

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,*}

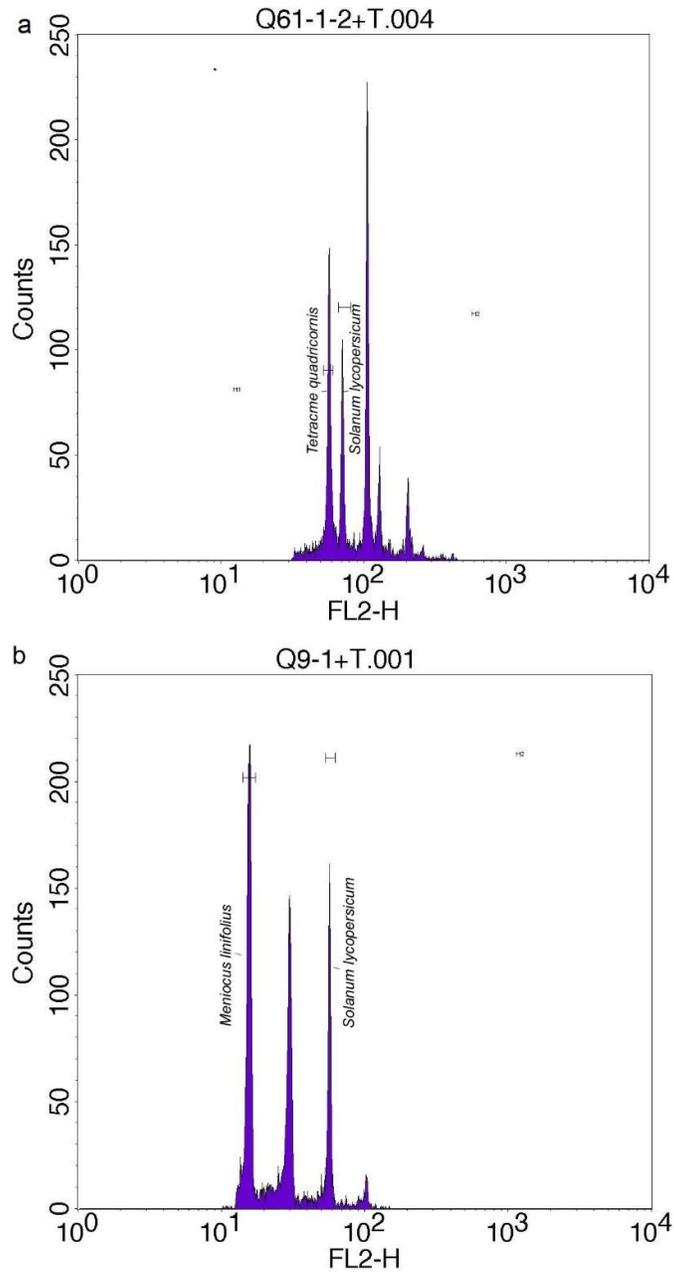
1. CAS Key Laboratory for Plant Diversity and Biogeography of East Asia & Yunnan Key Laboratory of Crop Wild Relatives Omics, Kunming Institute of Botany, Chinese Academy of Sciences, Kunming 650201, China

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

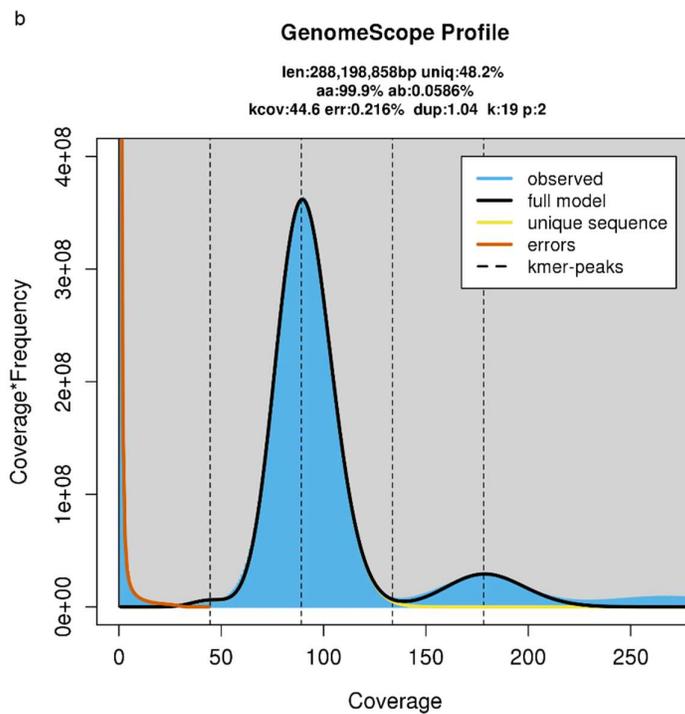
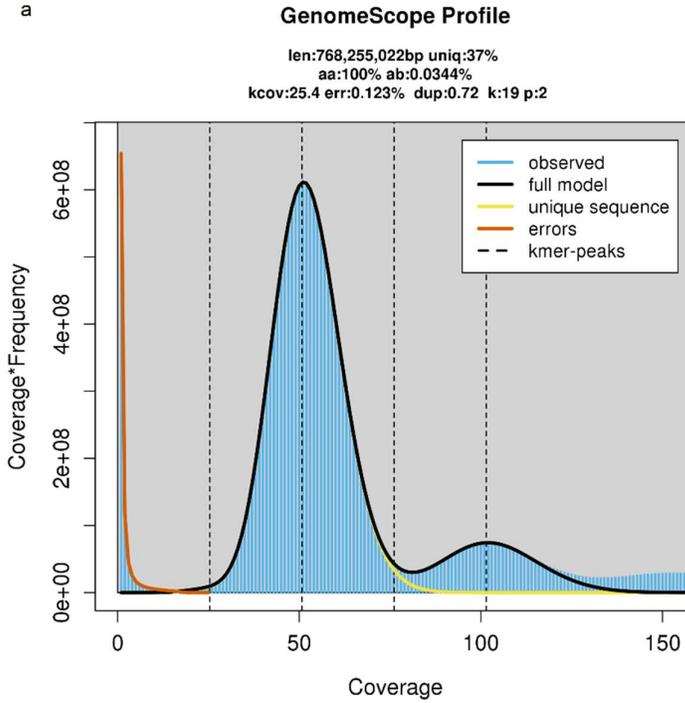
3. College of Life Sciences, Xinjiang Agricultural University, Ürümqi 830052, China

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

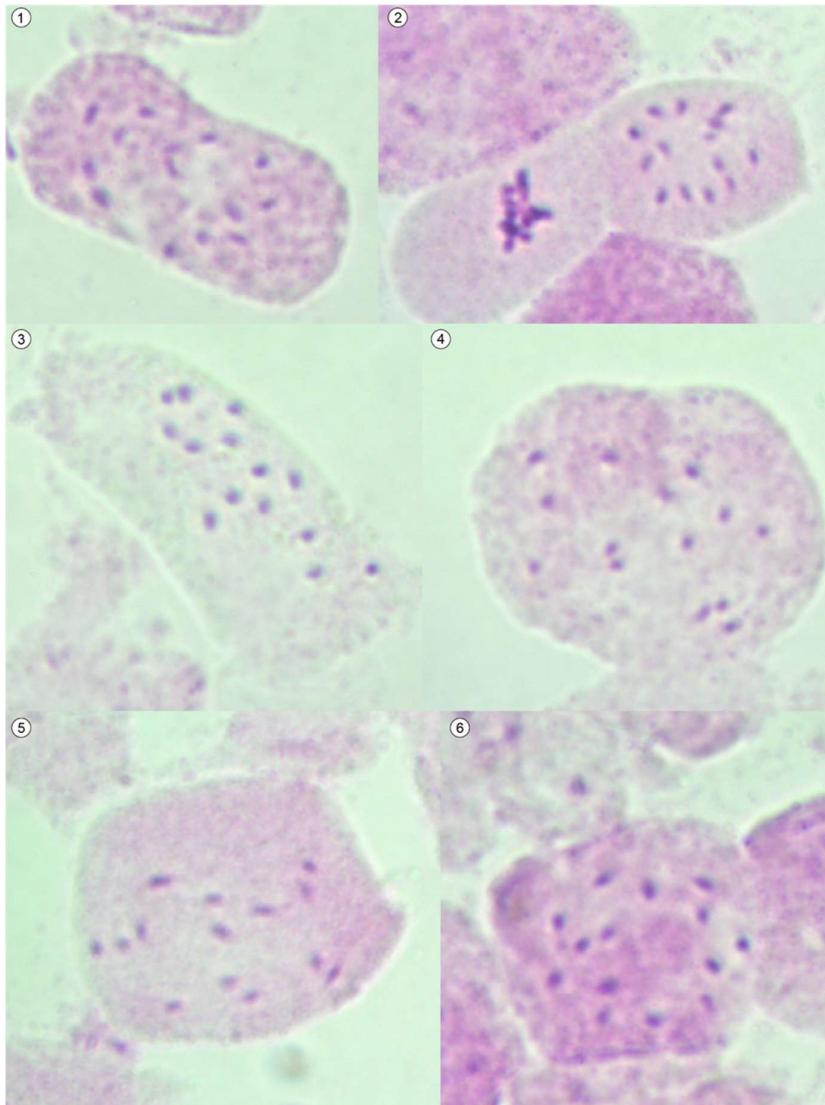
Supplemental Figures 1-54.



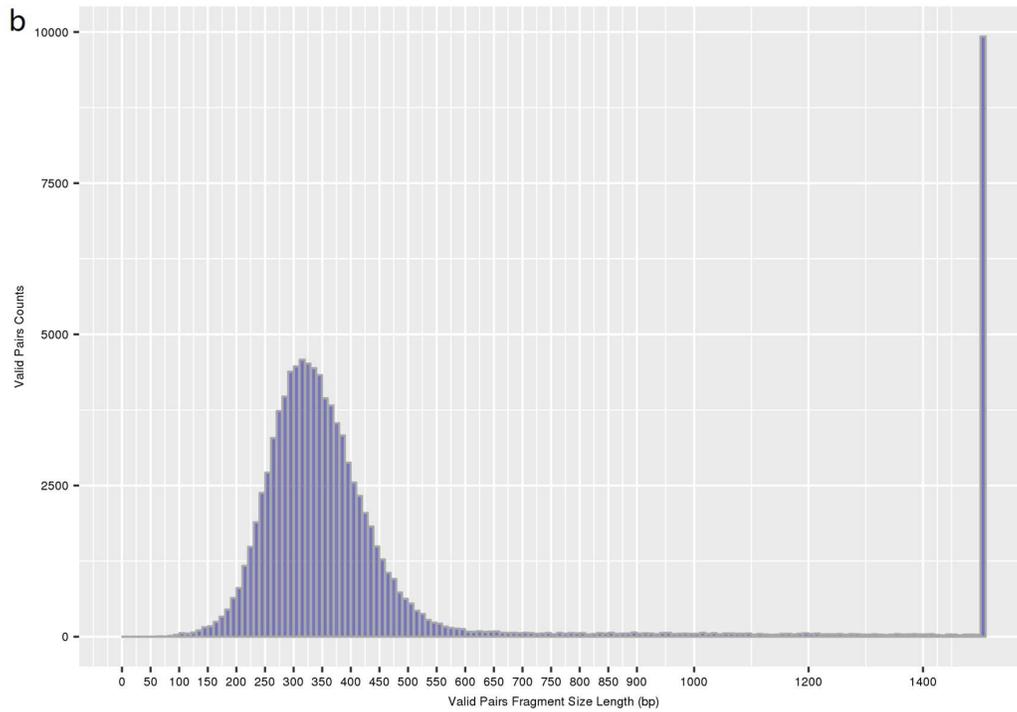
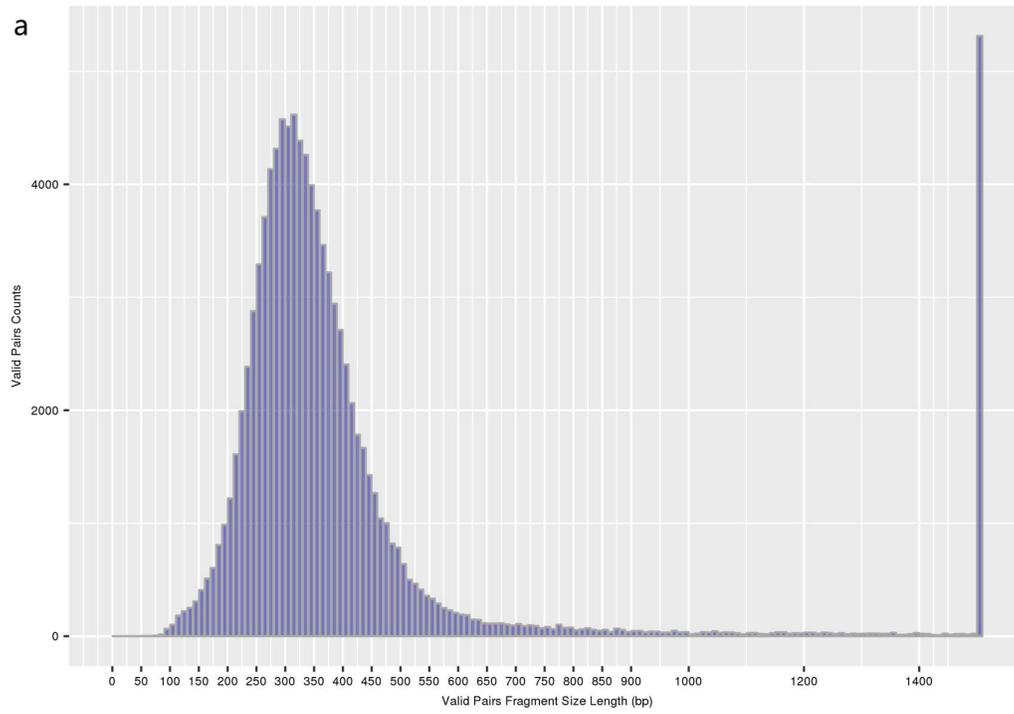
Supplemental Figure 1. Genome sizes of the *Tetracme quadricornis* (a; Q61-1-2; *Tqu*) and *Menicocus linifolius* (b; Q9-1; *Mli*) estimated by flow cytometry with *Solanum lycopersicum* as reference.



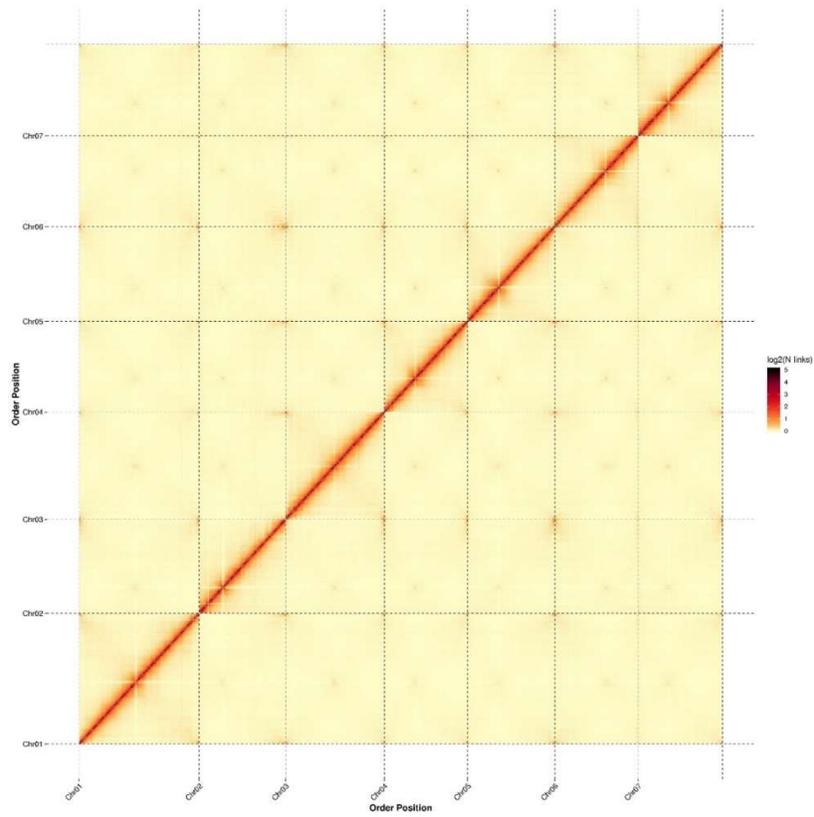
Supplemental Figure 2. Genome size estimation using *GenomeScope2*. 19 K-mer size was used to survey both *Tqu* (a) and *Mli* (b). X-axis shows the depth of read coverage while Y-axis is the frequency at certain depth divided by the total frequency of all depths. Estimated genome size, heterozygosity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



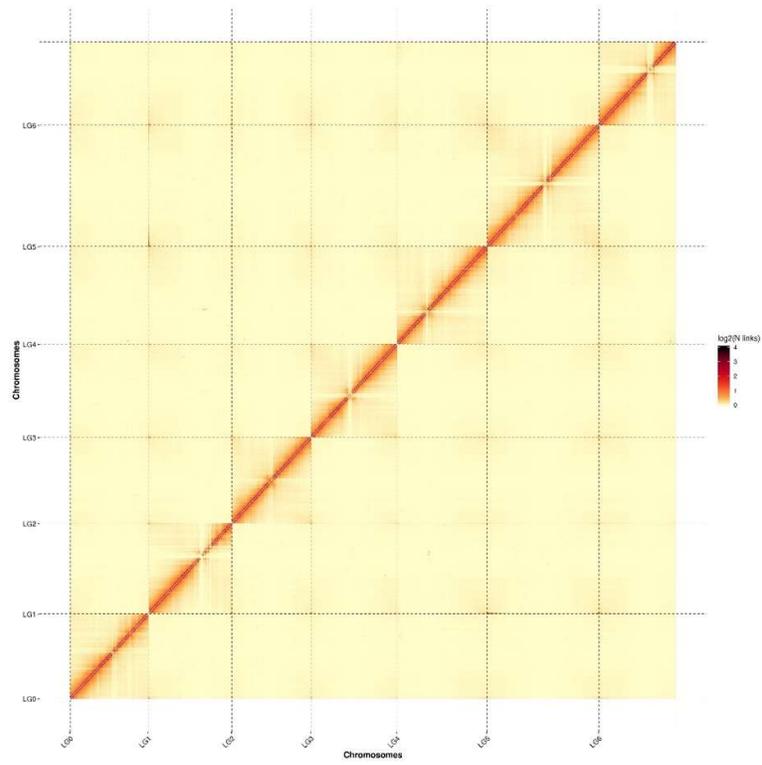
Supplemental Figure 3. Cytological analyses of *M. linifolius* karyotypes.



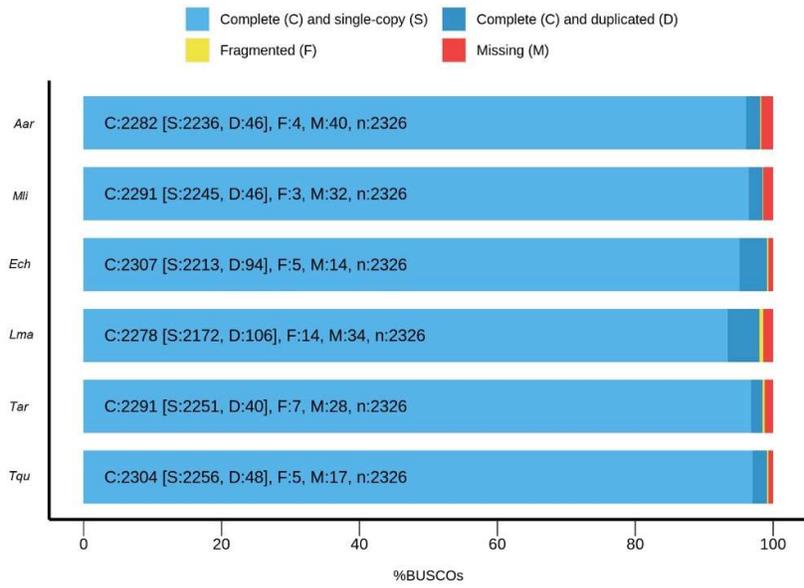
Supplemental Figure 4. Distribution insert Hi-C reads for *Tqu* (a) and *Mli* (b). X-axis is the sum of the distance between the paired end reads on the assembled genome, while Y-axis shows the number of 100000 pair reads randomly selected in insert fragments of different lengths.



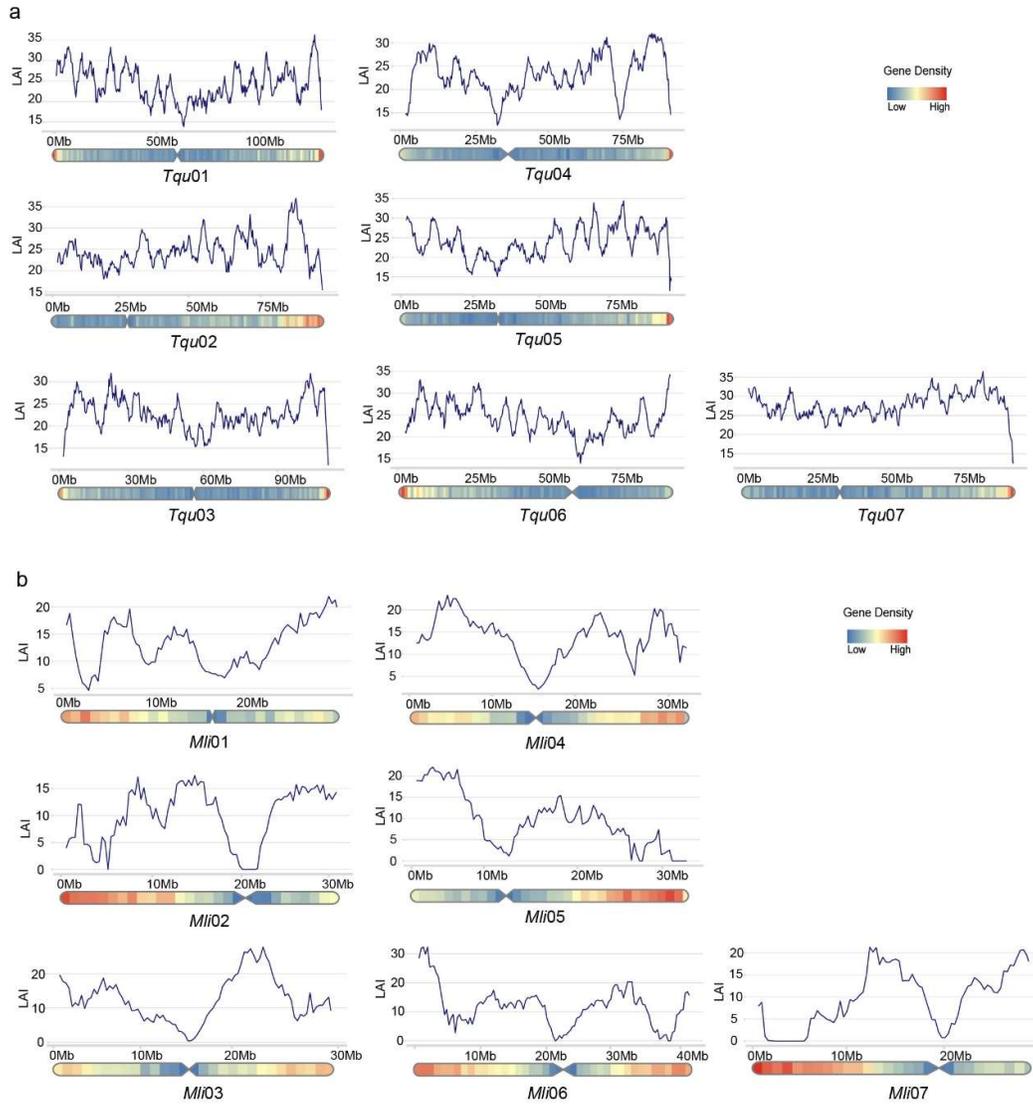
Supplemental Figure 5. Distribution of interaction frequency for Hi-C links among chromosomes of *Tqu* at 100 kb resolution. The X- and Y- axes present the order of the positions of scaffolds on the corresponding pseudochromosomes (the same for Supplemental Figure 6).



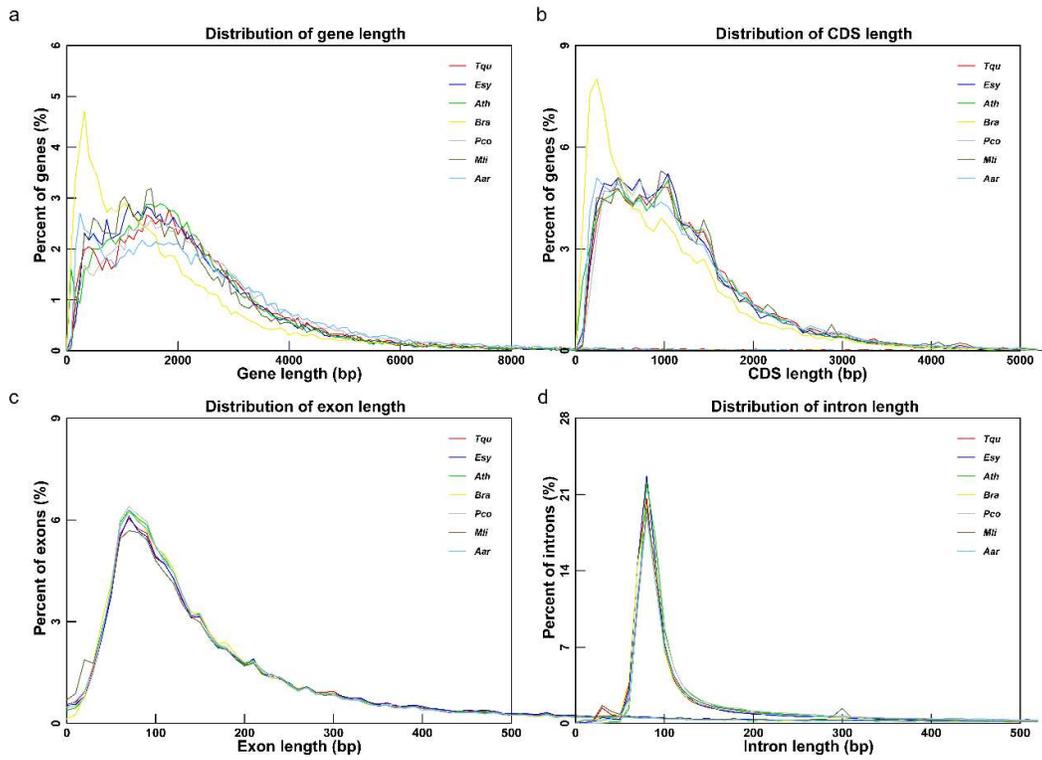
Supplemental Figure 6. Distribution of interaction frequency for Hi-C links among chromosomes of *Mli* at 100 kb resolution.



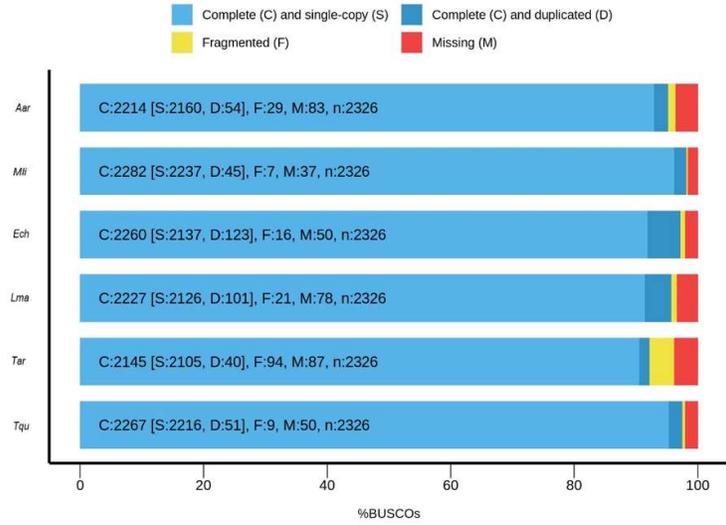
Supplemental Figure 7. BUSCO assessments (Eudicots_odb10) at the genome level for six representative species. *Aar*, *Aethionema arabicum*; *Ech*, *Erysimum cheiranthoides*; *Lma*, *Lobularia maritima*; *Tar*, *Thlaspi arvense*.



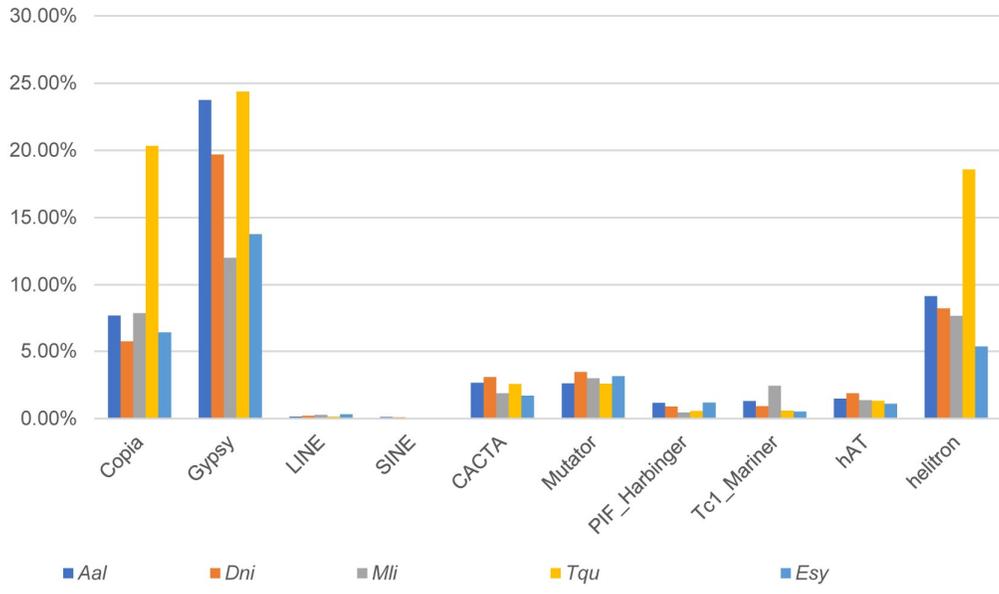
Supplemental Figure 8. Distribution of the LTR Assembly Index (LAI) along the chromosomes for *Tqu* (a) and *Mli* (b).



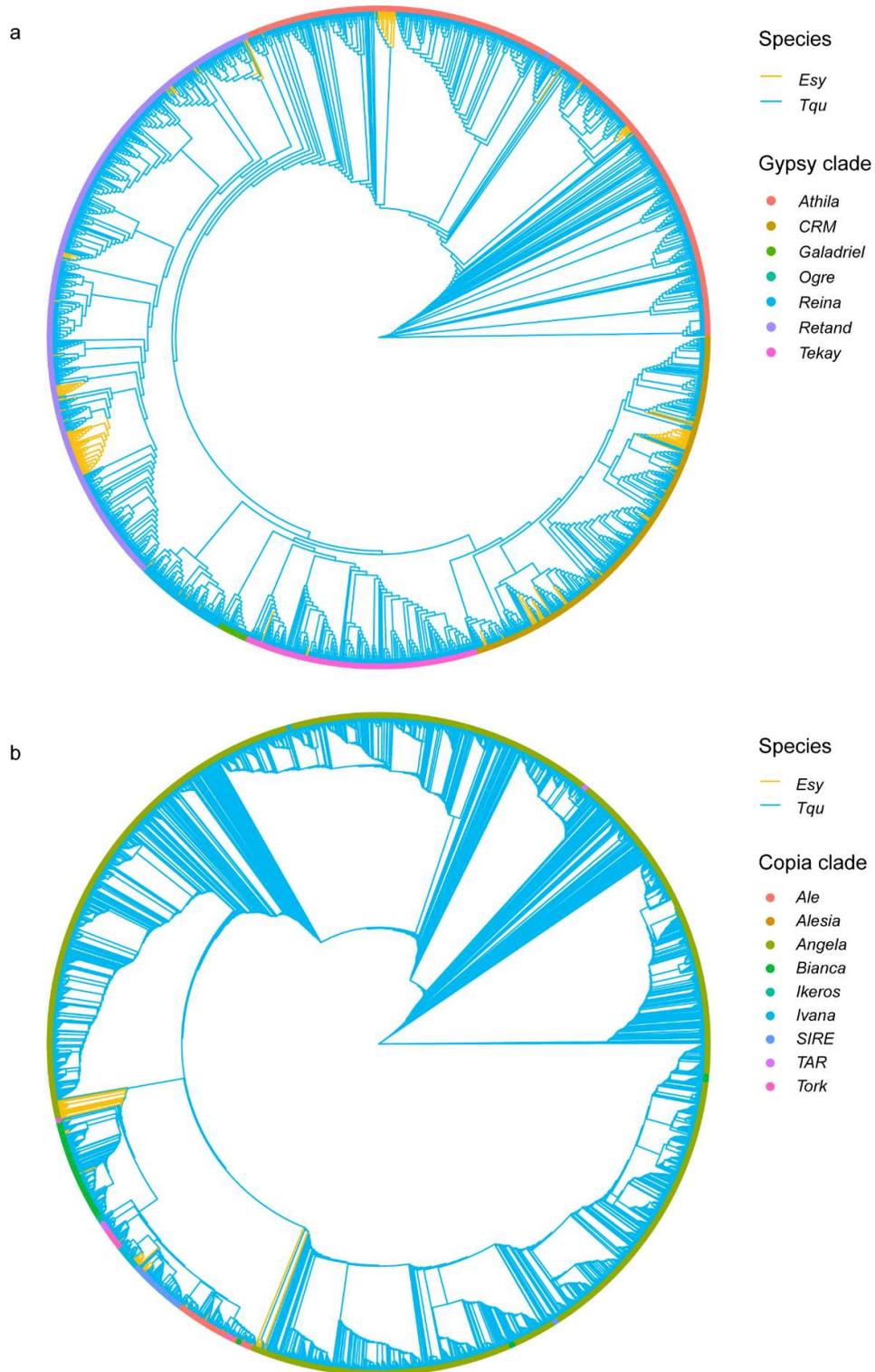
Supplemental Figure 9. Gene structure features for selected Brassicaceae species. **a**, gene length; **b**, coding length; **c**, exon length; **d**, intron length.



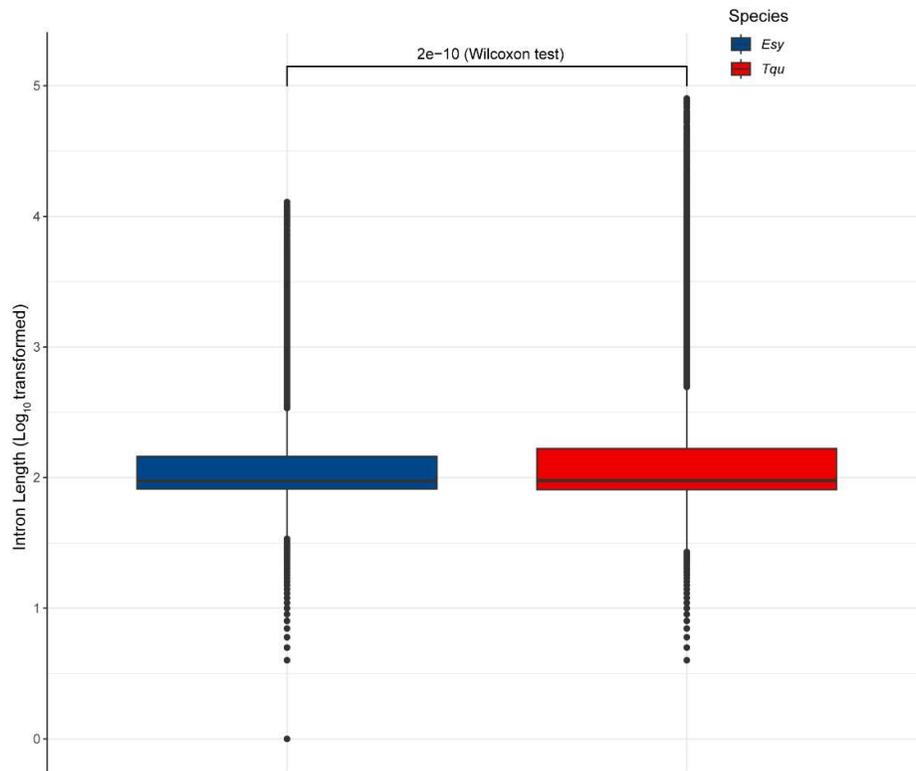
Supplemental Figure 10. BUSCO assessment (Eudicots_odb10) at the gene level for six representative species.



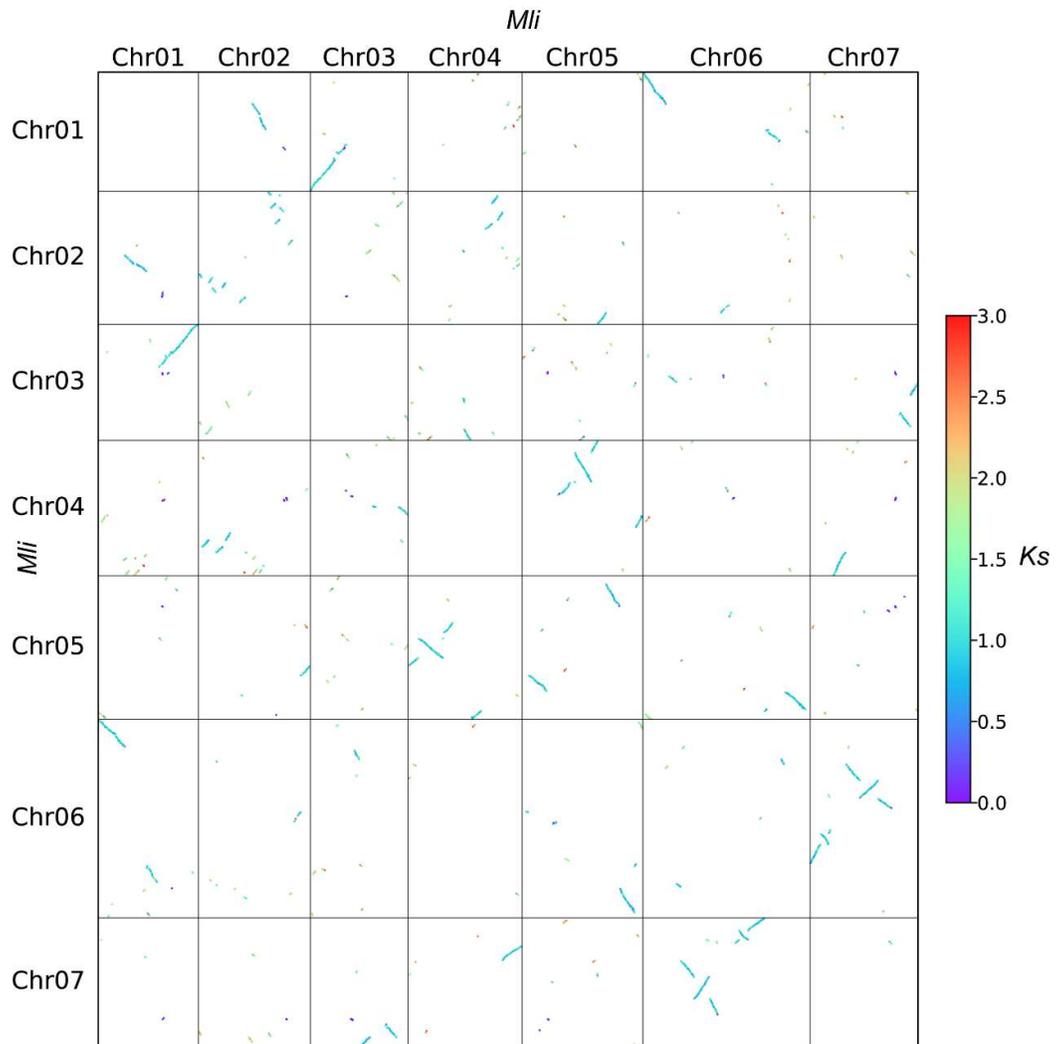
Supplemental Figure 11. Proportion of different repeat types in five genomes.



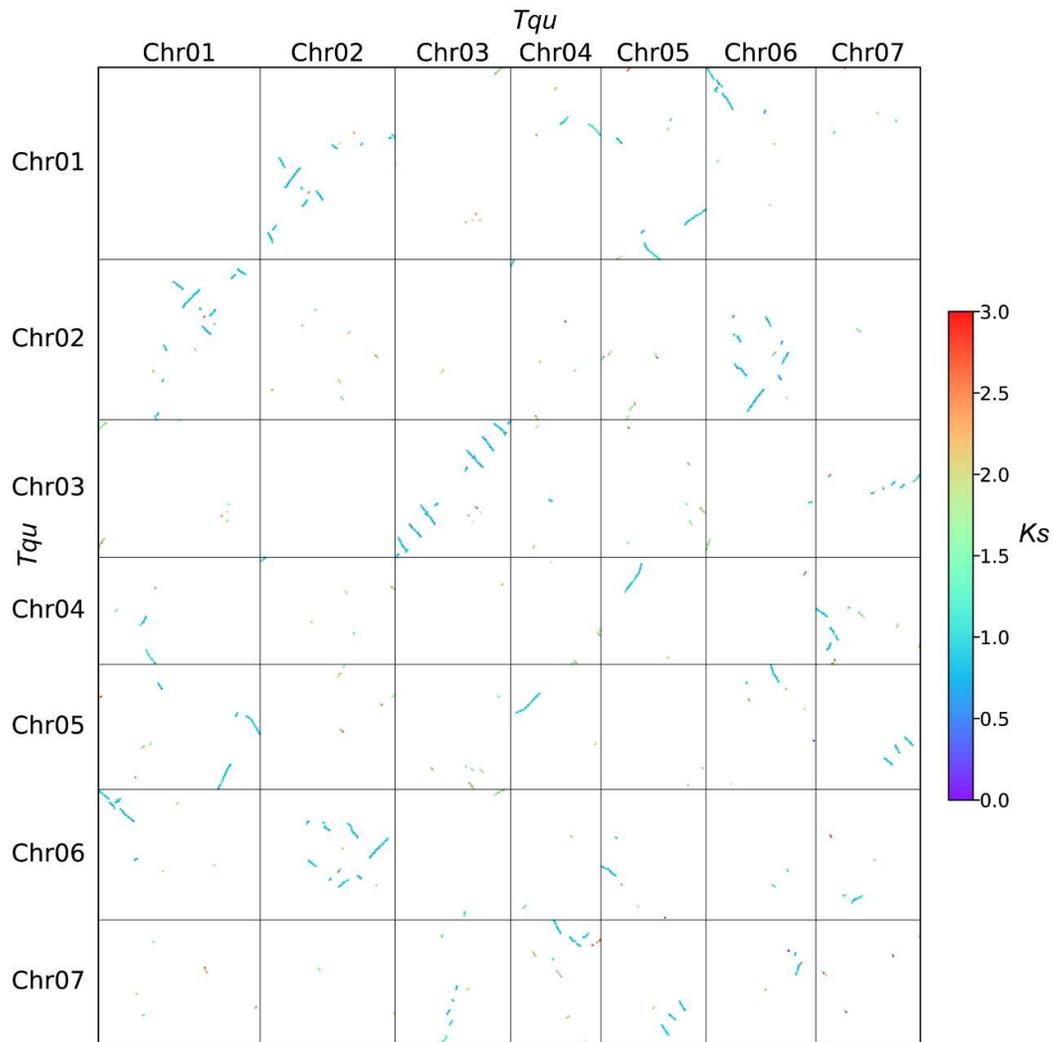
Supplemental Figure 12. Phylogenetic clustering of *Ty3/Gypsy* (a) and *Ty1/Copia* (b) retrotransposons in *Tqu* and *E. syriacum* (*Esy*).



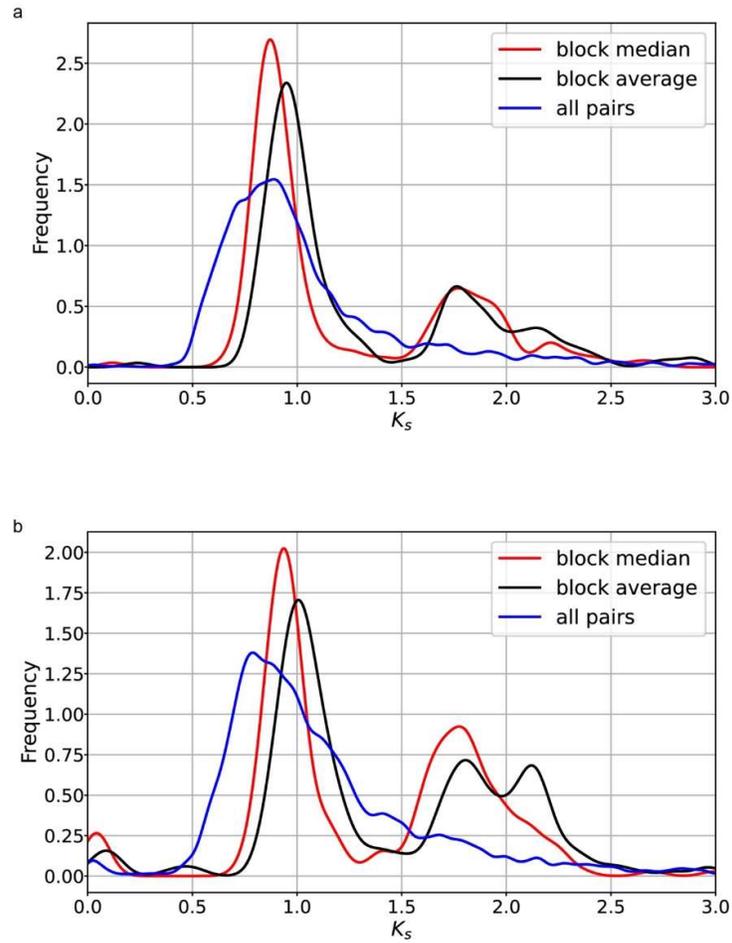
Supplemental Figure 13. The gene intron length distribution of *Tqu* and *Esy*.



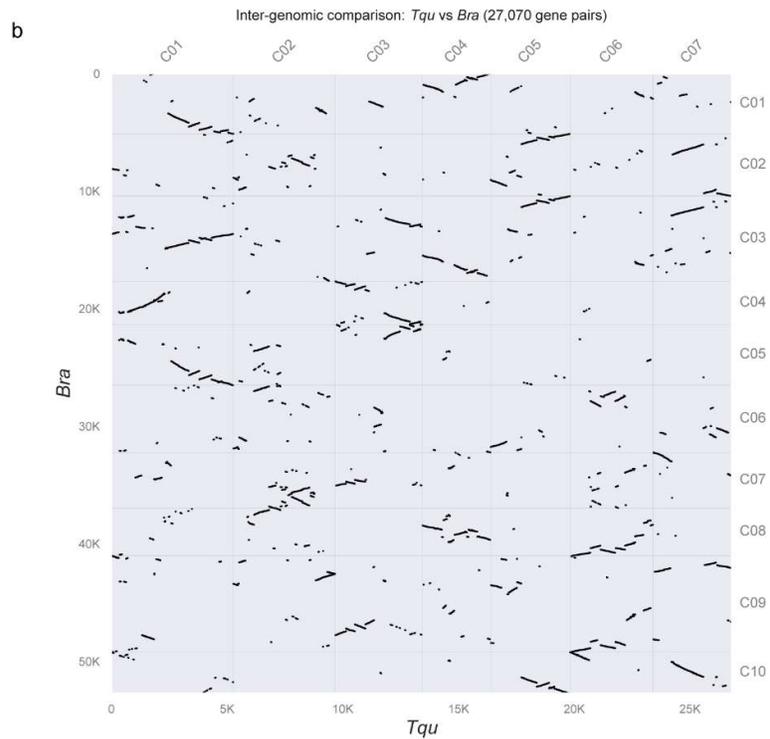
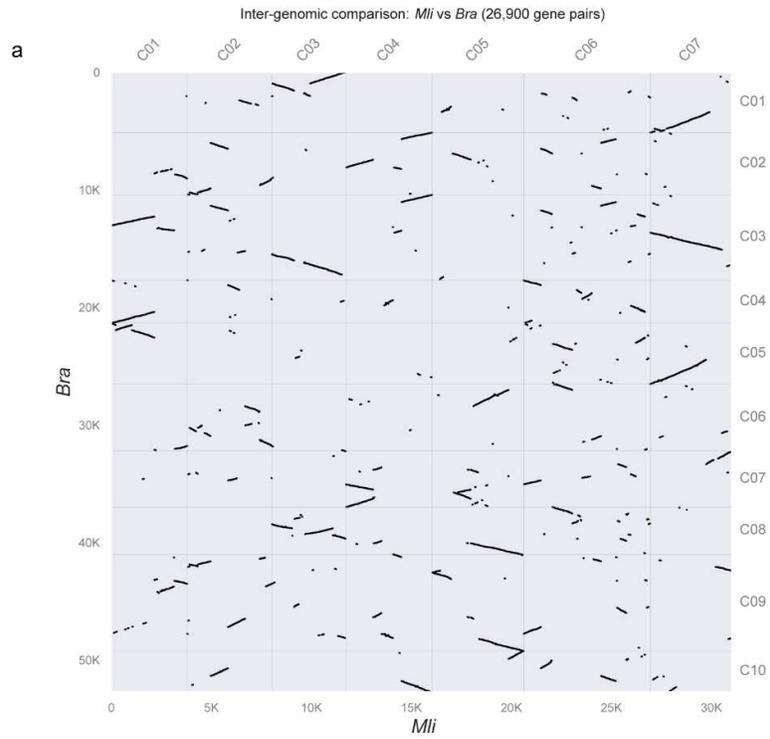
Supplemental Figure 14. Syntenic relationships within *Mli*. Syntenic blocks were colored based on the *Ks* values per syntenic gene pair between syntenic blocks of each genome.



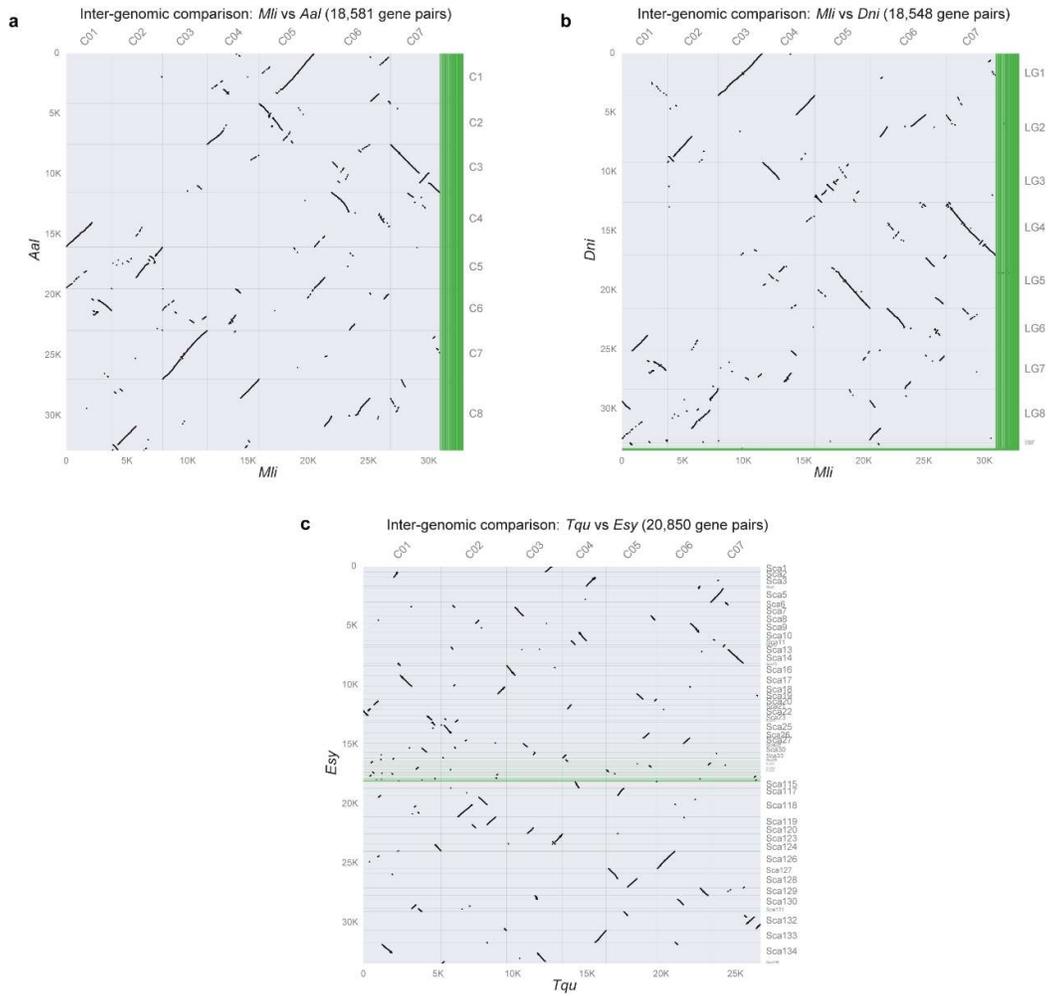
Supplemental Figure 15. Syntenic relationships within *Tqu*.



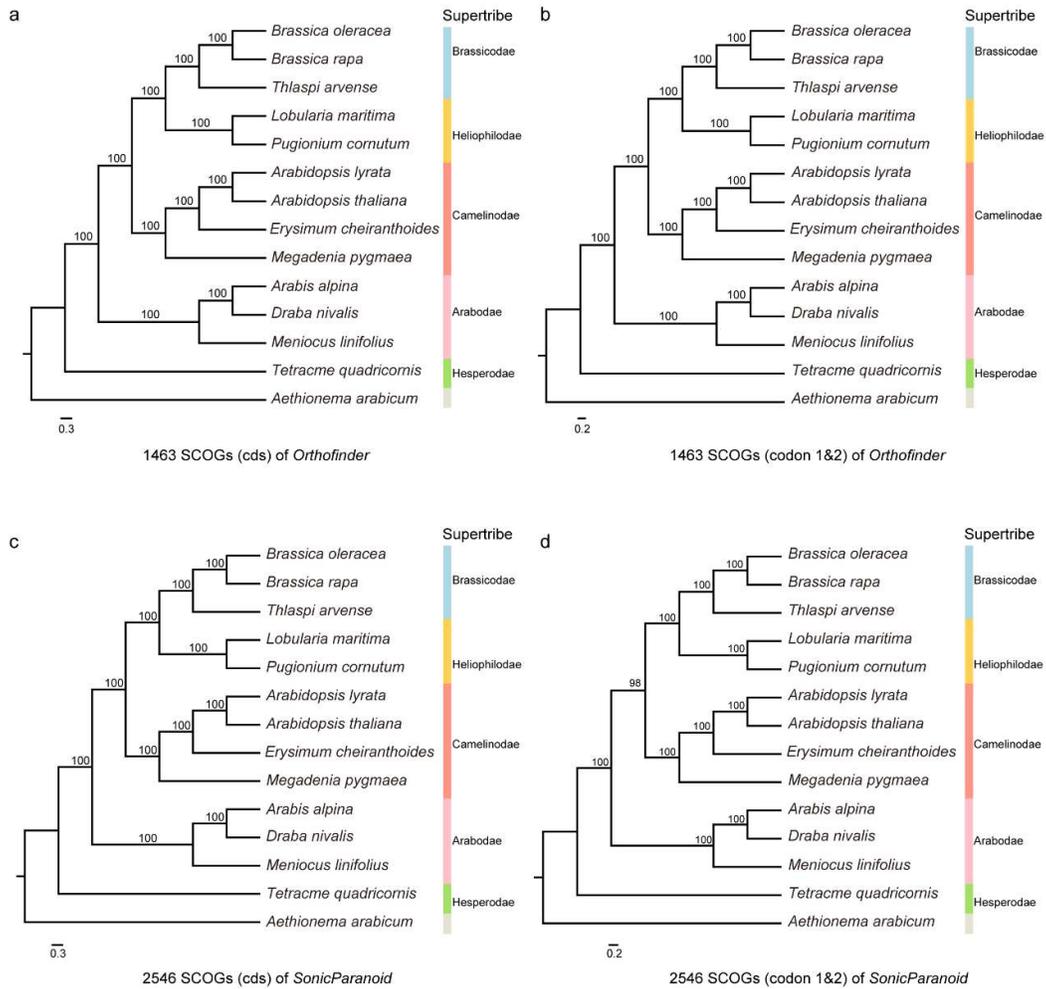
Supplemental Figure 16. Distribution of K_s of intragenomic syntenic blocks for *Tqu* (a) and *Mli* (b). Block median and average values were estimated based on the median and average K_s for each collinear block.



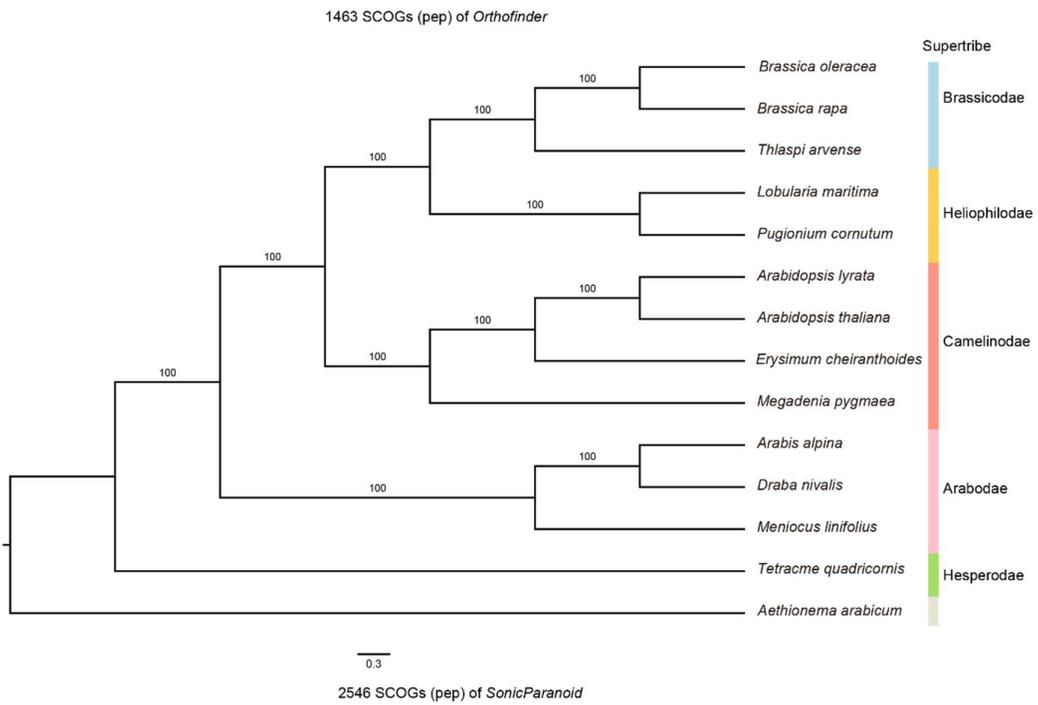
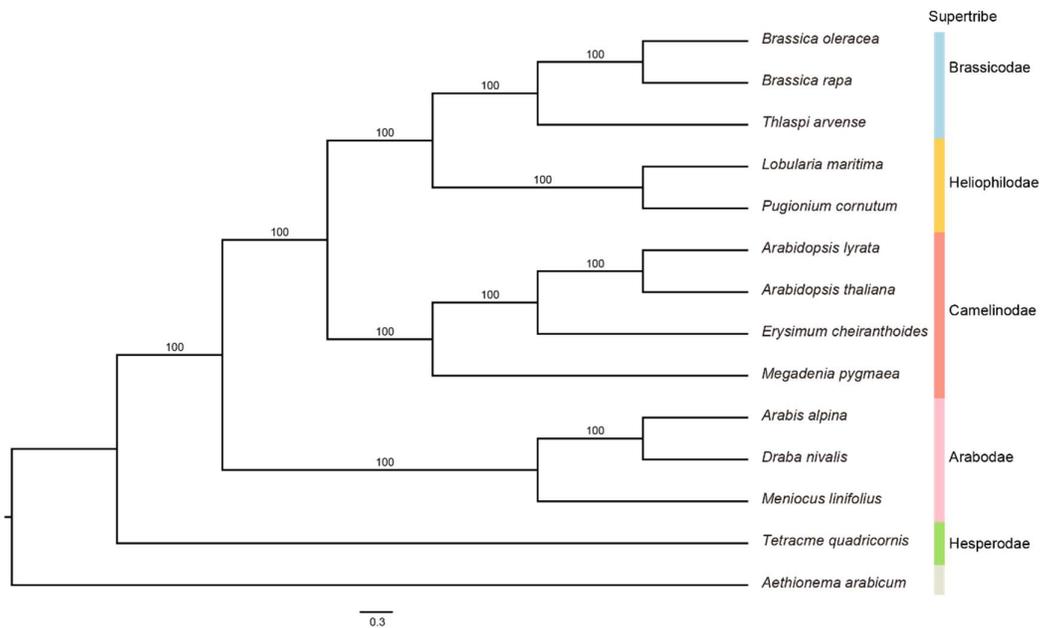
Supplemental Figure 17. Genomic synteny comparisons between *B. rapa* and *Tqu* (a) or *Mli* (b).



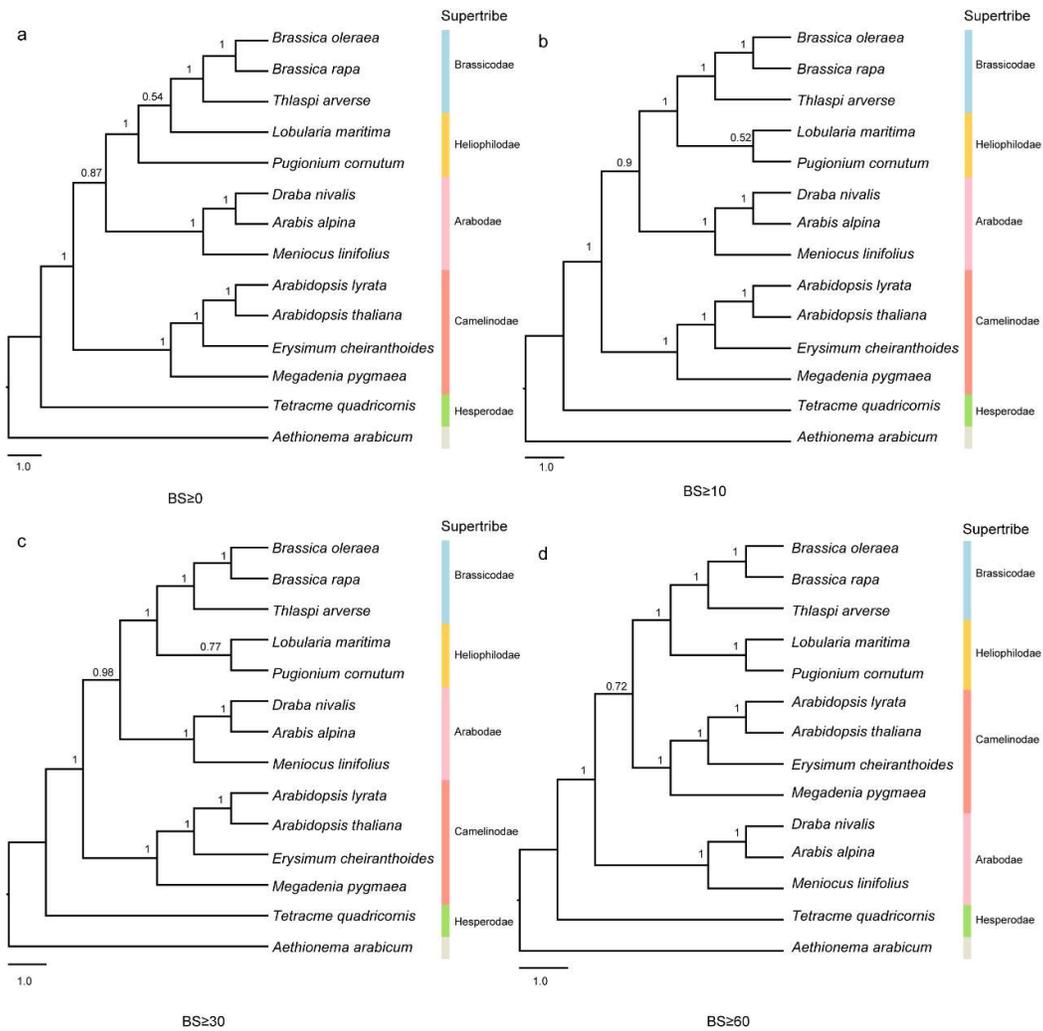
Supplemental Figure 18. Synteny comparisons of *A. alpina* (*Aal*) to *Mli* (a), *D. nivalis* (*Dni*) to *Mli* (b), and *Esy* to *Tqu* (c).



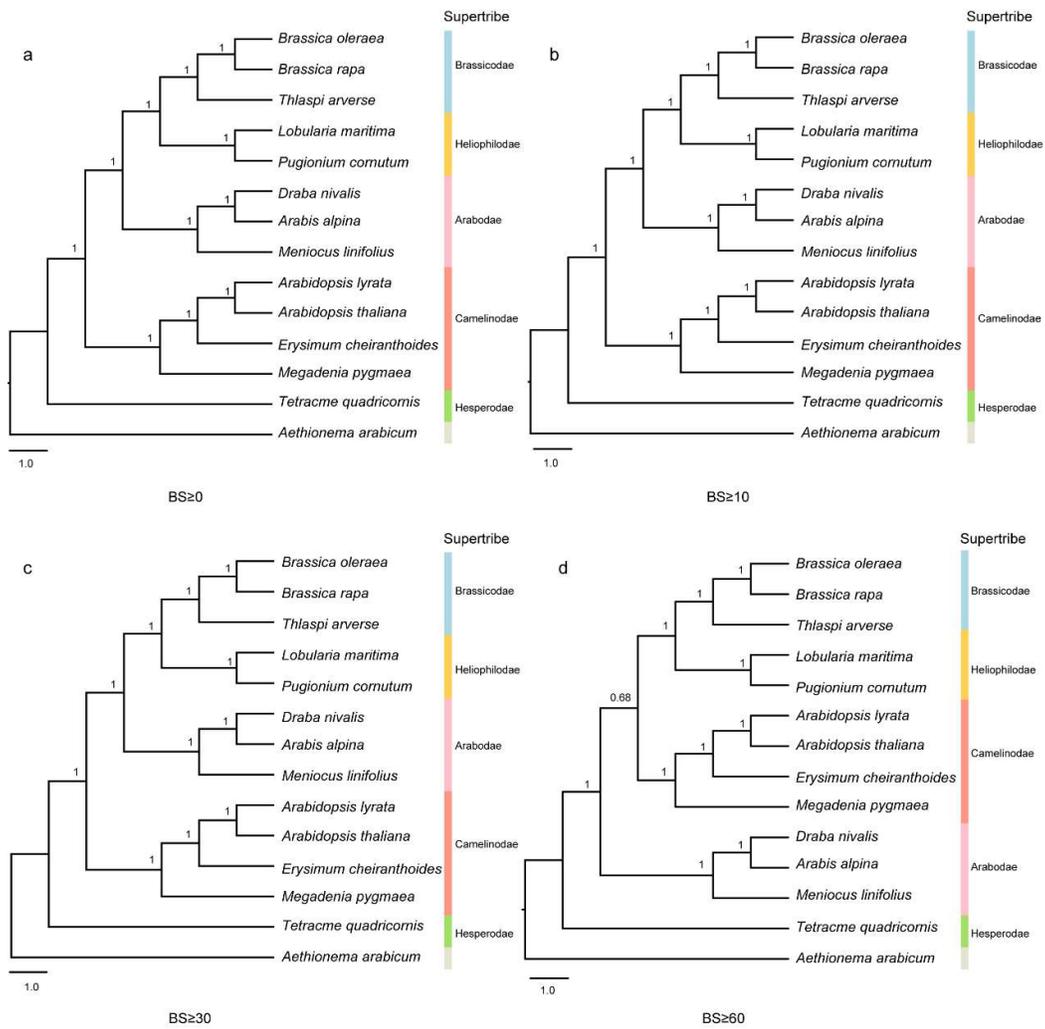
Supplemental Figure 19. Maximum likelihood (ML) phylogenies based on concatenated nucleotide sequences using *Orthofinder* (a for CDS and b for codons 1 & 2) and *SonicParanoid* (c for CDS and d for codons 1 & 2) with bootstrap (BS) supports shown.



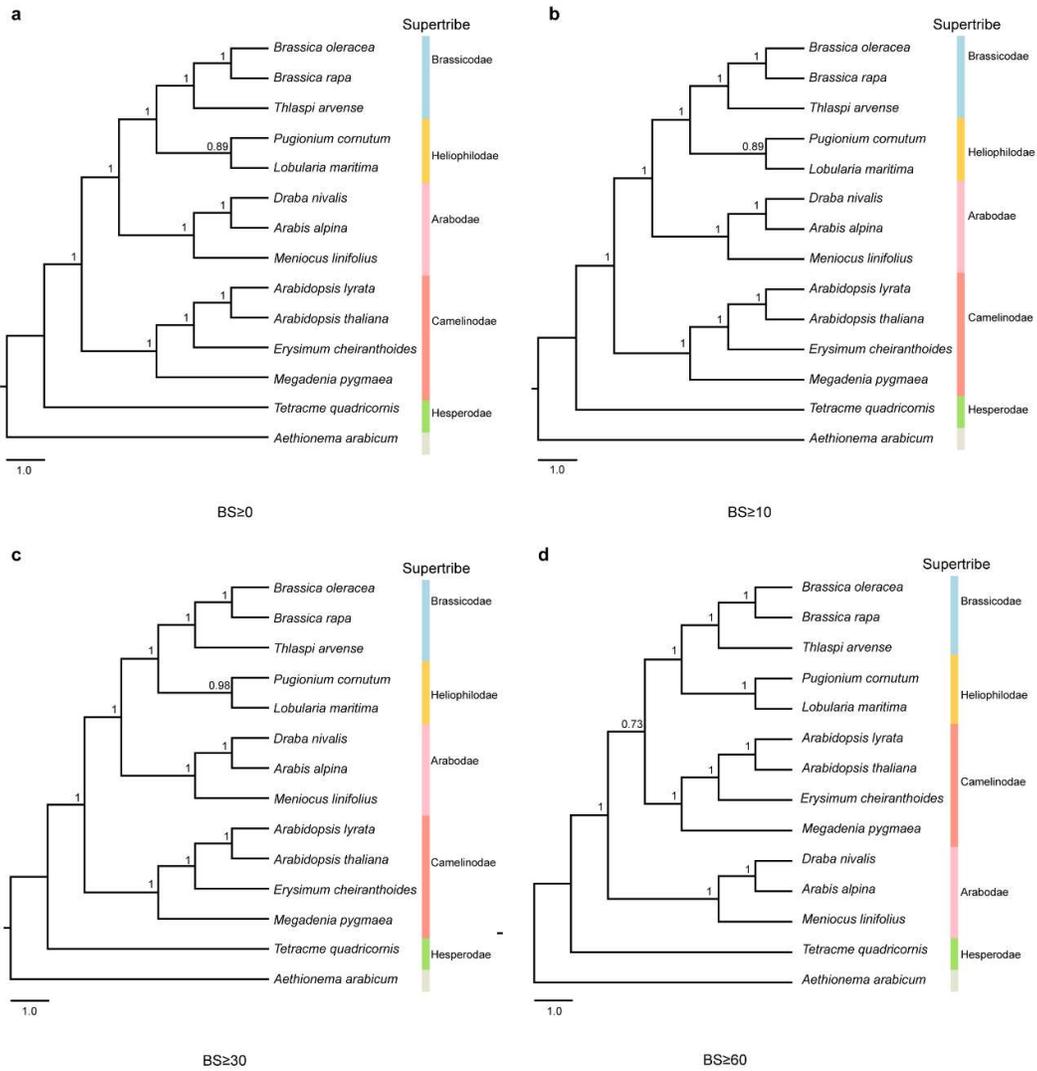
Supplemental Figure 20. Maximum likelihood (ML) phylogenies based on concatenated amino acids using *Orthofinder* (upper) and *SonicParanoid* (lower) with BS supports shown.



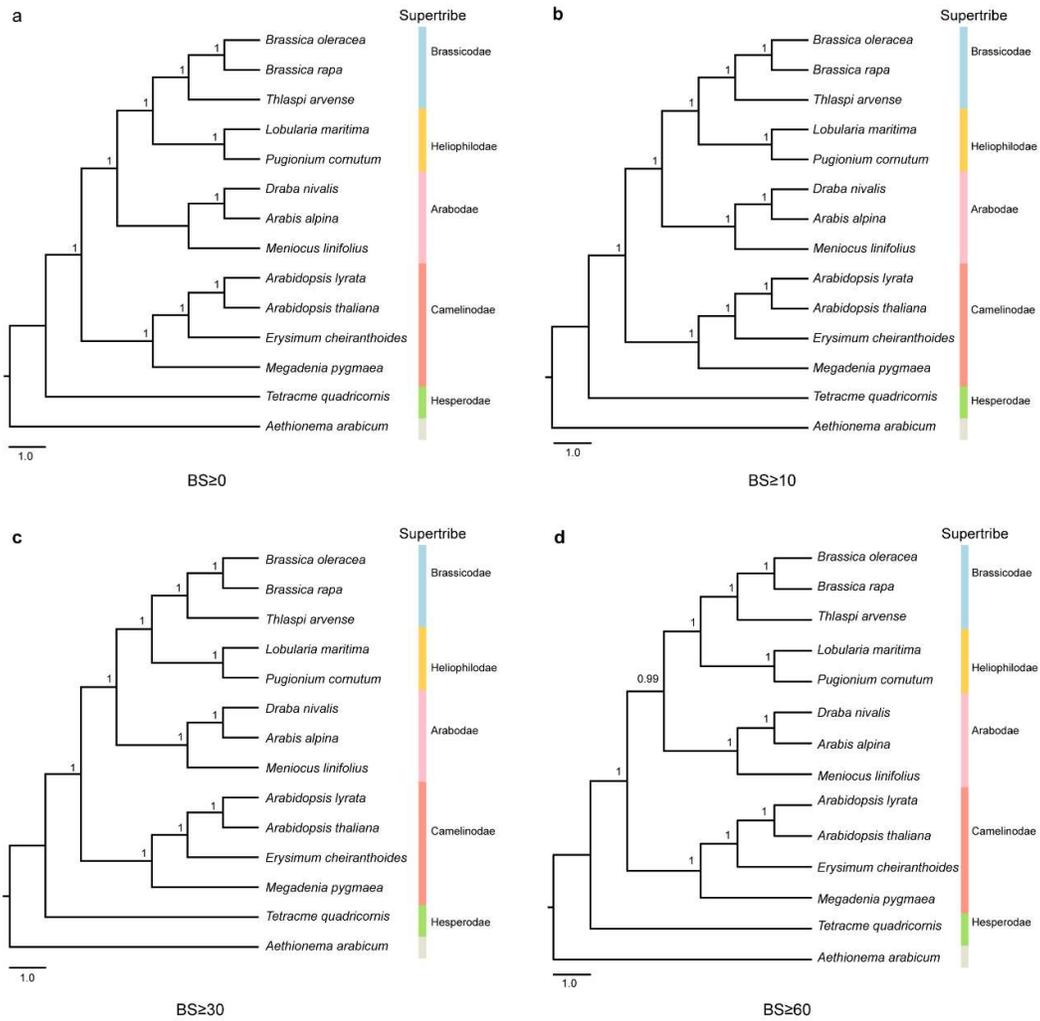
Supplemental Figure 21. Effect of BS values of the 1463 input individual gene trees for coalescent-based phylogenetic analyses using coding sequences. a, BS ≥ 0; b, BS ≥ 10; BS ≥ 30; BS ≥ 60. Phylogenetic trees inferred based on 1463 individual gene trees with CD S detected through *orthofinder*. Local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



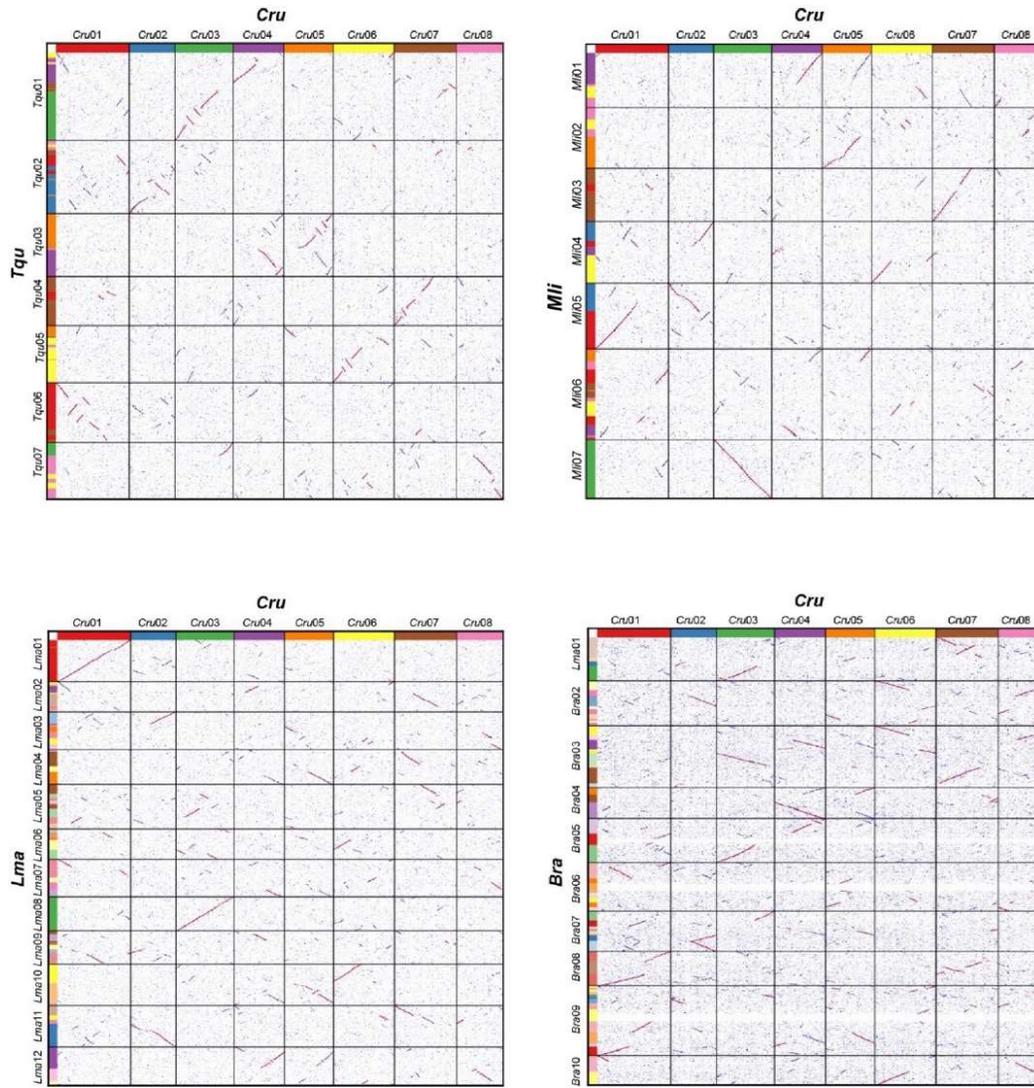
Supplemental Figure 22. Effect of BS values of the 1463 input individual gene trees for coalescent-based phylogenetic analyses using amino acid sequences. A, $BS \geq 0$; b, $BS \geq 10$; $BS \geq 30$; $BS \geq 60$. Phylogenetic trees inferred based on 1463 individual gene trees with amino acid sequences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



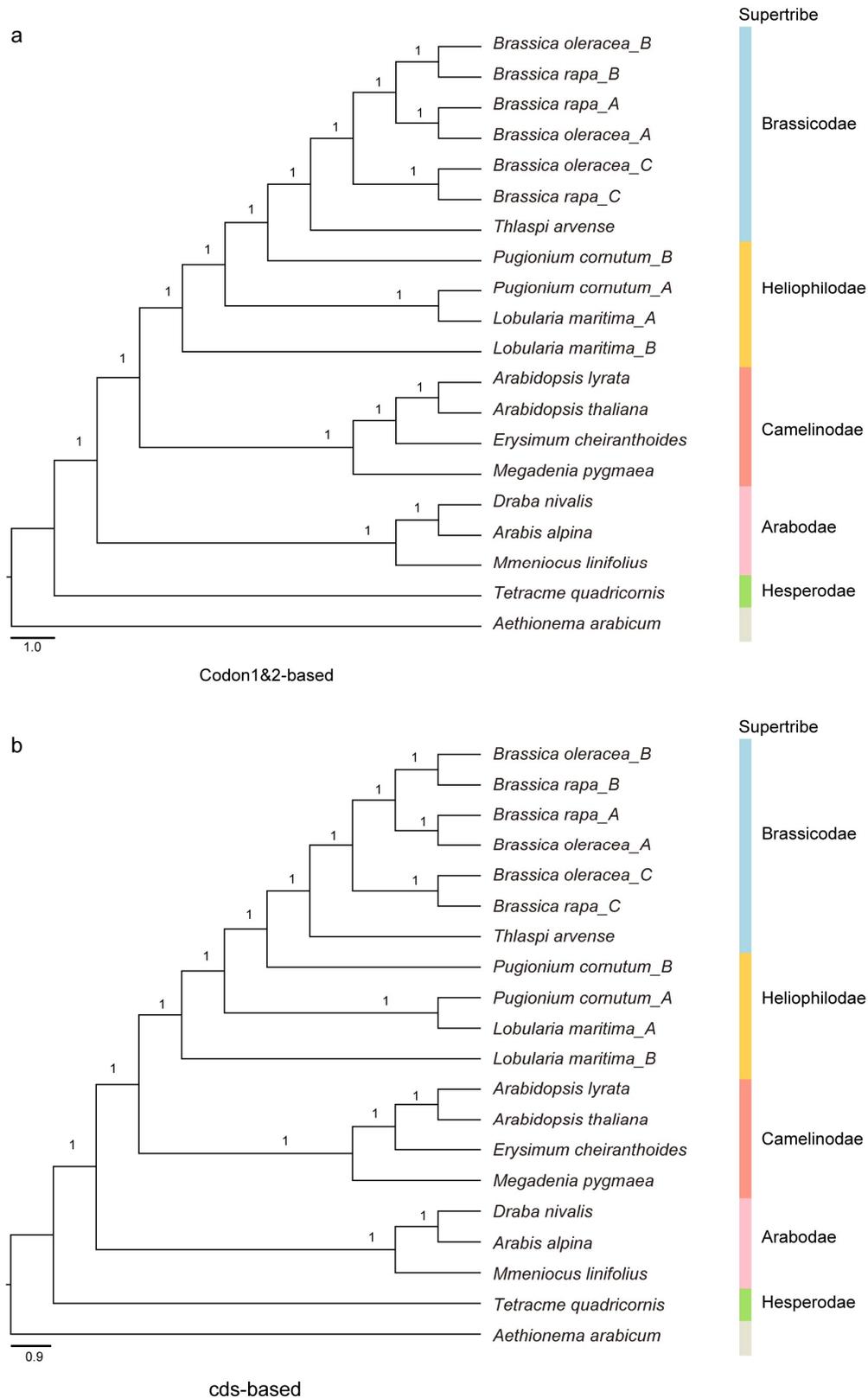
Supplemental Figure 23. Effect of BS values of the 2546 input individual gene trees for coalescent-based phylogenetic analyses using coding sequences. A, $BS \geq 0$; b, $BS \geq 10$; $BS \geq 30$; $BS \geq 60$. Phylogenetic trees inferred based on 2546 individual gene trees with CDS detected through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



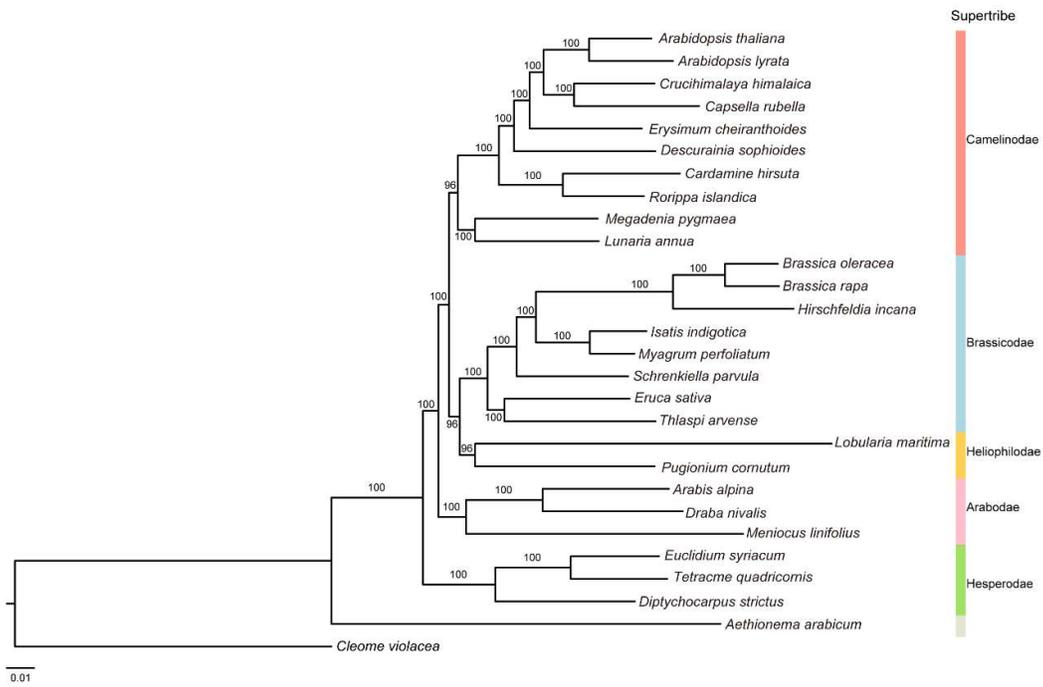
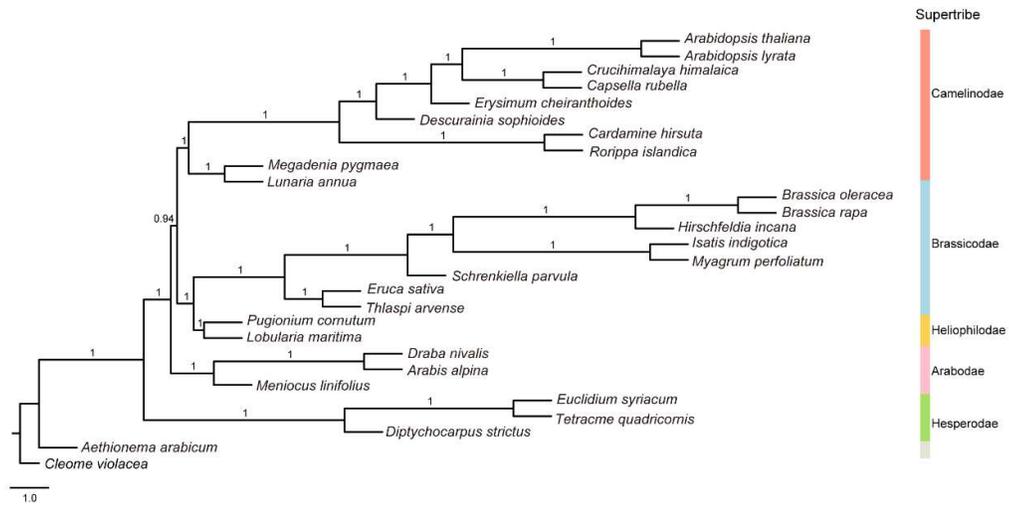
Supplemental Figure 24. Effect of BS values of the 2546 input individual gene trees for coalescent-based phylogenetic analyses using amino acid sequences. A, BS \geq 0; b, BS \geq 10; BS \geq 30; BS \geq 60. Phylogenetic trees inferred based on 2546 individual gene trees with amino acid sequences detected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



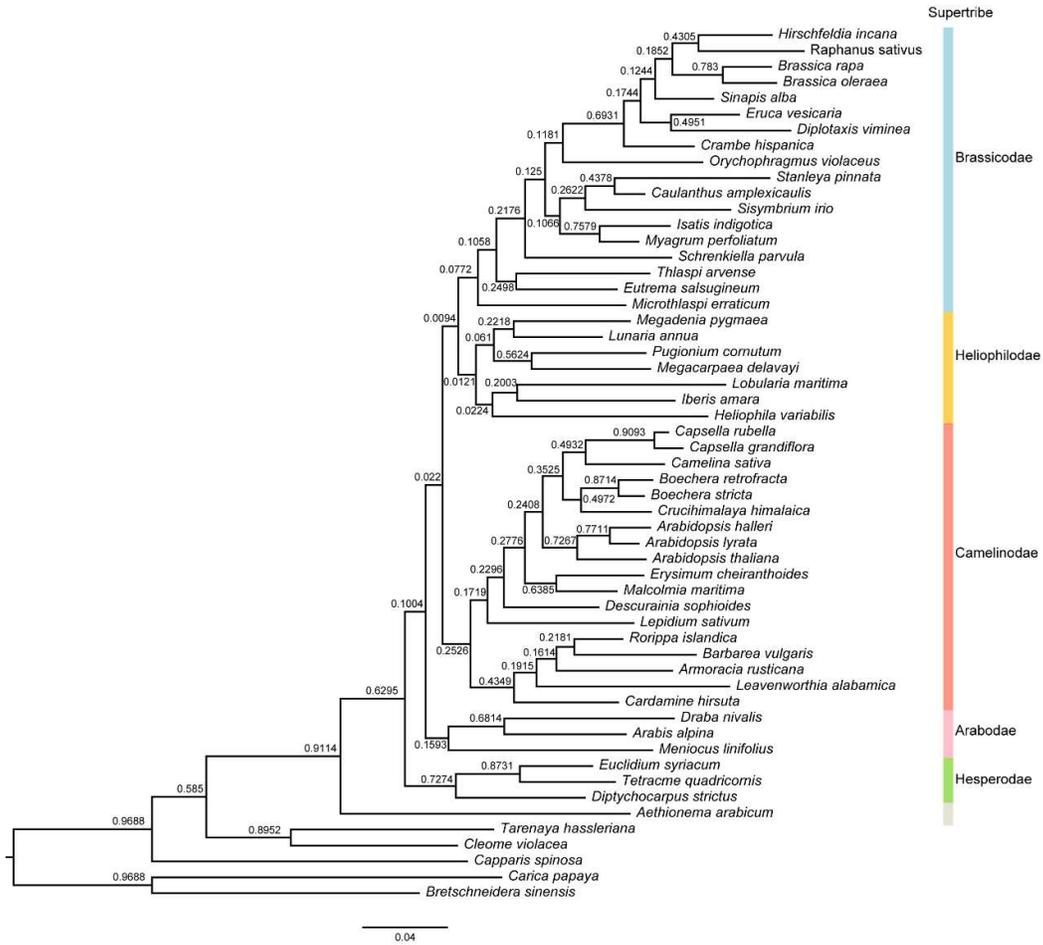
Supplemental Figure 25. Genomic synteny between *Capsella rubella* (Cru) and four Brassicaceae species.



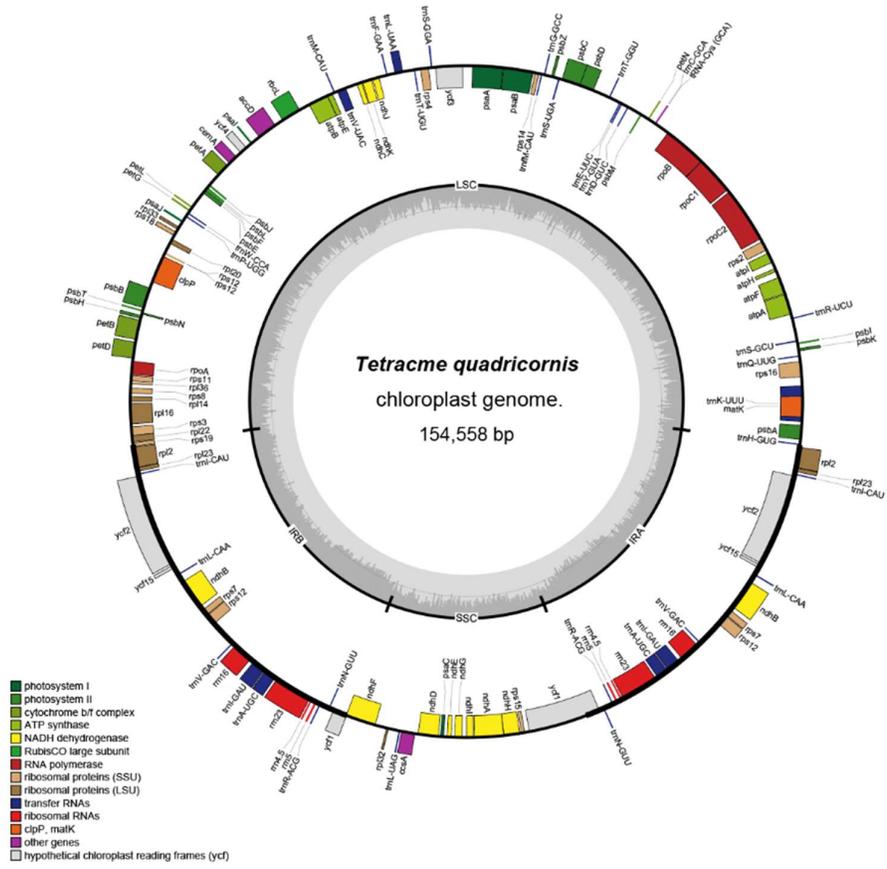
Supplemental Figure 26. Synteny-based coalescent trees using 4434 collinear genes (a, codon 1&2 based; b, CDS based) with LPP values shown.



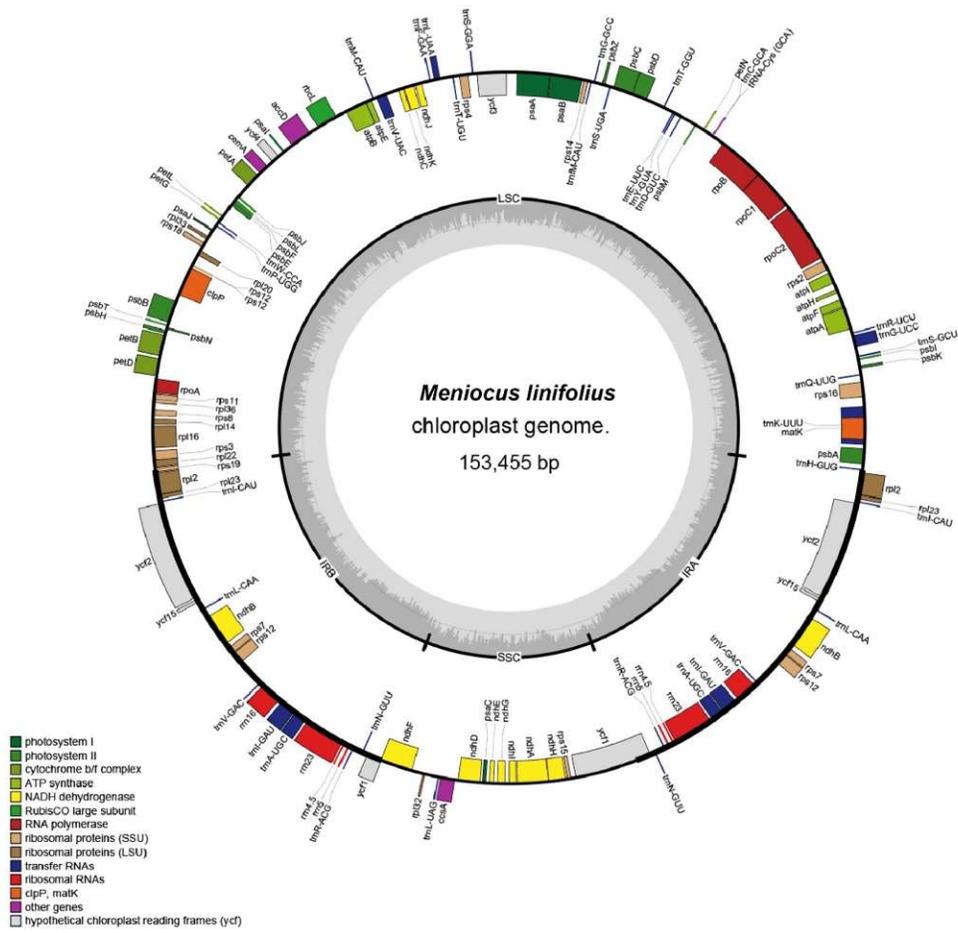
Supplemental Figure 27. Phylogenies using 1092 single copy orthologues for 28 species (a, coalescent based; b, concatenation based) with *Cleome violacea* (Cleomaceae) as outgroup. LPP values (a) and BS values (b) are shown.



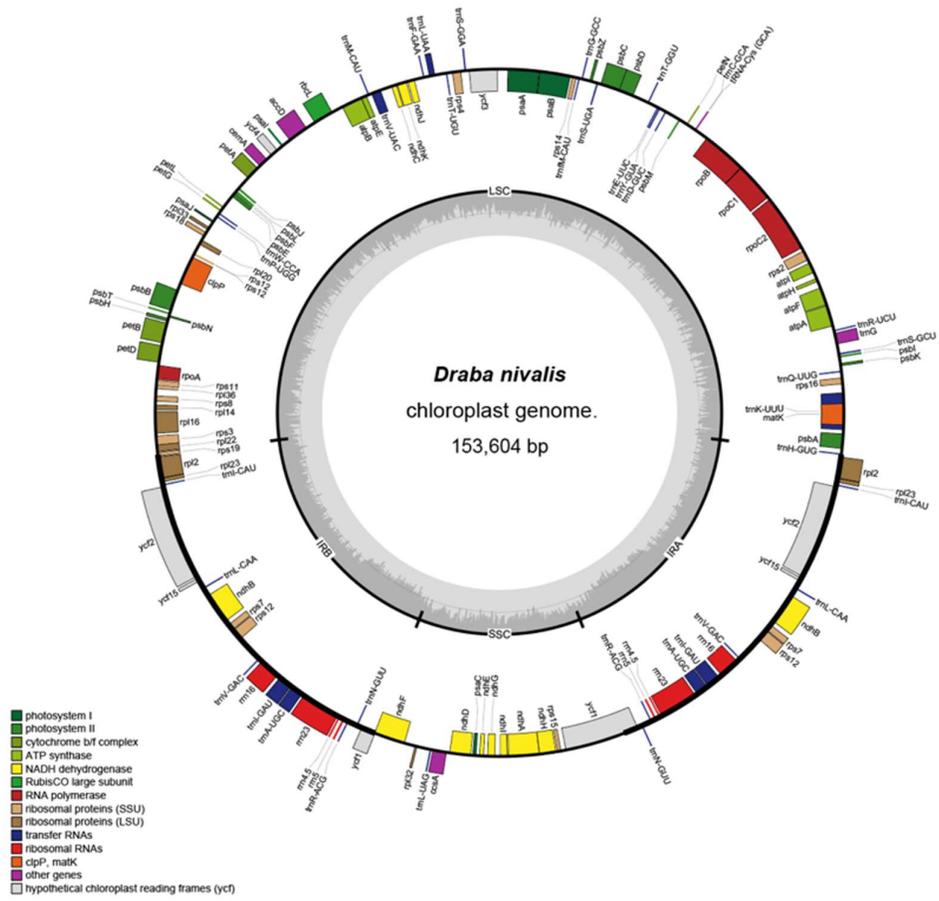
Supplemental Figure 28. Phylogeny inferred with 5217 low-copy ortholog genes (shared ortholog groups) for 55 Brassicaceae species. Support values (similar to LPP) generated via STAG (v1.0.0) are shown.



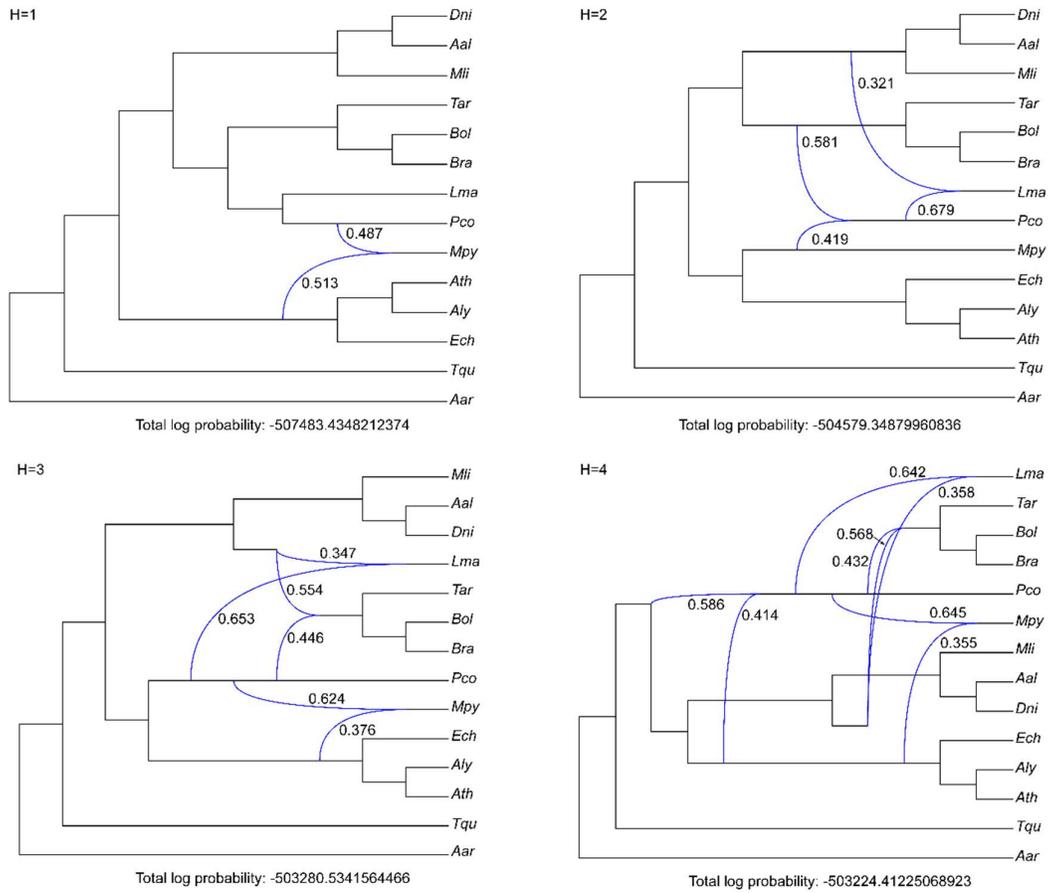
Supplemental Figure 30. Chloroplast genome structure for *Tqu*.



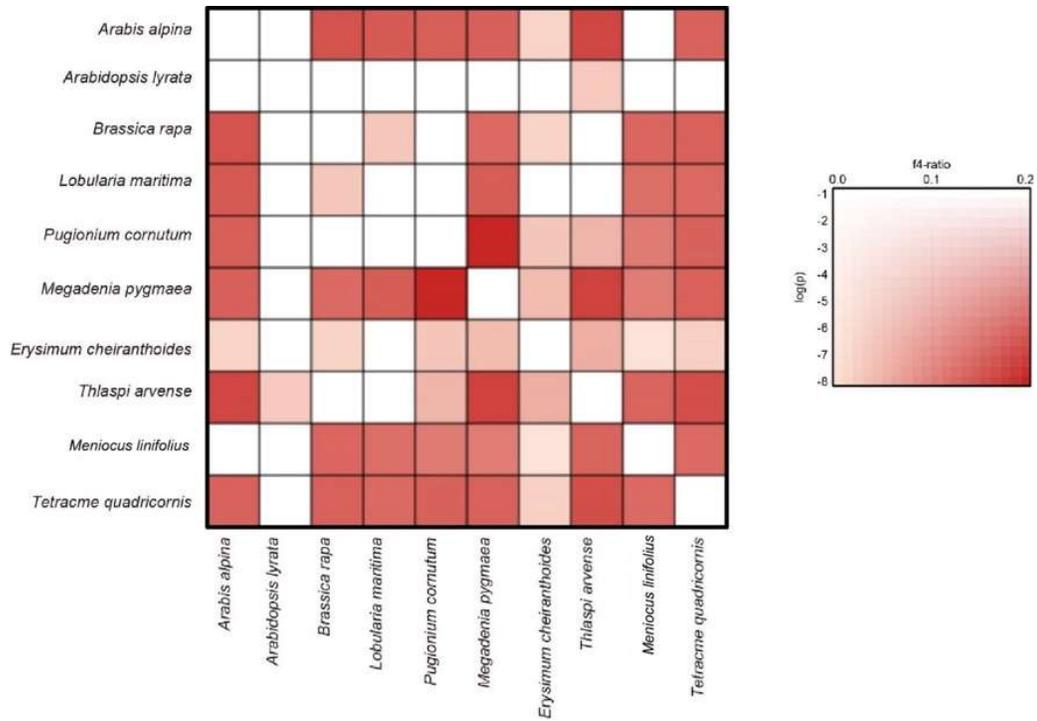
Supplemental Figure 31. Chloroplast genome structure for *Mli*.



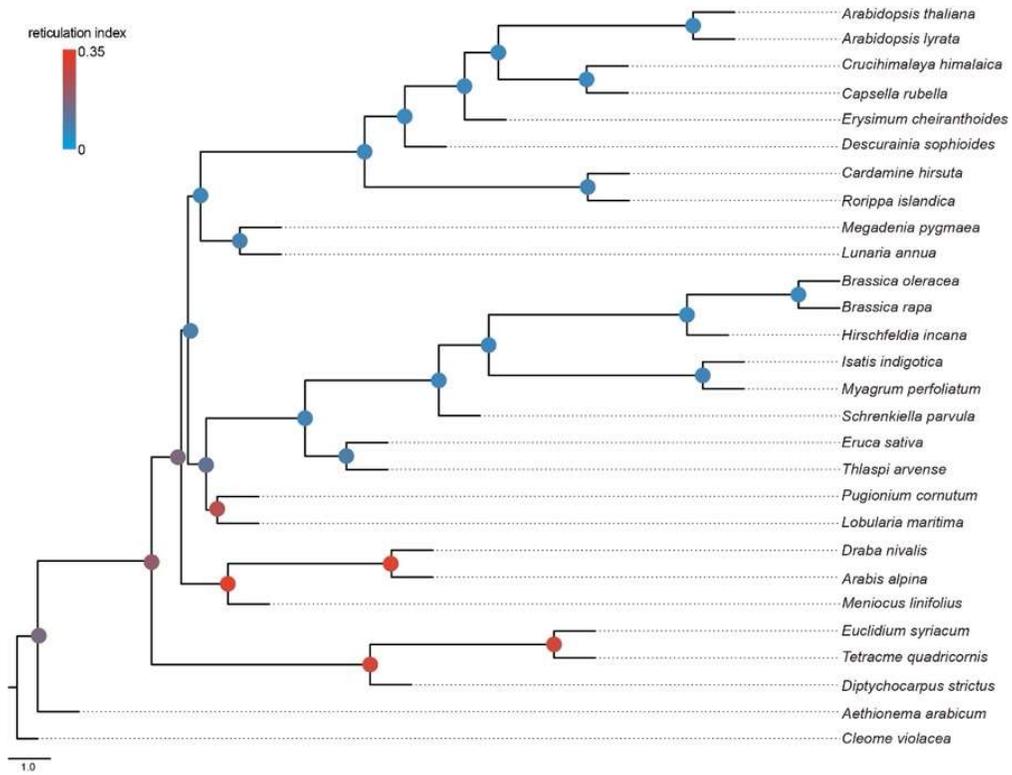
Supplemental Figure 32. Chloroplast genome structure for *Dni*.



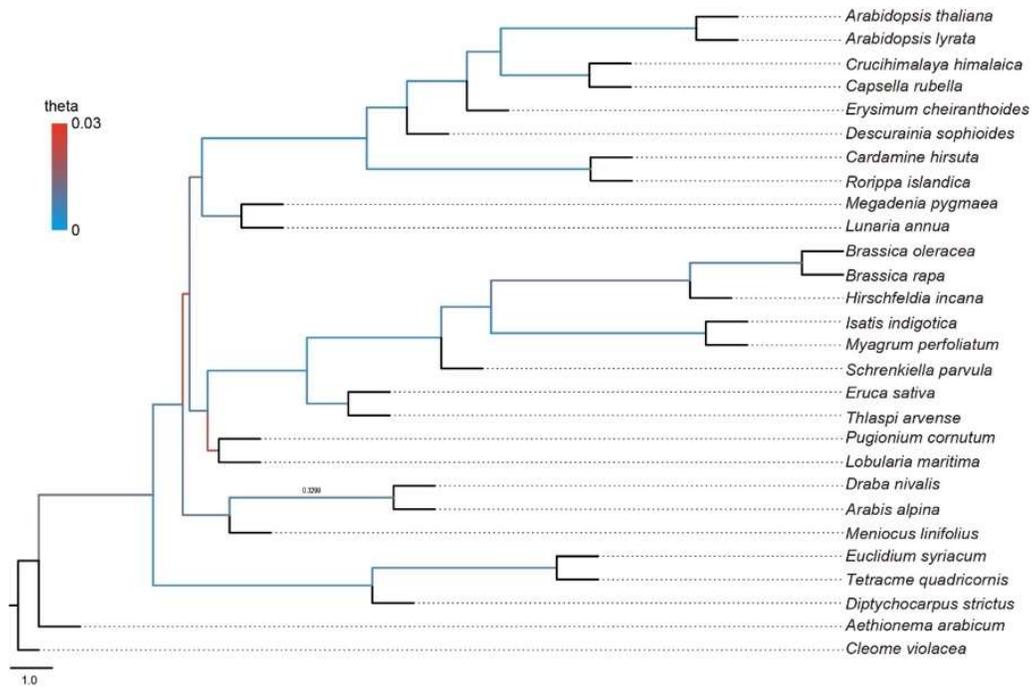
Supplemental Figure 33. Reticulation history analyses using *PhyloNet* with the number of reticulations set to 1 (a), 2 (b), 3 (c), and 4 (d), respectively.



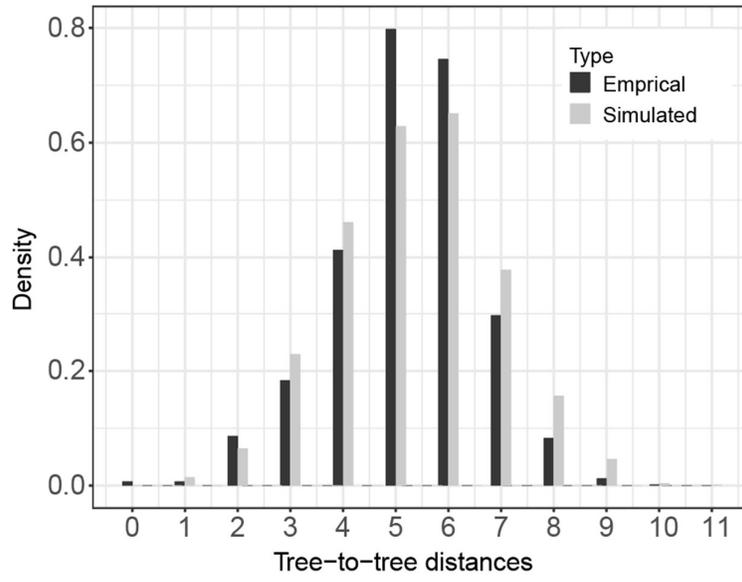
Supplemental Figure 34. Maximum pairwise f_4 heatmaps between pairs of species across all combinations of trios. The depth of color (from white to red) shows the level of f_4 values.



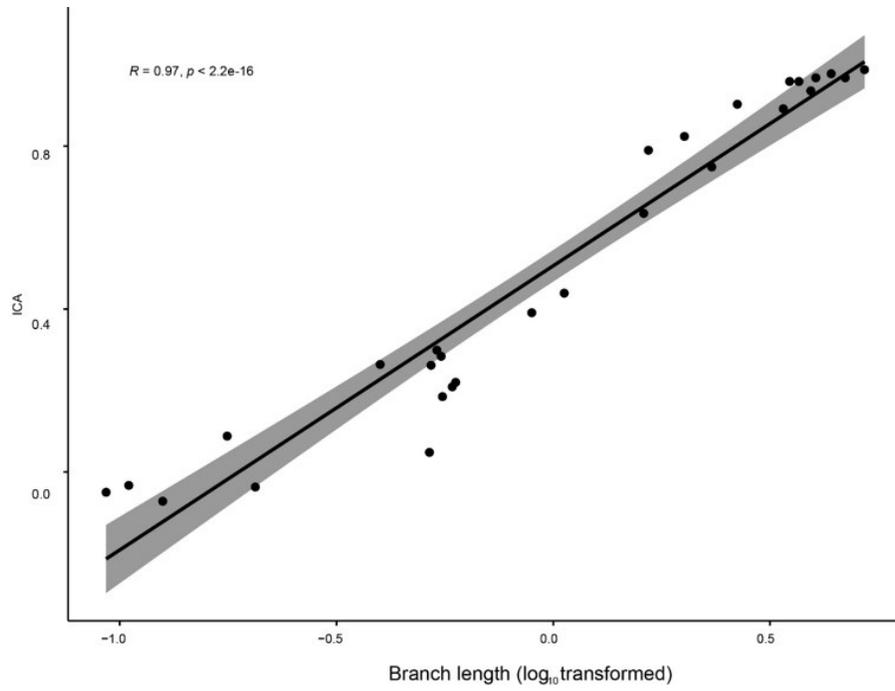
Supplemental Figure 35. Reticulation Index among Brassicaceae. Node color close to red indicates higher probabilities of hybridization.



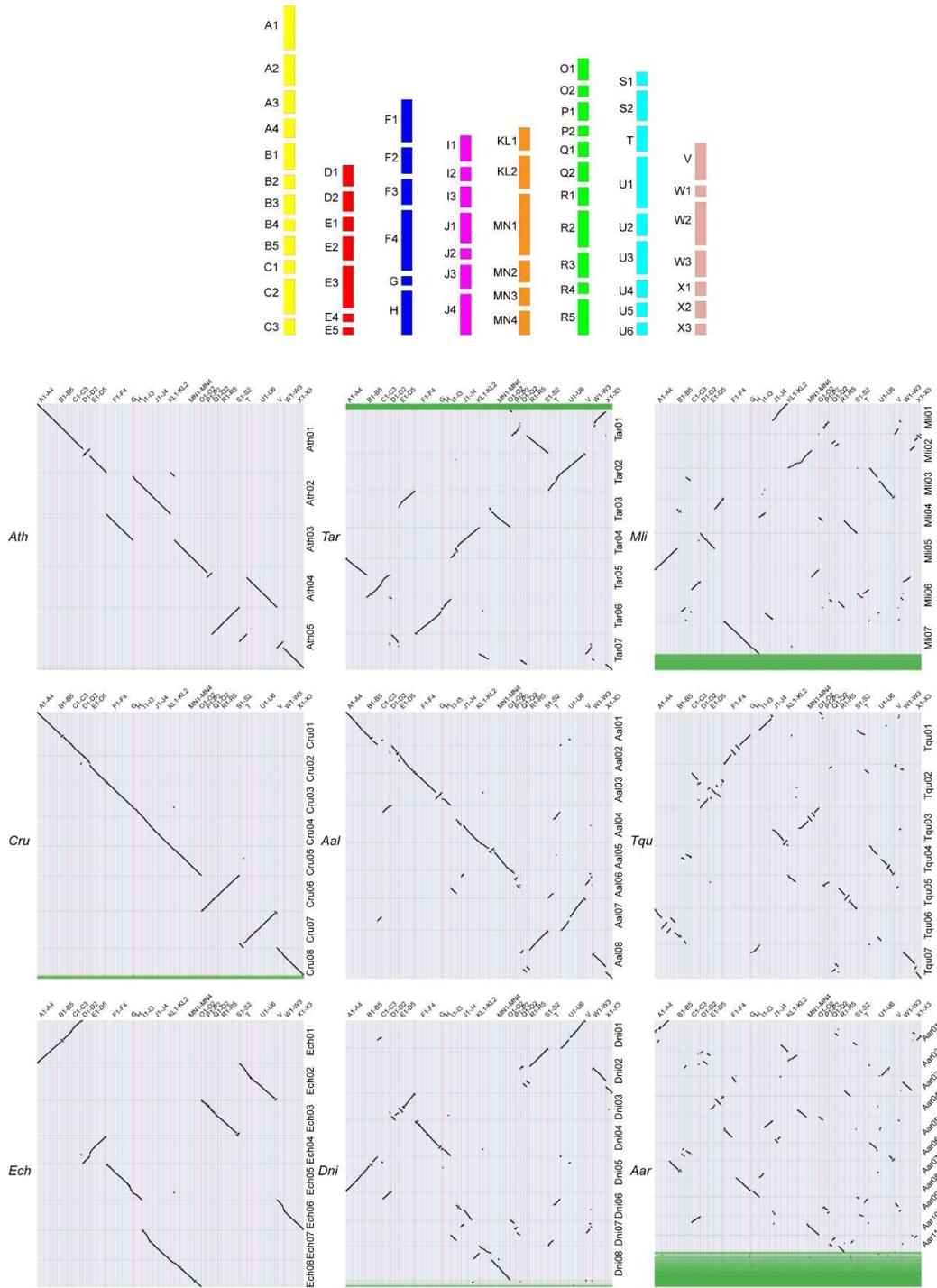
Supplemental Figure 36. Estimated theta values for each internal branch of Brassicaceae. Phylogenetic tree with branches was colored by the inferred population mutation parameter theta, which reflects the population polymorphism generated by dividing the mutation units for each internal branch by coalescent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



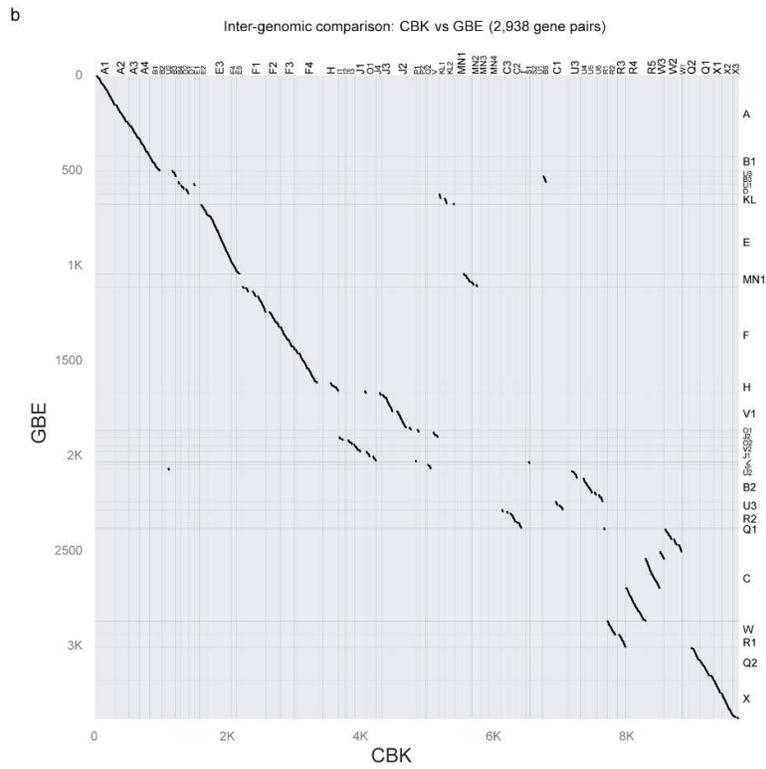
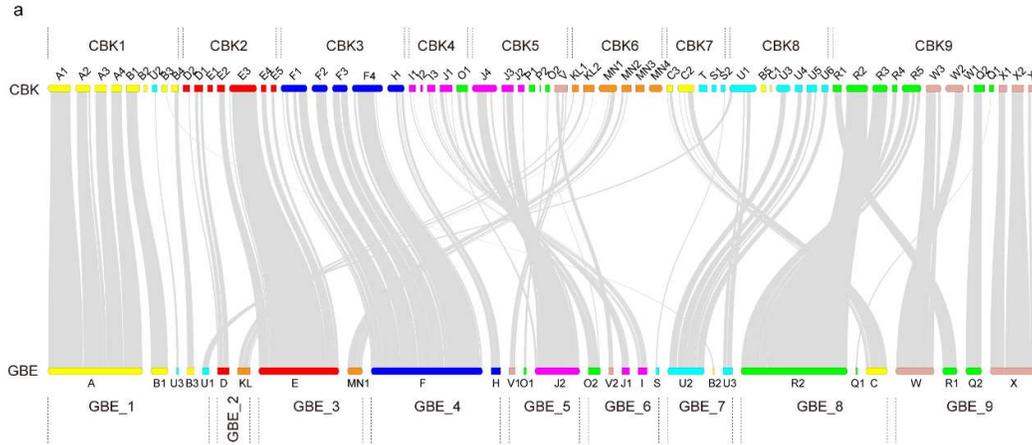
Supplemental Figure 37. Coalescent simulation shows the distribution of tree-to-tree distances between empirical gene trees and the *ASTRAL* species tree (black bars) compared to those from the coalescent simulation (grey bars).



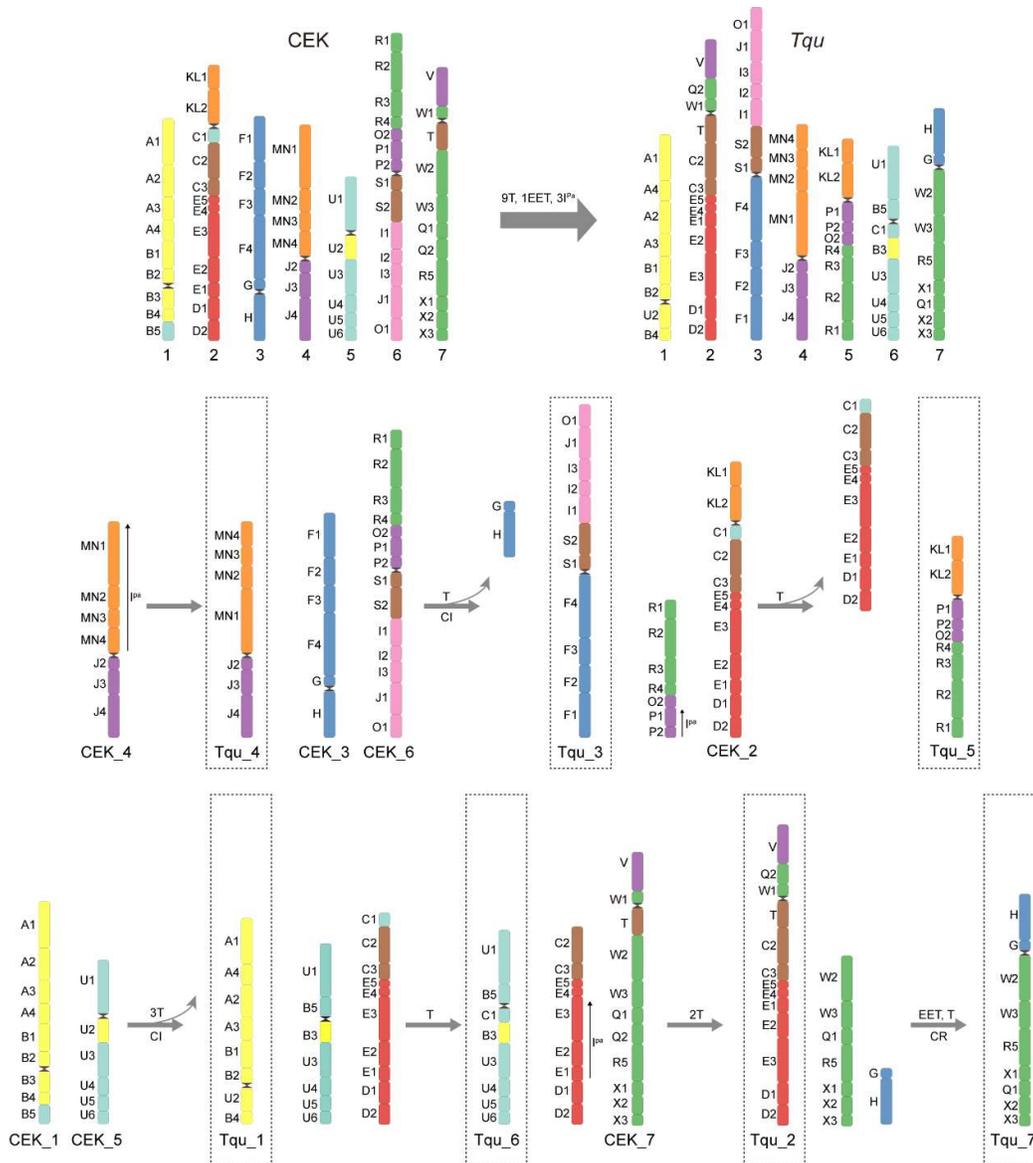
Supplemental Figure 38. Correlation between length of branch and internode certainty (ICA) value tested with a *Pearson's correlation test (two-sided)*.



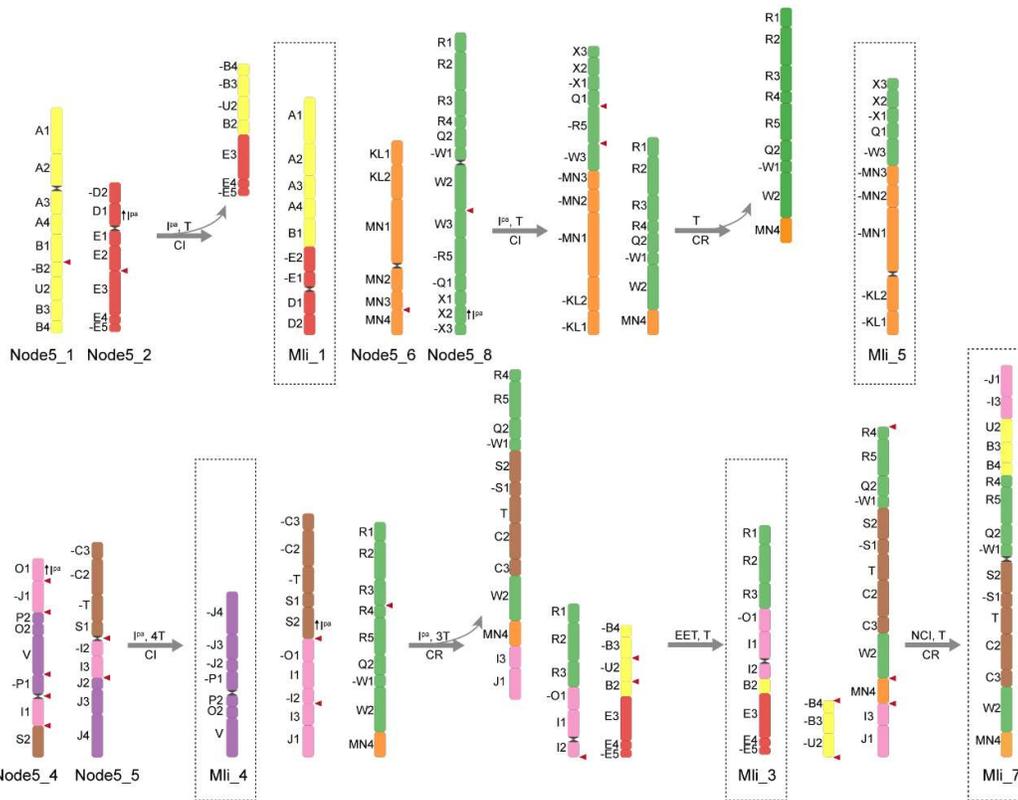
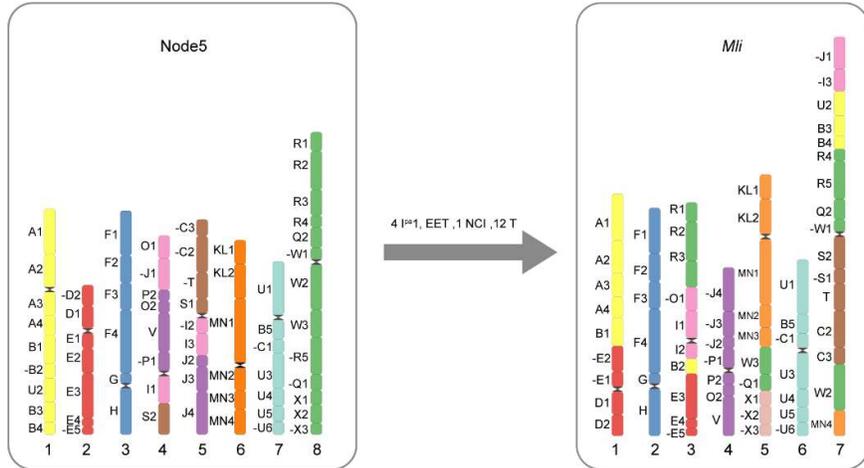
Supplemental Figure 39. Reconstruction of the 65 CBK genomic blocks based on the syntenic relationships between ACK and nine extant species.



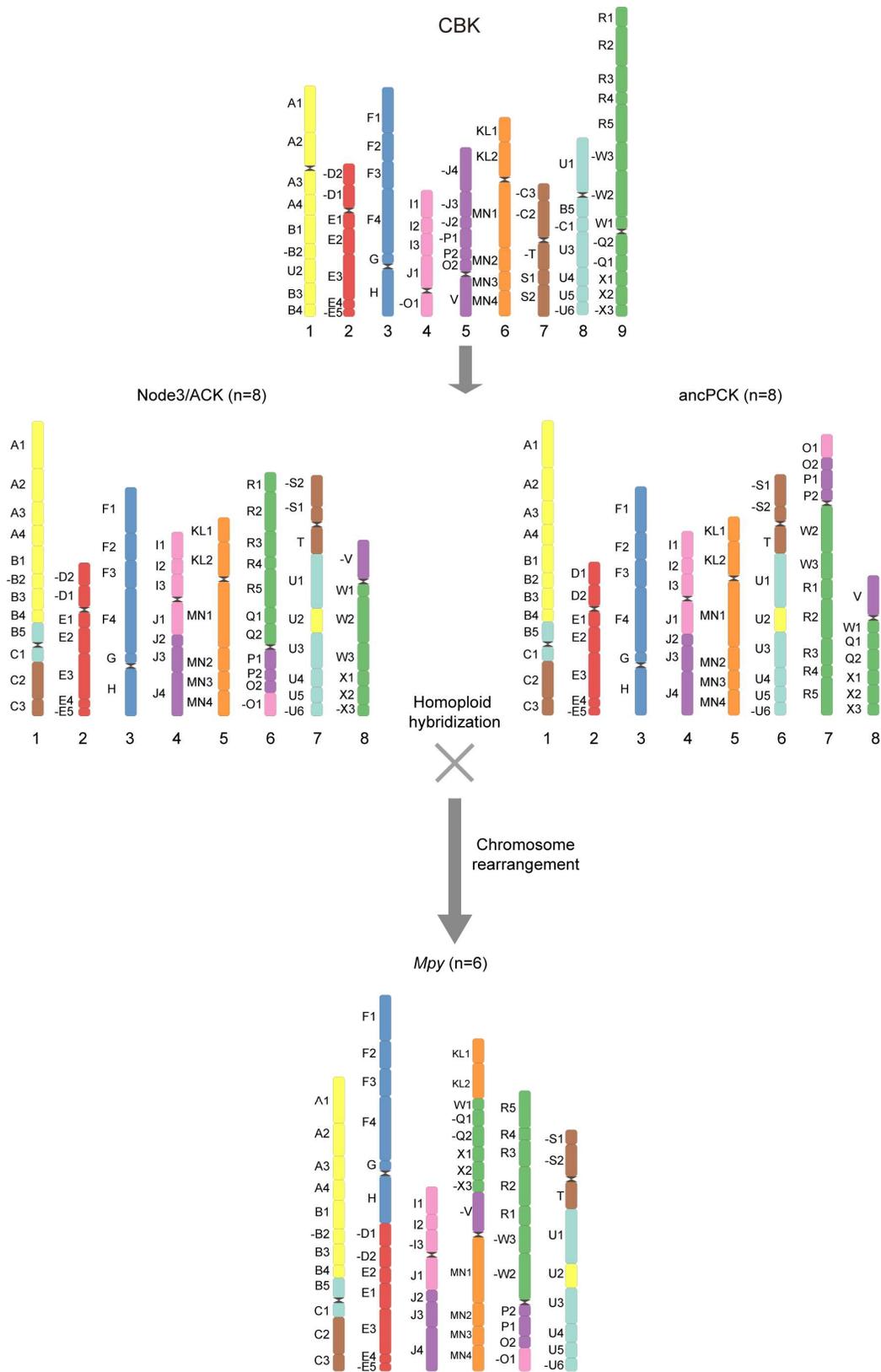
Supplemental Figure 40. The comparison between CBK and GBE (Walden and Schranz, 2023, Genome Biology and Evolution). a, The ancestral genomes of CBK (left) and GBE (Right). b, The syntenic relationship between CBK (upper) and GBE (lower). c, Dot plots between CBK and GBE.



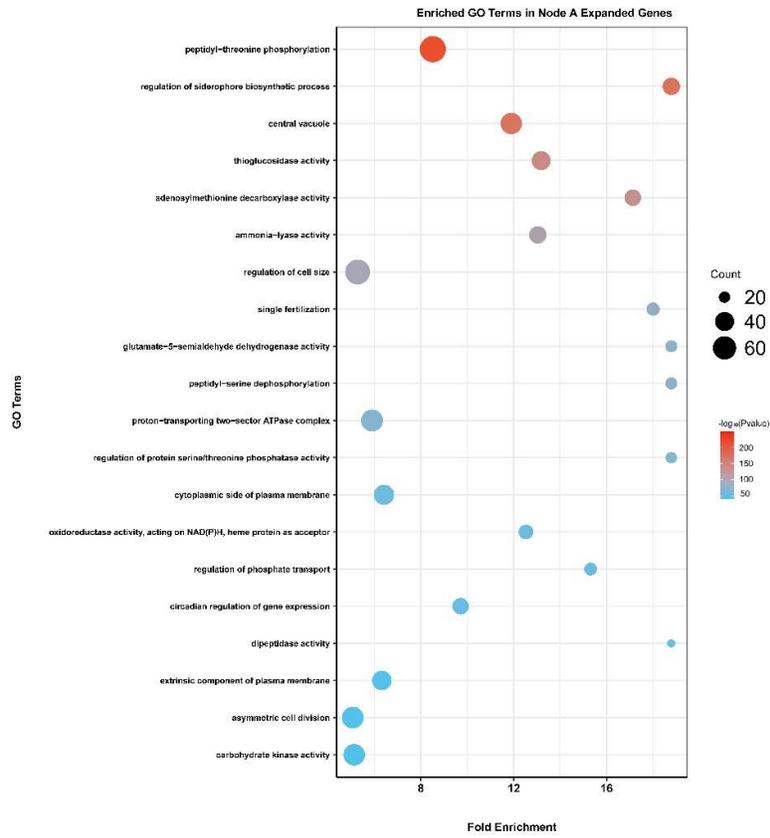
Supplemental Figure 41. Deduced scenario for the chromosome evolution from CEK to *Tqu*. Position of block T was determined based on its location in *Tqu*. Considering that CEK was procured via cytological methods (comparative chromosome painting, CCP), the direction of each genomic block was not factored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



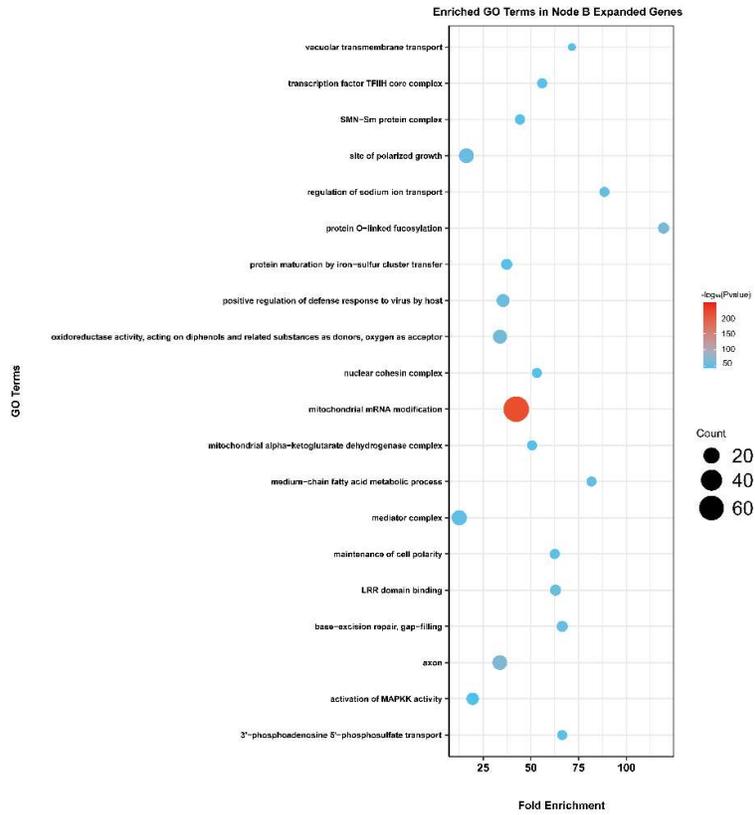
Supplemental Figure 42. Deduced scenario for the chromosome evolution in *Mli*.



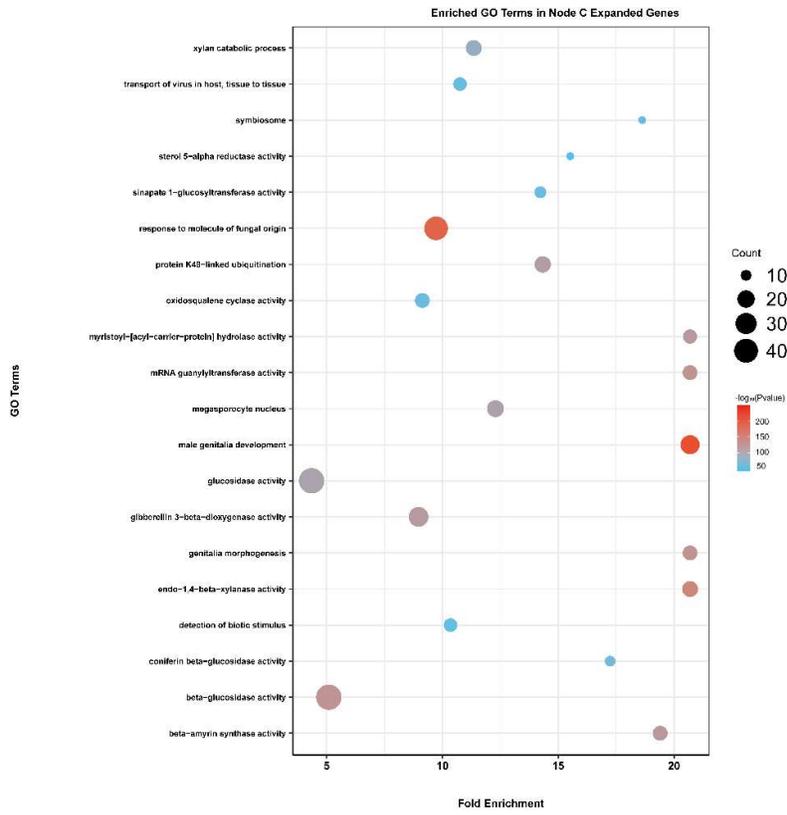
Supplemental Figure 43. Deduced scenario of the hybridization origin of *Megadenia pygmaea*.



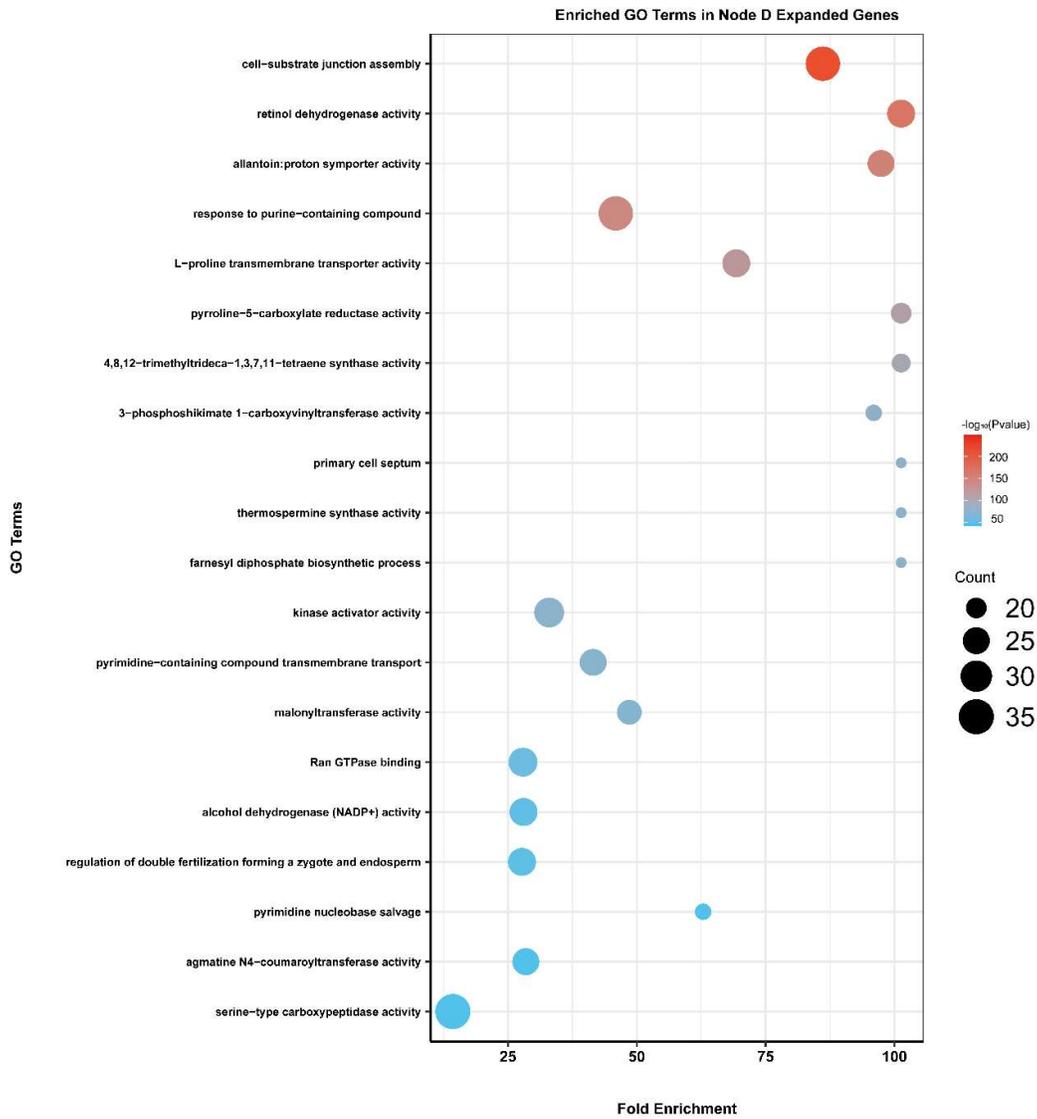
Supplemental Figure 44. Top 20 enriched Gene Ontology (GO) terms of the expanded gene families at node A of Figure 1F.



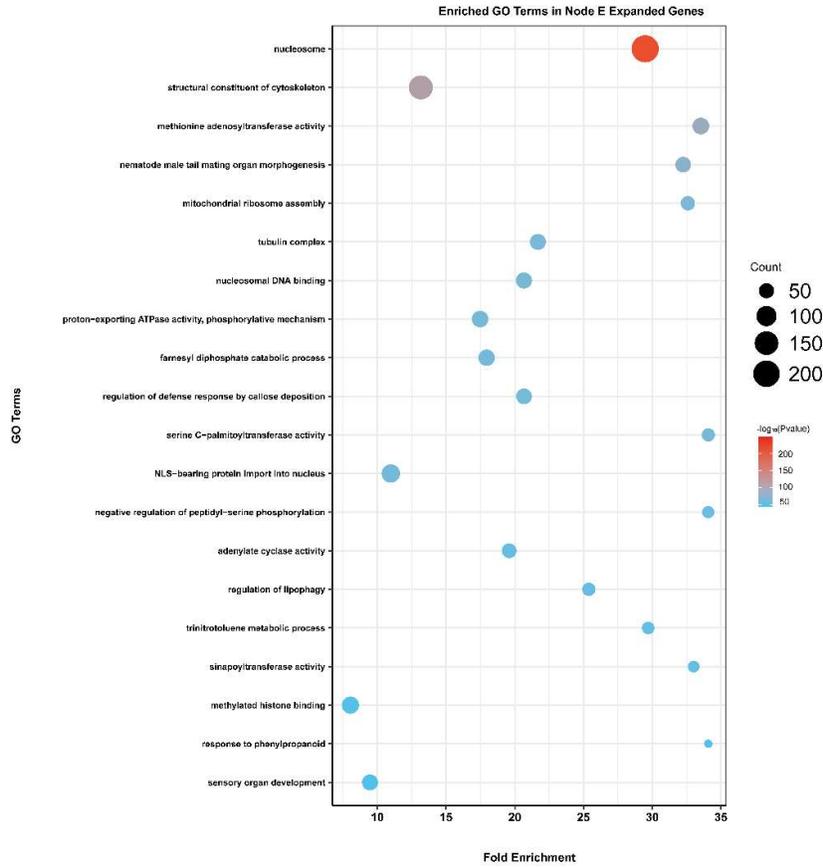
Supplemental Figure 45. Top 20 enriched Gene Ontology (GO) terms of the expanded gene families at node B of Figure 1F.



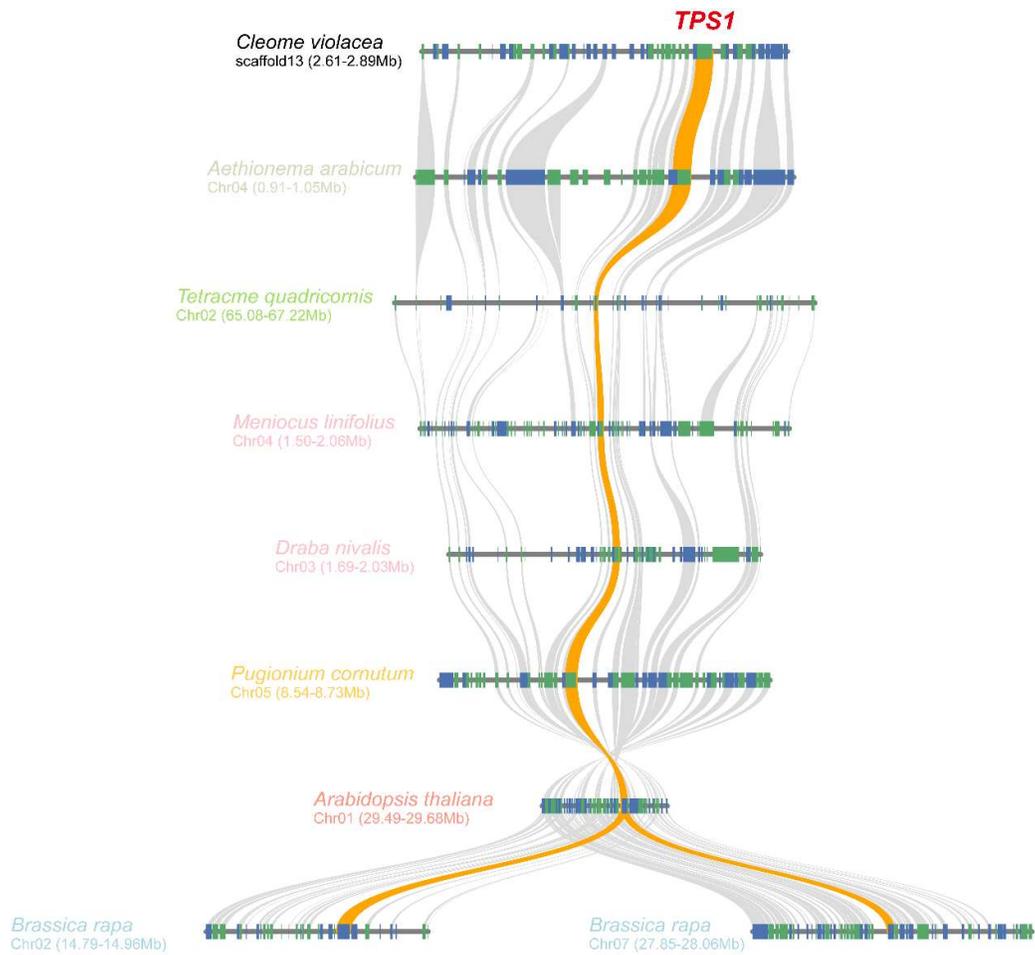
Supplemental Figure 46. Top 20 enriched Gene Ontology (GO) terms of the expanded gene families at node C of Figure 1F.



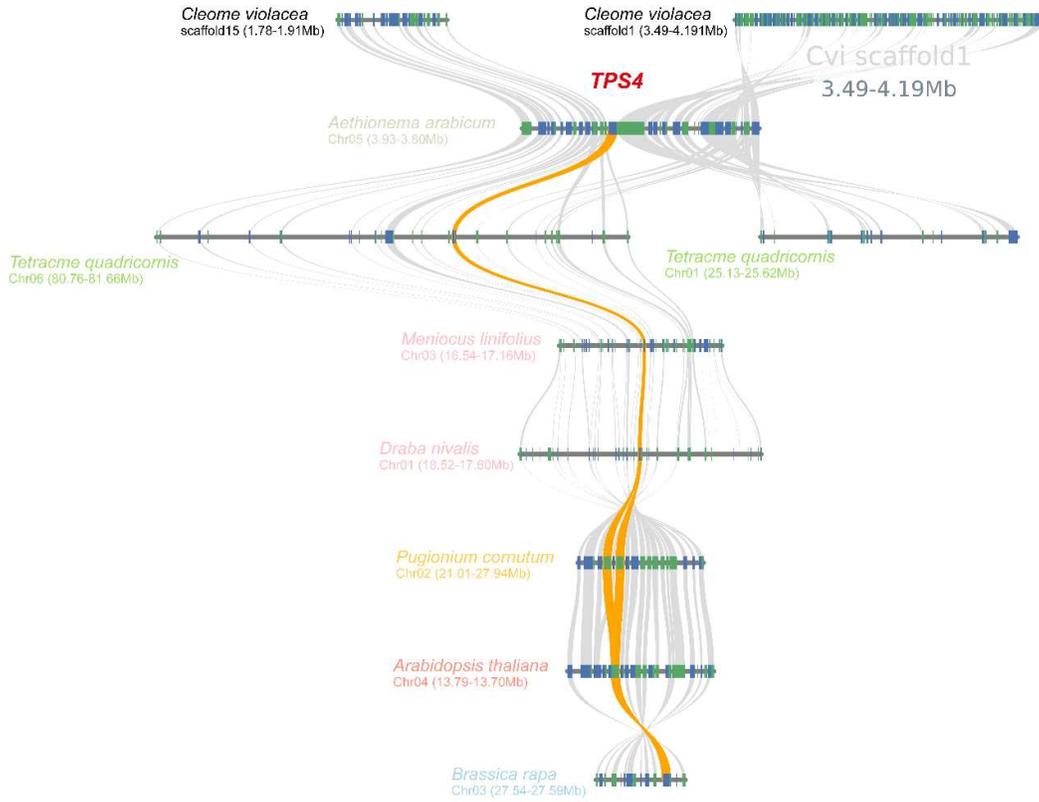
Supplemental Figure 47. Top 20 enriched Gene Ontology (GO) terms of the expanded gene families at node D of Figure 1F.



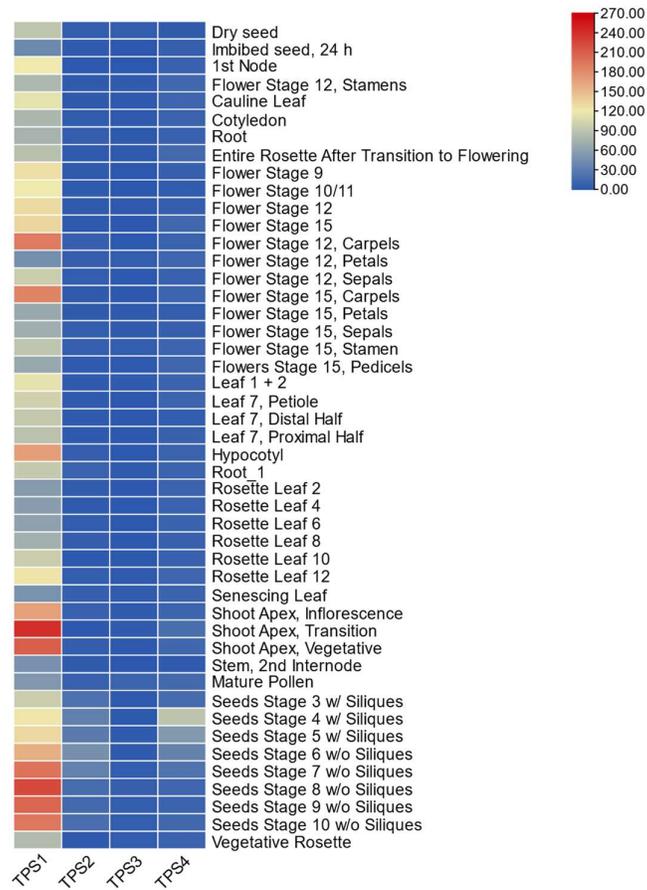
Supplemental Figure 48. Top 20 enriched Gene Ontology (GO) terms of the expanded gene families at node E of Figure 1F.



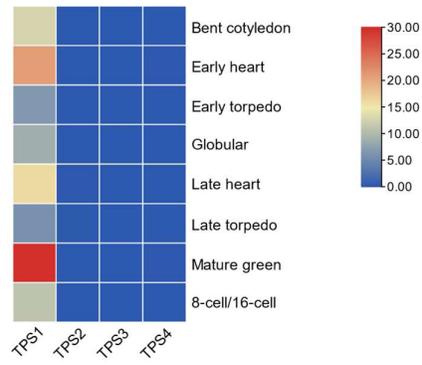
Supplemental Figure 49. The collinearity relationships of subgroups *TPS1* in Brassicaceae and *Cleome violacea* (Cleomaceae).



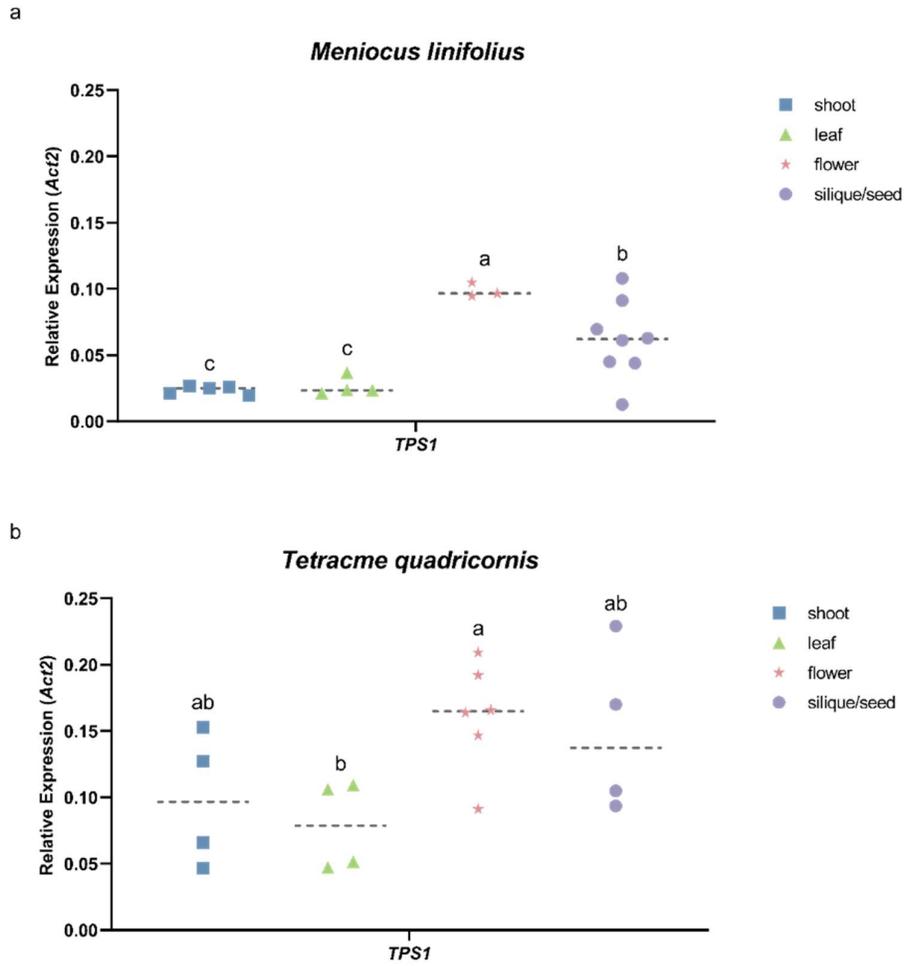
Supplemental Figure 50. The collinearity relationships of subgroups *TPS4* in Brassicaceae and *C. violacea* (Cleomaceae).



Supplemental Figure 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).