common ancestor of Camelinodae and Brassicodae) is highly consistent with the ACK and also with the extant *Cru* and *Ech*, indicating the accuracy of our analyses. However, to fully resolve the ancestral karyotype for the whole family, a suitable high-quality genome at the chromosome-level for Cleomaceae is still necessary. Nonetheless, our data provide a valuable resource and a foundation for more in-depth comparative analyses of genome evolution in Brassicaceae.

The available high-quality genomes and phylogeny enabled us to more comprehensively evaluate the molecular features associated with the Brassicaceae radiation, especially during the late Oligocene to early Miocene. Our analyses in angiosperms revealed that, after the At-α WGD, Brassicaceae specifically duplicated its class I *TPS1* into *TPS4*s (Figures 5 and 6). However, this expansion seems not to align well with the known WGD radiation lag-time model (Schranz et al., 2012). The divergence between core Brassicaceae and the basal Aethionemeae during the late Oligocene and early Miocene is coincident with the accelerated uplift of the Tibetan Plateau. The uplift was followed by increasing aridity in the Asian interior, intensifying Eastern Asian monsoons, fluctuating daily temperatures (Zachos et al., 2001; Kagale et al., 2014; Ding et al., 2020; Miao et al., 2022), as well as increased diversification in Brassicaceae. This geologic period also saw two independent random duplications: *TPS4-like* produced the *TPS2/3-like* genes in Hesperodae and the successive supertribes Camelinodae, Brassicodae, Arabodae, and Heliophilodae, respectively (Figure 6). It should be noted that duplication of class I *TPS* genes seems to be an evolving and highly dynamic process, as demonstrated by a further tandem duplication of *TPS2/3-like* genes in Camelinodae and Brassicodae. Intriguingly, the duplicated *TPS2/3-like* and *TPS4-like* genes exhibit highly varied expression under both natural and disturbed temperatures, a signal of functional diversification (Figure 5). However, the precise molecular and genetic connections between these duplications and paleoclimatic changes require further in-depth study.

The assembly of high-quality genomes for Hesperodae and Arabodae provides fresh insights into the ancestral karyotypes and molecular features associated with the complex evolutionary history of Brassicaceae. The lineage-specific and dynamic expansion of key flowering-time regulators may have served as an evolutionary gate, with more efficiency and precision, to sense and respond to fluctuating energy availability under ever-changing environmental conditions.

MATERIALS AND METHODS

Plant material and DNA sequencing

Seeds of *M. linifolius* (*Mli*) (voucher ID TanDY0110) and *T. quadricornis* (*Tqu*) (voucher ID TanDY0709) were collected in Karamay City and Fukang City, respectively, in Xinjiang Autonomous Region, China. The seeds were preserved in the Germplasm Bank of Wild Species, Kunming Institute of Botany, Chinese Academy of Sciences. For genome sequencing, seeds were cultured on half-strength Murashige and Skoog medium and grown under long-day conditions. Fresh leaves were harvested from each single plant for genomic DNA extraction.

For short-read sequencing, 150-bp paired-end reads were generated using the Illumina NovaSeq platform (Illumina, USA) at ∼50×. For long-read sequencing, Single-Molecule Real-Time PacBio Genome Sequencing was performed using the PacBio Sequel II platform with the circular consensus sequencing model (Pacific Biosciences). After quality control, we obtained 38.3 and 52.4 Gb of high-quality HiFi reads for *Mli* (∼147.3×) and *Tqu* (∼72.3×), respectively (supplemental Table 2). For scaffolding, Hi-C sequencing was performed to generate ∼100× data for both species. For genome annotation, total RNA from mixed samples (including young leaves, stems, roots, and flowers) was used for both Illumina paired-end (150 bp) and full-length Nanopore long-read (Oxford Nanopore Technologies, UK) sequencing.

Genome size estimation and assembly

Genome sizes were initially estimated using flow cytometry with *Solanum lycopersicum* (0.88 Gb) as a reference. In brief,

approximately 20 mg of fresh leaves were harvested and placed in a Petri dish containing 1 ml of ice-cold nuclei isolation buffer (45 mM $MgCl_2 \cdot 6H_2O$, 20 mM MOPS, 30 mM $Na_3C_6H_5O_7$, 1% [w/v] PVP 40, 0.2% [v/v] Triton X-100, 10 mM Na_2 EDTA [pH 7.0]). The tissues were minced with a sterile razor blade, mixed, and filtered through a 42-μm nylon mesh into a sample tube. Later, a stock DNA fluorochrome solution was added together with 50 mg/ml PI and 50 mg/ml RNase, and the mixture was incubated on ice prior to analysis. Sample measurement and data analysis were performed using a BD FACScalibur and *Modifit* (v3.0), respectively (Duda et al., 1999). All measurements were carried out in triplicate, and mean values are reported. Genomic surveys were performed using *Jellyfish* (v2.2.10) (Marcais and Kingsford, 2011) and *GenomeScope2* (Vurture et al., 2017) with k-mer = 19 and 50× Illumina short reads. High-quality circular consensus sequencing reads were used for genome assembly with default parameters of the *hifiasm* (v0.12) pipeline (Cheng et al., 2021). Contigs were scaffolded using ~100× Hi-C data according to previously described methods. In brief, the Hi-C data were filtered, mapped, and evaluated using *HiC-pro* (v2.10.0) (Servant et al., 2015) and *BWA* (v0.7.10-r789) (Li and Durbin, 2009). Next, *LACHESIS* was used to cluster and reorder the corrected contigs into pseudochromosomes (Burton et al., 2013), using the following parameters: *cluster min re sites* = 22, *cluster max link density* = 2, *cluster non-informative ratio* = 2, *order min n res intrun* = 10, *order min n res in shreds* = 10.

Genome quality assessment

The quality and completeness of both genome assemblies were assessed using three approaches. We first judged assembly quality by mapping back the Illumina reads to the genomes using default parameters of *BWA-MEM* (Li, 2013). Next, the assemblies were subjected to BUSCO (v5.4) analysis (Simao et al., 2015). Finally, the LTR assembly indices of the two genomes were evaluated using published procedures (Ou et al., 2018; Ou and Jiang, 2018).

Annotation of repetitive sequences

Tandem repeats were identified with *TRF* (v4.07) using the parameters: *1 1 2 80 5 200 2000 -d -h* (Benson, 1999). Simple sequence repeats were identified with *MISA* (v2.1) (Beier et al., 2017). Next, we performed *de novo* and homology-based prediction to detect TEs. The *Extensive de novo TE Annotator* (*EDTA*) pipeline (v2.0.1) (Ou et al., 2019), integrating *RepeatMasker* (Tarailo-Graovac and Chen, 2009), *LTR_Finder* (Xu and Wang, 2007), *LTRharvest* (Ellinghaus et al., 2008), and *LTR_retriever* (Ou and Jiang, 2018), was used to classify the TEs as DNA transposons or retrotransposons. Lineages for the *Copia* and *Gypsy* superfamilies were identified using *Tesorter* (Zhang et al., 2022). The insertion time (*T*) of each type of intact LTR retrotransposon was estimated using the formula $T = K/2r$ with a substitution rate (r) of 8.22×10^{-9} substitutions per site per year (Kagale et al., 2014) and K representing the genetic distance.

Gene annotation

Gene annotation was performed using *ab initio*, homology-based, and transcriptome-based predictive methods after masking of all repetitive regions. For *ab initio* prediction, *Augustus* (v2.4) (Stanke and Waack, 2003) and *SNAP* (v2006-07–28) (Korf, 2004) were used with default parameters. Five Brassicaceae species (*A. thaliana* [Arabidopsis Genome, 2000], *B. napus* [Song et al., 2020], *B. nigra* [Perumal et al., 2020], *D. nivalis* [Nowak et al., 2021], and *Isatis indigotica* [Kang et al., 2020]) were used for homology-based prediction with *GeMoMa* (v1.7) (Keilwagen et al., 2018). For transcriptome-based prediction, *Trinity* (v2.11) (Grabherr et al., 2011) and *PASA* (v2.0.2) (Haas et al., 2003) were used to generate *de novo* assemblies, and *Hisat* (v2.0.4) (Kim et al., 2015), *Stringtie* (v1.2.3) (Pertea et al., 2015), and *GeneMarkS-T* (v5.1) (Tang et al., 2015) were used to predict transcripts and genes. All gene models were integrated using *EVidenceModeler* (v1.1.1) (Haas et al., 2008) to generate a consensus gene set. To obtain untranslated regions and alternatively spliced isoforms, we used *PASA* to update the *gff3* file (two rounds). Finally, BUSCO (v5.4) was used to assess the completeness of the gene set.

Functional annotations were assigned to protein-coding genes by performing *BLAST* (v2.2.31) searches against the NR, GO, KEGG, KOG, and TrEMBL databases (Ye et al., 2006) with an e value cutoff of 10^{-5}. We also used *Blast2GO* (v4.1) (Conesa and Gotz, 2008) to search the GO and KEGG databases.

Four types of non-coding RNAs (microRNAs, transfer RNAs, ribosomal RNAs, and small nuclear RNAs) were annotated using *tRNAscan-SE* (v1.3) (Lowe and Eddy, 1997) for tRNA, *Rfam* (v12.0) (Kalvari et al., 2018) and *Infernal* (v1.1) (Nawrocki and Eddy, 2013) for snoRNA and snRNA, *Rfam* (v12.0) (Kalvari et al., 2018) and *Barrnap* (v 0.9) for rRNA, and *miRbase* (Kozomara et al., 2019) for miRNA. Pseudogenes were predicted with *GeBlastA* (v1.0.4) (She et al., 2009) and *GeneWise* (v2.4.1) (Birney et al., 2004).

Gene family classification and analyses

On the basis of recently published tree topologies (Huang et al., 2016; German et al., 2023; Hendriks et al., 2023), we selected 14 diploid species with high-quality chromosome-level assemblies (supplemental Table 16) representing all 5 supertribes of core Brassicaceae and the basal Aethionemeae. These included *A. thaliana* (Arabidopsis Genome, 2000), *A. lyrata* (Hu et al., 2011), *M. pygmaea* (Yang et al., 2021), and *E. cheiranthoides* (Zust et al., 2020) in Camelinodae; *B. rapa* (Belser et al., 2018), *B. oleracea* (Belser et al., 2018), and *T. arvense* (Geng et al., 2021) in Brassicodae; *Lobularia maritima* (Huang et al., 2020a) and *P. cornutum* (Hu et al., 2021) in Heliophilodae; *M. linifolius* (this study), *A. alpina* (Willing et al., 2015), and *D. nivalis* (Nowak et al., 2021) in Arabodae; *T. quadricornis* (this study) in Hesperodae; and the Aethionemeae species *A. arabicum* (Nguyen et al., 2019).

Orthofinder (v2.5.4) (Emms and Kelly, 2019) was used to identify putative gene families. For genes with alternative splicing variants, the longest transcript was selected. Protein sequence alignments were obtained with *MAFFT* (v7.475) (Katoh and Standley, 2013) and converted into the corresponding codon alignments with *PAL2NAL* (Suyama et al., 2006). *TrimAL* (v1.4, -gt 0.8 -cons 80) (Capella-Gutierrez et al., 2009) was used to extract the conserved sites of the multiple sequence alignments. The *PhyloSuite* (v1.2.2) pipeline (Zhang et al., 2020)

was used to concatenate the alignments into a super matrix. ML trees were generated with *IQ-TREE* (v2.1.2, -m MFP -bb 1000) (Nguyen et al., 2015). We skipped the divergence time estimation and followed the results of Hendriks et al. (2023), with the exception of the divergence time between *B. rapa* and *B. oleracea* (2.5–3.2 mya), which was obtained from *TimeTree* (Kumar et al., 2017). Gene family expansion and contraction were evaluated using *CAFÉ* (v5) (De Bie et al., 2006) based on a Poisson distribution. Genes in significantly expanded families were used for GO enrichment analysis with *clusterProfiler* (Wu et al., 2021).

WGD history prediction

Six species (*A. thaliana*, *B. rapa*, *P. cornutum*, *M. linifolius*, *T. quadricornis*, and *A. arabicum*) representing each Brassicaceae clade, as well as *C. papaya* (Yue et al., 2022) representing the basal Brassicales, were selected for WGD analyses. Syntenic blocks and collinear genes within and between species were identified using *WGDI* (0.5.6) with the parameter "*-icl*" (Sun et al., 2022). *Ks* values between collinear genes were estimated using the *Nei-Gojobori* approach implemented in *PAML* (v4.9) (Yang, 2007). Genes with $Ks < 0.15$ were excluded from further analysis. Median *Ks* values in each syntenic block were fitted for peak identification in *WGDI* using the "*-pf*" option. To justify the ploidy levels, we used dot plots of collinear genes and syntenic blocks to determine syntenic ratios between species.

Phylogenic analyses in Brassicaceae

We used both *Orthofinder* (v2.5.4) (Emms and Kelly, 2019) and *SonicParanoid* (Cosentino and Iwasaki, 2019) to identify orthologous genes for phylogenic reconstruction with both coalescent- and concatenation-based analyses at the nucleotide and protein levels. In total, six different alignments for each gene family were used to perform phylogenetic analyses: (1) amino acid alignments of *Orthofinder* genes, (2) amino acid alignments of *SonicParanoid* genes, (3) nucleotide alignments of *Orthofinder* genes, (4) nucleotide alignments of *SonicParanoid* genes, (5) nucleotide codon alignments with the third position removed for *Orthofinder* genes, and (6) nucleotide codon alignments with the third position removed for *SonicParanoid* genes. Because concatenated datasets do not account for the stochasticity of the coalescent process, we used *ASTRAL* (v5.7.8) (Zhang et al., 2018) to reconstruct the coalescent tree. The *ASTRAL* pipeline is statistically consistent under the multi-species coalescent model and is thus useful for handling ILS. Individual ML gene trees were constructed using *IQ-TREE*, with the same parameters listed above. For stem group nodes, we checked the bootstrap support values with 1000 simulations and summarized the topologies with bootstrap support values $\geq$0%, 10%, 30%, or 60%. To minimize errors resulting from poor taxon sampling, up to 28 species were added, including 27 Brassicaceae species and one Cleomaceae species (*C. violacea*) as an outgroup. Another species tree containing an additional 55 Brassicales species was analyzed with *STAG* (Emms and Kelly, 2018) using the low-copy gene set (shared ortholog groups) to further verify the stability of our phylogenetic structure. First, we used *Orthofinder* to cluster gene families based on the annotated proteins of 55 species, resulting in 5217 gene families shared among all species. Next, we built gene trees for each of these 5217 gene families using *FastTree* (Price et al., 2009). Lastly, we used *STAG* (Emms and Kelly, 2018) to build the final species tree based on the 5217 gene trees.

To minimize ortholog identification errors, synteny-based orthologous gene relationships were evaluated with *WGDI*, which does not require gene-family clustering. In addition to the 14 species mentioned previously, *C. rubella* (Slotte et al., 2013) was also included as a reference because its karyotype has been suggested to be similar to the ACK (Lysak et al., 2016). We identified intergenomic syntenic blocks between *C. rubella* and each species and intragenomic syntenic blocks within each species. On the basis of the similarity (*Ks* and *BLAST* scores) and completeness (covered genes and spanned gene length) of each syntenic block, the *WGDI* pipeline (*-bi* and *-a*) was used to assign syntenic blocks into putative sets. Genes exhibiting collinear relationships to *C. rubella* were used to infer the collinear gene tree with *IQ-TREE*. Finally, the synteny-based species tree was constructed with *ASTRAL*.

Plastome assembly and phylogenetic reconstruction

To investigate maternal phylogenetic relationships, we assembled and annotated the plastomes using *GetOrganelle* (v1.7.2a) (Jin et al., 2020) and *PGA* (Qu et al., 2019). The inverted repeat and coding region boundaries of each annotated gene were determined with *Geneious* (v9.0.2) (Kearse et al., 2012). The protein-coding sequences of the 14 species mentioned above, as well as other Brassicales species, were extracted, aligned, and concatenated using *Phylosuite*. ML trees were constructed with *IQ-TREE* using the settings *-bb 1000 -m MFP*.

Phylogenetic discordance assessment

To evaluate nuclear gene-tree discordance, we calculated the ICA to quantify the degree of conflict between the species and gene trees at each node. ICA values close to 1 indicate strong concordance for the bipartition defined by a given internode, whereas ICA values close to 0 indicate strong conflict. Negative ICA values suggest a conflict for the defined bipartition with other high-frequency bipartitions (Salichos et al., 2014; Zhou et al., 2022). ICA values were estimated in *RAxML* (v8.2.12) (Stamatakis, 2014), and the tree generated with *ASTRAL* was used as the species tree. Quartet frequencies of the internal branches of the species tree were calculated using *ASTRAL* ($t = 8$), and gene-tree compatibility (whether sister groups of each other) was analyzed with *DiscoVista* (v1.0) for each combination. *Phyparts* (Smith et al., 2015) was used to summarize the number of conflicting and concordant bipartitions between the species tree and the individual gene tree. To visualize the conflicts, we built a cloudogram using *DensiTree* (v2.2.6) (Bouckaert, 2010), and the input gene trees were time-calibrated with *TreePL* (v1.0) (Smith and O'Meara, 2012).

Detection of ancient hybridization, gene flow, and ILS

We first inferred the species network for modeling both ILS and introgression with the maximum pseudo-likelihood pipeline *PhyloNet* (v3.8.4) using the *InferNetwork_MPL* command (Than et al., 2008). Network searches were performed by allowing between one and four reticulation events and using 10 runs for each search. To estimate the optimum number of reticulations, we optimized the branch lengths and inheritance probabilities and computed the likelihood of the best-scored network for each of

the three maximum reticulation event searches. Next, we used *QuIBL*, which compares branch length distributions across gene trees, to test hypotheses regarding whether phylogenetic discordance between all possible triplets could be explained by ILS alone or by a combination of ILS and introgression (Edelman et al., 2019). The Bayesian information criterion (BIC) was used to test whether the discordance between gene and species trees was due more to ILS or introgression. We used a stringent cutoff of $\Delta BIC < -10$ to accept the ILS+ introgression model, as suggested by the authors.

Next, we performed an ABBA-BABA test using *Dsuite* (v0.5) (Malinsky et al., 2021) with a four-taxon statement ((H1, H2)H3) H4). Using H4 as the outgroup (herein *A. arabicum*), H1 and H2 were treated as sister species, and H3 was tested for potential gene flow or introgression with H1 or H2. The number of sites with ABBA and BABA allele patterns were tallied. *D* was calculated as (*n*ABBA − *n*BABA)/(*n*ABBA + *n*BABA), in which *n*ABBA and *n*BABA represent the total number of sites with ABBA and BABA patterns, respectively. A negative *D* indicates gene flow between H1 and H3, a positive *D* indicates gene flow between H2 and H3, and $D = 0$ indicates no gene flow. The *f4*-admixture ratio and f_b statistic were calculated for all trios using Fbranch in the *Dsuite* pipeline. The f_b statistic is a heuristic strategy to summarize *f4*-admixture ratios over the entire tree topology, including internal branches, to detect introgression events and excess allele sharing across a dataset (Malinsky et al., 2021). For these analyses, all genome sequences were first aligned to the reference *Arabidopsis* genome using *BWA* (v0.7.10-r789) and then sorted and converted with *SAMtools* (v0.1.19) (Danecek et al., 2021) to generate summary statistics. Single-nucleotide polymorphisms were identified with *GATK* (v3.7) (McKenna et al., 2010). The resulting VCF file was used as input for *Dsuite* and further analyzed using *A. arabicum* as an outgroup. Finally, a "Reticulation Index" was generated to measure the frequency of the asymmetrical triplets in all combinations at each node to quantify the intensity of introgression using 28 species datasets via a recently reported pipeline (Cai et al., 2021).

However, ILS could not be excluded as a contributor to the observed gene tree discordances. We tested the population mutation parameter theta of each internal branch using mutation units inferred by *IQ-TREE* and coalescent units inferred by *ASTRAL*. Theta reflects the level of ILS, with a high theta value indicating a large ancestral population size and hence a high level of ILS (Cai et al., 2021). Next, 10 000 gene trees were simulated under the ILS hypothesis with the MSC model using *Phybase* (v1.4) in *R* (v4.2.1) (Liu and Yu, 2010). The Robinson-Foulds distance distribution (Bogdanowicz et al., 2012) between the species tree and the simulated gene trees, and those between the species tree and the empirical gene trees, were compared. Finally, correlations between branch lengths and ICA values were calculated to determine the contribution of ILS to tree conflicts. In general, the shorter the branch length, the more ILS and conflicts among gene trees (Zhou et al., 2022).

Reconstruction of the CBK

To reconstruct the CBK, we selected nine diploid species without additional WGDs. These species represent four major supertribes (Camelinodae, Brassicodae, Hesperodae, and Arabodae) and the basal Aethionemeae: *A. thaliana*, *C. rubella*, *E. cheiranthoides*, *T. arvense*, *A. alpina*, *D. nivalis*, *M. linifolius*, *T. quadricornis*, and *A. arabicum*. Given its complex evolutionary history and potential hybrid origin, Heliophilodae was excluded from the reconstruction of the CBK (Hendriks et al., 2023). The CBK was reconstructed as follows. Initially, syntenic gene pairs were identified between each pair from the nine species using *MCscan* (*Python* version) in "*–full*" mode to extract one-to-one quota syntenic blocks (Tang et al., 2008). Subsequently, these syntenic gene pairs were categorized into syntenic groups with the synteny-based gene-family clustering pipeline *SynPan*, using *merge.pl* for clustering and *transform_stat.pl* for summarization (Wu et al., 2023). Considering the syntenic groups and the phylogenetic tree (Figure 3) of the species under investigation, we deduced the strictly conserved syntenic genes (putative protogenes, or pPGs) at each node of the phylogenetic tree. This facilitated the identification of an MRCA for the Brassicaceae pool, which comprises 9702 core pPGs conserved across all investigated species. Third, by collinearity analysis between 9 extant species and the ACK with *MCscan*, we revisited 22 conserved syntenic GBs of the ACK (Schranz et al., 2006; Lysak et al., 2016) in these species. Moreover, 43 additional breakpoints were identified by the addition of genomic data (supplemental Figure 39). Then, using 9702 pPGs, we refined the boundaries of these syntenic blocks to obtain 65 conserved GBs. We built the karyotypes of extant species based on these 65 GBs. Finally, with the topology and 65 conserved GBs, we reconstructed the CBK and traced the evolutionary scenario of karyotypes in Brassicaceae using *MLGO*, which is based on a *PMAG* method (Hu et al., 2014; Feng et al., 2017). In addition, centromere positions were predicted using *quartet* (Lin et al., 2023) and previous reports (Willing et al., 2015; Lysak et al., 2016; Mandakova et al., 2020; Naish et al., 2021; Nowak et al., 2021; Yang et al., 2021).

RNA extraction and gene expression analyses

For tissue-specific expression assays, leaves, shoots, flowers, and siliques were collected in early May 2022 from plants growing wild in Fukang City, Xinjiang Autonomous Region, China. For temperature-response analyses, seeds of *M. linifolius* and *T. quadricornis* were collected from plants grown in Fukang City, Xinjiang, China, and *A. thaliana* seeds (Col-0) were collected from plants grown under long-day conditions (16-h light and 8-h dark, 22°C). After disinfection with 75% ethanol, the seeds were sown on half-strength Murashige and Skoog medium and stratified at 4°C for 3 days. Ten days after germination under long-day conditions, control plants were maintained under the same growth conditions, whereas experimental plants were subjected to either 6 or 24 h of low temperature (4°C) in a refrigerator or 3 h of high temperature (38°C) in a water bath. After treatment, a portion of the plants were immediately sampled, and the other portion was placed back under the same conditions as the control group for an additional 6 or 24 h before sampling. Each assay was carried out using 3 to 5 biological replicates, with approximately 10 plants per replicate.

Total RNA isolation and quantitative reverse-transcription PCR (qRT–PCR) were performed as described previously (Zhong et al., 2021; Dong et al., 2023). In brief, total RNA was isolated from shoots and leaves with TRI Reagent TR118 (MRC,

USA) and from flowers and siliques with an OmniPlant RNA Kit (CWBio, China). A total of 1 μg of RNA was treated with gDNA Purge (Novoprotein, China) to remove contaminating genomic DNA and reverse transcribed using NovoScript Plus All-in-one 1st Strand cDNA Synthesis SuperMix (Novoprotein, China). Gene expression analyses were carried out with gene-specific primers (supplemental Table 27) at an amplification efficiency of between 90% and 110% using NovoStart SYBR qPCR SuperMix Plus on a QuantStudio Real-Time PCR apparatus. *ACT2* was used as a reference. Primer information can be found in supplemental Table 31.

DATA AND CODE AVAILABILITY

All data supporting the results of this study are included in the manuscript and its supplemental files. Genome assemblies, annotations, and sequence data were deposited in the CNSA database (https://db.cngb.org) under project number CNP0003993.

SUPPLEMENTAL INFORMATION

Supplemental information is available at *Plant Communications Online*.

FUNDING

This work was supported by the Priority Research Program of the Chinese Academy of Sciences (CAS) (Grant No. XDA0440000 and XDB31000000).

AUTHOR CONTRIBUTIONS

J.-Y.H. and D.-Z.L. conceptualized and coordinated the project. J.L. and B.-Y.Z. collected the samples and performed the main experiments. J.L., S.-Z.Z., Y.-L.L., and B.-Y.Z. analyzed and visualized the data with the help of all other authors. J.-Y.H. wrote the manuscript with input from J.L. and D.-Z.L. All authors read and approved the final manuscript.

ACKNOWLEDGMENTS

We thank Guli Mayem, Haiyang Ma, Kaiqing Xie, Hoja Tore, and Musen Lin from Xinjiang Agricultural University for their help with field work. We are grateful to Jun He, Chun-Fang Li, and Shu-Lan Chen for their help in nursing and propagating the plants. We appreciate useful discussions with Lian-Ming Gao, Jun-Bo Yang, Peng-Fei Ma, Yan-Xia Jia, Rong Zhang, Shui-Yin Liu, and Sheng-Yuan Qin. We greatly appreciate the four anonymous reviewers for their very insightful comments and suggestions on improving the manuscript. This work was partially facilitated by GBOWS and the iFlora HPC Center of GBOWS. No conflict of interest declared.

Received: March 20, 2023
Revised: March 3, 2024
Accepted: March 11, 2024
Published: March 11, 2024

REFERENCES

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.

Alonso-Blanco, C., Aarts, M.G.M., Bentsink, L., Keurentjes, J.J.B., Reymond, M., Vreugdenhil, D., and Koornneef, M. (2009). What has natural variation taught us about plant development, physiology, and adaptation? Plant Cell **21**:1877–1896. https://doi.org/10.1105/tpc.109.068114.

Andrés, F., and Coupland, G. (2012). The genetic basis of flowering responses to seasonal cues. Nat. Rev. Genet. **13**:627–639. https://doi.org/10.1038/nrg3291.

Anselmetti, Y., Luhmann, N., Bérard, S., Tannier, E., and Chauve, C. (2018). Comparative Methods for Reconstructing Ancient Genome Organization. Methods Mol. Biol. **1704**:343–362. https://doi.org/10.1007/978-1-4939-7463-4_13.

Arabidopsis Genome Initiative. (2000). Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. Nature **408**:796–815. https://doi.org/10.1038/35048692.

Avonce, N., Mendoza-Vargas, A., Morett, E., and Iturriaga, G. (2006). Insights on the evolution of trehalose biosynthesis. BMC Evol. Biol. **6**:109. https://doi.org/10.1186/1471-2148-6-109.

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.

Bayat, S., Lysak, M.A., and Mandáková, T. (2021). Genome structure and evolution in the cruciferous tribe Thlaspideae (Brassicaceae). Plant J. **108**:1768–1785. https://doi.org/10.1111/tpj.15542.

Beier, S., Thiel, T., Münch, T., Scholz, U., and Mascher, M. (2017). MISA-web: a web server for microsatellite prediction. Bioinformatics **33**:2583–2585. https://doi.org/10.1093/bioinformatics/btx198.

Beilstein, M.A., Al-Shehbaz, I.A., and Kellogg, E.A. (2006). Brassicaceae phylogeny and trichome evolution. Am. J. Bot. **93**:607–619. https://doi.org/10.3732/ajb.93.4.607.

Beilstein, M.A., Al-Shehbaz, I.A., Mathews, S., and Kellogg, E.A. (2008). Brassicaceae phylogeny inferred from phytochrome A and *ndhF* sequence data: tribes and trichomes revisited. Am. J. Bot. **95**:1307–1327. https://doi.org/10.3732/ajb.0800065.

Belser, C., Istace, B., Denis, E., Dubarry, M., Baurens, F.C., Falentin, C., Genete, M., Berrabah, W., Chèvre, A.M., Delourme, R., et al. (2018). Chromosome-scale assemblies of plant genomes using nanopore long reads and optical maps. Nat. Plants **4**:879–887. https://doi.org/10.1038/s41477-018-0289-4.

Benson, G. (1999). Tandem repeats finder: a program to analyze DNA sequences. Nucleic Acids Res. **27**:573–580. https://doi.org/10.1093/nar/27.2.573.

Birney, E., Clamp, M., and Durbin, R. (2004). GeneWise and Genomewise. Genome Res. **14**:988–995. https://doi.org/10.1101/gr.1865504.

Bogdanowicz, D., Giaro, K., and Wróbel, B. (2012). TreeCmp: Comparison of Trees in Polynomial Time. Evol Bioinform **8**:475–487. https://doi.org/10.4137/Ebo.S9657.

Bouckaert, R.R. (2010). DensiTree: making sense of sets of phylogenetic trees. Bioinformatics **26**:1372–1373. https://doi.org/10.1093/bioinformatics/btq110.

Brock, J.R., Mandáková, T., McKain, M., Lysak, M.A., and Olsen, K.M. (2022). Chloroplast phylogenomics in *Camelina* (Brassicaceae) reveals multiple origins of polyploid species and the maternal lineage of C. sativa. Hortic. Res. **9**, uhab050. https://doi.org/10.1093/hr/uhab050.

Burton, J.N., Adey, A., Patwardhan, R.P., Qiu, R., Kitzman, J.O., and Shendure, J. (2013). Chromosome-scale scaffolding of de novo genome assemblies based on chromatin interactions. Nat. Biotechnol. **31**:1119–1125. https://doi.org/10.1038/nbt.2727.

Cai, L., Xi, Z., Lemmon, E.M., Lemmon, A.R., Mast, A., Buddenhagen, C.E., Liu, L., and Davis, C.C. (2021). The Perfect Storm: Gene Tree Estimation Error, Incomplete Lineage Sorting, and Ancient Gene Flow Explain the Most Recalcitrant Ancient Angiosperm Clade. Syst. Biol. **70**:491–507. https://doi.org/10.1093/sysbio/syaa083.

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.

Cheng, F., Wu, J., and Wang, X. (2014). Genome triplication drove the diversification of *Brassica* plants. Hortic. Res. **1**, 14024. https://doi.org/10.1038/hortres.2014.24.

Cheng, F., Mandáková, T., Wu, J., Xie, Q., Lysak, M.A., and Wang, X. (2013). Deciphering the diploid ancestral genome of the Mesohexaploid *Brassica rapa*. Plant Cell **25**:1541–1554. https://doi.org/10.1105/tpc.113.110486.

Cheng, H., Concepcion, G.T., Feng, X., Zhang, H., and Li, H. (2021). Haplotype-resolved de novo assembly using phased assembly graphs with hifiasm. Nat. Methods **18**:170–175. https://doi.org/10.1038/s41592-020-01056-5.

Conesa, A., and Götz, S. (2008). Blast2GO: A comprehensive suite for functional analysis in plant genomics. Int. J. Plant Genomics **2008**, 619832. https://doi.org/10.1155/2008/619832.

Cosentino, S., and Iwasaki, W. (2019). SonicParanoid: fast, accurate and easy orthology inference. Bioinformatics **35**:149–151. https://doi.org/10.1093/bioinformatics/bty631.

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.

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.

De Bie, T., Cristianini, N., Demuth, J.P., and Hahn, M.W. (2006). CAFE: a computational tool for the study of gene family evolution. Bioinformatics **22**:1269–1271. https://doi.org/10.1093/bioinformatics/btl097.

Ding, W.N., Ree, R.H., Spicer, R.A., and Xing, Y.W. (2020). Ancient orogenic and monsoon-driven assembly of the world's richest temperate alpine flora. Science **369**:578–581. https://doi.org/10.1126/science.abb4484.

Dong, X., Zhang, L.P., Tang, Y.H., Yu, D., Cheng, F., Dong, Y.X., Jiang, X.D., Qian, F.M., Guo, Z.H., and Hu, J.Y. (2023). Arabidopsis AGAMOUS-LIKE16 and SUPPRESSOR OF CONSTANS1 regulate the genome-wide expression and flowering time. Plant Physiol. **192**:154–169. https://doi.org/10.1093/plphys/kiad058.

Duda, M., Gasińska, A., and Gregoraszczuk, E.L. (1999). Flow cytometric cell cycle analysis of two subpopulations of porcine granulosa cells. Exp. Clin. Endocrinol. Diabetes **107**:203–207. https://doi.org/10.1055/s-0029-1212099.

Edelman, N.B., Frandsen, P.B., Miyagi, M., Clavijo, B., Davey, J., Dikow, R.B., García-Accinelli, G., Van Belleghem, S.M., Patterson, N., Neafsey, D.E., et al. (2019). Genomic architecture and introgression shape a butterfly radiation. Science **366**:594–599. https://doi.org/10.1126/science.aaw2090.

Edger, P.P., Hall, J.C., Harkess, A., Tang, M., Coombs, J., Mohammadin, S., Schranz, M.E., Xiong, Z., Leebens-Mack, J., Meyers, B.C., et al. (2018). Brassicales phylogeny inferred from 72 plastid genes: A reanalysis of the phylogenetic localization of two paleopolyploid events and origin of novel chemical defenses. Am. J. Bot. **105**:463–469. https://doi.org/10.1002/ajb2.1040.

Ellinghaus, D., Kurtz, S., and Willhoeft, U. (2008). LTRharvest, an efficient and flexible software for de novo detection of LTR retrotransposons. BMC Bioinf. **9**:18. https://doi.org/10.1186/1471-2105-9-18.

Emms, D.M., and Kelly, S. (2018). STAG: Species Tree Inference from All Genes. Preprint at bioRxiv. https://doi.org/10.1101/267914.

Emms, D.M., and Kelly, S. (2019). OrthoFinder: phylogenetic orthology inference for comparative genomics. Genome Biol. **20**:238. https://doi.org/10.1186/s13059-019-1832-y.

Feng, B., Zhou, L., and Tang, J. (2017). Ancestral Genome Reconstruction on Whole Genome Level. Curr. Genomics **18**:306–315. https://doi.org/10.2174/1389202918666170307120943.

Fernandez-Mazuecos, M., Vargas, P., McCauley, R.A., Monjas, D., Otero, A., Chaves, J.A., Guevara Andino, J.E., and Rivas-Torres, G. (2020). The Radiation of Darwin's Giant Daisies in the Galapagos Islands. Curr. Biol. **30**:4989–4998.e4987. https://doi.org/10.1016/j.cub.2020.09.019.

Fernandez, O., Béthencourt, L., Quero, A., Sangwan, R.S., and Clément, C. (2010). Trehalose and plant stress responses: friend or foe? Trends Plant Sci. **15**:409–417. https://doi.org/10.1016/j.tplants.2010.04.004.

Fichtner, F., and Lunn, J.E. (2021). The Role of Trehalose 6-Phosphate (Tre6P) in Plant Metabolism and Development. Annu. Rev. Plant Biol. **72**:737–760. https://doi.org/10.1146/annurev-arplant-050718-095929.

Fichtner, F., Dissanayake, I.M., Lacombe, B., and Barbier, F. (2021). Sugar and Nitrate Sensing: A Multi-Billion-Year Story. Trends Plant Sci. **26**:352–374. https://doi.org/10.1016/j.tplants.2020.11.006.

Fichtner, F., Olas, J.J., Feil, R., Watanabe, M., Krause, U., Hoefgen, R., Stitt, M., and Lunn, J.E. (2020). Functional Features of TREHALOSE-6-PHOSPHATE SYNTHASE1, an Essential Enzyme in Arabidopsis. Plant Cell **32**:1949–1972. https://doi.org/10.1105/tpc.19.00837.

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.

Gao, S., Yang, X., Sun, J., Zhao, X., Wang, B., and Ye, K. (2022). IAGS: Inferring Ancestor Genome Structure under a Wide Range of Evolutionary Scenarios. Mol. Biol. Evol. **39**, msac041. https://doi.org/10.1093/molbev/msac041.

Garcia-Porta, J., Irisarri, I., Kirchner, M., Rodríguez, A., Kirchhof, S., Brown, J.L., MacLeod, A., Turner, A.P., Ahmadzadeh, F., Albaladejo, G., et al. (2019). Environmental temperatures shape thermal physiology as well as diversification and genome-wide substitution rates in lizards. Nat. Commun. **10**:4077. https://doi.org/10.1038/s41467-019-11943-x.

Gaudinier, A., and Blackman, B.K. (2020). Evolutionary processes from the perspective of flowering time diversity. New Phytol. **225**:1883–1898. https://doi.org/10.1111/nph.16205.

Geiser, C., Mandáková, T., Arrigo, N., Lysak, M.A., and Parisod, C. (2016). Repeated Whole-Genome Duplication, Karyotype Reshuffling, and Biased Retention of Stress-Responding Genes in Buckler Mustard. Plant Cell **28**:17–27. https://doi.org/10.1105/tpc.15.00791.

Geng, Y., Guan, Y., Qiong, L., Lu, S., An, M., Crabbe, M.J.C., Qi, J., Zhao, F., Qiao, Q., and Zhang, T. (2021). Genomic analysis of field pennycress (*Thlaspi arvense*) provides insights into mechanisms of adaptation to high elevation. BMC Biol. **19**:143. https://doi.org/10.1186/s12915-021-01079-0.

German, D.A., Friesen, N., Neuffer, B., Al-Shehbaz, I.A., and Hurka, H. (2009). Contribution to ITS phylogeny of the Brassicaceae, with special reference to some Asian taxa. Plant Syst. Evol. **283**:33–56. https://doi.org/10.1007/s00606-009-0213-5.

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.

Grabherr, M.G., Haas, B.J., Yassour, M., Levin, J.Z., Thompson, D.A., Amit, I., Adiconis, X., Fan, L., Raychowdhury, R., Zeng, Q., et al. (2011). Full-length transcriptome assembly from RNA-Seq data

without a reference genome. Nat. Biotechnol. **29**:644–652. https://doi.org/10.1038/nbt.1883.

Guo, C., Luo, Y., Gao, L.M., Yi, T.S., Li, H.T., Yang, J.B., and Li, D.Z. (2023). Phylogenomics and the flowering plant tree of life. J. Integr. Plant Biol. **65**:299–323. https://doi.org/10.1111/jipb.13415.

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.

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

Haas, B.J., Salzberg, S.L., Zhu, W., Pertea, M., Allen, J.E., Orvis, J., White, O., Buell, C.R., and Wortman, J.R. (2008). Automated eukaryotic gene structure annotation using EVidenceModeler and the Program to Assemble Spliced Alignments. Genome Biol. **9**:R7. https://doi.org/10.1186/gb-2008-9-1-r7.

Haas, B.J., Delcher, A.L., Mount, S.M., Wortman, J.R., Smith, R.K., Hannick, L.I., Maiti, R., Ronning, C.M., Rusch, D.B., Town, C.D., et al. (2003). Improving the genome annotation using maximal transcript alignment assemblies. Nucleic Acids Res. **31**:5654–5666. https://doi.org/10.1093/nar/gkg770.

Hendriks, K.P., Kiefer, C., Al-Shehbaz, I.A., Bailey, C.D., Hooft van Huysduynen, A., Nikolov, L.A., Nauheimer, L., Zuntini, A.R., German, D.A., Franzke, A., et al. (2023). Global Brassicaceae phylogeny based on filtering of 1,000-gene dataset. Curr. Biol. **33**:4052–4068.e6. https://doi.org/10.1016/j.cub.2023.08.026.

Hloušková, P., Mandáková, T., Pouch, M., Trávníček, P., and Lysak, M.A. (2019). The large genome size variation in the Hesperis clade was shaped by the prevalent proliferation of DNA repeats and rarer genome downsizing. Ann. Bot. **124**:103–120. https://doi.org/10.1093/aob/mcz036.

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.

Hu, F., Lin, Y., and Tang, J. (2014). MLGO: phylogeny reconstruction and ancestral inference from gene-order data. BMC Bioinform **15**:354. https://doi.org/10.1186/s12859-014-0354-6.

Hu, H., Sun, P., Yang, Y., Ma, J., and Liu, J. (2023). Genome-scale angiosperm phylogenies based on nuclear, plastome, and mitochondrial datasets. J. Integr. Plant Biol. **65**:1479–1489. https://doi.org/10.1111/jipb.13455.

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.

Hu, T.T., Pattyn, P., Bakker, E.G., et al. (2011). The *Arabidopsis lyrata* genome sequence and the basis of rapid genome size change. Nat. Genet. **43**:476–481. https://doi.org/10.1038/ng.807.

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.

Huang, L., Ma, Y., Jiang, J., Li, T., Yang, W., Zhang, L., Wu, L., Feng, L., Xi, Z., Xu, X., et al. (2020a). A chromosome-scale reference genome of *Lobularia maritima*, an ornamental plant with high stress tolerance. Hortic. Res. **7**:197. https://doi.org/10.1038/s41438-020-00422-w.

Huang, X.C., German, D.A., and Koch, M.A. (2020b). 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.

Hyun, Y., Richter, R., and Coupland, G. (2017). Competence to Flower: Age-Controlled Sensitivity to Environmental Cues. Plant Physiol. **173**:36–46. https://doi.org/10.1104/pp.16.01523.

Iordachescu, M., and Imai, R. (2008). Trehalose biosynthesis in response to abiotic stresses. J. Integr. Plant Biol. **50**:1223–1229. https://doi.org/10.1111/j.1744-7909.2008.00736.x.

Jiao, W.B., Accinelli, G.G., Hartwig, B., Kiefer, C., Baker, D., Severing, E., Willing, E.M., Piednoel, M., Woetzel, S., Madrid-Herrero, E., et al. (2017). Improving and correcting the contiguity of long-read genome assemblies of three plant species using optical mapping and chromosome conformation capture data. Genome Res. **27**:778–786. https://doi.org/10.1101/gr.213652.116.

Jin, J.J., Yu, W.B., Yang, J.B., Song, Y., dePamphilis, C.W., Yi, T.S., and Li, D.Z. (2020). GetOrganelle: a fast and versatile toolkit for accurate de novo assembly of organelle genomes. Genome Biol. **21**:241. https://doi.org/10.1186/s13059-020-02154-5.

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.

Kalvari, I., Nawrocki, E.P., Argasinska, J., Quinones-Olvera, N., Finn, R.D., Bateman, A., and Petrov, A.I. (2018). Non-Coding RNA Analysis Using the Rfam Database. Curr. Protoc. Bioinformatics **62**:e51. https://doi.org/10.1002/cpbi.51.

Kang, M., Wu, H., Yang, Q., et al. (2020). A chromosome-scale genome assembly of *Isatis indigotica*, an important medicinal plant used in traditional Chinese medicine: An Isatis genome. Hortic. Res. **7**:18. https://doi.org/10.1038/s41438-020-0240-5.

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. https://doi.org/10.1093/molbev/mst010.

Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., Buxton, S., Cooper, A., Markowitz, S., Duran, C., et al. (2012). Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics **28**:1647–1649. https://doi.org/10.1093/bioinformatics/bts199.

Keilwagen, J., Hartung, F., Paulini, M., Twardziok, S.O., and Grau, J. (2018). Combining RNA-seq data and homology-based gene prediction for plants, animals and fungi. BMC Bioinform **19**:189. https://doi.org/10.1186/s12859-018-2203-5.

Kiefer, C., Willing, E.M., Jiao, W.B., Sun, H., Piednoël, M., Hümann, U., Hartwig, B., Koch, M.A., and Schneeberger, K. (2019). Interspecies association mapping links reduced CG to TG substitution rates to the loss of gene-body methylation. Nat. Plants **5**:846–855. https://doi.org/10.1038/s41477-019-0486-9.

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.

Kim, D., Langmead, B., and Salzberg, S.L. (2015). HISAT: a fast spliced aligner with low memory requirements. Nat. Methods **12**:357–360. https://doi.org/10.1038/nmeth.3317.

Korf, I. (2004). Gene finding in novel genomes. BMC Bioinf. **5**:59. https://doi.org/10.1186/1471-2105-5-59.

Kozomara, A., Birgaoanu, M., and Griffiths-Jones, S. (2019). miRBase: from microRNA sequences to function. Nucleic Acids Res. **47**:D155–D162. https://doi.org/10.1093/nar/gky1141.

Kumar, S., Stecher, G., Suleski, M., and Hedges, S.B. (2017). TimeTree: A Resource for Timelines, Timetrees, and Divergence Times. Mol. Biol. Evol. **34**:1812–1819. https://doi.org/10.1093/molbev/msx116.

Leyman, B., Van Dijck, P., and Thevelein, J.M. (2001). An unexpected plethora of trehalose biosynthesis genes in *Arabidopsis thaliana*. Trends Plant Sci. **6**:510–513. https://doi.org/10.1016/s1360-1385(01)02125-2.

Li, H. (2013). Aligning Sequence Reads, Clone Sequences and Assembly Contigs with BWA-MEM. Preprint at arXiv. https://doi.org/10.48550/arXiv.1303.3997.

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.

Lin, Y., Ye, C., Li, X., Chen, Q., Wu, Y., Zhang, F., Pan, R., Zhang, S., Chen, S., Wang, X., et al. (2023). quarTeT: a telomere-to-telomere toolkit for gap-free genome assembly and centromeric repeat identification. Hortic. Res. **10**:uhad127. https://doi.org/10.1093/hr/uhad127.

Liu, J., Hu, J.Y., and Li, D.Z. (2024). Remarkable mitochondrial genome heterogeneity in *Meniocus linifolius* (Brassicaceae). Plant Cell Rep. **43**:36. https://doi.org/10.1007/s00299-023-03102-w.

Liu, L., and Yu, L. (2010). Phybase: an R package for species tree analysis. Bioinformatics **26**:962–963. https://doi.org/10.1093/bioinformatics/btq062.

Liu, L.M., Du, X.Y., Guo, C., and Li, D.Z. (2020). Resolving robust phylogenetic relationships of core Brassicaceae using genome skimming data. J. Syst. Evol. **59**:442–453. https://doi.org/10.1111/jse.12666.

Lowe, T.M., and Eddy, S.R. (1997). tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. Nucleic Acids Res. **25**:955–964. https://doi.org/10.1093/nar/25.5.955.

Lunn, J.E. (2007). Gene families and evolution of trehalose metabolism in plants. Funct. Plant Biol. **34**:550–563. https://doi.org/10.1071/FP06315.

Lysak, M.A., Mandáková, T., and Schranz, M.E. (2016). Comparative paleogenomics of crucifers: ancestral genomic blocks revisited. Curr. Opin. Plant Biol. **30**:108–115. https://doi.org/10.1016/j.pbi.2016.02.001.

Ma, Y., Mao, X., Wang, J., et al. (2022). Pervasive hybridization during evolutionary radiation of *Rhododendron* subgenus *Hymenanthes* in mountains of southwest China. Natl. Sci. Rev. **9**:nwac276. https://doi.org/10.1093/nsr/nwac276.

Malinsky, M., Matschiner, M., and Svardal, H. (2021). Dsuite - Fast D-statistics and related admixture evidence from VCF files. Mol. Ecol. Resour. **21**:584–595. https://doi.org/10.1111/1755-0998.13265.

Malinsky, M., Svardal, H., Tyers, A.M., Miska, E.A., Genner, M.J., Turner, G.F., and Durbin, R. (2018). Whole-genome sequences of Malawi cichlids reveal multiple radiations interconnected by gene flow. Nat. Ecol. Evol. **2**:1940–1955. https://doi.org/10.1038/s41559-018-0717-x.

Mandáková, T., and Lysak, M.A. (2008). Chromosomal phylogeny and karyotype evolution in x=7 crucifer species (Brassicaceae). Plant Cell **20**:2559–2570. https://doi.org/10.1105/tpc.108.062166.

Mandáková, T., Hloušková, P., German, D.A., and Lysak, M.A. (2017). Monophyletic Origin and Evolution of the Largest Crucifer Genomes. Plant Physiol. **174**:2062–2071. https://doi.org/10.1104/pp.17.00457.

Mandáková, T., Hloušková, P., Koch, M.A., and Lysak, M.A. (2020). Genome Evolution in Arabideae Was Marked by Frequent Centromere Repositioning. Plant Cell **32**:650–665. https://doi.org/10.1105/tpc.19.00557.

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.

Marçais, G., and Kingsford, C. (2011). A fast, lock-free approach for efficient parallel counting of occurrences of k-mers. Bioinformatics **27**:764–770. https://doi.org/10.1093/bioinformatics/btr011.

McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimella, K., Altshuler, D., Gabriel, S., Daly, M., and DePristo, M.A. (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.

Miao, Y., Fang, X., Sun, J., Xiao, W., Yang, Y., Wang, X., Farnsworth, A., Huang, K., Ren, Y., Wu, F., et al. (2022). A new biologic paleoaltimetry indicating Late Miocene rapid uplift of northern Tibet Plateau. Science **378**:1074–1079. https://doi.org/10.1126/science.abo2475.

Murat, F., Armero, A., Pont, C., Klopp, C., and Salse, J. (2017). Reconstructing the genome of the most recent common ancestor of flowering plants. Nat. Genet. **49**:490–496. https://doi.org/10.1038/ng.3813.

Murat, F., Louis, A., Maumus, F., Armero, A., Cooke, R., Quesneville, H., Roest Crollius, H., and Salse, J. (2015). Understanding Brassicaceae evolution through ancestral genome reconstruction. Genome Biol. **16**:262. https://doi.org/10.1186/s13059-015-0814-y.

Naciri, Y., and Linder, H.P. (2020). The genetics of evolutionary radiations. Biol. Rev. Camb. Philos. Soc. **95**:1055–1072. https://doi.org/10.1111/brv.12598.

Naish, M., Alonge, M., Wlodzimierz, P., Tock, A.J., Abramson, B.W., Schmücker, A., Mandáková, T., Jamge, B., Lambing, C., Kuo, P., et al. (2021). The genetic and epigenetic landscape of the *Arabidopsis* centromeres. Science **374**, eabi7489. https://doi.org/10.1126/science.abi7489.

Nawrocki, E.P., and Eddy, S.R. (2013). Infernal 1.1: 100-fold faster RNA homology searches. Bioinformatics **29**:2933–2935. https://doi.org/10.1093/bioinformatics/btt509.

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.

Nguyen, T.P., Mühlich, C., Mohammadin, S., van den Bergh, E., Platts, A.E., Haas, F.B., Rensing, S.A., and Schranz, M.E. (2019). Genome Improvement and genetic map construction for *Aethionema arabicum*, the first divergent branch in the Brassicaceae family. G3 (Bethesda) **9**:3521–3530. https://doi.org/10.1534/g3.119.400657.

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.

Nowak, M.D., Birkeland, S., Mandáková, T., Roy Choudhury, R., Guo, X., Gustafsson, A.L.S., Gizaw, A., Schrøder-Nielsen, A., Fracassetti, M., Brysting, A.K., et al. (2021). The genome of *Draba nivalis* shows signatures of adaptation to the extreme environmental stresses of the Arctic. Mol. Ecol. Resour. **21**:661–676. https://doi.org/10.1111/1755-0998.13280.

Ou, S., and Jiang, N. (2018). LTR_retriever: A Highly Accurate and Sensitive Program for Identification of Long Terminal Repeat

Retrotransposons. Plant Physiol. **176**:1410–1422. https://doi.org/10.1104/pp.17.01310.

Ou, S., Chen, J., and Jiang, N. (2018). Assessing genome assembly quality using the LTR Assembly Index (LAI). Nucleic Acids Res. **46**:e126. https://doi.org/10.1093/nar/gky730.

Ou, S., Su, W., Liao, Y., Chougule, K., Agda, J.R.A., Hellinga, A.J., Lugo, C.S.B., Elliott, T.A., Ware, D., Peterson, T., et al. (2019). Benchmarking transposable element annotation methods for creation of a streamlined, comprehensive pipeline. Genome Biol. **20**:275. https://doi.org/10.1186/s13059-019-1905-y.

Pertea, M., Pertea, G.M., Antonescu, C.M., Chang, T.C., Mendell, J.T., and Salzberg, S.L. (2015). StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. Nat. Biotechnol. **33**:290–295. https://doi.org/10.1038/nbt.3122.

Perumal, S., Koh, C.S., Jin, L., Buchwaldt, M., Higgins, E.E., Zheng, C., Sankoff, D., Robinson, S.J., Kagale, S., Navabi, Z.K., et al. (2020). A high-contiguity Brassica nigra genome localizes active centromeres and defines the ancestral Brassica genome. Nat. Plants **6**:929–941. https://doi.org/10.1038/s41477-020-0735-y.

Price, M.N., Dehal, P.S., and Arkin, A.P. (2009). FastTree: computing large minimum evolution trees with profiles instead of a distance matrix. Mol. Biol. Evol. **26**:1641–1650. https://doi.org/10.1093/molbev/msp077.

Qiao, X., Li, Q., Yin, H., Qi, K., Li, L., Wang, R., Zhang, S., and Paterson, A.H. (2019). Gene duplication and evolution in recurring polyploidization-diploidization cycles in plants. Genome Biol. **20**:38. https://doi.org/10.1186/s13059-019-1650-2.

Qu, X.J., Moore, M.J., Li, D.Z., and Yi, T.S. (2019). PGA: a software package for rapid, accurate, and flexible batch annotation of plastomes. Plant Methods **15**:50. https://doi.org/10.1186/s13007-019-0435-7.

Salichos, L., Stamatakis, A., and Rokas, A. (2014). Novel information theory-based measures for quantifying incongruence among phylogenetic trees. Mol. Biol. Evol. **31**:1261–1271. https://doi.org/10.1093/molbev/msu061.

Schranz, M.E., Lysak, M.A., and Mitchell-Olds, T. (2006). The ABC's of comparative genomics in the Brassicaceae: building blocks of crucifer genomes. Trends Plant Sci. **11**:535–542. https://doi.org/10.1016/j.tplants.2006.09.002.

Schranz, M.E., Mohammadin, S., and Edger, P.P. (2012). Ancient whole genome duplications, novelty and diversification: the WGD Radiation Lag-Time Model. Curr. Opin. Plant Biol. **15**:147–153. https://doi.org/10.1016/j.pbi.2012.03.011.

Servant, N., Varoquaux, N., Lajoie, B.R., Viara, E., Chen, C.J., Vert, J.P., Heard, E., Dekker, J., and Barillot, E. (2015). HiC-Pro: an optimized and flexible pipeline for Hi-C data processing. Genome Biol. **16**:259. https://doi.org/10.1186/s13059-015-0831-x.

She, R., Chu, J.S.C., Wang, K., Pei, J., and Chen, N. (2009). GenBlastA: enabling BLAST to identify homologous gene sequences. Genome Res. **19**:143–149. https://doi.org/10.1101/gr.082081.108.

Simão, F.A., Waterhouse, R.M., Ioannidis, P., Kriventseva, E.V., and Zdobnov, E.M. (2015). BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. Bioinformatics **31**:3210–3212. https://doi.org/10.1093/bioinformatics/btv351.

Slotte, T., Hazzouri, K.M., Ågren, J.A., Koenig, D., Maumus, F., Guo, Y.L., Steige, K., Platts, A.E., Escobar, J.S., Newman, L.K., et al. (2013). The *Capsella rubella* genome and the genomic consequences of rapid mating system evolution. Nat. Genet. **45**:831–835. https://doi.org/10.1038/ng.2669.

Slovák, M., Melichárková, A., Štubňová, E.G., Kučera, J., Mandáková, T., Smyčka, J., Lavergne, S., Passalacqua, N.G., Vďačný, P., and Paun, O. (2023). Pervasive Introgression During Rapid Diversification of the European Mountain Genus *Soldanella* (L.) (Primulaceae). Syst. Biol. **72**:491–504. https://doi.org/10.1093/sysbio/syac071.

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.

Smith, S.A., Moore, M.J., Brown, J.W., and Yang, Y. (2015). Analysis of phylogenomic datasets reveals conflict, concordance, and gene duplications with examples from animals and plants. BMC Evol. Biol. **15**:150. https://doi.org/10.1186/s12862-015-0423-0.

Soltis, P.S., Marchant, D.B., Van de Peer, Y., and Soltis, D.E. (2015). Polyploidy and genome evolution in plants. Curr. Opin. Genet. Dev. **35**:119–125. https://doi.org/10.1016/j.gde.2015.11.003.

Song, J.M., Guan, Z., Hu, J., et al. (2020). Eight high-quality genomes reveal pan-genome architecture and ecotype differentiation of *Brassica napus*. Nat. Plants **6**:34–45. https://doi.org/10.1038/s41477-019-0577-7.

Španiel, S., Kempa, M., Salmerón-Sánchez, E., Fuertes-Aguilar, J., Mota, J.F., Al-Shehbaz, I.A., German, D.A., Olšavská, K., Šingliarová, B., Zozomová-Lihová, J., and Marhold, K. (2015). AlyBase: database of names, chromosome numbers, and ploidy levels of Alysseae (Brassicaceae), with a new generic concept of the tribe. Plant Syst. Evol. **301**:2463–2491. https://doi.org/10.1007/s00606-015-1257-3.

Stamatakis, A. (2014). RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics **30**:1312–1313. https://doi.org/10.1093/bioinformatics/btu033.

Stanke, M., and Waack, S. (2003). Gene prediction with a hidden Markov model and a new intron submodel. Bioinformatics **19** (*Suppl 2*), ii215-225. https://doi.org/10.1093/bioinformatics/btg1080.

Sun, P., Jiao, B., Yang, Y., Shan, L., Li, T., Li, X., Xi, Z., Wang, X., and Liu, J. (2022). WGDI: A user-friendly toolkit for evolutionary analyses of whole-genome duplications and ancestral karyotypes. Mol. Plant **15**:1841–1851. https://doi.org/10.1016/j.molp.2022.10.018.

Suyama, M., Torrents, D., and Bork, P. (2006). PAL2NAL: robust conversion of protein sequence alignments into the corresponding codon alignments. Nucleic Acids Res. **34**:W609–W612. https://doi.org/10.1093/nar/gkl315.

Tang, H., Bowers, J.E., Wang, X., Ming, R., Alam, M., and Paterson, A.H. (2008). Synteny and collinearity in plant genomes. Science **320**:486–488. https://doi.org/10.1126/science.1153917.

Tang, S., Lomsadze, A., and Borodovsky, M. (2015). Identification of protein coding regions in RNA transcripts. Nucleic Acids Res. **43**:e78. https://doi.org/10.1093/nar/gkv227.

Tarailo-Graovac, M., and Chen, N. (2009). Using RepeatMasker to identify repetitive elements in genomic sequences. Curr. Protoc. Bioinformatics **4**:4.10.1–4.10.14. https://doi.org/10.1002/0471250953.bi0410s25.

Than, C., Ruths, D., and Nakhleh, L. (2008). PhyloNet: a software package for analyzing and reconstructing reticulate evolutionary relationships. BMC Bioinf. **9**:322. https://doi.org/10.1186/1471-2105-9-322.

Van Dijck, P., Mascorro-Gallardo, J.O., De Bus, M., Royackers, K., Iturriaga, G., and Thevelein, J.M. (2002). Truncation of *Arabidopsis thaliana* and *Selaginella lepidophylla* trehalose-6-phosphate synthase unlocks high catalytic activity and supports high trehalose levels on expression in yeast. Biochem. J. **366**:63–71. https://doi.org/10.1042/BJ20020517.

Vandesteene, L., Ramon, M., Le Roy, K., Van Dijck, P., and Rolland, F. (2010). A single active trehalose-6-P synthase (TPS) and a family of putative regulatory TPS-like proteins in *Arabidopsis*. Mol. Plant **3**:406–419. https://doi.org/10.1093/mp/ssp114.

Vurture, G.W., Sedlazeck, F.J., Nattestad, M., Underwood, C.J., Fang, H., Gurtowski, J., and Schatz, M.C. (2017). GenomeScope: fast reference-free genome profiling from short reads. Bioinformatics **33**:2202–2204. https://doi.org/10.1093/bioinformatics/btx153.

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.

Walden, N., Nguyen, T.P., Mandáková, T., Lysak, M.A., and Schranz, M.E. (2020a). Genomic Blocks in *Aethionema arabicum* Support Arabideae as Next Diverging Clade in Brassicaceae. Front. Plant Sci. **11**:719. https://doi.org/10.3389/fpls.2020.00719.

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. (2020b). 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.

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.

Willing, E.M., Rawat, V., Mandáková, T., Maumus, F., James, G.V., Nordström, K.J.V., Becker, C., Warthmann, N., Chica, C., Szarzynska, B., et al. (2015). Genome expansion of *Arabis alpina* linked with retrotransposition and reduced symmetric DNA methylation. Nat. Plants **1**, 14023. https://doi.org/10.1038/nplants.2014.23.

Wu, D., Xie, L., Sun, Y., Huang, Y., Jia, L., Dong, C., Shen, E., Ye, C.Y., Qian, Q., and Fan, L. (2023). A syntelog-based pan-genome provides insights into rice domestication and de-domestication. Genome Biol. **24**:179. https://doi.org/10.1186/s13059-023-03017-5.

Wu, T., Hu, E., Xu, S., Chen, M., Guo, P., Dai, Z., Feng, T., Zhou, L., Tang, W., Zhan, L., et al. (2021). clusterProfiler 4.0: A universal enrichment tool for interpreting omics data. Innovation **2**, 100141. https://doi.org/10.1016/j.xinn.2021.100141.

Xia, X.M., Yang, M.Q., Li, C.L., Huang, S.X., Jin, W.T., Shen, T.T., Wang, F., Li, X.H., Yoichi, W., Zhang, L.H., et al. (2022). Spatiotemporal Evolution of the Global Species Diversity of Rhododendron. Mol. Biol. Evol. **39**, msab314. https://doi.org/10.1093/molbev/msab314.

Xu, Z., and Wang, H. (2007). LTR_FINDER: an efficient tool for the prediction of full-length LTR retrotransposons. Nucleic Acids Res. **35**:W265–W268. https://doi.org/10.1093/nar/gkm286.

Yang, W., Zhang, L., Mandáková, T., Huang, L., Li, T., Jiang, J., Yang, Y., Lysak, M.A., Liu, J., and Hu, Q. (2021). The chromosome-level genome sequence and karyotypic evolution of *Megadenia pygmaea* (Brassicaceae). Mol. Ecol. Resour. **21**:871–879. https://doi.org/10.1111/1755-0998.13291.

Yang, Z. (2007). PAML 4: phylogenetic analysis by maximum likelihood. Mol. Biol. Evol. **24**:1586–1591. https://doi.org/10.1093/molbev/msm088.

Ye, J., McGinnis, S., and Madden, T.L. (2006). BLAST: improvements for better sequence analysis. Nucleic Acids Res. **34**:W6–W9. https://doi.org/10.1093/nar/gkl164.

Yue, J., VanBuren, R., Liu, J., Fang, J., Zhang, X., Liao, Z., Wai, C.M., Xu, X., Chen, S., Zhang, S., et al. (2022). SunUp and Sunset genomes revealed impact of particle bombardment mediated transformation and domestication history in papaya. Nat. Genet. **54**:715–724. https://doi.org/10.1038/s41588-022-01068-1.

Zachos, J.C., Shackleton, N.J., Revenaugh, J.S., Pälike, H., and Flower, B.P. (2001). Climate Response to Orbital Forcing Across the Oligocene-Miocene Boundary. Science **292**:274–278. https://doi.org/10.1126/science.1058288.

Zhang, C., Rabiee, M., Sayyari, E., and Mirarab, S. (2018). ASTRAL-III: polynomial time species tree reconstruction from partially resolved gene trees. BMC Bioinf. **19**:153. https://doi.org/10.1186/s12859-018-2129-y.

Zhang, D., Gao, F., Jakovlić, I., Zou, H., Zhang, J., Li, W.X., and Wang, G.T. (2020). PhyloSuite: An integrated and scalable desktop platform for streamlined molecular sequence data management and evolutionary phylogenetics studies. Mol. Ecol. Resour. **20**:348–355. https://doi.org/10.1111/1755-0998.13096.

Zhang, R.G., Li, G.Y., Wang, X.L., Dainat, J., Wang, Z.X., Ou, S., and Ma, Y. (2022). TEsorter: an accurate and fast method to classify LTR-retrotransposons in plant genomes. Hortic. Res. **9**, uhac017. https://doi.org/10.1093/hr/uhac017.

Zhong, M.C., Jiang, X.D., Yang, G.Q., Cui, W.H., Suo, Z.Q., Wang, W.J., Sun, Y.B., Wang, D., Cheng, X.C., Li, X.M., et al. (2021). Rose without prickle: genomic insights linked to moisture adaptation. Natl. Sci. Rev. **8**:nwab092. https://doi.org/10.1093/nsr/nwab092.

Zhou, B.F., Yuan, S., Crowl, A.A., Liang, Y.Y., Shi, Y., Chen, X.Y., An, Q.Q., Kang, M., Manos, P.S., and Wang, B. (2022). Phylogenomic analyses highlight innovation and introgression in the continental radiations of Fagaceae across the Northern Hemisphere. Nat. Commun. **13**:1320. https://doi.org/10.1038/s41467-022-28917-1.

Züst, T., Strickler, S.R., Powell, A.F., et al. (2020). Independent evolution of ancestral and novel defenses in a genus of toxic plants (*Erysimum*, Brassicaceae). Elife **9**, e51712. https://doi.org/10.7554/eLife.51712.

Plant Communications, Volume *5*

Supplemental information

Genomes of *Meniocus linifolius* and *Tetracme quadricornis* reveal the ancestral karyotype and genomic features of core Brassicaceae

Jie Liu, Shi-Zhao Zhou, Yun-Long Liu, Bin-Yan Zhao, Dongmei Yu, Mi-Cai Zhong, Xiao-Dong Jiang, Wei-Hua Cui, Jiu-Xia Zhao, Juan Qiu, Liang-Min Liu, Zhen-Hua Guo, Hong-Tao Li, Dun-Yan Tan, Jin-Yong Hu, and De-Zhu Li

Genomes of *Meniocus linifolius* and *Tetracme quadricornis* unveil the ancestral karyotype and genomic features of core Brassicaceae

Jie Liu [1,4,#], Shi-Zhao Zhou [1,4,#], Yun-Long Liu [2,#], Bin-Yan Zhao [1,4,#], Dongmei Yu [1], Mi-Cai Zhong [1], Xiao-Dong Jiang [1], Wei-Hua Cui [1], Jiu-Xia Zhao [1,4], Juan Qiu [3], Liang-Min Liu [2,4], Zhen-Hua Guo [2], Hong-Tao Li [2], Dun-Yan Tan [3], Jin-Yong Hu [1,*], De-Zhu Li [2,*]

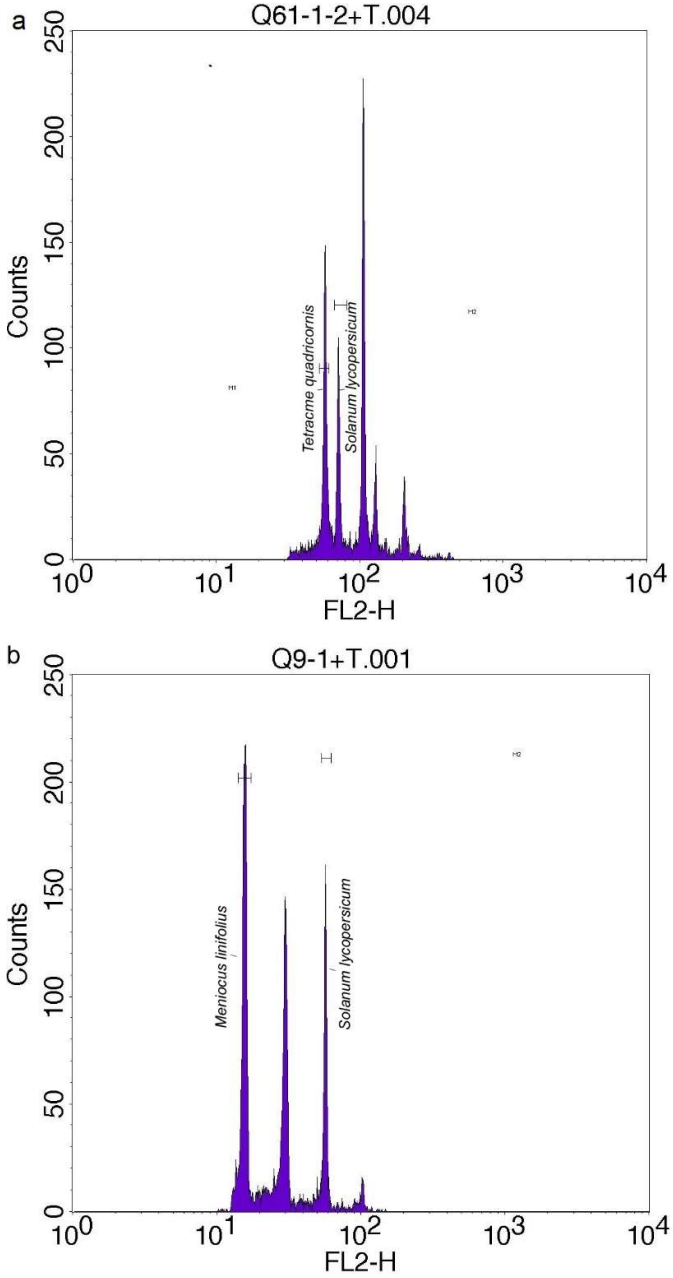
1. CAS Key Laboratory for Plant Diversity and Biogeography of East Asia & Yunnan Key Laboratory of Crop Wild Relatives Omics, Kunming Institute of Botany, Chinese Academy of Sciences. Kunming 650201, China

2. Germplasm Bank of Wild Species & Yunnan Key Laboratory for Crop Wild Relatives Omics, Kunming Institute of Botany, Chinese Academy of Sciences, Kunming, 650201, China

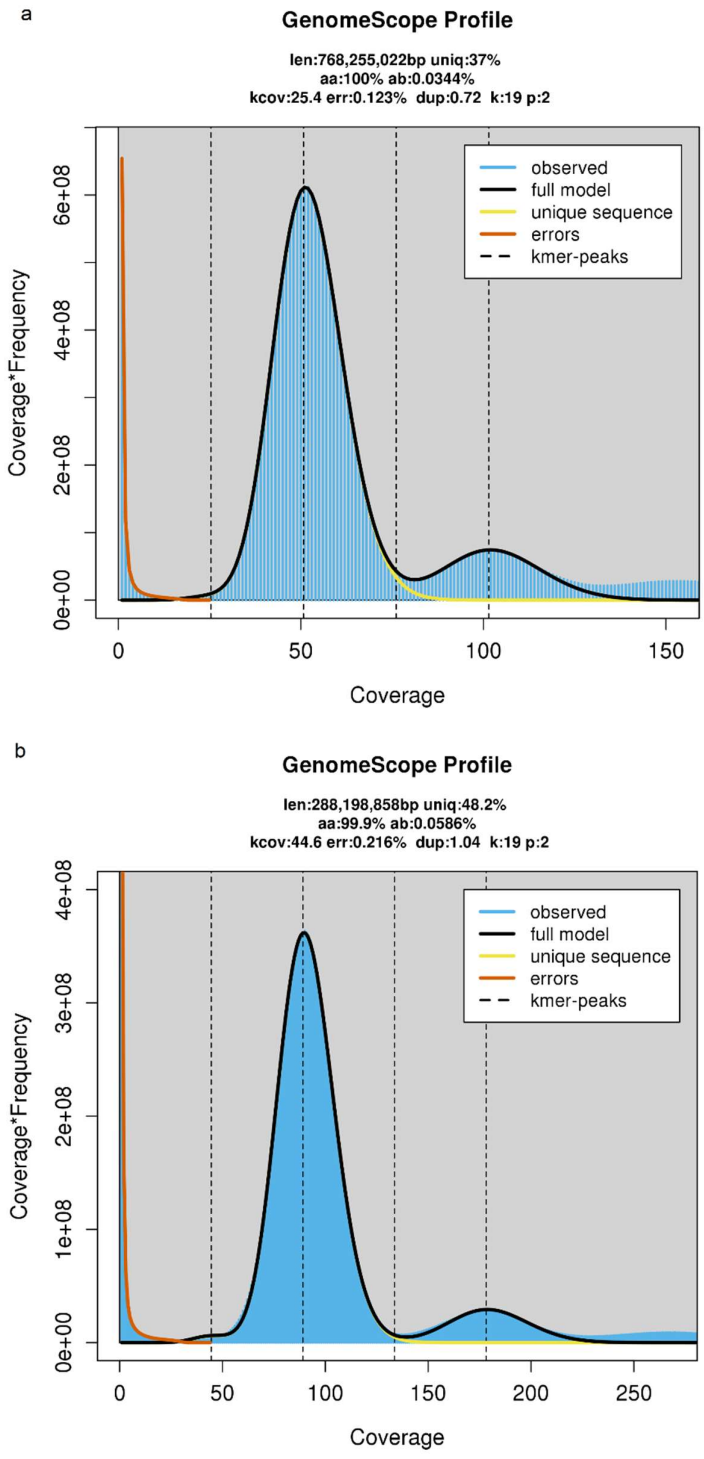
3. College of Life Sciences, Xinjiang Agricultural University, Ürümqi 830052, China

4. University of Chinese Academy of Sciences, Beijing 100049, China

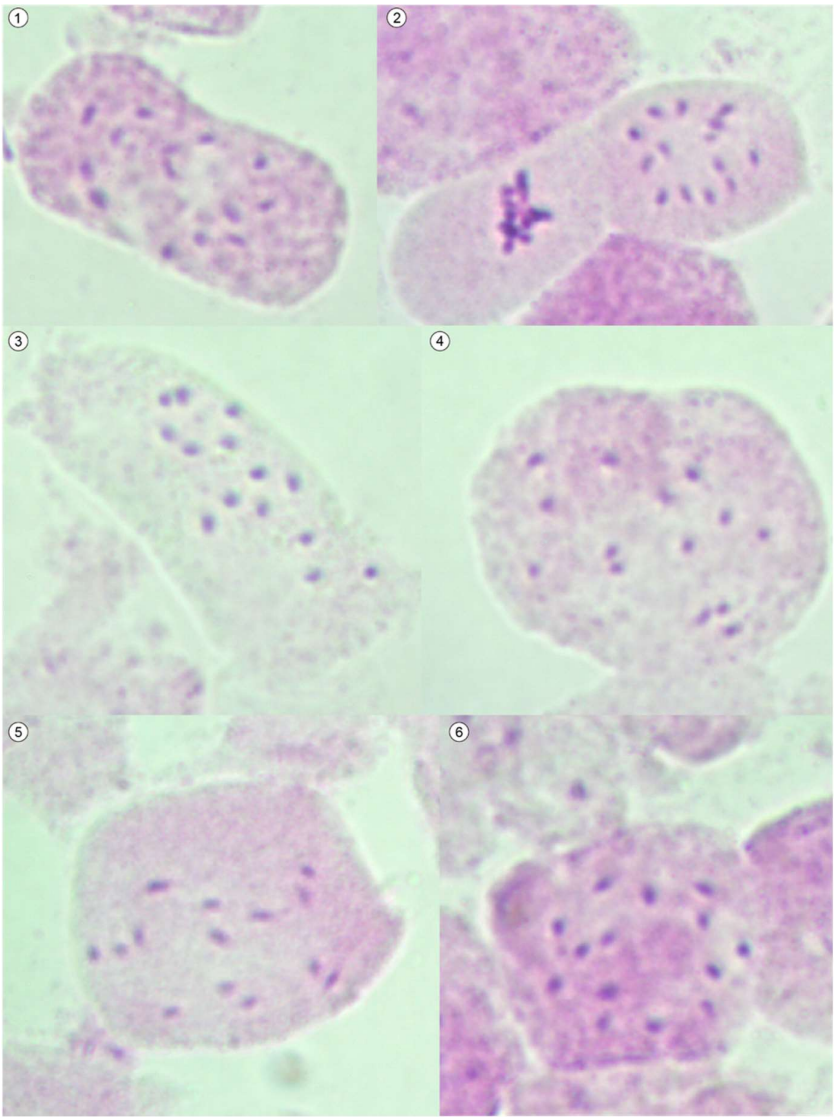
Supplemental Figures 1-54.



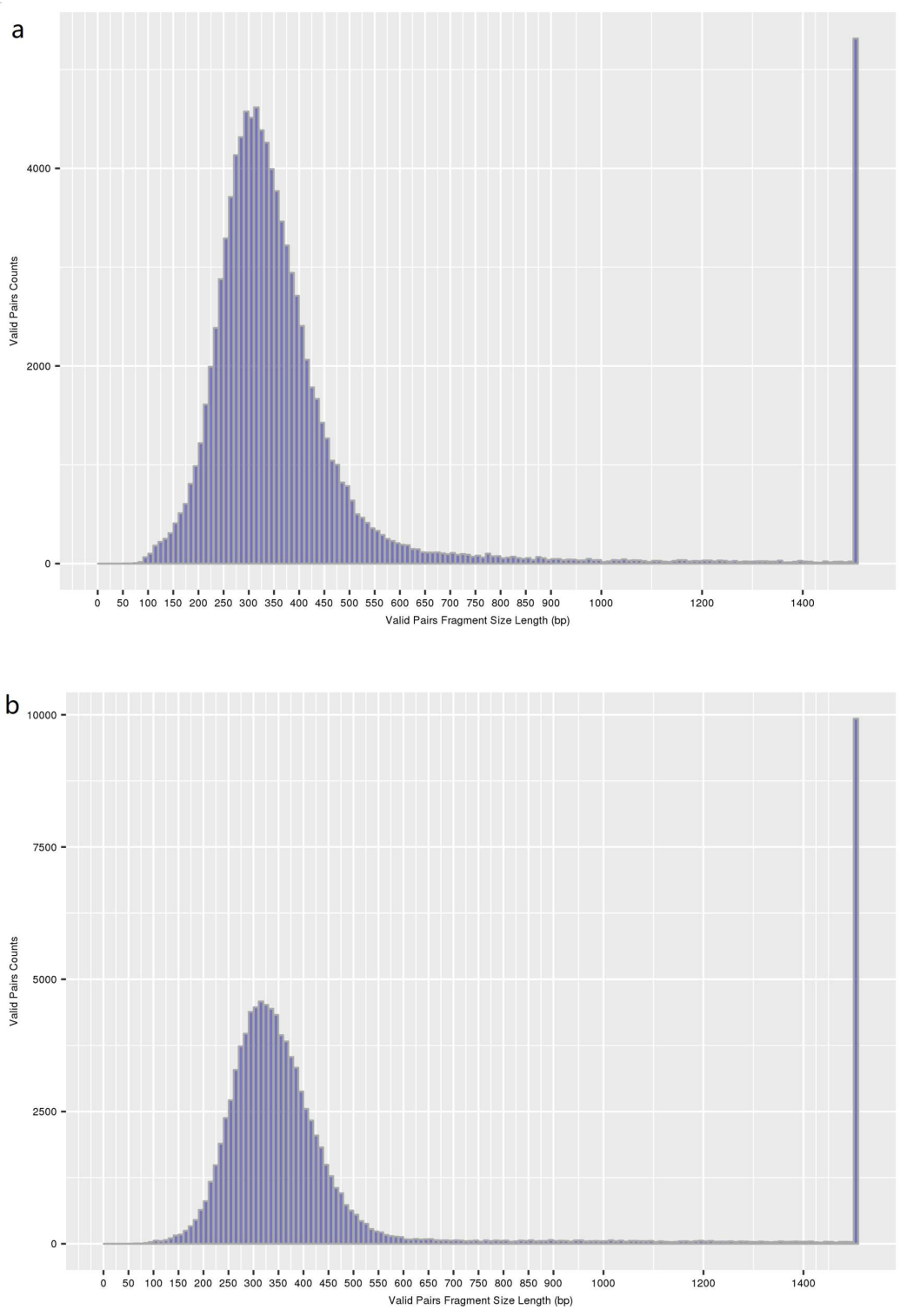
Supplemental Figure 1. Genome sizes of the *Tetracme quadricornis* (a; Q61-1-2; *Tqu*) and *Meniocus linifolius* (b; Q9-1; *Mli*) estimated by flow cytometry with *Solanum lycopersicum* as reference.



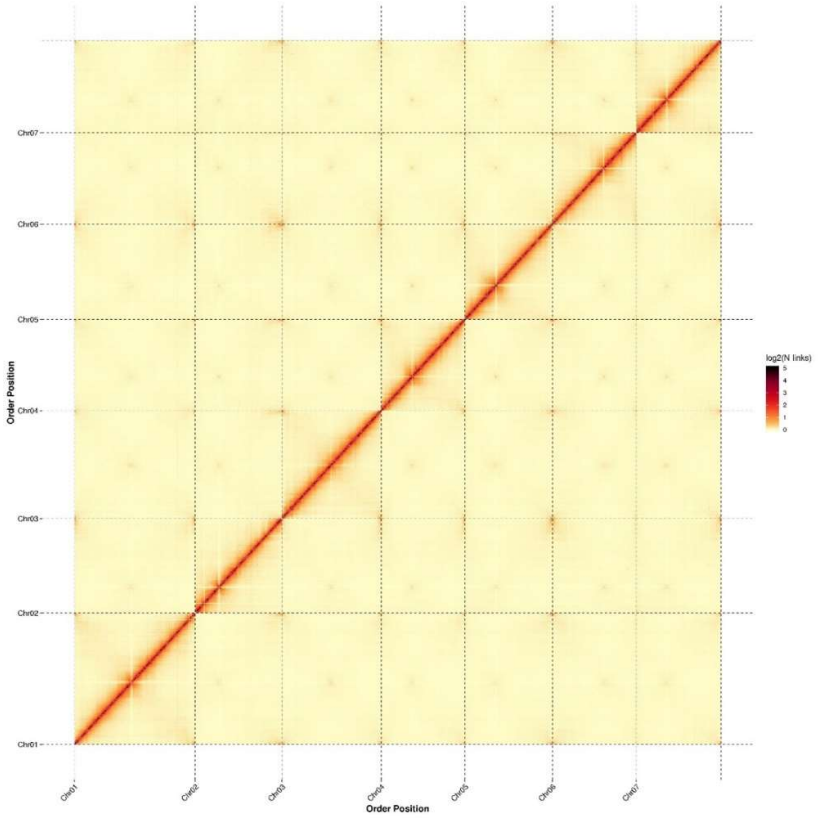
Supplemental Figure 2. Genome size estimation using *GenomeScope2*. 19 K-mer size was used to survey both *Tqu* (a) and *Mli* (b). X-axis shows the depth of read coverage while Y-axis is the frequency at certain depth divided by the total frequency of all depths. Estimated genome size, heterozygosity and repeat content are shown above each graph.



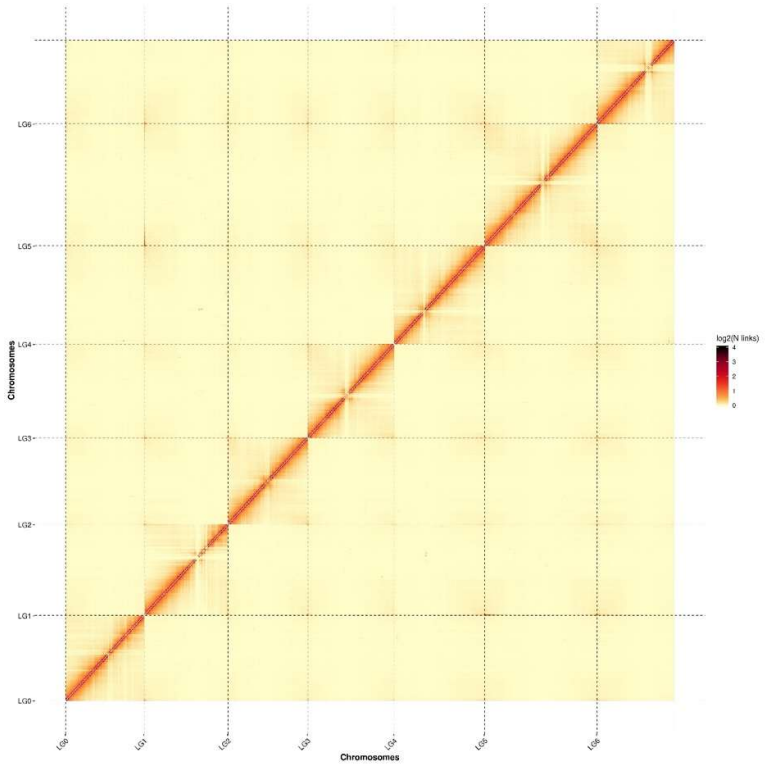
Supplemental Figure 3. Cytological analyses of *M. linifolius* karyotypes.



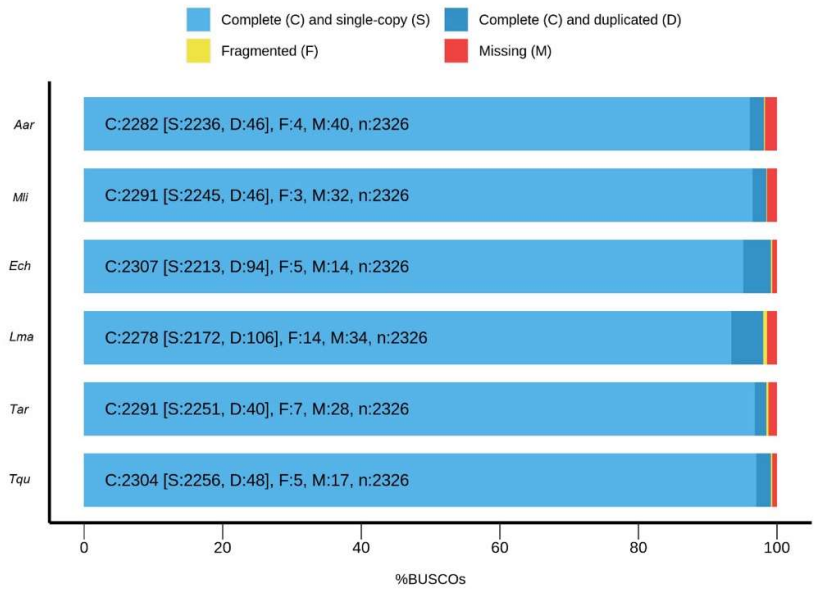
Supplemental Figure 4. Distribution insert Hi-C reads for *Tqu* (a) and *Mli* (b). X-axis is the sum of the distance between the paired end reads on the assembled genome, while Y-axis shows the number of 100000 pair reads randomly selected in insert fragments of different lengths.



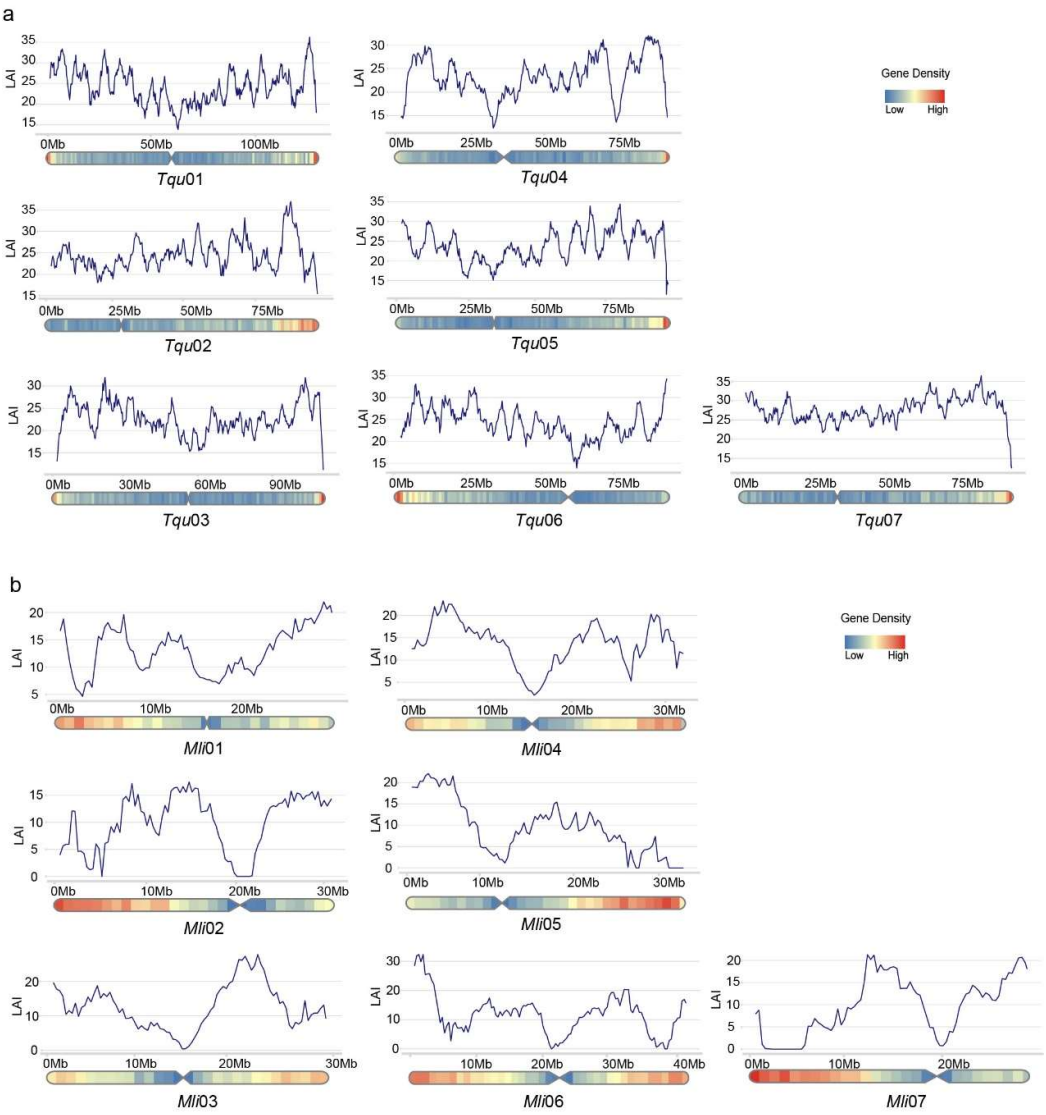
Supplemental Figure 5. Distribution of interaction frequency for Hi-C links among chromosomes of *Tqu* at 100 kb resolution. The X- and Y- axes present the order of the positions of scaffolds on the corresponding pseudochromosomes (the same for Supplemental Figure 6).



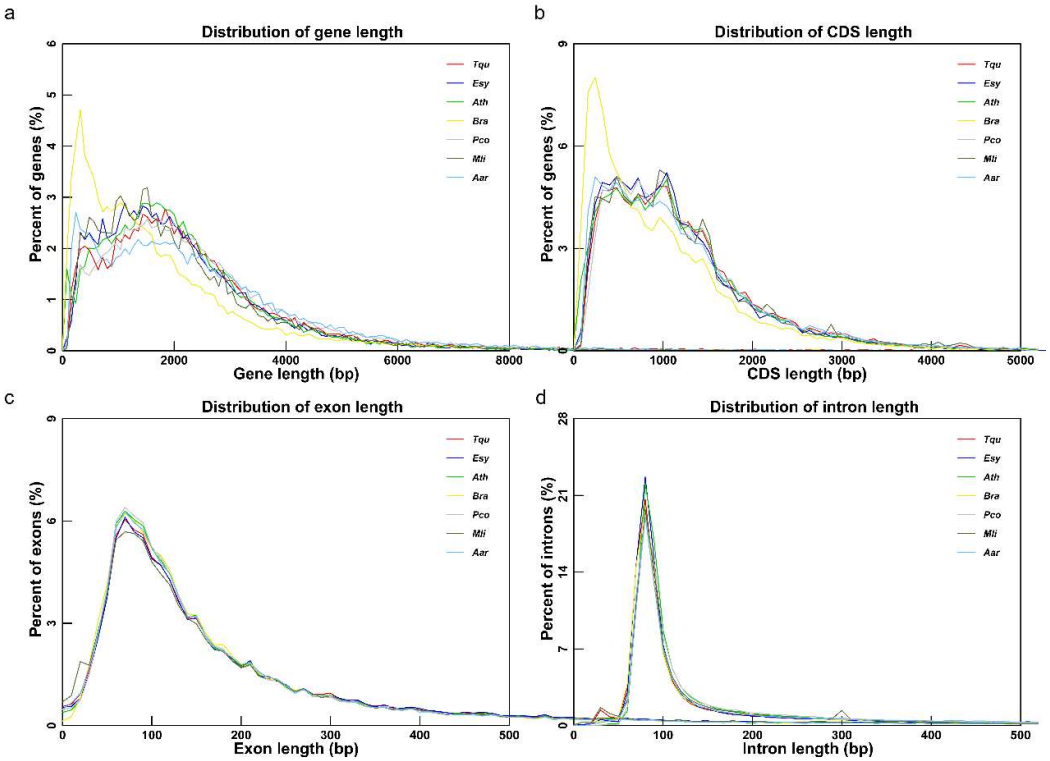
Supplemental Figure 6. Distribution of interaction frequency for Hi-C links among chromosomes of *Mli* at 100 kb resolution.



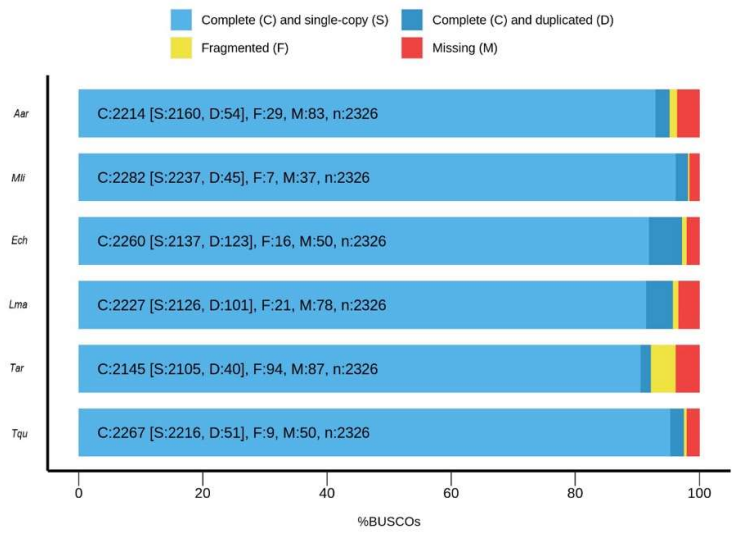
Supplemental Figure 7. BUSCO assessments (Eudicots_odb10) at the genome level for six representative species. *Aar*, *Aethionema arabicum*; *Ech*, *Erysimum cheiranthoides*; *Lma, Lobularia maritima*; *Tar*, *Thlaspi arvense*.



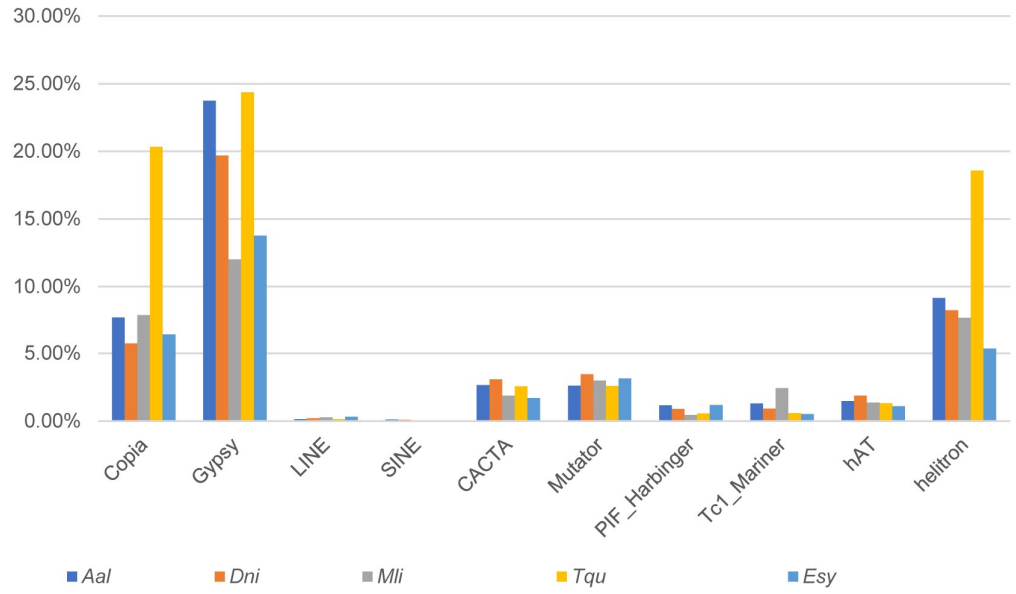
Supplemental Figure 8. Distribution of the LTR Assembly Index (LAI) along the chromosomes for *Tqu* (a) and *Mli* (b).



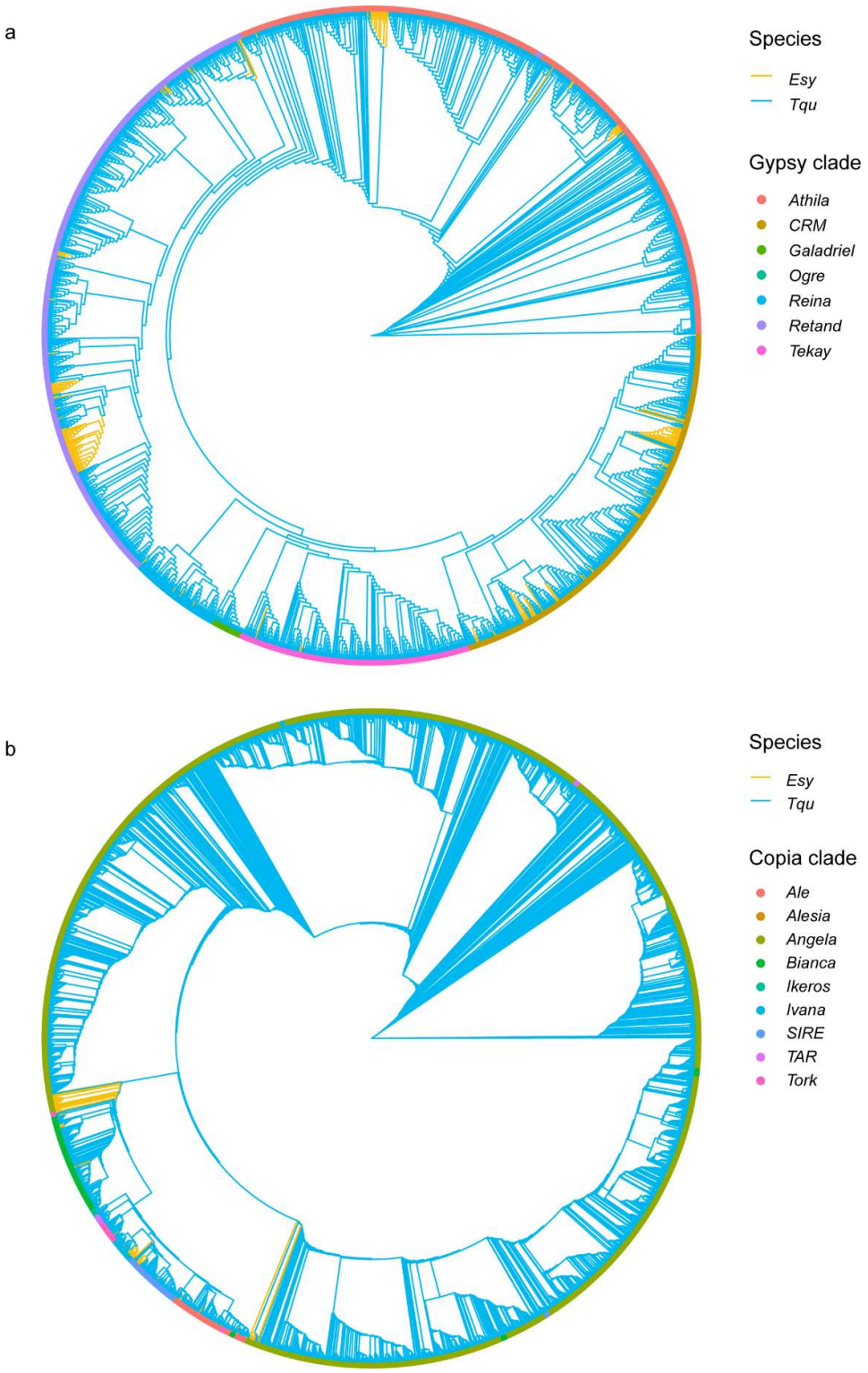
Supplemental Figure 9. Gene structure features for selected Brassicaceae species. **a**, gene length; **b**, coding length; **c**, exon length; **d**, intron length.



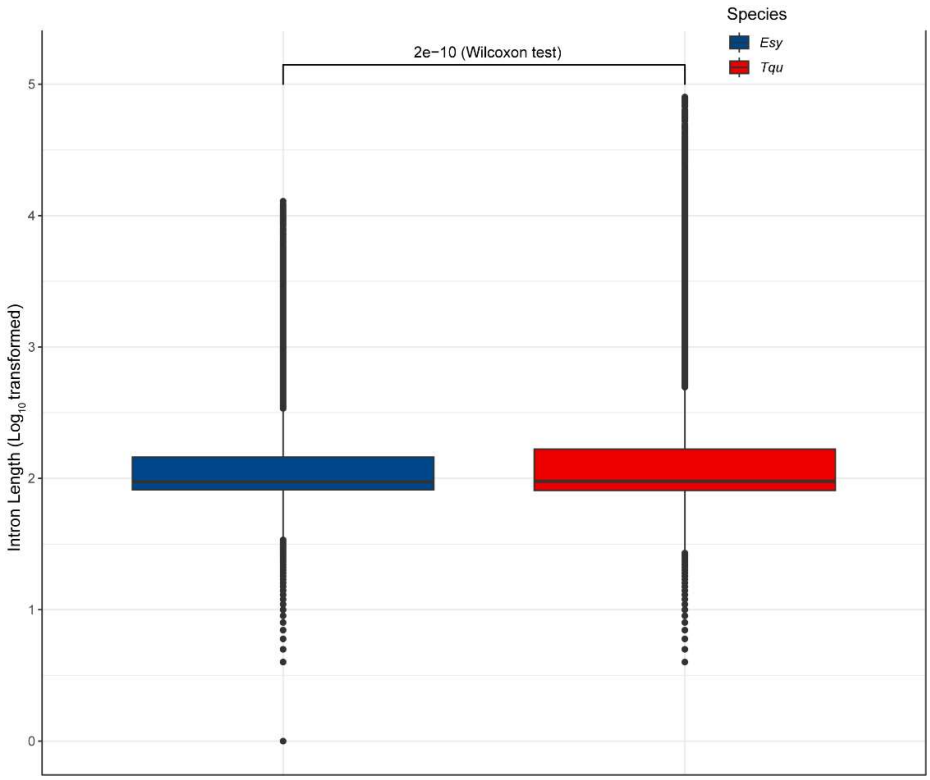
Supplemental Figure 10. BUSCO assessment (Eudicots_odb10) at the gene level for six representative species.



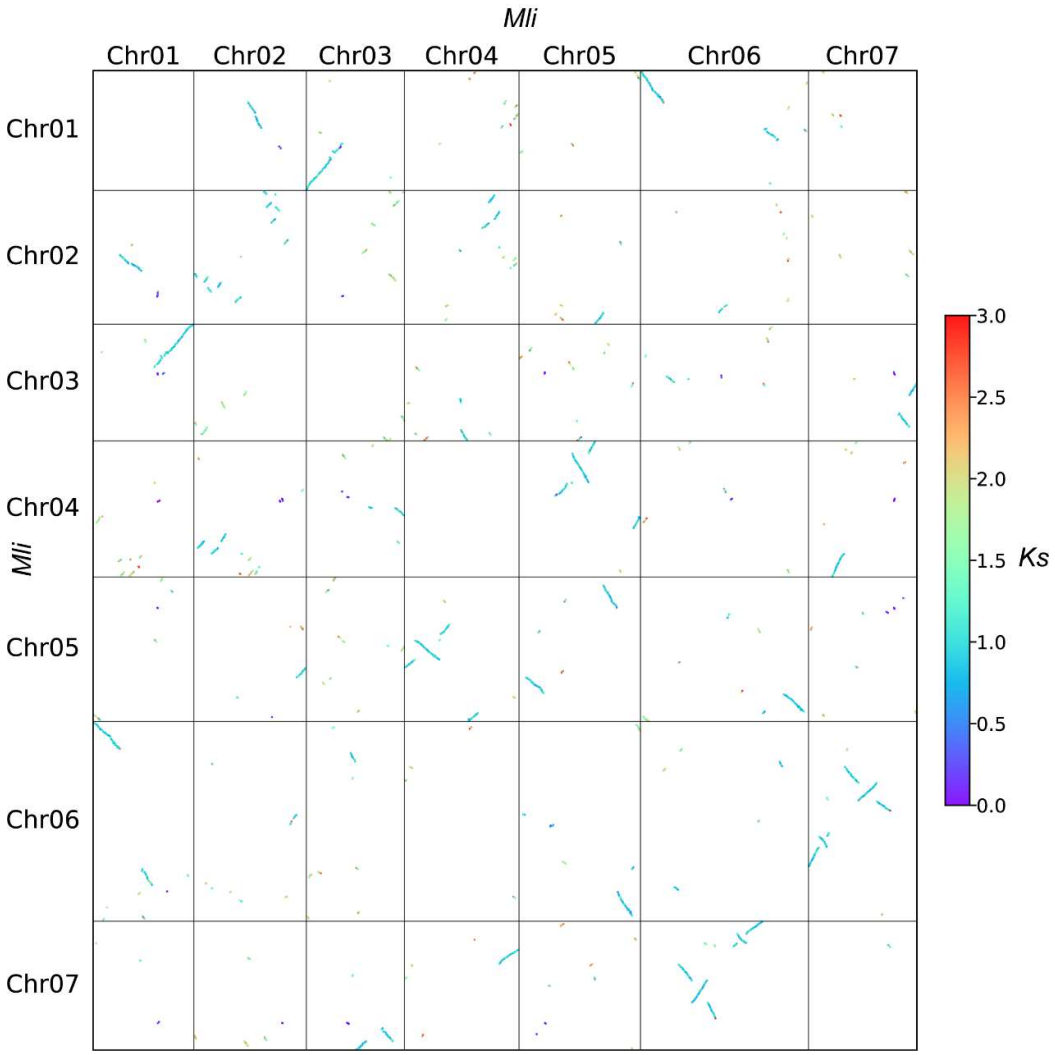
Supplemental Figure 11. Proportion of different repeat types in five genomes.



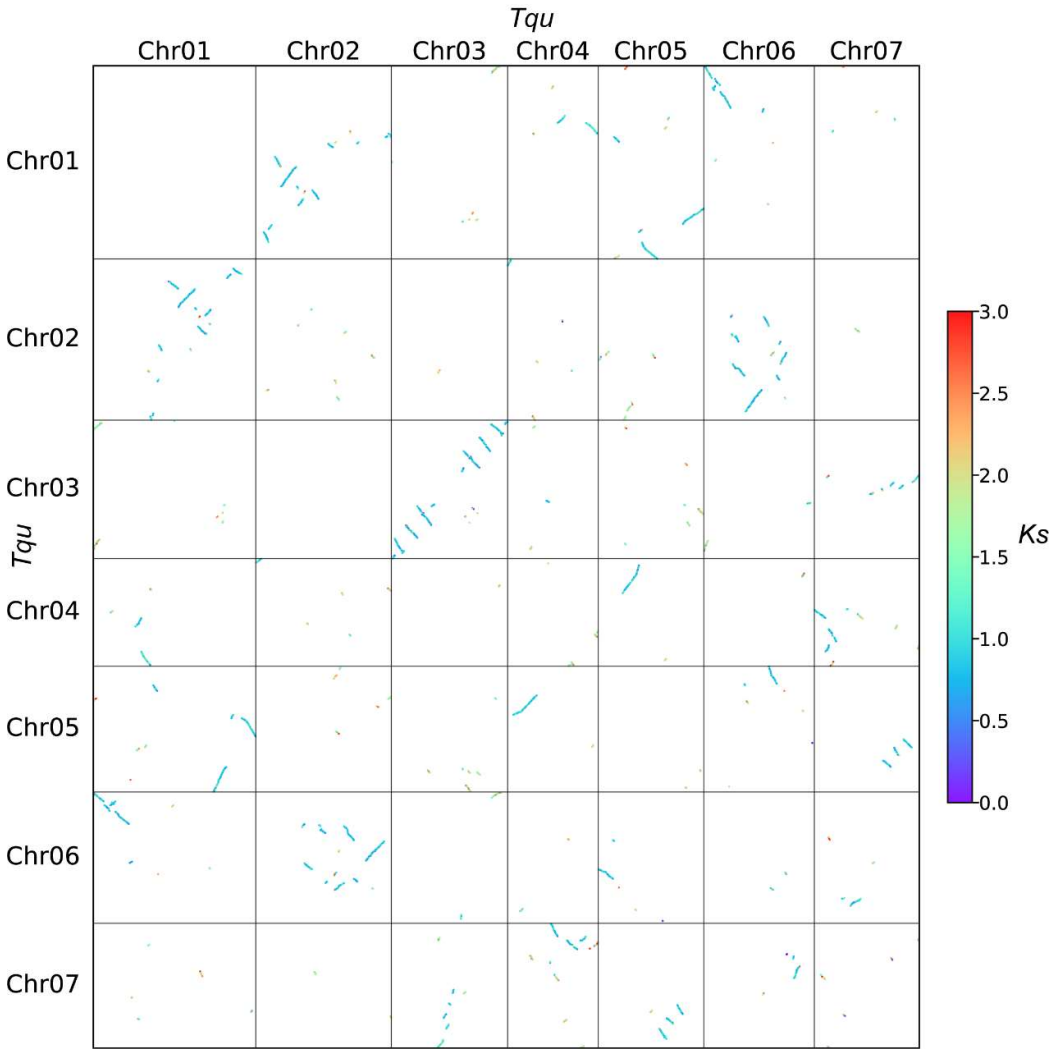
Supplemental Figure 12. Phylogenetic clustering of *Ty3/Gypsy* (a) and *Ty1/Copia* (b) retrotransposons in *Tqu* and *E. syriacum* (*Esy*).



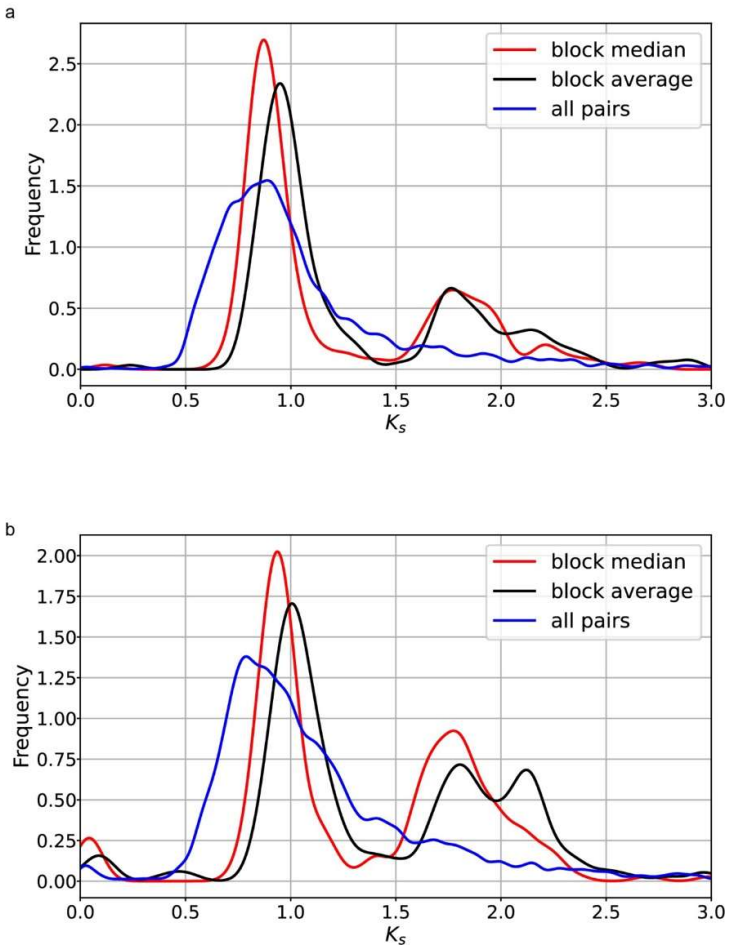
Supplemental Figure 13. The gene intron length distribution of *Tqu* and *Esy*.



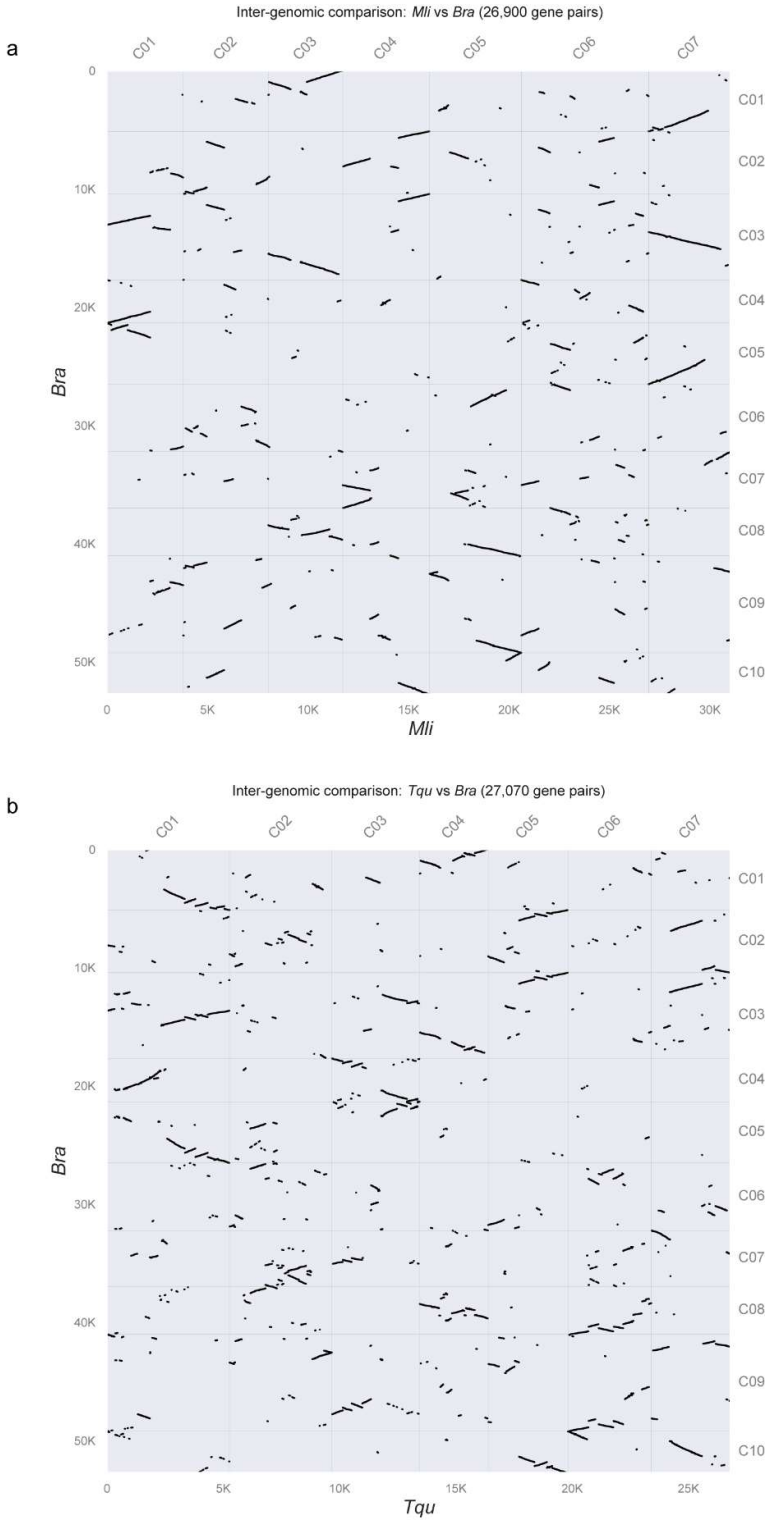
Supplemental Figure 14. Syntenic relationships within *Mli.* Syntenic blocks were colored based on the *Ks* values per syntenic gene pair between syntenic blocks of each genome.



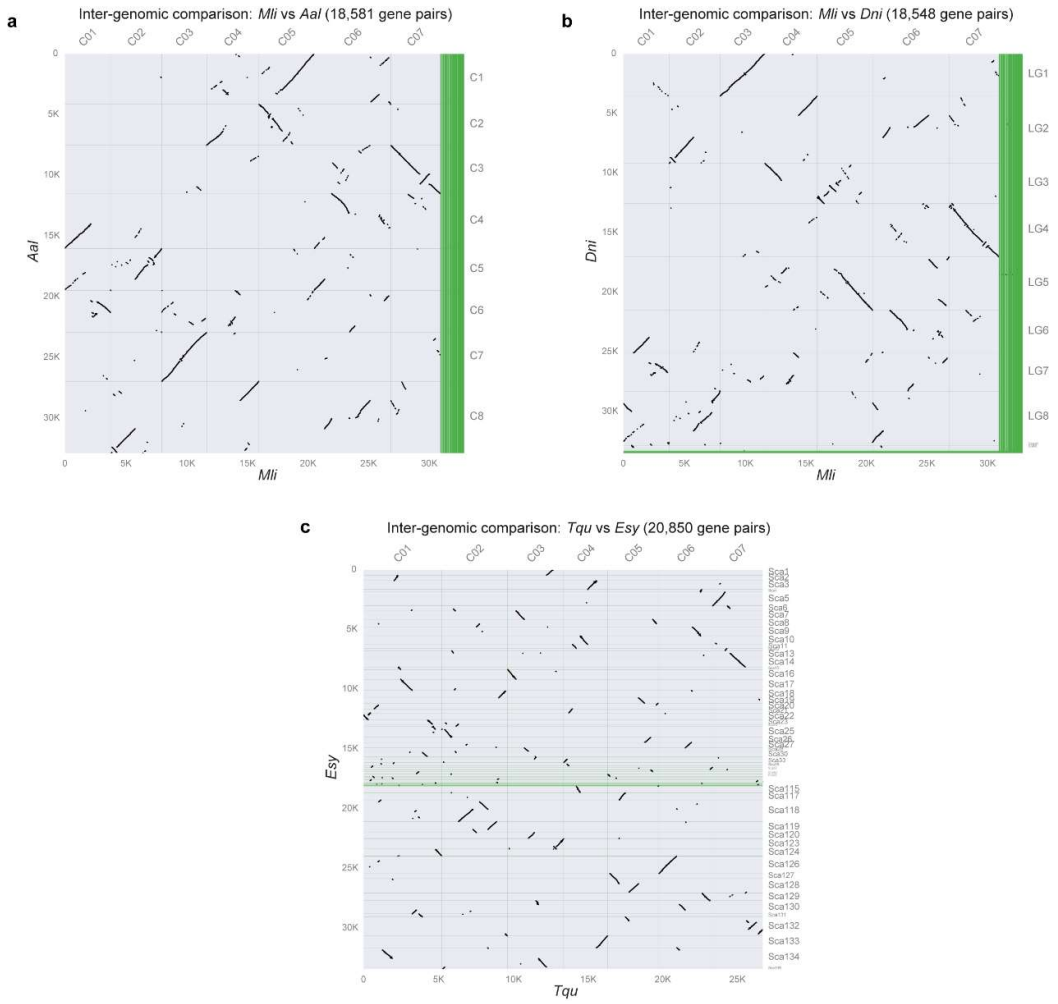
Supplemental Figure 15. Syntenic relationships within *Tqu.*



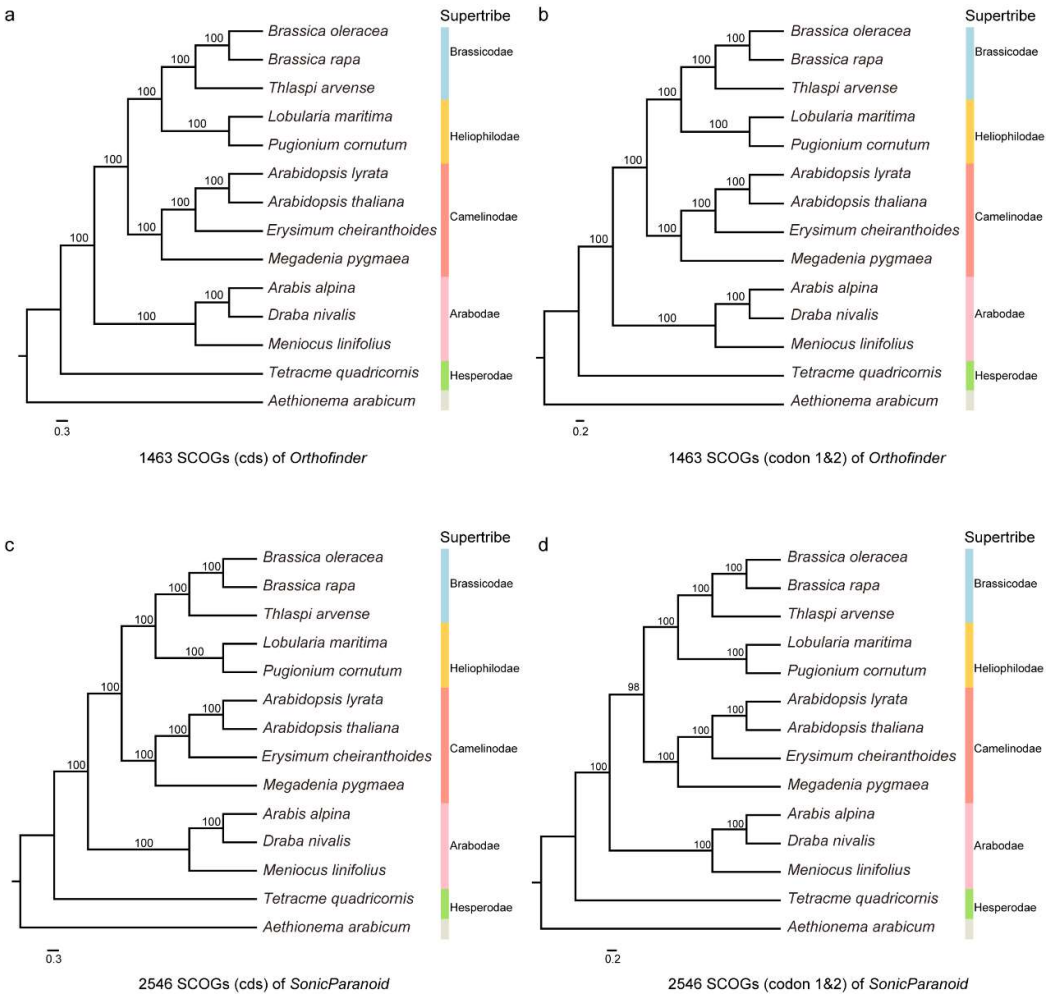
Supplemental Figure 16. Distribution of *Ks* of intragenomic syntenic blocks for *Tqu* (a) and *Mli* (b). Block median and average values were estimated based on the median and average *Ks* for each collinear block.



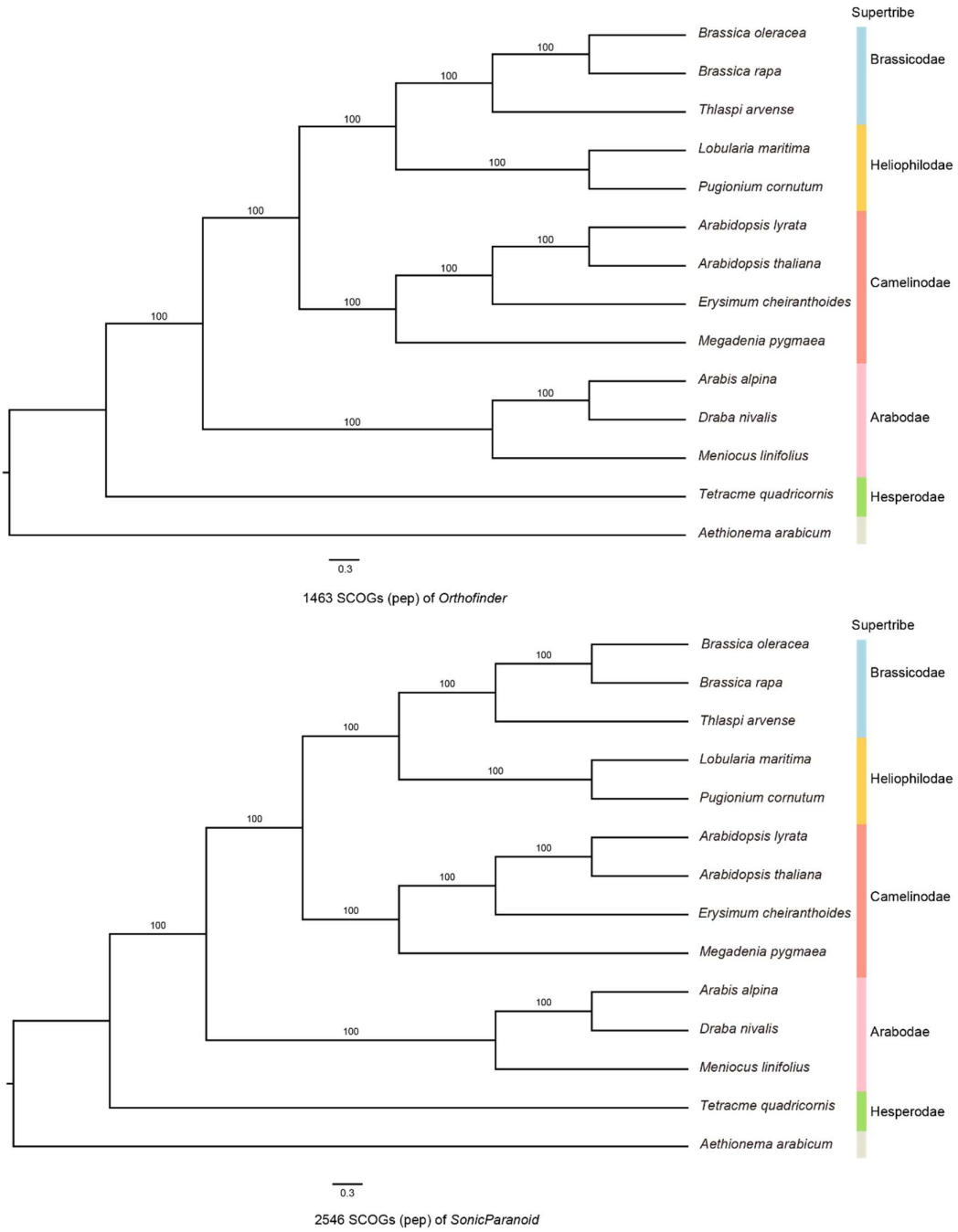
Supplemental Figure 17. Genomic synteny comparisons between *B. rapa* and *Tqu* (a) or *Mli* (b).



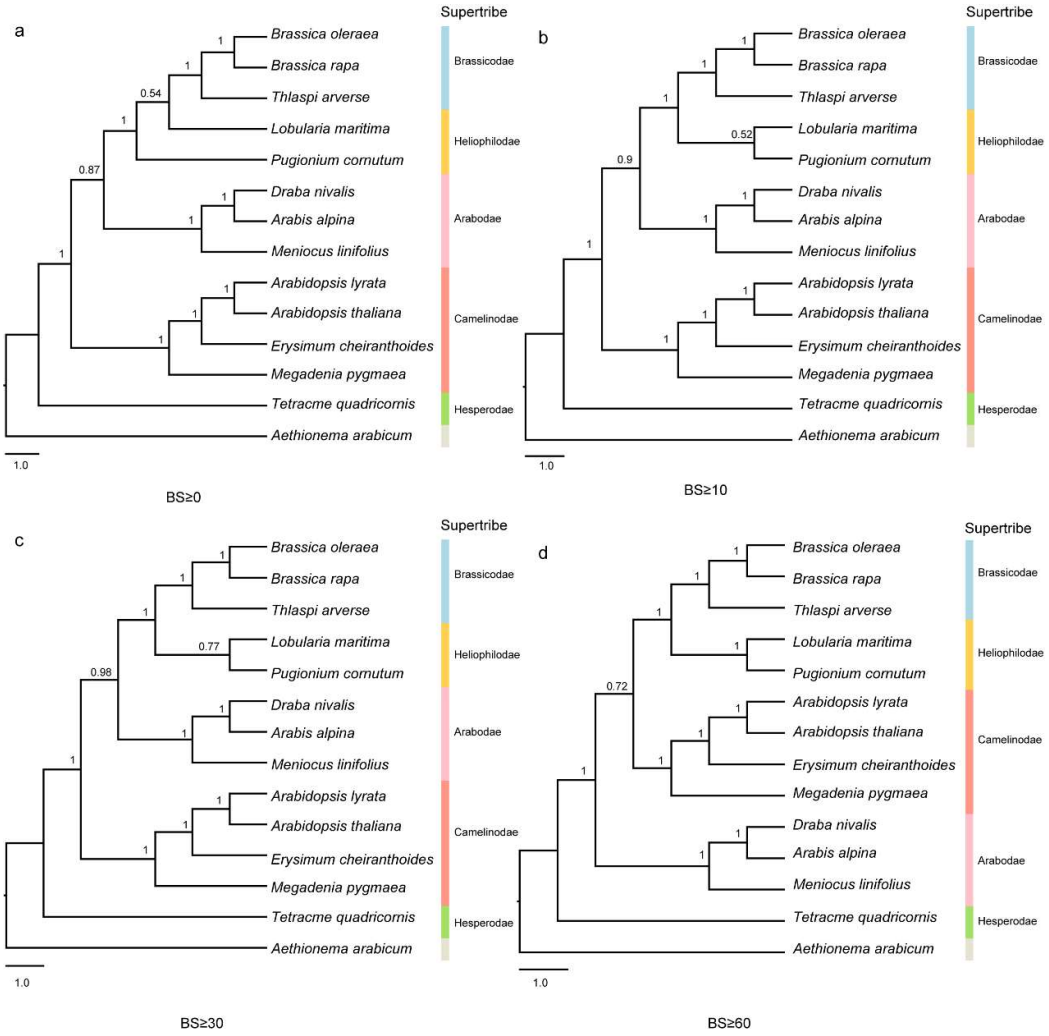
Supplemental Figure 18. Synteny comparisons of *A. alpina* (*Aal*) to Mli (a), *D. nivalis* (*Dni*) to *Mli* (b), and *Esy* to *Tqu* (c).



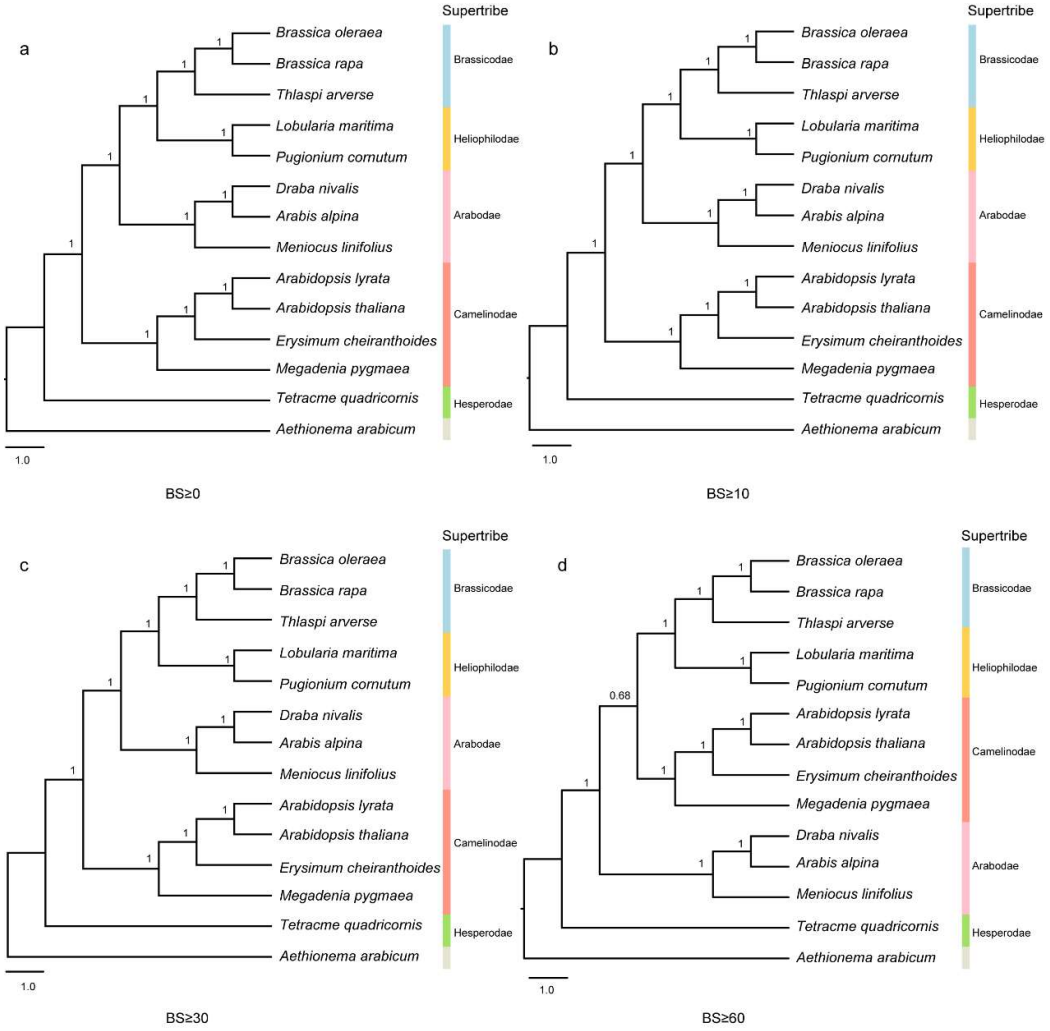
Supplemental Figure 19. Maximum likelihood (ML) phylogenies based on concatenated nucleotide sequences using *Orhofinder* (a for CDS and b for codons 1 & 2) and *SonicParanoid* (c for CDS and d for codons 1 & 2) with bootstrap (BS) supports shown.



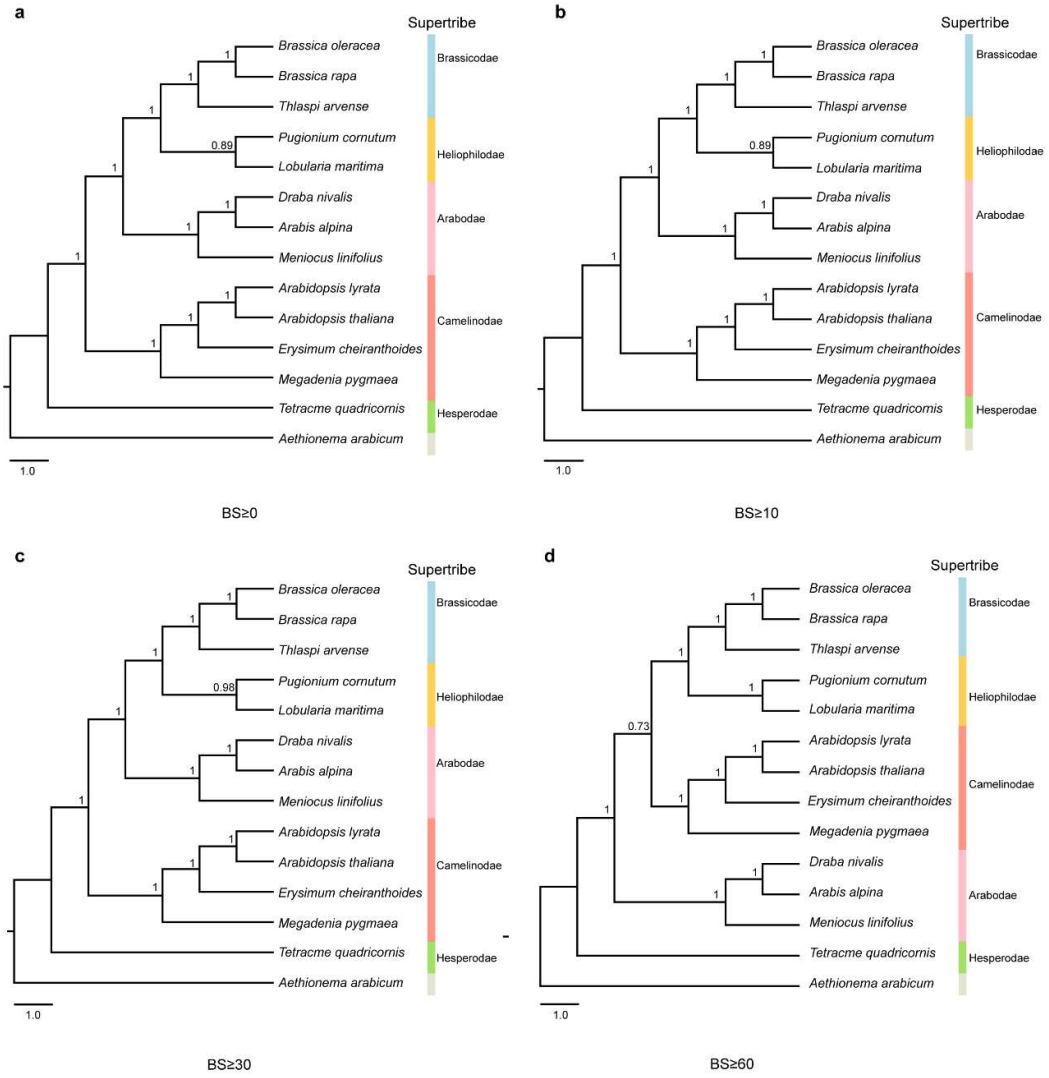
Supplemental Figure 20. Maximum likelihood (ML) phylogenies based on concatenated amino acids using *Orhofinder* (upper) and *SonicParanoid* (lower) with BS supports shown.



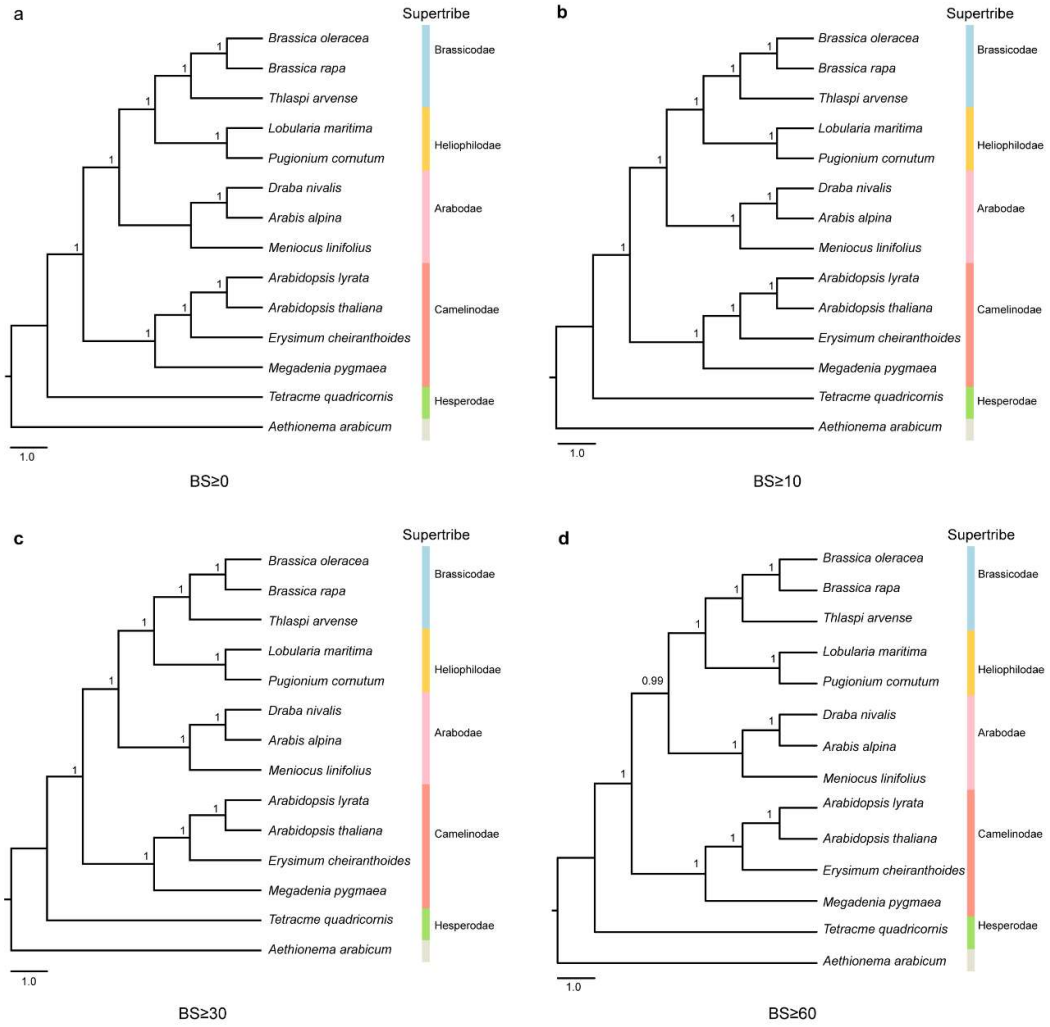
Supplemental Figure 21. Effect of BS values of the 1463 input individual gene trees for coalescent-based phylogenetic analyses using coding sequences. a, BS≥0; b, BS≥10; BS≥30; BS≥60. Phylogenetic trees inferred based on 1463 individual gene trees with CD S detected through *orthofinder*. Local posterior probability (LPP) values for the coalescent-based analyses are shown on the branch to each node.



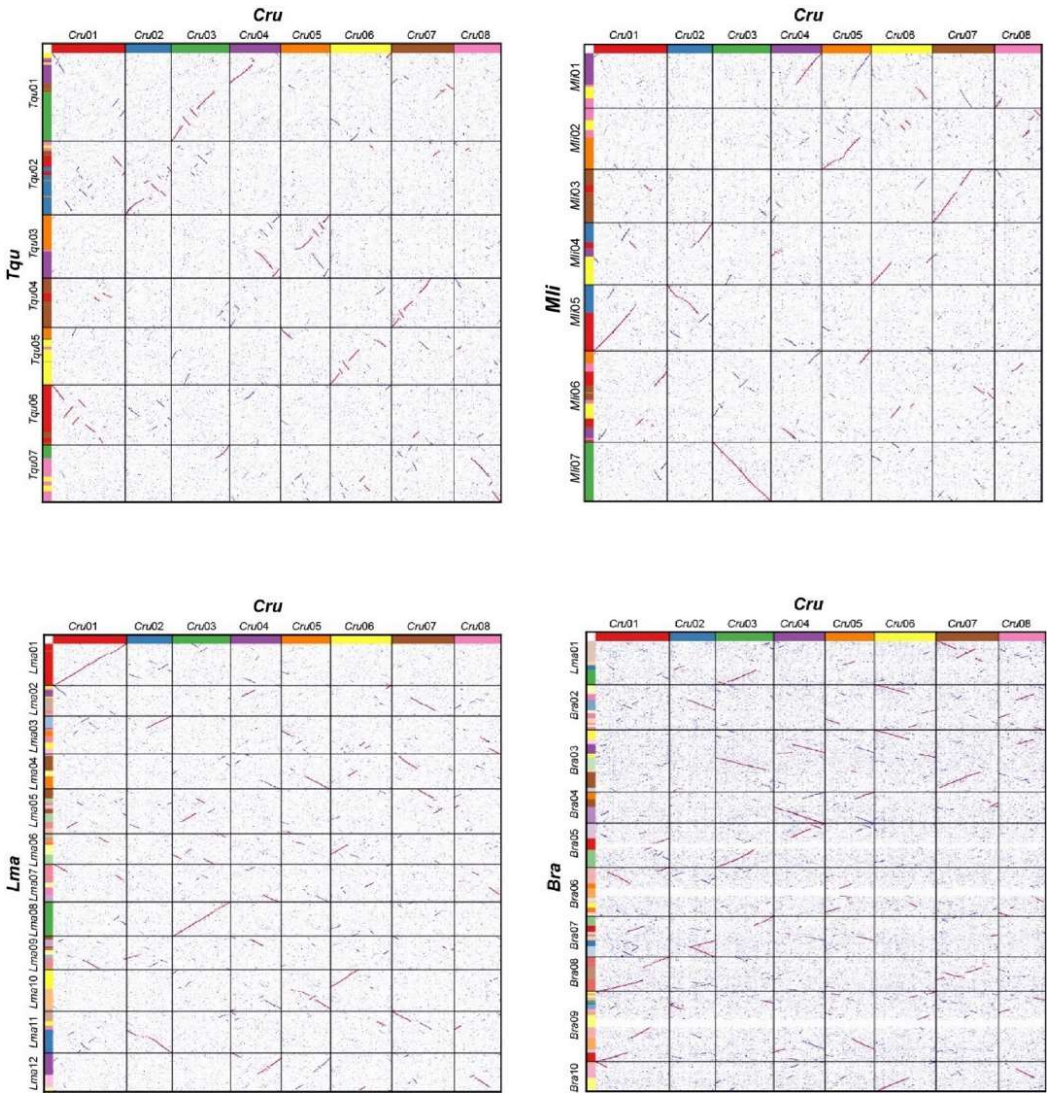
Supplemental Figure 22. Effect of BS values of the 1463 input individual gene trees for coalescent-based phylogenetic analyses using amino acid sequences. A, BS≥0; b, BS≥10; BS≥30; BS≥60. Phylogenetic trees inferred based on 1463 individual gene trees with amino acid sequences detected through *orthofinder*. LPP values for the coalescent-based analyses are shown on the branch to each node.



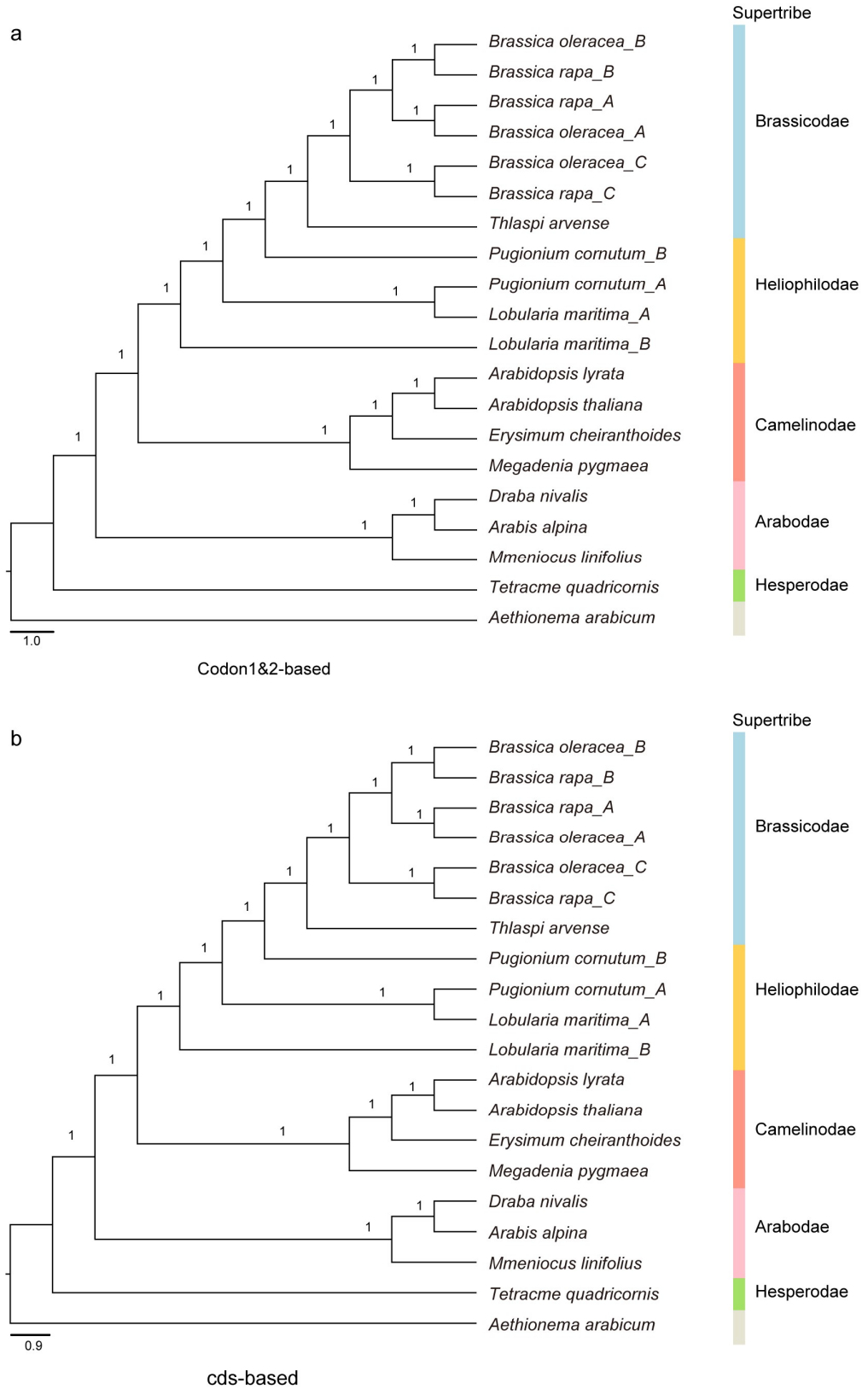
Supplemental Figure 23. Effect of BS values of the 2546 input individual gene trees for coalescent-based phylogenetic analyses using coding sequences. A, BS≥0; b, BS≥10; BS≥30; BS≥60. Phylogenetic trees inferred based on 2546 individual gene trees with CDS detected through *SonicParanoid*. LPP values for the coalescent-based analyses are shown on the branch to each node.



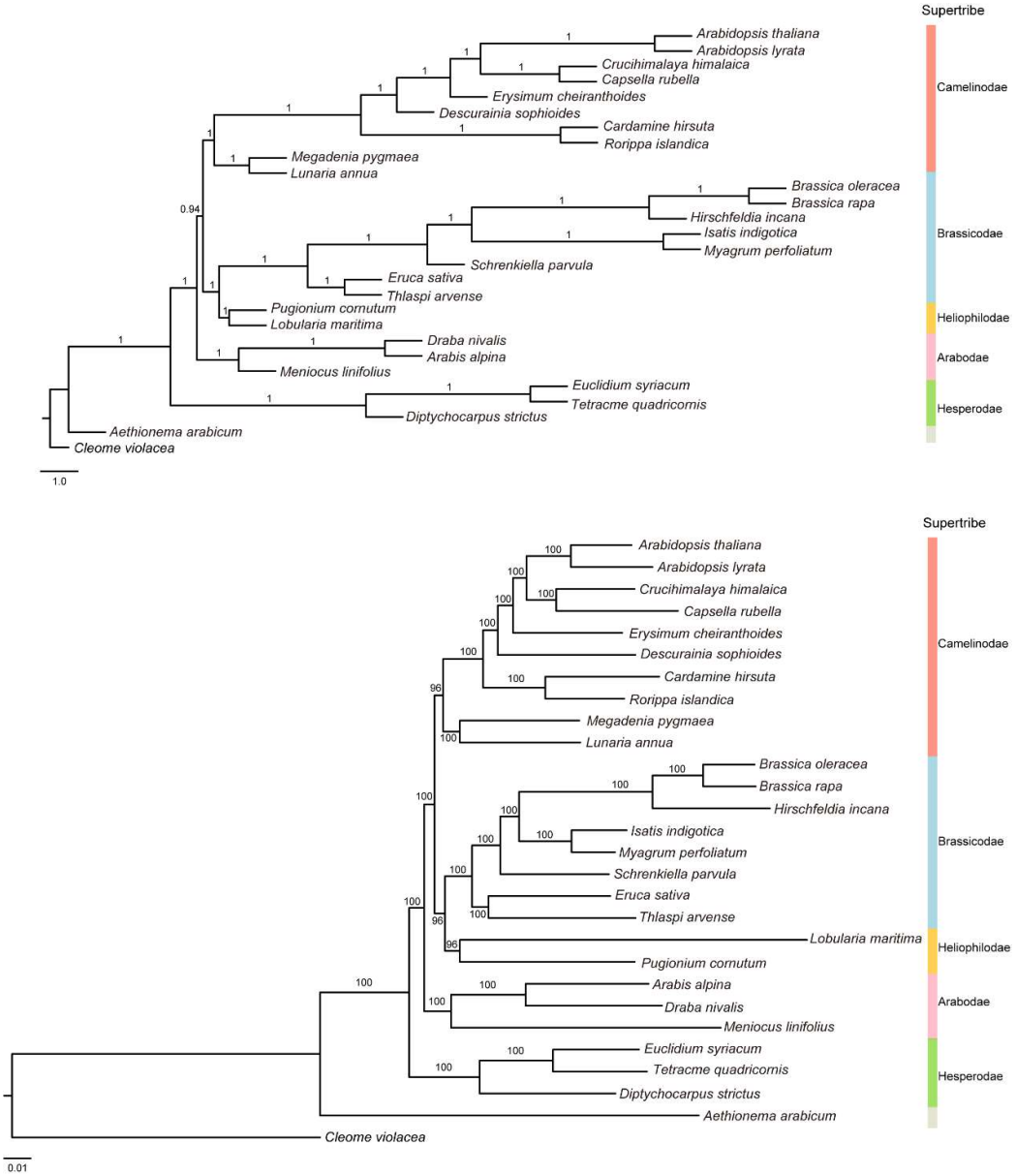
Supplemental Figure 24. Effect of BS values of the 2546 input individual gene trees for coalescent-based phylogenetic analyses using amino acid sequences. A, BS≥0; b, BS≥10; BS≥30; BS≥60. Phylogenetic trees inferred based on 2546 individual gene trees with amino acid sequences detected through *SonicParanoid*. LPP values for the coalescent-based analyses are shown on the branch to each node.



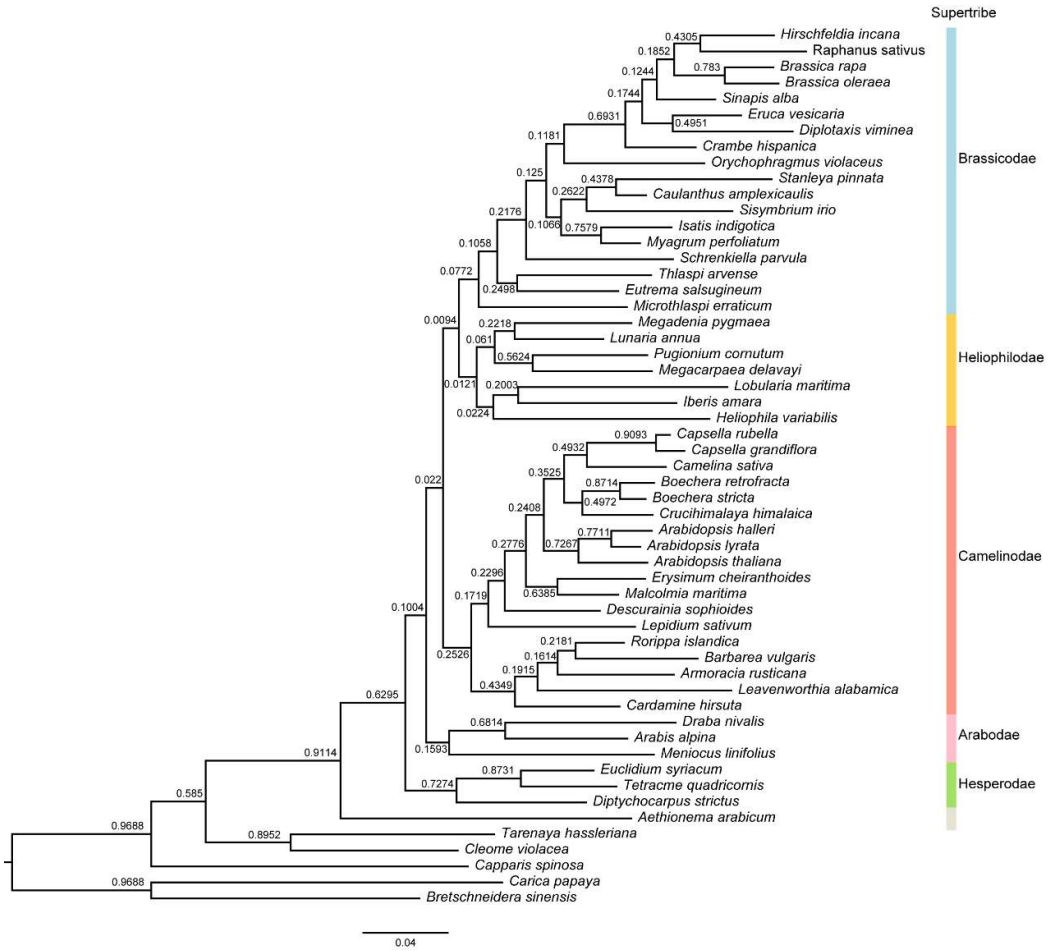
Supplemental Figure 25. Genomic synteny between *Capsella rubella* (*Cru*) and four Brassicaceae species.



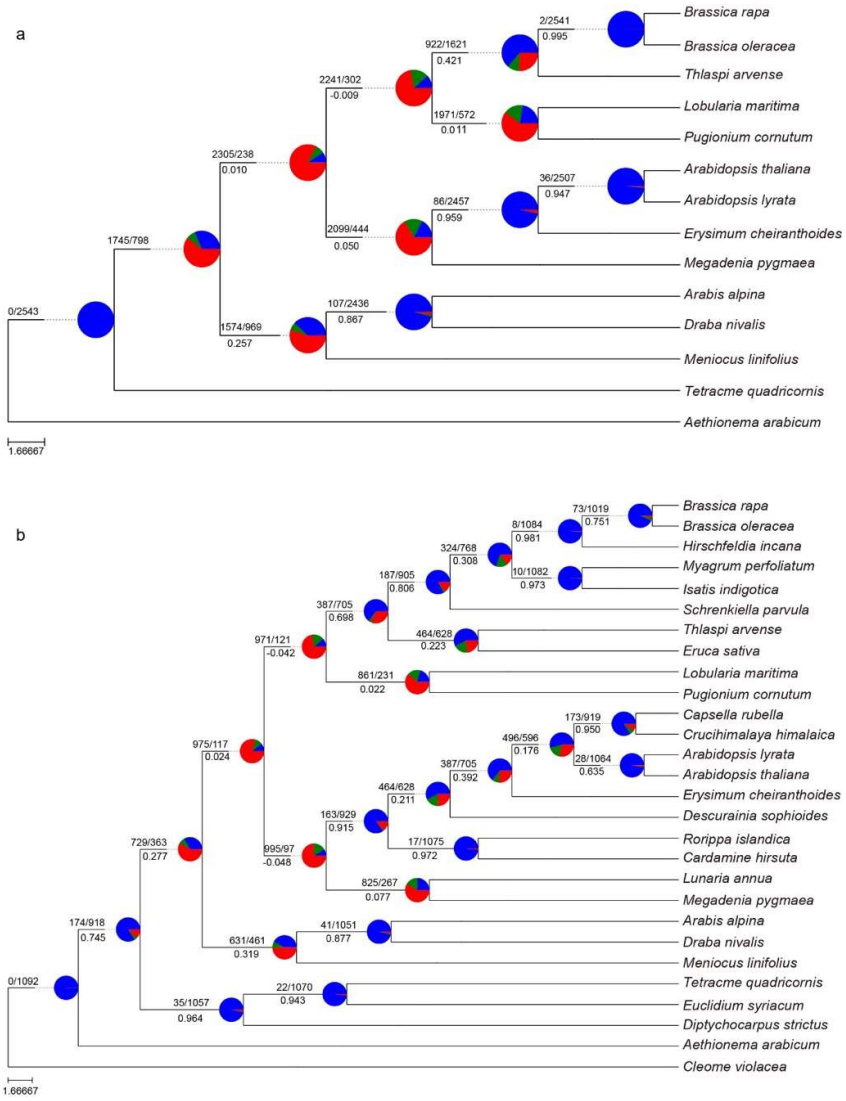
Supplemental Figure 26. Synteny-based coalescent trees using 4434 collinear genes (a, codon 1&2 based; b, CDS based) with LPP values shown.



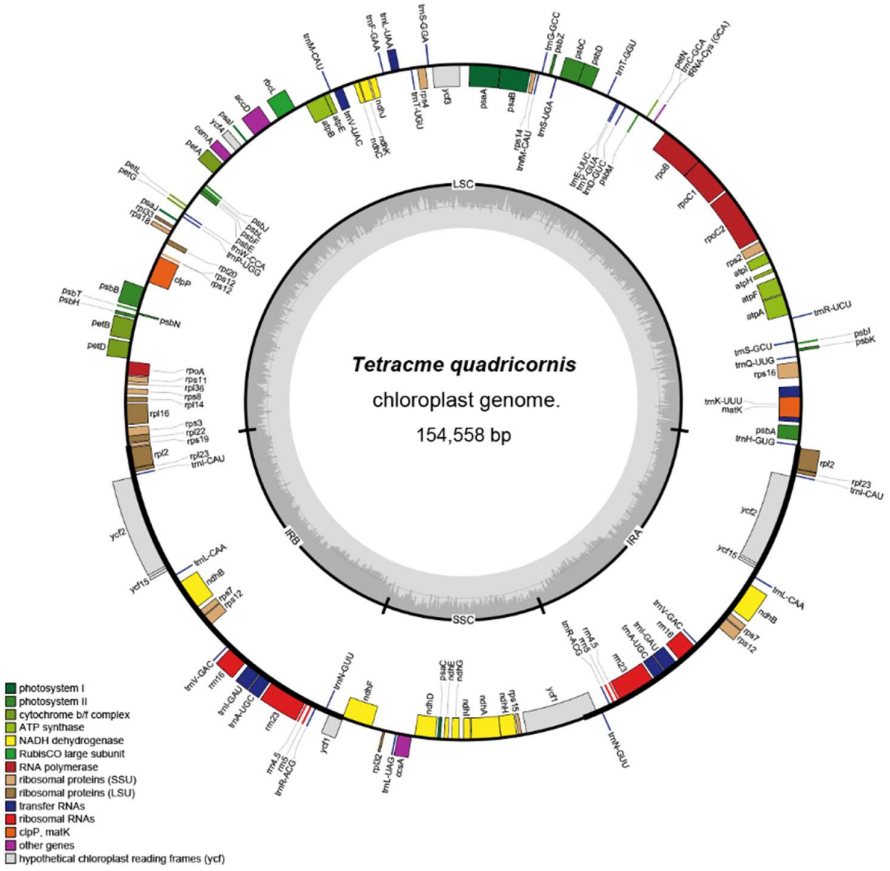
Supplemental Figure 27. Phylogenies using 1092 single copy orthologues for 28 species (a, coalescent based; b, concatenation based) with *Cleome violacea* (Cleomaceae) as outgroup. LPP values (a) and BS values (b) are shown.



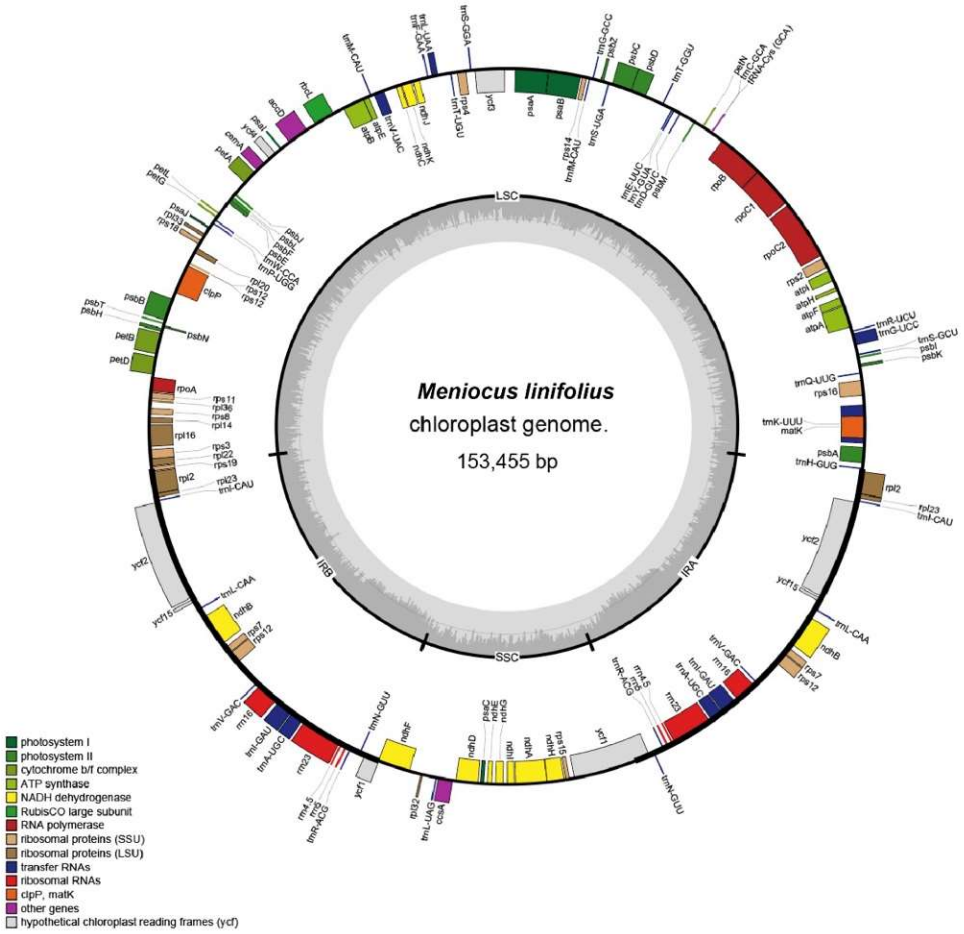
Supplemental Figure 28. Phylogeny inferred with 5217 low-copy ortholog genes (shared ortholog groups) for 55 Brassicale species. Support values (similar to LPP) generated via *STAG* (v1.0.0) are shown.



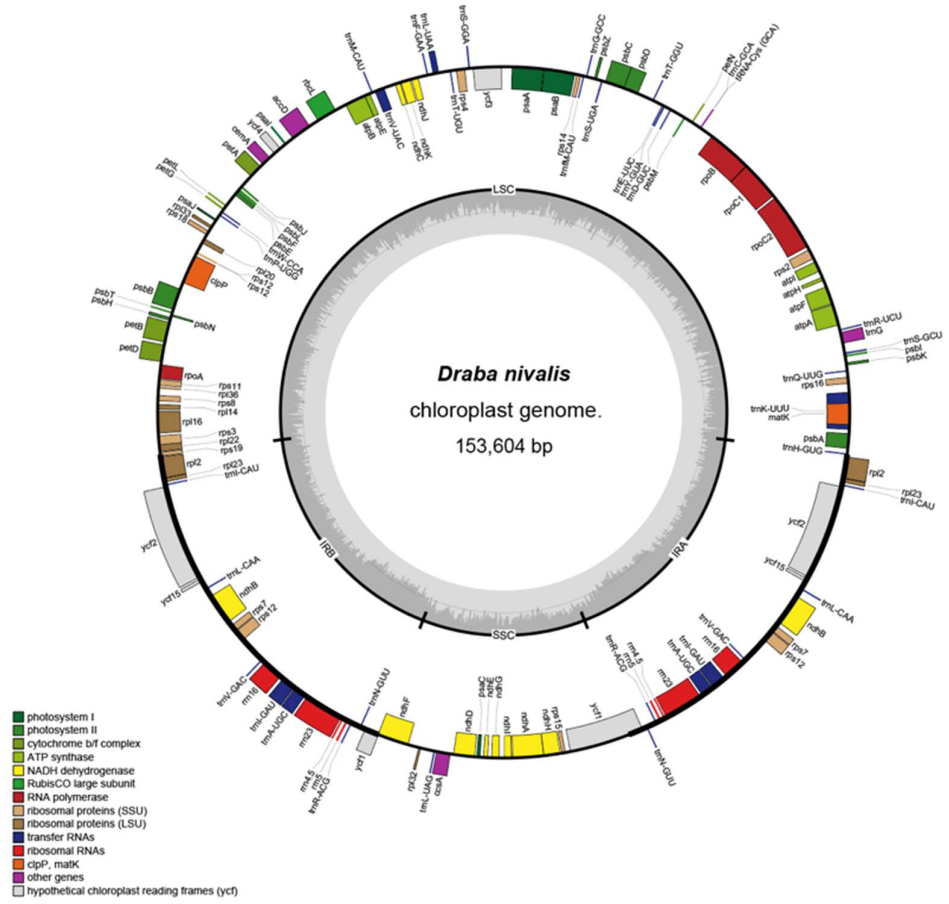
Supplemental Figure 29. Concordances and conflicts between nuclear gene and species trees with *PHYPARTS* (a for 14 species and b for 28 species). Pie charts next to the nodes indicate the proportions of gene trees supporting the relationship shown (blue), the main alternative bipartition for that clade (green), or the remaining alternative bipartitions (red). Numbers of gene tree conflicting (before the '/') and concordant (after the '/') with a bipartition on the species trees are indicated under the branch. Values under the branches show internode certainty (ICA) values.



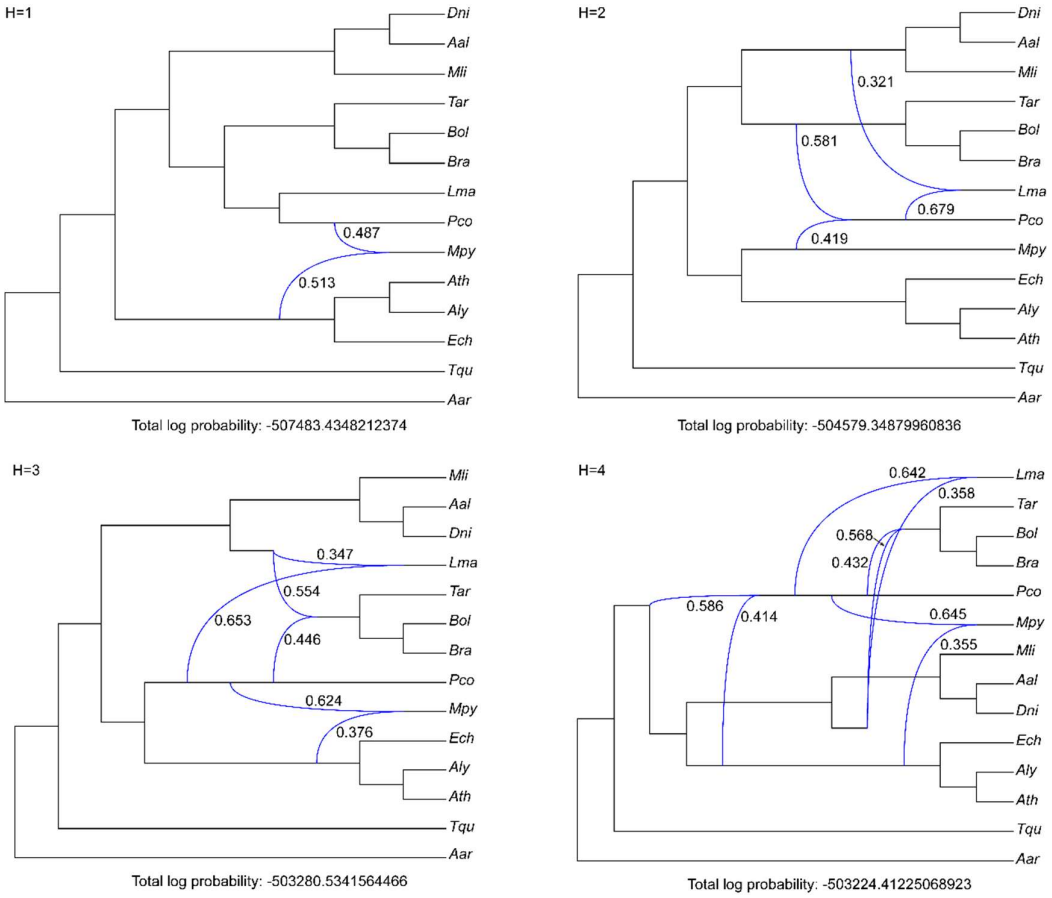
Supplemental Figure 30. Chloroplast genome structure for *Tqu*.



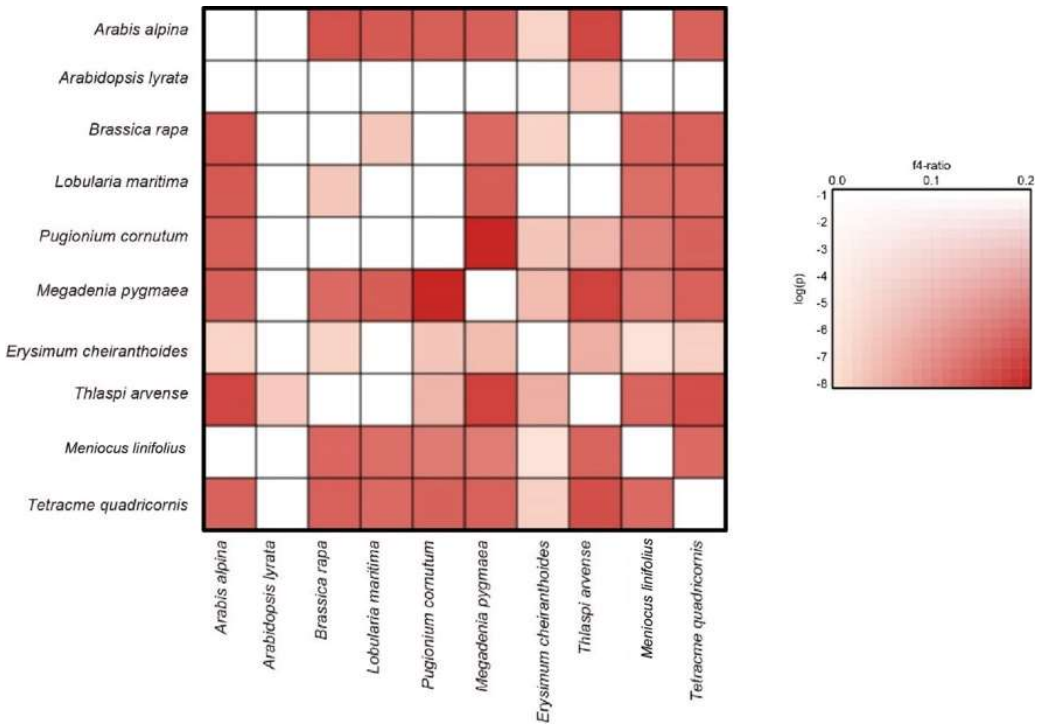
Supplemental Figure 31. Chloroplast genome structure for *Mli*.



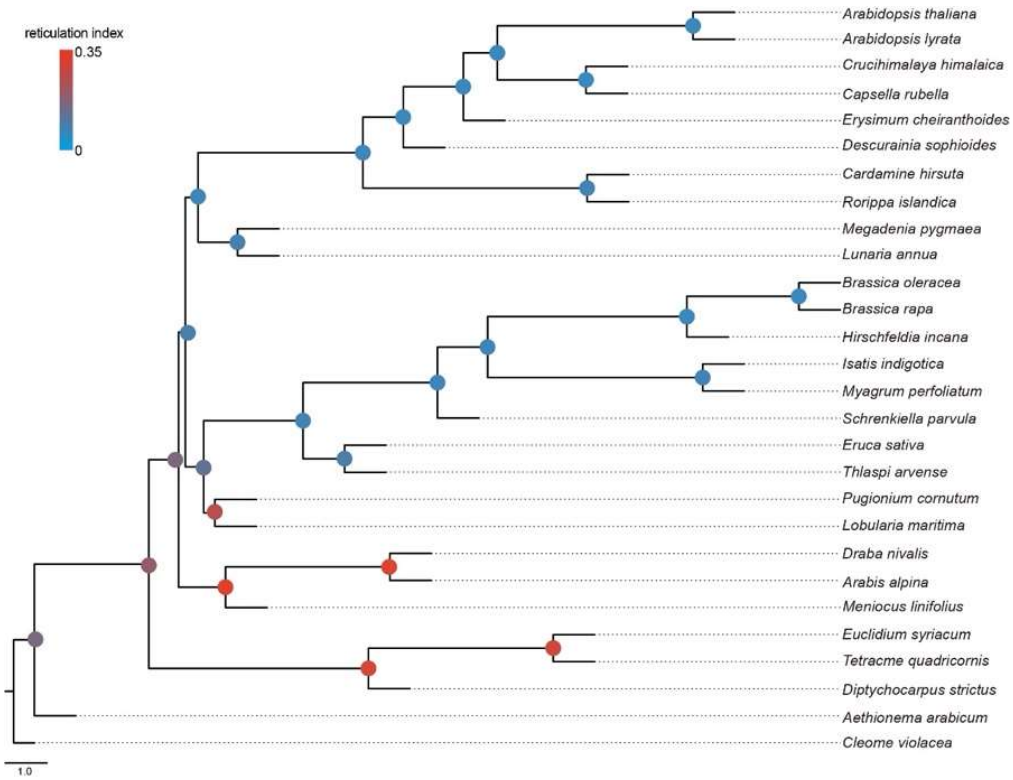
Supplemental Figure 32. Chloroplast genome structure for *Dni*.



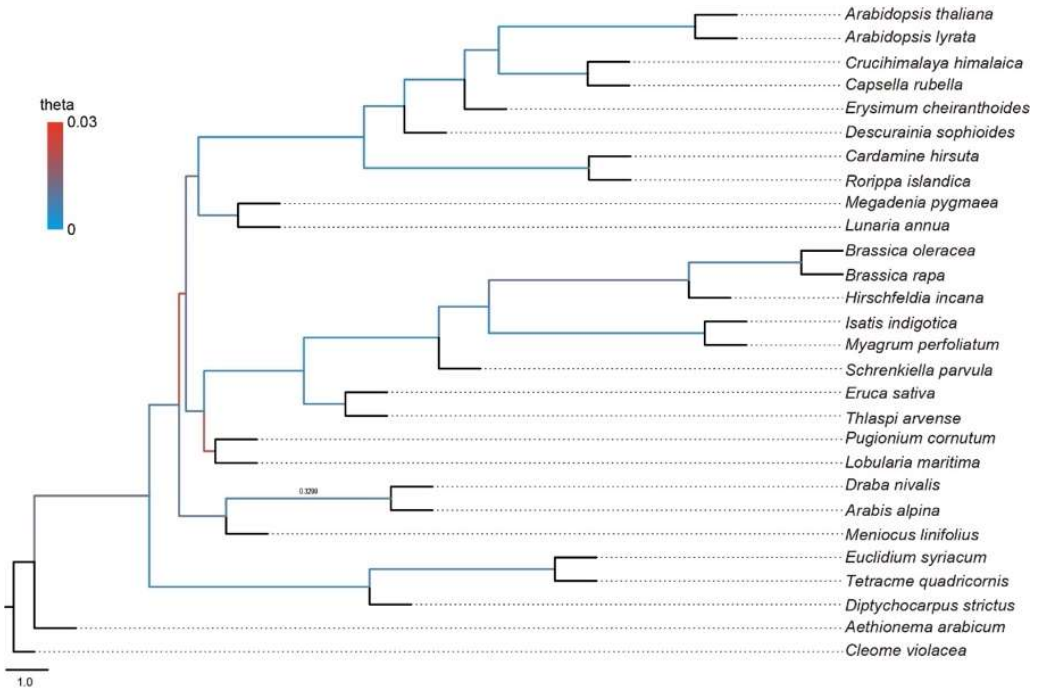
Supplemental Figure 33. Reticulation history analyses using *PhyloNet* with the number of reticulations set to 1 (a), 2 (b), 3 (c), and 4 (d), respectively.



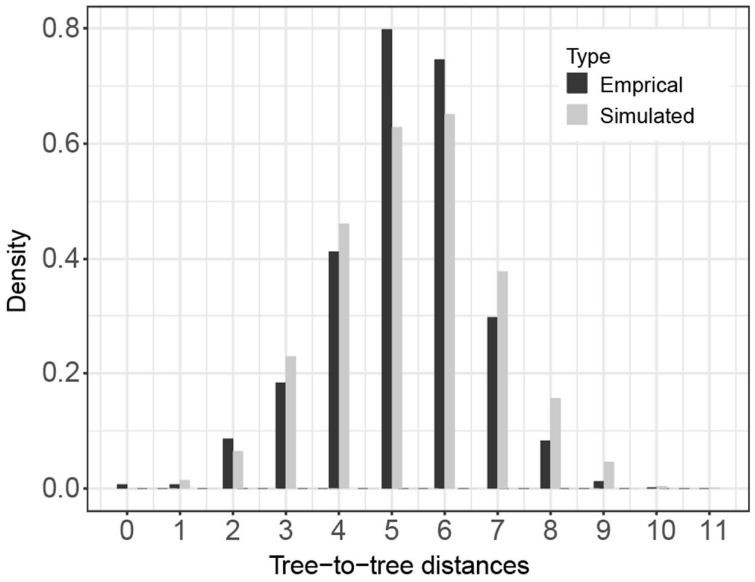
Supplemental Figure 34. Maximum pairwise *f4* heatmaps between pairs of species across all combinations of trios. The depth of color (from white to red) shows the level of f_4 values.



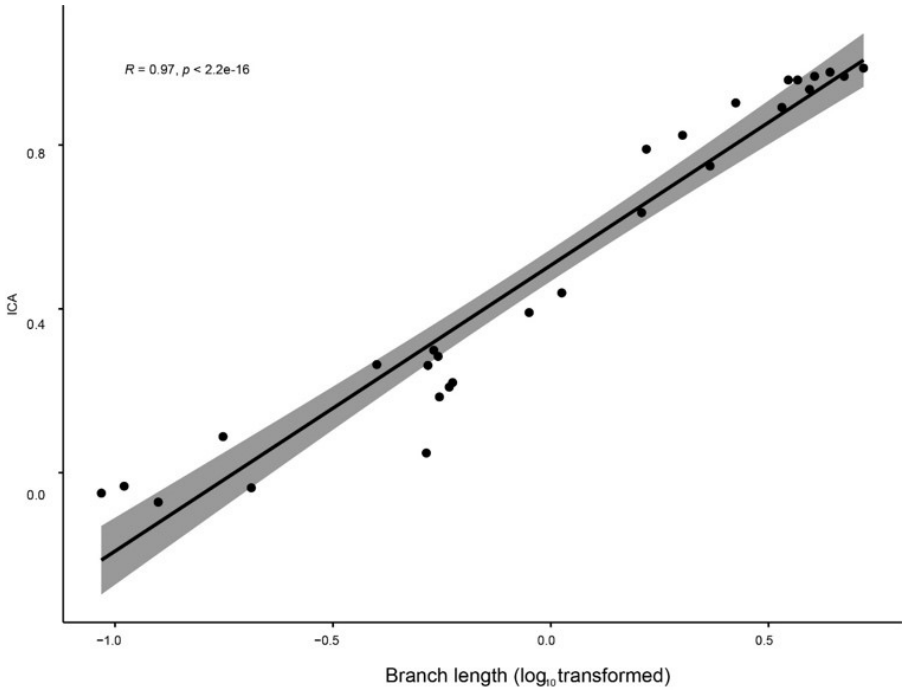
Supplemental Figure 35. Reticulation Index among Brassicaceae. Node color close to red indicates higher probabilities of hybridization.



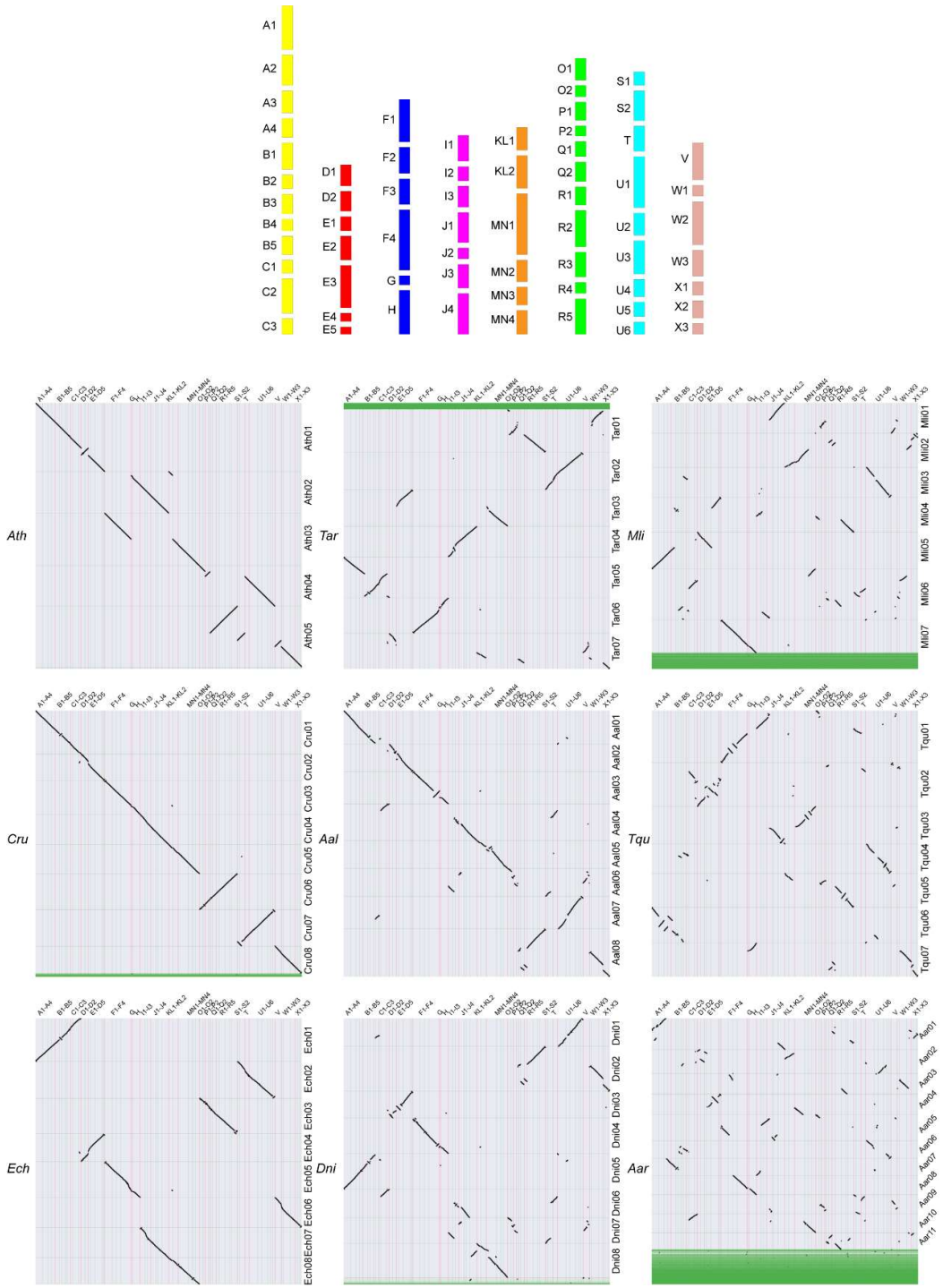
Supplemental Figure 36. Estimated theta values for each internal branch of Brassicaceae. Phylogenetic tree with branches was colored by the inferred population mutation parameter theta, which reflects the population polymorphism generated by dividing the mutation units for each internal branch by coalescent units. Grey colors denote branches where theta is unable to be computed due to lack of data.



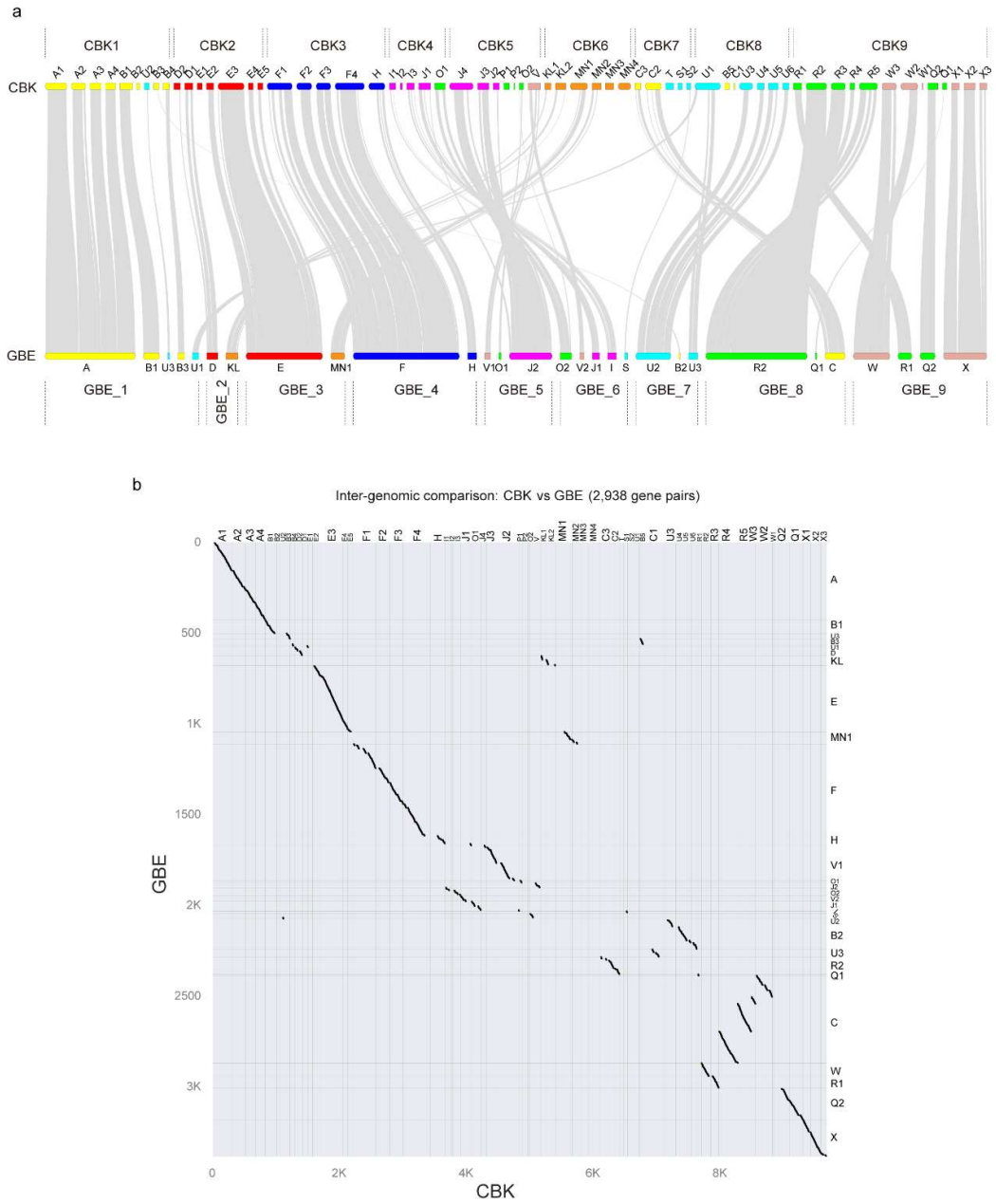
Supplemental Figure 37. Coalescent simulation shows the distribution of tree-to-tree distances between empirical gene trees and the *ASTRAL* species tree (black bars) compared to those from the coalescent simulation (grey bars).



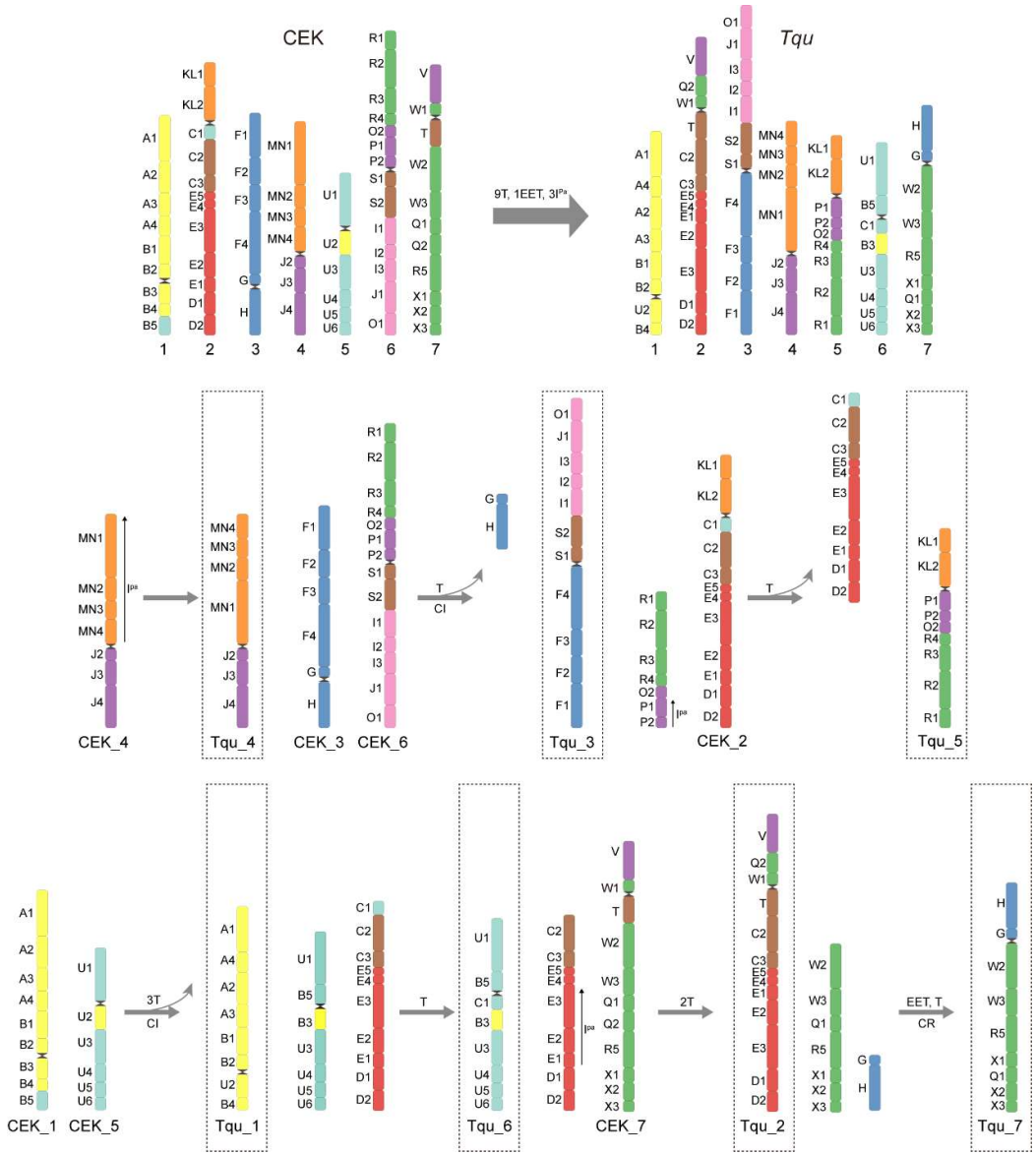
Supplemental Figure 38. Correlation between length of branch and internode certainty (ICA) value tested with a *Pearson's correlation test* (*two-sided*).



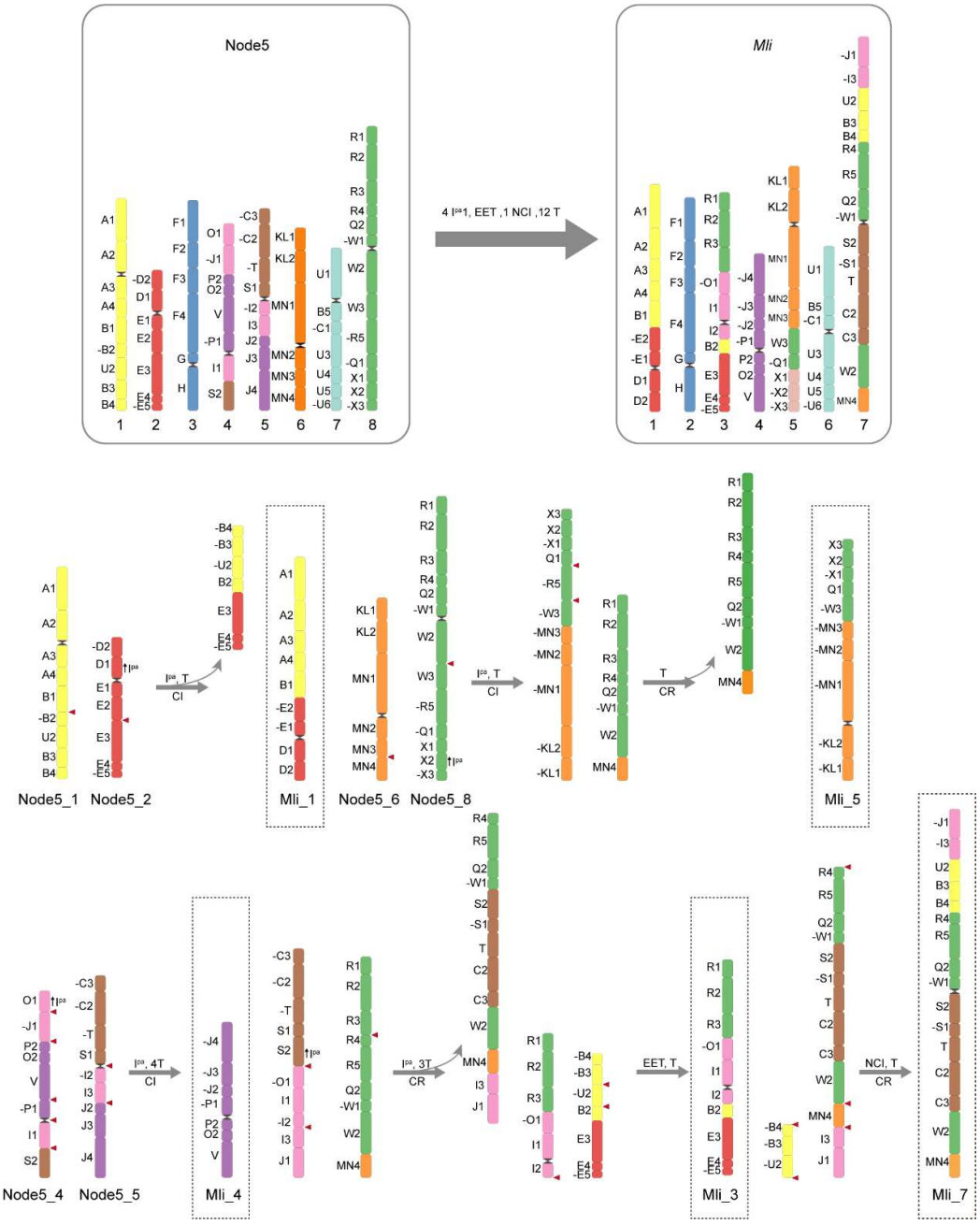
Supplemental Figure 39. Reconstruction of the 65 CBK genomic blocks based on the syntenic relationships between ACK and nine extant species.



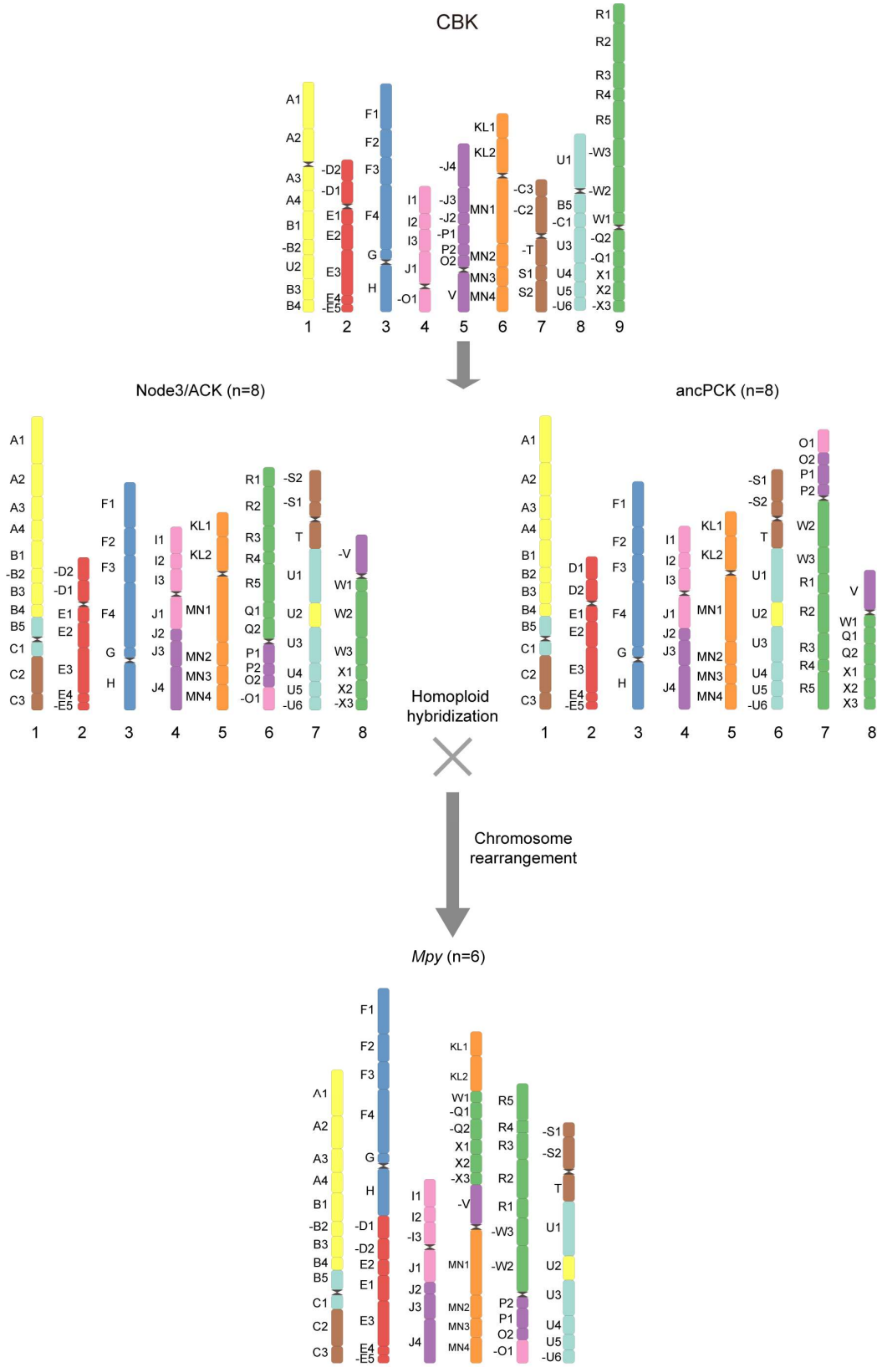
Supplemental Figure 40. The comparison between CBK and GBE (Walden and Schranz, 2023, Genome Biology and Evolution). a, The ancestral genomes of CBK (left) and GBE (Right). b, The syntenic relationship between CBK (upper) and GBE (lower). c, Dot plots between CBK and GBE.



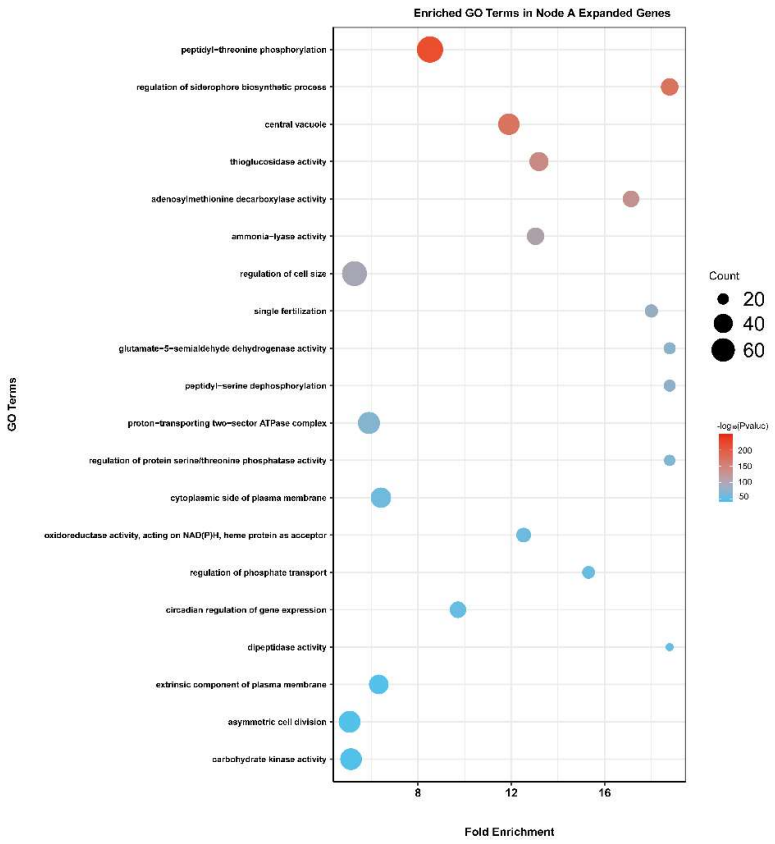
Supplemental Figure 41. Deduced scenario for the chromosome evolution from CEK to *Tqu*. Position of block T was determined based on its location in *Tqu*. Considering that CEK was procured via cytological methods (comparative chromosome painting, CCP), the direction of each genomic block was not factored into this deduction.



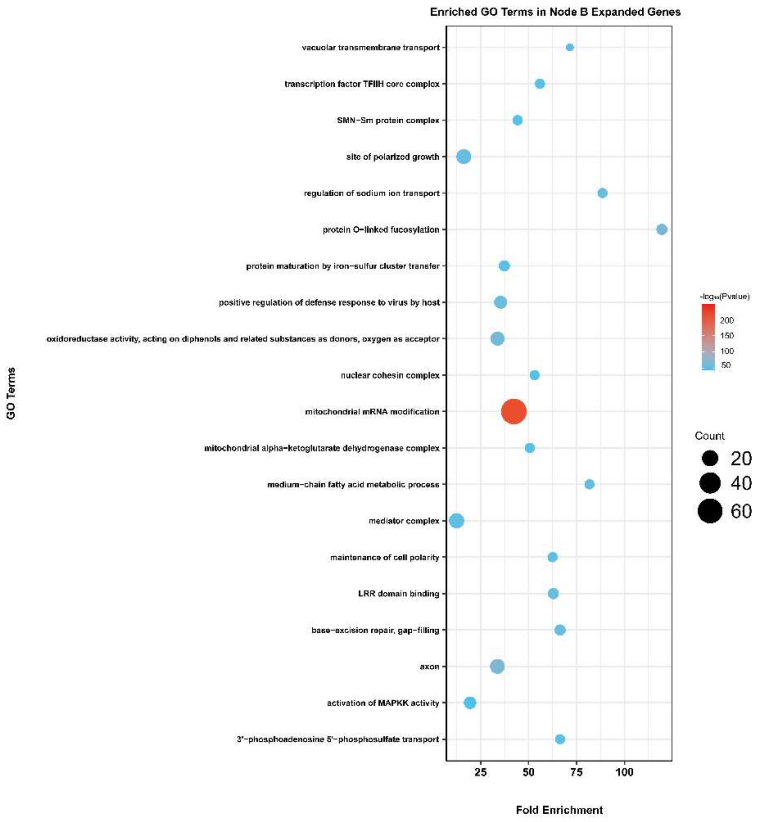
Supplemental Figure 42. Deduced scenario for the chromosome evolution in *Mli*.



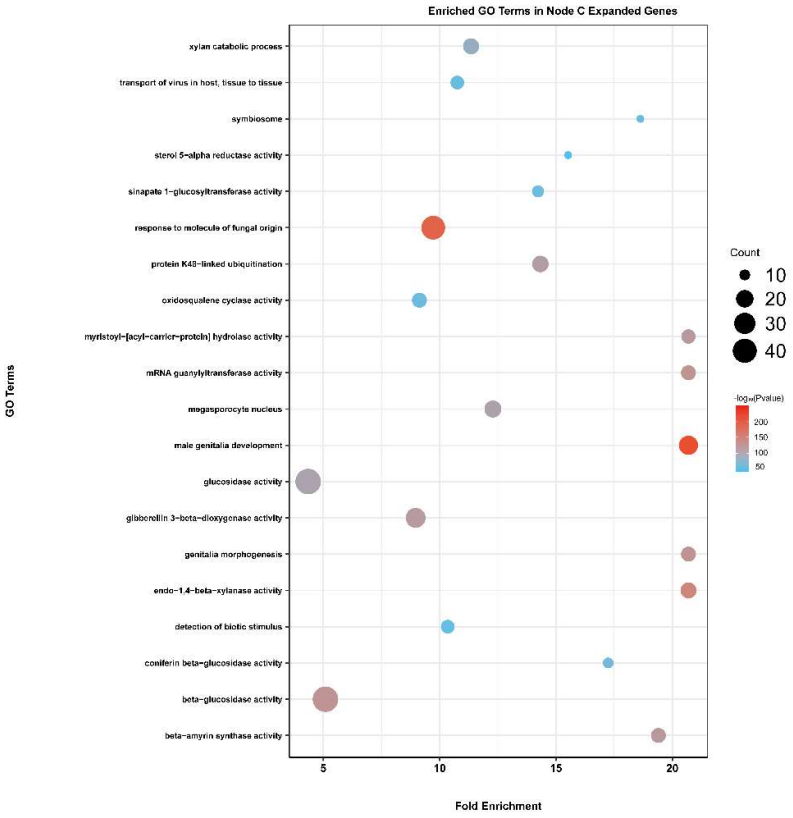
Supplemental Figure 43. Deduced scenario of the hybridization origin of *Megadenia pygmaea*.



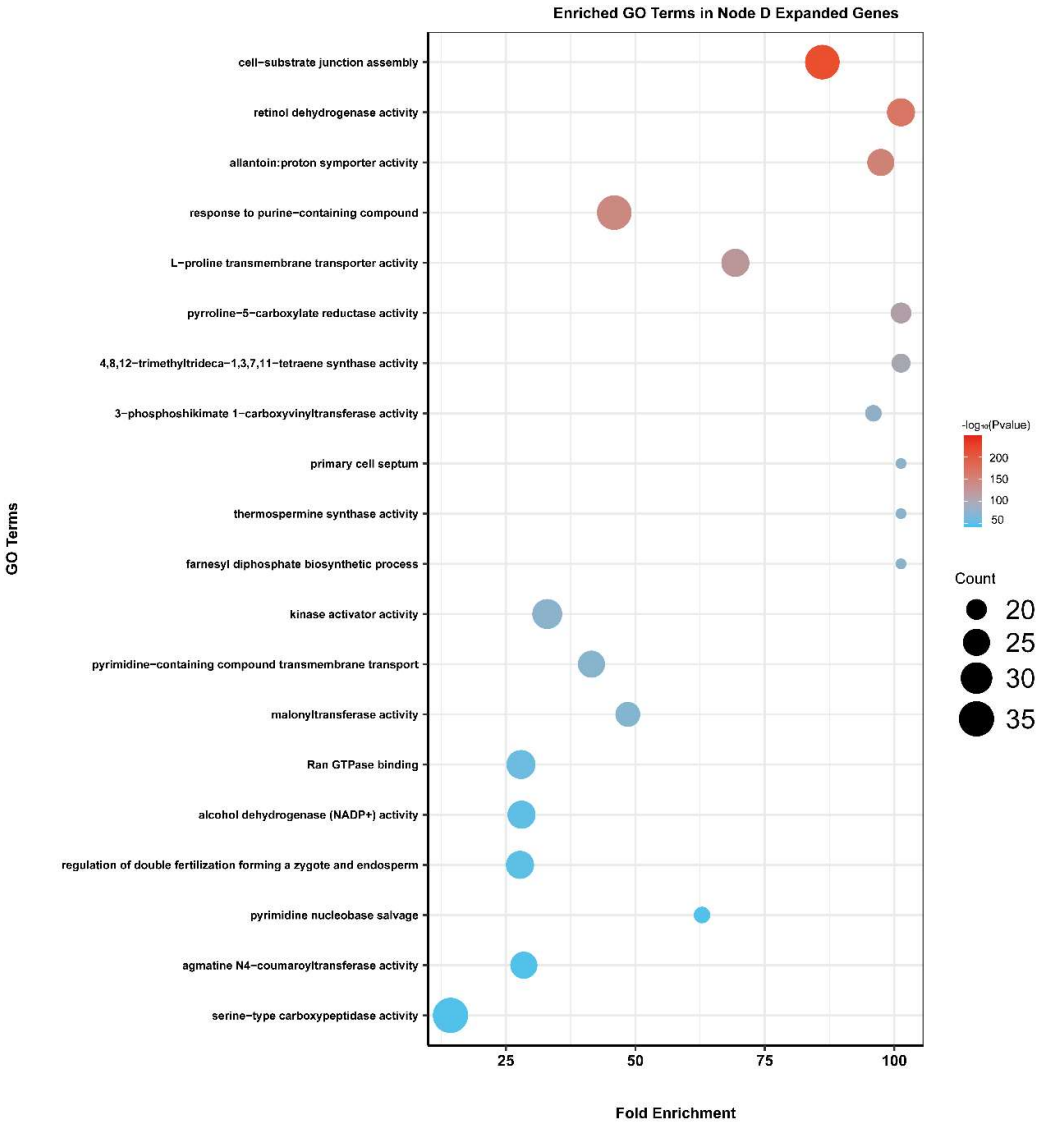
Supplemental Figure 44. Top 20 enriched Gene Ontology (GO) terms of the expanded gene families at node A of Figure 1F.



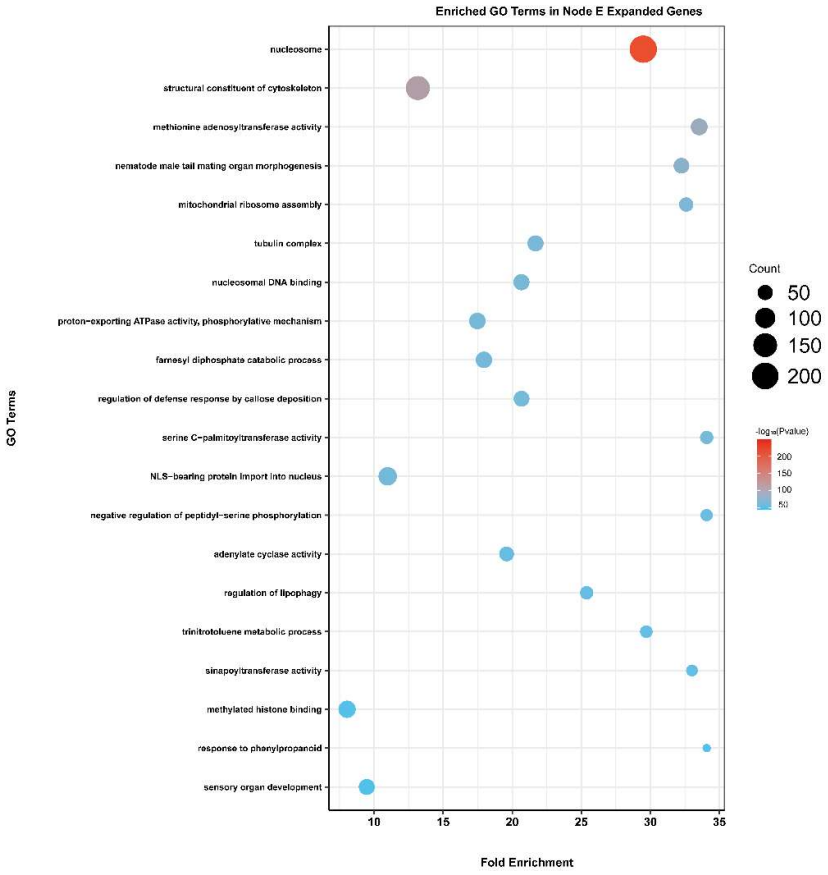
Supplemental Figure 45. Top 20 enriched Gene Ontology (GO) terms of the expanded gene families at node B of Figure 1F.



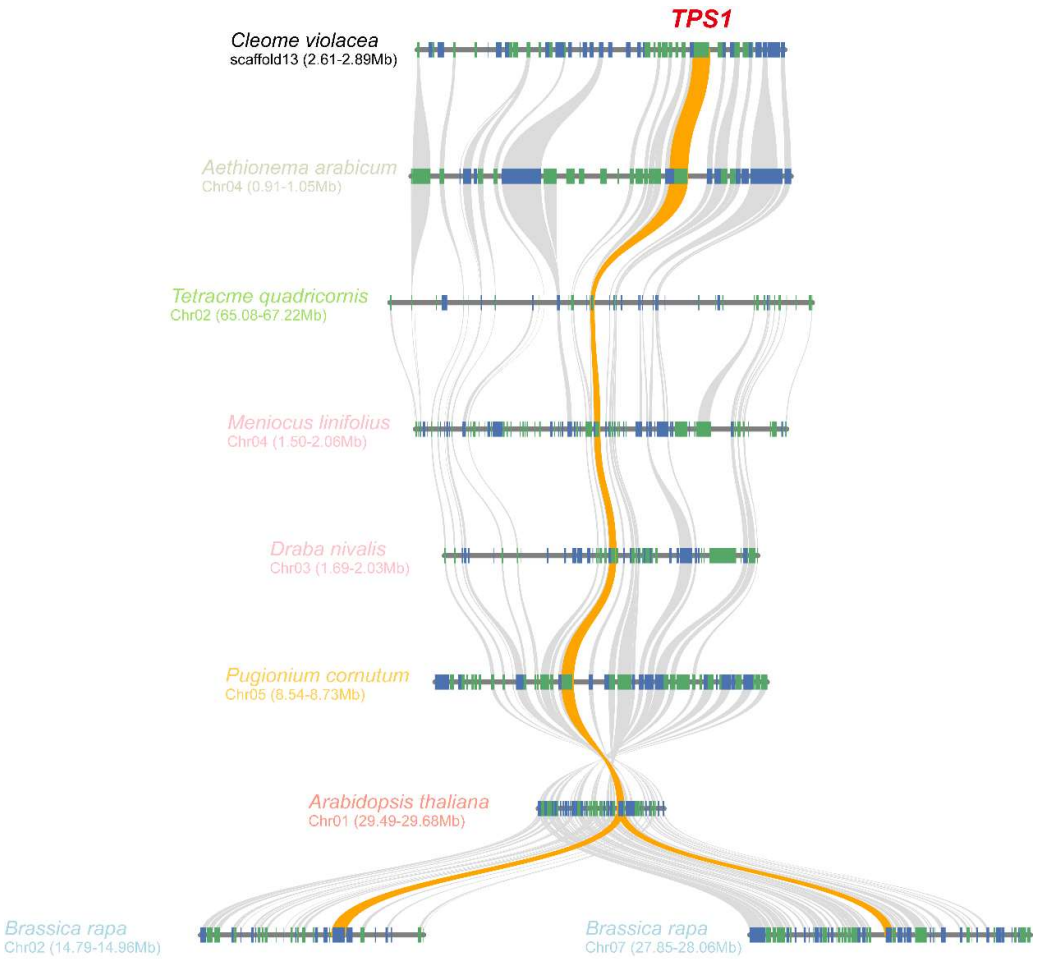
Supplemental Figure 46. Top 20 enriched Gene Ontology (GO) terms of the expanded gene families at node C of Figure 1F.



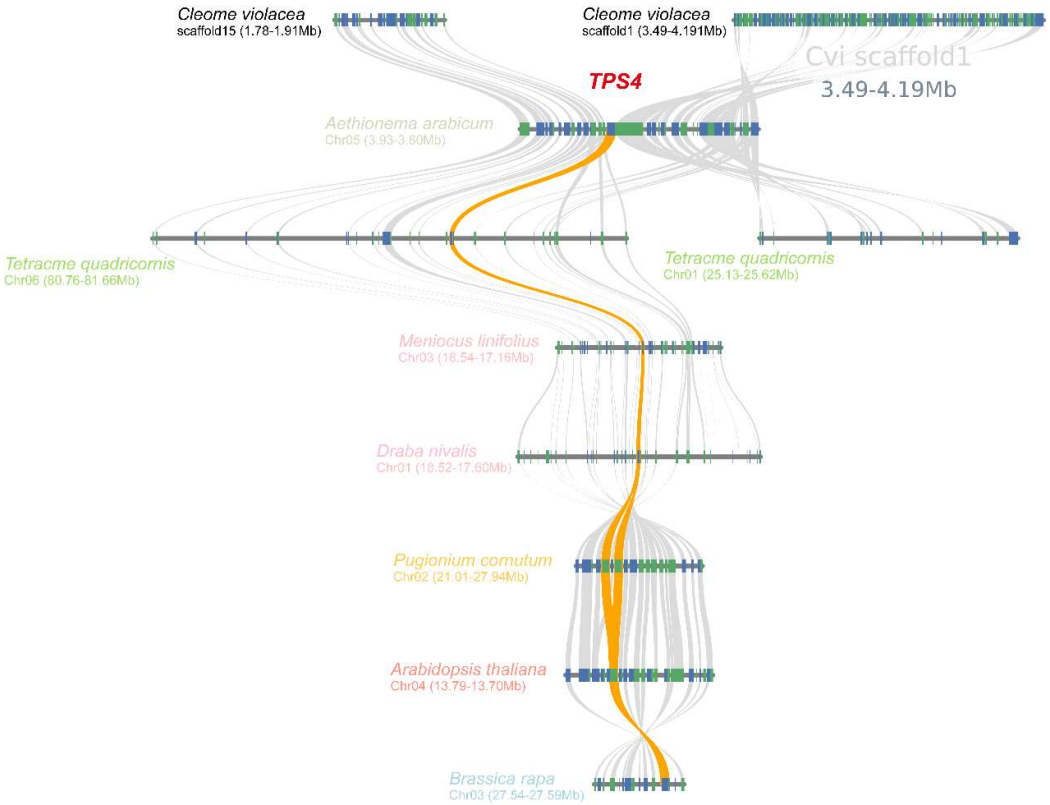
Supplemental Figure 47. Top 20 enriched Gene Ontology (GO) terms of the expanded gene families at node D of Figure 1F.



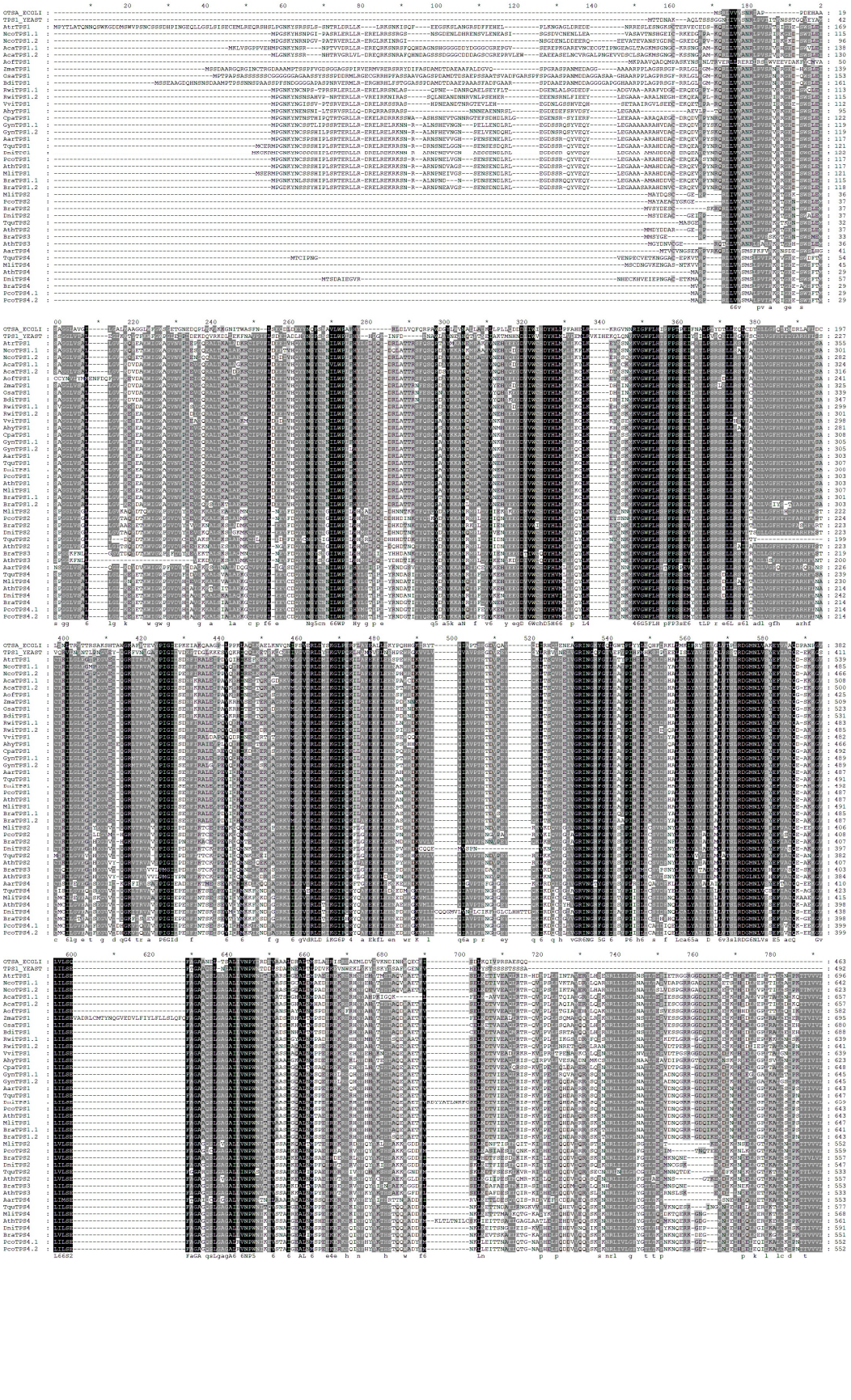
Supplemental Figure 48. Top 20 enriched Gene Ontology (GO) terms of the expanded gene families at node E of Figure 1F.



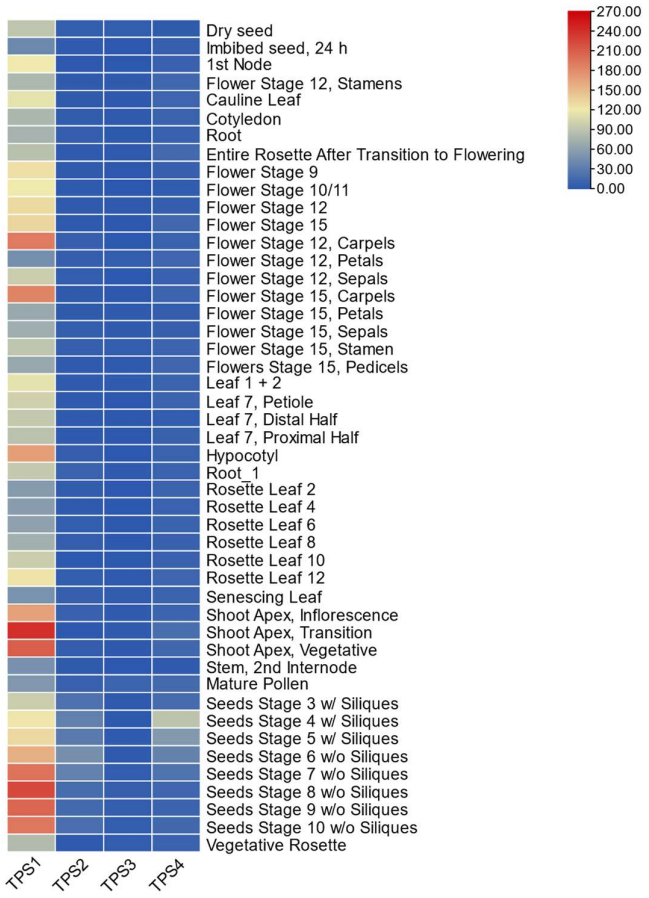
Supplemental Figure 49. The collinearity relationships of subgroups *TPS1* in Brassicaceae and *Cleome violacea* (Cleomaceae).



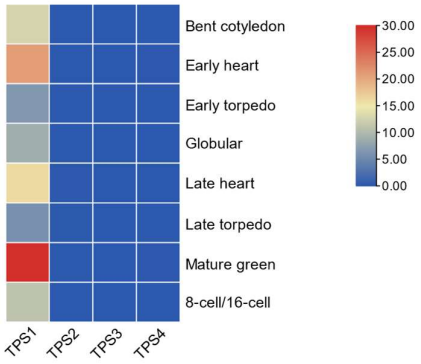
Supplemental Figure 50. The collinearity relationships of subgroups *TPS4* in Brassicaceae and *C. violacea* (Cleomaceae).



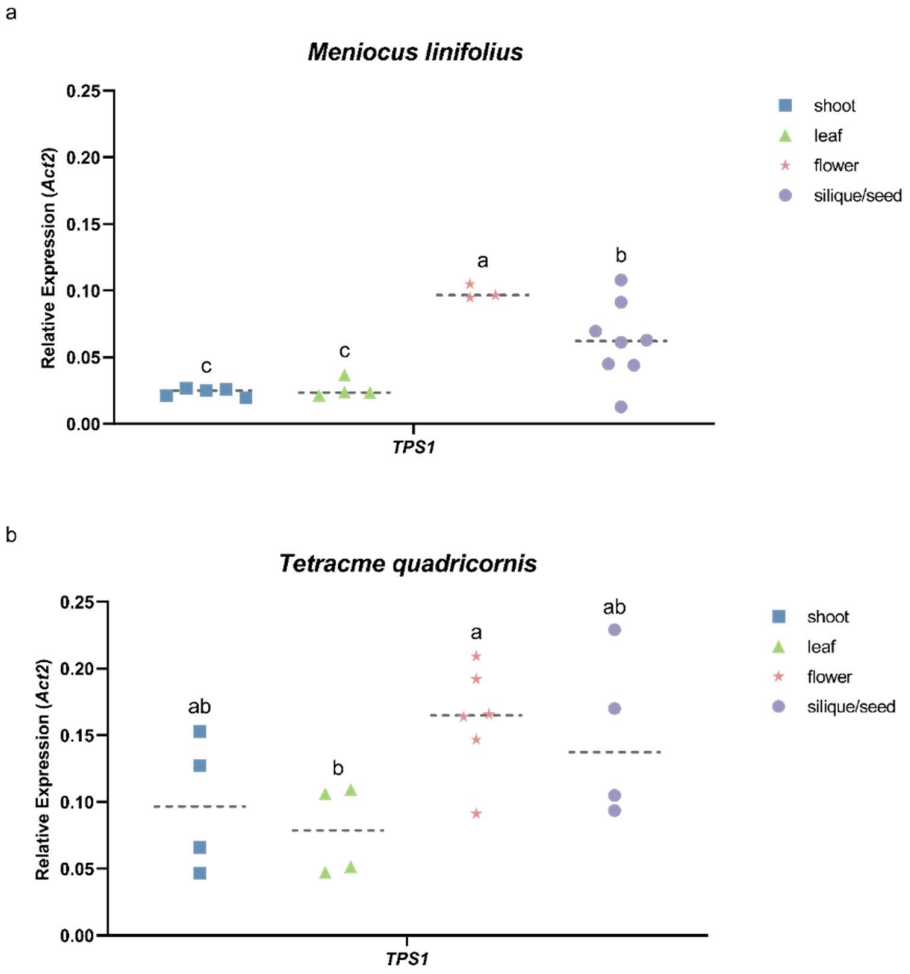
Supplemental Figure 51. Sequence alignments for *TPS1-4* in 21 selected Eukarya species.



Supplemental Figure 52. Gene expression level of *TPS1-4* along development stages in *A. thaliana* (source data: https://bar.utoronto.ca/efp//cgi-bin/efpWeb.cgi).



Supplemental Figure 53. Gene expression level of *TPS1-4* in *A. thaliana* embryo (source data: https://bar.utoronto.ca/efp//cgi-bin/efpWeb.cgi).



Supplemental Figure 54. *TPS1* expression determined with RT-qPCR using materials collected in field conditions in Xinjiang, China for *Mli* (a) and *Tqu* (b).