3-13-2023

Genome assembly of the Brassicaceae diploid *Orychophragmus violaceus* reveals complex whole-genome duplication and evolution of dihydroxy fatty acid metabolism

Fan Huang

Peng Chen

Xinyu Tang

Ting Zhong

Taihua Yang

See next page for additional authors

Authors

Fan Huang, Peng Chen, Xinyu Tang, Ting Zhong, Taihua Yang, Chinedu Charles Nwafor, Chao Yang, Xianhong Ge, Hong An, Zaiyun Li, Edgar B. Cahoon, and Chunyu Zhang

Plant Communications
Research article

CellPress
Partner Journal

Genome assembly of the Brassicaceae diploid *Orychophragmus violaceus* reveals complex whole-genome duplication and evolution of dihydroxy fatty acid metabolism

Fan Huang[1,5], Peng Chen[1,5], Xinyu Tang[1], Ting Zhong[1], Taihua Yang[1], Chinedu Charles Nwafor[1], Chao Yang[1], Xianhong Ge[1], Hong An[4], Zaiyun Li[1], Edgar B. Cahoon[2,3,*] and Chunyu Zhang[1,*]

[1]National Key Lab of Crop Genetic Improvement and College of Plant Science and Technology, Huazhong Agricultural University, Wuhan, China

[2]Center for Plant Science Innovation, University of Nebraska-Lincoln, Lincoln, NE, USA

[3]Department of Biochemistry, University of Nebraska-Lincoln, Lincoln, NE, USA

[4]Bioinformatics and Analytics Core, University of Missouri-Columbia, Columbia, MO, USA

[5]These authors contributed equally

***Correspondence: Edgar B. Cahoon (ecahoon2@unl.edu), Chunyu Zhang (zhchy@mail.hzau.edu.cn)**

https://doi.org/10.1016/j.xplc.2022.100432

Orychophragmus violaceus **is a Brassicaceae species widely cultivated in China, particularly as a winter cover crop in northern China because of its low-temperature tolerance and low water demand. Recently, *O. violaceus* has also been cultivated as a potential industrial oilseed crop because of its abundant 24-carbon dihydroxy fatty acids (diOH-FAs), which contribute to superior high-temperature lubricant properties. In this study, we performed *de novo* assembly of the *O. violaceus* genome. Whole-genome synteny analysis of the genomes of its relatives demonstrated that *O. violaceus* is a diploid that has undergone an extra whole-genome duplication (WGD) after the Brassicaceae-specific α-WGD event, with a basic chromosome number of x = 12. Formation of diOH-FAs is hypothesized to have occurred after the WGD event. Based on the genome and the transcriptome data from multiple stages of seed development, we predicted that *OvDGAT1-1* and *OvDGAT1-2* are candidate genes for the regulation of diOH-FA storage in *O. violaceus* seeds. These results may greatly facilitate the development of heat-tolerant and eco-friendly plant-based lubricants using *O. violaceus* seed oil and improve our understanding of the genomic evolution of Brassicaceae.**

Keywords: *Orychophragmus violaceus*, genome evolution, dihydroxy fatty acids, polyestolides, lubricant oil, oilseed, Brassicaceae

Huang F., Chen P., Tang X., Zhong T., Yang T., Nwafor C.C., Yang C., Ge X., An H., Li Z., Cahoon E.B., and Zhang C. (2023). Genome assembly of the Brassicaceae diploid *Orychophragmus violaceus* reveals complex whole-genome duplication and evolution of dihydroxy fatty acid metabolism. Plant Comm. **4**, 100432.

INTRODUCTION

Orychophragmus violaceus is an ornamental Brassicaceae species with small purple flowers that bloom in the early spring. It has the common name of Chinese violet cress (Zhou et al., 1987). This plant typically grows in the wild in East Asia, particularly Korea and northern China, where it is known as "er-yue-lan" (Zhang and Dai, 2005). It is also cultivated as a leafy vegetable ("zhuge") in China and as a cover crop in northern China because of its low-temperature tolerance and high water use efficiency (Liu et al., 2012; Wen et al., 2020). *O. violaceus* has emerged as a potential industrial oilseed crop because its seed oil contains abundant dihydroxy fatty acids (diOH-FAs) (nebraskanic acid, 7,18-OH-24:1Δ^{15}; wuhanic acid, 7,18-OH-24:2$\Delta^{15,21}$) in the form of polyestolides, which can contribute to superior lubrication properties (Li et al., 2018). These fatty acids and their special storage form (triacylglycerol [TAG] polyestolides) are unique in the plant kingdom and give *O. violaceus* oil even better high-temperature lubricant properties than castor oil, a valuable plant-based lubricant

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

(Romsdahl et al., 2019). However, the specific metabolism of C24 diOH-FAs (fatty acid chains with 24 carbons) and TAG polyestolides in *O. violaceus* is largely unknown.

Our previous study demonstrated that fatty acid desaturase 2 (FAD2) and fatty acid elongase 1 (FAE1) in *O. violaceus* have developed specific enzymatic activities critical for diOH-FA biosynthesis (Li et al., 2018). Rather than catalyzing fatty acid desaturation, the enzyme encoded by *OvFAD2-2* functions as a fatty acid hydroxylase to generate the terminal hydroxyl group of nebraskanic and wuhanic acids (Li et al., 2018). The functional variants encoded by *OvFAE1-1* produce the carboxyl-terminal hydroxyl group of diOH-FA through a ''discontinuous elongation'' process (Li et al., 2018). The enzymatic origin of TAG polyestolides, which account for nearly all fatty acid storage in *O. violaceus* seeds, remains unclear. These molecules consist of high-molecular-weight TAG species containing a diOH-FA at the *sn*-3 or *sn*-1 position and an additional diOH-FA linked to the 18-OH of esterified nebraskanic and wuhanic acids (Romsdahl et al., 2019). It is presumed that polyestolides are formed by an acyltransferase, such as diacylglycerol acyltransferase (DGAT), that is assumed to have novel activity. Genomics and seed transcriptomics data for *O. violaceus* are expected to enhance our understanding of the evolution and the ''missing'' steps in the pathways of diOH-FA-containing polyestolide biosynthesis in Brassicaceae.

Previous studies have reported the unusual chromosome pairing behavior during meiosis in *O. violaceus* (Li and Liu, 1995; Li et al., 1996; Yin et al., 2020). Cytological observation has clearly revealed that *O. violaceus* has a total of 24 chromosomes (Li et al., 1996). However, it remains controversial whether the 24 chromosomes are derived from a tetraploid with a basic chromosome number of six or an octoploid with a basic chromosome number of three (Li et al., 1996; Lysak et al., 2007; Yin et al., 2020). The phylogenetic position of *O. violaceus* remains disputable, and the main concern involves the relationship of the *Orychophragmus* genus to other branches such as the *Conringia* genus and the rest of the genera in different Cruciferae lineages (Lysak et al., 2007; Zhou et al., 2009; Liu et al., 2011; Hu et al., 2016; Mandáková et al., 2017; Guo et al., 2021; Huang et al., 2020). A previous study using mitochondrial *NAD7* revealed that the *Orychophragmus* genus is a branch in Brassicaceae that is paralleled by Isatideae, Sisymbrieae, Iberideae, Arabideae, Calepineae, Thalspideae, Alysseae, and *Eutrema* to form lineage II of Cruciferae (Couvreur et al., 2010). A comprehensive plastome-based genus-level phylogenetic study of a collection of Brassicaceae species updated the disparity among evolutionary lineages and suggested defined terms for genera and tribes that were improperly assigned previously (Walden et al., 2020). Apparently, dissection of genome information can greatly help to resolve these controversies.

In this study, we performed *de novo* assembly of the *O. violaceus* genome at the chromosome level. Based on the high-quality genome, we confirmed that the basic chromosome number of *O. violaceus* is x = 12. An analysis of its genome structural characteristics indicated that *O. violaceus* did not undergo the Brassicaceae whole-genome triplication (WGT) (Lysak et al., 2007) but experienced a unique whole-genome duplication (WGD) event, consistent with the results of previous studies (Lysak et al., 2007; Franzke et al., 2011). The present *O. violaceus* genome can be considered a diploid with an evolutionary track of polyploidy. Some ancient chromosomes have been well retained, whereas other chromosomes have experienced fragmentation and rearrangement, thus explaining the formation of the multivalent configuration in regenerated haploids from pollen mother cells of *O. violaceus* (Yin et al., 2020). This genomic information, together with a transcriptome of the developing seed, also provides important information about the evolution of genes associated with diOH-FA biosynthesis, including *OvFAE1-1*. Our study also reveals that variation in the transcript structure of diacylgylcerol acyltransferase 1 (DGAT1) may be associated with TAG polyestolide biosynthesis. The findings of the present study may improve our understanding of Brassicaceae evolution and variant fatty acid and TAG biosynthesis and contribute to the genetic improvement of *O. violaceus* as a new and high-value industrial oilseed crop.

RESULTS

De novo assembly of the *O. violaceus* genome

Based on the NovaSeq 6000 and PacBio Sequel II platforms, approximately 180 Gb of Illumina short reads and 48 Gb of PacBio circular consensus sequencing (CCS) long reads were obtained. Genome analysis according to the distribution of K-mers (Marçais and Kingsford, 2011) revealed that the genome of *O. violaceus* is highly complicated and the estimated size is approximately 1.27 Gb, including 74.89% repetitive regions and 1.45% genome heterozygosity (Supplemental Figure 1; Table 1). Hifiasm software was used for initial assembly of the *O. violaceus* genome (Cheng et al., 2021). The draft genome contained 4328 contigs with an N50 (read length metric) value of 1.96 Mb and a total length of 1.87 Gb. After filtering of the redundant contigs with Purge Haplotigs (Roach et al., 2018), 3D DNA pipelines (Durand et al., 2016; Dudchenko et al., 2017) were used to scaffold the genome and filter the heterozygous region by integrating approximately 150 Gb of Hi-C (whole genome chromosome conformation capture) data (Figure 1B; Supplemental Table 1). Finally, a chromosome-scale genome with a total length of 1.25 Gb was obtained that contained 12 pseudochromosomes and 6 Mb of unplaced scaffold. The final genome contained 97.8% complete Benchmarking Universal Single-Copy Orthologs (BUSCO) genes, and the long terminal repeat (LTR) assembly index (LAI) value was 36.15. Based on the final assembly, a total of 61 097 protein-coding genes were annotated (Table 1).

In the chromosome synteny analysis, two pairs of chromosomes (Chr01–Chr02 and Chr09–Chr10) exhibited a similar chromosome structure, suggesting that *O. violaceus* was possibly derived from a tetraploid (Figure 1C).

Genome collinearity and evolution analysis of the *O. violaceus* genome

To verify this conclusion, we performed a genome synteny analysis between *O. violaceus*, *Arabidopsis thaliana*, and *Brassica napus* because these species can represent different nodes on the evolutionary route (The Arabidopsis Genome Inititiative, 2000; Song et al., 2020), particularly *B. napus*, which has experienced the WGT of *Brassica* (Figure 2). As shown in Figure 2A, some *A. thaliana* chromosomal fragments showed collinearity with

Assembly feature	Statistics
Assembled genome size	1.25 Gb
Estimated genome size	1.27 Gb
Estimated genome heterozygosity	1.45%
Contig N50	1.96 Mb
BUSCO coverage	97.8%
LAI	36.15
Chromosome number	12
Assembled % of genome	99.50%
Repeat region % of assembly	74.89%
GC content	39.10%
Number of protein-coding genes	61 097
Average gene length	2097.4 bp

Table 1. Assembly features of the *O. violaceus* genome.

two fragments in *O. violaceus* and six fragments in *B. napus*, indicating the tetraploid nature of the *O. violaceus* genome (Figure 2A, orange, green, and blue labels). We then compared the 12 pseudochromosomes of *O. violaceus* with the seven chromosomes of *Isatis indigotica*, whose genome has an intact ancestral translocation proto-Calepineae karyotype (tPCK) (Kang et al., 2020). The linear order and continuity of homologous genes on the OvChr01 and OvChr02 chromosome pair were highly consistent with those on *I. indigotica* Chr01. Similarly, OvChr09 and OvChr10 also aligned well with *I. indigotica* Chr06 (Figures 2C and 2D). Except for chr03, which showed high collinearity with *I. indigotica* Chr02 (Figure 2E, orange label), the other *O. violaceus* pseudochromosomes showed interlaced collinearity with *I. indigotica* chromosomes. For several *I. indigotica* chromosomal fragments, one fragment might share collinearity with one or two chromosomal fragments in *O. violaceus* (Figure 2E, orange and mazarine). It could therefore be inferred that *O. violaceus* possibly evolved through a tetraploid, consistent with the hypothesis proposed by Lysak et al. (2007). Our data demonstrated that, through chromosomal fragmentation and rearrangement, *O. violaceus* has evolved from a tetraploid into a diploid with a basic chromosome number of n = 12 (2n = 24).

To confirm this evolutionary path, 818 single-copy gene families were used to infer the phylogenetic position and divergence time of *O. violaceus* and 12 other Brassicaceae species (Figure 3A). Although *O. violaceus* was clustered between *I. indigotica* and the Brassicaceae species, it diverged from the Brassicaceae branch around 18.0 million years ago (mya), before the *Brassica* WGT event (Figure 2F), and underwent a separate WGD at around 6.22 mya (Supplemental Figure 2). Because *O. violaceus* has undergone one more WGD than *I. indigotica*, its predicted genome size would be around 600 Mb, which is about twice that of *I. indigotica* (around 284 Mb, Kang et al., 2020). However, the total genome size of *O. violaceus* was 1.25 Gb, with an average chromosome size of around 100 Mb, which is much larger than that of most cruciferous species (Supplemental Table 2; Shan et al., 2021). This may be due to the burst of LTR retrotransposons in the *O. violaceus* genome (Rensing et al., 2008; Nystedt et al., 2013), as about 55.18% of the genome sequence was annotated as LTR retrotransposons (Supplemental Figure 3).

Karyotype analysis of *O. violaceus* pseudochromosomes

To better study the evolution of chromosome structure in Brassicaceae, a model containing 24 genomic blocks (GBs; named from A to X) was simulated using comparative chromosome painting (CCP) (Supplemental Figure 4A; Lysak et al., 2007, 2016; Schranz et al., 2006; Schranz et al., 2007; Mandáková and Lysak, 2016). Based on CCP, the karyotype evolution in eight species with x = 7 (2n = 14, 28) chromosomes from six Brassicaceae tribes (Calepineae, Conringieae, Noccaeeae, Eutremeae, Isatideae, and Sisymbrieae) was reconstructed with an ancestral PCK (n = 7). Among them, Eutremeae, Isatideae, and Sisymbrieae showed an additional translocation between the second and seventh chromosomes (tPCK, n = 7) (Supplemental Figure 4A; Mandáková and Lysak, 2008; Lysak et al., 2016).

To decipher the karyotype of *O. violaceus*, we performed a whole-genome collinearity comparison between *O. violaceus* and *A. thaliana* and determined the order and orientation of the 24 ancestral GBs of *A. thaliana* along the *O. violaceus* pseudochromosomes (Figure 3B). Among the 12 *O. violaceus* pseudochromosomes, five displayed a karyotype similar to the ancestral karyotype: Chr01 and Chr02 were similar to tPCK1, Chr03 was similar to tPCK2, and Chr09 and Chr10 were similar to tPCK6 (Figure 3B, red stars). Translocation had occurred in the rest of the pseudochromosomes (Figure 3B). To explore how the chromosomes are rearranged and reduced, we propose a brief model (Supplemental Figure 5). The whole model consists of three progressive processes (Supplemental Figure 5): 1) GBs in OvtPCK3, OvtPCK2, OvtPCK4, and OvtPCK5 were rearranged to form new chromosomes OvNew1–OvNew6; 2) GBs in OvtPCK7 were rearranged with OvNew02 and OvNew04 to form OvNew07–OvNew10; and 3) OvNew10 was rearranged with OvtPCK4 to form OvNew12, and OvNew08 was rearranged with OvNew11 to form OvNew13 (Supplemental Figure 5).

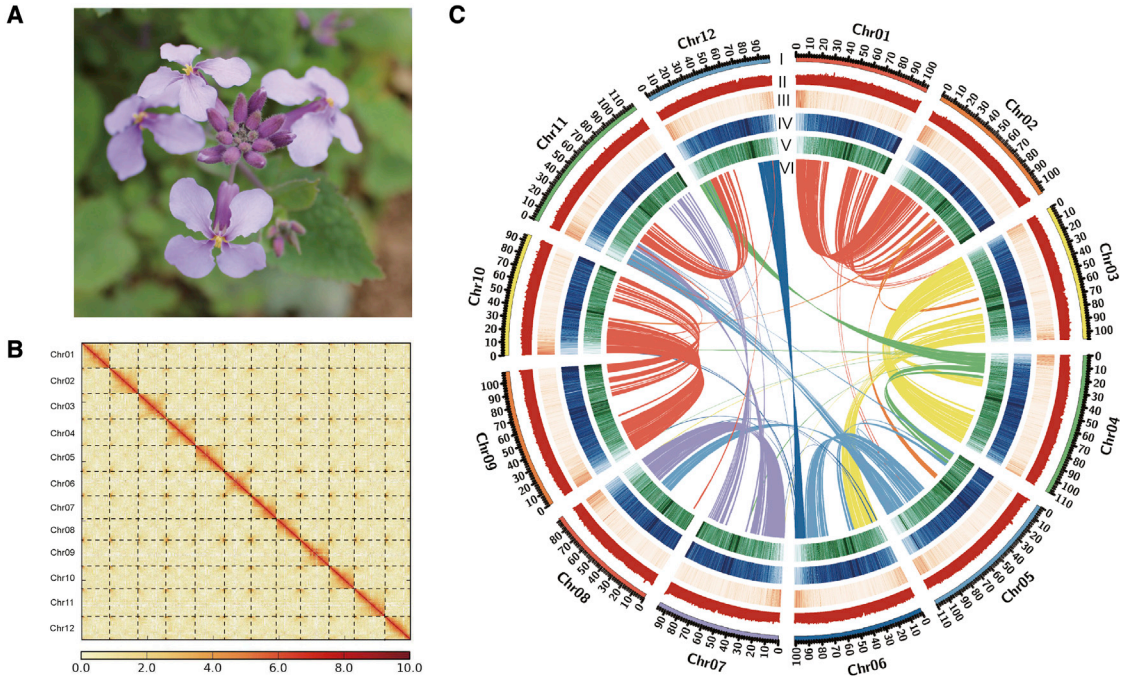


Figure 1. Hi-C interaction heatmap and genome features of *O. violaceus*.
(A) The flower of *O. violaceus*.
(B) Hi-C interaction heatmap of *O. violaceus*.
(C) Genome features of *O. violaceus*. I, chromosomes; II, GC content; III, gene density; IV, repeat sequences; V, LTRs; VI, intra-genomic synteny within *O. violaceus*.

Previous cytological observations suggested that *O. violaceus* chromosomes may have undergone rearrangement (Li et al., 1996; Lysak et al., 2007; Yin et al., 2020). This hypothesis is supported by our genome assembly and collinearity analysis (Figures 1C and 2A). Our genome assembly and synteny relationships may also explain the formation of multivalents observed during meiosis (Supplemental Figure 4B). Because of inter-chromosomal fragmental homology, the non-homologous chromosomes of *O. violaceus* can pair with each other, thus explaining the circular chromosome structure during meiosis (Li and Liu, 1995; Li et al., 1996; Yin et al., 2020).

Transcriptomic analysis of developing *O. violaceus* seeds

To identify genes involved in diOH-FA biosynthetic pathways, we collected *O. violaceus* seeds over a time course from 22–44 days after flowering (DAF) to obtain fatty acid and transcriptomic information (Supplemental Table 3). At the early stages of seed development, no diOH-FA was detected until 32 DAF (Figure 4B and Supplemental Figure 6). Apparent diOH-FA began to be observed from 32 DAF, and there was a more significant increase in the content of C24:2-diOH than of C24:1-diOH at 40 DAF (Figures 4B and Supplemental Figure 6). At 44 DAF, the total diOH-FA content accounted for 40.55% of the total fatty acids (Figure 4B and Supplemental Figure 6).

Developing *O. violaceus* seeds at 22, 26, 32, and 40 DAF were collected for transcriptome analysis, and three biological replicates were obtained for each time point. A total of 6747 differentially expressed genes (DEGs) were identified, which were further divided into 10 modules according to their expression levels at the four developmental stages (Supplemental Figure 7 and Supplemental Figure 8). The two most relevant modules, which are marked in brown and turquoise (Supplemental Figure 9), comprised a total of 2332 genes whose expression was upregulated at 32 DAF (brown, Supplemental Figure 9) or 40 DAF (turquoise, Supplemental Figure 9).

To determine whether there are unique genes for diOH-FA synthesis that are not present in other oil crops, we extracted the transcriptome data of two *B. napus* materials, one with high oil content and the other with low oil content, and compared them with the transcriptome data from *O. violaceus* seeds (Figure 4C). Using the same criteria described above, we identified genes from the two *B. napus materials* whose expression patterns were similar to those in the brown or turquoise modules (Supplemental Figure 10). By overlapping these three datasets, we identified 165 *O. violaceus*-specific gene families and 953 singletons (Figure 4D and Supplemental Figure 11). Although Kyoto Encyclopedia of Genes and Genomes (KEGG) enrichment identified 60 genes associated with lipid metabolism, most of these genes were neither highly expressed at the corresponding seed developmental stages during diOH-FA synthesis nor annotated as being connected with very-long-chain fatty acid metabolism, except for *OvFAE1-1* and *OvFAD2-2* (Supplemental Table 4). Because the candidate genes were expected to be upregulated at the initiation of diOH-FA biosynthesis, these results suggested that, instead of novel genes, some ''known'' enzymes normally involved in fatty acid metabolism may have acquired new functions during evolution to catalyze the biosynthesis of diOH-FAs and polyestolides.

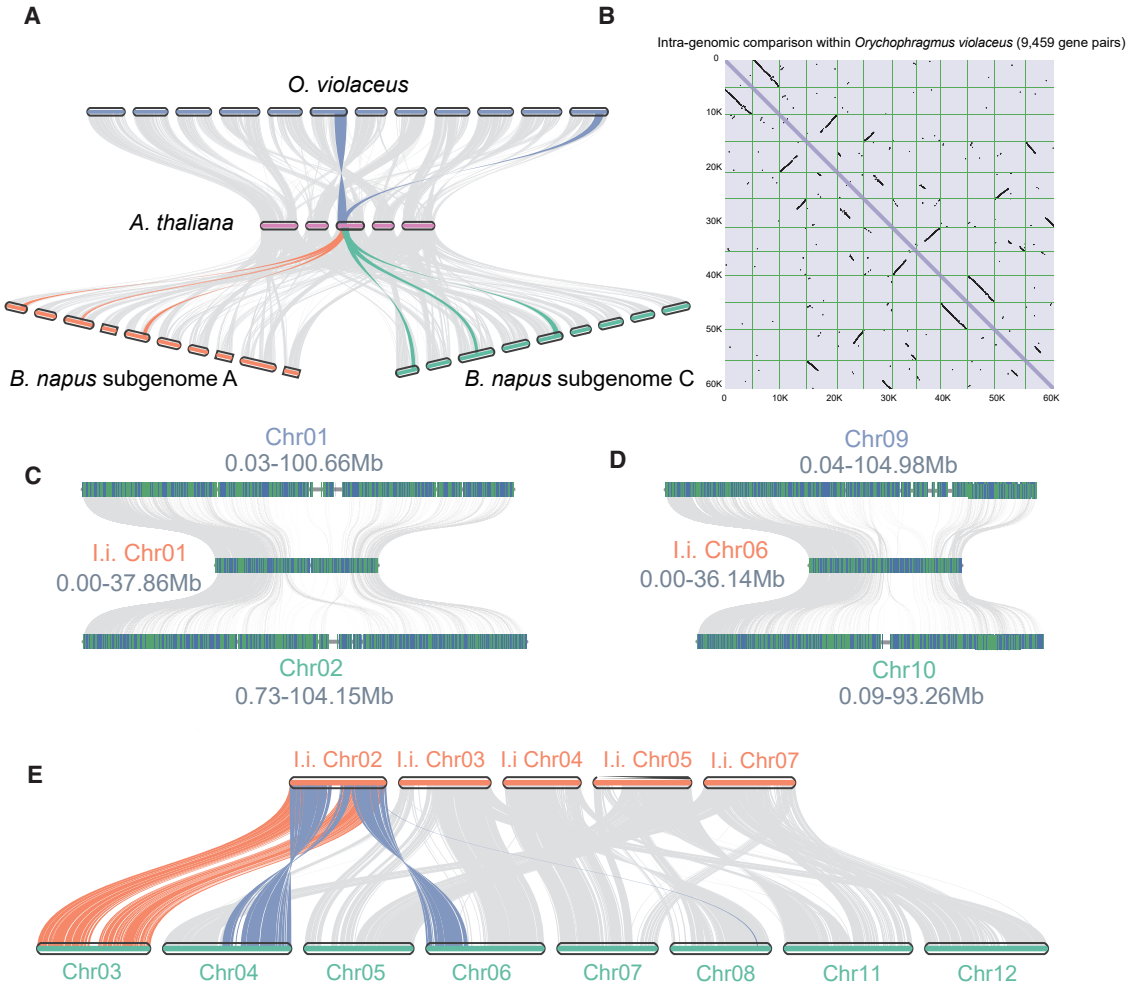


Figure 2. Genome collinearity and phylogenetic analysis.
(A) Genome collinearity of *A. thaliana*, *B. napus*, and *O. violaceus*.
(B) Intra-genomic comparison within *O. violaceus*.
(C) Chromosome comparison between *O. violaceus* Chr01 and Chr02 and *I. indigotica* Chr01.
(D) Chromosome comparison between *O. violaceus* Chr09 and Chr10 and *I. indigotica* Chr06.
(E) Chromosome comparison between the other chromosomes of *O. violaceus* and *I. indigotica* Chr02, Chr03, Chr04, Chr05, and Chr07.

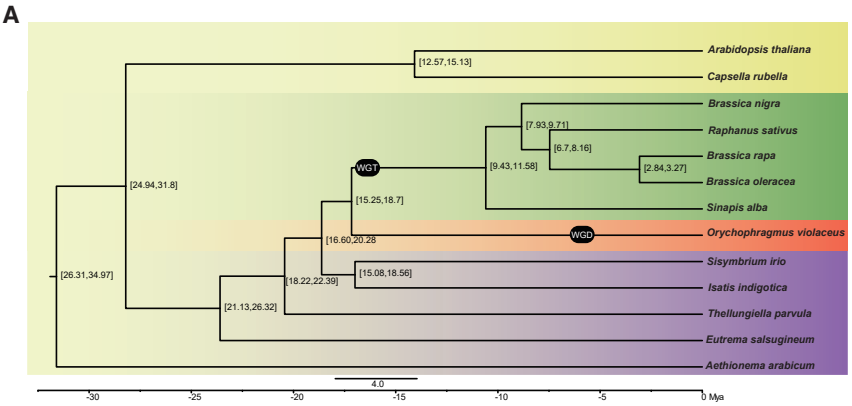
Evolutionary analysis of the *OvFAE1* and *OvFAD2* gene families

Previous studies have suggested that *OvFAD2-2*, *OvFAE1-1*, and *OvFAE1-2* participate in the "discontinuous elongation" pathway of diOH-FAs (Li et al., 2018). Our transcriptome data also associated these three genes with the two modules most relevant to diOH-FA synthesis. A BLAST search revealed the presence of five copies of *OvFAD2* and three copies of *OvFAE1* in the *O. violaceus* genome (Figures 5A and 5B), but only *OvFAD2-1*, *OvFAD2-2*, *OvFAE1-1*, and *OvFAE1-2* were highly expressed (Figure 5C; Supplemental Table 4).

Because *I. indigotica* has not experienced WGT or WGD (Figure 3A), it may retain the tPCK karyotype of the ancestor species of *O. violaceus*. We therefore used *I. indigotica* as a reference and compared gene-to-gene *Ks* values between *FAD2*/*FAE1* orthologs (Supplemental Table 5). The results suggested that the *FAD2* and *FAE1* gene families had not undergone positive selection (Supplemental Table 5). For the *FAD2* genes, there was no significant difference in *Ks* value from *IiFAD2*, although, according to the Ks value, *OvFAD2-5* might be the most distant member (Supplemental Table 5). When we compared the *OvFAE1* orthologs with *IiFAE1*, *OvFAE1-1* and *OvFAE1-2* had much lower *Ks* values than *OvFAE1-3* (Supplemental Table 5), suggesting a two-stage evolution of *FAE1* genes, in which *OvFAE1-1* and *OvFAE1-2* appeared first, followed by *OvFAE1-3*, which may have originated from a local duplication of *OvFAE1-2* (Figures 5B and 5D).

OvDGAT1 as a candidate gene for the biosynthesis of polyestolides

Fatty acid acyltransferases are involved in TAG biosynthesis. Among these enzymes, DGAT1 or DGAT2 catalyzes the addition of the third fatty acid to the glycerol backbone of diacylglycerol (DAG) to form TAG (Li-Beisson et al., 2013). The storage of primary fatty acids in the form of TAG polyestolides in *O. violaceus* suggested additional acyltransferase speciation to esterify fatty acids to the terminal hydroxyl group of TAG-linked



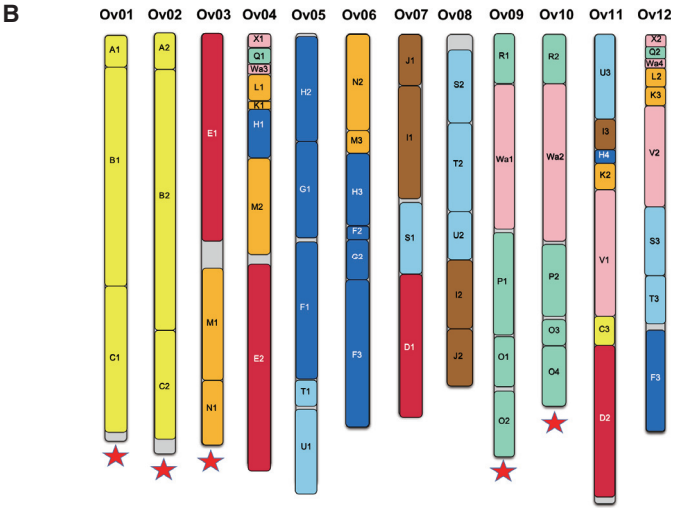


Figure 3. Evolution and chromosome structure analysis of *O. violaceus*.

(A) Phylogenetic analysis of *O. violaceus* and other cruciferous species. The numbers on the tree represent the estimated differentiation times (million years ago [mya]). WGD, whole-genome duplication; WGT, whole-genome triplication.

(B) Karyotype of *O. violaceus* (2n = 24) based on 24 ancestral GBs of *A. thaliana*. Red stars represent chromosomes that retain the ancestral karyotype. Chr01–Chr12 represent the 12 pseudochromosomes of *O. violaceus*.

diOH-FA. We found two *OvDGAT1*-related genes, *OvDGAT1-1* and *OvDGAT1-2*, in the *O. violaceus* seed transcriptome (Supplemental Table 6). Phylogenetic analysis of the *DGAT* gene family suggested that *OvDGAT1-1* and *OvDGAT1-2* were in a unique branch distinct from *DGAT* genes of other species (Supplemental Figure 12A). Through protein sequence alignment of OvDGAT1-1/1-2, AtDGAT1, BnDGAT1, and RcDGAT1, we found that the amino acids in the catalytic center were highly conserved, but some residues in the acyl-coenzyme A (CoA) binding site showed variations in OvDGAT1-1 or OvDGAT1-2 (Figures 4E and 4F and Supplemental Figure 12B; Sui et al., 2020). In particular, OvDGAT1-1 contained an insertion of ~28 amino acids, likely from alternative transcript splicing, that was not present in all known plant DGAT1s (Supplemental Figure 12B). We speculated that these protein sequence and structural variations might be associated with the acquisition of new functions in *OvDGAT1-1* and *OvDGAT1-2* for polyestolide biosynthesis.

DISCUSSION

In this study, we performed *de novo* assembly of the *O. violaceus* genome and characterized its evolutionary position and the key candidate genes responsible for diOH-FA biosynthesis in seeds. The results lay a solid foundation for genetic improvement of *O. violaceus* for use as a high-value industrial oil crop in the future.

O. violaceus is a new oilseed crop whose genome has undergone an *Orychophragmus*-specific WGD event

Previous studies have demonstrated that *O. violaceus* is a close relative of the *Brassica* species (Lysak et al., 2007). Based on the CCP technique, Lysak et al. (2007) proposed that *O. violaceus* has experienced a duplication event rather than a triplication event because one copy of the ancestral GBs from *A. thaliana* corresponds to two GBs in *O. violaceus* (Lysak et al., 2007). In the present study, we provided direct evidence for this WGD event in *O. violaceus*, which is independent of the *Brassica*-specific WGT proposed previously (Lysak et al., 2005; Wang et al., 2011; Cai et al., 2021). The genome information revealed that *O. violaceus* is a diploid (2n = 24) evolving from an ancient tetraploid. The haploid genome of this ancestor contained 14 pseudochromosomes; the karyotype of five was retained, and the remaining nine underwent fragmentation and rearrangement, eventually leading to 12 chromosomes in the *O. violaceus* haploid genome (Supplemental Figure 5). The genome block homology also explains the previous observation of multivalent synapsis conformation during meiosis (Li and Liu, 1995; Yin et al., 2020). Our genome structure comparison suggested that *O. violaceus* may have evolved from one or two very similar parental diploid species close to *I. indigotica*. The present diploid genome is apparently stable because the pollen fertility is normal (data not shown). Considering the wide distribution of *O. violaceus* in mainland China, this species can adapt well to different environmental niches, again demonstrating the stability and plasticity of its genome.

O. violaceus gene neo-functionalization is associated with diOH-FA synthesis

Polyploidization in higher plants is frequently associated with adaptation and diversification (Zuo et al., 2022). The seed oil of *O. violaceus* has unique diOH-FAs, and to date, only several related genes have been identified, some of which are dual-function proteins (Li et al., 2018). Based on genome assembly and comparison, we proposed a hypothesis regarding the evolution of the *OvFAE1* gene family (Figure 5D). By sequence alignment, we found that *OvFAE1-2* was closer to *IiFAE1*,

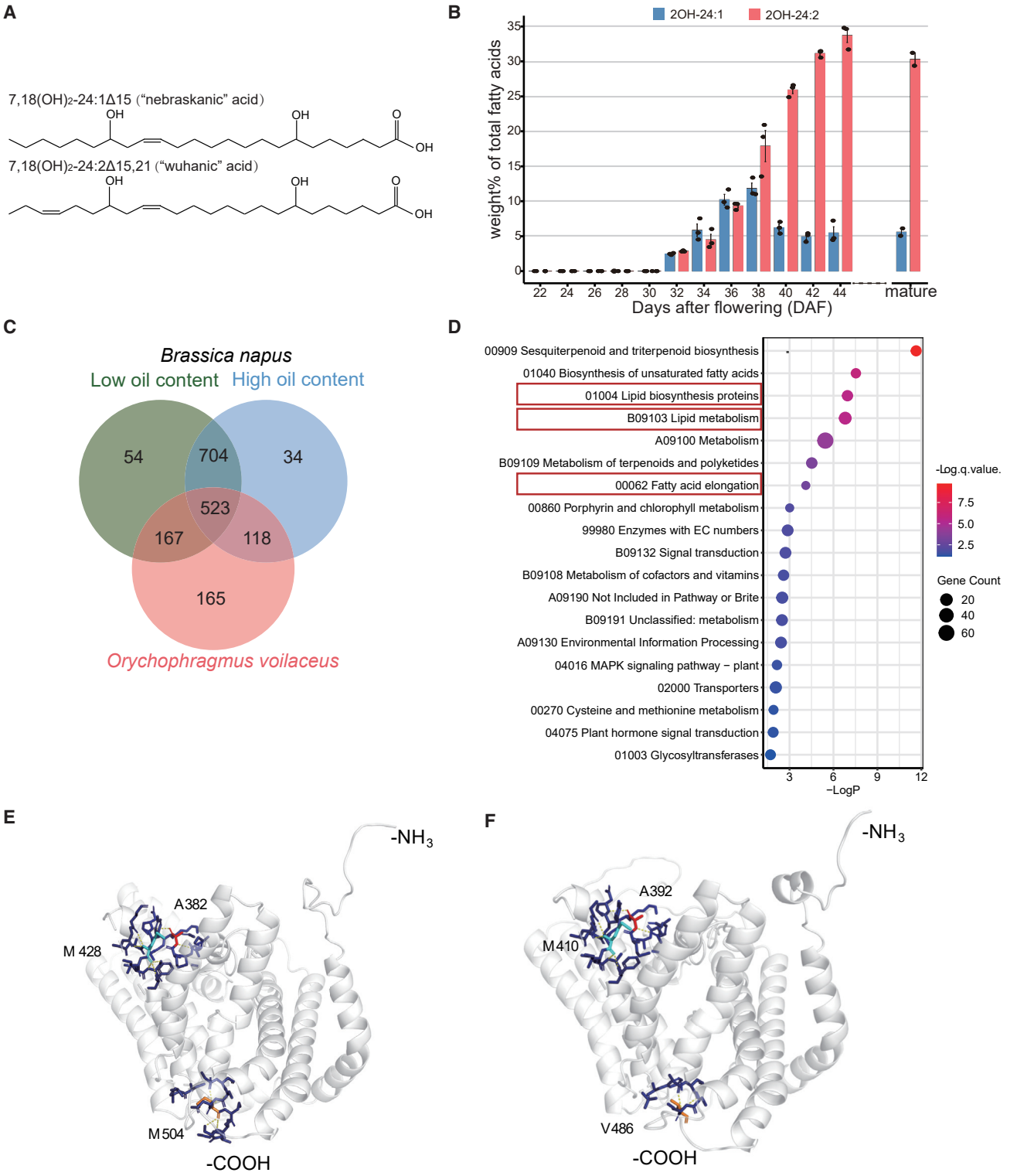


Figure 4. Fatty acid profiles and transcriptome profiling of *O. violaceus* seeds at different developmental stages in comparison with *B. napus* with low and high oil content.

(A) Structure of nebraskanic and wuhanic acids.

(B) Weight percentage of diOH-FAs in developing *O. violaceus* seeds from 22–44 DAF based on three biological replicates. Error bars represent the standard error.

(C) Venn diagram of genes with similar expression profiles in seeds from developing siliques of high-oil-content *B. napus*, low-oil-content *B. napus*, and *O. violaceus*.

(legend continued on next page)

although *OvFAE1-1* has been identified previously to be functional for fatty acid discontinuous elongation (Li et al., 2018). The third gene, *OvFAE1-3*, was far not only from *IiFAE1* but also from the other two *OvFAE1* genes. We therefore hypothesized that it may have come from local fragmental insertion (Figure 5D). These results indicated that a duplicated ancestral pair of *FAE1* genes was generated by WGD, followed by mutation accumulation and diversification. Because only one copy of *FAE1* is present in *I. indigotica*, and no diOH-FA was found in its seed oil, the new function of *OvFAE1-1* for diOH-FA biosynthesis was presumably acquired after the WGD event.

Based on the seed transcriptome and weighted gene co-expression network analysis (WGCNA), we identified two modules highly relevant to diOH-FA synthesis (Supplemental Figure 9). Several genes were identified as candidates, including the previously known *FAD2* and *FAE1* genes and the *DGAT1* gene (Figure 2E). Based on the phylogeny and protein sequence comparison with *B. napus* genes and our previous studies, we can speculate that the biosynthetic genes of diOH-FAs and polyestolides may not be new genes but may instead be derived from structural changes in the domains of proteins normally involved in fatty acid metabolism. Previously, we found that the variant forms of *FAE1* (*OvFAE1-1*) and *FAD2* (*OvFAD2-2*) are required for the biosynthesis of hydroxyl groups in the diOH-FAs (Li et al., 2018). Given the sequence variation and perhaps also alternative splicing variants, OvDGAT1 is a promising candidate for the production of polyestolides, but direct experimental evidence is still lacking. Based on these results, we simulated the pathways of diOH-FA biosynthesis and storage in *O. violaceus* seeds (Figure 6 and Supplemental Figure 15): OvFAD2-2 catalyzes the hydroxylation of oleoyl-phosphatidylcholine (PC) into ricinoleyl-PC; OvLCAT-PLA hydrolyses ricinoleyl-PC and generates ricinoleyl-CoA; and OvFAE1-1/1-2 catalyzes the formation of 7,18-OH-24:1Δ15-CoA through discontinuous chain elongation (Figure 6). After OvGPAT1- and OvLPAT2-catalyzed acylation, the resulting DAG is acylated at the *sn*-3 position by OvDGAT1-1 to form a TAG species with the diOH-FA (Supplemental Figure 15, first compound). The acyl-transferring activity of OvDGAT1 adds one or more diOH-FAs to the hydroxyl group of the TAG-esterified fatty acids to form polyestolides (Supplemental Figure 15).

The reference genome presented here provides not only an important resource for future use of *O. violaceus* as a new industrial oil crop but also a better understanding of the cytological behaviors of chromosomes during meiosis at the genome level. The unique phylogenetic position of *O. violaceus* relative to other Brassicaceae species provides a new perspective for understanding the appearance of diOH-FAs during evolution. Because these special fatty acids are not present in seed oils from *A. thaliana* (Li-Beisson et al., 2013), *Brassica* spp. (Cacciola et al., 2016; Rout et al., 2018; Cartea et al., 2019; Tang et al., 2021), or *I. indigotica* (Supplemental Figure 16), it remains unclear where and when the genes responsible for diOH-FAs arose during evolution. The genomic data reveal that only one copy of the *FAD2* gene is present in *I. indigotica*, but multiple copies are present in the *O. violaceus* genome, indicating the possibility of neo-functionalization by mutation. Finally, we propose a most promising candidate gene, *OvDGAT1-1*, which can contribute to the accumulation of up to 40% of diOH-FAs in *O. violaceus* seed oil to enable the utilization of *O. violaceus* in the plant-based lubricant industry.

METHODS

O. violaceus materials and sample collection

The *O. violaceus* plants were cultivated in the experimental fields at the campus of Huazhong Agricultural University, Wuhan, China. Flowering-stage plants were used for seed collection at different developmental stages. The appearance of the first flower was marked, and DAF were used as time points. Siliques of four different developmental stages (22–44 DAF) were collected at around 8:00 a.m. The seeds were removed from the siliques and immediately frozen in liquid nitrogen until further use.

Seed oil extraction and fatty acid composition analysis

Fatty acids were extracted using 2.5% (w/v) sulfuric acid-methanol and 0.01% (w/v) 2,6-Di-tert-butyl-4-methylphenol (BHT) as described previously (Li et al., 2018). Fatty acids were analyzed by gas chromatography using an HP-INNOWax column (30 m × 0.25 mm, 0.25-μm particle size, Agilent Technologies, USA) and flame ionization detector (Li et al., 2018).

Whole-genome sequencing

Young *O. violaceus* leaves were used to construct the library for sequencing on an Illumina paired-end high-throughput sequencing platform (NovaSeq 6000) with a read length of 150 bp following the standard library building process by Novogene (Cuddapah et al., 2009).

For the construction of PacBio libraries, DNA samples were sheared with a Covaris ultrasonic crusher. Magnetic beads were used to enrich and purify large fragments of DNA. Stem-loop sequencing connectors were then added to both ends of the DNA fragments, and exonucleases were used to remove the fragments that failed to connect. Constructed libraries were sequenced using the PacBio Sequel II platform.

Young leaves were also used to construct the Hi-C library. First, the young leaves were fixed in mass-spectrometry (MS) buffer containing 1% formaldehyde solution. Leaf DNA was then extracted and digested by the *Dpn*II restriction enzyme. An Illumina paired-end sequencing library with a 350-bp insert size was constructed and sequenced with the HiSeq X Ten sequencer.

Transcriptome sequencing

For the transcriptome analysis, RNA sequencing (RNA-seq) was performed by the Beijing Genomics Institute (Shenzhen, China).

(D) KEGG enrichment of 165 gene families clustered only in *O. violaceus*.

(E and F) Predicted 3D structures of OvDGAT1-1 **(E)** and OvDGAT1-2 **(F)**. Critical residues that are potentially important for diOH biosynthesis are shown as stick models. Numbers represent amino acid positions in OvDGAT1-1 or OvDGAT1-2. M428 from OvDGAT1-1 and M410 from OvDGAT1-2 are shown in cyan. A382 (OvDGAT1-1)/A392 (OvDGAT1-2) are shown in red. M504 from OvDGAT1-1 and V486 from OvDGAT1-2 are shown in orange. Blue sticks indicate neighboring residues within 6 Å. Yellow lines represent segment polar contacts to atoms.

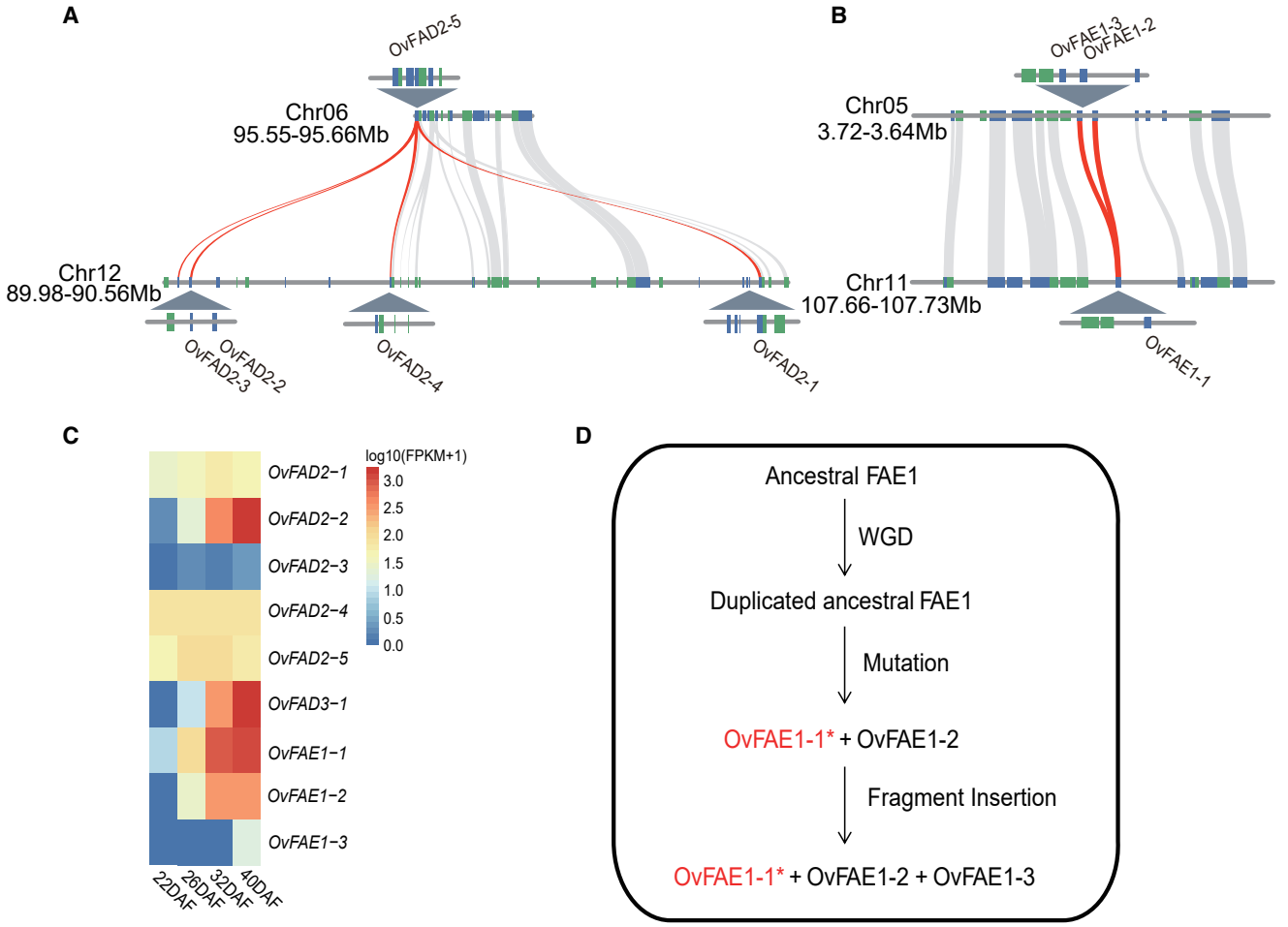


Figure 5. *OvFAD2* and *OvFAE1* gene family analysis.
(A) Collinearity of the *OvFAD2* gene family.
(B) Collinearity of the *OvFAE1* gene family.
(C) Gene expression profile of the *OvFAD2* and *OvFAE1* gene families at different seed developmental stages.
(D) Proposed evolutionary line of OvFAE1. Asterisk-labeled OvFAE1-1 functions as the discontinuous fatty acid elongase.

For genome annotation, root, stem, leaf, flower, and silique tissues were collected from the experimental fields at the flowering stage. About 0.5 g of tissue was used to extract RNA for transcriptomics analysis. To identify DEGs in seeds from different developmental stages, we selected seeds from siliques at 22, 26, 32, and 40 DAF for RNA-seq, using three biological replicates for each stage.

Genome assembly and quality assessment

The Genome Characteristics Estimation software package was first used to conduct a genome survey (Liu et al., 2013). HiFi reads sequenced on the PacBio platform were then used for *de novo* assembly with the Hifiasm software package (Cheng et al., 2021). High fidelity (HiFi) reads were aligned to the draft assembly using the minimap2 software package (Li, 2018) and polished three times according to the alignment results using the Racon software package (Vaser et al., 2017). Next, BWA-MEM (maximal exact matches) was used to map the Illumina paired-end reads to the corrected primarily assembled draft (Li, 2013), and SAMtools was used to filter out low-quality reads (Li et al., 2009a, 2009b). Pilon was then used with default parameters to correct the assembled contigs using the short reads (Walker et al., 2014). The Purge Haplotigs pipeline was used to identify and reassign the duplicate contigs of the polished draft, with the parameter "align_cov" set to 65 (Roach et al., 2018).

For Hi-C scaffolding, around 60 Gb of clean Illumina paired-end Hi-C reads were first mapped to the contigs using Juicer (Durand et al., 2016). The contigs were then corrected, clustered, ordered, and oriented using the 3D DNA pipeline (Dudchenko et al., 2017). HiC-Pro was used to draw the Hi-C interaction matrix (Servant et al., 2015).

The BUSCO software package was used to assess the integrity of single-copy gene clusters (Simao et al., 2015). The LTR_FINDER and LTRharvest software packages were used to annotate LTRs in the assembly (Xu and Wang, 2007; Ellinghaus et al., 2008). These results were integrated and used to calculate the LAI with LTR_retriever (Ou and Jiang, 2018; Ou et al., 2018).

Genome annotation

RepeatModeler (http://www.repeatmasker.org/RepeatModeler/) was used to predict and construct the repeat sequence library

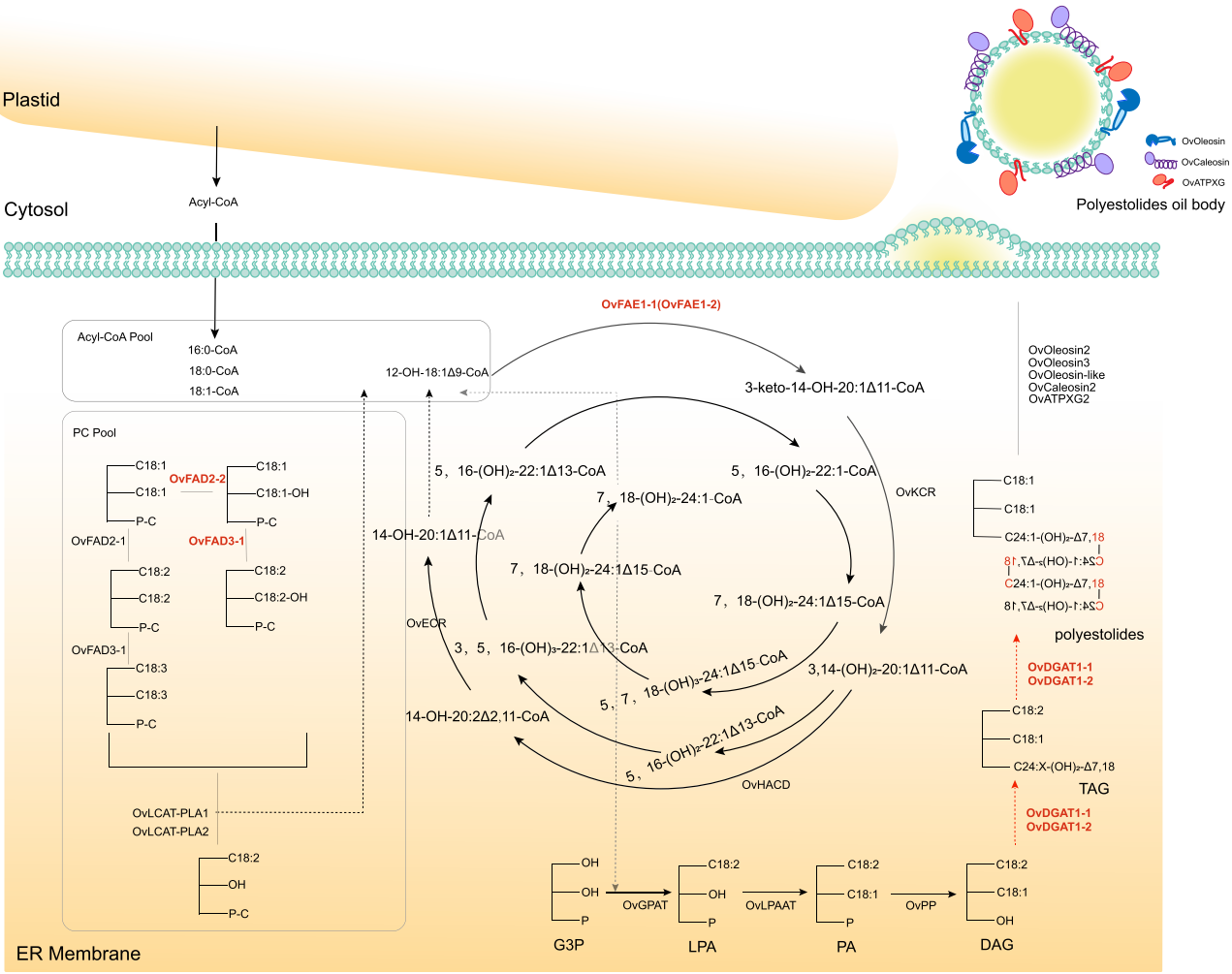


Figure 6. Proposed metabolic pathway of diOH-FAs in seeds of *O. violaceus*.

Oleoyl-phosphatidylcholine (oleoyl-PC) is hydroxylated by OvFAD2-2, and the resulting ricioleoyl-PC is desaturated by OvFAD3-1 or hydrolyzed directly. The free hydroxy acyl-CoA is elongated by OvFAE1-1. The elongated 3-keto-14-hydroxy-20:1Δ11-CoA (3-keto-14-hydroxy-20:2Δ11,14-CoA, not shown) is reduced to 3,14-dihydroxy-20:1Δ11-CoA (3,14-dihydroxy-20:2Δ11,14-CoA, not shown). The intermediates are elongated again by OvFAE1-1 rather than dehydrated by 3-hydroxyacyl-CoA dehydratase. diOH-FA is generated by the discontinuous biosynthesis pathway. Non-hydroxylated fatty acids are assembled on the glycerol backbone. OvDGAT1-related enzymes transfer the dihydroxy acyl chain to the *sn*-3 position of DAG. We propose that OvDGAT1-related enzymes will continue to transfer acyl chains to the Δ18 hydroxy group of dihydroxy acyl on the *sn*-3 of TAG. The final polyestolides can contain one to three extra acyl chains (normal acyl chains or dihydroxy acyl chains). The assembled polyestolides are then stored in the oil body. Red labels, key genes related to diOH-FA metabolism. Pathway modified from Li-Bession et al. (2013) and Li et al. (2018).

of our assembly. RepeatMasker software was used to mask the repeat sequences in the assembly using the parameters "-E Wublast -GFF -S -xsmall" (Tarailo-Graovac et al., 2009).

For structure annotation and function annotation, we performed *de novo* annotation. First, the RNA-seq data from root, stem, leaf, flower, and silique tissues were aligned to the assembled genome using HISAT2 (Kim et al., 2015). Based on the alignment results, Trinity was used for transcriptome assembly (Borodina et al., 2011; Grabherr et al., 2011). We used StringTie to annotate gene structure (Pertea et al., 2015) and obtain transcriptome evidence. For homology annotation, we downloaded the published protein sequences of Chiifu (Zhang et al., 2018), Zhongshuang 11 (Song et al., 2020), and *Arabidopsis* (TAIR10) as references. For *ab initio* prediction, we used BRAKER2 and AUGUSTUS for training and carried out *ab initio* prediction of gene structures (Stanke et al., 2004; Stanke et al., 2006; Stanke et al., 2008; Bruna et al., 2020; Bruna et al., 2021; Hoff et al., 2016; Hoff et al., 2019; Lomsadze et al., 2005; Lomsadze et al., 2014). Finally, we used the MAKER (annotation software, http://www.yandell-lab.org/software/index.htm) pipeline to integrate the results of these three methods (transcriptome evidence, homologous proteins, and *ab initio* prediction) to obtain the final annotation (Cantarel et al., 2008).

InterProScan was used to perform functional annotation of the protein-coding genes (Mulder and Apweiler, 2007). The Blastp software package was used to compare protein sequences with the Gene Ontology, KEGG (Ogata et al., 1999; Kanehisa and Goto, 2000), and other protein sequence databases.

Phylogenetic tree construction and species divergence time

We first downloaded the genomic information for 12 species closely related to *O. violaceus*: *Aethionema arabicum*, *A. thaliana*, *Brassica nigra*, *Brassica rapa*, *Brassica oleracea*, *Capsella rubella*, *Thellungiella parvula*, *I. indigotica*, *Sinapis alba*, *Raphanus sativus*, *Eutrema salsugineum*, and *Sisymbrium irio* from The Arabidopsis Information Resource (https://www.arabidopsis.org/) and the Brassica database (http://brassicadb.cn/). The protein sequences of the 12 species closely related to *O. violaceus* were extracted using gffread (Pertea and Pertea, 2020). Gene family cluster analysis was performed using OrthoFinder (Emms and Kelly, 2019). Protein sequences from the single-copy gene families obtained in the previous step were compared using the MUSCLE software package (Edgar, 2004). Based on the comparison results, the RAxML software package was used to construct a phylogenetic tree of the 13 species, including *O. violaceus*, with the parameters "-f a -x 12 345 -# 1000 -p 12 345 -m PROTGAMMAAUTO" (Stamatakis, 2014). Finally, the species divergence time was estimated using MCMCTree in the PAML software package with the parameter "-p 0.05" (Yang, 2007).

Chromosome karyotype analysis and collinearity analysis

Using the JCVI (collinearity software, https://github.com/tanghaibao/jcvi) software package, we first obtained the collinear relationship between the genomes of *O. violaceus* and *A. thaliana* (Tang et al., 2015). Then the genome was divided according to the 24 GBs of *A. thaliana* (Mandakova and Lysak, 2008). For collinearity analysis of the genomes of *O. violaceus* and *I. indigotica*, we extracted coding sequence (CDS) sequences and annotation information from the genomes of *O. violaceus* and *I. indigotica* and drew a collinearity map with JCVI (Tang et al. 2015).

WGD analysis

The ksd program in the wgd software package was used to calculate the values of the non-synonymous substitution rate *Ka* and the synonymous substitution rate *Ks* for homologous genes (Zwaenepoel and Van de Peer, 2019), and the probability distribution curve of the *O. violaceus* synonymous substitution rate was visualized using the R language. The divergence times of known species were obtained from the TimeTree evolutionary timescale website (http://www.timetree.org). Six time points were selected for correction of divergence time estimates of the phylogenetic tree between *O. violaceus* and the other 12 cruciferous plants. According to TimeTree, the divergence time of *B. rapa* and *B. oleracea* was between 2.02 mya and 3.212 mya, that of *B. nigra* and *R. sativus* was between 7.6 mya and 15.9 mya, that of *S. alba* and *B. nigra* was between 4.6 mya and 21.9 mya, that of *S. irio* and *I. indigotica* was between 11.4 mya and 43.8 mya, that of *A. thaliana* and *C. rubella* was between 7.9 mya and 14.6 mya, and that of *A. thaliana* and *E. salsugineum* was between 19.7 mya and 32.3 mya. The value of the synonymous substitution rate r was calculated according to the formula time = *Ks*/2r (Goldman and Yang, 1994; Hurst, 2002; Li et al., 2009a, 2009b; Tiley et al., 2018) and was around 8.7E−9. Finally, according to the known *Ks* peak value and the synonymous replacement rate, the times of the WGT event of cruciferous plants and the WGD event of *O. violaceus* were calculated.

Transcriptome analysis

We first used the HISAT2 software package to align RNA-seq data of seeds from siliques at 22, 26, 32, and 40 DAF to the assembly (Kim et al., 2015). Gene expression data were then obtained using the featureCounts software package. Finally, differential gene expression data were obtained according to the DEGSeq2 library in R with an adjusted *P* value of 0.05 (Wang et al., 2010; Zhu et al., 2019). WGCNA was performed with the WGCNA library in R using a powerEstimate value of 18 (Langfelder and Horvath, 2008).

Based on the differential expression information, we first performed visual analysis of the number of DEGs. GO and KEGG enrichment analysis of the DEGs was then performed using TBtools software (Chen et al., 2020). According to the enrichment results, visualization was performed using R.

Bioinformatics analysis of the *DGAT1* gene family

RcDGAT1 (NW_017871090.1), *BrDGAT1* (NC_024803.2), *BoDGAT1* (NC_027756.1), and *BnDGAT1* (NC_027765.2) were downloaded from NCBI and *AtDGAT1* (AT2G19450.1) from TAIR. The CDS sequences of *IiDGAT1-1* and *IiDGAT1-2* were extracted from the genome of *I. indigotica*. The CDS sequences of *OvDGAT1-1* and *OvDGAT1-2* were extracted from our assembly and amplified by PCR from the cDNA of developing *O. violaceus* seeds.

MEGA X was used to perform phylogenetic analysis of the *DGAT1* genes (Kumar et al., 1994). The statistical method was set to "maximum likelihood." The test of phylogeny was set to "bootstrap method," with 1000 bootstrap replications. The peptide sequences of these genes were then obtained, and AtDGAT1, BnDGAT1, RcDGAT1, and OvDGAT1-1/2 were aligned using ClustalX (Jeanmougin et al., 1998). Finally, PyMOL (Delano, 2002) was used to predict the structure of OvDGAT1-1/2.

Data availability

The data supporting the findings of this work are available in the paper and its supplemental information. The whole-genome shotgun sequencing data, PacBio CCS sequencing data (HiFi reads), Hi-C data, and transcriptomes of different *O. violaceus* tissues have been deposited at NCBI under BioProject number PRJNA828624 and at the China National Genomics Data Center (https://ngdc.cncb.ac.cn) under accession ID CRA008040. The nucleotide sequencing data for OvDGAT1-related genes identified in this study have been deposited at NCBI GenBank under accession numbers ON325585 (*OvDGAT1-1*) and ON325586 (*OvDG*AT1-2).

SUPPLEMENTAL INFORMATION

Supplemental information is available at *Plant Communications Online*.

FUNDING

This work was supported by the National Natural Science Foundation of China (U20A2034 and 31871659) and the China Agriculture Research System (CARS-12) (to C.Z.). E.B.C. was supported by funding from the National Science Foundation (Plant Genome IOS-13-39385).

AUTHOR CONTRIBUTIONS

F.H., X.T., and T.Z. performed the experiments. F.H. and P.C. prepared figures and drafted the manuscript. T.Y. helped with genome annotation. C.C.N. and H.A. helped to edit the manuscript. C.Y. and X.G. helped with cytological and genome evolution analysis. Z.L. provided the *O. violaceus* seeds and conceived the study. C.Z. and E.B.C. conceived the study and coordinated discussions and editing of the manuscript. All authors have read and approved the final manuscript.

ACKNOWLEDGMENTS

We thank Lun Guan for help with 3D structure analysis of the DGAT1 protein. No conflict of interest is declared.

Received: April 20, 2022
Revised: July 27, 2022
Accepted: September 5, 2022
Published: September 7, 2022

REFERENCES

Borodina, T., Adjaye, J., and Sultan, M. (2011). A strand-specific library preparation protocol for RNA sequencing. Methods Enzymol. **500**:79–98.

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

Brůna, T., Lomsadze, A., and Borodovsky, M. (2020). GeneMark-EP+: eukaryotic gene prediction with self-training in the space of genes and proteins. NAR Genom. Bioinform. **2**:lqaa026.

Cacciola, F., Beccaria, M., Oteri, M., Utczas, M., Giuffrida, D., Cicero, N., Dugo, G., Dugo, P., and Mondello, L. (2016). Chemical characterisation of old cabbage (*Brassica oleracea* L. var. acephala) seed oil by liquid chromatography and different spectroscopic detection systems. Nat. Prod. Res. **30**:1646–1654.

Cai, X., Chang, L., Zhang, T., Chen, H., Zhang, L., Lin, R., Liang, J., Wu, J., Freeling, M., and Wang, X. (2021). Impacts of allopolyploidization and structural variation on intraspecific diversification in *Brassica rapa*. Genome Biol. **22**:166.

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

Cartea, E., De Haro-Bailón, A., Padilla, G., Obregón-Cano, S., Del Rio-Celestino, M., and Ordás, A. (2019). Seed oil quality of *Brassica napus* and *Brassica rapa* germplasm from northwestern Spain. Foods **8**:292.

Chen, C., Chen, H., Zhang, Y., Thomas, H.R., Frank, M.H., He, Y., and Xia, R. (2020). TBtools: an integrative toolkit developed for interactive analyses of big biological data. Mol. Plant **13**:1194–1202.

Cheng, H., Concepcion, G.T., Feng, X., Zhang, H., and Li, H. (2021). Haplotype-resolved de novo assembly using phased assembly graphs with hifiasm. Nat. Methods **18**:170–175.

Couvreur, T.L.P., Franzke, A., Al–Shehbaz, I.A., Bakker, F.T., Koch, M.A., and Mummenhoff, K. (2010). Molecular phylogenetics, temporal diversification, and principles of evolution in the Mustard Family (Brassicaceae). Mol. Biol. Evol. **27**:55–71.

Cuddapah, S., Barski, A., Cui, K., Schones, D.E., Wang, Z., Wei, G., and Zhao, K. (2009). Native chromatin preparation and Illumina/Solexa library construction. Cold Spring Harb. Protoc. **2009**. pdb.prot5237.

Delano, W.L. (2002). The PyMOL molecular graphics System version 1.(schrödinger). http://www.pymol.org.

Dudchenko, O., Batra, S.S., Omer, A.D., Nyquist, S.K., Hoeger, M., Durand, N.C., Shamim, M.S., Machol, I., Lander, E.S., Aiden, A.P., et al. (2017). De novo assembly of the Aedes aegypti genome using Hi-C yields chromosome-length scaffolds. Science **356**:92–95.

Durand, N.C., Shamim, M.S., Machol, I., Rao, S.S.P., Huntley, M.H., Lander, E.S., and Aiden, E.L. (2016). Juicer provides a one-click System for analyzing loop-resolution Hi-C experiments. Cell Syst. **3**:95–98.

Edgar, R.C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. **32**:1792–1797.

Ellinghaus, D., Kurtz, S., and Willhoeft, U. (2008). *LTRharvest*, an efficient and flexible software for de novo detection of LTR retrotransposons. BMC Bioinf. **9**:18.

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

Franzke, A., Lysak, M.A., Al-Shehbaz, I.A., Koch, M.A., and Mummenhoff, K. (2011). Cabbage family affairs: the evolutionary history of Brassicaceae. Trends Plant Sci. **16**:108–116.

Goldman, N., and Yang, Z. (1994). A codon-based model of nucleotide substitution for protein-coding DNA sequences. Mol. Biol. Evol. **11**:725–736.

Grabherr, M.G., Haas, B.J., Yassour, M., Levin, J.Z., Thompson, D.A., Amit, I., Adiconis, X., Fan, L., Raychowdhury, R., Zeng, Q., et al. (2011). Full-length transcriptome assembly from RNA-seq data without a reference genome. Nat. Biotechnol. **29**:644–652.

Guo, X., Mandáková, T., Trachtová, K., Özüdoğru, B., Liu, J., and Lysak, M.A. (2021). Linked by ancestral bonds: multiple whole-genome duplications and reticulate evolution in a Brassicaceae tribe. Mol. Biol. Evol. **38**:1695–1714.

Hoff, K.J., Lomsadze, A., Borodovsky, M., and Stanke, M. (2019). Whole-genome annotation with BRAKER. Methods Mol. Biol. **1962**:65–95.

Hoff, K.J., Lange, S., Lomsadze, A., Borodovsky, M., and Stanke, M. (2016). BRAKER1: unsupervised RNA-seq-based genome annotation with GeneMark-ET and AUGUSTUS. Bioinformatics **32**:767–769.

Hu, H., Hu, Q., Al-Shehbaz, I.A., Luo, X., Zeng, T., Guo, X., and Liu, J. (2016). Species delimitation and interspecific relationships of the genus *Orychophragmus (Brassicaceae)* inferred from whole chloroplast genomes. Front. Plant Sci. **7**:1826.

Huang, X.C., German, D.A., and Koch, M.A. (2020). Temporal patterns of diversification in Brassicaceae demonstrate decoupling of rate shifts and mesopolyploidization events. Ann. Bot. **125**:29–47.

Hurst, L.D. (2002). The *Ka/Ks* ratio: diagnosing the form of sequence evolution. Trends Genet. **18**:486.

Jeanmougin, F., Thompson, J.D., Gouy, M., Higgins, D.G., and Gibson, T.J. (1998). Multiple sequence alignment with Clustal X. Trends Biochem. Sci. **23**:403–405.

Kanehisa, M., and Goto, S. (2000). KEGG: kyoto encyclopedia of genes and genomes. Nucleic Acids Res. **28**:27–30.

Kang, M., Wu, H., Yang, Q., Huang, L., Hu, Q., Ma, T., Li, Z., and Liu, J. (2020). A chromosome-scale genome assembly of *Isatis indigotica*, an important medicinal plant used in traditional Chinese medicine. Hortic. Res. **7**:18.

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

Kumar, S., Tamura, K., and Nei, M. (1994). MEGA: molecular evolutionary genetics analysis software for microcomputers. Comput. Appl. Biosci. **10**:189–191.

Langfelder, P., and Horvath, S. (2008). WGCNA: an R package for weighted correlation network analysis. BMC Bioinf. **9**:559.

Li, J., Zhang, Z., Vang, S., Yu, J., Wong, G.K.S., and Wang, J. (2009a). Correlation between Ka/Ks and Ks is related to substitution model and evolutionary lineage. J. Mol. Evol. **68**:414–423.

Li, H. (2013). Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. Genomics **1303**.

Li, H. (2018). Minimap2: pairwise alignment for nucleotide sequences. Bioinformatics **34**:3094–3100.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., and Durbin, R. (2009b). The sequence alignment/map (SAM) format and SAMtools. Bioinformatics **25**:2078–2079.

Li, X., Teitgen, A.M., Shirani, A., Ling, J., Busta, L., Cahoon, R.E., Zhang, W., Li, Z., Chapman, K.D., Berman, D., et al. (2018). Discontinuous fatty acid elongation yields hydroxylated seed oil with improved function. Nat. Plants **4**:711–720.

Li, Z.Y., and Liu, H.L. (1995). A study on meiotic pairing of *Orychophragmus violaceus*. J. Huazhong Agric. Univ. **14**:435–439. (in Chinese with English abstract).

Li, Z.Y., Liu, H.L., and Heneen, W.K. (1996). Meiotic behaviour in intergeneric hybrids between *Brassica napus* and *Orychophragmus violaceus*. Hereditas **125**:69–75.

Li-Beisson, Y., Shorrosh, B., Beisson, F., Andersson, M.X., Arondel, V., Bates, P.D., Baud, S., Bird, D., DeBono, A., Durrett, T.P., et al. (2013). Acyl-lipid metabolism. Arabidopsis Book **11**:e0161.

Liu, B., Shi, Y., Yuan, J.Y., Hu, X.S., Zhang, H., Li, N., Li, Z.Y., Chen, Y.X., Mu, D.S., and Fan, W. (2013). Estimation of genomic characteristics by analyzing k-mer frequency in de novo genome projects. Quant. Biol. **35**:62–67.

Liu, J., Cao, W.D., Rong, X.N., and Liang, J.F. (2012). Nutritional characteristics of *Orychophragmus violaceus* in north China. Soil and Fertilizer Sciences in China **1**:78–82. (In Chinese).

Liu, L., Zhao, B., Tan, D., and Wang, J. (2011). Phylogenetic relationships of Brassicaceae in China: insights from a non-coding chloroplast, mitochondrial, and nuclear DNA data set. Biochem. Syst. Ecol. **39**:600–608.

Lomsadze, A., Burns, P.D., and Borodovsky, M. (2014). Integration of mapped RNA-Seq reads into automatic training of eukaryotic gene finding algorithm. Nucleic Acids Res. **42**:e119.

Lomsadze, A., Ter-Hovhannisyan, V., Chernoff, Y.O., and Borodovsky, M. (2005). Gene identification in novel eukaryotic genomes by self-training algorithm. Nucleic Acids Res. **33**:6494–6506.

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

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

Lysak, M.A., Mandáková, T., and Schranz, M.E. (2016). Comparative paleogenomics of crucifers: ancestral genomic blocks revisited. Curr. Opin. Plant Biol. **30**:108–115.

Mandáková, T., and Lysak, M.A. (2016). Painting of Arabidopsis chromosomes with chromosome-specific BAC clones. Curr. Protoc. Plant Biol. **1**:359–371.

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

Mandáková, T., Li, Z., Barker, M.S., and Lysak, M.A. (2017). Diverse genome organization following 13 independent mesopolyploid events in Brassicaceae contrasts with convergent patterns of gene retention. Plant J. **91**:3–21.

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

Mulder, N., and Apweiler, R. (2007). InterPro and InterProScan. Methods Mol. Biol. **396**:59–70.

Nystedt, B., Street, N.R., Wetterbom, A., Zuccolo, A., Lin, Y.C., Scofield, D.G., Vezzi, F., Delhomme, N., Giacomello, S., Alexeyenko, A., et al. (2013). The Norway spruce genome sequence and conifer genome evolution. Nature **497**:579–584.

Ogata, H., Goto, S., Sato, K., Fujibuchi, W., Bono, H., and Kanehisa, M. (1999). KEGG: Kyoto Encyclopedia of Genes and Genomes. Nucleic Acids Res. **27**:29–34.

Ou, S., and Jiang, N. (2018). LTR_retriever: a highly accurate and sensitive program for identification of long terminal repeat retrotransposons. Plant Physiol. **176**:1410–1422.

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

Pertea, G., and Pertea, M. (2020). GFF Utilities: GffRead and GffCompare. F1000Res. **9**. ISCB Comm J-304.

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

Rensing, S.A., Lang, D., Zimmer, A.D., Terry, A., Salamov, A., Shapiro, H., Nishiyama, T., Perroud, P.F., Lindquist, E.A., Kamisugi, Y., et al. (2008). The Physcomitrella genome reveals evolutionary insights into the conquest of land by plants. Science **319**:64–69.

Roach, M.J., Schmidt, S.A., and Borneman, A.R. (2018). Purge Haplotigs: allelic contig reassignment for third-gen diploid genome assemblies. BMC Bioinf. **19**:460.

Romsdahl, T., Shirani, A., Minto, R.E., Zhang, C., Cahoon, E.B., Chapman, K.D., and Berman, D. (2019). Nature-guided synthesis of advanced bio-lubricants. Sci. Rep. **9**:11711.

Rout, K., Yadav, B.G., Yadava, S.K., Mukhopadhyay, A., Gupta, V., Pental, D., and Pradhan, A.K. (2018). QTL landscape for oil content in Brassica juncea: analysis in multiple Bi-parental populations in high and "0" erucic background. Front. Plant Sci. **9**:1448.

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

Schranz, M.E., Song, B.H., Windsor, A.J., and Mitchell-Olds, T. (2007). Comparative genomics in the Brassicaceae: a family-wide perspective. Curr. Opin. Plant Biol. **10**:168–175.

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

Shan, W., Kubová, M., Mandáková, T., and Lysak, M.A. (2021). Nuclear organization in crucifer genomes: nucleolus-associated telomere clustering is not a universal interphase configuration in *Brassicaceae*. Plant J. **108**:528–540.

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

Song, J.M., 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.

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

Stanke, M., Diekhans, M., Baertsch, R., and Haussler, D. (2008). Using native and syntenically mapped cDNA alignments to improve *de novo* gene finding. Bioinformatics **24**:637–644.

Stanke, M., Schöffmann, O., Morgenstern, B., and Waack, S. (2006). Gene prediction in eukaryotes with a generalized hidden Markov model that uses hints from external sources. BMC Bioinf. **7**:62.

Stanke, M., Steinkamp, R., Waack, S., and Morgenstern, B. (2004). AUGUSTUS: a web server for gene finding in eukaryotes. Nucleic Acids Res. **32**:W309–W312.

Sui, X., Wang, K., Gluchowski, N.L., Elliott, S.D., Liao, M., Walther, T.C., and Farese, R.V., Jr. (2020). Structure and catalytic mechanism of a human triacylglycerol-synthesis enzyme. Nature **581**:323–328.

Tang, H.B., Vivek, K., and Li, J.P. (2015). Jcvi: JCVI Utility Libraries (Zenodo).

Tang, S., Zhao, H., Lu, S., Yu, L., Zhang, G., Zhang, Y., Yang, Q.Y., Zhou, Y., Wang, X., Ma, W., et al. (2021). Genome- and transcriptome-wide association studies provide insights into the genetic basis of natural variation of seed oil content in *Brassica napus*. Mol. Plant **14**:470–487.

Tarailo-Graovac, M., Chen, N., and Chen, N. (2009). Using RepeatMasker to identify repetitive elements in genomic sequences. Curr. Protoc. Bioinformatics. Chapter **4**:Unit 4.10.

The Arabidopsis Genome Inititiative. (2000). Analysis of the genome sequence of the flowering plant. Arabidopsis thaliana Nature **408**:796–815.

Tiley, G.P., Barker, M.S., and Burleigh, J.G. (2018). Assessing the performance of Ks plots for detecting ancient whole genome duplications. Genome Biol. Evol. **10**:2882–2898.

Vaser, R., Sović, I., Nagarajan, N., and Šikić, M. (2017). Fast and accurate *de novo* genome assembly from long uncorrected reads. Genome Res. **27**:737–746.

Walden, N., German, D.A., Wolf, E.M., Kiefer, M., Rigault, P., Huang, X.C., Kiefer, C., Schmickl, R., Franzke, A., Neuffer, B., et al. (2020). Nested whole-genome duplications coincide with diversification and high morphological disparity in Brassicaceae. Nat. Commun. **11**:3795.

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

Wang, L., Feng, Z., Wang, X., Wang, X., and Zhang, X. (2010). DEGseq: an R package for identifying differentially expressed genes from RNA-seq data. Bioinformatics **26**:136–138.

Wang, X., Wang, H., Wang, J., Sun, R., Wu, J., Liu, S., Bai, Y., Mun, J.H., Bancroft, I., Cheng, F., et al. (2011). The genome of the mesopolyploid crop species *Brassica rapa*. Nat. Genet. **43**:1035–1039.

Wen, Y., Du, J., Yu, S., Zhao, F., Li, Z., Liu, J., Cao, G., and Xiang, C. (2020). Comparative study on low temperature germination ability of overwintering green manure. IOP Conf. Ser. Earth Environ. Sci. **598**:012068.

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

Yang, Z. (2007). Paml 4: phylogenetic analysis by maximum likelihood. Mol. Biol. Evol. **24**:1586–1591.

Yin, J.M., Zhong, R.Q., Lin, N., Tang, Z.L., and Li, J.N. (2020). Microspore culture and observations on meiotic chromosome pairing of the haploid in Orychophragmus violaceus. Crop J. **46**:194–203. (in Chinese with English abstract).

Zhang, L., Cai, X., Wu, J., Liu, M., Grob, S., Cheng, F., Liang, J., Cai, C., Liu, Z., Liu, B., et al. (2019). Improved *Brassica rapa* reference genome by single-molecule sequencing and chromosome conformation capture technologies. Hortic. Res. **6**:124.

Zhang, L.J., and Dai, S.L. (2005). The value of development of Orychophragmus violaceus and its landscape utilization. Beijing Landscape **4**:43–45.

Zhou, L.R., Yu, Y., Song, R.X., He, X.J., Jiang, Y., Li, X.F., and Yang, Y. (2009). Phylogenetic relationships within the *Orychophragmus violaceus* complex (Brassicaceae) endemic to China. Acta Bot. Yunnanica **31**:127–137.

Zhou, T.Y., Guan, K.J., and Guo, R.L. (1987). Flora Reipublicae Popularis Sinicae, 33 (Science Press), pp. 40–43.

Zhu, A., Ibrahim, J.G., and Love, M.I. (2018). Heavy-tailed prior distributions for sequence count data: removing the noise and preserving large differences. Bioinformatics **35**:2084–2092.

Zuo, S., Guo, X., Mandáková, T., Edginton, M., Al-Shehbaz, I.A., and Lysak, M.A. (2022). Genome diploidization associates with cladogenesis, trait disparity, and plastid gene evolution. Plant Physiol. **190**:403–420.

Zwaenepoel, A., and Van de Peer, Y. (2019). wgd-simple command line tools for the analysis of ancient whole-genome duplications. Bioinformatics **35**:2153–2155.