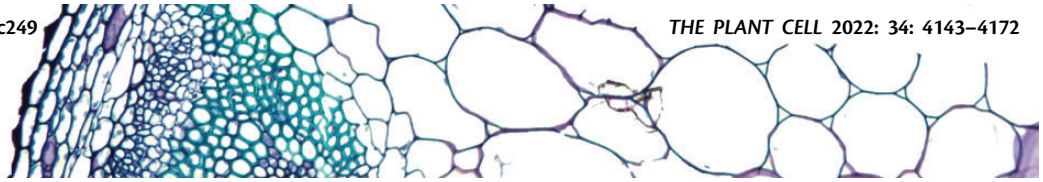
https://doi.org/10.1093/plcell/koac249 THE PLANT CELL 2022: 34: 4143–4172



Research Article

The final piece of the Triangle of U: Evolution of the tetraploid *Brassica carinata* genome

Won Cheol Yim [1,2,*], Mia L. Swain [1,*], Dongna Ma [3], Hong An [4], Kevin A. Bird [5], David D. Curdie [1], Samuel Wang [1], Hyun Don Ham [1], Agusto Luzuriaga-Neira [6], Jay S. Kirkwood [7], Manhoi Hur [7], Juan K.Q. Solomon [8], Jeffrey F. Harper [1], Dylan K. Kosma [1], David Alvarez-Ponce [6], John C. Cushman [1], Patrick P. Edger [5], Annaliese S. Mason [9], J. Chris Pires [10], Haibao Tang [3] and Xingtan Zhang [3]

1 Department of Biochemistry and Molecular Biology, University of Nevada, Reno, Nevada 89557, USA
2 Department of Life Science, Dongguk University, Goyang-si, Gyeonggi-do 10326, Republic of Korea
3 Fujian Provincial Key Laboratory of Haixia Applied Plant Systems Biology, Key Laboratory of Ministry of Education for Genetics, Breeding and Multiple Utilization of Crops, Key Laboratory of National Forestry and Grassland Administration for Orchid Conservation and Utilization, Fujian Agriculture and Forestry University, Fuzhou, China
4 Division of Biological Sciences, University of Missouri, Columbia, Missouri 65201, USA
5 Department of Horticulture, Michigan State University, East Lansing, Michigan 48824, USA
6 Department of Biology, University of Nevada, Reno, Nevada 89557, USA
7 Metabolomics Core Facility, Institute for Integrative Genome Biology, University of California, Riverside, California 92521, USA
8 Department of Agriculture, Veterinary & Rangeland Sciences, University of Nevada, Reno, Nevada 89557, USA
9 Plant Breeding Department, INRES, The University of Bonn, Bonn 53115, Germany
10 Division of Biological Sciences, Bond Life Sciences Center, , University of Missouri, Columbia, Missouri 65211, USA

*Author for correspondence: wyim@unr.edu
These authors contributed equally (W.C.Y. and M.L.S.).

W.C.Y. led and managed the project. H.D.H., M.L.S., and W.C.Y. collected, prepared, and sequenced the plant material. A.S.M., P.E., J.C.C., J.H., J.C.P., J.S., H.T., W.C.Y., and X.Z. devised the main conceptual ideas and proof outline. W.C.Y., H.T., and X.Z. assembled and annotated the genome. D.M., S.W., W.C.Y., and X.Z. performed gene family and gene enrichment analysis. D.D.C. and W.C.Y. performed Hi-C analysis. A.L.N., D.A.P., and W.C.Y. performed comparative evolutionary analysis. A.S.M., H.A., K.B., P.E., J.C.P., H.T., and W.C.Y. designed and performed homoeolog exchange analysis. All authors contributed to the article and approved the submitted version.

The authors responsible for distribution of materials integral to the findings presented in this article in accordance with the policy described in the Instructions for Authors (https://academic.oup.com/plcell) are: Mia L. Swain (mias@unr.edu) and Won Cheol Yim (wyim@unr.edu).

Abstract

Ethiopian mustard (*Brassica carinata*) is an ancient crop with remarkable stress resilience and a desirable seed fatty acid profile for biofuel uses. *Brassica carinata* is one of six *Brassica* species that share three major genomes from three diploid species (AA, BB, and CC) that spontaneously hybridized in a pairwise manner to form three allotetraploid species (AABB, AACC, and BBCC). Of the genomes of these species, that of *B. carinata* is the least understood. Here, we report a chromosome scale 1.31-Gbp genome assembly with 156.9-fold sequencing coverage for *B. carinata*, completing the reference genomes comprising the classic Triangle of U, a classical theory of the evolutionary relationships among these six species. Our assembly provides insights into the hybridization event that led to the current *B. carinata* genome and the genomic features that gave rise to the superior agronomic traits of *B. carinata*. Notably, we identified an expansion of transcription factor networks and agronomically important gene families. Completion of the Triangle of U comparative genomics platform has allowed us to examine the dynamics of polyploid evolution and the role of subgenome dominance in the domestication and continuing agronomic improvement of *B. carinata* and other *Brassica* species.

Received January 03, 2022. Accepted June 24, 2022. Advance access publication August 12, 2022

IN A NUTSHELL

Background: *Brassica carinata* (Ethiopian mustard) is an ancient crop from the Ethiopian highlands with remarkable heat and drought tolerance that has potential as a sustainable oil source for biofuel production. The resilience of this species might be due to hybrid vigor, as *B. carinata* is a species derived from a hybridization between Brassica *nigra* (black mustard) and *Brassica oleracea* (kale, broccoli, etc.). Thus, the *B. carinata* genome is allotetraploid with two parental genomes, or subgenomes, merged in one nucleus. We present a high-quality, chromosome-scale reference assembly of the *B. carinata* genome, which is the last of six genomes comprising the classic Triangle of U model used to study hybridization and polyploid evolution.

Question: Here, we compare *B. carinata* to the other Triangle of U genomes for insight into the remarkable heat and drought tolerance of this crop. We investigate the evolutionary trajectory of the *B. carinata* genome as it returns to the diploid state to elucidate the mechanisms that act on duplicated genes, such as functional divergence of gene families and the biased fractionation of one subgenome.

Findings: The *B. carinata* genome is the largest among the Triangle of U with notable expansions in repetitive DNA sequences and gene families related to transcriptional regulation and stress tolerance. We characterized patterns of subgenome bias, finding that the subgenome derived from *B. nigra* is likely dominant over the subgenome from *B. oleracea*. Furthermore, we comprehensively characterize subgenomic bias in homoeologous exchanges, or meiotic crossover between subgenomes, in the Triangle of U allotetraploids.

Next steps: The presented *B. carinata* genome is a crucial resource for its expanded use as a biofuel feedstock and insight into polyploid evolution. Unraveling the genomic basis of the stress resilience of *B. carinata* provides an opportunity to introgress these traits to other cruciferous vegetables, which are used worldwide as vegetable and oil sources.

Introduction

Brassica carinata has been traditionally cultivated as a dual oilseed and leafy vegetable crop in the Ethiopian highlands where it has served as a staple food in rural areas (Ojiewo et al., 2013). Among the oldest Ethiopian crops and locally referred to as *gomenzer* in the Amharic language (Hagos et al., 2020), *B. carinata* has also recently gained traction as an alternative biofuel feedstock in countries such as Canada, India, and Italy for its adaptability and notable productivity on marginal land (Cardone et al., 2003; Taylor et al., 2010). *Brassica carinata* is a highly productive mustard and one of the most drought- and heat-tolerant species within the economically important Brassicaceae family (Kumar et al., 1984; Malik, 1990). Its seed oil has significant erucic acid (C22:1) content, which is toxic to humans but ideal for biofuel production due to the resulting high cetane number and oxidative stability (Cardone et al., 2003; Folayan et al., 2019). Biofuels could offer sustainable energy security while decreasing greenhouse gas emissions; however, more efficient crop and land use are necessary to reap such climate benefits while avoiding effects on food security (Kazamia and Smith, 2014; Searchinger et al., 2018). Its high productivity under low input conditions and high seed erucic acid contents make *B. carinata* a more sustainable alternative to current biodiesel feedstocks.

The Triangle of U describes the evolutionary relationships among six globally important *Brassica* species that share independently evolved versions of three core genomes, termed A, B, and C (Figure 1) (Nagaharu, 1935). The three diploid species *Brassica rapa* (Ar, $2n = 20$), Brassica *nigra* (Bn, $2n = 16$), and *Brassica oleracea* (Co, $2n = 18$) participated in pairwise hybridizations to form three allotetraploids: *Brassica juncea* (AjBj, $2n = 36$), *Brassica napus* (AnCn, $2n = 38$), and *B. carinata* (BcCc, $2n = 34$). According to the classic Triangle of U model, the *B. carinata* genome spontaneously arose from hybridization between two diploid progenitor species, *B. nigra* and *B. oleracea* (Nagaharu, 1935). The overlapping genomic relationships among these six *Brassica* species allow researchers to compare the evolution of a given genome in different genomic environments. This ability to investigate the evolutionary trajectories of the same core genome in divergent species is ideal for elucidating the mechanisms that act on duplicated genes during diploidization, such as functional divergence of gene families and the biased fractionation of one parental genome or subgenome (Thomas et al., 2006).

Biased fractionation, when diverged genomes within one nucleus display asymmetric gene expression, occurs in many allopolyploids and leads to greater fractionation of the recessive, less expressed subgenome (Cheng et al., 2018). The molecular mechanisms that reconcile the effects of merging divergent genomes with species-specific genetic and epigenetic differences have drawn much attention due in part to the close association between polyploidization and crop domestication (Bird et al., 2018). Elucidating the genetic

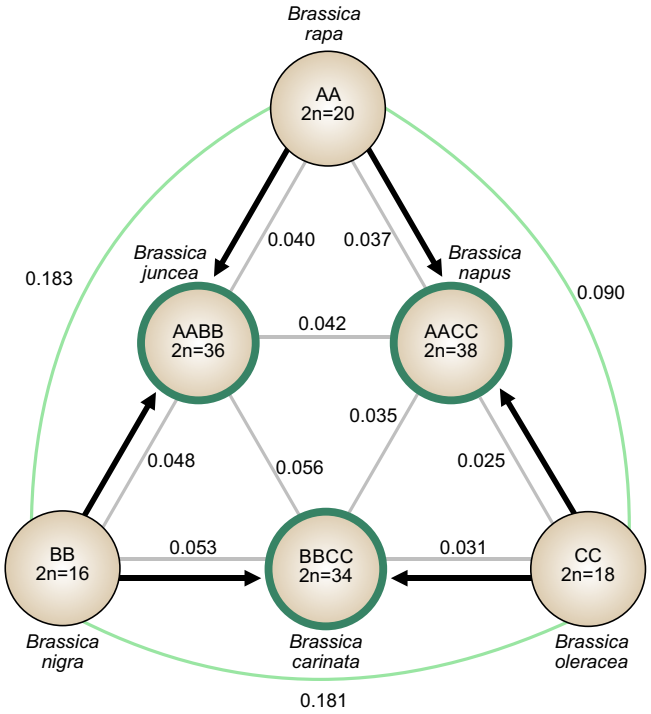


Figure 1 Genetic relationships between the *Brassica* species in the Triangle of U model. The Triangle of U describes the genetic relationships among six *Brassica* species that share the same three core genomes. The three diploid species (*B. rapa*, *B. nigra*, and *B. oleracea*) representing the AA, BB, and CC genomes, respectively, are shown at the nodes of the triangle. The arrows along the sides of the triangle indicate the pairwise hybridization events that occurred to generate the three allotetraploid species (*B. juncea*, *B. napus*, and *B. carinata*). The synonymous divergence (K_s) between each pair of species is shown to convey the degree of differentiation of the shared genomes. As K_s is proportional to divergence time, a larger K_s value indicates a longer time since the species diverged.

consequences of polyploidization and the genomic features that determine subgenome dominance will be essential for understanding crop domestication, directing crop improvement, and exploring angiosperm evolution on a larger scale. Furthermore, the importance of these *Brassica* species as global vegetable sources, including crops such as canola (*B. rapa*), broccoli and cabbage (*B. oleracea*), and mustard seeds (*B. nigra* or *B. juncea*), provides a strong economic motive for elucidating the genomic features that contribute to their different traits. *Brassica* lines carrying advantageous alleles at the genomic regions defined here could be useful as introgression donors in breeding for improved climate resilience and oil production, which could decrease the costs of renewable diesel and jet fuels (Wei et al., 2016; Zhang et al., 2017; Basili and Rossi, 2018).

Here, we generated a complete and comprehensive characterization of the *B. carinata* genome to help illustrate the genetic basis of its numerous valuable traits. Our chromosome-scale assembly of the var. Gomenzer genome represents a significant improvement compared to the previous *B. carinata* Zd-1 assembly (Song et al., 2021). This higher quality genome assembly has allowed us to build upon the findings of Song et al. to explore the unique features of the *B. carinata* genome and confirm an expansion of repetitive sequences as well as reveal remarkable increases in the size and gene content of the *B. carinata* genome. Notably, we report here the role of subgenome dominance and transposable element (TE) insertions in the domestication of this species. We also present a robust comparative analysis of genome evolution trajectories among the Triangle of U species using available genome assemblies (Chalhoub et al., 2014; Liu et al., 2014; Parkin et al., 2014; Yang et al., 2016; Zhang et al., 2018; Belser et al., 2018; Rousseau-Gueutin et al., 2020; Lee et al., 2020; Song et al., 2020; Perumal et al., 2020; Cai et al., 2020; Lv et al., 2020; Paritosh et al., 2021).

Results

Our high-quality assembly improves our understanding of the genome and domestication of *B. carinata*

We assembled the genome of *B. carinata* A. Braun var. Gomenzer (USDA accession Gomenzer, PI 273640) using PacBio, Illumina, and Hi-C sequencing data. The initial PacBio read-based assembly used 140.76 Gbp of sequencing data and produced a 1.33 Gbp draft genome comprising 11,531 contigs with an N50 value of 602.9 kb (Supplemental Data Sets 1 and 2). Next, the contigs were segregated according to their subgenomes, Bc or Cc, using the *B. nigra* (Bn) and *B. oleracea* (Co) genomes as references (Liu et al., 2014; Wang et al., 2019). The subgenome-clustered contigs were then ordered and oriented into scaffolds according to a contact frequency matrix of Hi-C loci that allowed anchoring of the scaffolded sequences onto pseudochromosomes. Lastly, gaps in the assembly were filled and sequencing errors were corrected using the PacBio data and 63.4 G bp of Illumina sequencing data (Supplemental Data Sets 3 and 4). The final assembly resulted in a 1.31-Gbp genome with a 156.9-fold sequencing depth (Figure 2; Supplemental Data Set 5) with chromosome-scale scaffolds that span the lengths of each of the 17 *B. carinata* chromosomes. The remarkable scaffold N50 value of 78.8 Mbp is the largest among the Triangle of U genome assemblies (Supplemental Data Set 6). This nearly complete final assembly covers a 95% confidence interval of the estimated size of the *B. carinata* genome, which was 1.28 Gbp by flow cytometry and 1.31 Gbp by *k*-mer analysis (Supplemental Figure S1; Supplemental Data Set 7). Genome-wide analysis of the Hi–C chromatin interactions and syntenic comparisons shows well-organized sequences with a high degree of synteny with those of shared Triangle of U genomes (Figure 3; Supplemental Figure S2).

The completeness and contiguity of our genome assembly were assessed in terms of gene space and repetitive sequences. Our survey of Benchmarking Universal Single-Copy Orthologs (BUSCOs) identified 97.0% of the 425 conserved genes in the Brassicales dataset, 83.1% of which were

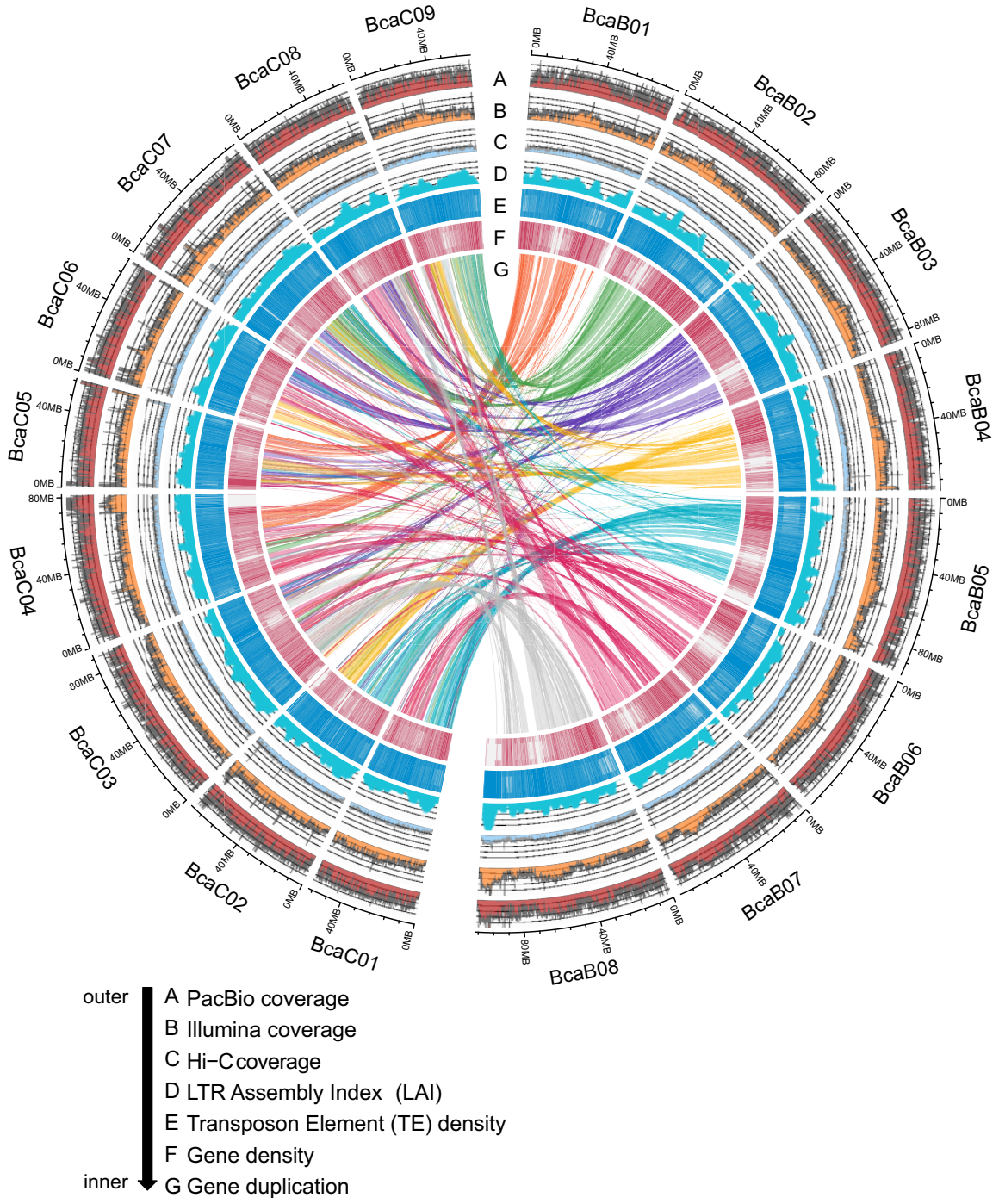


Figure 2 Chromosomal features of the *B. carinata* Gomenzer genome assembly. The *B. carinata* genome comprises a total of 17 chromosomes with eight in the Bc subgenome (BcaB01-BcaB08, right-hand semicircle) and nine in the Cc subgenome (BcaC01-BcaC09, left-hand semicircle). The chromosomes are scaled according to their assembled length. Homoeologous relationships between the two subgenomes are indicated with links connecting Bc subgenome homoeologs with Cc subgenome homoeologs. The tracks, from outermost to innermost, display the sequences coverage depths for PacBio, Illumina, and Hi-C data, LAI, TE density, gene density, and gene duplication. All of the plots are drawn in a 100-kb sliding window.

duplicated (Supplemental Figure S3; Simão et al., 2015). Identification of the majority of BUSCOs indicates a high-quality genome assembly, and the magnitude of duplications reflects the relatively recent hybridization event that the *B. carinata* genome experienced.

We also assessed the number of intact long terminal repeat (LTR) retrotransposons identified in the genome assembly as an indicator of repetitive sequence contiguity using the LTR assembly index (LAI; Ou et al., 2018). Lower quality assemblies often collapse the large repetitive regions of plant genomes,

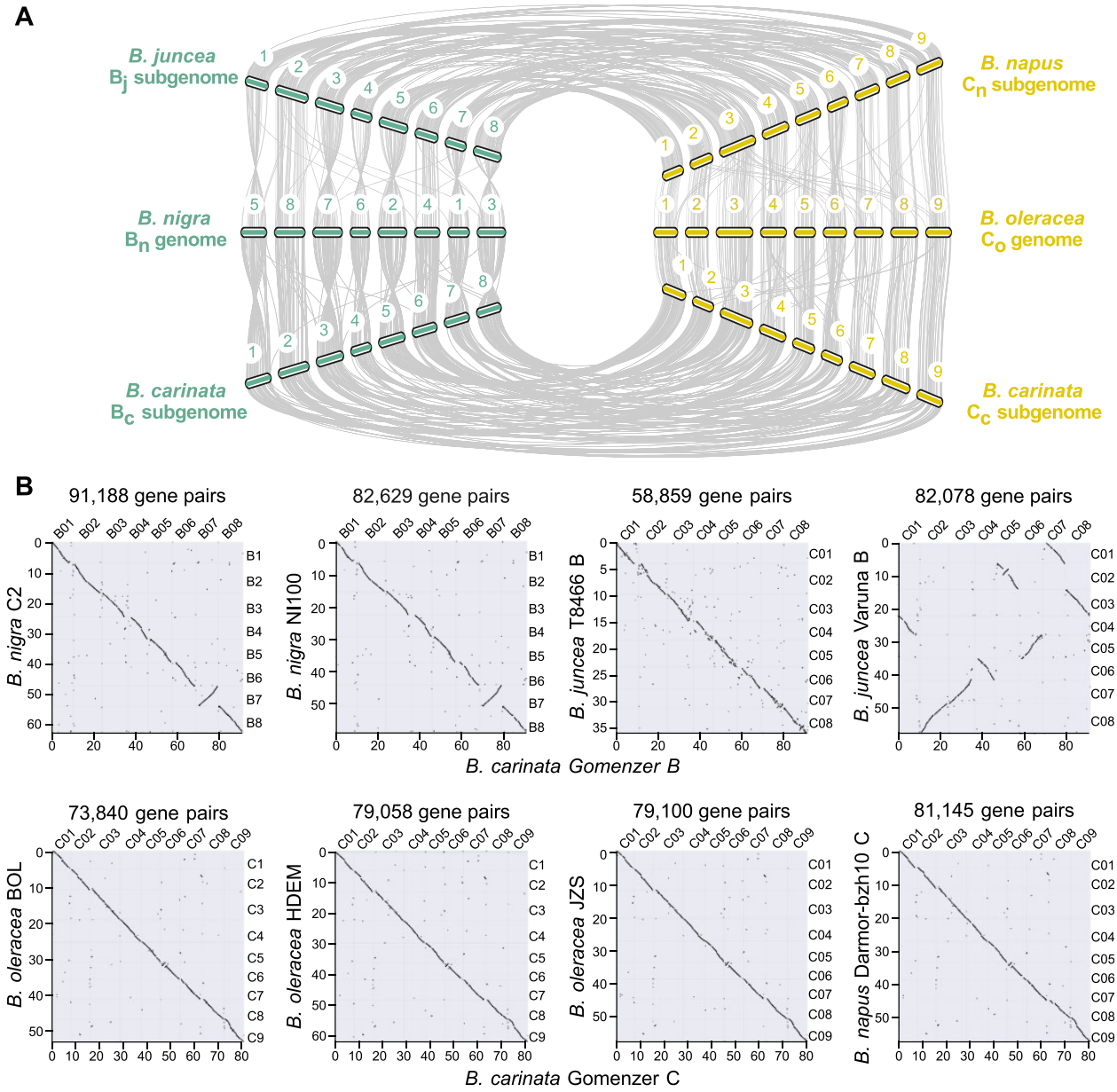


Figure 3 Genomic alignment between progenitor *Brassica* species (*B. rapa*, *B. nigra*, and *B. oleracea*) and the hybridized allotetraploid species (*B. carinata*, *B. juncea*, and *B. napus*). A, Visualization of macrosynteny among *B. carinata* subgenomes and shared genome species generated using MCscan. The karyotypes of the B genomes (Bj, Bn, and Bc) of the Triangle of U species are on the left and the karyotypes of the C genome are on the right. Lines connecting the chromosomes represent regions of syntenic gene arrangement between the progenitor genomes (Bn and Co) and their shared subgenomes in the allotetraploid species (Bc, Bj and Cc, Cn). B, Pairwise intergenomic comparison dot plots between the two *B. carinata* subgenomes and their respective shared Triangle of U genomes (Bn, Co, Bj, and Cn). The dots indicate the syntenic genes conserved as a block among the two genomes. Longer lines represent a high abundance of genomic collinearity.

whose accurate reconstruction can shed light on their effect on genome evolution (Veeckman et al., 2016). Our Gomenzer assembly has an LAI score of 12.23 for the whole genome, and the Bc and Cc subgenomes have LAI values of 11.78 and 12.71, respectively (Supplemental Data Set 8). The whole-genome LAI for Gomenzer is consistent with those of the other Triangle of U allotetraploid assemblies, which have average LAI values of 12.32 for *B. juncea* and 12.60 for *B. napus*.

The *B. carinata* genome has the largest gene content among the Triangle of U allotetraploids, containing 133,667 annotated gene models predicted to generate 143,117 transcripts (Figure 4A; Supplemental Data Set 9). A total of 105,596 genes, or 78.9% of the total annotated gene models, were supported by RNA-Seq alignments and the rest were supported by encoded Pfam domains or by homology according to BLAST. Each subgenome in *B. carinata* also contains more gene models than the respective shared genomes in the Triangle of U (Bn, Bj, Co, and Cn; Figure 4D). We identified asymmetry in gene content between the two subgenomes, with the larger Bc subgenome containing 54.1% of the

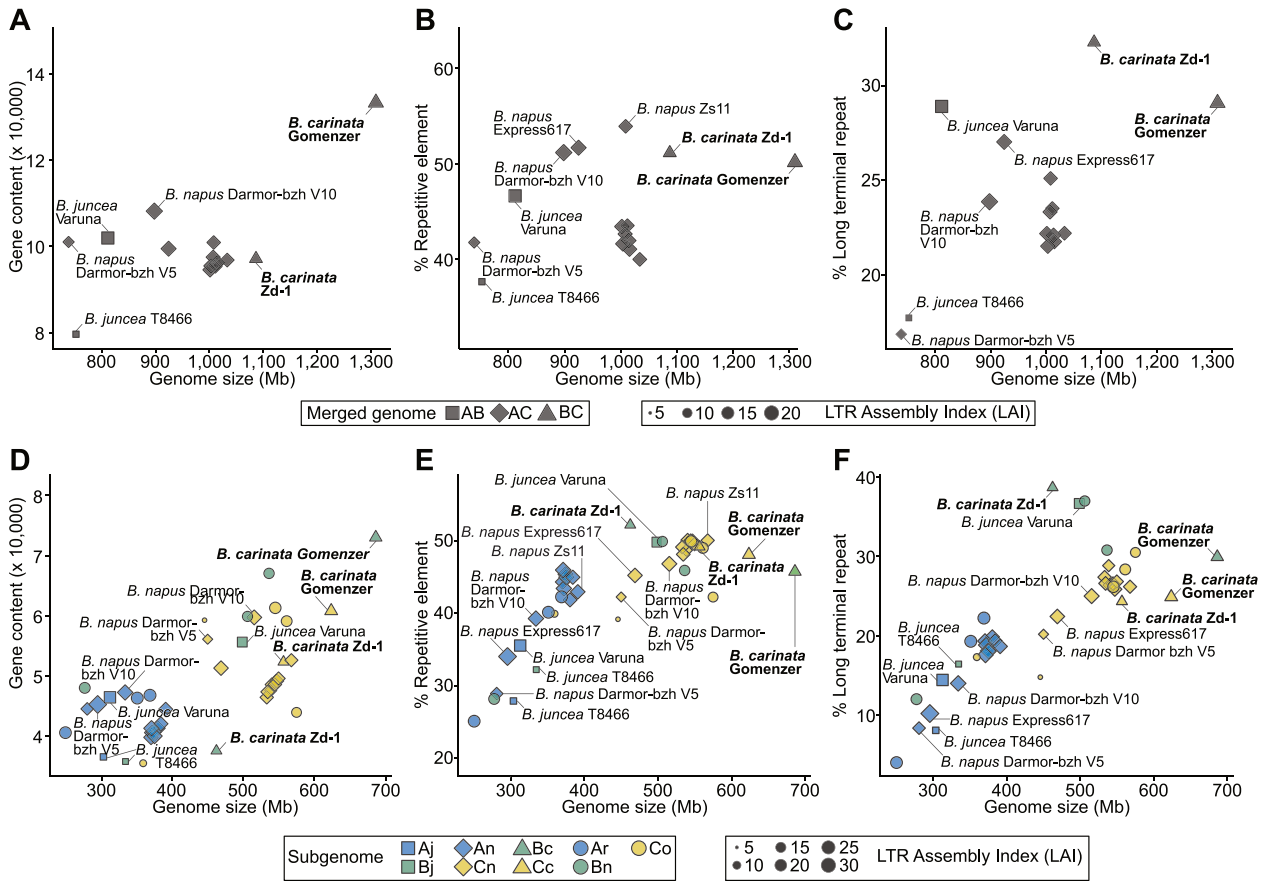


Figure 4 Dot plot comparisons of genome size and genetic features among assembled *Brassica* genomes. The upper parts (A, B, and C) compare whole-genome assemblies of the Triangle of U allotetraploids and the lower parts (D, E, and F) compare subgenome assemblies among all six species. The *x*-axes represent the sizes of the *Brassica* genome assemblies in Mbp, and the *y*-axes show either gene content (left), repetitive element content (middle), or LTR content (right). The shapes of the data points indicate the species and the sizes of the data points are scaled according to the assembly's LAI, with larger LAI values indicating a higher quality assembly. For the bottom, the data points for the three subgenomes are assigned as blue (A), green (B), or yellow (C). The three allotetraploid species are represented as unique shapes: square (*B. juncea*), diamond (*B. napus*), and triangle (*B. carinata*), with their progenitor species (*B. rapa*, *B. nigra*, and *B oleracea*) represented as circles. The *B. carinata* assemblies Gomenzer and Zd-1 are highlighted in bold, and other assemblies of interest are marked on each part.

annotated gene models (Bc: 72,373 Cc: 61,294; Supplemental Figure S4B, Supplemental Data Set 10). The Bc subgenome contains 12.6% exonic sequences, while the Cc subgenome only contains 8.7% exonic sequences. To assess whether the asymmetry in gene content between the two subgenomes could be a legacy of the parental genomes, we compared the average gene content of the progenitor assemblies. The *B. nigra* (Bn) assemblies had an average of 58,275 genes, and the *B. oleracea* (Co) assemblies had an average of 51,767 genes (Supplemental Data Set 8). Considering the average gene numbers of the Bc progenitor species genome assemblies, we cannot rule out that the higher gene content of the Bc subgenome had already been established at the time of hybridization.

Gene fractionation, or the process through which homoeologous subgenomes lose genes to return to the diploid state, was also examined in *B. carinata* (Renny-Byfield et al., 2017). The Triangle of U genomes are highly duplicated, with diploids having undergone an aggregate 36× multiplication (3× 2 × 2 × 3) and the allotetraploids having undergone an aggregate 72× multiplication (3 × 2 × 2 × 3 × 2) since the most recent common ancestor of all eudicots (Figure 5). Although the *B. Carinata* genome has undergone diploidization to some extent, a simple whole-genome comparison did not indicate substantial gene loss. We found that the Bc subgenome has retained 28,328 pairs of whole-genome duplication (WGD)-derived genes, the most among the Triangle of U allotetraploid subgenomes (Supplemental Data Set 11). A total of 78.3% of the Bc subgenome genes are WGD-derived, whereas only 49.85% of the Cc subgenome genes are WGD derived. Our RNA-Seq analysis resulted in high rates of RNA-Seq alignment (~82.1%; Supplemental Figure S5A) but lower rates (~78.4%; Supplemental Figure S5B) of assignment of expressed sequences to the chromosomes, likely due to ambiguities in read mapping caused by the high similarity between the two subgenomes. These results denote apparent retention of duplicated genes in *B. carinata* similar to that observed for many other paleopolyploidization events in plants (Bowers et al., 2003). Our results strengthen the hypothesis that the retention of duplicate genes in neopolyploids might give rise

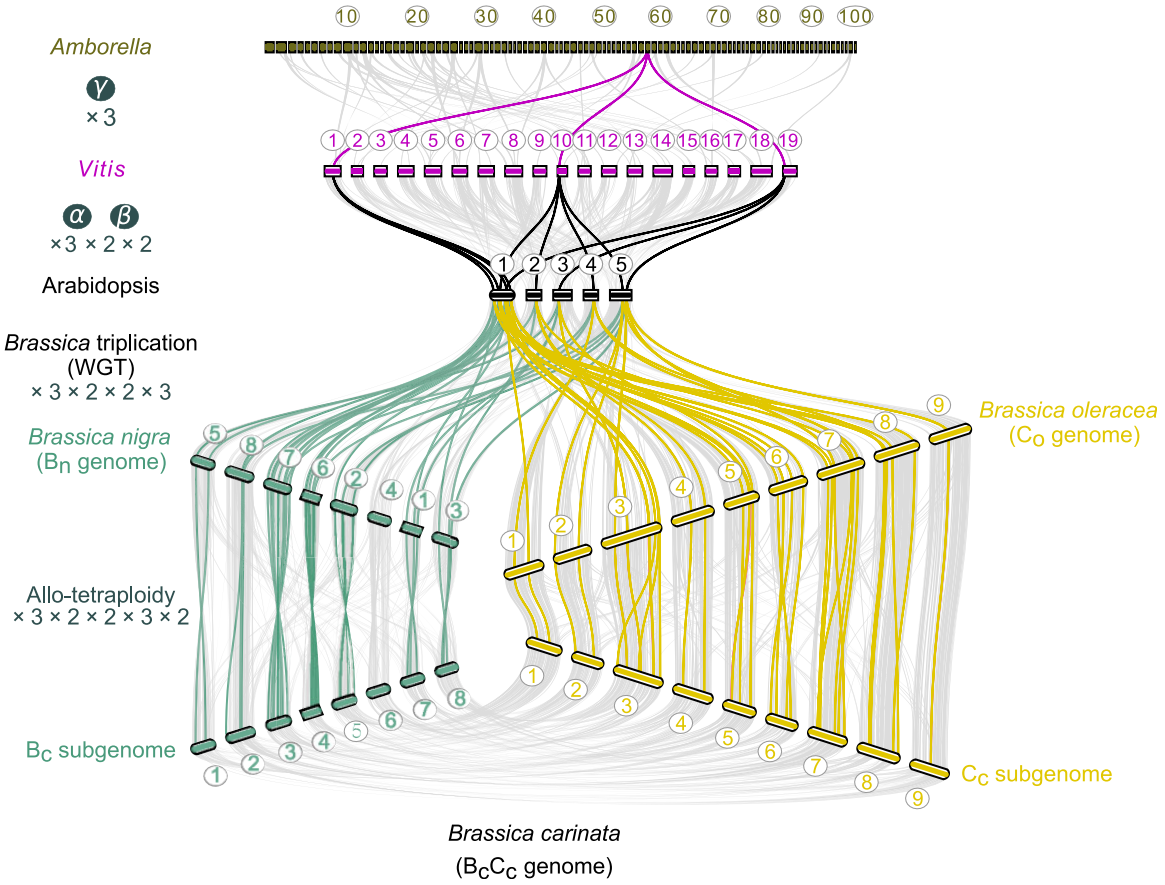


Figure 5 Representation of local syntenic regions in support of WGD events affecting *Brassica* species. Syntenic gene arrangements are shown for the genomes of *Amborella trichopoda*, *V. vinifera* (2*n* = 19), *A. thaliana* (2*n* = 5), *B. carinata*'s two progenitor species, *B. nigra* (Bn), *B. oleracea* (Co), and the Bc and Cc subgenomes of *B. carinata*. Lines connecting the species indicate conserved syntenic blocks containing more than 30 orthologs. The highlighted lines that radiate from the extant *A. trichopoda* genome represent the WGD-derived amplification of homoeologs retained in the rediploidized genomes of paleopolyploid species and in the more recently formed polyploid, *B. carinata*. The concept of this figure was adapted from Chalhoub et al. (2014).

to adaptive evolution through the subfunctionalization or neofunctionalization of redundant genes (Lynch and Conery, 2000) by providing the genetic plasticity required to generate novel traits such as the diversification of glucosinolates in the Brassicales (Edger et al., 2015; Qi et al., 2021).

The *B. carinata* Zd-1 genome assembly was published in 2021 (Song et al., 2021). Notably, our Gomenzer assembly shows superior quality metrics and is likely a more reliable reference genome. Our Gomenzer assembly was constructed using 57.7 Gbp more long-read sequencing data and has a scaffold N50 value that is 19.9 Mbp longer than that of the Zd-1 assembly (Supplemental Data Sets 2 and 6; Song et al., 2021). Our BUSCO analysis identified more complete BUSCOs in the Gomenzer assembly (Supplemental Figure S3) than the Zd-1 assembly. We found that the discrepancies are most apparent in the Bc subgenome, which contains 26.2% more complete BUSCOs in Gomenzer than in the Zd-1 Bc subgenome assembly. Our Gomenzer assembly is also a better reference in terms of the contiguity of repetitive sequences. The LAI value calculated for our Gomenzer assembly falls within the reference quality range of the LAI classification system ($10 \leqslant 12.23 < 20$), while that of the Zd-1 assembly lies within the draft quality range ($0 \leqslant 9.57 < 10$) (Supplemental Data Set 8; Ou et al., 2018). An intergenomic alignment of the two assemblies identified large-scale disruptions in collinearity, such as chromosomes that were truncated in the Zd-1 assembly (Zd1B06, Zd1B08, and Zd1B05), large inversions of chromosome segments (Zd1B03, Zd1C01, and Zd1C02), and a duplication of chromosome GomB06 (Supplemental Figure S6). Our 1.31-Gbp assembly is estimated to be 22.2% larger and contains 36,518 more gene models than the 1.09-Gbp Zd-1 assembly (Supplemental Data Sets 6 and 9). In line with our collinearity assessment, the Bc subgenome assemblies showed the most pronounced differences. The Gomenzer Bc and Cc subgenomes are 224.2 and 67.2 Mbp larger than the Zd-1 Bc and Cc subgenomes, respectively, and contain 34,903 and 9,015 more gene models (Supplemental Data Sets 6 and 9).

Our high-quality genome assembly refines estimates of the timing of Triangle of U hybridization events

We deduced that the hybridization event between *B. nigra* (Bn) and *B. oleracea* (Co) that led to the speciation of *B. carinata* occurred about 11,000–29,900 years ago based on our

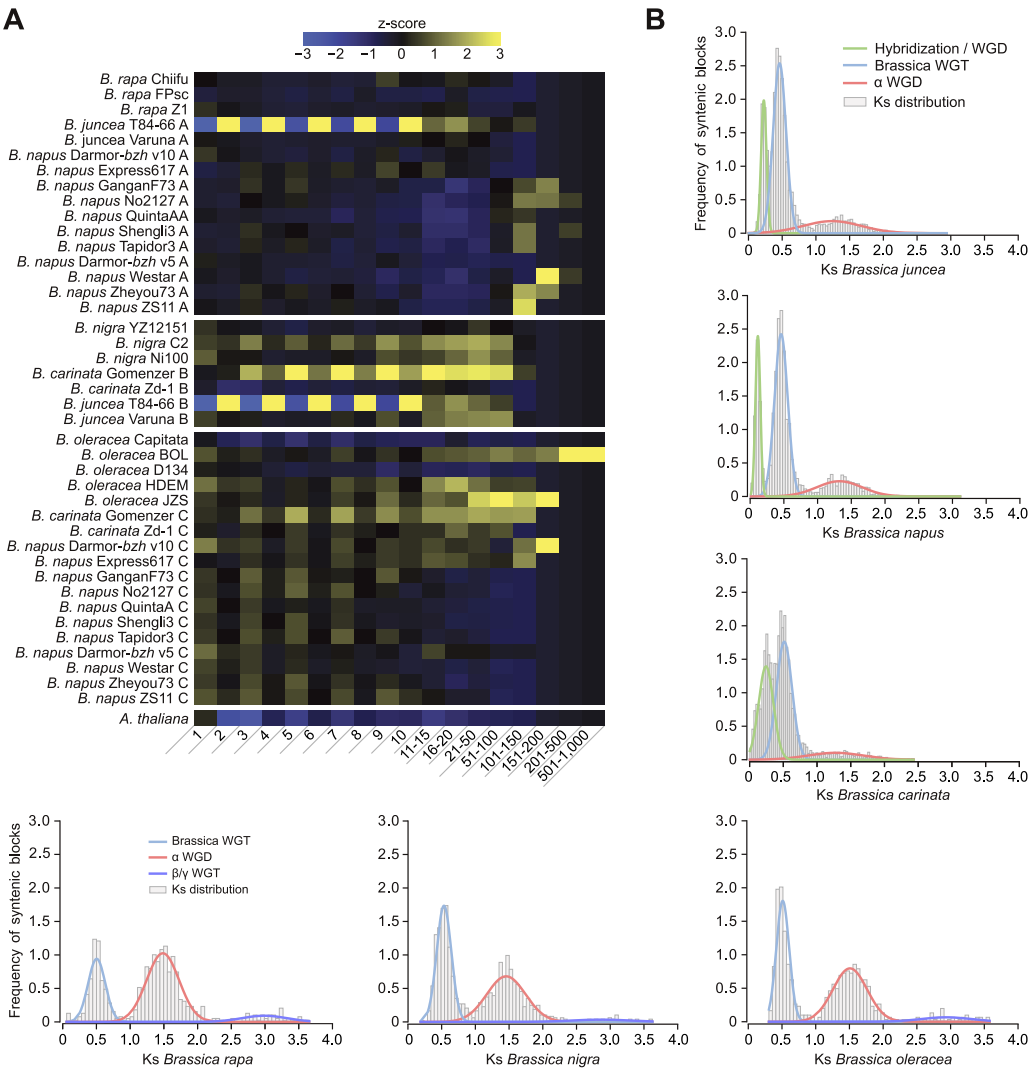


Figure 6 Gene families and WGD events. A, Heatmap depicting the relative number of genes belonging to gene families binned by the number of genes they contain. The *z*-scores increase with increasing relative gene count. The *y*-axis describes the genome assembly with A genomes in the top, B genomes in the middle, and C genomes in the bottom. The *x*-axis describes the binned gene families with the number of genes the binned gene families contain. B, Ks plots used in timing Triangle of U WGD events. Peaks of established events are highlighted, including the allotetraploid hybridization events, the Brassica-specific whole-genome triplication (WGT), the α/β WGD events, and the γ WGT.

synonymous divergence (K_s) rate distribution (K_s plot) analysis (Figure 6B). Our estimate is more recent than the estimate of 41,000–45,000 years ago provided by Song et al. (2021). As for the other *Brassica* allotetraploids, the hybridization events leading to the speciation of *B. juncea* and *B. napus* occurred an estimated 11,000–28,600 and 5,600–7,000 years ago, respectively. Our estimated time frame for *B. juncea* is more recent than the initial estimate of 39,000–55,000 years ago (Yang et al., 2016) and overlaps with the more recently calculated K_s plot estimate of 8,000–14,000 years ago (Kang et al., 2021). Our hybridization estimate for *B. napus* is also more recent than both the initial estimate of 7,500–12,500 years ago (Chalhoub et al., 2014) and the estimate of 38,000–51,000 years ago (Yang et al., 2016).

We then constructed a maximum-likelihood phylogeny of the Triangle of U genomes to resolve the origins of the Bc and Cc subgenomes (Figure 7A). According to our fossil clock calibration, the Bc subgenome diverged from the *B. juncea* Bj subgenome ~5.44 million years ago (MYA) and the last common ancestor of the two diverged from the extant *B. nigra* (Bn) genomes ~6.61 MYA (Figure 7A). These results imply that *B. carinata* (BcCc) and *B. juncea* (AjBj) arose from the same B progenitor lineage. On the other hand, the Cc subgenome diverged from the *B. oleracea* (Co) kale-type lineage (BOL) ~2.88 MYA (Figure 7A; Parkin et al., 2014). The *B. napus* (AnCn) Cn subgenomes are rooted in the cabbage-type *B. oleracea* (Co) lineage with a more recent divergence estimate of ~2.63 MYA (Figure 7A; Liu et al., 2014; Lv et al., 2020; Cai et al., 2020). Thus, *B. carinata* (BcCc) and *B. napus* (AnCn) appear to have arisen from distinct *B. oleracea* lineages. Furthermore, the Bc subgenome is more diverged from the extant *B. nigra* (Bn) genome assemblies than the Cc subgenome is from the extant *B. oleracea* (Co) genome assemblies (Figure 1).

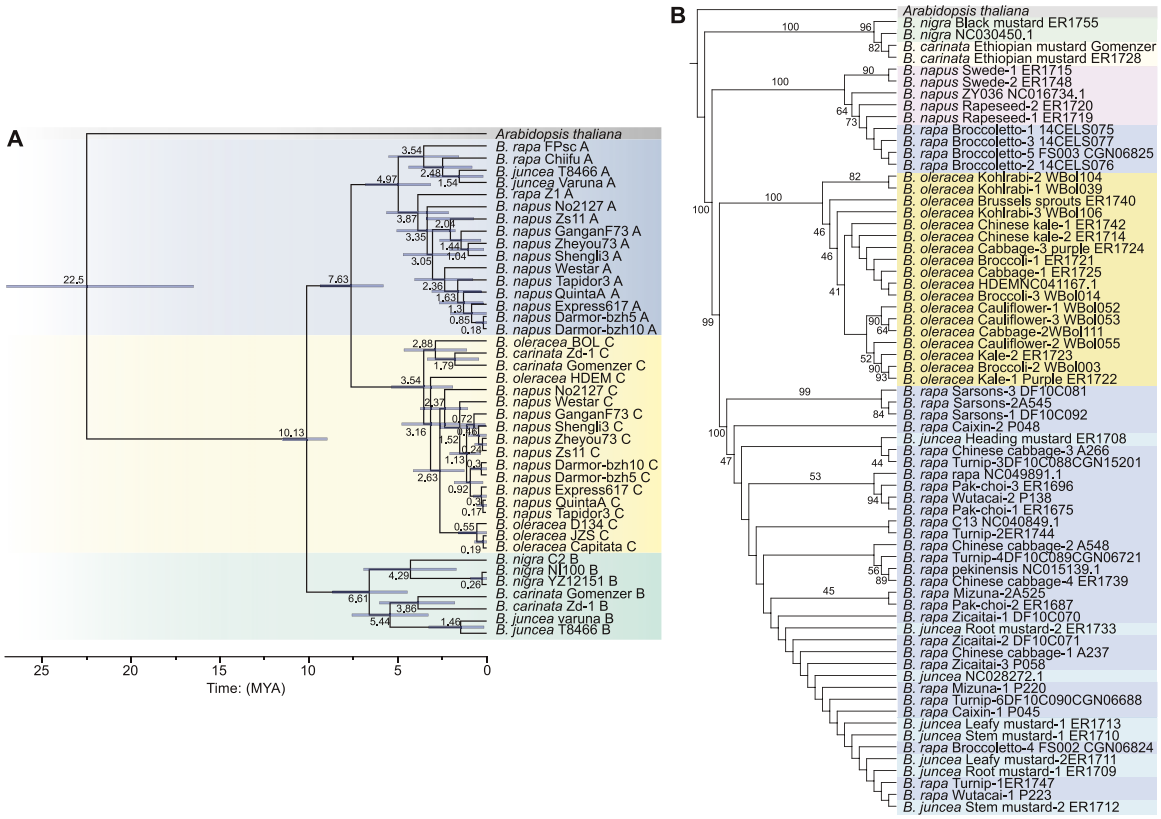


Figure 7 Evolutionary relationships among Triangle of U genomes and divergence time estimates. A, Time-calibrated phylogenetic tree of individual Triangle of U genomes with Arabidopsis as the outgroup. The divergence time estimates (MYA) are placed at each node with the uncertainty bar. B, Phylogenetic tree of the Triangle of U chloroplast genomes using the *Arabidopsis* chloroplast genome as the outgroup. Bootstrap values are shown at each node.

To determine the maternal progenitor of *B. carinata*, we assembled the chloroplast genome from our sequencing data and used the concatenated coding sequence (CDS) matrix to infer phylogenetic relationships with other Triangle of U chloroplast genome assemblies (Supplemental Figure S7). Identification of a distinct clade comprising *B. nigra* (Bn) and *B. carinata* (BcCc) allowed us to designate *B. nigra* as the maternal progenitor (Figure 7B), as supported by a previous study (Xue et al., 2020). Similarly, *B. napus* (AnCn) and *B. juncea* (AjBj) cluster independently with *B. rapa* (Ar), allowing us to designate *B. rapa* as the maternal progenitor of both of those species, as reported in other studies (Li et al., 2017; Kim et al., 2018; Xue et al., 2020). These observations are consistent with previous work suggesting a turnip-type *B. rapa* (ssp. *rapa*) as an ancestor of *B. napus* (Lu et al., 2019) and a yellow sarson-type *B. rapa* (ssp. *trilocularis*) as a parent of *B. juncea* (Yang et al., 2016). Our finding that *B. napus* is a more recent allotetraploid than *B. juncea* also supports divergence of the maternal genotypes that participated in these hybridization events.

The *B. carinata* genome is the largest among the Triangle of U

Our 1.31-Gbp *B. carinata* Gomenzer genome assembly is larger than those of the other two Triangle of U allotetraploids, with the *B. napus* (AnCn) assemblies having an average size of 982.0 Mbp and the *B. juncea* (AjBj) genome assemblies having an average size of 888.6 Mbp (Supplemental Data Set 6). However, assembly quality must be considered when comparing genome assembly sizes, as lower quality assemblies tend to collapse repetitive regions and thus underestimate the actual genome size (Veeckman et al., 2016). As our Gomenzer assembly has the largest N50 value among the Triangle of U genome assemblies, the size of the *B. carinata* genome assembly reflects its high contiguity in addition to true biological differences among the Triangle of U species.

An asymmetry in size was identified between the Bc and Cc subgenomes, as the apparent size of the Bc subgenome (686.6 Mbp) is 9.1% larger than that of the Cc subgenome (623.9 Mbp) (Supplemental Figure S4D). We compared the assembly sizes of the subgenomes with the genome size estimates for the extant progenitor species to gain insights into whether the observed asymmetry in size could have arisen from differences in the parental genomes. As our phylogenetic analysis did not resolve the *B. nigra* (Bn) lineage from which the Bc subgenome arose, we then referred to the Ni100 and C2 long-read genome assemblies. The flow cytometry-based size estimates for the Ni100 and C2 genomes are 570 and 607.8 Mbp, respectively (Perumal

et al., 2020). The Bc subgenome is ~14.2% larger than the average of the Ni100 and C2 assemblies, implying an expansion in the size of this genome relative to those of the modern *B. nigra* (Bn) genomes. Our phylogeny did reveal that the Cc subgenome is in a distinct clade together with the *B. oleracea* var. Capitata BOL assembly, which was estimated at 648 Mbp using *k*-mer distribution (Parkin et al., 2014). The 623.9-Mbp Cc subgenome is 3.7% smaller, implying a contraction in size compared to the most closely related *B. oleracea* (Co) genome. These results reflect the earlier divergence of the Bc subgenome from the extant *B. nigra* genomes and reveal a Bc-biased expansion in genome size following the hybridization event.

TE propagation facilitated the increased size of the *B. carinata* genome

TEs make up a large portion of plant genomes and play a crucial role in their structural and functional evolution (Waminal et al., 2018; Anderson et al., 2019; Catlin and Josephs, 2022). Our repeat analysis identified 663.5 Mbp of repetitive DNA sequences making up 50.7% of the whole-genome assembly (Supplemental Data Set 12). Our Gomenzer assembly contains the greatest amount of repetitive sequence among the Triangle of U species, and both of its subgenomes contain more repetitive sequence than their respective shared genomes do (Bn, Bj, Co, and Cn). As with genome size, the observed abundance of TEs could be partially due to the improved quality of the presented assembly, which would provide a more accurate representation of the TE content in *Brassica* genomes. Still, the average of the Gomenzer and Zd-1 *B. carinata* assemblies reveal the higher average TE contents of the Bc and Cc subgenomes ($\bar{x}$ = 49.9% and $\bar{x}$ = 50.61%, respectively) relative to their respective shared genomes in the Triangle of U (Bn, Bj, Co, and Cn; Supplemental Data Set 12).

An asymmetric distribution of repetitive DNA content was also identified between the Bc and Cc subgenomes. Although the two subgenomes share similar TE content overall (Bc: 47.3%, Cc: 48.7%; Supplemental Data Set 12), the distributions of TEs along the chromosomes differ remarkably. TEs in the Bc subgenome exhibit typical pericentromeric and telomeric enrichment, but those in the Cc genome are more evenly distributed across the apparently more fractionated Cc chromosomes (Figure 8, A

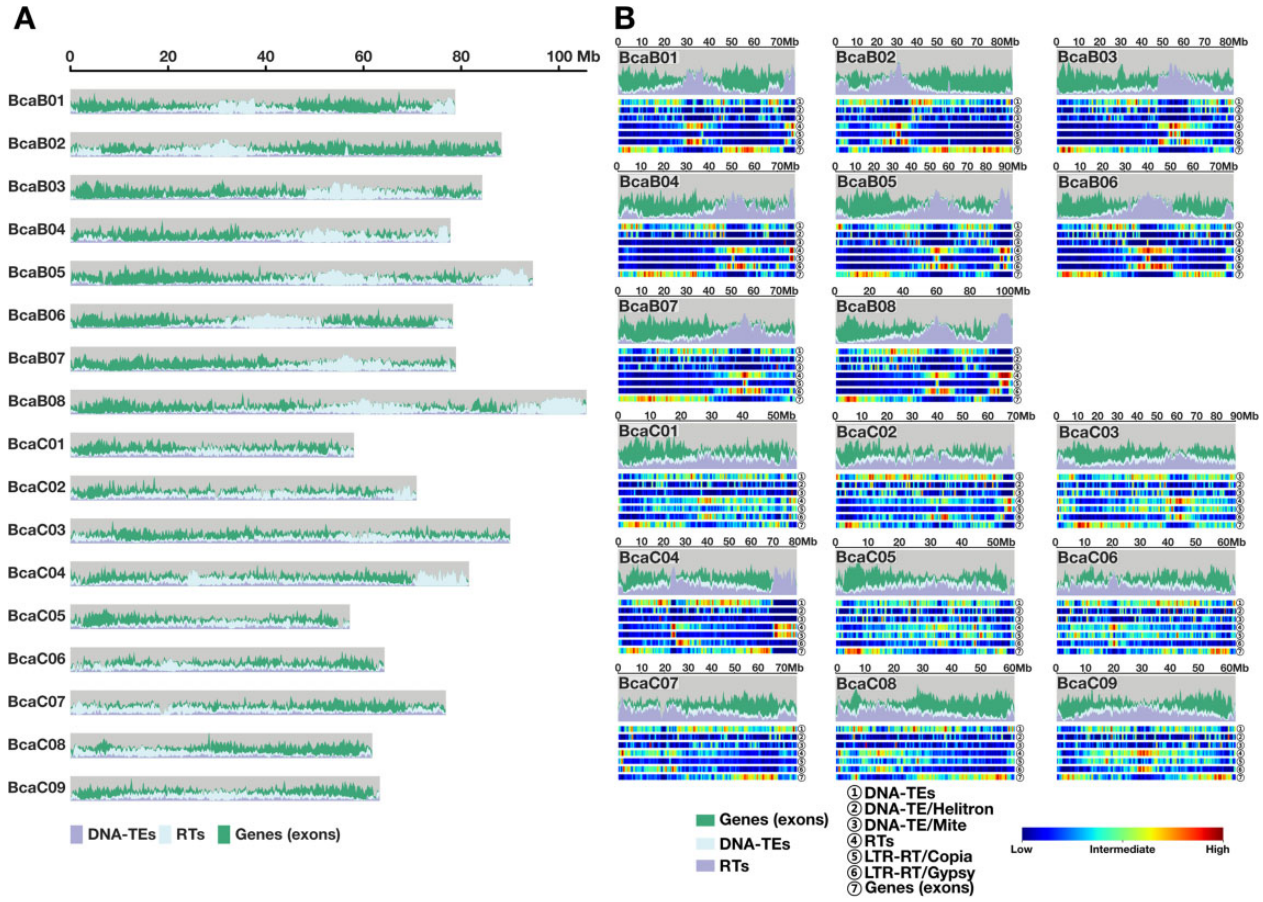


Figure 8 Genomic landscape of *B. carinata* chromosomes. A, Distribution of genomic features for each chromosome scaled according to their assembled length. The area chart indicates the DNA transposon (DNA-TE) content, retrotransposon (RT) content, and exon content. B, Detailed genomic landscapes for each chromosome with distributions of gene exons, DNA-TEs, and RTs are shown in the area chart and the heat map tracks.

and B). The Cc subgenome also exhibits a higher overall repeat content (64.9%) than the Bc subgenome does (59.8%), indicating a higher level of retention of tandem repeats, such as microsatellite sequences (Supplemental Figure S4C). Tandem repeats are both substrates and byproducts of ectopic recombination, and their presence is closely related to chromosomal rearrangements (Waminal et al., 2021).

We then compared TE classes to identify repetitive elements or trends specific to the *B. carinata* genome. As with other plant genomes, the LTR retrotransposons contribute the largest fraction of repetitive elements, covering 29.1% of the whole genome (Supplemental Data Set 12). The two major LTR superfamilies, *Copia* and *Gypsy*, cover 8.8% and 14.2% of the genome, respectively (Supplemental Data Set 12). The Gomenzer and Zd-1 assemblies show that the *B. carinata* genome has the highest average LTR content among the Triangle of U allotetraploids (Figure 4C). An asymmetry in LTR content was observed with LTR sequences comprising 29.9% and 26.0% of the Bc and Cc subgenomes, respectively (Supplemental Figure S4D). These results are consistent with the LTR content of the progenitor species, *B. nigra* (Bn) and *B. oleracea* (Co), whose genome assemblies have an average LTR content of 26.6% and 23.4%, respectively (Figure 4F; Supplemental Data Set 12). Our TE analysis also identified asymmetric LTR distribution in the other two allotetraploids, *B. napus* (AnCn) and *B. juncea* (AjBj, Supplemental Figure S4D). The larger Cn and Bj subgenomes also comprise a higher average LTR proportion ($\bar{x}$ = 25.7% and $\bar{x}$ = 26.6%, respectively) than do the smaller An and Aj subgenomes (17.6% and 11.3%, respectively).

From the whole-genome perspective, we observed similar LTR dynamics among the Triangle of U allotetraploid genomes (Supplemental Figure S8A). However, the more LTR-rich Bc subgenome has experienced more recent bursts in *Gypsy* element activity than the *B. nigra* (Bn) genomes (Supplemental Figure S8B) and has also accumulated more *Gypsy* elements than the Cc subgenome (Supplemental Data Set 12). Similarly, the Cc subgenome has experienced more recent bursts in both *Copia* and *Gypsy* elements compared to *B. oleracea* (Co, Supplemental Figure S8C).

Helitrons are the most prevalent DNA transposon in our Gomenzer assembly and represent 107.8 Mbp or 13.2% of the genome assembly (Supplemental Data Set 12). The *Helitron* content in plant genomes is generally lower and more stable than that of retrotransposons, and *Helitron*-related data could thus be more informative of plant genome features (Hu et al., 2019). Therefore, we investigated the *Helitron* content of the *B. carinata* genome relative to that of the other Triangle of U genomes and found that the Bc subgenome contains more *Helitrons* per Mbp than the Cc subgenome and the other Triangle of U B genomes do (Bn and Bj; Supplemental Data Set 12).

Gene family expansion facilitated adaptation during *B. carinata* domestication

Gene duplications have long been known to contribute to the genesis of novel traits, phenotypic variation, and adaptation to changes in the environment (Ohno, 1970; Rizzon et al., 2006; Kliebenstein, 2008). Thus, we investigated orthogroups among the Triangle of U species to investigate changes in gene family counts that might have spurred genetic novelty and provided the basis for adaptation in *B. carinata*. A total of 97.8% of the gene models annotated in the *B. carinata* genome were assigned to orthogroups (Supplemental Data Set 9). *Brassica carinata* was included in 88.1% of all of these orthogroups, the most among the Triangle of U genome assemblies. A total of 416 of these orthogroups, which cumulatively comprise 1,092 genes, were specific to *B. carinata*. The Bc subgenome contains more species-specific orthogroups than the Cc subgenome does (Bc: 254, Cc: 162). In line with expectations, we found that the more gene-rich Bc subgenome contains more genes per orthogroup than Cc subgenome does (Figure 6A). The numbers of genes in both subgenomes were higher than the average across the Triangle of U species complex, which indicates lower levels of gene loss and possible gene expansions of these gene families.

The retention of successive duplicated genes in multiple pathways might have contributed to the improved environmental adaptability and novel trait development observed in *B. carinata* relative to other *Brassica* species. Intensive Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway analysis of expanded gene families revealed enrichment of pathways involved in photosynthesis, suberin and wax biosynthesis, hormone signal transduction, plant–pathogen interactions, diterpenoid biosynthesis, ubiquitin-mediated proteolysis, and ascorbate and aldarate metabolism (Supplemental Figure S9A). The expansion of these gene families suggests that *B. carinata* might have developed effective adaptive responses to biotic and abiotic stresses. For example, the expansion of genes involved in wax biosynthesis has likely facilitated the thick layer of cuticular wax that coats the stems and leaves of the plant. The accumulation of cuticular wax helps reduce transpiration and provides photoprotection for drought-tolerant plants to maintain stomatal conductance and photosynthesis during drought stress (Guo et al., 2016). These and other genes involved in the expression of agriculturally important traits could serve as potential targets for future engineering and breeding efforts to improve climate resilience in other crops, especially other Triangle of U species (Wang et al., 2020).

The contraction of some gene families also provides insight into the evolution of a species, as gene loss is another recognized driver of evolutionary novelty and adaptation (Meyer and Purugganan, 2013; Helsen et al., 2020; Monroe et al., 2021). Although the number of expanded gene families in *B. carinata* is similar to that of *B. napus* (AnCn), 14.9-fold more gene families were contracted in the *B. carinata* genome than in *B. napus* (Figure 6A). Gene ontology (GO)

analysis of contracted gene families revealed enrichments for gene families involved in the "response to ATP," "toxin response and catabolism," and "glutathione metabolism" categories (Supplemental Figure S9D). Interestingly, gene families involved in cell wall functions and components, including "adherens junction assembly," "xyloglucan biosynthetic process," and "fucosylation," were contracted. Importantly, the removal of xyloglucans has been shown to loosen cell walls and promote cell expansion and plant growth (Park et al., 2003, 2004; Naoumkina et al., 2017). The possible functional relevance of the contraction of this and the other gene families could be revealed in future studies.

GO analysis of expanded gene families revealed significant enrichments for transcriptional regulation and the regulation of biosynthetic processes (Supplemental Figure S9C). Transcription factor (TF) family expansions are closely related to the evolution of novel traits (Lang et al., 2010) and have been correlated with the generation of critical processes such as the WGD-derived expansion of MADS-box TFs and the origination of flowers (Li et al., 2015). We identified 9,599 TF-encoding genes in *B. carinata*, with 5,494 and 4,113 in the Bc and Cc subgenomes, respectively (Supplemental Data Set 14). MYB, NAC, and basic helix-loop-helix (bHLH) represent the most expanded TF families in the *B. carinata* genome.

Asymmetric distribution of genomic features between the Bc and Cc subgenomes shaped by Bc subgenome bias

As the *B. carinata* genome might be in the early stages of diploidization, 82.6% of its genes have maintained collinearity with those in at least one of the respective progenitor assemblies (Bn and Co; Figure 3; Supplemental Data Set 15). The Bc subgenome maintains the most collinearity with the *B. nigra* Ni100 assembly (72.2%) and the Cc subgenome shares the most colinear genes with the *B. oleracea* HDEM assembly (74.3%). While there is no statistical difference between the subgenomes in the number of collinear genes they have maintained with their respective progenitors, the Bc subgenome shares more collinearity blocks with shared genomes in the Triangle of U than the Cc subgenome does (analysis of variance (ANOVA), $P = 3.0 \times 10^{-4}$; Supplemental Data Set 16). In conjunction with the results obtained from our repeat content analysis, these data indicate that the Bc subgenome has undergone fewer genomic rearrangements than the Cc subgenome, which exhibits high levels of fractionation. These findings are also supported by the differences in transposon distributions previously observed between the respective progenitors (Parkin et al., 2014; Perumal et al., 2020). We also hypothesize that the higher degree of gene retention in the Bc subgenome might be correlated to genome dominance, in light of the determination that a dominant genome has experienced less fractionation (Schnable et al., 2011), as will be further discussed below.

Asymmetric retention of genes following HE the Bc and Cc subgenomes

Homoeologous exchanges (HEs), which are reciprocal exchanges of homoeologous chromosomal regions between subgenomes, occur due to crossovers between these regions during meiosis (Gaeta and Chris Pires, 2010). Although the HE mechanism is reciprocal, the two recombinant chromatids resulting from an initial HE event are normally not segregated into the same resulting gamete and are often detected as unbalanced exchanges or duplication–deletion events, wherein one homoeolog is entirely replaced with a duplicate of its counterpart from the other subgenome. Unbalanced exchanges are mechanistically the most frequently observed HE (Mason and Wendel, 2020) and are commonly observed in the established allopolyploid, *B. napus* (Chalhoub et al., 2014). On the other hand, balanced exchanges, wherein the homoeologs are reciprocally exchanged, may also become fixed following self-pollination and selection but are more challenging to detect using sequencing-based methods.

To investigate HE events in the *B. carinata* genome, we identified 32,443 colinear homoeologous gene quartets among the extant genomes of the diploid progenitors and the *B. carinata* subgenomes (Bn–Bc–Cc–Co, Supplemental Data Set 17). Based on relative sequence divergence ($K_s < 0.01$), we found evidence of 28 unbalanced HE events. Among these, 13 replaced the Bc homoeolog with the Cc homoeolog (Cc to Bc translocation), while 15 were classified as unknown due to the negligible divergence between the homoeologous pair ($K_s < 0.01$) and higher divergence from the progenitors. While we did not find evidence of balanced HEs in our Gomenzer genome assembly, we identified three balanced HEs in the *B. carinata* Zd-1 genome assembly and 244 unbalanced HEs among the 21,304 quartets (Supplemental Data Set 17). The same subgenomic bias in unbalanced HEs was observed in the *B. carinata* Zd-1 assembly, with 24 Bc to Cc exchanges and 68 Cc to Bc exchanges. However, the subgenomic origins of 152 HEs could not be assigned. As for the other two Triangle of U allotetraploids, our investigation revealed intraspecific variations in subgenomic bias for unbalanced HEs. The frequencies of HE events in the genome assemblies of the two *B. juncea* (AjBj) varieties var. Tumida T84-66 and Varuna, differed significantly (Yang et al., 2016; Paritosh et al., 2021). A total of 321 HEs out of 32,044 quartets were identified in the T84-66 genome assembly and 5 HEs out of 19,186 quartets were identified in the Varuna genome assembly (Supplemental Data Set 17). More Aj to Bj (26) than Bj to Aj (16) HEs were identified in the T84-66 genome, and 120 HEs were classified as unknown. Additionally, while we found three balanced HEs in the T84-66 genome, none were found in the Varuna genome assembly. One of the three unbalanced HEs observed in the Varuna genome assembly could be attributed to a Bj to Aj translocation, while the other four were classified as unknown. The *B. napus* (AnCn) assemblies, on average, contained the largest numbers of identified quartets (Ar–An–Cn–Co, $\bar{x} = 32{,}759$) among the Triangle of U species, which is evidence of unbalanced HE

events ($\bar{x}$ = 592.6) and balanced HEs ($\bar{x}$ = 31.9). An average of 61.1% ($\bar{x}$ = 362.4) of the unbalanced HEs could not be assigned to a subgenome, which makes determination of subgenomic bias difficult. Seven *B. napus* genome assemblies show an An to Cn bias, whereas four exhibit a Cn to An bias (Supplemental Data Set 17).

Lastly, homoeologs that have participated in HEs can eventually be removed through genome fractionation, leaving only one duplicated gene resulting from a HE. In order to investigate these instances, we constructed gene triplets comprising one retained homoeolog in a colinear block with the orthologs in both progenitor genomes (i.e. Bn–Bc–Co or Bn–Cc–Co, Supplemental Data Set 17). Analysis of these triplets allowed identification of homoeologous transfers (HTs), or instances wherein a homoeolog was removed from its original subgenome after transfer to the new subgenome in either a balanced or unbalanced manner. Among the 10,098 gene triplets in the *B. carinata* genome, we found evidence for 75 HTs, 86.7% of which had been transferred from the Cc subgenome to the Bc subgenome before being deleted from the Cc subgenome. We observed the same trend in our HT analysis in which Cc homoeologs had replaced Bc homoeologs. Similarly, the *B. carinata* Zd-1 genome assembly revealed the same bias, with 66.1% of the 174 HTs losing the Cc homoeolog after it was transferred to the Bc subgenome. We identified 156 and 2 HTs in the *B. juncea* T84-66 and Varuna genome assemblies, respectively (Supplemental Data Set 17). Both genomes showed the same pattern of subgenomic bias, with an average of 81.1% of HTs losing the Aj homoeolog after it had been transferred to the Bj subgenome. The *B. napus* (AnCn) genome assemblies showed intraspecific variations in HT bias, with seven and four genome assemblies retaining more An and Cn subgenome homoeologs after HTs, respectively. With respect to balanced and unbalanced HEs, as well as HTs, averages of 262, 163, and 761 HEs were detected in the *B. carinata* (BcCc), *B. juncea* (AjBj), and *B. napus* (AnCn) genomes, respectively.

Biased expression and selection pressures between homoeologs reveal Bc subgenome dominance

Higher transcript expression of homoeologs from one subgenome over those of another is one of the signatures of subgenome dominance. Homoeolog expression dominance has been detected in natural and resynthesized lines of *B. napus* (AnCn) with a bias for either the An subgenome (Wu et al., 2018; Li et al., 2020) or the Cn subgenome (Bird et al., 2020), but not in *B. juncea* (Yang et al., 2016). Therefore, we directly compared gene expression between colinear homoeologous gene pairs (Bc compared with Cc) using our RNA-Seq data. The majority of homoeologs (60.4%) showed similar expression patterns; however, an average of 22.6% of the homoeologs displayed Bc subgenome bias and 18.3% showed Cc subgenome bias (Supplemental Figure S10A; Supplemental Data Set 19). This bias was evident across all tissue types and was most prominent in floral tissues (Bc: 22.1%, Cc: 16.7%) and during heat stress (Bc: 23.44%, Cc: 18.1%). Homoeolog expression dominance was also most pronounced in floral tissue in *B. napus* (AnCn) with a bias for the An subgenome (Li et al., 2020).

Comparing sequence evolution rates of *B. carinata* homoeologs with their respective orthologs in the progenitor species revealed a higher average nonsynonymous to synonymous divergence ratio ($\omega = K_a/K_s$ or d_N/d_S) for the Cc homoeologs, indicating a greater degree of sequence evolution compared to the Bc homoeologs (Supplemental Figure S10B). Highly expressed genes tend to encode more slowly evolving proteins (Pál et al., 2001; Alvarez-Ponce, 2014), which might explain our observation. Thus, we compared the K_a/K_s ratios of highly expressed genes from the Bc and Cc subgenomes to test this possibility. The ω values were still higher for the Cc subgenome homoeologs across RNA-Seq samples (Mann–Whitney *U* test, $P = 8.374 \times 10^{-54}$), indicating that the higher rates of evolution of proteins encoded by the Cc subgenome are not due to their low expression levels. Our results thus indicate that the Cc subgenome is subject to weaker purifying selection pressures, a higher rate of adaptation than the Bc subgenome, or both, which is consistent with our overall findings of asymmetric subgenome evolution and Bc subgenome dominance in *B. carinata*.

Redundant gene evolution and selection pressures on small-scale duplications among *Brassica* genomes

Duplicated genes can diversify to take on new functional roles (Tang et al., 2010; Wu and Qi, 2010) and serve as substrates for adaptation to stressful environments (Alix et al., 2017). We observed a greater degree of sequence evolution for the *B. carinata* WGD-derived genes compared to the other allotetraploids. The *B. carinata* WGD-derived genes have a higher ω value ($\bar{x}$ = 0.31, σ = 0.28) than the other allotetraploids do. *Brassica napus* (AnCn) and *B. juncea* (AjBj) exhibit average ω values of 0.28 (σ = 0.27) and 0.26 (σ = 0.19), respectively (Supplemental Data Set 20). This increased rate of sequence divergence between *B. carinata* homoeologs relative to those in the other tetraploids suggests a greater degree of relaxation of purifying selection or greater adaptation in *B. carinata* that might have facilitated the high resilience of this species to environmental stresses.

Small-scale duplications, or duplicated genes derived from events other than WGD or HE, were also assessed in the *B. carinata* genome to identify 3,645 tandem, 5,594 proximal, 11,264 transposed, and 47,157 dispersed pairs of paralogs (Supplemental Data Set 11). Although the Bc subgenome contains more WGD-derived genes than the Cc subgenome, both subgenomes contain similar proportions of tandem and proximal duplicates (PDs) relative to the total number of genes. Our Gomenzer assembly contains the smallest proportion of tandem duplicates (TDs), or duplications that are linearly adjacent to each other, relative to its total gene content (2.7%) among the Triangle of U allotetraploids. These results indicate lower propagation or retention rates, or both, for

TDs in *B. carinata*. For all six species, the ω values of the TDs are about 1.8-fold larger than the ω values of the homoeologs, indicating higher diversifying positive selection acting on TDs (Supplemental Data Set 20). Although *B. napus* (AnCn) has the highest ω value (0.53) for TDs among the Triangle of U species, the ω value (0.39) for the *B. carinata* TDs is still larger than those of the respective progenitor species (0.33). One of the most overrepresented GO terms among the TDs was "DNA methylation on cytosine within a CG sequence" (Supplemental Figure S9E), which is associated with epigenetic control of stress adaptation through reprogramming of the transcriptome and altering genome stability to enhance resilience to stress (Tirnaz and Batley, 2019). Thus, because the TDs in the *B. carinata* genome appear to be under less purifying selection than those in the progenitor species genomes (Bn and Co), their gene variants might help *B. carinata* adapt to changes in its environment.

We identified 1.5-fold more PDs than TDs in the *B. carinata* genome, which contains more PDs than the other Triangle of U allotetraploids (Supplemental Data Set 11). Although we found similar sequence evolution rates for PDs and TDs, in agreement with the current literature (Supplemental Data Set 20; Qiao et al., 2019), we also found that PD sequences diversify more rapidly in *B. carinata* ($\omega = 0.52$) than in the other five Triangle of U species. This higher ω value appears to be due to an unusually low mean K_s value for *B. carinata* PDs (0.30) compared to the average K_s value of 0.52 for all six species, suggesting that the *B. carinata* PDs are relatively young. Gene duplication is often followed by a short period of accelerated protein evolution (Pegueroles et al., 2013). The GO terms enriched among PDs included "lipid metabolic processes" and "interspecies interactions," including those with viral pathogens (Supplemental Figure S9F). The ω values of the more distant transposed duplicates and dispersed duplicates lay at the median of the five duplication modes considered. However, the K_s values of these distant duplications were up to 3.5-fold greater than the WGD K_s values, indicating that more time has passed since their duplication.

Genes encoding key FA synthesis-related enzymes display Bc subgenome dominant expression

We then investigated the gene expansion and transcript abundance of agronomically important genes that have been targeted in *Brassica* improvement efforts to test whether the dominant Bc subgenome shows higher expression levels as the subgenome dominance hypothesis describes (Freeling et al., 2012). Genes in the fatty acid (FA) biosynthesis and elongation pathways have undergone notable expansions in *B. carinata* (Supplemental Data Set 21).

In higher plants, *de novo* FA synthesis begins with the conversion of glycolysis-derived pyruvate to acetyl-CoA by the chloroplast-localized acetyl-CoA carboxylase (ACCase) complex (Figure 9A). ACCase catalyzes the rate-limiting step in FA synthesis and has been regarded as a significant bottleneck in the accumulation of triacylglycerols (TAGs), the primary storage form of seed lipids (Ke et al., 2000). Our transcriptome data show that the three nuclear-encoded enzymes of the complex displayed Cc-biased expression, with the Cc subgenome copies contributing 54.0% of the total ACCase-related transcripts in the developing siliques (Figure 9B). Acetyl-CoA is then converted to malonyl-acyl carrier protein (ACP), with ACP acting as an essential shuttle for the acyl intermediate as it is elongated by the cyclical FA synthase (FAS) complex (Huang et al., 2017). ACP is encoded by *ACP1*, which exists in eight copies in the Bc subgenome and only one copy in the Cc subgenome. The Bc subgenome copies contribute 98.4% of the *ACP1* transcripts in the siliques.

Beginning with an initial acetyl-CoA, each FAS cycle uses one malonyl-ACP to add two carbons to the growing acyl-CoA chain. This complex is composed of four enzymes that perform sequential reactions. The initial condensing enzyme, ß-ketoacyl-ACP synthase (KAS) III, displays Bc-biased expression, with the Bc subgenome contributing 64.4% of the *KASIII* transcripts in the developing siliques. The rest of the enzymes involved in the FAS cycle do not show an overall bias, with *KASII* and ß-hydroxyacyl-ACP dehydrogenase (*HAD*) displaying Bc subgenome bias and ß-ketoacyl-ACP reductase (*KAR*) and enoyl-ACP reductase (*ENR*) displaying Cc subgenome bias. Seven FAS cycles produce stearic acid (C18:0), which is then desaturated by a stearoyl-ACP desaturase (SAD) before export to the cytosol. The *Arabidopsis thaliana* genome encodes four SADs (FA BIOSYNTHESIS [FAB] 2, ACYL-ACP DESATURASE [AAD]1, AAD5, and AAD6) that work redundantly and play a role in determining the resulting FA composition and total concentration of seed oil (Kazaz et al., 2020). The *B. carinata* homologs of these four SADs are significantly expanded in the Bc subgenome (15 total copies) compared to those in the Cc subgenome (5 copies). The genes encoding these desaturases also show Bc-biased expression with *SAD*, *FAB2*, and the *AADs* contributing 70.3%, 67.4%, and 85.7% of those transcripts, respectively, in the developing siliques.

TE insertions in *FAE1*, which encodes a regulator of erucic acid synthesis

After export to the cytosol, oleoyl-CoA (C18:1) encounters the endoplasmic reticulum (ER) membrane-localized FA elongation (FAE) complex, which extends FAS-derived FAs into very-long-chain FAs (VLCFAs) such as erucic acid (C22:1, Figure 9A). The FAE complex comprises four membrane-bound proteins that perform sequential reactions, with each cycle adding two carbons to the growing acyl-CoA chain. The first of these four enzymes, β-ketoacyl-CoA synthase (KCS), facilitates the condensation of an acyl-CoA chain with malonyl-CoA and is the rate-limiting step in VLCFA production. The resulting VLCFA chain lengths are determined by the substrate specificity of the numerous KCS isoforms (Millar and Kunst, 2003). FA ELONGASE 1 (FAE1) is the isoform that targets oleic acid for elongation and governs erucic acid production in *Brassica* species (Millar and Kunst, 2003; Saini et al., 2019).

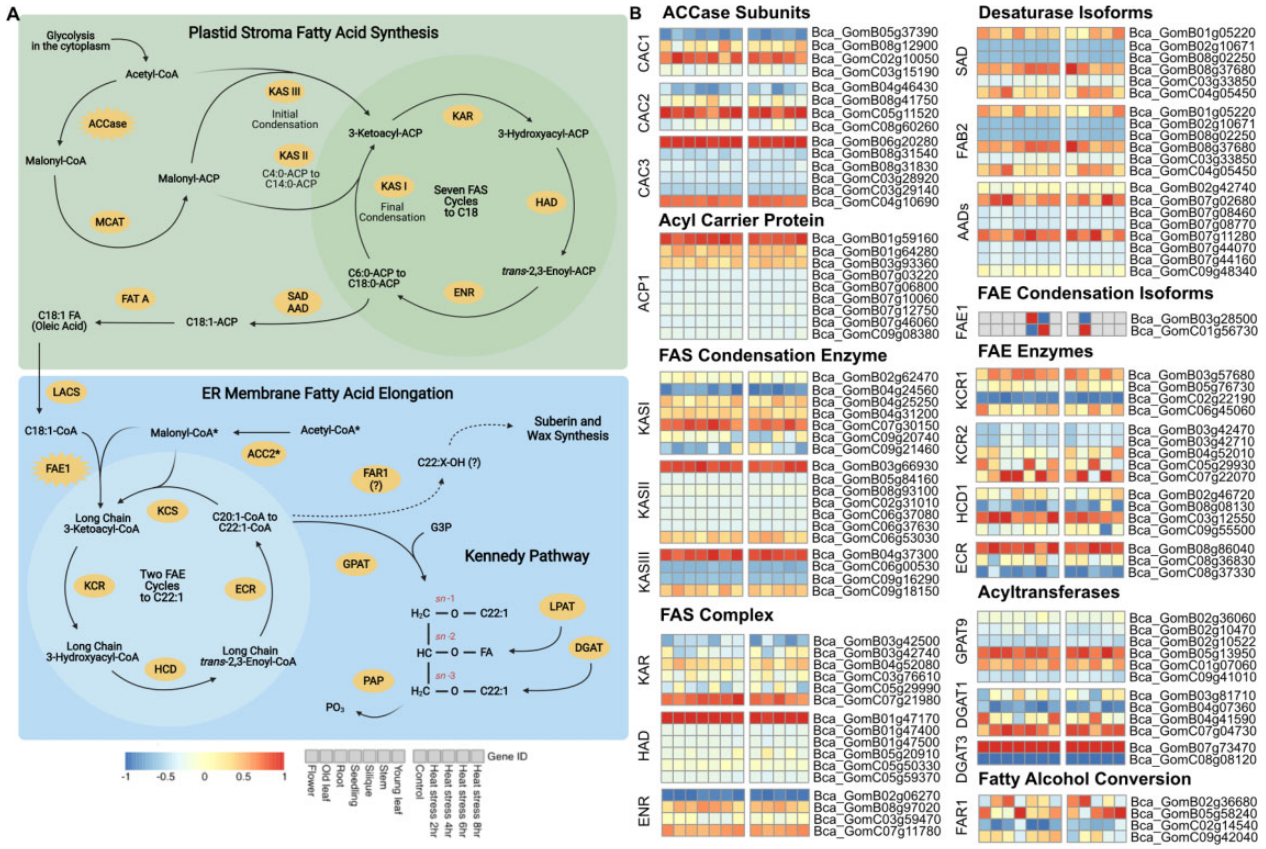


Figure 9 Erucic acid TAG biosynthesis pathway. A, Visualization of the processes of fatty acid synthesis, elongation, TAG assembly, and fatty acid recycling made using BioRender (B) Relative expression of genes involved in erucic acid biosynthesis. The relative transcript abundances of genes involved in erucic acid biosynthesis are visualized as heat maps. The *x*-axes of the heat maps describe either the tissue or heat stress time point for which the RNA was extracted and the *y*-axis provides the gene IDs. MCAT, malonyl-CoA ACP S-malonyltransferase; KAS, β-ketoacyl-ACP synthase; KAR, β-ketoacyl-ACP reductase; HAD, β-hydroxyacyl-ACP dehydrase; ENR, enoyl-ACP reductase; SAD, stearoyl-ACP Δ9-desaturase; FAT A, fatty acyl-ACP thioesterase A; LACS, long-chain acyl-CoA synthetase; ACC2, acetyl-CoA carboxylase 2; KCS, β-ketoacyl-CoA synthase; HCD, β-hydroxyacyl-CoA dehydrase; ECR, enoyl-CoA reductase; G3P, glycerol-3-phosphate; GPAT, glycerol phosphate acyltransferase; LPAT, lysophosphatidyl acyltransferase; PAP, phosphatidic acid phosphohydrolase; DGAT, diacylglycerol acyltransferase.

The *B. carinata* genome has retained one *FAE1* homolog per subgenome (Figure 9). The other KCS-encoding genes (*KCS1-3*) are relatively more expanded, suggesting purifying selection pressure on the number of *FAE1* copies. The other Triangle of U allotetraploids also contain one *FAE1* copy per subgenome and our phylogenetic analysis shows that the *FAE1* homologs cluster according to their genome of origin (A, B, or C; Supplemental Figure S11). We found that the Bc subgenome copy exhibits stronger codon usage bias (effective number of codons, ENC = 56.24) and thus is expected to be translated at a higher rate than the Cc subgenome copy (ENC = 58.26; Wright, 1990). In addition, transcription of the Bc subgenome homolog does appear to be more active, contributing 57.7% of the *FAE1* transcripts in the developing siliques (Supplemental Data Set 22). Interestingly, the *FAE1* locus in the Cc subgenome is also expressed in the stem.

After polyploidization, unprecedented retention of TE insertions may occur in genic regions and thereby introduce some of the intraspecific diversity required for crop domestication (Baduel et al., 2019). TE insertions are significant drivers of phenotypic variation and have been shown to contribute to variation in agronomically important traits such as self-compatibility in *B. napus* (Gao et al., 2016). TE insertions in promoter and genic regions can have multiple effects on the expression of a gene, including silencing mediated by methylated TEs (Hollister and Gaut, 2009). As *FAE1* is believed to be the regulator of erucic acid synthesis, we investigated the gene structure of the two *B. carinata* homologs. We found that the *FAE1* copy in the dominant Bc subgenome contains a 206-bp *Tc1/Mariner* insertion in its promoter region (Supplemental Figure S12A). *Tc1/Mariner* transposons can carry "blurry" promoters that function in a wide range of genomic environments and might contribute to its increased transcript abundance (Palazzo et al., 2019). Interestingly, the Cc subgenome *FAE1* copy that is expressed in stems carries multiple *PIF/Harbinger* insertions in its promoter region (Supplemental Figure S13A). However, our results suggest that the dominant Bc subgenome contributes more to seed oil biosynthesis. In contrast, the Cc subgenome might be more involved in modulating leaf

membrane lipid composition in response to heat stress and stem membrane lipid composition under normal development (Zhukov and Shumskaya, 2020; Zoong Lwe et al., 2021). Notably, *FAE1* expression in plants is typically restricted to developing seeds (Zeng and Cheng, 2014). Thus, *FAE1* activity in the stems might contribute to the production of VLCFAs as a component of barrier lipids, as proposed by Chiron et al. (2015).

FAE1 homologs among the A, B, and C genomes

We found that the *FAE1* homologs are similarly located among the Triangle of U genomes, with most of the A, B, and C subgenome homologs appearing on chromosomes A08, B03, and C03, respectively, suggesting conservation of both copy number and locus (Supplemental Figures S12–S14). Each A subgenome homolog is positioned within an 8-Mbp region on chromosome A08 (Supplemental Figure S14). The *B. rapa* (Ar) Z1 assembly is the only A subgenome homolog lacking a promoter TE insertion (Supplemental Figure S14). Like the A subgenome homologs, all of the C subgenome homologs are located within a 19-Mbp region of chromosome C03, except for the *B. carinata* homologs, which have been transferred to chromosome C01 (Supplemental Figure S13).

While upstream *Helitrons* represent the prevailing insertions among the three subgenomes, the A subgenomes are enriched for *hAT* insertions and the C subgenomes are enriched for *PIF/Harbinger* insertions, both of which are classified as terminal inverted repeats. Eight of the ten *B. napus* (AnCn) genomes we investigated, including No. 2127, the only accession with high erucic acid content we studied (Liu et al., 2004), share a 3,074-bp *hAT* insertion in the promoter region of the A08 *FAE1* homolog. Considering that this *hAT* insertion is present in the high and low erucic acid accessions, the Cn subgenome copy of *FAE1* in No. 2127 likely facilitates erucic acid production and is contained within a *Helitron* that might carry promoter or enhancer sequences.

The genome of the *B. juncea* (AjBj) T84-66 vegetable accession contains two *FAE1* homologs, one on chromosome B03 and one on chromosome B04. The genic regions of both contain an upstream *Helitron* insertion (Supplemental Figure S12, D and E; Yang et al., 2016). We also confirmed the existence of two *FAE1* homologs in the oleiferous Varuna accession (Gupta et al., 2004). One homolog is located on chromosome A08 preceded by a *Helitron* insertion and the other is on chromosome B07 with no TE insertions in its promoter region (Supplemental Figures S12C and S14K; Paritosh et al., 2021). InDels in AT-rich regions of the promoter of the *FAE1* copy on the A subgenome have been correlated with low erucic acid content in *B. rapa* (Ar; Yan et al., 2015) and *B. juncea* (AjBj, Saini et al., 2019). AT-rich promoter sequences can function to enhance the expression of genes expressed in a tissue-specific manner (Sandhu et al., 1998), such as *FAE1*. Further, *Helitrons* display an insertion preference for AT-rich regions (Grabundzija et al., 2016) and can thereby disrupt an enhancer sequence. Unlike T84-66, Varuna seed oil contains high amounts of erucic acid (~47%, Gupta et al., 2004), which might have been facilitated by the lack of a *Helitron* insertion in the promoter of the *FAE1* copy in the dominant Bj subgenome. Other types of TE insertions in the *FAE1* promoter region have been associated with low erucic content, such as a *Tc1/Mariner* transposon insertion in *B. juncea* (Saini et al., 2019) and a *PIF/Harbinger*-like insertion in *Sinapis alba* (Zeng and Cheng, 2014). Because the Bj subgenome *FAE1* homolog in the Varuna genome also exhibits stronger codon usage bias (ENC = 56.06), we would expect its transcript to be translated at a higher rate than that of the Aj subgenome copy (ENC = 57.30).

Association of fatty acyl-CoA reductases with seed oil erucic acid content

Two FAE cycles produce erucoyl-CoA (C22:1) from oleoyl-CoA (C18:1) in the ER membrane (Figure 9A). Erucoyl-CoA is then transferred to a glycerol-3-phosphate (G3P) backbone to form seed storage TAGs, which are packaged into an oil body that is then pinched off from the ER membrane. Alternatively, VLCFA acyl-CoAs can be converted to fatty alcohols by ER-localized alcohol-forming fatty acyl-CoA reductases (FARs) and used for suberin biosynthesis or their wax esters can be used for cuticular wax synthesis (Rowland and Domergue, 2012). A study in *B. napus* showed an enrichment for *Helitron* insertions in *FAR* genes although *Helitrons* generally accumulate in gene-poor regions (Hu et al., 2019). Hu et al. found a *Helitron* insertion in the first intron of *FAR1* in two low erucic acid cultivars that might contribute to low erucic acid trait.

The *B. carinata* genome contains four *FAR1* homologs with two copies in each subgenome (Supplemental Data Set 23). Our phylogenetic analysis of the Triangle of U *FAR1* homologs revealed two clades that cluster according to their subgenome of origin (Supplemental Figure S15). Of the four *B. carinata FAR1* homologs, Bca_GomB02g36680 and Bca_GomC09g42040 show the highest expression levels in the developing siliques, contributing 38.8% and 45.9% of the *FAR1* transcripts in the siliques (Supplemental Data Set 23). The codon usage bias of the four *B. carinata FAR1* homologs is largely similar with ENC values that range from 55.34 to 56.79. However, the two *FAR1* homologs that are lowly expressed in the siliques have intronic TE insertions (Supplemental Figure S16). Because the highly expressed *FAR1* homologs are free of TE insertions in their coding sequence (CDS) regions, the intronic TE insertions might reduce the *FAR1* expression in the siliques. Thus, the expression of the *FAR1* homologs without TE insertions might contribute to the high erucic acid phenotype in *B. carinata*.

Gene family expansion and TE signatures in *FLC* homologs facilitate adaptation to different climates for crop expansion

Of the Triangle of U allotetraploids, *B. carinata* and *B. juncea* are both winter crop species (Paritosh et al., 2021), while *B. napus* originated as a winter crop from which spring and semi-winter ecotypes have been developed through traditional breeding methods (Song et al., 2020). Late-flowering

(winter) ecotypes overwinter before flowering to coordinate their reproductive development with the spring season. This process of cold exposure-induced flowering, or vernalization, is partly facilitated by strong expression of *FLOWERING LOCUS C (FLC)*. FLC, a MADS-box TF that is a key repressor of flowering, acts in a dosage-dependent manner and is one of the most important genes controlling the initiation of flowering in *Brassica* species (Song et al., 2020). Winter ecotypes typically display strong *FLC* expression and therefore a strong vernalization requirement, while spring types flower earlier in the growing season due to weak *FLC* expression. Thus, spring types with a reduced need for vernalization can adapt to climates that experience summer drought. The weak vernalization requirement in spring types is facilitated partly by TE insertions in *FLC*, which is particularly prone to accumulate TE insertions due to selection for these variants (Quadrana, 2020). TE insertions in the promoter and CDS of *FLC* can either upregulate (Sheldon et al., 2002) or downregulate (Gazzani et al., 2003) its expression. Song et al. (2020) found that variations in one of the nine *B. napus FLC* copies (*BnaA10.FLC*) were useful for classifying accessions by ecotype, a crucial aspect of crop breeding (Song et al., 2020).

We identified 10 *FLC* homologs in the *B. carinata* genome including four in the Bc subgenome and six in the Cc subgenome (Supplemental Data Set 24). The Bc subgenome contributes 68.4% of the *FLC* transcripts across all tissues and one copy, Bca_GomB08g03090, with the highest transcript abundance contributes 74.9% of the *FLC* transcripts in mature leaves. Bca_GomB08g03090 and three other *FLC* homologs on chromosome B08 all carry the same *Gypsy* insertion (TE_00008622) in their promoter regions (Supplemental Figures S17 and S18). Song et al. (2020) found that a MITE insertion in the promoter region of *Bna*A10.*FLC* was correlated with its increased transcript abundance, and might have facilitated the development of winter ecotypes. As the Bc subgenome contributes 2.2-fold more *FLC* transcripts than the Bc subgenome, the vernalization requirement of *B. Carinata* as a winter crop appears to be facilitated by the dominant subgenome (Supplemental Data Set 24).

Discussion

Brassica carinata displays great potential as a climate-resilient crop for use in semi-arid and tropical sub-humid environments. Our improved reference genome not only provides a thorough examination of the unique features of the *B. carinata* genome, but also further insights into the genetic diversity between the Gomenzer and Zd-1 accessions (Song et al., 2021). Our Gomenzer assembly is 22.2% larger than the Zd-1 assembly, with 95.4 Mbp more repetitive sequence and 50,195 more genes (Supplemental Data Sets 6, 9, and 12). The differences between these assemblies are particularly apparent between the Bc subgenomes, which have a 156.9 Mbp greater difference in size than the Cc subgenomes and a 3.2-fold higher difference in the number of annotated gene models (Supplemental Data Sets 6 and 9). Although the Gomenzer assembly is more complete, recent evidence suggests that the variations in genome size between accessions have biological relevance and are not just an effect of variations in genome assembly quality. For example, a comparative study of the *B. rapa* Z1 and Chiifu genotypes cytogenetically validated several large chromosomal variants, which together contributed to a 16% difference in their genome sizes (Boutte et al., 2020).

A previous population structure analysis of 620 *B. carinata* accessions reported two distinct subpopulations of this species, SP1 and SP2, that appear to have been developed through targeted selection (Khedikar et al., 2020). The Gomenzer accession belongs to the larger SP1 population comprising the Ethiopian lines, and we propose that the Zd-1 accession belongs to the SP2 population of breeding lines. That study also identified bias toward the Bc subgenome, which exhibited longer linkage disequilibrium (LD) decay and contained selective sweep regions harboring genes involved in FA and glucosinolate biosynthesis. The longer LD decay of the Bc subgenome reflects the preferential retention of Bc alleles.

Subgenomic selection bias during the domestication of *B. juncea* has also been identified. A population structure analysis of 480 *B. juncea* accessions reported that the Aj subgenome underwent stronger selection than the Bj subgenome during the development of new crop types (Kang et al., 2021). However, subgenomic bias resulting from selection pressure is not as clear in *B. napus*. The An subgenome exhibits more genetic diversity (Huang et al., 2013; Qian et al., 2014; Sun et al., 2017) and has evolved at a faster rate (Chen et al., 2021) than the Cn subgenome, but exhibits faster LD decay (Qian et al., 2014; Rahman et al., 2022). The increased genetic diversity and recombination rates in the An genome could have been influenced by the prevalence of introgressions with *B. rapa* (Ar) during the breeding history of *B. napus* (Qian et al., 2014; An et al., 2019).

The disparities between the Gomenzer and Zd-1 genomes could be not only due to differences in assembly quality, but also due to actual biological differences that could have arisen from Bc subgenome selection bias during domestication. Alternatively, the Gomenzer and Zd-1 accessions might have arisen from separate hybridization events with distinct Bc subgenome progenitor genotypes. While *B. juncea* (AjBj) likely arose from a single hybridization event (Yang et al., 2016; Kang et al., 2021), *B. napus* (AnCn) might have arisen from more than one hybridization event involving different maternal genotypes (Song and Osborn, 1992; Allender and King, 2010). However, past introgressions of *B. napus* with *B. rapa* (Ar) make the origins of *B. napus* difficult to elucidate (An et al., 2019). In support of this hypothesis, it is plausible that the Gomenzer and Zd-1 Bc subgenomes diverged earlier (~2.07 MYA) than the Cc subgenomes (Figure 7A). Our maximum likelihood (ML) analysis of the supermatrix and analysis of node ages shows the greatest overlaps between subgenome species. Although we used 1,181 low-copy number genes from

subgenomes to construct the supermatrix, we note that a considerable amount of incomplete lineage sorting is expected for a recent divergence (Pamilo and Nei, 1988). However, our divergence time estimates not only provide indications of the degree of divergence of subgenomes, but also give some notion of the timing of the Triangle of U allotetraploid hybridization events. Contrary to previous *B. juncea* research (Yang et al., 2016; Kang et al., 2021), our analysis allows us to infer subsequent independent hybridization events within allotetraploid *Brassica*. Additional analysis of genotypes, construction of a *Brassica* pangenome, and production of linkage mapping populations will aid in resolving this hypothesis precisely and will also contribute to identifying functionally important genome fragments, such as the evolutionarily constrained elements controlling polyploidy among the Triangle of U species.

Although estimates of the timing of hybridization events can help us understand the evolution of allotetraploid genomes, such estimates are not entirely reliable. The timing of recent polyploidization events, such as the Triangle of U hybridizations, is more difficult to estimate as there has not yet been enough time for synonymous mutations to accumulate (Doyle and Egan, 2010). Furthermore, because the timing of the artificial hybridizations, such as those leading to the reconstruction of the *B. napus* synthetic line No. 2127, is particularly challenging to accurately estimate, the timing of the Triangle of U allotetraploid hybridization events is still under debate. Previous estimates were calculated using K_s plots or phylogenetic analysis combined with Bayesian estimation (Chalhoub et al., 2014; Yang et al., 2016; Song et al., 2021). The distribution K_s values among individual ortholog pairs can be highly variable, which prompted Yang et al., (2016) to use a phylogenetic analysis instead (Zhang et al., 2002). However, there is still no guarantee that combined phylogenetic and Bayesian methods are more precise or accurate as these methods can also provide variable estimates for the timing of hybridization events (Doyle and Egan, 2010).

Chalhoub et al. (2014) provided the first assessment of a Triangle of U allotetraploid, using a K_s plot analysis to estimate the *B. napus* hybridization event at 7,500–12,500 years ago. Yang et al. (2016) then introduced the *B. juncea* genome assembly and used phylogenetic analysis to estimate the timing of the hybridization event for *B. juncea* at 39,000–55,000 years ago. However, their estimate for the hybridization event for *B. napus* was much earlier at 38,000–51,000 years ago. Song et al. (2021) followed suit with their *B. carinata* Zd-1 assembly, using a phylogenetic approach to estimate the hybridization times of all three allotetraploids. Their estimate for *B. napus* was similar to the estimate of Yang et al. (2016) at 41,000–45,000 years ago, but they provided an even earlier estimate for *B. juncea* at 72,000–80,000 years ago. The differences in these estimates are likely due to differences in the estimated divergence time of the outgroup species used for calibration. For example, Chalhoub et al. (2014) based their estimate of the *B. napus* hybridization event on an estimate of the divergence of *A. thaliana* and other *Brassica* species at ~12–17MYA. In contrast, by using an estimate of ~29.5 MYA as their molecular clock for the divergence of *A. thaliana* and other *Brassica* species, Song et al. (2021) increased their estimate of the *B. napus* hybridization event at 41,000–45,000 years ago.

Polyploidization triggers genome fractionation to compensate for the evolutionarily instantaneous duplication of all genes and cis-regulatory elements (Zhang et al., 2021). Several reports have emphasized that most lineages have undergone extreme gene loss or genome size reduction (Bennett and Leitch, 2005; Leitch and Leitch, 2008) and diploidization (Wolfe, 2001) following polyploidization. The *B. carinata* genome might be in the early stages of the diploidization process and thus still retains many of the genes provided by the parental genomes. Our Gomenzer assembly shows that the *B. carinata* genome is the largest among the Triangle of U species, with significant expansions of repetitive DNA sequences and gene families (Supplemental Data Sets 6, 9, and 12). These expansions are particularly apparent in the Bc subgenome, which is larger and contains 55.1% of the total gene content for the *B. carinata* genome (Supplemental Data Set 9). While the two subgenomes have similar TE contents, the Bc subgenome has a higher LTR content (Supplemental Figure S4D). *Gypsy* elements showed the most pronounced bias between the two subgenomes and were also preferentially expanded in the larger Bj subgenome of *B. juncea* (Supplemental Data Set 12; Paritosh et al., 2021). Interestingly, *Gypsy*-specific proliferation has accompanied the expansions of particularly large genomes, as observed for the *Hesperis* clade, which comprises the largest genomes among Brassicaceae species, and the larger *Gossypium* genomes (Hawkins et al., 2006; Hloušková et al., 2019).

The "genomic shock" hypothesis suggests that interspecific hybridization events can be followed by bursts of TE activity and a relaxation of selection pressures that can facilitate their retention (Parisod et al., 2010). Patterns of TE amplification and elimination then restructure the polyploid genome through the diploidization process, giving rise to species-specific TE repertoires (Du et al., 2010). We hypothesize that the larger genome size of *B. carinata* might have been facilitated by more propagation and retention of TEs, particularly LTRs (Kidwell, 2002; Canapa et al., 2015), in *B. carinata* than in the other *Brassica* allotetraploids partly due to the relatively limited breeding efforts so far for this orphan crop. The linear relationship between genome size and repetitive sequence content is well established and LTR proliferation is regarded as the primary driver of genome expansions in plants (Figure 4F; Bennetzen, 2002; Ammiraju et al., 2007; Huang et al., 2020; Hu et al., 2022). Song et al. (2020) also identified a remarkable expansion of *Gypsy* elements in *B. carinata* relative to the other Triangle of U species and reported that the LTRs in *B. carinata* are younger and less diverged compared to those in the other Triangle of U genomes.

We also found that the Bc genome has the highest *Helitron* content among the B genomes of the Triangle of U species, indicating higher *Helitron* activity, retention, or a combination of the two (Bn, Bj; Supplemental Data Set 12). *Helitrons* are "exon shuffling machines" that generate genomic diversity by disrupting cis-regulatory elements as well as capturing and translocating genes and gene fragments around the genome (Feschotte and Wessler, 2001; Morgante et al., 2007; Gao et al., 2016; Muyle et al., 2021). *Helitrons* have been recognized as key players in plant evolution and have contributed substantially to the domestication of maize (*Zea mays*) and wheat (*Triticum aestivum*) (Morgante et al., 2005; Barbaglia et al., 2012; Wang et al., 2022). We hypothesize that the notable expansion of *Helitrons* in the Bc subgenome might have facilitated its higher levels of gene family expansion through the diversification of gene and regulatory sequences.

Although concept of the Triangle of U model was introduced in 1935, the utility of this well-established comparative genomics platform was not truly harnessed until these large, highly syntenic, allotetraploid genomes could be assembled using long-read sequencing technologies. The previous assemblies of the genomes of the other two Triangle of U allotetraploids, *B. juncea* (AjBj) and *B. napus* (AnCn), have provided insights into the asymmetric evolution of their subgenomes as they return to the diploid state. *Brassica napus* Darmor-bzh was the first of the allotetraploid genomes to be assembled, enabling deeper analysis of the allotetraploid genomes and elucidating the differences between its constituent An and Cn subgenomes (Chalhoub et al., 2014). The *B. napus* genome assemblies consistently show that the Cn subgenome is larger, has higher gene and TE contents, and is more susceptible to homoeologous gene loss than the An subgenomes (Chalhoub et al., 2014; Rousseau-Gueutin et al., 2020; Song et al., 2020; Sun et al., 2017). Conversely, the An subgenome of *B. napus* has higher single nucleotide polymorphism (SNP) density and is more prone to replacing syntenic homoeologous segments in the Cn subgenome (Sun et al., 2017; Lu et al., 2019). Additionally, asymmetric subgenomic selection during the domestication of *B. napus* has resulted in the An subgenome contributing more to stress resistance and oil accumulation and the Cn subgenome contributing more to development and flowering time (Lu et al., 2019). The *B. juncea* (AjBj) assemblies also show consistent patterns of subgenome dominance. The Bj subgenome is larger, has higher gene and TE contents, and exhibits a lower degree of genetic recombination (Yang et al., 2016; Paritosh et al., 2021). While global homoeolog expression dominance has not been detected in *B. juncea*, Bj subgenome-biased expression has been observed at specific developmental stages and in particular tissue types (Yang et al., 2016). Overall, the An and Bj subgenomes of *B. napus* and *B. juncea*, respectively, show signs of subgenome dominance (Bird et al., 2018).

Polyploidy is present in other lineages, such as gymnosperms, fish, and amphibians, but rapid genomic downsizing is unique to angiosperms and might be facilitated by subgenome dominance (Cheng et al., 2018). Some speculate that genome dominance might be at the core of Darwin's "abominable mystery": the abrupt appearance of diverse flowering plant lineages in fossils dating to the mid-Cretaceous period (Friedman, 2009; Schnable and Freeling, 2011). Bc subgenome dominance in *B. carinata* was evident from its increased gene retention (Supplemental Data Sets 9 and 11), homoeolog expression dominance (Supplemental Figure S10A), and pericentromeric and telomeric clustering of TEs (Figure 8; Bird et al., 2018). At the same time, the Cc subgenome homoeologs show a higher rate of sequence evolution (Supplemental Figure S10B) and the Cc subgenome TEs are dispersed throughout the chromosome arms (Figure 8). Higher TE density around genes is a sign of subgenome submissiveness (Edger et al., 2019), as the presence of methylated TEs is correlated with lower expression levels of nearby genes (Hollister and Gaut, 2009). Overall, we found that the Cc subgenome has undergone higher levels of positive selection or relaxed purifying selection, or both. Bc subgenome dominance might have effects on biological function and phenotypic traits (Schnable and Freeling, 2011), as we found that key genes involved in seed oil biosynthesis and flowering time display Bc-biased expression (Figure 9).

As subgenome dominance in *B. carinata* is likely a legacy of the genomic features of the diploid progenitor genomes (Edger et al., 2017), we compared the observed subgenomic asymmetries to the genomic features of the extant progenitor species genomes. We found that the genome assemblies of the Bc subgenome progenitor species *B. nigra* (Bn) generally have higher gene and LTR contents than the assemblies of the Cc subgenome progenitor species *B. oleracea* do (Co, Supplemental Data Sets 9 and 12). Thus, the higher gene and LTR contents of the Bc genome could have been established upon hybridization. However, because the *B. nigra* (Bn) genomes are generally smaller than the *B. oleracea* assemblies, the expanded size of the Bc subgenome might be unique to the *B. carinata* genome or the Gomenzer accession.

HEs act as substrates for selection and can profoundly affect genome structure and gene expression levels, giving rise to inter- and intra-specific phenotypic variations (He et al., 2017; Stein et al., 2017; Lloyd et al., 2018). Unbalanced translocations often show subgenomic bias following selection, such as the preferential replacement of Cn subgenome fragments by the homoeologous portion of the An subgenome observed in *B. napus* (AnCn) (Nicolas et al., 2012; Chalhoub et al., 2014; Samans et al., 2017; Higgins et al., 2018). Among the Triangle of U allotetraploids, the *B. napus* (AnCn) assemblies display the most HEs, with an average of 32 balanced translocations and 593 unbalanced translocations (Supplemental Data Set 17). As the *Brassica* B genome diverged before the A and C genomes (Figure 7A), the lower level of divergence between the An and Cn genomes likely led to the observed higher frequency of HEs in the *B. napus* (AnCn) genomes compared to those of *B. carinata* (BcCc)

and *B. juncea* (AjBj). We found that the *B. napus* genome assemblies showed intraspecific variations in HE subgenomic bias. For example, preferential replacement of Cn homoeologs with An homoeologs occurred in the synthetic *B. napus* accession No. 2127 (Supplemental Data Set 17). However, the Darmor-bzh v5 genome assembly displayed a bias for An to Cc translocations, while the Darmor-bzh version 10 genome assembly showed a bias for Cc to An translocations. HE is a continual phenomenon that appears to cause intraspecific variations, but we acknowledge that genome assembly quality might be a factor affecting HE bias (Glover et al., 2016). We also postulate that the observed variations in bias among these *B. napus* assemblies could be due to the prevalence of past introgressions of *B. napus* (AnCn) with *B. rapa* (Ar) in Chinese *B. napus* accessions (Qian et al., 2006). Moreover, 63.1% of the unbalanced translocations found in all three allotetraploids could not be attributed to a particular subgenome. Consequently, these results illustrate the difficulty of accurately characterizing HE events in these closely related genomes. As a result of multiple HE events, homoeologs might have been assimilated into one another.

As part of homoeology inference and HE analysis, we propose identifying HT events using gene triplets. Considering the disparities noted during the characterization of HE subgenomic bias, the analysis of HTs would provide a more comprehensive picture of HE selection bias because these HEs are inaccessible to gene quartet analysis. Our HT results clearly show, as did our unbalanced translocation results, that homoeologs from the recessive Cc subgenome more commonly replaced homoeologs from the dominant Bc subgenome in the *B. carinata* genome (Supplemental Data Set 17). The higher mutation rate of these Cc subgenome homoeologs (Supplemental Figure S10B) could have allowed for the selection of favorable mutations following HE events.

In conclusion, the high-quality *B. carinata* genome assembly reported here will help to improve our comprehension of the genetics underlying favorable agronomic traits. We observed evidence of the early stages of diploidization that will contribute to our understanding of interspecific hybridization and the establishment of subgenome dominance in *Brassicas*. The *B. carinata* subgenomes (Bc and Cc) are distinct from the current genomes of their extant respective progenitor species (Bn and Co). In line with the findings from Song et al. (2021), we found that both subgenomes show higher levels of divergence (K_s) from their respective progenitor species (Bn and Co) than their shared genomes in the other allotetraploids do from their progenitor species (i.e. comparing Bj to Bn and Cn to Co). These higher mutation rates might also have helped promote the extensive gene family expansions in the *B. carinata* genome. Our comprehensive characterization of the *B. carinata* genome could advance introgression-based crop improvement efforts, including those in other *Brassica* species. The addition of this high-quality chromosome-scale *B. carinata* genome assembly to complete the Triangle of U will further accelerate the pace of genetic and genomic research in *Brassica* species.

Materials and methods

Library preparation and sequencing

Brassica carinata A. Braun var. Gomenzer (PI 273640, USDA) seeds from the USDA North Central Regional Plant Introduction Station were sown and selfed for three generations in a greenhouse at the University of Nevada, Reno to decrease the heterozygosity of the plants sampled for sequencing. High molecular weight DNA was isolated from plant leaves according to a CTAB protocol modified from previous reports (Japelaghi et al., 2011; Healey et al., 2014) for PacBio sequencing. On the Sequel system, 30-kb and 8-kb insert SMRTbell libraries were prepared for sequencing using five and two cells, respectively. Short-read DNA was extracted using a Zymo *Quick*-DNA Plant/Seed Miniprep Kit (Zymo Research, Irvine, CA, USA) following the manufacturer's recommendations.

Hi-C experiments were performed with young leaves fixed with a 2% formaldehyde solution using the Arima Hi-C Kit according to the manufacturer's specifications with some modifications (Arima Genomics, San Diego, CA, USA). Two Hi-C libraries were prepared using the Accel-NGS 2S Plus DNA Library Kit (Swift Biosciences, Ann Arbor, MI, USA) following the manufacturer's recommendations.

Seven cDNA libraries for RNA sequencing were constructed from the following *B. carinata* tissues in triplicate: (1) 13-day-old seedlings; (2) roots; (3) young leaves; (4) mature leaves; (5) stems; (6) flowers; and (7) green siliques. Fifteen cDNA libraries were generated from the leaves of plants exposed to 40°C heat stress at durations of 2, 4, 6, and 8 h. Total RNA was isolated from the ground tissues using the *Quick*-RNA Plant Miniprep Kit following the manufacturer's specifications (Zymo Research, Irvine, CA, USA, USA). The short-read DNA, Hi-C, and cDNA libraries each had a 300-bp average insert size for paired-end sequencing by Novogene Corporation, Inc. (Davis, CA, USA) on an Illumina NovaSeq 6000 platform.

De novo genome assembly

We used two independent methods to estimate the size of the *B. carinata* genome: (1) three replicates of flow cytometry and (2) *k*-mer based estimation with GenomeScope version 2.0 (Ranallo-Benavidez et al., 2020). Contigs were assembled from the PacBio sequencing data using the Canu version 2.0 pipeline with the following parameters: "*corThreads = 64 batMemory = 186 ovbMemory = 24 ovbThreads = 12 corOutCoverage = 120 ovsMemory = 32-186 maxMemory = 249 ovsThreads = 20 oeaMemory = 32 executiveMemory = 64*" (Koren et al., 2017). The contigs were clustered according to their respective subgenomes using a combined reference genome of the diploid progenitor species *B. nigra* (Bn) and *Brassica oleracea* (Co) with RaGOO (Alonge et al., 2019).

The Hi-C reads were trimmed using TrimGalore (K and the HiC-Pro pipeline (Servant et al., 2015) was used to filter out low-quality reads from the Hi-C sequencing data. Loci

contact frequency matrices were generated using HiC-Pro (Servant et al., 2015) and Hi-C Explorer (Wolff et al., 2020). We used the valid and trimmed read pairs as input for the ALLHiC pipeline (Zhang et al., 2019) to scaffold the pseudochromosome-clustered contigs according to the linkage information from the Hi-C data. Only read pairs that aligned to separate contigs were used for the initial scaffolding and misjoined contigs were inspected using the 3D-DNA pipeline (Dudchenko et al., 2017) and manually corrected.

Gaps in the scaffolded assembly were filled in using the PacBio sequencing data and four iterative runs of the LR_Gapcloser program (Xu et al., 2019) with the following parameters "*-a 0.2 -m 600 -g 300.*" Three programs, Pilon (Walker et al., 2014), Arrow (Chin et al., 2013), and FreeBayes (Garrison and Marth, 2012) were used consecutively to correct errors in the assemblies. First, inaccurate base calls, InDels, and gaps in the PacBio-based assembly were identified and corrected with Pilon using the Illumina sequencing data. The raw PacBio data was then used with Arrow (Chin et al., 2013) for further correction of call variants and Indels. Lastly, FreeBayes (Garrison and Marth, 2012) was implemented to process the resulting consensus sequence and mapped Hi-C reads in 100-kb chunks to reveal additional call variants. These data were then compressed and indexed with Tabix (Li et al., 2011) and a further improved consensus sequence was generated using BCFtools (Danecek et al., 2021).

Repeat sequence annotation

Repetitive elements within the *B. carinata* genome were identified using the MAKER Advanced Repeat Library Construction document (Campbell et al., 2014) with some minor deviations, including the substitution of MITE-Hunter with MITE Tracker (Crescente et al., 2018), the use of Repbase (Bao et al., 2015) as the transposase protein and DNA sequence database, and the use of an updated release of UniProt as the plant protein database (UniProt Consortium, 2014; Boutet et al., 2016). The Extensive *de novo* TE Annotator (EDTA) was used with the default parameters to predict TEs (Ou et al., 2018). We used EDTA to validate the repeat sequence annotations, and to detect structure-based genomic repeat contents. The results were combined with the previous repeat identification results for complementary repeat annotation and were then passed to the MAKER annotation software (Cantarel et al., 2008).

Lastly, each of the repeat sequence libraries was used as a query for NCBI BLASTX (Camacho et al., 2009) against the UniProt/SwissProt database (Boutet et al., 2016) to identify any potential protein-CDS fragments that should be excluded. ProtExcluder version 1.2 (Campbell et al., 2014) was also used to process each of the BLASTX results in text files and the corresponding repeat sequence library to generate a library containing no protein-coding gene sequence fragments.

The LAI was calculated in 3-Mbp steps and a sliding window of 300 kb as the length of intact LTRs divided by the total length of LTRs (Intact LTR length/Total LTR length) (Ou et al., 2018).

Transcriptome alignment and gene annotation

The cDNA reads were trimmed with TrimGalore (https://github.com/FelixKrueger/TrimGalore) and the transcriptome was generated by assembling the trimmed transcriptome cDNA reads using the Trinity assembler (Grabherr et al., 2011) with the "*–jaccard_clip*" option. CDSs of the resulting transcriptome were predicted using TransDecoder (Haas et al., 2003). The filtered reads were aligned to the assembled genome using the Spliced Transcripts Alignment to a Reference Aligner (Dobin et al., 2013), then assembled via genome-guided Trinity assembly. The *de novo* transcriptome and genome-guided assembly were processed using Program to Assemble Spliced Alignments with *A. thaliana*, *B. nigra*, *B. oleracea*, *B. napus*, and *B. juncea* protein sequences as references (Haas et al., 2003). Read counts were obtained using featureCount (Liao et al., 2014) to identify quantitative differential expression estimates using DESeq2 (Love et al., 2014). We used the transcripts per million normalization method for downstream analysis gene expression (Robinson et al., 2010). We identified differentially expressed genes (DEGs) as those with a ±two-fold change in expression with a false discovery rate cutoff of <0.001.

We used a MAKER pipeline to annotate protein-CDS s within the genome based on multiple forms of evidence. Six iterative rounds of MAKER (Cantarel et al., 2008) were run with SNAP (Korf, 2004), and AUGUSTUS (Stanke et al., 2006), and FGENESH (Salamov and Solovyev, 2000) for ab initio gene prediction, each using the ab initio training set from the previous run and the *B. carinata*-specific repeat libraries. The transcripts constructed using *de novo* and genome-guided methods were used to train SNAP (Korf, 2004) after evaluating them against a protein database comprising full-length candidates from Arabidopsis, *B. nigra*, *B. oleracea*, *B. napus*, and *B. juncea* with $>95\%$ coverage of core conserved genes. GeneMark (Besemer and Borodovsky, 2005) and AUGUSTUS (Stanke et al., 2006) were trained directly from the transcriptome data using BRAKER2 (Brůna et al., 2021). FGENESH (Salamov and Solovyev, 2000) with pretrained Arabidopsis models. The completeness of our gene annotation was assessed after each round of the MAKER (Cantarel et al., 2008) process by analyzing the presence of BUSCOs. The results from BUSCO were then used for AUGUSTUS retraining. For SNAP, we used the options "*-x 0.80 -l 50*" for retraining. Predicted proteins longer than 30 amino acids were retained and alternative splicing was allowed. Among the alternative splicing isoforms, the most highly expressed transcripts were designated as primary transcripts with the suffix "t1."

For annotation of functional descriptions, DCBLAST (Yim and Cushman, 2017) was used to identify the homologous genes in the Arabidopsis protein database (Krishnakumar et al., 2015), UniProt-SwissProt, and UniProt-TrEMBL (Boutet et al., 2016). Descriptions of protein functions were inferred using automated assignment of human readable descriptions (AHRD)

(https://github.com/groupschoof/AHRD) based on the best hit human readable protein descriptions for each predicted *B. carinata* protein from the above three protein databases. We used InterProScan (Jones et al., 2014) to identify the GO terms and KEGG pathway information associated with all predicted proteins, in addition to the Pfam domains contained in each predicted protein. The Pfam domains were then used to identify genes encoding putative TFs using plantTF_identifier (http://github.com/tangerzhang/plantTF_identifier). The relevant pathways were identified using the Ensemble Enzyme Prediction Pipeline (E2P2 v3.1) (Schläpfer et al., 2017). We used GO terms from InterProScan and created a *B. carinata*-specific GO database, then we used GOATOOLS (Klopfenstein et al., 2018) to analyze the enrichment of GO terms associated with our set of predicted *B. carinata* genes.

Chloroplast genome assembly

The *B. carinata* chloroplast genome was assembled by aligning the Illumina gDNA reads to the Arabidopsis chloroplast genome using Bowtie2 (Langmead and Salzberg, 2012). The mapped reads were converted to a FASTQ file and then individually assembled using the SPAdes (Bankevich et al., 2012) assembler. Palindromic sequences in the Arabidopsis chloroplast genome were identified using a BLASTN search against itself. The Arabidopsis chloroplast genome contains a pair of inverted repeat (IRA and IRB), a long single-copy region, and a small single-copy region. These four regions were fed to SPAdes (Bankevich et al., 2012) as trusted contigs and reassembled. The final assembly was oriented with the *psbA* gene as the starting point. Genome annotation was then performed using CPGAVAS2 (Shi et al., 2019).

Comparison of orthologous genes and gene families

We used OrthoFinder (Emms and Kelly, 2019) software to identify orthologous genes among 27 representative plant genomes, including several in the Brassicaceae (*B. carinata*, *B. juncea*, *B. napus*, *B. nigra*, *B. oleracea*, *B. rapa*, *Aethionema arabicum*, *Arabidopsis lyrata*, *A. thaliana*, *Arabis alpina*, *Camelina sativa*, *Capsella rubella*, *Raphanus raphanistrum* subsp. *sativus*, *Sisymbrium irio*, *Thellungiella parvula*, *Thellungiella halophila*, *Eutrema salsugineum*, *Leavenworthia alabamica*, and *Tarenaya hassleriana*), the Caricaceae (*Carica papaya*), the Rutaceae (*Citrus clementina*), the Malvaceae (*Theobroma cacao*, *Corchorus olitorius*, *Gossypium raimondii*), the Myrtaceae (*Eucalyptus grandis*), the Vitaceae (*Vitis vinifera*), and the Poaceae (*Oryza sativa*) (Chalhoub et al., 2014; Jaillon et al., 2007; Ming et al., 2008; Argout et al., 2011; Hu et al., 2011; Cheng et al., 2013; Haudry et al., 2013; Slotte et al., 2013; Yang et al., 2013; Kagale et al., 2014; Kitashiba et al., 2014; Willing et al., 2015; Wu et al., 2012; Hinze et al., 2016). Divergence time estimates were calculated using r8s (Sanderson, 2003) with calibration points obtained from the TimeTree database (Kumar et al., 2017).

We used the online platform OmicShare (www.omicshare.com/tools) to conduct GO term enrichment and KEGG pathway analyses. Gene duplications were identified using the DupGen_finder pipeline (Qiao et al., 2019), with the DupGen_finder-unique.pl script and were classified into tandem, proximal, transposed, or dispersed duplication types. The respective progenitors were taken as outgroups for all species in the Triangle of U. A genome wide BLASTP search was performed (e-value $< 10^{-10}$, maximum five matches, and tabular format) using the DCBLAST (Yim and Cushman, 2017) pipeline followed by MCScanX (Wang et al., 2012) analysis to identify gene pairs that had resulted from a WGD event.

Phylogenetic analysis

Molecular dating was carried out using a stringent set of 1,181 low-copy orthologs (411,367 sites) in all of the *Brassica* subgenomes (An, Ar, Aj, Bc, Bn, Bj, Cc, Cn, and Co), and the genomes of *A. thaliana*, *A. lyrata*, *A. alpina*, and *Raphanus sativus*, together with six fossil-based age constraints on internal nodes of the tree. A ML analysis was performed with the 1,181-sequence supermatrix using RaxML (Stamatakis, 2014) under the GTR + GAMMA + I model defined as the best-fit evolutionary model according to the Bayesian information criterion (BIC) from IQ-TREE (Nguyen et al., 2015). We calculated the posterior probability distribution of node ages and calculated evolutionary rates using the MCMCtree program in the PAML package (Yang, 1997). Phylogenetic trees were visualized with iTOL (Letunic and Bork, 2021).

To identify maternal inheritance from the *B. carinata* chloroplast genome assembly, we retrieved assembled chloroplast genomes from NCBI Organelle Genome Resources (Wolfsberg et al., 2001), BRAD database (Cheng et al., 2011) and previous research (Li et al., 2017). The CDSs and protein sequences of the chloroplast genome assemblies were aligned using MUSCLE (Edgar, 2004). Next, each alignment was adjusted using PAL2NAL (Suyama et al., 2006) according to the protein sequence alignments. We used a single matrix comprised of the aligned sequences to determine the best-fit nucleotide substitution model under a BIC in IQ-TREE (Nguyen et al., 2015). Thus, the phylogenetic tree was reconstructed using RAxML under the GTR + CAT model with 1,000 bootstrap replicates.

Dating of WGDs and species divergence

Both the peptide sequences and CDS of these orthogroups identified with OrthoFinder (Emms and Kelly, 2019) were aligned using MUSCLE (Edgar, 2004). The CDSs were aligned onto the amino acid alignments using PAL2NAL (Suyama et al., 2006). Any invalid alignments were then filtered out using the following trimAl criteria (Capella-Gutiérrez et al., 2009). First, if there were gaps in over 90% of the sequences, sequence bases in the alignments were removed. Second, if the translations of the transcript coverage comprised $< 30\%$ of the total alignment length of a gene family, they were also filtered out.

The hybridization date for the *B. carinata* genome was estimated from the average of two calculations: $K_{sB}/2T_B$ and $K_{sC}/2T_C$, where K_{sB} is the K_s rate estimated for ortholog pairs in the *B. nigra* and Bc subgenomes, K_{sC} is the K_s rate estimated for ortholog pairs in a given *B. oleracea* and Cc

subgenome, and T is the divergence time proposed in previous reports (Lysak et al., 2005; Navabi et al., 2013; Arias et al., 2014; Cheng et al., 2017). For instance, *B. carinata* divtime = $K_{sB(Bni,BcaB)}/2T_{BBni} + K_{sC(Bol,BcaC)}/2T_{CBol}$. K_{sB} and K_{sC} were determined as the mean K_s value from all orthologous pairs analyzed. The K_s values of ortholog pairs were calculated using the synonymous_calc.py script (https://github.com/tanghaibao/bio-pipeline).

Collinearity analyses

The assembled *B. carinata* genome was compared in a pairwise fashion with the related Triangle of U genomes among the two allotetraploids (*B. napus* and *B. juncea*) and six reference genomes from two diploid progenitors (*B. nigra* YZ12151, *B. nigra* C2, *B. nigra* NI100, *B. oleracea* Capitata, *B. oleracea* BOL, and *B. oleracea* HDEM). Genome-wide syntenic regions in these *Brassica* species were identified using LAST (Frith et al., 2010) with cscore cut-off of 99% and a distance cutoff of 20 neighboring genes. Each 1x1 synteny block was screened with QUOTA-ALIGN according to the best cscore set from LAST. The process of synteny identification was implemented in the Python version of MCScan (Tang et al., 2008). Results were visualized as dot plots using the JCVI utility libraries (https://github.com/tanghaibao/jcvi).

Identification of homoeologs and duplication-deletion events

We performed local gene collinearity analysis to identify homoeologs in *B. carinata* (BcCc), *B. juncea* (AjBj), *B. napus* (AnCn), *B. nigra* (Bn), and *B. oleracea* (Co). We performed an all-against-all sequence similarity analysis using LAST (Frith et al., 2010). Collinearity blocks were identified as at least four homoeologs pairs with a distance cutoff of 20 genes using a MCScanX (Wang et al., 2012). For each of the resulting homoeologous pairs, we then calculated the K_s value using yn00 in the PAML package (Yang, 1997). The peaks in *Ks* value distributions for each species were identified with Gaussian mixture models using the DupGen_finder pipeline (Qiao et al., 2019).

To compare homoeologous gene expression, we directly compared gene expression within the quartets (Bc–Bn–Cc–Co) and between homoeologous gene pairs (Bc compared with Cc) using our RNA-Seq samples. We assigned a new ID to each pair of homoeologs. Raw read counts were divided into the Bc and Cc subgenome matrixes then combined into a super matrix according to their pair ID. For DEG analysis, we set up a contrast of the raw read counts derived from the Bc and Cc subgenomes to determine whether measured differences in gene expression differed from zero.

Accession numbers

Sequence data from this article can be found in the NCBI SRA data libraries under the BioProject number PRJNA 835891 and accession number(s): PacBio (SRR19221947-SRR19221953), Hi-C (SRR19201389 and SRR19201390), DNA-Seq (SRR19207379), RNA-Seq (SRR19134636-SRR19134670). The genome assembly and annotation for *B. carinata* is available from CoGe under id 63922 and name *B. carinata* Genome Assembly from Plant Genomics Lab (University of Nevada, Reno). Phylogenetic trees in this article can be found in Figshare under the accession number: FAE1 (19775764), FLC (19775749), and FAR1 (19775797).

Supplemental data

The following materials are available in the online version of this article.

Supplemental Figure S1. Illumina sequencing data *k*-mer read length distribution for *B. carinata* Gomenzer.

Supplemental Figure S2. Hi-C-based scaffolding and syntenic alignment of *B. carinata* Gomenzer assembly with shared Triangle of U genomes.

Supplemental Figure S3. Comparison of BUSCO assessments for Triangle of U genomes using the Viridiplantae dataset.

Supplemental Figure S4. Distribution of genomic features among the subgenomes of the Triangle of U allotetraploid species.

Supplemental Figure S5. Transcriptome read mapping and assignment statistics for *B. carinata* Gomenzer.

Supplemental Figure S6. Syntenic dot plots showing alignment of the *B. carinata* var. Gomenzer genome against the *B. carinata* var. Zd-1 genome.

Supplemental Figure S7. Gene map of the *B. carinata* chloroplast genome.

Supplemental Figure S8. Insertion times of intact LTR groups in Triangle of U species.

Supplemental Figure S9. KEGG and GO analysis for the *B. carinata* Gomenzer genome.

Supplemental Figure S10. Patterns of gene expression dominance and relative selection pressure between homoeologous gene pairs.

Supplemental Figure S11. *FA Elongase 1* phylogenetic tree.

Supplemental Figure S12. Gene structures of *FA Elongase 1* homologs in the Triangle of U B genomes.

Supplemental Figure S13. Gene structures of *FA Elongase 1* homologs in the Triangle of U C genomes.

Supplemental Figure S14. Gene structures of *Fatty Acyl Elongase 1* homologs in the Triangle of U A genomes.

Supplemental Figure S15. *FAR 1* phylogenetic tree.

Supplemental Figure S16. Gene structures of *FARe 1* homologs in the *B. carinata* genome.

Supplemental Figure S17. Gene structure of *Flowering Locus C* homologs in the *B. carinata* Bc subgenome.

Supplementary Figure S18. Gene structure of *Flowering Locus C* homologs in the *B. carinata* Cc subgenome.

Supplemental Data Set 1. PacBio sequencing read summary for *B. carinata* Gomenzer from the SMRT portal.

Supplemental Data Set 2. Draft genome scaffolding results and error correction for *B. carinata* Gomenzer.

Supplemental Data Set 3. Illumina sequencing read summary from MultiQC for *B. carinata* Gomenzer.

Supplemental Data Set 4. Error correction results.

Supplemental Data Set 5. Scaffolding results for the final *B. carinata* Gomenzer genome assembly.

Supplemental Data Set 6. Quality statistics for the Triangle of U genome assemblies.

Supplemental Data Set 7. Estimation of *B. carinata* Gomenzer genome size by flow cytometry.

Supplemental Data Set 8. Comparison of LAI values among Triangle of U genome assemblies.

Supplemental Data Set 9. Comparison of gene annotations among Triangle of U genome assemblies.

Supplemental Data Set 10. CDS annotation statistics.

Supplemental Data Set 11. Number of gene duplications among Triangle of U genome assemblies.

Supplemental Data Set 12. Comparison of repeat analysis among Triangle of U species using *de novo* libraries.

Supplemental Data Set 13. Repeat content per chromosome for the *B. carinata* genome.

Supplemental Data Set 14. Summary of TF families detected in the Triangle of U genomes.

Supplemental Data Set 15. List of collinear genes between *B. carinata* and its progenitor species.

Supplemental Data Set 16. Comparison of syntenic analyses of *B. carinata* subgenomes and their respective shared Triangle of U genomes.

Supplemental Data Set 17. Homoeolog exchange analysis among the Triangle of U allotetraploids.

Supplemental Data Set 18. Gene IDs for *B. carinata* homoeologs involved in HEs and their colinear orthologs in the progenitor species.

Supplemental Data Set 19. Homoeolog expression bias between *B. carinata* subgenomes.

Supplemental Data Set 20. Sequence evolution rates for duplicated genes among the Triangle of U species.

Supplemental Data Set 21. Genes involved in acyl lipid metabolism in Arabidopsis and their homologous sequences in Triangle of U species.

Supplemental Data Set 22. Transcript abundance of *FAE1* homologs in the *B. carinata* genome.

Supplemental Data Set 23. Transcript abundance of *FAR1* homologs in the *B. carinata* genome.

Supplemental Data Set 24. Transcript abundance of *FLC* homologs in the *B. carinata* genome.

Supplemental Data Set 25. RNA-Seq read alignment and assignment rates.

Acknowledgments

We gratefully acknowledge the support of the Nevada Agricultural Experiment Station (Grant No. NEV00384) and VPRI research funding (University of Nevada, Reno). The authors would like to thank the Germplasm Resources Information Network of the USDA-ARS for their help in providing genetic resources. The authors would also like to thank Maggie Weitzman at the Genomics & Cell Characterization Core Facility (GC3F) at the University of Oregon for performing the PacBio library preparation and sequencing. Further thanks go to Diana Burkart-Waco for carrying out the Illumina sequencing at the DNA Technologies and Expression Analysis Cores at the UC Davis Genome Center, supported by NIH Shared Instrumentation (NIH 1S10OD010786-01). We would also like to thank Sitharam Ramaswami at the Genome Technology Center at NYU Langone Health for performing Hi-C sequencing. Lastly, we would like to thank the Office of Information Technology and Research & Innovation at the University of Nevada, Reno for providing paid access to the Pronghorn High-Performance Computing Cluster.

Funding

We gratefully acknowledge the support of the Nevada Agricultural Experiment Station (Grant No. NEV00384) and VPRI research funding (University of Nevada, Reno).The Pires lab is funded by the National Science Foundation (NSF IOS 1339156) and the Department of Energy Defense Threat Reduction Agency (HDTRA 1-16-1-0048). The Edger lab is funded by the National Science Foundation (NSF IOS 2029959). The Mason lab is partially funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy (EXC 2070 - 390732324). The Alvarez-Ponce lab is funded by the National Science Foundation (NSF MCB 1818288).

Conflict of interest statement. None declared.

References

Alix K, Gérard PR, Schwarzacher T, Heslop-Harrison JS (2017) Polyploidy and interspecific hybridization: partners for adaptation, speciation and evolution in plants. Ann Bot **120**: 183–194

Allender CJ, King GJ (2010) Origins of the amphiploid species *Brassica napus* L. investigated by chloroplast and nuclear molecular markers. BMC Plant Biol **10**: 54

Alonge M, Soyk S, Ramakrishnan S, Wang X, Goodwin S, Sedlazeck FJ, Lippman ZB, Schatz MC (2019) RaGOO: fast and accurate reference-guided scaffolding of draft genomes. Genome Biol **20**: 224

Alvarez-Ponce D (2014) Why proteins evolve at different rates: The determinants of proteins' rates of evolution. Natural Selection. Taylor & Francis Group, Abingdon, pp 126–178

Ammiraju JSS, Zuccolo A, Yu Y, Song X, Piegu B, Chevalier F, Walling JG, Ma J, Talag J, Brar DS, et al. (2007) Evolutionary dynamics of an ancient retrotransposon family provides insights into evolution of genome size in the genus *Oryza*. Plant J **52**: 342–351

An H, Qi X, Gaynor ML, Hao Y, Gebken SC, Mabry ME, McAlvay AC, Teakle GR, Conant GC, Barker MS, et al. (2019) Transcriptome and organellar sequencing highlights the complex origin and diversification of allotetraploid *Brassica napus*. Nat Commun **10**: 2878

Anderson SN, Stitzer MC, Brohammer AB, Zhou P, Noshay JM, O'Connor CH, Hirsch CD, Ross-Ibarra J, Hirsch CN, Springer NM (2019) Transposable elements contribute to dynamic genome content in maize. Plant J **100**: 1052–1065

Argout X, Salse J, Aury J-M, Guiltinan MJ, Droc G, Gouzy J, Allegre M, Chaparro C, Legavre T, Maximova SN, et al. (2011) The genome of *Theobroma cacao*. Nat Genetics **43**: 101–108

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

Baduel P, Quadrana L, Hunter B, Bomblies K, Colot V (2019) Relaxed purifying selection in autopolyploids drives transposable element over-accumulation which provides variants for local adaptation. Nat Commun **10**: 5818

Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Prjibelski AD, et al. (2012) SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. J Comput Biol **19**: 455–477

Barbaglia AM, Klusman KM, Higgins J, Shaw JR, Hannah LC, Lal SK (2012) Gene capture by Helitron transposons reshuffles the transcriptome of maize. Genetics **190**: 965–975

Basili M, Rossi MA (2018) *Brassica carinata*-derived biodiesel production: economics, sustainability and policies. The Italian case. J Cleaner Prod **191**: 40–47

Bao W, Kojima KK, Kohany O (2015) Repbase update, a database of repetitive elements in eukaryotic genomes. Mobile DNA **6**: 11

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

Bennett MD, Leitch IJ (2005) Nuclear DNA amounts in angiosperms: progress, problems and prospects. Ann Bot **95**: 45–90

Bennetzen JL (2002) Mechanisms and rates of genome expansion and contraction in flowering plants. Genetica **115**: 29–36

Besemer J, Borodovsky M (2005) GeneMark: web software for gene finding in prokaryotes, eukaryotes and viruses. Nucleic Acids Res **33**: W451–W454

Bird KA, Niederhuth C, Ou S, Gehan M, Pires JC, Xiong Z, VanBuren R, Edger PP (2020) Replaying the evolutionary tape to investigate subgenome dominance in allopolyploid *Brassica napus*. New Phytol **230**: 354–371

Bird KA, VanBuren R, Puzey JR, Edger PP (2018) The causes and consequences of subgenome dominance in hybrids and recent polyploids. New Phytologist **220**: 87–93

Boutet E, Lieberherr D, Tognolli M, Schneider M, Bansal P, Bridge AJ, Poux S, Bougueleret L, Xenarios I (2016) UniProtKB/Swiss-Prot, the manually annotated section of the UniProt knowledgebase: how to use the entry view. *In* D Edwards, ed, Plant Bioinformatics: Methods and Protocols, Methods in Molecular Biology. Springer, New York, NY, pp 23–54

Boutte J, Maillet L, Chaussepied T, Letort S, Aury J-M, Belser C, Boideau F, Brunet A, Coriton O, Deniot G, et al. (2020) Genome size variation and comparative genomics reveal intraspecific diversity in *Brassica rapa*. Front Plant Sci **11**: 577536

Bowers JE, Chapman BA, Rong J, Paterson AH (2003) Unravelling angiosperm genome evolution by phylogenetic analysis of chromosomal duplication events. Nature **422**: 433–438

Brůna T, Hoff KJ, Lomsadze A, Stanke M, Borodovsky M (2021) BRAKER2: automatic eukaryotic genome annotation with GeneMark-EP + and AUGUSTUS supported by a protein database. NAR Genom Bioinform **3**: lqaa108

Cai X, Wu J, Liang J, Lin R, Zhang K, Cheng F, Wang X (2020) Improved *Brassica oleracea* JZS assembly reveals significant changing of LTR-RT dynamics in different morphotypes. Theor Appl Genet **133**: 3187–3199

Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL (2009) BLAST + : architecture and applications. BMC Bioinformatics **10**: 421

Campbell MS, Law M, Holt C, Stein JC, Moghe GD, Hufnagel DE, Lei J, Achawanantakun R, Jiao D, Lawrence CJ, et al. (2014) MAKER-P: a tool kit for the rapid creation, management, and quality control of plant genome annotations. Plant Physiol **164**: 513–524

Canapa A, Barucca M, Biscotti MA, Forconi M, Olmo E (2015) Transposons, Genome Size, and Evolutionary Insights in Animals. Cytogenet Genome Res **147**: 217–239

Cantarel BL, Korf I, Robb SMC, Parra G, Ross E, Moore B, Holt C, Sánchez Alvarado A, Yandell M (2008) MAKER: an easy-to-use annotation pipeline designed for emerging model organism genomes. Genome Res **18**: 188–196

Capella-Gutiérrez S, Silla-Martínez JM, Gabaldón T (2009) trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. Bioinformatics **25**: 1972–1973

Cardone M, Mazzoncini M, Menini S, Rocco V, Senatore A, Seggiani M, Vitolo S (2003) *Brassica carinata* as an alternative oil crop for the production of biodiesel in Italy: agronomic evaluation, fuel production by transesterification and characterization. Biomass Bioenergy **25**: 623–636

Catlin NS, Josephs EB (2022) The important contribution of transposable elements to phenotypic variation and evolution. Curr Opin Plant Biol **65**: 102140

Chalhoub B, Denoeud F, Liu S, Parkin IAP, Tang H, Wang X, Chiquet J, Belcram H, Tong C, Samans B, et al. (2014) Early allopolyploid evolution in the post-Neolithic *Brassica napus* oilseed genome. Science **345**: 950–953

Chen X, Tong C, Zhang X, Song A, Hu M, Dong W, Chen F, Wang Y, Tu J, Liu S, et al. (2021) A high-quality *Brassica napus* genome reveals expansion of transposable elements, subgenome evolution and disease resistance. Plant Biotechnol J **19**: 615–630

Cheng F, Liang J, Cai C, Cai X, Wu J, Wang X (2017) Genome sequencing supports a multi-vertex model for Brassiceae species. Curr Opin Plant Biol **36**: 79–87

Cheng F, Liu S, Wu J, Fang L, Sun S, Liu B, Li P, Hua W, Wang X (2011) BRAD, the genetics and genomics database for *Brassica* plants. BMC Plant Biol **11**: 136

Cheng F, Wu J, Cai X, Liang J, Freeling M, Wang X (2018) Gene retention, fractionation and subgenome differences in polyploid plants. Nat Plants **4**: 258–268

Cheng S, van den Bergh E, Zeng P, Zhong X, Xu J, Liu X, Hofberger J, Bruijn S de Bhide AS, Kuelahoglu C, et al. (2013) The *Tarenaya hassleriana* genome provides insight into reproductive trait and genome evolution of crucifers. Plant Cell **25**: 2813–2830

Chin CS, Alexander DH, Marks P, Klammer AA, Drake J, Heiner C, Clum A, Copeland A, Huddleston J, Eichler EE, et al. (2013) Nonhybrid, finished microbial genome assemblies from long-read SMRT sequencing data. Nat Methods **10**: 563–569

Chiron H, Wilmer J, Lucas MO, Nesi N, Delseny M, Devic M, Roscoe TJ (2015) Regulation of Fatty Acid Elongation1 expression in embryonic and vascular tissues of *Brassica napus*. Plant Mol Biol **88**: 65–83

Crescente JM, Zavallo D, Helguera M, Vanzetti LS (2018) MITE Tracker: an accurate approach to identify miniature inverted-repeat transposable elements in large genomes. BMC Bioinformatics **19**: 348

Cultrone A, Domínguez YR, Drevet C, Scazzocchio C, Fernández-Martín R (2007) The tightly regulated promoter of the xanA gene of *Aspergillus nidulans* is included in a Helitron. Mol Microbiol **63**: 1577–1587

Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M, Gingeras TR (2013) STAR: ultrafast universal RNA-seq aligner. Bioinformatics **29**: 15–21

Doyle JJ, Egan AN (2010) Dating the origins of polyploidy events. New Phytologist **186**: 73–85

Du J, Tian Z, Bowen NJ, Schmutz J, Shoemaker RC, Ma J (2010) Bifurcation and enhancement of autonomous-nonautonomous retrotransposon partnership through LTR swapping in soybean. Plant Cell **22**: 48–61

Danecek P, Bonfield JK, Liddle J, Marshall J, Ohan V, Pollard MO, Whitwham A, Keane T, McCarthy SA, Davies RM, et al. (2021) Twelve years of SAMtools and BCFtools. GigaScience **10**: giab008

Dudchenko O, Batra SS, Omer AD, Nyquist SK, Hoeger M, Durand NC, Shamim MS, Machol I, Lander ES, Aiden AP, et al. (2017) *De novo* assembly of the *Aedes aegypti* genome using Hi-C yields chromosome-length scaffolds. Science **356**: 92–95

Edgar RC (2004) MUSCLE: a multiple sequence alignment method with reduced time and space complexity. BMC Bioinformatics **5**: 113

Edger PP, Poorten TJ, VanBuren R, Hardigan MA, Colle M, McKain MR, Smith RD, Teresi SJ, Nelson ADL, Wai CM, et al. (2019) Origin and evolution of the octoploid strawberry genome. Nat Genet **51**: 541–547

Edger PP, Smith R, McKain MR, Cooley AM, Vallejo-Marin M, Yuan Y, Bewick AJ, Ji L, Platts AE, Bowman MJ, et al. (2017) Subgenome dominance in an interspecific hybrid, synthetic allopolyploid, and a 140-year-old naturally established neo-allopolyploid monkeyflower. Plant Cell **29**: 2150–2167

Edger PP, Heidel-Fischer HM, Bekaert M, Rota J, Glöckner G, Platts AE, Heckel DG, Der JP, Wafula EK, Tang M, et al. (2015) The butterfly plant arms-race escalated by gene and genome duplications. Proc Natl Acad Sci USA **112**: 8362–8366

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

Feschotte C, Wessler SR (2001) Treasures in the attic: rolling circle transposons discovered in eukaryotic genomes. Proc Natl Acad Sci **98**: 8923–8924

Folayan AJ, Anawe PAL, Aladejare AE, Ayeni AO (2019) Experimental investigation of the effect of fatty acids configuration, chain length, branching and degree of unsaturation on biodiesel fuel properties obtained from lauric oils, high-oleic and high-linoleic vegetable oil biomass. Energy Rep **5**: 793–806

Freeling M, Woodhouse MR, Subramaniam S, Turco G, Lisch D, Schnable JC (2012) Fractionation mutagenesis and similar consequences of mechanisms removing dispensable or less-expressed DNA in plants. Curr Opin Plant Biol **15**: 131–139

Friedman WE (2009) The meaning of Darwin's "abominable mystery." Am J Bot **96**: 5–21

Frith MC, Wan R, Horton P (2010) Incorporating sequence quality data into alignment improves DNA read mapping. Nucleic Acids Res **38**: e100

Gaeta RT, Chris Pires J (2010) Homoeologous recombination in allopolyploids: the polyploid ratchet. New Phytologist **186**: 18–28

Gao C, Zhou G, Ma C, Zhai W, Zhang T, Liu Z, Yang Y, Wu M, Yue Y, Duan Z, et al. (2016) Helitron-like transposons contributed to the mating system transition from out-crossing to self-fertilizing in polyploid *Brassica napus* L. Sci Rep **6**: 33785

Garrison E, Marth G (2012) Haplotype-based variant detection from short-read sequencing. arXiv:1207.3907 [q-bio.GN]

Gazzani S, Gendall AR, Lister C, Dean C (2003) Analysis of the molecular basis of flowering time variation in *Arabidopsis* accessions. Plant Physiol **132**: 1107–1114

Glover NM, Redestig H, Dessimoz C (2016) Homoeologs: what are they and how do we infer them? Trend Plant Sci **21**: 609–621

Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, Adiconis X, Fan L, Raychowdhury R, Zeng Q, et al. (2011) Trinity: reconstructing a full-length transcriptome without a genome from RNA-Seq data. Nat Biotechnol **29**: 644–652

Grabundzija I, Messing SA, Thomas J, Cosby RL, Bilic I, Miskey C, Gogol-Döring A, Kapitonov V, Diem T, Dalda A, et al. (2016) A Helitron transposon reconstructed from bats reveals a novel mechanism of genome shuffling in eukaryotes. Nat Commun **7**: 10716

Guo J, Xu W, Yu X, Shen H, Li H, Cheng D, Liu A, Liu J, Liu C, Zhao S, et al. (2016) Cuticular wax accumulation is associated with drought tolerance in wheat near-isogenic lines. Front Plant Sci **7**: 1809

Gupta V, Mukhopadhyay A, Arumugam N, Sodhi YS, Pental D, Pradhan AK (2004) Molecular tagging of erucic acid trait in oilseed mustard (*Brassica juncea*) by QTL mapping and single nucleotide polymorphisms in FAE1 gene. Theor Appl Genet **108**: 743–749

Haas BJ, Delcher AL, Mount SM, Wortman JR, Smith RK, Hannick LI, Maiti R, Ronning CM, Rusch DB, Town CD, et al. (2003) Improving the *Arabidopsis* genome annotation using maximal transcript alignment assemblies. Nucleic Acids Res **31**: 5654–5666

Hagos R, Shaibu AS, Zhang L, Cai X, Liang J, Wu J, Lin R, Wang X (2020) Ethiopian mustard (*Brassica carinata* A. Braun) as an alternative energy source and sustainable crop. Sustainability **12**: 7492

Haudry A, Platts AE, Vello E, Hoen DR, Leclercq M, Williamson RJ, Forczek E, Joly-Lopez Z, Steffen JG, Hazzouri KM, et al. (2013) An atlas of over 90,000 conserved noncoding sequences provides insight into crucifer regulatory regions. Nat Genetics **45**: 891–898

Hawkins JS, Kim H, Nason JD, Wing RA, Wendel JF (2006) Differential lineage-specific amplification of transposable elements is responsible for genome size variation in *Gossypium*. Genome Res **16**: 1252–1261

He Z, Wang L, Harper AL, Havlickova L, Pradhan AK, Parkin IAP, Bancroft I (2017) Extensive homoeologous genome exchanges in allopolyploid crops revealed by mRNAseq-based visualization. Plant Biotechnol J **15**: 594–604

Healey A, Furtado A, Cooper T, Henry RJ (2014) Protocol: a simple method for extracting next-generation sequencing quality genomic DNA from recalcitrant plant species. Plant Methods **10**: 21

Helsen J, Voordeckers K, Vanderwaeren L, Santermans T, Tsontaki M, Verstrepen KJ, Jelier R (2020) Gene loss predictably drives evolutionary adaptation. Mol Biol Evol **37**: 2989–3002

Higgins EE, Clarke WE, Howell EC, Armstrong SJ, Parkin IAP (2018) Detecting *de novo* homoeologous recombination events in cultivated *Brassica napus* using a genome-wide SNP array. G3 Genes|Genomes|Genetics **8**: 2673–2683

Hinze LL, Gazave E, Gore MA, Fang DD, Scheffler BE, Yu JZ, Jones DC, Frelichowski J, Percy RG (2016) Genetic diversity of the two commercial tetraploid cotton species in the *Gossypium* diversity reference set. J Heredity **107**: 274–286

Hloušková P, Mandáková T, Pouch M, Trávníček P, Lysak MA (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

Hollister JD, Gaut BS (2009) Epigenetic silencing of transposable elements: a trade-off between reduced transposition and deleterious effects on neighboring gene expression. Genome Res **19**: 1419–1428

Hu K, Xu K, Wen J, Yi B, Shen J, Ma C, Fu T, Ouyang Y, Tu J (2019) Helitron distribution in Brassicaceae and whole Genome Helitron density as a character for distinguishing plant species. BMC Bioinform **20**: https://doi.org/10.1186/s12859-019-2945-8

Hu TT, Pattyn P, Bakker EG, Cao J, Cheng J-F, Clark RM, Fahlgren N, Fawcett JA, Grimwood J, Gundlach H, et al. (2011) The *Arabidopsis lyrata* genome sequence and the basis of rapid genome size change. Nat Genet **43**: 476–481

Hu Y, Wu X, Jin G, Peng J, Leng R, Li L, Gui D, Fan C, Zhang C (2022) Rapid genome evolution and adaptation of *Thlaspi arvense* mediated by recurrent RNA-based and tandem gene duplications. Front Plant Sci **12**: 772655.

Huang G, Wu Z, Percy RG, Bai M, Li Y, Frelichowski JE, Hu J, Wang K, Yu JZ, Zhu Y (2020) Genome sequence of *Gossypium herbaceum* and genome updates of *Gossypium arboreum* and *Gossypium hirsutum* provide insights into cotton A-genome evolution. Nat Genet **52**: 516–524

Huang J, Xue C, Wang H, Wang L, Schmidt W, Shen R, Lan P (2017) Genes of acyl carrier protein family show different expression profiles and overexpression of acyl carrier protein 5 modulates fatty acid composition and enhances salt stress tolerance in *Arabidopsis*. Front Plant Sci **8**: 987

Huang K, Li CF, Wu J, Wei JH, Zou Y, Han MJ, Zhou ZY (2016) Enhancer activity of Helitron in sericin-1 gene promoter from *Bombyx mori*. Insect Sci **23**: 396–405

Huang S, Deng L, Guan M, Li J, Lu K, Wang H, Fu D, Mason AS, Liu S, Hua W (2013) Identification of genome-wide single

nucleotide polymorphisms in allopolyploid crop *Brassica napus*. BMC Genomics **14**: 717

Jaillon O, Aury J-M, Noel B, Policriti A, Clepet C, Casagrande A, Choisne N, Aubourg S, Vitulo N, Jubin C, et al. (2007) The grapevine genome sequence suggests ancestral hexaploidization in major angiosperm phyla. Nature **449**: 463–467

Japelaghi RH, Haddad R, Garoosi GA (2011) Rapid and efficient isolation of high quality nucleic acids from plant tissues rich in polyphenols and polysaccharides. Mol Biotechnol **49**: 129–137

Jones P, Binns D, Chang H-Y, Fraser M, Li W, McAnulla C, McWilliam H, Maslen J, Mitchell A, Nuka G, et al. (2014) InterProScan 5: genome-scale protein function classification. Bioinformatics **30**: 1236–1240

Kagale S, Koh C, Nixon J, Bollina V, Clarke WE, Tuteja R, Spillane C, Robinson SJ, Links MG, Clarke C, et al. (2014) The emerging biofuel crop *Camelina sativa* retains a highly undifferentiated hexaploid genome structure. Nat Commun **5**: 1–11

Kang L, Qian L, Zheng M, Chen L, Chen H, Yang L, You L, Yang B, Yan M, Gu Y, et al. (2021) Genomic insights into the origin, domestication and diversification of *Brassica juncea*. Nat Genet **53**: 1392–1402

Kazamia E, Smith AG (2014) Assessing the environmental sustainability of biofuels. Trend Plant Sci **19**: 615–618

Kazaz S, Barthole G, Domergue F, Ettaki H, To A, Vasselon D, De Vos D, Belcram K, Lepiniec L, Baud S (2020) Differential activation of partially redundant Δ9 stearoyl-ACP desaturase genes is critical for omega-9 monounsaturated fatty acid biosynthesis during seed development in *Arabidopsis*. Plant Cell **32**: 3613–3637

Ke J, Wen TN, Nikolau BJ, Wurtele ES (2000) Coordinate regulation of the nuclear and plastidic genes coding for the subunits of the heteromeric acetyl-coenzyme A carboxylase. Plant Physiol **122**: 1057–1072

Khedikar Y, Clarke WE, Chen L, Higgins EE, Kagale S, Koh CS, Bennett R, Parkin IAP (2020) Narrow genetic base shapes population structure and linkage disequilibrium in an industrial oilseed crop, Brassica carinata A. Braun Sci Rep **10**: 12629

Kidwell MG (2002) Transposable elements and the evolution of genome size in eukaryotes. Genetica **115**: 49–63

Kim CK, Seol Y-J, Perumal S, Lee J, Waminal NE, Jayakodi M, Lee S-C, Jin S, Choi B-S, Yu Y, et al. (2018) Re-exploration of U's Triangle *Brassica* species based on chloroplast genomes and 45S nrDNA sequences. Sci Rep **8**: 7353

Kitashiba H, Li F, Hirakawa H, Kawanabe T, Zou Z, Hasegawa Y, Tonosaki K, Shirasawa S, Fukushima A, Yokoi S, et al. (2014) Draft sequences of the radish (*Raphanus sativus* L.) genome. DNA Res **21**: 481–490

Kliebenstein DJ (2008) A role for gene duplication and natural variation of gene expression in the evolution of metabolism. PLoS One **3**: e1838

Klopfenstein DV, Zhang L, Pedersen BS, Ramírez F, Vesztrocy AW, Naldi A, Mungall CJ, Yunes JM, Botvinnik O, Weigel M, et al. (2018) GOATOOLS: a python library for gene ontology analyses. Sci Rep **8**: 10872

Koren S, Walenz BP, Berlin K, Miller JR, Bergman NH, Phillippy AM (2017) Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. Genome Res **27**: 722–736

Korf I (2004) Gene finding in novel genomes. BMC Bioinformatics **5**: 59

Krishnakumar V, Hanlon MR, Contrino S, Ferlanti ES, Karamycheva S, Kim M, Rosen BD, Cheng C-Y, Moreira W, Mock SA, et al. (2015) Araport: the *Arabidopsis* information portal. Nucleic Acids Res **43**: D1003–D1009

Kumar A, Singh P, Singh DP, Singh H, Sharma HC (1984) Differences in osmoregulation in *Brassica* species. Ann Bot **54**: 537–542

Kumar S, Stecher G, Suleski M, Hedges SB (2017) TimeTree: a resource for timelines, timetrees, and divergence times. Mol Biol Evol **34**: 1812–1819

Lang D, Weiche B, Timmerhaus G, Richardt S, Riaño-Pachón DM, Corrêa LGG, Reski R, Mueller-Roeber B, Rensing SA (2010) Genome-wide phylogenetic comparative analysis of plant transcriptional regulation: a timeline of loss, gain, expansion, and correlation with complexity. Genome Biol Evol **2**: 488–503

Langmead B, Salzberg SL (2012) Fast gapped-read alignment with Bowtie 2. Nat Methods **9**: 357–359

Lee H, Chawla HS, Obermeier C, Dreyer F, Abbadi A, Snowdon R (2020) Chromosome-scale assembly of winter oilseed rape *Brassica napus*. Front Plant Sci **11**: 496

Leitch AR, Leitch IJ (2008) Genomic plasticity and the diversity of polyploid plants. Science **320**: 481–483

Letunic I, Bork P (2021) Interactive Tree Of Life (iTOL) v5: an online tool for phylogenetic tree display and annotation. Nucleic Acids Res **49**: W293–W296

Li H (2011) Tabix: fast retrieval of sequence features from generic TAB-delimited files. Bioinformatics **27**: 718–719

Li L, Yu XX, Guo CC, Duan XS, Shan HY, Zhang R, Xu GX, Kong HZ (2015) Interactions among proteins of floral MADS-box genes in *Nuphar pumila* (Nymphaeaceae) and the most recent common ancestor of extant angiosperms help understand the underlying mechanisms of the origin of the flower. J Syst Evol **53**: 285–296

Li M, Wang R, Wu X, Wang J (2020) Homoeolog expression bias and expression level dominance (ELD) in four tissues of natural allotetraploid *Brassica napus*. BMC Genomics **21**: 330

Li P, Zhang S, Li F, Zhang S, Zhang H, Wang X, Sun R, Bonnema G, Borm TJA (2017) A phylogenetic analysis of chloroplast genomes elucidates the relationships of the six economically important *Brassica* species comprising the Triangle of U. Front Plant Sci **8**: 111

Liao Y, Smyth GK, Shi W (2014) featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. Bioinformatics **30**: 923–930

Liu S, Liu Y, Yang X, Tong C, Edwards D, Parkin IAP, Zhao M, Ma J, Yu J, Huang S, et al. (2014) The *Brassica oleracea* genome reveals the asymmetrical evolution of polyploid genomes. Nat Commun **5**: 1–11

Liu XP, Tu JX, Chen BY, Fu TD (2004) Identification of the linkage relationship between the flower colour and the content of erucic acid in the resynthesized *Brassica napus* L. Yi Chuan Xue Bao **31**: 357–362

Lloyd A, Blary A, Charif D, Charpentier C, Tran J, Balzergue S, Delannoy E, Rigaill G, Jenczewski E (2018) Homoeologous exchanges cause extensive dosage-dependent gene expression changes in an allopolyploid crop. New Phytologist **217**: 367–377

Love MI, Huber W, Anders S (2014) Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol **15**: 550

Lu K, Wei L, Li X, Wang Y, Wu J, Liu M, Zhang C, Chen Z, Xiao Z, Jian H, et al. (2019) Whole-genome resequencing reveals *Brassica napus* origin and genetic loci involved in its improvement. Nat Commun **10**: 1154

Lv H, Wang Y, Han F, Ji J, Fang Z, Zhuang M, Li Z, Zhang Y, Yang L (2020) A high-quality reference genome for cabbage obtained with SMRT reveals novel genomic features and evolutionary characteristics. Sci Rep **10**: 12394

Lynch M, Conery JS (2000) The evolutionary fate and consequences of duplicate genes. Science **290**: 1151–1155

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

Malik RS (1990) Prospects for *Brassica carinata* as an oilseed crop in India. Ex Agric **26**: 125–129

Mason AS, Wendel JF (2020) Homoeologous exchanges, segmental allopolyploidy, and polyploid genome evolution. Front Genet **11**: 1014

Meyer RS, Purugganan MD (2013) Evolution of crop species: genetics of domestication and diversification. Nat Rev Genet **14**: 840–852

Millar AA, Kunst L (2003) Very-long-chain fatty acid biosynthesis is controlled through the expression and specificity of the condensing enzyme. Plant J **12**: 121–131

Ming R, Hou S, Feng Y, Yu Q, Dionne-Laporte A, Saw JH, Senin P, Wang W, Ly BV, Lewis KLT, et al. (2008) The draft genome of the transgenic tropical fruit tree papaya (Carica papaya Linnaeus). Nature **452**: 991–996

Monroe JG, McKay JK, Weigel D, Flood PJ (2021) The population genomics of adaptive loss of function. Heredity **126**: 383–395

Morgante M, Brunner S, Pea G, Fengler K, Zuccolo A, Rafalski A (2005) Gene duplication and exon shuffling by helitron-like transposons generate intraspecies diversity in maize. Nat Genet **37**: 997–1002

Morgante M, De Paoli E, Radovic S (2007) Transposable elements and the plant pan-genomes. Curr Opin Plant Biol **10**: 149–155

Muyle A, Seymour D, Darzentas N, Primetis E, Gaut BS, Bousios A (2021) Gene capture by transposable elements leads to epigenetic conflict in maize. Mol Plant **14**: 237–252

Nagaharu U (1935) Genome analysis in *Brassica* with special reference to the experimental formation of *B. napus* and peculiar mode of fertilization. Jpn J Bot **7**: 389–452

Naoumkina M, Hinchliffe DJ, Fang DD, Florane CB, Thyssen GN (2017) Role of xyloglucan in cotton (*Gossypium hirsutum* L.) fiber elongation of the short fiber mutant Ligon lintless-2 (*Li2*). Gene **626**: 227–233

Navabi ZK, Huebert T, Sharpe AG, O'Neill CM, Bancroft I, Parkin IA (2013) Conserved microstructure of the *Brassica* B Genome of *Brassica nigra* in relation to homologous regions of *Arabidopsis thaliana*, *B. rapa* and *B. oleracea*. BMC Genomics **14**: 250

Nguyen LT, Schmidt HA, von Haeseler A, Minh BQ (2015) IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. Mol Biol Evol **32**: 268–274

Nicolas SD, Monod H, Eber F, Chèvre AM, Jenczewski E (2012) Non-random distribution of extensive chromosome rearrangements in *Brassica napus* depends on genome organization. Plant J **70**: 691–703

Ohno S (1970) Evolution by Gene Duplication, 1st ed. Springer, Berlin, Heidelberg, Germany

Ojiewo C, Teklewold A, Weyesa B, Tesfaye M, Wakjira A, Samali S (2013) Good agricultural practices for production of Ethiopian mustard (*Brassica carinata* A. Braun) in sub-Saharan Africa. Scripta Hortic **15**: 103–114

Ou S, Chen J, Jiang N (2018) Assessing genome assembly quality using the LTR Assembly Index (LAI). Nucleic Acids Res **46**: e126

Pál C, Papp B, Hurst LD (2001) Highly expressed genes in yeast evolve slowly. Genetics **158**: 927–931

Palazzo A, Lorusso P, Miskey C, Walisko O, Gerbino A, Marobbio CMT, Ivics Z, Marsano RM (2019) Transcriptionally promiscuous "blurry" promoters in Tc1/mariner transposons allow transcription in distantly related genomes. Mobile DNA **10**: 13

Pamilo P, Nei M (1988) Relationships between gene trees and species trees. Mol Biol Evol **5**: 568–583

Parisod C, Alix K, Just J, Petit M, Sarilar V, Mhiri C, Ainouche M, Chalhoub B, Grandbastien MA (2010) Impact of transposable elements on the organization and function of allopolyploid genomes. New Phytologist **186**: 37–45

Paritosh K, Yadava SK, Singh P, Bhayana L, Mukhopadhyay A, Gupta V, Bisht NC, Zhang J, Kudrna DA, Copetti D, et al. (2021) A chromosome-scale assembly of allotetraploid *Brassica juncea* (AABB) elucidates comparative architecture of the A and B genomes. Plant Biotechnol J **19**: 602–614

Park YW, Baba K, Furuta Y, Iida I, Sameshima K, Arai M, Hayashi T (2004) Enhancement of growth and cellulose accumulation by overexpression of xyloglucanase in poplar. FEBS Lett **564**: 183–187

Park YW, Tominaga R, Sugiyama J, Furuta Y, Tanimoto E, Samejima M, Sakai F, Hayashi T (2003) Enhancement of growth by expression of poplar cellulase in *Arabidopsis thaliana*. Plant J **33**: 1099–1106

Parkin IA, Koh C, Tang H, Robinson SJ, Kagale S, Clarke WE, Town CD, Nixon J, Krishnakumar V, Bidwell SL, et al. (2014) Transcriptome and methylome profiling reveals relics of genome dominance in the mesopolyploid *Brassica oleracea*. Genome Biol **15**: R77

Pegueroles C, Laurie S, Albà MM (2013) Accelerated evolution after gene duplication: a time-dependent process affecting just one copy. Mol Biol Evol **30**: 1830–1842

Perumal S, Koh CS, Jin L, Buchwaldt M, Higgins EE, Zheng C, Sankoff D, Robinson SJ, 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

Qi X, An H, Hall TE, Di C, Blischak PD, McKibben MTW, Hao Y, Conant GC, Pires JC, Barker MS (2021) Genes derived from ancient polyploidy have higher genetic diversity and are associated with domestication in *Brassica rapa*. New Phytologist **230**: 372–386

Qian L, Qian W, Snowdon RJ (2014) Sub-genomic selection patterns as a signature of breeding in the allopolyploid *Brassica napus* genome. BMC Genomics **15**: 1170

Qian W, Meng J, Li M, Frauen M, Sass O, Noack J, Jung C (2006) Introgression of genomic components from Chinese *Brassica rapa* contributes to widening the genetic diversity in rapeseed (*B. napus* L.), with emphasis on the evolution of Chinese rapeseed. Theor Appl Genet **113**: 49–54

Qiao X, Li Q, Yin H, Qi K, Li L, Wang R, Zhang S, Paterson AH (2019) Gene duplication and evolution in recurring polyploidization–diploidization cycles in plants. Genome Biol **20**: 38

Quadrana L (2020) The contribution of transposable elements to transcriptional novelty in plants: the FLC affair. Transcription **11**: 192–198

Rahman M, Hoque A, Roy J (2022) Linkage disequilibrium and population structure in a core collection of *Brassica napus* (L.). PLoS One **17**: e0250310

Ranallo-Benavidez TR, Jaron KS, Schatz MC (2020) GenomeScope 2.0 and Smudgeplot for reference-free profiling of polyploid genomes. Nat Commun **11**: 1432

Renny-Byfield S, Rodgers-Melnick E, Ross-Ibarra J (2017) Gene fractionation and function in the ancient subgenomes of maize. Mol Biol Evol **34**: 1825–1832

Rizzon C, Ponger L, Gaut BS (2006) Striking similarities in the genomic distribution of tandemly arrayed genes in *Arabidopsis* and rice. PLoS Comput Biol **2**: e115

Robinson MD, McCarthy DJ, Smyth GK (2010) edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics **26**: 139–140

Rousseau-Gueutin M, Belser C, Da Silva C, Richard G, Istace B, Cruaud C, Falentin C, Boideau F, Boutte J, Delourme R, et al. (2020) Long-read assembly of the *Brassica napus* reference genome Darmor-bzh. GigaScience **9**: giaa137

Rowland O, Domergue F (2012) Plant fatty acyl reductases: enzymes generating fatty alcohols for protective layers with potential for industrial applications. Plant Sci **193–194**: 28–38

Saini N, Yashpal Koramutla MK, Singh N, Singh S, Singh R, Yadav S, Bhattacharya R, Vasudev S, Yadava DK (2019) Promoter polymorphism in FAE1.1 and FAE1.2 genes associated with erucic acid content in *Brassica juncea*. Mol Breed **39**: 75

Salamov AA, Solovyev VV (2000) *Ab initio* gene finding in *Drosophila* genomic DNA. Genome Res **10**: 516–522

Samans B, Chalhoub B, Snowdon RJ (2017) Surviving a genome collision: genomic signatures of allopolyploidization in the recent crop species *Brassica napus*. Plant Genome **10**: https://doi.org/10.3835/plantgenome2017.02.0013

Sanderson MJ (2003) r8s: inferring absolute rates of molecular evolution and divergence times in the absence of a molecular clock. Bioinformatics **19**: 301–302

Sandhu JS, Webster CI, Gray JC (1998) A/T-rich sequences act as quantitative enhancers of gene expression in transgenic tobacco and potato plants. Plant Mol Biol **37**: 885–896

Schläpfer P, Zhang P, Wang C, Kim T, Banf M, Chae L, Dreher K, Chavali AK, Nilo-Poyanco R, Bernard T, et al. (2017) Genome-wide prediction of metabolic enzymes, pathways, and gene clusters in plants. Plant Physiol **173**: 2041–2059

Schnable JC, Freeling M (2011) Genes identified by visible mutant phenotypes show increased bias toward one of two subgenomes of maize. PLoS One **6**: e17855

Schnable JC, Springer NM, Freeling M (2011) Differentiation of the maize subgenomes by genome dominance and both ancient and ongoing gene loss. Proc Natl Acad Sci **108**: 4069–4074

Searchinger TD, Wirsenius S, Beringer T, Dumas P (2018) Assessing the efficiency of changes in land use for mitigating climate change. Nature **564**: 249–253

Servant N, Varoquaux N, Lajoie BR, Viara E, Chen CJ, Vert JP, Heard E, Dekker J, Barillot E (2015) HiC-Pro: an optimized and flexible pipeline for Hi-C data processing. Genome Biol **16**: 259

Sheldon CC, Conn AB, Dennis ES, Peacock WJ (2002) Different regulatory regions are required for the vernalization-induced repression of flowering locus C and for the epigenetic maintenance of repression. Plant Cell **14**: 2527–2537

Shi L, Chen H, Jiang M, Wang L, Wu X, Huang L, Liu C (2019) CPGAVAS2, an integrated plastome sequence annotator and analyzer. Nucleic Acids Res **47**: W65–W73

Simão FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM (2015) BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. Bioinformatics **31**: 3210–3212

Slotte T, Hazzouri KM, Ågren JA, Koenig D, Maumus F, Guo Y-L, Steige K, Platts AE, Escobar JS, Newman LK, et al. (2013) The *Capsella rubella* genome and the genomic consequences of rapid mating system evolution. Nat Genet **45**: 831–835

Song JM, Guan Z, Hu J, Guo C, Yang Z, Wang S, Liu D, Wang B, Lu S, Zhou R, et al. (2020) Eight high-quality genomes reveal pan-genome architecture and ecotype differentiation of *Brassica napus*. Nat Plants **6**: 34–45

Song K, Osborn TC (1992) Polyphyletic origins of *Brassica napus*: new evidence based on organelle and nuclear RFLP analyses. Genome **35**: 992–1001

Song X, Wei Y, Xiao D, Gong K, Sun P, Ren Y, Yuan J, Wu T, Yang Q, Li X, et al. (2021) *Brassica carinata* genome characterization clarifies U's triangle model of evolution and polyploidy in *Brassica*. Plant Physiol.

Stamatakis A (2014) RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics **30**: 1312–1313

Stanke M, Keller O, Gunduz I, Hayes A, Waack S, Morgenstern B (2006) AUGUSTUS: *ab initio* prediction of alternative transcripts. Nucleic Acids Res **34**: W435–W439

Stein A, Coriton O, Rousseau-Gueutin M, Samans B, Schiessl SV, Obermeier C, Parkin IAP, Chèvre AM, Snowdon RJ (2017) Mapping of homoeologous chromosome exchanges influencing quantitative trait variation in *Brassica napus*. Plant Biotechnol J **15**: 1478–1489

Sun F, Fan G, Hu Q, Zhou Y, Guan M, Tong C, Jiana L, Du D, Qi C, Jiang L, et al. (2017) The high-quality genome of *Brassica napus* cultivar 'ZS11' reveals the introgression history in semi-winter morphotype. Plant J **92**: 452–468

Suyama M, Torrents D, Bork P (2006) PAL2NAL: robust conversion of protein sequence alignments into the corresponding codon alignments. Nucleic Acids Res **34**: W609–W612

Tang H, Bowers JE, Wang X, Paterson AH (2010) Angiosperm genome comparisons reveal early polyploidy in the monocot lineage. Proc Natl Acad Sci USA **107**: 472–477

Tang H, Bowers JE, Wang X, Ming R, Alam M, Paterson AH (2008) Synteny and collinearity in plant genomes. Science **320**: 486–488

Taylor DC, Falk KC, Palmer CD, Hammerlindl J, Babic V, Mietkiewska E, Jadhav A, Marillia E-F, Francis T, Hoffman T, et al. (2010) *Brassica carinata* – a new molecular farming platform for delivering bio-industrial oil feedstocks: case studies of genetic modifications to improve very long-chain fatty acid and oil content in seeds. Biofuels Bioprod Biorefining **4**: 538–561

Thomas BC, Pedersen B, Freeling M (2006) Following tetraploidy in an *Arabidopsis* ancestor, genes were removed preferentially from one homeolog leaving clusters enriched in dose-sensitive genes. Genome Res **16**: 934–946

Tirnaz S, Batley J (2019) DNA methylation: toward crop disease resistance improvement. Trend Plant Sci **24**: 1137–1150

UniProt Consortium (2014) UniProt: a hub for protein information. Nucleic Acids Res **43**: D204–D212

Veeckman E, Ruttink T, Vandepoele K (2016) Are we there yet? Reliably estimating the completeness of plant genome sequences. Plant Cell **28**: 1759–1768

Walker BJ, Abeel T, Shea T, Priest M, Abouelliel A, Sakthikumar S, Cuomo CA, Zeng Q, Wortman J, Young SK, et al. (2014) Pilon: an integrated iool for comprehensive microbial variant detection and genome assembly improvement. PLoS One **9**: e112963

Waminal NE, Pellerin RJ, Kang SH, Kim HH (2021) Chromosomal mapping of tandem repeats revealed massive chromosomal rearrangements and insights into *Senna tora* dysploidy. Front Plant Sci **12**: 629898

Waminal NE, Perumal S, Liu S, Chalhoub B, Kim HH, Yang TJ (2018) Quantity, distribution, and evolution of major repeats in *Brassica napus*. *In* S Liu, R Snowdon, B Chalhoub, eds, The Brassica napus Genome, Compendium of Plant Genomes. Springer International Publishing, Cham, Switzerland, pp 111–129.

Wang W, Guan R, Liu X, Zhang H, Song B, Xu Q, Fan G, Chen W, Wu X, Liu X, et al. (2019) Chromosome level comparative analysis of *Brassica* genomes. Plant Mol Biol **99**: 237–249

Wang Y, Jin S, Xu Y, Li S, Zhang S, Yuan Z, Li J, Ni Y (2020) Overexpression of BnKCS1-1, BnKCS1-2, and BnCER1-2 promotes cuticular wax production and increases drought tolerance in *Brassica napus*. Crop J **8**: 26–37

Wang Y, Tang H, DeBarry JD, Tan X, Li J, Wang X, Lee T, Jin H, Marler B, Guo H, et al. (2012) MCScanX: a toolkit for detection and evolutionary analysis of gene synteny and collinearity. Nucleic Acids Res **40**: e49

Wang Z, Zhao G, Yang Q, Gao L, Liu C, Ru Z, Wang D, Jia J, Cui D (2022) Helitron and CACTA DNA transposons actively reshape the common wheat - AK58 genome. Genomics **114**: 110288

Wei Z, Wang M, Chang S, Wu C, Liu P, Meng J, Zou J (2016) Introgressing subgenome components from *Brassica rapa* and *B. carinata* to *B. juncea* for broadening Its genetic base and exploring intersubgenomic heterosis. Front Plant Sci **7**: 1677

Willing EM, Rawat V, Mandáková T, Maumus F, James GV, Nordström KJV, 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**: 1–7

Wolfe KH (2001) Yesterday's polyploids and the mystery of diploidization. Nat Rev Genet **2**: 333–341

Wolff J, Rabbani L, Gilsbach R, Richard G, Manke T, Backofen R, Grüning BA (2020) Galaxy HiCExplorer 3: a web server for reproducible Hi-C, capture Hi-C and single-cell Hi-C data analysis, quality control and visualization. Nucleic Acids Res **48**: W177–W184

Wolfsberg TG, Schafer S, Tatusov RL (2001) Organelle genome resource at NCBI. Trends Biochem Sci **26**: 199–203

Wright F (1990) The 'effective number of codons' used in a gene. Gene **87**: 23–29

Wu HJ, Zhang Z, Wang J-Y, Oh D-H, Dassanayake M, Liu B, Huang Q, Sun H-X, Xia R, Wu Y, et al. (2012) Insights into salt tolerance from the genome of *Thellungiella salsuginea*. Proc Natl Acad Sci USA **109**: 12219–12224

Wu J, Lin L, Xu M, Chen P, Liu D, Sun Q, Ran L, Wang Y (2018) Homoeolog expression bias and expression level dominance in resynthesized allopolyploid *Brassica napus*. BMC Genomics **19**: 586

Wu X, Qi X (2010) Genes encoding hub and bottleneck enzymes of the *Arabidopsis* metabolic network preferentially retain homeologs through whole genome duplication. BMC Evol Biol **10**: 145

Xu GC, Xu TJ, Zhu R, Zhang Y, Li SQ, Wang HW, Li JT (2019) LR_Gapcloser: a tiling path-based gap closer that uses long reads to complete genome assembly. Gigascience **8**: giy157

Xue JY, Wang Y, Chen M, Dong S, Shao ZQ, Liu Y (2020) Maternal Inheritance of U's Triangle and evolutionary process of *Brassica* mitochondrial genomes. Front Plant Sci **11**: 805

Yan G, Li D, Cai M, Gao G, Chen B, Xu K, Li J, Li F, Wang N, Qiao J, et al. (2015) Characterization of *FAE1* in the zero erucic acid germplasm of *Brassica rapa* L. Breed Sci **65**: 257–264

Yang J, Liu D, Wang X, Ji C, Cheng F, Liu B, Hu Z, Chen S, Pental D, Ju Y, et al. (2016) The genome sequence of allopolyploid *Brassica juncea* and analysis of differential homoeolog gene expression influencing selection. Nat Genetics **48**: 1225–1232

Yang R, Jarvis DJ, Chen H, Beilstein M, Grimwood J, Jenkins J, Shu S, Prochnik S, Xin M, Ma C, et al. (2013) The reference genome of the halophytic plant *Eutrema salsugineum*. Front Plant Sci **4**: 46

Yang Z (1997) PAML: a program package for phylogenetic analysis by maximum likelihood. Comput Appl Biosci **13**: 555–556

Yim WC, Cushman JC (2017) Divide and Conquer (DC) BLAST: fast and easy BLAST execution within HPC environments. PeerJ **5**: e3486

Zeng F, Cheng B (2014) Transposable element insertion and epigenetic modification cause the multiallelic variation in the expression of *FAE1* in *Sinapis alba*. Plant Cell **26**: 2648–2659

Zhang L, Cai X, Wu J, Liu M, Grob S, Cheng F, Liang J, Cai C, Liu Z, Liu B, et al. (2018) Improved *Brassica rapa* reference genome by single-molecule sequencing and chromosome conformation capture technologies. Hortic Res **5**: 1–11

Zhang L, Vision TJ, Gaut BS (2002) Patterns of nucleotide substitution among simultaneously duplicated gene pairs in *Arabidopsis thaliana*. Mol Biol Evol **19**: 1464–1473

Zhang W, Hu D, Raman R, Guo S, Wei Z, Shen X, Meng J, Raman H, Zou J (2017) Investigation of the genetic diversity and quantitative trait loci accounting for important agronomic and seed quality traits in *Brassica carinata*. Front Plant Sci **8**: 615

Zhang X, Zhang S, Zhao Q, Ming R, Tang H (2019) Assembly of allele-aware, chromosomal-scale autopolyploid genomes based on Hi-C data. Nat Plants **5**: 833–845

Zhang Y, Yu Z, Zheng C, Sankoff D (2021) Integrated synteny- and similarity-based inference on the polyploidization–fractionation cycle. Interface Foc **11**: 20200059

Zhukov AV, Shumskaya M (2020) Very-long-chain fatty acids (VLCFAs) in plant response to stress. Functional Plant Biol **47**: 695–703

Zoong Lwe Z, Sah S, Persaud L, Li J, Gao W, Raja Reddy K, Narayanan S (2021) Alterations in the leaf lipidome of *Brassica carinata* under high-temperature stress. BMC Plant Biol **21**: 404