Plant Communications
Research article

CellPress
Partner Journal

Genomic evolution and insights into agronomic trait innovations of *Sesamum* species

Hongmei Miao[1,7], Lei Wang[2,7], Lingbo Qu[3,7], Hongyan Liu[4], Yamin Sun[2], Meiwang Le[5], Qiang Wang[6], Shuangling Wei[1], Yongzhan Zheng[1], Wenchao Lin[2], Yinghui Duan[1], Hengchun Cao[1], Songjin Xiong[2], Xuede Wang[3], Libin Wei[1], Chun Li[1], Qin Ma[1], Ming Ju[1], Ruihong Zhao[1], Guiting Li[1], Cong Mu[1], Qiuzhen Tian[1], Hongxian Mei[1], Tide Zhang[1], Tongmei Gao[1] and Haiyang Zhang[1,*]

[1]Henan Sesame Research Center, Henan Academy of Agricultural Sciences, Zhengzhou 450002, China

[2]TEDA School of Biological Sciences and Biotechnology, Nankai University, Tianjin 300457, China

[3]College of Food Science and Technology, Henan Technology University, Zhengzhou 450001, China

[4]Institute of Plant Protection Research, Henan Academy of Agricultural Sciences, Zhengzhou 450002, China

[5]Crops Research Institute, Jiangxi Academy of Agricultural Sciences, Nanchang 330200, China

[6]Crop Research Institute, Anhui Academy of Agricultural Sciences, Hefei 230031, China

[7]These authors contributed equally to this article.

***Correspondence: Haiyang Zhang (zhanghaiyang@zzu.edu.cn)**

https://doi.org/10.1016/j.xplc.2023.100729

ABSTRACT

Sesame is an ancient oilseed crop with high oil content and quality. However, the evolutionary history and genetic mechanisms of its valuable agronomic traits remain unclear. Here, we report chromosome-scale genomes of cultivated sesame (*Sesamum indicum* L.) and six wild *Sesamum* species, representing all three karyotypes within this genus. Karyotyping and genome-based phylogenic analysis revealed the evolutionary route of *Sesamum* species from n = 13 to n = 16 and revealed that allotetraploidization occurred in the wild species *Sesamum radiatum*. Early divergence of the *Sesamum* genus (48.5–19.7 million years ago) during the Tertiary period and its ancient phylogenic position within eudicots were observed. Pan-genome analysis revealed 9164 core gene families in the 7 *Sesamum* species. These families are significantly enriched in various metabolic pathways, including fatty acid (FA) metabolism and FA biosynthesis. Structural variations in *SiPT1* and *SiDT1* within the phosphatidyl ethanolamine-binding protein gene family lead to the genomic evolution of plant-architecture and inflorescence-development phenotypes in *Sesamum*. A genome-wide association study (GWAS) of an interspecific population and genome comparisons revealed a long terminal repeat insertion and a sequence deletion in *DIR* genes of wild *Sesamum angustifolium* and cultivated sesame, respectively; both variations independently cause high susceptibility to *Fusarium* wilt disease. A GWAS of 560 sesame accessions combined with an overexpression study confirmed that the *NAC1* and *PPO* genes play an important role in upregulating oil content of sesame. Our study provides high-quality genomic resources for cultivated and wild *Sesamum* species and insights that can improve molecular breeding strategies for sesame and other oilseed crops.

Key words: *Sesamum*, genome evolution, structural variation, plant architecture, *Fusarium* wilt disease, oil content

Miao H., Wang L., Qu L., Liu H., Sun Y., Le M., Wang Q., Wei S., Zheng Y., Lin W., Duan Y., Cao H., Xiong S., Wang X., Wei L., Li C., Ma Q., Ju M., Zhao R., Li G., Mu C., Tian Q., Mei H., Zhang T., Gao T., and Zhang H. (2024). Genomic evolution and insights into agronomic trait innovations of *Sesamum* species. Plant Comm. **5**, 100729.

INTRODUCTION

Sesame (*Sesamum indicum* L., 2n = 2x = 26) is the only cultivated species in the *Sesamum* genus of the Pedaliaceae family and an early eudicot crop that diverged from the Asterid branch (Joshi, 1961; Zhang et al., 2013a, 2013b) (Supplemental Figure 1). The history of sesame cultivation can be traced back to 3000–3050 BC in the Harappa Valley of the Indian subcontinent (Bedigian and Harlan, 1986). Sesame has the highest oil content (OC;

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.

average 55%) of all annual oilseed crops (Zhang et al., 2013b). Moreover, its seeds contain approximately 0.8% lignans, and it is regarded as "the queen of the plant oil crop seeds" for its high resistance to oxidation and rancidity (Bedigian and Harlan, 1986; Zhang et al., 2019). Sesame, cultivated mainly in tropical and subtropical regions, exhibits a higher tolerance to drought, high temperature, and soil infertility and a lower resistance to *Fusarium* wilt and stem rot diseases, waterlogging, and low temperature than most wild *Sesamum* species (Nimmakayala et al., 2011). Sesame's short life cycle (70–90 days), abundant seeds (~6000 seeds per plant), and small genome (354 Mb) make it a valuable model for studying angiosperm evolution and biological traits (Zhang et al., 2019).

The *Sesamum* genus comprises 36 species that can be categorized into three groups according to their basic chromosome numbers (n = 13, 16, or 32) (Kobayashi, 1991; Zhang et al., 2019, 2021). Certain wild *Sesamum* species, such as *Sesamum alatum*, *Sesamum calycinum*, *Sesamum latifolium*, and *Sesamum radiatum*, exhibit high resistance to diseases, waterlogging, drought, and insects, earning them a prominent position in academic research (Nimmakayala et al., 2011; Zhang et al., 2019; Miao et al., 2021a). The Sesame Genome Working Group initiated the Sesame Genome Project in 2010 to address the evolutionary history of the *Sesamum* genus and provide a comprehensive genetic foundation for accelerating genomics-assisted breeding (Zhang et al., 2013b, 2021). Although several preliminary sesame genomes have been reported (Zhang et al., 2013b; Wang et al., 2014, 2022; Uncu et al., 2015; Li et al., 2020), a more accurate chromosome-scale genome analysis is required to achieve the goals of the Sesame Genome Project. Moreover, an understanding of the genetic mechanisms underlying disease resistance, oil biosynthesis, and metabolic regulation in *Sesamum* is necessary to effectively utilize the genetic data obtained from wild *Sesamum* species for sesame breeding.

Here, we report chromosome-scale genomes of cultivated sesame and six wild *Sesamum* species. We identify the genome evolution features of *Sesamum* and analyze their utility in understanding the evolutionary history of the *Sesamum* genus, providing genomic insights into key agronomic traits such as plant architecture, disease resistance, and OC.

RESULTS AND DISCUSSION

Sequencing, assembly, and annotation

We first sequenced the genome of sesame (*S. indicum* var. Yuzhi 11, n = 13) using an Illumina/Solexa (Zhang et al., 2013b) + PacBio SMRT + Roche/454 (Zhang et al., 2016) + bacterial artificial chromosome (BAC) ends (ABI 3730) strategy. The genome was then assembled using a Hi-C library + super-dense SNP genetic map (Zhang et al., 2016) + BAC-fluorescence *in situ hybridization* (FISH) cytogenetic map strategy (Supplemental Figures 2 and 3; supplemental information). The resulting sesame genome assembly (Figure 1; Supplemental Table 1) spanned 312.95 Mb and contained 2210 scaffolds (largest: 28.36 Mb) with a scaffold N50 number and N50 scaffold length of 7 and 20.33 Mb, respectively. We assembled and annotated approximately 281.40 Mb (89.92%) of 13 pseudomolecule sequences as 13 chromosomes. An additional 31.55 Mb (10.08%) within unanchored scaffolds was composed mainly of repetitive sequences.

Our draft sesame genome contains 31 462 predicted protein-coding genes with an average gene length of 2619 bp; 29 166 (92.7%) of these genes mapped to the 13 chromosomes. Interspersed repetitive sequences occupied 47.75% of the sesame genome (supplemental information; Supplemental Table 2). This is comparable to other angiosperm genomes, such as those of potato (*Solanum tuberosum*, 62.2%) (Xu et al., 2011), cacao (*Theobroma cacao*, 25.7%) (Argout et al., 2011), and grapevine (*Vitis vinifera*, 41.4%) (Jaillon et al., 2007). The telomeric repeats in sesame shared the same conserved repeat-unit sequence (TTTAGGG) as those found in other plants (supplemental information). The 45S rDNA and 5S rDNA repeats were detected in most satellite pseudochromosomes, reflecting their fine integrity. In addition, we detected a 153-bp centromeric repeat sequence (named *Si*Cen1) using BAC-FISH (15 specific BACs hybridized at the centromeric and pericentromeric regions of all 13 chromosome pairs) (supplemental information; Supplemental Tables 3 and 7). This enabled subsequent comparisons of species-specific centromeric repeats, independently revealing the features of chromosome karyotype evolution in *Sesamum*.

To explore the origin and evolution of sesame within the *Sesamum* genus, we assembled chromosome-level genome sequences of six wild *Sesamum* species (*S. alatum* [var. 3651, n = 13], *S. latifolium* [var. KEN1, n = 16], *S. angolense* [var. K16, n = 16], *S. calycinum* [var. KEN8, n = 16], *S. angustifolium* [var. G01, n = 16], and *S. radiatum* [var. G02, n = 2x = 32]) representing the three basic chromosome sets using a complicated assembly strategy (Supplemental Figures 1, 4, and 5; supplemental information; Supplemental Tables 4–6). Interspecific SNP genetic maps of *S. calycinum* and *S. angustifolium* were also constructed and used to evaluate the genome assemblies (Supplemental Table 7; Supplemental Figure 6). In addition, we used BioNano mapping to obtain high-quality genomic assemblies of *S. latifolium* and *S. radiatum* (Supplemental Figure 4). The genome sizes of the six wild species varied from 300.7 Mb (*S. angustifolium*) to 668.4 Mb (*S. radiatum*) (Supplemental Table 6). Benchmarking Universal Single-Copy Orthologs annotation (Supplemental Table 8) suggested that these genome assemblies and the annotations of sesame (98.5%) and wild species (91.2–98.1%) were relatively complete and of high quality.

Chromosome groups and phylogeny

To clarify *Sesamum* evolution, we investigated the synonymous substitution rate (Ks) distribution of paralogous genes (Figure 2A). Similar to most dicots, the seven *Sesamum* species underwent whole-genome duplication (WGD) and whole-genome triplication (WGT-γ) events (Figure 2A–a1; supplemental information). *S. radiatum* also underwent lineage-specific WGD-2 (Figure 2A), resulting in a chromosome number of 2n = 4x = 64. Genome fluorescence *in situ* hybridization (GISH) of the seven *Sesamum* species and centromeric repeat FISH indicated that the 32 pairs of chromosomes in *S. radiatum* were derived from a group of 2n = 2x = 32 diploid species (here named AA), which included *S. latifolium*, and another group of 2n = 2x = 32 species (named BB), which included *S. angolense*, *S. calycinum*, and *S. angustifolium* (Figure 2B and Supplemental

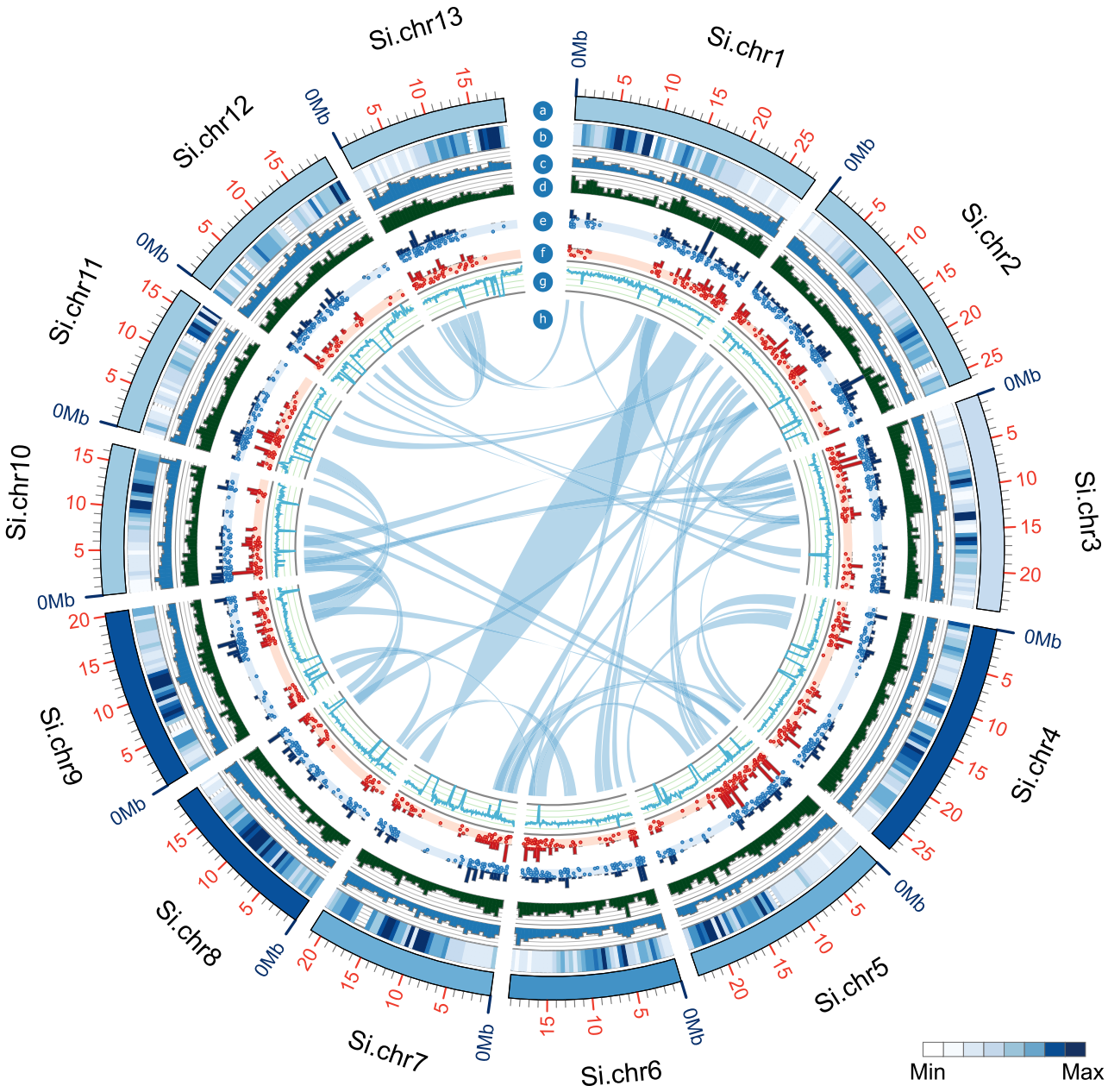

Figure 1. The sesame genome.
Shown are chromosome set and length **(a)**, density of interspersed repetitive sequences (repeat number per 500 kb) **(b)** gene density of pseudomolecule strand + **(c)**, gene density of pseudomolecule strand − (gene number per 500 kb) **(d)**, density (upper line) and location (lower line) of FA biosynthesis- and metabolism-related genes **(e)**, density (upper line) and location (lower line) of disease resistance genes (R genes) **(f)**, GC content **(g)**, and alignment of duplicated blocks between chromosomes **(h)**.

Figure 7; Supplemental Tables 9 and 10). Therefore, *S. radiatum* is an allotetraploid species with two genome subgroups designated AA_{Sra} and BB_{Sra}. Because of the high incompatibility of *S. indicum* with *S. alatum* and the other 2n = 32 *Sesamum* species (Supplemental Table 5) and their weak associated GISH hybridization (Figure 2B), the *S. indicum* and *S. alatum* genomes were designated CC and DD, respectively. These results indicated that typical polyploidization events contributed to species divergence within the *Sesamum* genus, reaffirming the high frequency of polyploidization in angiosperms (Li et al., 2015).

To further investigate the evolution of *Sesamum* within the angiosperms, we performed a phylogenetic analysis using 645 single-copy orthologous genes common to cultivated sesame, five diploid *Sesamum* species, and 15 representative plant species from monocots, eudicots, diploid oilseed crops, and closely related taxa (Figure 2C; supplemental information). The phylogenetic tree indicated that the branch containing Pedaliaceae (*Sesamum* and other genera) and Pharymaceae (*Mimulus* and other genera) diverged from the Solanaceae family (*Solanum* and other genera) approximately 86.0 million

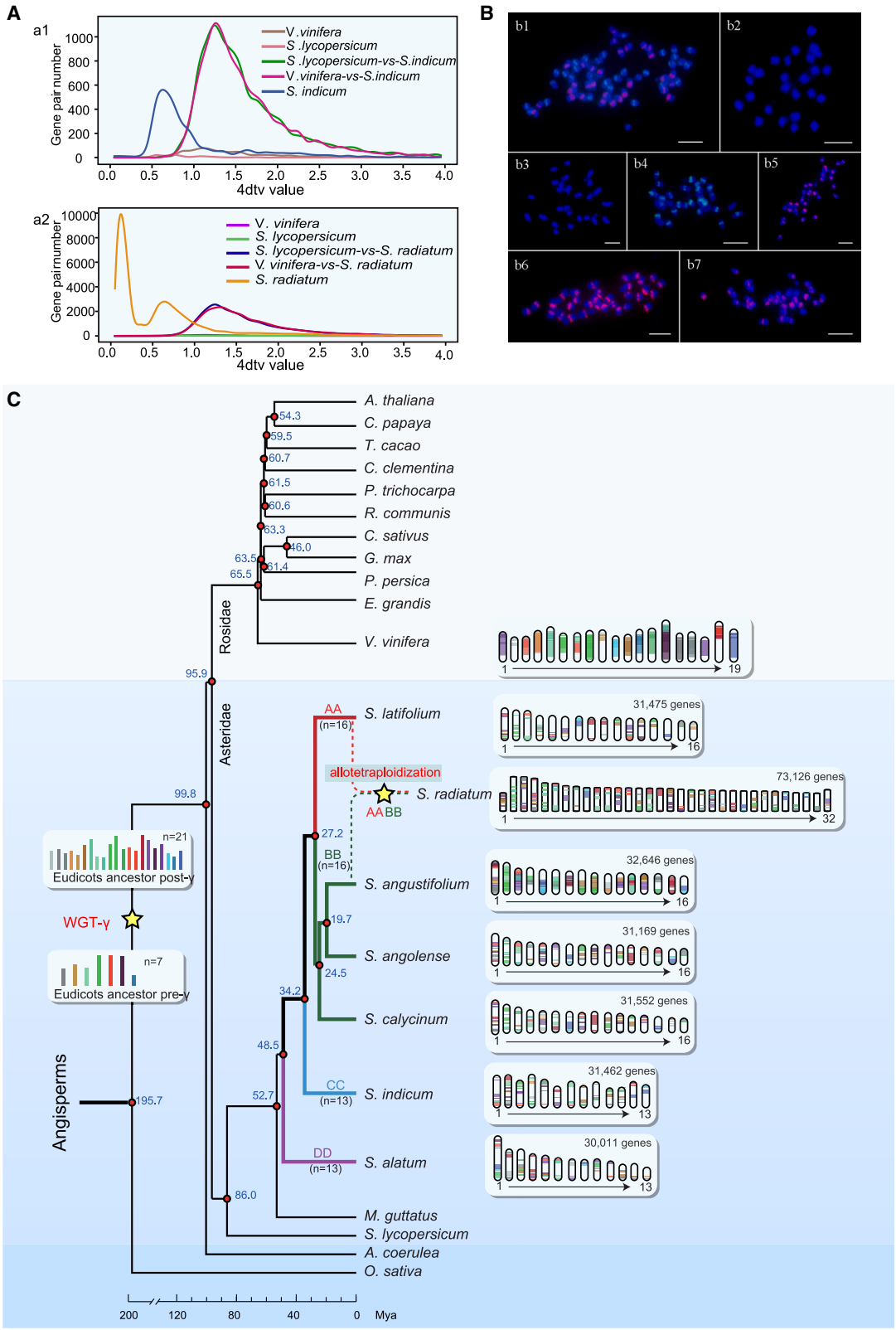

Figure 2. Genome evolution of *Sesamum*.

(A) 4DTv distribution for duplicated gene pairs of sesame **(a1)** and the allotetraploid species *S. radiatum* **(a2)** compared with grapevine and tomato. **(B)** FISH hybridization of the two centromere repeats of *S. radiatum* with the chromosomes of *S. radiatum* **(b1)**, *S. alatum* **(b2)**, *S. indicum* **(b3)**, *S. latifolium* **(b4)**, *S. calycinum* **(b5)**, *S. angolense* **(b6)**, and *S. angustifolium* **(b7)**.

(legend continued on next page)

years ago (mya) during the Cretaceous period (145.50–65.50 mya; Figure 2C). The Pedaliaceae branch then diverged from Pharymaceae approximately 52.7 mya during the early Tertiary period (65.5–23.3 mya), which corresponds to the recently reported evolutionary rhythm of flowering plants (Zhang et al., 2020). Cultivated sesame and wild diploid species diversified approximately 48.5–19.7 mya. *S. alatum* (n = 13, DD) was the first species to diverge, followed by *S. indicum* (n = 13, CC). Despite a lack of prior knowledge regarding the progenitor of modern cultivated sesame, our analysis indicated that *S. indicum* (cultivated sesame) diverged from *S. latifolium* before 34.2 mya, with the other four diploid species (n = 16) diverging 27.2–19.7 mya. The formation of the allotetraploid species *S. radiatum* likely occurred at a later stage, possibly ~5 mya, which coincides with the period in which polyploidy events have been identified in other crops (Myburg et al., 2014). Thus, we inferred that the evolutionary route of *Sesamum* species involved an increase in chromosome number from n = 13 to n = 16. The early divergence of *Sesamum* species, together with the large time scale (48.5–19.7 mya) of divergence within this genus, highlights their specific and early evolutionary position among eudicots (Zhang et al., 2013b).

Pseudochromosomes of the *Sesamum* progenitor

Genome synteny analysis can be used to evaluate genome conservation and identify the characteristics of progenitor species by clustering collinearly matched orthologous or paralogous gene pairs (Zhang et al., 2020). Genome synteny was used to compare the *Sesamum* species with grapevine (Jaillon et al., 2007), which has undergone a WGT-γ event but no WGD events (Wood et al., 2009; supplemental information). The ratio of orthologous genes in syntenic blocks between *V. vinifera* and individual *Sesamum* species varied from 25.35% (*V. vinifera* vs. *S. latifolium*) to 31.82% (*V. vinifera* vs. *S. calycinum*; $P < 1E{-}10$; Supplemental Table 11). The genome synteny percentages between *V. vinifera* and *Sesamum* were higher than those obtained by comparing *V. vinifera* with various angiosperm species such as poplar (9.98%), papaya (11.17%), cacao (15.53%), and soybean (8.64%) (Salse, 2012), suggesting that *Sesamum* is a highly conserved and ancient genus.

We then established the genome structure of *Sesamum* spp. using the ancestral eudicot karyotype (AEK post-γ; 9022 contiguous ancestral genes [protogenes] in 21 paleochromosomes) (Murat et al., 2017; supplemental information; Figure 2C). We detected 106 syntenic blocks in the cultivated sesame genome, which contained 4653 (15.95%) collinear orthologous genes ($P < 1E{-}10$; Source Data 3 for Figure 2; Supplemental Tables 12 and 13). The ratio of orthologous genes in the 5 wild diploid species varied from 12.84% (4986 orthologous genes in *S. latifolium*) to 18.27% (4489 in *S. angustifolium*). A high ratio of syntenic blocks in the chromosomes of the ancestral eudicot and the seven tested *Sesamum* species (Supplemental Figure 8) further confirmed that the *Sesamum* genus experienced WGD and tetraploidization events. This finding highlights the high level of evolutionary conservation within this genus.

To determine the basic chromosome number and possible karyotype structure of the *Sesamum* progenitor, we aligned the syntenic blocks and collinear orthologous genes of four representative genomes, sesame (CC group), *S. latifolium* (AA_{Sla}), *S. calycinum* (BB_{Sca}), and *S. angustifolium* (BB_{Sau}), whose genome assemblies were supported by SNP maps or BioNano assembly platforms (Figure 3; supplemental information; Supplemental Tables 7 and 10). Overall, we detected seven chromosomal fissions and four chromosomal fusions between sesame (CC) and *S. latifolium* (AA_{Sla}) (Figure 3A). Only 1 fission and 1 fusion event were detected within the three n = 16 diploid species (AA_{Sla}, BB_{Sca}, and BB_{Sau}) (Figure 3B). For the AA and BB genomes, the number of orthologous gene pairs between AA_{Sla} and BB_{Sca} and BB_{Sau} were 3501 and 3687, respectively. The chromosome pairs of *S. angustifolium* (BB_{Sau}) aligned well with the corresponding chromosomes in *S. calycinum* (BB_{Sca}), demonstrating the structural conservation of these two genomes.

To investigate the chromosomal relationships between tetraploid and diploid *Sesamum* species, we aligned the 32 pseudochromosomes of *S. radiatum* with the AA_{Sla} and BB_{Sau} genomes using MCScanX ($E < 1E{-}05$) (Figure 3C; supplemental information). The *S. radiatum* genome was successfully grouped into the AA_{Sra} and BB_{Sra} subgenomes, confirming the evolutionary relationships between these species. This finding aligned with the prior results of subgenome duplication analysis (Figure 2A and Supplemental Figure 8) and the distribution of the two centromeric repeats (*Sra*Cen1 and *Sra*Cen2) (Figure 2B and Supplemental Figure 7). Moreover, this further confirmed the high degree of conservation between the AA_{Sra} and BB_{Sra} subgenomes.

These results indicate that although speciation within the ancient and highly conserved *Sesamum* genus primarily involved WGD events, polyploidization and chromosomal fusions and fissions have also played a vital role in the formation of species diversity. High incompatibility and weak GISH hybridization among most *Sesamum* species (Supplemental Table 5; Figure 2B), alongside relatively low identity of interspecific genetic maps between the two closely related species *S. calycinum* and *S. angustifolium* (Supplemental Table 7; Supplemental Figure 6), also indicated the distant and species-specific diversity of *Sesamum* spp.

Sesamum- and species-specific gene families

To systematically explore the genome evolution of *Sesamum*, we first compared the six diploid *Sesamum* genomes with the common ancestor of angiosperms, analyzing the expansion and contraction of orthologous gene clusters (Supplemental Figure 9). A total of 1172 gene families were expanded and 1549 families were contracted in sesame compared with the other 15 crops (Supplemental Figure 9A). As reported for *Aquilegia coerulea*, grapevine, *Eucalyptus grandis*, cucumber,

(C) Phylogenetic evolutionary position and comparison of *Sesamum* with AEK using ancestral syntenic blocks. Light gray and light blue blocks cover the Rosidae and Asteridae branches of the eudicots, respectively. The blue block covers monocots. Chromosomal syntenic blocks on the right show the ancestral synteny results of *V. vinifera* and the seven *Sesamum* species with the AEK. 4DTv, four-fold degenerate site; WGT, whole-genome triplication; mya, million years ago.

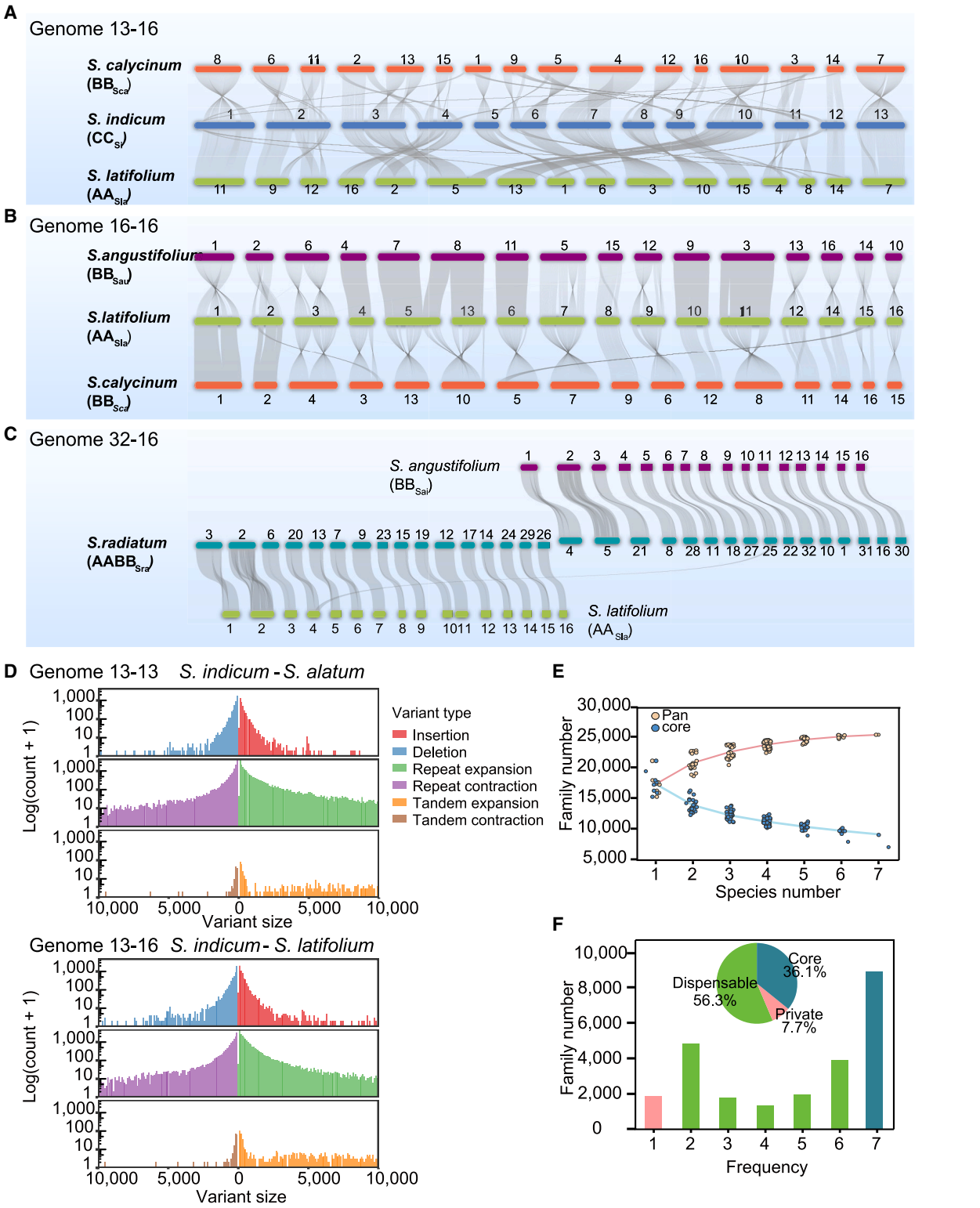

Figure 3. Genome evolution of the *Sesamum* genus.

(A) Correspondence of genome group CC (n = 13) to groups AA and BB (n = 16).

(B) Correspondence of genome group AA (n = 13) to group BB (n = 16).

(C) Correspondence of subgenome groups AA_{Sra} and BB_{Sra} of *S. radiatum* (n = 32) to groups AA_{Sla} and BB_{Sau} (n = 16), respectively.

(D) Distribution of SVs between *S. indicum* and *S. alatum* and between *S. indicum* and *S. latifolium*.

(E) Pan-genome analysis of *Sesamum* species.

(F) Core, dispensable, and private gene families in pan-genomes of *Sesamum*.

Prunus persica, castor bean, citrus, cacao, and papaya, loss-of-function mutations may have played an important role in *Sesamum* evolution (Supplemental Figure 9A and 9B). Notably, dozens of expanded and contracted gene families in *Sesamum* are involved in disease resistance and fatty acid (FA) biosynthesis and metabolism (Supplemental Tables 15 and 17; $P < 0.05$), and these genes may contribute to the adaptation of species to specific habitats (Wu et al., 2018).

We performed pan-genome analysis to identify structural variations (SVs) in *Sesamum* species, using *S. indicum* and *S. latifolium* as reference genomes (Figure 3D–3F; Source Data 4 for Figure 3). The number of SVs relative to cultivated sesame in the 5 diploid *Sesamum* species varied from 44 698 (*S. alatum* vs. *S. indicum*) to 50 063 (*S. latifolium* vs. *S. indicum*), with total SV lengths of 48.88 and 43.66 Mb, respectively. Overall, 202 685 putative genes were detected in 25 401 putative gene families in *Sesamum* (Figure 3E; Source Data 5 for Figure 3). Among these gene families, 9164 (36.1%) were classified as core families, 14 292 (56.3%) as dispensable, and 1945 (7.7%) as private. *Sesamum* exhibited a lower percentage of conserved collinear gene families than did in *Glycine* (50.1%) (Liu et al., 2020b) and *Gossypium* (61.8%) (Li et al., 2021a). These results reflect the evolution of *Sesamum* across an extended period.

Pfam enrichment and Gene Ontology (GO) analyses demonstrated that core genes in sesame were enriched in various biological processes, cellular components, and molecular functions ($P < 0.05$). Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway analyses also revealed enrichment of core genes in *Sesamum* within metabolic pathways such as "Biosynthesis of amino acids" (map01230), "Carbon metabolism" (map01200), "FA metabolism" (map00564), and "FA biosynthesis" (map00061) (supplemental information; Supplemental Figure 10; $P < 0.05$). The FA gene enrichment and high FA biosynthesis and metabolism in *Sesamum* demonstrate the long-term genomic evolution of this genus. By contrast, the private genes in all seven *Sesamum* genomes were enriched mainly in ribosomes, oxidative phosphorylation, DNA replication, and small molecule biosynthesis and metabolism (supplemental information), which may have conferred the genus-specific and conserved characteristics of *Sesamum* in response to environmental change (Wei et al., 2019; Miao et al., 2021b).

SVs associated with plant architecture

Plant architecture (uniculm or branching) is an important trait that affects seed yield and cultivation models in sesame and other crops (Mei et al., 2017). Branching type is a key agronomic trait for *Sesamum*, with branch numbers ranging from 1 to 18 (Supplemental Table 2). We performed a genome-wide association study (GWAS) on branch number traits in 560 core sesame accessions (supplemental information; Supplemental Figure 11) across two environments (Es; 2017–2018) using an 896 745-variant matrix with a mixed linear model (MLM) program (Figure 4A). Overall, SNP_3364138 in *Si*Chr.10, with the lowest *P* value of 6.58E−20, was determined to be the target variant significantly associated with plant architecture traits in sesame. SNP_3364138 was located in Sindi_2199200 (*SiPT1*; GenBank: MT362742), which encodes a Centroradialis-like protein 2 isoform X1 (CEN-like 2). For the allelic gene *Sipt1* in the uniculm genotype, insertion of a Copia long terminal repeat (LTR) retrotransposon fragment (4166 bp) resulted in a frameshift of 22 amino acids and formation of a premature terminator at 1261 bp, which subsequently effected gene function (Figure 4B). *SiPT1* overexpression in T_2 transgenic *Arabidopsis* resulted in the formation of multiple branches on the stem branch and a significant delay in flowering time (>20 d; Figure 4C). Homolog analysis indicated that *SiPT1* belongs to the Terminal flower1 (TFL) like subfamily of the phosphatidylethanolamine-binding protein (*PEBP*) gene family in flowering plants (Figure 4D), which are ancestral genes (homologous to protogene P01033829001) (Murat et al., 2017). Furthermore, the SiPT1 and PT1 homologs were conserved in *Sesamum* (resemblance coefficient $\geq$ 93.5%; $P < 1E{-}50$; Supplemental Figure 12).

Interestingly, we found that the DT1 subfamily, which controls inflorescence determination in sesame (Zhang et al., 2016), belongs to the TFL-like subfamily (Figure 4E). Sesame exhibits indeterminate growth, and there are up to 63 stem capsule internodes in cultivated sesame (Miao et al., 2020a). *Sidt1-1*, a SNP_{397} mutation (*dt1*) in exon 2 of *Sidt1* (GenBank: KU240042), resulted in a S79N amino acid substitution, leading to determinate inflorescence growth (Figure 4E). Because deletion of a 24.9-kb SV sequence occurred in *DT1*, the mutant *Sidt1-2* exhibited entirely determinate characteristics, with only 3–5 capsule nodes (Figure 4E; Zhang et al., 2016). These results demonstrated the conservation and specific function of the PEPB gene family in plant-architecture and inflorescence-development traits.

SVs related to *Fusarium* wilt resistance

Fusarium wilt, caused by *Fusarium oxysporum* (FO), is a major fungal disease that affects sesame and hundreds of other crops (Ma et al., 2013; Qiu et al., 2014). In contrast to the highly resistant (HR) *S. calycinum*, *S. latifolium*, *S. angolense*, and *S. radiatum* species, cultivated sesame (var. Yuzhi 11) and *S. angustifolium* are highly susceptible (HS) to highly pathogenic strains of FO f. sp. *sesami* (FOS) (Supplemental Figure 13A; Duan et al., 2020). Proteins involved in disease resistance responses (encoded by R genes) participate in the recognition of nonspecific transmembrane pattern recognition receptors and pathogen effectors (Yue et al., 2012). Therefore, we screened the *Sesamum* genomes against the plant resistance gene database (PRGdb 3.0) (Osuna-Cruz et al., 2018) to identify R genes ($P < 1E{-}6$; Source Data 1 for Figure 1; Supplemental Table 14). As expected, the cultivated sesame genome harbored fewer R genes (1198, comprising 14 types) than the 5 diploid wild *Sesamum* genomes, which had R gene counts ranging from 1457 (*S. alatum*) to 1570 (*S. latifolium*) (Supplemental Table 14). The ratios of nucleotide-binding subdomain (NBS) pathogen effector recognition genes in *Sesamum* species, which ranged from 16.69% in cultivated sesame to 23.75% in *S. alatum*, were similar to those observed in 11 crop species analyzed in this study (from 12.56% in *Linum usitatissimum* to 27.86% in *V. vinifera*), most of which were susceptible to FO infection.

Similar to the basal eudicot species *A. coerulea*, sesame contains a gene rarely observed in angiosperms, *CNLK* (Osuna-Cruz et al., 2018). However, *Sesamum* species lack Toll-interleukin-1 receptor

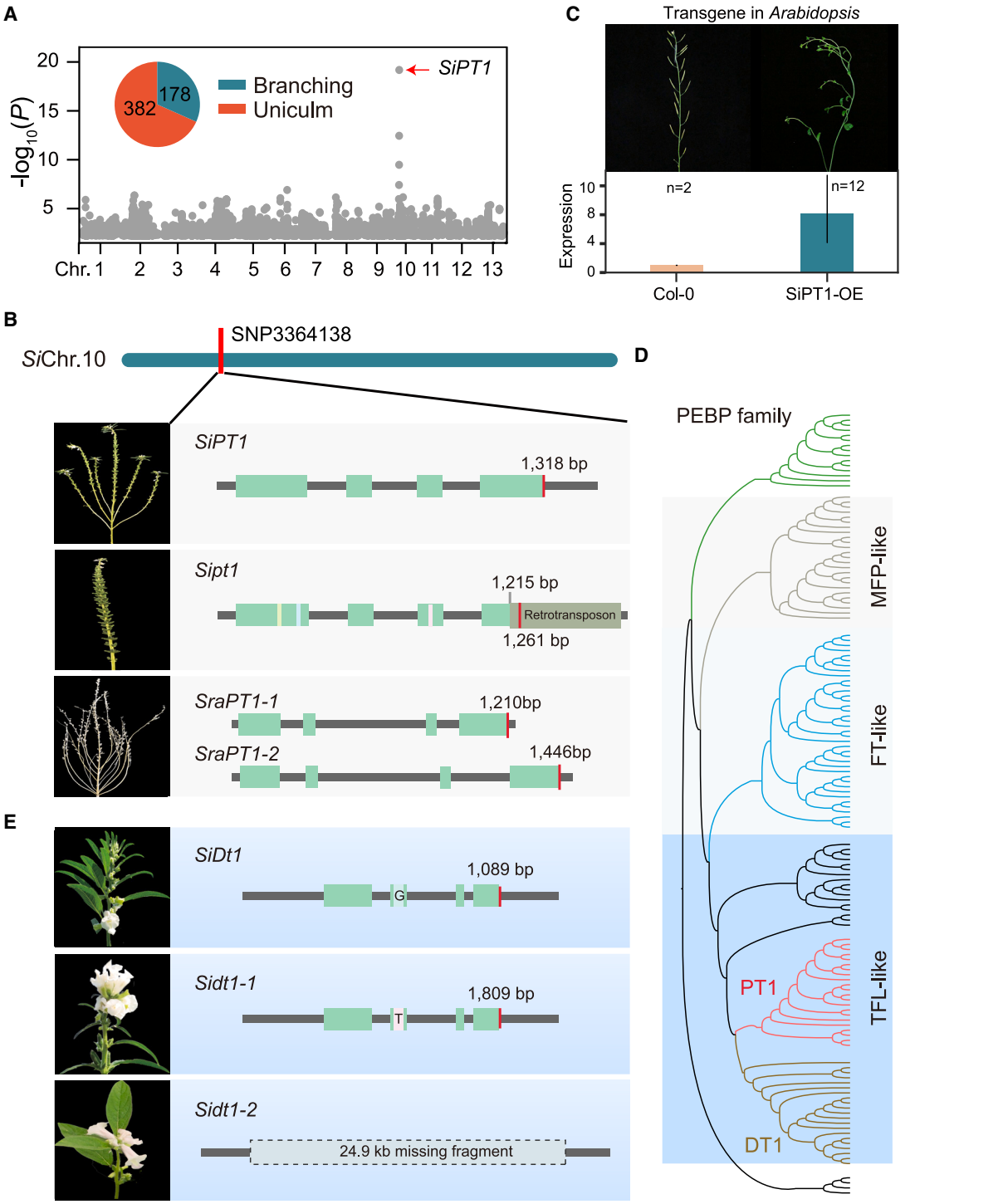

Figure 4. The PEBP gene family regulates plant architecture and inflorescence development in *Sesamum*.

(A) GWAS analysis of the branch number trait in 560 sesame accessions. The target gene *SiPT1* (red arrow) was identified on *Si*Chr.10.

(B) Gene structure and evolution of *SiPT1* and orthologous genes in sesame and *S. radiatum*. The target SNP (SNP_3364138) associated with branch number is present in the target gene *SiPT1*. Compared with *SiPT1* in branched accessions, *Sipt1* in uniculm accessions contains a Copia LTR retrotransposon fragment insertion at 1261 bp. *S. radiatum* contains two PT1 genes, *SraPT1-1* and *SraPT1-2*, with conserved sequences.

(C) *Arabidopsis* overexpressing *SiPT1* exhibits a higher branch number (up to 12) and delayed flowering time.

(D) Phylogenetic tree of the PEBP family in *Sesamum*. The PEBP family comprises the TFL-like, FT-like, and MFP-like subfamilies. PT1 and DT1 proteins are included in the TFL-like subfamily (blue background).

(E) Polymorphic character of *SiDT1* genes in sesame. *Sidt1-1* carries a SNP in a determinate mutant. *Sidt1-2* is lost in an entirely determinate accession owing to deletion of a 24.9-kb DNA sequence.

(TIR)-like NBS-leucine-rich repeat (LRR), TIR-NBS, coiled-coil TIR-NBS-LRR, and TIR-NBS kinase R genes. Absence of TIR-NBS-containing NBS genes has also been observed in the early-diverging species *A. coerulea*, *Mimulus guttatus*, *Oryza sativa*, and other monocot crops (Collier et al., 2011; Shao et al., 2016). Thus, preservation or loss of specific R-gene types may have occurred during early angiosperm evolution, perhaps reflecting differences in fusion and fission of R-gene domains during their formation over different evolutionary periods (Osuna-Cruz et al., 2018). Additional R genes in *Sesamum* were found to have undergone tandem duplication, resulting in an uneven distribution across chromosomes (supplemental information; Figure 1 and Supplemental Figure 5). A significant contraction of 19 R gene clusters, including *RLK* (receptor-like kinase), *KIN* (kinase), *RLP* (receptor-like transmembrane protein), and NBS-type genes, was observed in sesame and its related branches, including monkey flower and tomato. This reduction may have contributed to variations in resistance to FO and other pathogens (Supplemental Table 15).

Histological observations indicated that the FOS pathogen 8PA73 (HSFO 08027 transformed with the GFP gene) invaded the root epidermal cells of resistant and susceptible *Sesamum* spp. within approximately 24 h (Supplemental Figure 13B). After 6 d of inoculation, FOS was detected in the pith tissues of HS species (*S. indicum* and *S. angustifolium*) but was restricted to the epidermal tissues of HR species (*S. calycinum* and *S. radiatum*). To identify target genes that regulate FOS resistance in *Sesamum*, we performed association mapping using F_2 interspecific crossing populations of *S. calycinum* (0% infected) and *S. angustifolium* (100% infected) that were exposed to the highly pathogenic FOS HSFO10175 (Duan et al., 2020; Figure 5A; supplemental information). Resistance to *Fusarium* wilt disease was determined to be controlled by a single and dominant gene pair in the interspecific crossing population (χ^2 = 0.0, $P < 0.01$) (Supplemental Table 18). Association analysis identified two significantly associated SNPs with the lowest *P* values: SNP1738690 (P = 3.53E−17, R^2 = 69.4%) on *Sca*Chr.16 of *S. calycinum* and SNP131559 (P = 1.56E−14, R^2 = 62.4%) on *Sau*Chr.14 of *S. angustifolium* (Figure 5B). SV screening of the target intervals in *S. calycinum* and *S. angustifolium*, together with analysis of syntenic blocks and presence-absence variations in gene families of 4 HS and HR species, revealed 52 syntenic gene pairs within the target 400-kb intervals (Figure 5C; Source Data 6 for Figure 5). Coupled with comparison of assembled genome sequences from 12 homogenous F_2 interspecific hybrid lines, SV screening revealed a clear association between SV_297658 (−1376 bp in *S. angustifolium*) and FOS resistance (Figure 5C; Supplemental Table 19). Sanger sequencing revealed a 1411-bp Copia LTR retrotransposon insertion at −210 bp within the 5′ untranslated region (5′ UTR) of Sangu_2112300 in *S. angustifolium* (HS). Genome annotation indicated that Sangu_2112300 (named *Saudir32*) encodes a dirigent protein (DIR). In contrast to the high expression level of the allelic gene Scaly_2470000 (named *ScaDIR40*) in *S. calycinum* (HR), *Saudir32* exhibited low expression in root and leaf tissues of *S. angustifolium* (HS) inoculated with FOS (Figure 5D).

Syntenic comparison also revealed the loss of −1 bp to −1400 bp in the 5′-UTR sequence and partial loss of the coding sequence (CDS) in the sesame ortholog Sindi_2089700 (*Sindir37*) (Figure 5E). No transcripts of *Sindir37* (Figure 5G) or the two duplicated *DIR* genes were detected in FOS-inoculated sesame (var. Yuzhi 11). Amplification of *DIR* genes within the tested population validated the consistency between genotypes and phenotypes (Supplemental Figure 14; Supplemental Table 20).

DIR proteins control stereoselective coniferyl alcohol coupling in lignan and lignin formation and also enhance biotic and abiotic resistance during secondary cell wall development in plants (Paniagua et al., 2017; Li et al., 2021b; Liu et al., 2021). The DIR-b/d subfamily enhances resistance to fungal pathogens such as *Fusarium* crown rot pathogens (Yang et al., 2021), *Fusarium solani* (Deng et al., 2022), and *Marssonina brunnea* (Li et al., 2021b). To elucidate the resistance mechanism of *DIR* genes in *Sesamum* spp., we analyzed the structural and evolutionary characteristics of the *DIR* family (Supplemental Figure 15). Overall, gene numbers in this family varied between 40 and 43, and it could be divided into 6 subfamilies (Figure 5F). Phylogenetic analysis revealed that all eight *DIR* genes related to *Fusarium* wilt resistance in *Sesamum* belonged to the *DIR*-b/d family (Supplemental Figure 15A). In addition, Pfam domain comparisons indicated that these DIR family proteins possessed similar structures with a high resemblance coefficient of greater than 64.7% ($P < 1E-15$), excluding *Sindir37* (Supplemental Figure 15C). Within the *DIR*-b/d sesame subfamilies, three genes were adjacent to each other on *Si*Chr.9, highlighting the tandem duplication characteristics of this family (Paniagua et al., 2017; Liu et al., 2021). However, no *DIR* genes were expressed in plantlets inoculated with FOS (data not shown).

Subsequent analysis of 5′-UTR flanking sequences demonstrated that the *DIR* genes contained elements such as antioxidant response element, as-1, MYB binding site, and W-box, as well as various other elements related to stress response, hormone signaling, development, and site binding (Supplemental Figure 15B). Unlike that of *Sindir37* and *Saudir32* in HS species, expression of *ScaDIR40* and *SlaDIR36* was significantly induced in HR species upon infection with highly pathogenic FOS strains (Figure 5G and Supplemental Figure 16A). These results revealed the induction characteristics of *DIR* promoters under biotic and abiotic stress. In the allotetraploid *S. radiatum* (Supplemental Figures 13 and 15), two *DIR* genes, *SraDir83* and *SraDir85*, were highly expressed after FOS inoculation (data not shown), suggesting that the ancestral species of the *Sesamum* genus was HR to *Fusarium* wilt. Although the occurrence of SVs in *DIR* genes may be incidental, they can still influence susceptibility to *Fusarium* wilt. The evolutionary characteristics of the *DIR* family in *Sesamum* species reflect the core and conserved functions of *DIR* genes in regulating resistance to *Fusarium* wilt.

To further examine the responses and regulatory network of *DIR* genes after FOS inoculation, we compared the root transcriptomes of two resistant species (*S. latifolium* and *S. calycinum*) and two susceptible species (*S. angustifolium* and *S. indicum*) inoculated with FOS strain HSFO 08027 for 0–168 h (Source Data Table 2). Overall, high *DIR* gene expression in *Sesamum* was correlated with disease resistance-related genes such as laccase 6, pentatricopeptide repeat 1, and papain family cysteine protease (peptidase) genes, as well as transcription factor (TF) genes such as those

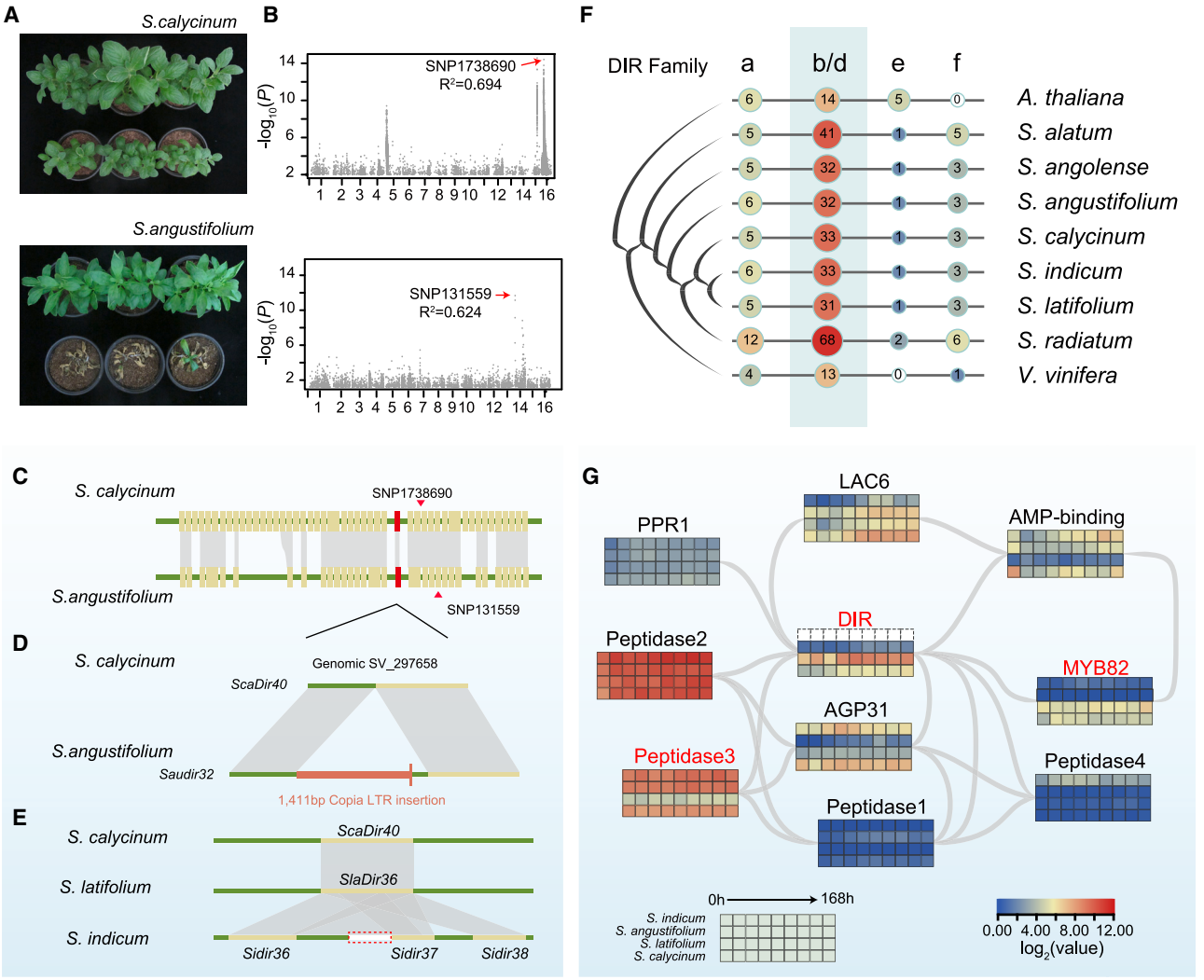

Figure 5. *DIR* genes regulate resistance responses to FO in *Sesamum*.

(A) Phenotype of *Fusarium* wilt resistance in *S. calycinum* (HR) and *S. angustifolium* (HS). Top: controls (top row) and FOS 08027 treatments (bottom row) of *S. calycinum*. Bottom: controls (top row) and FOS08027 treatments (bottom row) of *S. angustifolium*.

(B) GWAS analysis of *Fusarium* wilt disease resistance using an interspecific cross population from *S. calycinum* (KEN8) (HR) (top) and *S. angustifolium* (G01) (HS) (bottom). Red arrows indicate the target SNPs with the lowest *P* values on the chromosome sets of *S. calycinum* and *S. angustifolium*.

(C) Synteny comparison of collinear genes in target intervals associated with *Fusarium* wilt resistance between *S. calycinum* and *S. angustifolium*. Red arrowheads indicate target SNPs with the lowest *P* values. Red bars indicate a gene pair with an SV. Yellow bars indicate gene sequences. Gray lines between genes indicate syntenic relationships.

(D) Structure of the target SV_297658 and *DIR* gene pair in *S. calycinum* and *S. angustifolium*. Segments in green and yellow indicate promoter and CDS sequences of the *DIR* genes. The red segment indicates a 1411-bp Copia LTR sequence inserted in the promoter of *Saudir32*.

(E) Syntenic comparison of the target interval in sesame, *S. calycinum* (HR), and *S. latifolium* (HR). Tandem duplication of 3 *DIR* genes (*Sidir36*, *Sidir37*, *Sidir38*) has occurred on *Si*Chr.9 of sesame.

(F) Phylogenetic tree of the *DIR* family in *Sesamum*, *Arabidopsis*, and *V. vinifera*. *DIR* genes related to *Fusarium* wilt resistance belong to the *DIR*-b/d subfamily (light blue background). There are many more genes in *DIR*-b/d subfamily in *Sesamum* than in *Arabidopsis* and *V. vinifera*.

(G) Regulatory network of *DIR* genes in *Sesamum*. The normalized expression profiles of key genes in two HS (sesame and *S. angustifolium*) and two HR (*S. calycinum* and *S. latifolium*) species inoculated with FOS 08027 for 0–168 h are shown as a heatmap.

encoding MYB82 and AMP-binding proteins. These TFs participate in lignin and lignan biosynthesis (Liu et al., 2020a), stress responses (Shockey and Browse, 2011; Li et al., 2022), and plant–pathogen interaction pathways (Park et al., 2014; Liu et al., 2016b) (Figure 5G; Source Data 10 for Figure 7). Weighted correlation network analysis of the transcriptome of the HR species *S. latifolium* identified *MYB82* (KO9422) as a node gene that was upregulated and co-expressed with 673 R genes in roots after FOS inoculation (Bisque 4 model) (Figure 5G and Supplemental Figure 16B and 16C). By contrast, no transcripts of *MYB82* were detected in HS sesame and *S. angustifolium*; however, the peptidase3 (KO1373) gene was highly expressed after 12-h inoculation and conferred high susceptibility to pathogen infection (Figure 5G). Therefore, we postulate that deletion of the 5′ UTR and partial CDS sequence of *SiDIR* may be essential for the high susceptibility of cultivated sesame to highly pathogenic FOS

strains. Early retention with later disruption or loss of key atypical pattern recognition receptor genes is a common evolutionary mechanism that gives rise to variation in crop pathogen resistance. Hence, a promising strategy for enhancing *Fusarium* wilt disease resistance could involve transfer of *DIR* genes through interspecific hybridization or genetic modification methods or regulation of key node genes such as *MYB82*, peptidase3, and crucial R genes via artificial stimulation.

Genome evolution related to high OC regulation

The high OC of sesame seeds was greater than 50%, whereas that of the 6 wild *Sesamum* species ranged from 29.46% (*S. angustifolium*) to 46.99% (*S. radiatum*) after removal of crude fiber (Supplemental Table 21). Compared with soybean (∼20%) (Gai, 2003), cotton (∼30%) (Liu et al., 2016a), sunflower (30%–47%) (Li et al., 2017), flax (∼40%) (Li et al., 2019), and peanut (50.6%) (Chen et al., 2012), the wild *Sesamum* species exhibited a relatively high OC.

To investigate the evolution of FA biosynthesis and metabolism genes in *Sesamum* and to understand the reasons for the high OC of sesame, we screened the genomes of *Sesamum* species against a reference dataset of FA genes related to acyl lipid metabolism ($P < 1E{-}5$; protein identity > 60%) (Figure 6A; Source Data 3 for Figure 2; Supplemental Tables 16 and 17). The percentage of FA genes in the cultivated sesame genome (2.15%) was similar to that in wild *Sesamum* (from 1.66% in *S. latifolium* to 2.54% in *S. calycinum*) and other oilseed crops (soybean, 2.82%; sunflower, 2.75%; castor bean, 1.92%). However, variations in the copy numbers of key FA genes were observed among the oilseed crops (Supplemental Table 16; Li et al., 2017, 2019). For example, the number of stearoyl-ACP desaturase (SAD) genes (7–11) in diploid *Sesamum* species (C18:1, 35.5%–42.1%) exceeded that in diploid oilseed crops with high linoleic (C18:2) or linolenic (C18:3) acid content, including soybean (C18:2, 47%; 5 SAD genes), sunflower (C18:2, 60.78%; 2 SAD genes), flax (C18:2, 12.8%; C18:3, 39.5%; 4 SAD genes), and castor bean (ricinolic acid, C18:1, 80%–88%; 4 SAD genes) (Figure 6A; Source Data 2 for Figure 1).

Cluster analysis of orthologous and paralogous FA genes revealed 5 significantly expanded and 7 contracted FA gene families ($P < 0.05$) in the *Sesamum* genus (Supplemental Table 17).For example, the carboxyltransferase beta subunit of heteromeric ACCase (β-CT) (Enzyme Commission: 6.4.1.2 and 6.3.4.14; Figure 6A) is the first rate-limiting enzyme in FA synthesis and participates in the formation of malonyl-coenzyme A (Roesler et al., 1997). The gene number of the β-CT gene family (Cluster1654) in cultivated sesame (eight) was higher than that in five other oilseed crops and wild *Sesamum* species, likely enhancing the initiation of FA synthesis and increasing seed OC (Ohlrogge and Jaworski, 1997; Unver et al., 2017). Correspondingly, the expression level of ACCase genes was higher in sesame than in the wild species (Figure 6A). In addition, homeodomain leucine zipper IV TFs regulate the expression of *GL2* (*GLABRA2*), a negative regulator of seed oil accumulation (Shen et al., 2006; Dong et al., 2015; Lung et al., 2018). Sesame possessed 10 homeodomain leucine zipper IV TFs, fewer than the closely related species monkey flower and tomato (Supplemental Table 17).

To understand the development of the high OC in cultivated sesame compared with wild species, we evaluated the effect of selection pressure. However, comparisons of the ratio of nonsynonymous substitution/synonymous substitution (Ka/Ks) indicated that FA gene families were not under positive selection in sesame (Ka/Ks > 1; Supplemental Table 22), suggesting that the elevated OC likely arose from independent species-specific genomic evolution rather than environmental selection. To assess the impact of breeding for improved OC on FA genes and elucidate the mechanism of OC regulation in sesame, we conducted an analysis of the population structure of 560 sesame accessions, including 97 varieties, 401 Chinese landraces, and 62 international accessions (supplemental information). A total of 894 425 reliable variants, distributed across 13 805 genes with a mean fixation index (Fst) of 0.0022, were retained (Figure 6B and Supplemental Figure 17). Compared with the 463 germplasm accessions (mean nucleotide diversity, π = 0.0023), the mean π value of the 97 Chinese varieties was 0.00186, indicating low genetic diversity. Evidence of artificial selection in these 97 varieties was detected across 32 regions (top 1%, π ratio $\geq$ 2.97), which collectively contained 972 genes (Source Data 7 for Figure 6). These regions were designated candidate artificial selection regions and included four FA (*Si*Chr.3, *SiChr.*5, and *SiChr.*9) and eight R (*Si*Chr.1, *SiChr.*2, *SiChr.*3, *SiChr.*4, *SiChr.*5, and *SiChr.*11) gene blocks (Figure 6B; Source Data 8 for Figure 6). Ultimately, these selected regions may enhance oil metabolism and disease resistance in sesame varieties.

GWAS analysis was then used to identify genes regulating OC. The 560 sesame accessions exhibited a wide OC range, from 30.58% (2018 Pingyu) to 58.67% (2017 Pingyu), reflecting significant phenotypic variation (Supplemental Figure 18). The broad-sense heritability (H^2) of OC was determined to be 83.07%, with minor effects from environment and genotype by environment interaction ($P < 0.01$, ANOVA), indicating that *Sesamum* OC was primarily controlled by genotype. GWAS results further indicated that 10 SNP intervals ($P < 1E{-}5$ and $R^2 > 5\%$) were significantly associated with OC in at least 3 environments (Figure 7A; Source Data 9 for Figure 7). Of the 10 SNP intervals, 5 candidate genes with homozygous SNPs or insertions or deletions (indels) were detected. In particular, *SiNAC1* (Sindi_2309500; *Si*Chr.10) exhibited the highest explanatory value (R^2 up to 43.8%) for OC variation in the sesame population (Supplemental Figure 19B). *SiNAC1* is a NAC domain-containing TF that participates in various plant developmental processes, such as remobilization, shoot branching determination, and stress and defense responses (Puranik et al., 2012). As a TF, SiNAC1 localized exclusively to the nucleus (Supplemental Figure 20B). Structural analysis indicated that a T/G mutation occurred within the conserved N-terminal adhesion motif of *SiNAC1*. In addition, within the 560 sesame accessions, the 2 haplotypes of *SiNAC1* (Sindi_2309500) and *Sinac1* were highly correlated with OC variation (Figure 7B and Supplemental Figure 19B). For example, in 24 haplotype-1 accessions with high OC (average OC = 51.96%), *SiNAC1* was highly expressed primarily in developing seeds (8–20 d after flowering, DAF), as observed in the varieties Shaanzhi 3 and WanzhiC02-3. By contrast, in 23 haplotype-2 accessions with low OC (average OC = 45.16%), *Sinac1* exhibited low expression levels, as observed in the variety Black sesame (Supplemental Figure 23). The C333A mutation in

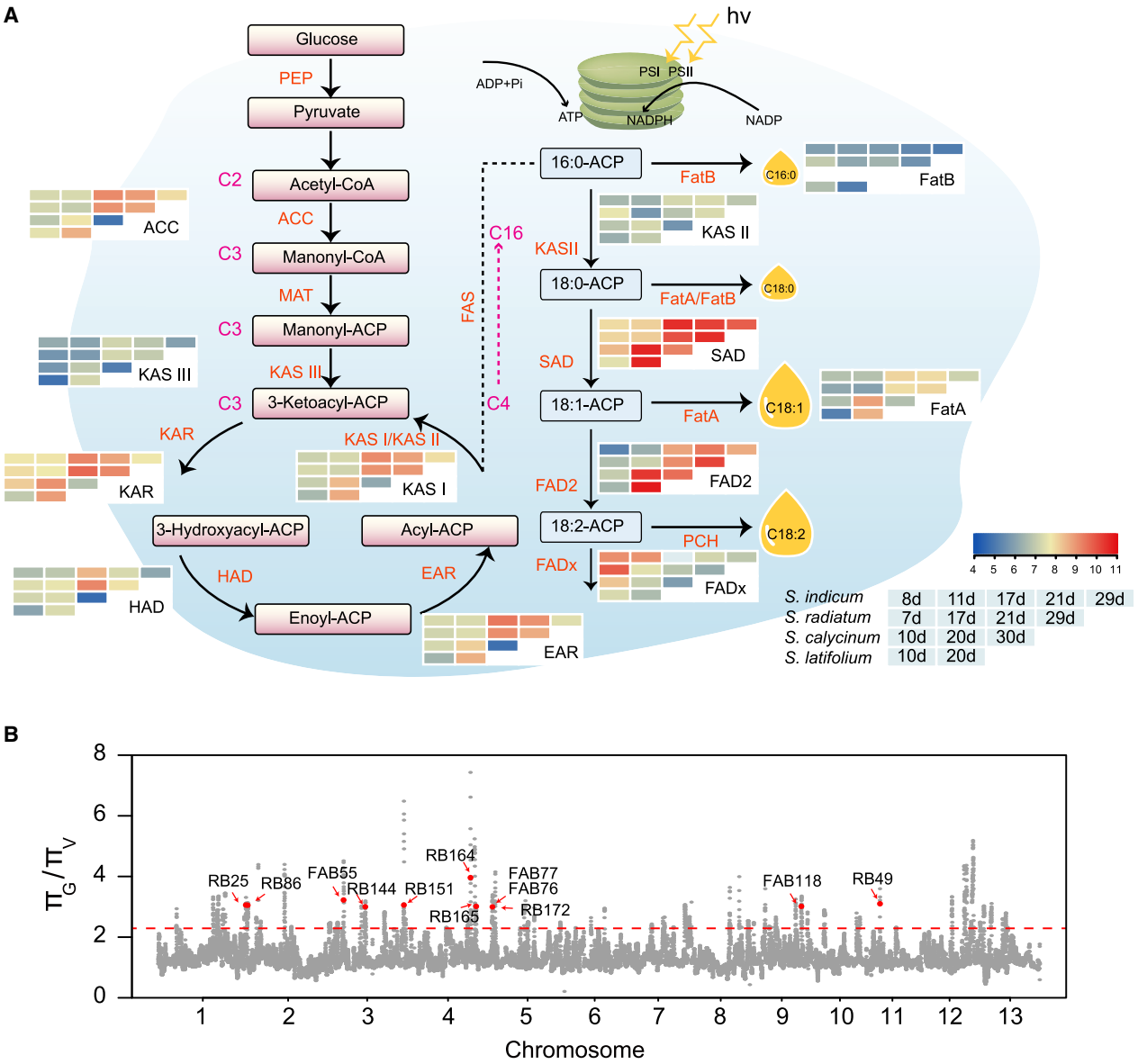

Figure 6. Oil biosynthetic process in sesame and domestication of FA- and R-gene blocks in the sesame population.

(A). Oil biosynthesis in *Sesamum*. Normalized expression profiles of key genes in developing seeds of sesame (var. Yuzhi11) and three wild *Sesamum* species are shown as a heatmap.

(B) Distribution of the π ratio between 97 Chinese sesame varieties and 463 worldwide germplasm accessions, and R-gene and FA-gene blocks in candidate artificial selection regions. Red arrows indicate four FA-gene blocks (FAB55, FAB76, FAB77, and FAB118) and eight R-gene blocks (RB25, RB49, RB86, RB144, RB151, RB164, RB165, and RB172) detected in candidate improvement regions. FAB, FA-gene block; RB, R-gene block.

SiNAC1 led to a significant increase in OC, resulting in a 15.06% boost (supplemental information; Supplemental Table 23). Overexpression of *SiNAC1* in 27 transgenic *Arabidopsis* lines (T_1) resulted in a significant increase in seed OC from 22.39% to 27.97% ($P < 1E{-}5$). We thus confirmed the regulatory function of *SiNAC1* in crop FA biosynthesis for the first time (Figure 7D).

We further analyzed the structural evolution of the *NAC1* family in *Sesamum*. Among all 6 wild species, 7 NAC1 homologs of SiNAC1 were conserved with high similarity (≥92.9%; $P < 1E{-}149$; Figure 7B). The seven *NAC1* orthologs carried the same haplotype as *SiNAC1* at nucleotide T_{333}. In addition, SaoNAC1 (Sango_1566400) in *S. angustifolium* and SalNAC1 (Slati_3919800) in *S. latifolium* displayed sequence deletions within the conserved N-terminal adhesion motif (Supplemental Figure 21). Transcriptome comparison indicated that most of the *NAC1* genes, excluding *SlaNAC1* and *SlaNAC2*, exhibited expression patterns in developing seeds similar to that of *SiNAC1* (Figure 7E). We hypothesize that the *SiNAC1*-dependent regulation of high OC content represents the original genotype of ancestral *Sesamum*. Furthermore, nucleotide mutagenesis and genomic variations occurred in conserved sequences of *NAC1* genes without strong natural selection pressure during *Sesamum*

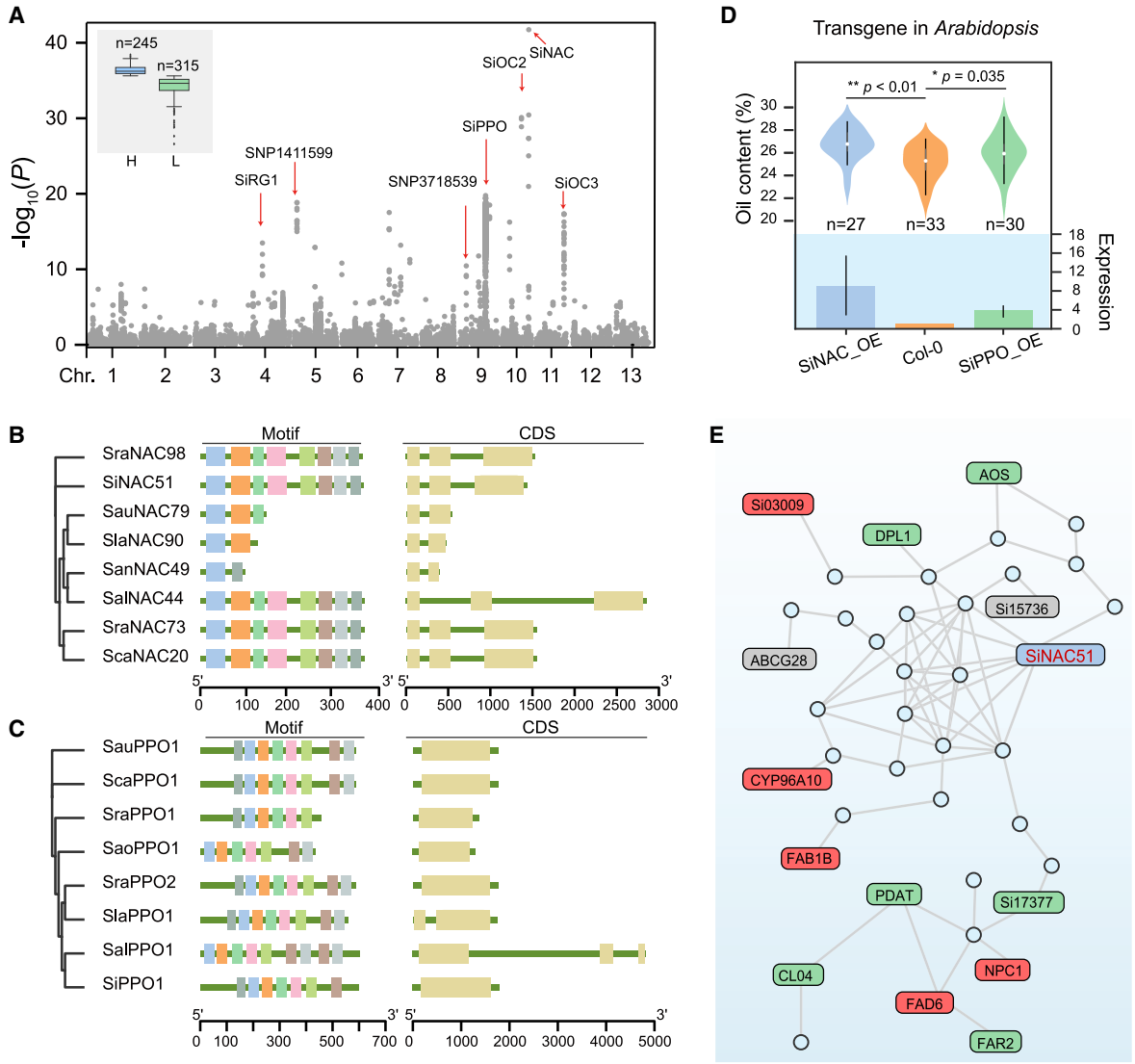

Figure 7. Candidate genes associated with OC in sesame.

(A) Candidate genes significantly associated with OC in eight environments. Of 560 accessions, 245 have high OC (average OC ≥ 50%) and 315 have low OC (average OC < 50%). –log*P* values of the SNP sites and candidate genes in the Sanya (2016) environment are shown.

(B) Phylogenic tree and structural comparison of NAC1 orthologs in *Sesamum*. Colored blocks indicate motifs in the promoter sequences of *NAC1* orthologs. Yellow blocks indicate gene CDSs.

(C) Phylogenic tree and structural comparison of PPO orthologs in *Sesamum*. Colored blocks indicate motifs in the promoter sequences of *PPO*. Yellow blocks indicate gene CDSs.

(D) OC comparison of transgenic *Arabidopsis* with *SiNAC1* and *SiPPO* overexpression vectors, and exogenous gene expression in developing seeds. Blue and green blocks indicate 27 transgenic *Arabidopsis* lines with *SiNAC1* and 30 transgenic *Arabidopsis* lines with *SiPPO*. Thirty-three wild-type plants were used as controls.

(E) Regulatory network related to SiNAC1 in high- and low-OC sesame accessions. Genes in red blocks are highly expressed in both high- and low-OC accessions. Genes in green and gray blocks are highly expressed in high-OC accessions and low-OC accessions, respectively.

evolution, reflecting the divergence of sesame with high OC from wild *Sesamum* species.

To further explore the association between *NAC1* and high OC, we compared the expression profiles of 59 groups of transcriptomes from developing seeds, including 2 high-OC accessions and 1 low-OC accession. We then analyzed the protein–protein interaction (PPI) network related to SiNAC1 in cultivated sesame (Figure 7E; Source Data 11 for Figure 7). In the PPI network, *SiNAC1* was coexpressed with *FAD6* (Sindi_1772100, KO10255), *NPC1* (Sindi_0852400, KO01114), *CYP96A10* (Sindi_2506800), Sindi_0300900, and *FAB1B* (Sindi_2220700, KO00921), all of which participate in unsaturated FA biosynthesis, phospholipid catabolism, metabolic regulation, oxidative stress response, and phosphatidylinositol phosphate biosynthesis. In particular, six genes (Figure 7E), including *FAR2* (Sindi_1809500, KO13356), *DPL1* (Sindi_0172700, KO01634), and *PDAT* (Sindi_1214000, KO00679), were involved in FA biosynthesis-related processes, and they were upregulated in the high-OC accessions. Conversely, Sindi_2186100 and Sindi_1573600 exhibited high expression in

low-OC accessions (Source Data 11 for Figure 7). These results suggested that the *NAC1* TFs play a significant role in the regulation of FA biosynthesis in developing seeds.

We also detected an association between *SiPPO*, which encodes polyphenol oxidase I (PPO; Sindi_2087700; *Si*Chr.9) and OC (with R^2 up to 17.66%) (Supplemental Figure 19A). *PPO* catalyzes the oxidation of o-diphenols to o-diquinones as well as the o-hydroxylation of monophenols (Vaughn and Duke, 1984; Demeke et al., 2001). Moreover, PPO is associated with OC and seed color in sesame (Wei et al., 2015). Among the 560 sesame accessions, *SiPPO* differed from the allelic gene *Sippo* because of a 243–245 bp deletion of the nucleotide TTT, which resulted in deletion of a phenylalanine (F_{82}) in low-OC sesame accessions (Supplemental Figure 19A). The allelic gene *Sippo* was highly expressed in developing seeds (15–20 DAF) of low-OC accessions such as var. Black sesame (Supplemental Figure 23). Furthermore, overexpression of *SiPPO* in all 6 transgenic *Arabidopsis* lines (T_1) significantly increased the OC (average OC = 25.41%, P = 0.035; Figure 7C).

To explore the regulatory function of the *PPO* family in oil biosynthesis, we analyzed the structural evolution of the *PPO* family in *Sesamum* using the Pfam database. Compared with those of 5 other plants, sesame PPO proteins were more closely related to monkey flower, with a resemblance coefficient of 48.5% or greater ($P < 5E{-}140$; Figure 7C). In *Sesamum*, all eight PPO orthologs had the same haplotype (deleted F_{82}) as Sippo (low OC), and most possessed PPO1_DWL, KFDV, and tyrosinase domains (Figure 7D and Supplemental Figure 22). Therefore, we postulated that the *SiPPO* haplotype with a high OC resulted from a mutation during modern sesame evolution. The high conservation of FA biosynthesis-related gene families in *Sesamum* explains the high OC commonly observed in progenitors of the *Sesamum* genus. Furthermore, nucleotide mutations in key genes, such as *SiNAC1* and *SiPPO*, represents an infrequent but feasible evolutionary mechanism that has affected the regulatory network of oil biosynthesis in modern cultivated sesame.

In summary, we successfully assembled chromosome-scale genomes for cultivated sesame and six wild *Sesamum* species. Our analysis revealed that the *Sesamum* genus diverged early, around 48.5–19.7 mya, during the Tertiary period. The evolutionary route of *Sesamum* species involved a transition from n = 13 to n = 16 chromosomes, and allotetraploidization occurred within the *Sesamum* genus. In addition, we observed SVs and PAVs in key gene families of *Sesamum*, highlighting the genomic evolutionary characteristics of sesame. Moreover, 32 artificial selection regions covering 4 FA- and 8 R-gene blocks were detected in Chinese sesame varieties. Notably, we identified four key genes (*PT1*, *DIR*, and *NAC1* and *PPO*) that controlled flowering and plant architecture, resistance to *Fusarium* wilt, and OC traits, respectively. These genomic resources and findings provide a robust foundation for future advances in the molecular breeding of sesame and other oilseed crops.

METHODS

Plant materials

The genome of the well-known Chinese sesame variety Yuzhi 11 (2n = 26) was sequenced as a reference genome, along with those of 6 wild *Sesamum* species (*S. alatum* [var. 3651, 2n = 26], *S. latifolium* [var. KEN1, 2n = 32], *S. angolense* [var. K16, 2n = 32], *S. calycinum* [var. KEN8, 2n = 32], *S. angustifolium* [var. G01, 2n = 32], and *S. radiatum* [var. G02, 2n = 64]) (Supplemental Figure 1; Supplemental Table 2). A core population of 560 sesame accessions (401 Chinese landraces, 97 Chinese varieties, and 62 international accessions from 20 countries) (Source Data Table 4) was used for genome re-sequencing and OC analysis. All materials are publicly available from the sesame germplasm collection at the Henan Sesame Research Center, Henan Academy of Agricultural Sciences.

Sesame genome sequencing

The sesame reference genome (BioProject PRJNA315784; v.3.0) was sequenced using a complex genome sequencing strategy (Supplemental Figure 2). In brief, DNA extracted from young leaf tissue was used for short-insert (~500 bp) and long-insert (~3 kb) paired-end (PE) library construction and sequencing according to the standard Illumina protocol (Illumina, USA). To increase the sequencing read length, ~8-kb and 20-kb PE tag libraries were prepared and sequenced using a Genome Sequencer FLX instrument (Roche Applied Science, Germany). BAC-end sequencing of 42 240 BAC clones from a sesame BAC library (CopyControl pCC1BAC vector) (Zhao et al., 2018) was performed by Sanger sequencing (ABI 3730, USA). A 20-kb SMRTbell library was constructed and sequenced using Pacific Biosciences RS II sequencing technology (USA). Sequencing reads were processed using Trimmomatic (v.0.33) (Bolger et al., 2014); after trimming, reads shorter than 50 nt were removed. High-throughput chromosome conformation capture (Hi-C) libraries were constructed and sequenced using the Illumina HiSeq 2500 platform. The constructed sesame genome is shown in Supplemental Figures 2 and 3.

High-density cytogenetic map construction

Hundreds of BACs with single stable hybridization signal pairs were screened from the sesame BAC library (Zhao et al., 2018). The consecutive BAC-FISH technique was used to group 210 BAC probes into 13 chromosome pairs. The hybridization signal positions of the BAC probes were calculated using more than 3 replicates (Fonsêca et al., 2010). The apex of the short arm of the chromosome was designated 0.00, and the end of the long arm was designated 1.00. The ultra-dense sesame SNP genetic map was integrated with a high-density cytogenetic map and used to guide genome assembly (Zhang et al., 2016).

Reference genome assembly

The strategy used for sesame genome assembly is outlined in Supplemental Figure 2. In brief, *de novo* assembly of the PacBio reads was performed using continuous long reads following the hierarchical genome assembly process (HGAP) workflow (PacBioDevNet). The PacBio Rs_PreAssembler.1 module was used to correct errors in the raw data using the following criteria: default minimum subread length $\geq$ 500 bp, minimum read quality $\geq$ 0.80, and minimum subread length $\geq$ 7500 bp. The reads from MiSeq platform were mapped to polish the assembly sequence from HGAP using Burrows–Wheeler Aligner (0.7.12) (Li and Durbin, 2009). SNPs and indels were called and corrected using SAMtools (Li et al., 2009) and an in-house script. Scaffolding and gap filling were performed on meta-paired Illumina and Roche 454 sequencing data using SSPACE 3.0 with default parameter values of x: 1 -m 50 -o 10 -z 200 -p 1. The PacBio-only and hybrid assembly methods were used to determine the best assembly strategy. Super scaffolds were grouped into contigs based on the SNP genetic map and BAC-FISH cytogenetic map. Hi-C data were mapped to the assembled genome sequences using HiC-Pro (Servant et al., 2015) and 3D-DNA (Dudchenko et al., 2017). Thirteen pseudomolecules were constructed using a Hi-C map.

Sequencing and assembly of wild *Sesamum* genomes

The genome sizes of the six wild *Sesamum* species were estimated by k-mer depth using Jellyfish (Marcais and Kingsford, 2011) and GenomeScope (Vurture et al., 2017). The genomes of these species were sequenced and

assembled using the strategy outlined in Supplemental Figure 4 (supplemental information; BioProject PRJNA625646). High-quality DNA was extracted and used to construct the 500-bp PE, ~20-kb SMRTbell, and Hi-C libraries.

High-molecular-weight DNA was isolated from *S. latifolium* and *S. radiatum* and labeled according to standard BioNano protocols using the single-stranded nicking endonuclease *Afl*II. Labeled DNA was imaged automatically using the Irys system (BioNano, USA) before *de novo* assembly into consensus physical maps using BioNano IrysView. Conflicts between the BioNano maps and the *de novo* assembly were manually resolved using BioNano Access.

Two interspecific genetic maps were constructed to evaluate the quality of the assembled *S. calycinum* and *S. angustifolium* genomes (supplemental information). In total, 126 F_2 individuals were randomly selected from the F_2 progeny and re-sequenced using the Illumina sequencing platform (Source Data Table 1). Putative SNPs were called using SAMtools with in-house script criteria: minimum read depth at a given position, 8; minimum supporting reads for an allele, 2; allele representing $\geq$ 20% of all observed alleles. Interspecific SNPs with allelic polymorphisms were identified using a custom script. JoinMap v.4.0 was used for marker segregation analyses and linkage map construction (Van Ooijen, 2011; Wang et al., 2015a; Santos et al., 2017). All linkage groups with more than 20 SNP markers were grouped into the genetic maps. The integrity of pseudomolecules with the linkage groups was determined using a custom script.

Genome annotation

Protein-coding regions and genes were predicted using a combination of *ab initio*, homology-based, and transcriptome-based prediction methods (supplemental information). *Ab initio* prediction of coding genes was performed with Augustus (v.2.5.5) (Stanke et al., 2006), GlimmerHMM (v.3.0.1) (Majoros et al., 2004), and SNAP15 (Korf, 2004). Homology-based prediction aligned the assembled *Sesamum* genomes to homologous proteins from the whole-genome sequences of *Arabidopsis thaliana*, millet, grapevine, castor bean, tomato, and potato (https://jgi.doe.gov/). GeneWise (v.2.2.0) (Stein, 2013) was used to generate gene structures based on homology alignments. Transcriptome-based prediction involved the generation of transcriptome data for the *Sesamum* species using the Illumina HiSeq 2500 platform; these were then mapped to the genome assembly using TopHat (v.2.0.8) (Kim et al., 2013). Cufflinks (v.2.1.1) (http://cufflinks.cbcb.umd.edu/) was used to identify spliced transcripts in the gene models. All evidence predicted from these approaches was integrated using EVidenceModeler (Haas et al., 2008), which gave a weighted and non-redundant (NR) consensus of gene structures. Gene models were generated and filtered according to the following criteria: coding region length $\leq$ 150 bp; exclusively *ab initio* methods; The fragments per kilobase of transcript per million mapped reads (FPKM) values < 5; hits with Uniref 90 database. Unigenes were searched against the NCBI NR, SwissProt, and Pfam databases using BLASTP with a cutoff E-value of 1E−5. Criteria included the following: identity $\geq$ 0.25; minimum alignment length 100 bp. InterProScan (Jones et al., 2014) was used to annotate motifs and domains in the gene sequences by comparison with publicly available databases. GO information for each gene code was extracted from the InterProScan results using BLAST2GO (Conesa et al., 2005). Putative gene pathways were derived by matching genes with the KEGG database (Kanehisa et al., 2004). Benchmarking Universal Single-Copy Orthologs (Waterhouse et al., 2018) and the core eukaryotic genes mapping approach (Parra et al., 2007) were used to evaluate the completeness of the *Sesamum* genomes.

Interspersed repeat sequences

RepeatModeler (http://www.repeatmasker.org) was used for *de novo* identification. RepeatMasker (v.4.0.6) was used to identify interspersed repetitive sequences by aligning the *Sesamum* genome sequences against the RepBase database (v.20150807) (Tarailo-Graovac and Chen, 2009) and the *de novo* database.

Telomeric repeats and rDNA loci

FISH of telomeres and rDNA repeats was performed as described previously (Chen et al., 2015a, 2015b; Zhao et al., 2018). Telomere and rDNA repeat probes were designed according to the universal plant telomeric repeat probe (5′-TTTAGGGTTTAGGGTTTAGGG-3′) and conserved sequences in *Arabidopsis* spp., respectively (Chen et al., 2015a, 2015b; Fuchs et al., 1995; Yang et al., 2017). Idiograms of 45S and 5S rDNA distribution patterns in the *Sesamum* chromosomes were constructed according to previously described methods (Zhang et al., 2012).

Centromeric repeats

Twenty-seven BACs with specific BAC-FISH hybridization signals at the centromeric and pericentromeric regions of each sesame chromosome pair were used to screen centromeric repeats (supplemental information; Supplemental Table 2). LTR_FINDER (Xu and Wang, 2007) and RepeatMasker (v.4.0.6) were used to detect and annotate LTR retrotransposon sequences. Specific repeats within the LTR retrotransposon sequences were aligned using MAFFT v.7.245 (Katoh et al., 2002; Katoh and Standley, 2013). Centromeric repeat sequences from wild *Sesamum* species were screened on the basis of the sequence alignment (Wang et al., 2015a) with identity $\geq$ 90% and E-value $\leq$ 1E−5. A phylogenetic tree of centromeric repeats from *Sesamum* species, *Zea mays* (154 bp) (Ananiev et al., 1998), and *S. tuberosum* (186 bp) (Tek and Jiang, 2004) was constructed using ClustalW2.

WGD, WGT-γ, and tetraploidization

Ks and four-fold degenerate site (4DTv) values for each *Sesamum* genome were calculated using the PAML software (Yang, 1997) and an in-house script. Grapevine and tomato were used to determine the WGD and WGT-γ events in *Sesamum*. To determine the effects of selection pressure on *Sesamum*, we estimated the Ka/Ks for each single-copy ortholog using Codeml with a free-ratio model. Pairwise whole-genome alignments were performed for the *Sesamum* genomes using MUMmer (v.3.23) (Kurtz et al., 2004) with default parameters.

To validate the relationships of different chromosome karyotypes in *Sesamum*, pairwise GISH and centromeric repeat FISH were performed on the seven *Sesamum* species using a consecutive FISH technique (Zhao et al., 2018).

Sesamum genome group nomenclature

To explore the characteristics of genome evolution in the *Sesamum* genus, we divided the seven assembled genomes into four genome groups according to their genome structures. Data from 4DTv analysis, centromere repeat hybridization signals, and GISH were used to differentiate the 2 subgenomes (AA_{Sra} and BB_{Sra}) of allotetraploid *S. radiatum* (2n = 4x = 64). Considering the high degree of incompatibility in interspecies hybridization and the weak GISH signals, the *S. indicum* (2n = 26) and *S. alatum* (2n = 26) genomes were designated CC and DD, respectively.

Phylogenetic position within the angiosperms

To elucidate the evolution of sesame within angiosperms, we aligned the genomes of sesame and the 5 diploid *Sesamum* species (*S. alatum*, *S. latifolium*, *S. angolense*, *S. calycinum*, and *S. angustifolium*) with those of 15 other plant species using BLASTP. The 645 single-copy orthologous genes identified were used to construct a phylogenetic tree (supplemental information). Divergence times of key branches between *A. thaliana* and *Solanum lycopersicum* (101.2–123.9 mya), *S. lycopersicum* and *M. guttatus* (70–90 mya), and *M. guttatus* and *V. vinifera* (110–140 mya) were determined based on the NODE TIME dataset (http://www.timetree.org). The estimated divergence times of all species were calculated using r8s software (https://sourceforge.net/projects/r8s/).

Syntenic blocks of *Sesamum* with post γ-AEK

The 9022 ancestral orthologs in the 21 paleochromosomes of the ancestral eudicot pseudochromosomes post-γ were used for the *Sesamum* ancestor analysis (Murat et al., 2017). Genome synteny analysis of the *Sesamum* species and grapevine (Jaillon et al., 2007) was performed using MCScanX (Wang et al., 2012) with the following parameters: MATCH_SCORE: 50, MATCH_SIZE: 20, GAP_PENALTY: -1, OVERLAP_WINDOW: 5, E_VALUE: 1e-05 and MAX GAPS: 25. Dot plot diagrams were produced using the Multi-Genome Synteny Viewer (v.2.1).

Four high-quality genomes (sesame, *S. latifolium*, *S. calycinum*, and *S. angustifolium*) were used to identify the ancestral pseudochromosomes of *Sesamum*. Collinear orthologous gene syntenic blocks (≥20 genes per block; $P < 1E{-}20$) shared among the 4 genomes were identified using CoGe (https://genomevolution.org/coge/). Translocation and rearrangement of the homologous chromosome fractions between genomes was determined using in-house scripts with the following parameters: ≥1.5 Mb, or >100 common ancestral orthologous genes per block.

Sub-genomes in tetraploid species

Syntenic block analysis of *S. radiatum* with the diploid species *S. latifolium* (AA_{Sla}) and *S. calycinum* (BB_{Sca}) was performed using MCScanX (Wang et al., 2012). Paralogous genes and duplication blocks between *S. radiatum* chromosomes were screened using BLASTP and MCScanX. The centromeric repeats *Sra*Cen1 and *Sra*Cen2 were identified in the assembled genome using BLASTN. The correspondence of homologous pairs between *S. radiatum* and *S. latifolium* (AA_{Sla}) and *S. calycinum* (BB_{Sca}), as well as the syntenic relationships within the two sub-genomes of *S. radiatum*, were visualized using VGSC Toolkit (Xu et al., 2016).

Orthologous gene clusters in *Sesamum*

Orthologous gene clusters between cultivated sesame and wild *Sesamum* species and 15 crop species were identified using OrthoFinder (v.2.3.3) (Emms and Kelly, 2019). The recommended settings were used for all-against-all BLASTP comparisons (v.2.3.05) and OrthoMCL analyses. In-house Perl scripts were used to process OrthoMCL outputs; data were visualized using InteractiVenn (http://www.interactivenn.net/). To gain further insight into the evolutionary dynamics of the *Sesamum* genes, expansion and contraction of gene families were analyzed using CAFE software (http://sourceforge.net/projects/cafehahnlab/). A phylogenetic tree based on the JTT matrix-based model was constructed using the maximum likelihood method with RaxML (Kozlov et al., 2019). GO and KEGG analyses of the protein-coding genes were performed using Omicshare CloudTools (http://www.omicshare.com/).

SV identification

Nucmer (MUMmer v.4.0, standard parameters) was used to align the reference genomes (*S. indicum* and *S. latifolium*) with the other *Sesamum* genomes. Alignments were filtered using a delta-filter (parameter: -1) and calculated using show-coords (parameters: -clrT -I 90 -L 300). Assemblytics (https://github.com/MariaNattestad/Assemblytics) (parameters: 500, 50, 10 000) was used to identify SVs from the nucmer results. SURVIVOR (v.1.0.5, convertAssemblytics, minimum size: 50) was used to show the SV distribution. SVs were annotated using ANNOVAR (v.2015-06-17). All six SV types, including Insertion, Deletion, Tandem_expansion, Tandem_contraction, Repeat_expansion, and Repeat_contraction (50–10 000 bp) were identified in individual *Sesamum* species (Source Data 4 for Figure 3). For each *Sesamum* species, >50% of the SV overlapping with the reference genome was considered the same SV. All SVs were grouped into core SVs, dispensable SVs, and private SVs for intergenomic analysis. Presence and absence of SVs in *Sesamum* were determined using custom scripts.

Key gene families in sesame

R genes in *Sesamum* were predicted using the Disease Resistance Analysis and Gene Orthology (DRAGO2) pipeline (https://github.com/sequentiabiotech/DRAGO2-API), using HMMER v.3, COILS 2.2, and TMHMM 2.0c (Osuna-Cruz et al., 2018). PRGdb 3.0 (http://prgdb.org/prgdb4/) was used as a reference. All putative and reference pathogen-recognition genes were pooled together for BLASTP analysis (2.2.30+) with max_hsps_per_subject 1 and $P < 1E{-}06$. The 11 plant species *A. coerulea* (GCA_002738505.1), *Glycine max* (GCF_000004515.4), *Gossypium raimondii* (GCF_000987745.1_ASM98774v1), *Helianthus annuus* (Ha412v1r1_genes_v1.0), *L. usitatissimum* (Lusitatissimum_200_v1.0), *M. guttatus* (GCF_000504015.1), *O. sativa* (GCF_000005425.2), *Ricinus communis* (GCF_000151685.1), *S. lycopersicum* (GCF_000188115.3), *T. cacao* (GCF_000208745.1), and *V. vinifera* (GCF_000003745.3) were used for R-gene family comparison ($P < 1E{-}06$).

FA biosynthesis and metabolism-associated genes were identified in *Sesamum* species by aligning all unigenes to the acyl lipid metabolism reference dataset (http://aralip.plantbiology.msu.edu/pathways/pathways) using BLASTP ($P < 1E{-}5$, protein identity > 60%). Five oilseed crops (*H. annuus*, *G. max*, *G. raimondii*, *L. usitatissimum*, and *R. communis*) and a close species, *S. lycopersicum*, were used to identify and compare FA homologs and families. Sesame FA- and R-gene blocks were identified according to the following criteria: distance between two FA (or R) genes < 200 kb and < 8 annotated non-target-type genes present between two consecutive FA- (or R-) gene sequences (Jupe et al., 2012).

Phylogenetic analyses of CEN-like protein 2, DIR, NAC, and PPO proteins in sesame and other plants were performed using MAFFT v.7.5 (Katoh and Standley, 2013). Branch support values were calculated using 1000 bootstrap replications. PPI networks related to DIR and NAC proteins were analyzed using STRING (https://string-db.org) (Szklarczyk et al., 2021) with an *Arabidopsis* dataset.

Genome re-sequencing and GWAS analysis

DNA was extracted from 560 sesame accessions for PE (100 and/or 150 bp) library construction and sequencing on the Illumina HiSeq 2500 platform (Source Data Table 5). Re-sequencing data were aligned to the sesame reference genome using Burrows–Wheeler Aligner 0.7.15 with default settings (Li and Durbin, 2009). Putative SNPs and indels were identified using the Genome Analysis Tool Kit (GATK3.7) according to GATK joint-calling best practices, with the following filter parameters: FS > 60.0, MQ < 40.0, QD < 2.0, SOR > 3.0, MQRankSum < −2.5, and ReadPosRankSum < −8.0 (Poplin et al., 2017). Variants in all re-sequencing samples were filtered according to stringent criteria: minimum variant count ≥ 100 and minimum frequency of 0.1. A phylogenetic tree was constructed using SnPhylo (v.20140701) (Lee et al., 2014). The K value was determined, and Ln likelihood scores were calculated using ADMIXTURE (v.1.3.0; https://dalexander.github.io/admixture/download.html). SNP density, pairwise nucleotide diversity (π), and Fst indices were calculated using VCFtools (0.1.15) (Danecek et al., 2011). Haploview (4.2; http://www.broadinstitute.org/haploview) and Java (v.1.0.0_131) were used to construct the haplotype maps. The minimum minor-allele frequency was set to 0.05, and the maximum linkage disequilibrium comparison distance was 100 kb. The mean Fst and π values for sesame varieties (π_V) and germplasm accessions (π_G) were calculated for all variants using VCFtools, and π_G/π_V was calculated with a window size of 200 kb and a step size of 10 kb. Association analysis between variants and phenotypes was performed using an MLM in TASSEL 5. A minimum of 100 permutation analyses were performed (cutoff: 10^{-6}). SNPs with r^2 >5 and $P < 1E{-}6$ were retained for target SNP region detection. Candidate SNPs and genes were identified according to a previously established method (Yano et al., 2016).

GWAS analysis of branch number and OC phenotypes of 560 sesame germplasms in different environments was performed using TASSEL 5 and MLM (Anderson and Weir, 2007; Bradbury et al., 2007) (supplemental information). NR and KEGG annotations were assessed using BLASTP and BLAST2GO. Gene evolution analysis of the target genes *SiNAC*, *SiPPO*, and *SiCEN-Like2-1* was performed (supplemental information).

FOS resistance assay

The FOS resistance of cultivated sesame and six wild species was evaluated using self-constructed evaluation methods (Qiu et al., 2014; Miao et al., 2020b). Three FOS strains (HSFO08027, HSFO09018, and HSFO09095) with different pathogenicity characteristics were used at a conidium concentration of 1×10^6 colony-forming units/ml. *S. indicum*, *S. latifolium*, *S. angustifolium*, and *S. radiatum* were inoculated with the FOS strain HSFO 08027 carrying the GFP gene for 0 h to 6 d before histological observation. Confocal laser-scanning microscopy (OLYMPUS FV1000; Olympus, Tokyo, Japan) was used to observe fluorescence intensity. The 488-nm argon laser was set to 26% intensity.

Transcriptome sequencing and analysis of FOS resistance

Seedlings of *S. indicum*, *S. latifolium*, and *S. calycinum* with 2–3 leaf pairs were inoculated with the FOS strain HSFO08027 according to a previously described method (Miao et al., 2020b). A total of 203 groups of root tissue samples, inoculated for 0–168 h and assessed at 9 time points, were collected for RNA extraction and transcriptome sequencing (Source Data Table 2). After filtering, the FPKM value was used to estimate gene expression levels (Mortazavi et al., 2008). Differentially expressed genes (DEGs) (adjusted $P < 0.05$, log2[FC] > 1 or < -1) were identified using the log2 ratio of mapped readswith Differential Expression software (https://github.com/griffithlab/rnaseq_tutorial/wiki/Differential-Expression).

Weighted correlation network analysis was used to construct a DEG network (https://www.plob.org/article/14994.html). GO annotation of the DEGs was performed against the NCBI NR database using BLAST2GO ($E < 10^{-5}$) (Conesa et al., 2005). KEGG enrichment analysis of the key modules was performed using Omicshare CloudTools (http://www.omicshare.com/). Orthologous gene pairs in *S. indicum*, *S. latifolium*, and *S. calycinum* were identified using OrthoVenn (Wang et al., 2015b). The R genes identified in the DEG data were compared between different treatments and species. A gene interaction network was constructed using Cytoscape v.3.7.2 (Smoot et al., 2011).

GWAS analysis of *Fusarium* wilt resistance using an interspecific F_2 population

Fusarium wilt resistance was evaluated in the parental lines, *S. calycinum* (P_1) and *S. angustifolium* (P_2), as well as the F_1, F_2, and $F_{2\text{-}3}$ lines of their interspecific crossing populations, using a previously described inoculation method (Miao et al., 2020b). SNP matrixes of the 126 individuals were independently constructed for *S. calycinum* (P_1) and *S. angustifolium* (P_2) (Source Data Table 1; Supplemental Table 5). GWAS analysis of all variants and phenotypes in the population was performed using an MLM in TASSEL 5 and the SNP interval screening method (Zhang et al., 2016, 2019). A minimum of 100 permutation analyses were performed (cutoff: 10^{-6}). The syntenic blocks and DEGs in the target SNP intervals between *S. calycinum* and *S. angustifolium* were screened to detect candidate resistance genes. SV analysis of target SNP intervals in the two parents and 12 homogenous F_2 lines with high resistance or susceptibility to *Fusarium* wilt was performed to determine the target gene using LUMPY (v.0.2.13) (Layer et al., 2014), Manta (v.1.6.0) (Chen et al., 2016), GRIDSS2 (v.2.5.2) (Cameron et al., 2021), and DELLY (v.0.7.8) (Rausch et al., 2012). SURVIVOR software (v.1.0.6) (Jeffares et al., 2017) was used for SV filtering. Only SVs detected simultaneously by 2 or more software programs were retained for association analysis. DIR homologs in *Sesamum* species and other crops were analyzed using BLASTP.

OC assay and seed transcriptome analysis

The OC of cultivated sesame and six wild species was measured using gas chromatography (Agilent 7890, Agilent Technologies, USA). Seeds from the 560 sesame accessions were collected from healthy plants grown in 8 environments (Yuanyang [113°97′E, 35°05′N; 2016 and 2017], Yangchang [113°6′E, 34°87′N; 2017 and 2018], Pingyu [114°63′E, 32°97′N; 2017 and 2018], Yuanmou [101°85′E, 25°07′N; 2018], and Sanya [109°50′E and 18°25′N; 2016]). The OC of each sample was measured using a previously established nondestructive NMR method with three replicates per sample. ANOVA of OC traits was performed using SAS (v.8.0). The OC trait distribution was normalized using the Shapiro–Wilk test with a significance level of $P < 0.05$. The effects of genotype, environment, and the genotype × environment interaction on phenotypic variation were evaluated using PROC GLM.

Eighteen groups of developing seed samples, ranging from 7–10 d to 29–30 d, were collected from *S. indicum*, *S. radiatum*, *S. latifolium*, and *S. calycinum* for RNA extraction and transcriptome sequencing (Source Data Table 3); each set included 3–5 time points. Forty-two groups of developing *Sesamum* seed samples were from the high-OC accessions var. Shannzhi 3 and WanzhiC02-3, spanning 4–28 DAF; 17 sets of the low-OC accession var. Black sesame, spanning 4–20 DAF, were also collected. All samples were used for RNA extraction and transcriptome sequencing (Source Data Table 4). The expression profiles of orthologous genes among the DEGs were compared using FPKM values (Smoot et al., 2011).

Transgenic functional validation of target genes

To validate the functions of target genes, overexpression vectors were constructed for *SiCEN-Like2-1* (NCBI: MT362742), *SiNAC* (NCBI: MT365249), and *SiPPO* (NCBI: MT365245) and transferred into *Arabidopsis* (LL4404) using the binary vector pFGC5941. The endogenous reference gene *At*β*-tubulin* was used as a reference for qRT–PCR. The transcript levels of *SiCEN-Like2-1*, *SiNAC*, *SiPPO*, and gene alleles were individually normalized against that of the β*-tubulin* gene. Comparisons were performed using the $2^{-\Delta\Delta Ct}$ method.

DATA AND CODE AVAILABILITY

The genome sequence of sesame has been deposited at GenBank under accession number MBSK00000000 in BioProject PRJNA315784. The assembled genome sequences and annotations of the six wild *Sesamum* species have been deposited at GenBank under accession numbers JACGWJ000000000, JACGWL000000000, JACGWM000000000, JACGWK000000000, JACGWN000000000, and JACGWO000000000 in BioProject PRJNA625646. Genome re-sequencing data for the 560 germplasm accessions is available at NCBI in BioProject PRJNA626474 (SAMN14693735–14694294). Sequences of three genes, *SiPT1* (*SiCEN-Like2-1*; NCBI: MT362742), *SiNAC* (NCBI: MT365249), and *SiPPO* (NCBI: MT365245), were uploaded to the NCBI database. Patents for *SiPT1* (*SiCEN-Like2-1*) have been applied for in China with patent application 2023102968853 (March 24, 2023).

SUPPLEMENTAL INFORMATION

Supplemental information is available at *Plant Communications Online*.

FUNDING

This work was supported by earmarked funding for the China Agricultural Research System of MOF and MARA (CARS-14), China; the China National "973" Project (2011CB109304), China; the Henan Zhongyuan Scientist Work Station Construction Fund (092101211100), China; the National Natural Science Foundation of China (U1204318, U1304321, 31301653, 31471537, and 32172094), China; the Key Project of Science and Technology of Henan Province (201300110600), China; the Key Research Project of the Shennong Laboratory (SN01-2022-04), China; the Key Research and Development Project of Henan Province (221111520400), China; and the Innovation Scientists and Technicians Troop Construction Project of the Henan Academy of Agricultural Sciences (2023TD04), China.

AUTHOR CONTRIBUTIONS

H.Z. conceived the project. H.Z., H. Miao and L.Wang supervised the research. H. Miao and L.Wang designed the experiments and managed particular components of the project. Y.S. led the bioinformatics analysis. Y.S., W.L., and S.X. performed genome assembly, genome structure, and genomic analyses. H. Miao and W.L. performed the genome synteny and evolution analyses. R.Z. and Q.M. established the consecutive BAC-FISH technique in sesame and constructed the cytogenetic map. L.Wei, G.L., H.C., C.L., Y.D., and Y.S. performed the genetic mapping and functional genomics analyses. Y.D., C.M., and Q.T. oversaw the histological imaging and gene cloning, gene expression, and genetic transformation analysis. H.L., M.L., and Q.W. participated in genetics analyses and shared their knowledge of sesame. H. Miao, H.Z., and Y.S. supervised the BioNano mapping. L.Wei, M.J., and Q.M. prepared materials for sequencing and trait investigation. Y.D., L.Wei., M.J., and Q.M. performed agronomic trait investigation and analyses. X.W. led the study and the analysis of seed quality. H.Z. led the phylogenetic analyses of the sesame germplasm accessions. S.W., Y.Z., H. Mei, T.G., and T.Z. participated in the project. H. Miao, L.Wang, and H.Z. organized the manuscript. H.Z., H. Miao, L.Wang, and L.Q. coordinated the Sesame Genome Project.

ACKNOWLEDGMENTS

We thank Dr. Joy Fleming (CAS Institute of Biophysics, Beijing, China) for constructive discussions and manuscript editing. This work was supported by the Henan Key Laboratory of Oilseed Crop Genomics. No conflict of interest is declared.

Received: June 25, 2023
Revised: August 17, 2023
Accepted: October 2, 2023
Published: October 5, 2023

REFERENCES

Ananiev, E.V., Phillips, R.L., and Rines, H.W. (1998). Chromosome-specific molecular organization of maize (*Zea mays* L.) centromeric regions. Proc. Natl. Acad. Sci. USA **95**:13073–13078. https://doi.org/10.1073/pnas.95.22.13073.

Anderson, A.D., and Weir, B.S. (2007). A maximum-likelihood method for the estimation of pairwise relatedness in structured populations. Genetics **176**:421–440. https://doi.org/10.1534/genetics.106.063149.

Argout, X., Salse, J., Aury, J.M., Guiltinan, M.J., Droc, G., Gouzy, J., Allegre, M., Chaparro, C., Legavre, T., Maximova, S.N., et al. (2011). The genome of Theobroma cacao. Nat. Genet. **43**:101–108. https://doi.org/10.1038/ng.736.

Bedigian, D., and Harlan, J.R. (1986). Evidence for cultivation of sesame in the ancient world. Econ. Bot. **40**:137–154. https://doi.org/10.1007/BF02859136.

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

Bradbury, P.J., Zhang, Z., Kroon, D.E., Casstevens, T.M., Ramdoss, Y., and Buckler, E.S. (2007). TASSEL: software for association mapping of complex traits in diverse samples. Bioinformatics **23**:2633–2635. https://doi.org/10.1093/bioinformatics/btm308.

Cameron, D.L., Baber, J., Shale, C., Valle-Inclan, J.E., Besselink, N., van Hoeck, A., Janssen, R., Cuppen, E., Priestley, P., and Papenfuss, A.T. (2021). GRIDSS2: comprehensive characterisation of somatic structural variation using single breakend variants and structural variant phasing. Genome Biol. **22**:202. https://doi.org/10.1186/s13059-021-02423-x.

Chen, C.B., Wang, C.G., and Song, W.Q. (2015a). A Method for Fluorescence in Situ Hybridization of 5S rDNA on Plant Chromosomes, CN103409523B.

Chen, C.B., Wang, C.G., and Song, W.Q. (2015b). A Method for Fluorescence in Situ Hybridization of 45S rDNA on Plant Chromosomes, CN103409524B.

Chen, S.L., Huang, J.Q., Lei, Y., Zhang, Y.T., Ren, X.P., Chen, Y.N., Jiang, H.F., Yan, L.Y., Li, Y.R., and Liao, B.S. (2012). Identification and characterization of a gene encoding a putative lysophosphatidyl acyltransferase from Arachis hypogaea. J. Bio. Sci. **37**:1029–1039. https://doi.org/10.1007/s12038-012-9277-4.

Chen, X., Schulz-Trieglaff, O., Shaw, R., Barnes, B., Schlesinger, F., Källberg, M., Cox, A.J., Kruglyak, S., and Saunders, C.T. (2016). Manta: rapid detection of structural variants and indels for germline and cancer sequencing applications. Bioinformatics **32**:1220–1222. https://doi.org/10.1093/bioinformatics/btv710.

Collier, S.M., Hamel, L.P., and Moffett, P. (2011). Cell death mediated by the N-terminal domains of a unique and highly conserved class of NB-LRR protein. Mol. Plant Microbe Interact. **24**:918–931. https://doi.org/10.1094/MPMI-03-11-0050.

Conesa, A., Götz, S., García-Gómez, J.M., Terol, J., Talón, M., and Robles, M. (2005). Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. Bioinformatics **21**:3674–3676. https://doi.org/10.1093/bioinformatics/bti610.

Danecek, P., Auton, A., Abecasis, G., Albers, C.A., Banks, E., DePristo, M.A., Handsaker, R.E., Lunter, G., Marth, G.T., Sherry, S.T., et al. (2011). The variant call format and VCFtools. Bioinformatics **27**:2156–2158. https://doi.org/10.1093/bioinformatics/btr330.

Demeke, T., Chang, H.G., and Morris, C.F. (2001). Effect of germination, seed abrasion and seed size on polyphenol oxidase assay activity in wheat. Plant Breed. **120**:369–373. https://doi.org/10.1046/j.1439-0523.2001.00625.x.

Deng, J., Guan, R., Liang, T., Su, L., Ge, F., Cui, X., and Liu, D. (2022). Dirigent gene family is involved in the molecular interaction between *Panax notoginseng* and root rot pathogen *Fusarium solani*. Ind. Crop. Prod. **178**, 114544. https://doi.org/10.1016/j.indcrop.2022.114544.

Dong, Z., Feng, B., Liang, H., Rong, C., Zhang, K., Cao, X., Qin, H., Liu, X., Wang, T., and Wang, D. (2015). Grain-specific reduction in lipoxygenase activity improves flour color quality and seed longevity in common wheat. Mol. Breed. **35**:150. https://doi.org/10.1007/s11032-015-0347-9.

Duan, Y., Qu, W., Chang, S., Li, C., Xu, F., Ju, M., Zhao, R., Wang, H., Zhang, H., and Miao, H. (2020). Identification of pathogenic groups and pathogenic molecular characterization of *Fusarium oxysporum* f. sp. *sesami* in China. Phytopathology **110**:1093–1104. https://doi.org/10.1094/PHYTO-09-19-0366-R.

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., and Aiden, E.L. (2017). De novo assembly of the *Aedes aegypti* genome using Hi-C yields chromosome-length scaffolds. Science **356**:92–95. https://doi.org/10.1126/science.aal3327.

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

Fonsêca, A., Ferreira, J., dos Santos, T.R.B., Mosiolek, M., Bellucci, E., Kami, J., Gepts, P., Geffroy, V., Schweizer, D., dos Santos, K.G.B., and Pedrosa-Harand, A. (2010). Cytogenetic map of common bean (*Phaseolus vulgaris* L.). Chromosome Res. **18**:487–502. https://doi.org/10.1007/s10577-010-9129-8.

Fuchs, J., Brandes, A., and Schubert, I. (1995). Telomere sequence localization and karyotype evolution in higher plants. Plant Systemat. Evol. **196**:227–241. https://doi.org/10.1007/BF00982962.

Gai, J.Y. (2003). Expanding and enhancing the research allocation on soybean breeding and genetics for the establishment of market supply based on domestic production. Eng. Sci. **5**:1–6.

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

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. https://doi.org/10.1038/nature06148.

Jeffares, D.C., Jolly, C., Hoti, M., Speed, D., Shaw, L., Rallis, C., Balloux, F., Dessimoz, C., Bähler, J., and Sedlazeck, F.J. (2017). Transient structural variations have strong effects on quantitative traits and reproductive isolation in fission yeast. Nat. Commun. **8**, 14061. https://doi.org/10.1038/ncomms14061.

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. https://doi.org/10.1093/bioinformatics/btu031.

Joshi, A.B. (1961). Sesamum (Indian Central Oilseeds Committee).

Jupe, F., Pritchard, L., Etherington, G.J., MacKenzie, K., Cock, P.J.A., Wright, F., Sharma, S.K., Bolser, D., Bryan, G.J., Jones, J.D.G., and Hein, I. (2012). Identification and localisation of the NB-LRR gene family within the potato genome. BMC Genom. **13**:75. https://doi.org/10.1186/1471-2164-13-75.

Kanehisa, M., Goto, S., Kawashima, S., Okuno, Y., and Hattori, M. (2004). The KEGG resource for deciphering the genome. Nucleic Acids Res. **32**:D277–D280. https://doi.org/10.1093/nar/gkh063.

Katoh, K., and Standley, D.M. (2013). MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol. Biol. Evol. **30**:772–780. https://doi.org/10.1093/molbev/mst010.

Katoh, K., Misawa, K., Kuma, K.i., and Miyata, T. (2002). MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. Nucleic Acids Res. **30**:3059–3066. https://doi.org/10.1093/nar/gkf436.

Kim, D., Pertea, G., Trapnell, C., Pimentel, H., Kelley, R., and Salzberg, S.L. (2013). TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. Genome Biol. **14**:R36. https://doi.org/10.1186/gb-2013-14-4-r36.

Kobayashi, T. (1991). Cytogenetics of sesame (*Sesamum indicum*). In Book: Developments in Plant Genetics and Breeding, T. Tsuchiya and P.K. Gupta, eds. (Elsevier), pp. 581–592. https://doi.org/10.1016/B978-0-444-88260-8.50036-7.

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

Kozlov, A.M., Darriba, D., Flouri, T., Morel, B., and Stamatakis, A. (2019). RAxML-NG: a fast, scalable and user-friendly tool for maximum likelihood phylogenetic inference. Bioinformatics **35**:4453–4455. https://doi.org/10.1093/bioinformatics/btz305.

Kurtz, S., Phillippy, A., Delcher, A.L., Smoot, M., Shumway, M., Antonescu, C., and Salzberg, S.L. (2004). Versatile and open software for comparing large genomes. Genome Biol. **5**:R12. https://doi.org/10.1186/gb-2004-5-2-r12.

Layer, R.M., Chiang, C., Quinlan, A.R., and Hall, I.M. (2014). LUMPY: a probabilistic framework for structural variant discovery. Genome Biol. **15**:R84. https://doi.org/10.1186/gb-2014-15-6-r84.

Lee, T.H., Guo, H., Wang, X., Kim, C., and Paterson, A.H. (2014). SNPhylo: a pipeline to construct a phylogenetic tree from huge SNP data. BMC Genom. **15**:162. https://doi.org/10.1186/1471-2164-15-162.

Li, C., Li, X., Liu, H., Wang, X., Li, W., Chen, M.S., and Niu, L.J. (2020). Chromatin architectures are associated with response to dark treatment in the oil crop *Sesamum indicum*, based on a high-quality genome assembly. Plant Cell Physiol. **61**:978–987. https://doi.org/10.1093/pcp/pcaa026.

Li, F., Fan, G., Lu, C., Xiao, G., Zou, C., Kohel, R.J., Ma, Z., Shang, H., Ma, X., Wu, J., et al. (2015). Genome sequence of cultivated Upland cotton (*Gossypium hirsutum* TM-1) provides insights into genome evolution. Nat. Biotechnol. **33**:524–530. https://doi.org/10.1038/nbt.3208.

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

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., and Durbin, R.; 1000 Genome Project Data Processing Subgroup (2009). The sequence alignment/map format and SAMtools. Bioinformatics **25**:2078–2079. https://doi.org/10.1093/bioinformatics/btp352.

Li, J., Yuan, D., Wang, P., Wang, Q., Sun, M., Liu, Z., Si, H., Xu, Z., Ma, Y., Zhang, B., et al. (2021a). Cotton pan-genome retrieves the lost sequences and genes during domestication and selection. Genome Biol. **22**:119. https://doi.org/10.1186/s13059-021-02351-w.

Li, L., Sun, W., Zhou, P., Wei, H., Wang, P., Li, H., Rehman, S., Li, D., and Zhuge, Q. (2021b). Genome-wide characterization of dirigent proteins in populus: gene expression variation and expression pattern in response to marssonina brunnea and phytohormones. Forests **12**:507. https://doi.org/10.3390/f12040507.

Li, P.J., Li, B.J., Nan, W., Mi, Y., Liao, F., and Li, G.R. (2017). Discriminatory analysis on the morphology and major components of kernels of the edible and oil-type sunflowers. J. Sichuan Univ. **54**:203–208.

Li, W., Zhong, J., Zhang, L., Wang, Y., Song, P., Liu, W., Li, X., and Han, D. (2022). Overexpression of a *Fragaria vesca* MYB transcription factor gene (*FvMYB82*) increases salt and cold tolerance in *Arabidopsis thaliana*. Int. J. Mol. Sci. **23**, 10538. https://doi.org/10.3390/ijms231810538.

Li, W.J., Qi, Y.N., Wang, L.M., Dang, Z., Zhao, L., Zhao, W., Xie, Y.P., Wang, B., Zhang, J.P., and Li, S.J. (2019). Correlation between oil content or fatty acid composition and expression levels of genes involved in TAG biosynthesis in flax. Acta Pratacult. Siniva **28**:138–149. https://doi.org/10.11686/cyxb2018321.

Liu, L., Yu, Y., Kong, X.H., Wang, J., Wang, X.W., and Hua, J.P. (2016a). Analysis on seed development characteristics and fatty acid composition of early-maturing varieties of upland cotton in Xinjiang. Southwest China J. Agric. Sci. **29**:765–769. https://doi.org/10.16213/j.cnki.scjas.2016.04.007.

Liu, M., Dong, H., Wang, M., and Liu, Q. (2020a). Evolutionary divergence of function and expression of laccase genes in plants. J. Genet. **99**:23. https://doi.org/10.1007/s12041-020-1184-0.

Liu, Y., Du, H., Li, P., Shen, Y., Peng, H., Liu, S., Zhou, G.A., Zhang, H., Liu, Z., Shi, M., et al. (2020b). Pan-genome of wild and cultivated soybeans. Cell **182**:162–176.e13. https://doi.org/10.1016/j.cell.2020.05.023.

Liu, Z., Li, Y., Li, J.B., and Yin, Y.L. (2016b). Expression of a tomato MYB gene in transgenic tobacco increases resistance to *Fusarium oxysporum* and *Botrytis cinerea*. Eur. J. Plant Pathol. **37**:607–614. https://doi.org/10.1007/s10658-015-0799-0.

Liu, Z., Wang, X., Sun, Z., Zhang, Y., Meng, C., Chen, B., Wang, G., Ke, H., Wu, J., Yan, Y., et al. (2021). Evolution, expression and functional analysis of cultivated allotetraploid cotton *DIR* genes. BMC Plant Biol. **21**:89. https://doi.org/10.1186/s12870-021-02859-0.

Lung, S.C., Liao, P., Yeung, E.C., Hsiao, A.S., Xue, Y., and Chye, M.L. (2018). Arabidopsis ACYL-COA-BINDING PROTEIN 1 interacts with STEROL C4-METHYL OXIDASE 1-2 to modulate gene expression of homeodomain-leucine zipper IV transcription factors. New Phytol. **218**:183–200. https://doi.org/10.1111/nph.14965.

Ma, L.J., Geiser, D.M., Proctor, R.H., Rooney, A.P., O'Donnell, K., Trail, F., Gardiner, D.M., Manners, J.M., and Kazan, K. (2013). Fusarium Pathogenomics. Annu. Rev. Microbiol. **67**:399–416. https://doi.org/10.1146/annurev-micro-092412-155650.

Majoros, W.H., Pertea, M., and Salzberg, S.L. (2004). TigrScan and GlimmerHMM: two open source ab initio eukaryotic gene-finders. Bioinformatics **20**:2878–2879. https://doi.org/10.1093/bioinformatics/bth315.

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

Mei, H., Liu, Y., Du, Z., Wu, K., Cui, C., Jiang, X., Zhang, H., and Zheng, Y. (2017). High-density genetic map construction and gene mapping of basal branching habit and flowers per leaf axil in sesame. Front. Plant Sci. **8**:636. https://doi.org/10.3389/fpls.2017.00636.

Miao, H.M., Langham, D.R., and Zhang, H.Y. (2021a). Botanical descriptions of sesame. In Book: The Sesame Genome. Compendium of Plant Genomes, H.M. Miao, H.Y. Zhang, and C. Kole, eds. (Springer), pp. 19–57. https://doi.org/10.1007/978-3-319-98098-0_2.

Miao, H.M., Sun, Y.M., Li, C., Wang, L., and Zhang, H.Y. (2021b). Genome annotation and gene families in sesame. In Book: The Sesame Genome. Compendium of Plant Genomes, H.M. Miao, H.Y. Zhang, and C. Kole, eds. (Springer), pp. 255–266. https://doi.org/10.1007/978-3-319-98098-0_15.

Miao, H., Li, C., Duan, Y., Wei, L., Ju, M., and Zhang, H. (2020a). Identification of a *Sidwf1* gene controlling short internode length trait in the sesame dwarf mutant *dw607*. Theor. Appl. Genet. **133**:73–86. https://doi.org/10.1007/s00122-019-03441-x.

Miao, H.M., Chang, S.X., Zhang, H.Y., Huang, J.Y., Duan, Y.H., and Qu, W.W. (2020b). An evalutation technique of sesame resistance to Fusarium wilt disease at vegetative stage. J. Plant Genet. Resour. **21**:330–337. https://doi.org/10.13430/j.cnki.jpgr.20190428003.

Mortazavi, A., Williams, B.A., McCue, K., Schaeffer, L., and Wold, B. (2008). Mapping and quantifying mammalian transcriptomes by RNA-Seq. Nat. Methods **5**:621–628. https://doi.org/10.1038/nmeth.1226.

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

Myburg, A.A., Grattapaglia, D., Tuskan, G.A., Hellsten, U., Hayes, R.D., Grimwood, J., Jenkins, J., Lindquist, E., Tice, H., Bauer, D., et al. (2014). The genome of *Eucalyptus grandis*. Nature **510**:356–362. https://doi.org/10.1038/nature13308.

Nimmakayala, P., Perumal, R., Mulpuri, S., and Reddy, U.K. (2011). In *Sesamum* in Book: Wild Crop Relatives - Genomic and Breeding Resources, C.K. Oilseeds, ed. (Springer Verlag), pp. 261–273. https://doi.org/10.1007/978-3-642-14871-2_16.

Ohlrogge, J.B., and Jaworski, J.G. (1997). Regulation of fatty acid synthesis. Annu. Rev. Plant Physiol. **48**:109–136. https://doi.org/10.1146/annurev.arplant.48.1.109.

Osuna-Cruz, C.M., Paytuvi-Gallart, A., Di Donato, A., Sundesha, V., Andolfo, G., Aiese Cigliano, R., Sanseverino, W., and Ercolano, M.R. (2018). PRGdb 3.0: a comprehensive platform for prediction and analysis of plant disease resistance genes. Nucleic Acids Res. **46**:D1197–D1201. https://doi.org/10.1093/nar/gkx1119.

Paniagua, C., Bilkova, A., Jackson, P., Dabravolski, S., Riber, W., Didi, V., Houser, J., Gigli-Bisceglia, N., Wimmerova, M., Budínská, E., et al. (2017). Dirigent proteins in plants: modulating cell wall metabolism during abiotic and biotic stress exposure. J. Exp. Bot. **68**:3287–3301. https://doi.org/10.1093/jxb/erx141.

Park, Y.J., Lee, H.J., Kwak, K.J., Lee, K., Hong, S.W., and Kang, H. (2014). MicroRNA400-guided cleavage of pentatricopeptide repeat protein mRNAs renders *Arabidopsis thaliana* more susceptible to pathogenic bacteria and fungi. Plant Cell Physiol. **55**:1660–1668. https://doi.org/10.1093/pcp/pcu096.

Parra, G., Bradnam, K., and Korf, I. (2007). CEGMA: a pipeline to accurately annotate core genes in eukaryotic genornes. Bioinformatics **23**:1061–1067. https://doi.org/10.1093/bioinformatics/btm071.

Poplin, R., Ruano-Rubio, V., DePristo, M.A., Fennell, T.J., Carneiro, M.O., Van der Auwera, G.A., Kling, D.E., Gauthier, L.D., Levy-Moonshine, A., Roazen, D., et al. (2017). Scaling accurate genetic variant discovery to tens of thousands of samples. Preprint at bioRxiv. https://doi.org/10.1101/201178.

Puranik, S., Sahu, P.P., Srivastava, P.S., and Prasad, M. (2012). NAC proteins: regulation and role in stress tolerance. Trends Plant Sci. **17**:369–381. https://doi.org/10.1016/j.tplants.2012.02.004.

Qiu, C.P., Zhang, H.Y., Chang, S.X., Wei, L.B., and Miao, H.M. (2014). Laboratory detecting method for pathogenicity of *Fusarium oxsporum* Schl. f. sp. *sesami* isolates. Acta Phytopathol. Sin. **44**:26–35. https://doi.org/10.13926/j.cnki.apps.2014.01.004.

Rausch, T., Zichner, T., Schlattl, A., Stütz, A.M., Benes, V., and Korbel, J.O. (2012). DELLY: structural variant discovery by integrated paired-end and split-read analysis. Bioinformatics **28**:i333–i339. https://doi.org/10.1093/bioinformatics/bts378.

Roesler, K., Shintani, D., Savage, L., Boddupalli, S., and Ohlrogge, J. (1997). Targeting of the Arabidopsis homomeric acetyl-coenzyme A carboxylase to plastids of rapeseeds. Plant Physiol. **113**:75–81. https://doi.org/10.1104/pp.113.1.75.

Salse, J. (2012). *In silico* archeogenomics unveils modern plant genome organisation, regulation and evolution. Curr. Opin. Plant Biol. **15**:122–130. https://doi.org/10.1016/j.pbi.2012.01.001.

Santos, C., Nelson, C.D., Zhebentyayeva, T., Machado, H., Gomes-Laranjo, J., and Costa, R.L. (2017). First interspecific genetic linkage map for *Castanea sativa* x *Castanea crenata* revealed QTLs for resistance to *Phytophthora cinnamomi*. PLoS One **12**, e0184381. https://doi.org/10.1371/journal.pone.0184381.

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

Shao, Z.Q., Xue, J.Y., Wu, P., Zhang, Y.M., Wu, Y., Hang, Y.Y., Wang, B., and Chen, J.Q. (2016). Large-scale analyses of angiosperm nucleotide-binding site-leucine-rich repeat genes reveal three anciently diverged classes with distinct evolutionary patterns. Plant Physiol. **170**:2095–2109. https://doi.org/10.1104/pp.15.01487.

Shen, B., Sinkevicius, K.W., Selinger, D.A., and Tarczynski, M.C. (2006). The homeobox gene *GLABRA2* affects seed oil content in *Arabidopsis*. Plant Mol. Biol. **60**:377–387. https://doi.org/10.1007/s11103-005-4110-1.

Shockey, J., and Browse, J. (2011). Genome-level and biochemical diversity of the acyl-activating enzyme superfamily in plants. Plant J. **66**:143–160. https://doi.org/10.1111/j.1365-313X.2011.04512.x.

Smoot, M.E., Ono, K., Ruscheinski, J., Wang, P.L., and Ideker, T. (2011). Cytoscape 2.8: new features for data integration and network visualization. Bioinformatics **27**:431–432. https://doi.org/10.1093/bioinformatics/btq675.

Stanke, M., Keller, O., Gunduz, I., Hayes, A., Waack, S., and Morgenstern, B. (2006). AUGUSTUS: *ab initio* prediction of alternative transcripts. Nucleic Acids Res. **34**:W435–W439. https://doi.org/10.1093/nar/gkl200.

Stein, L.D. (2013). Using GBrowse 2.0 to visualize and share next-generation sequence data. Briefings Bioinf. **14**:162–171. https://doi.org/10.1093/bib/bbt001.

Szklarczyk, D., Gable, A.L., Nastou, K.C., Lyon, D., Kirsch, R., Pyysalo, S., Doncheva, N.T., Legeay, M., Fang, T., Bork, P., et al. (2021). The STRING database in 2021: customizable protein-protein networks, and functional characterization of user-uploaded gene/measurement sets. Nucleic Acids Res. **49**:D605–D612. https://doi.org/10.1093/nar/gkaa1074.

Tarailo-Graovac, M., and Chen, N. (2009). Using RepeatMasker to identify repetitive elements in genomic sequences. Curr Protoc Bioinform **25**:4. https://doi.org/10.1002/0471250953.bi0410s25.

Tek, A.L., and Jiang, J. (2004). The centromeric regions of potato chromosomes contain megabase-sized tandem arrays of telomere-similiar sequence. Chromosoma **113**:77–83. https://doi.org/10.1007/s00412-004-0297-1.

Uncu, A.Ö., Gultekin, V., Allmer, J., Frary, A., and Doganlar, S. (2015). Genomic simple sequence repeat markers reveal patterns of genetic relatedness and diversity in sesame. Plant Genome **8**, eplantgenome2014.11.0087-12. https://doi.org/10.3835/plantgenome2014.11.0087.

Unver, T., Wu, Z., Sterck, L., Turktas, M., Lohaus, R., Li, Z., Yang, M., He, L., Deng, T., Escalante, F.J., et al. (2017). Genome of wild olive and the evolution of oil biosynthesis. Proc. Natl. Acad. Sci. USA **114**:E9413–E9422. https://doi.org/10.1073/pnas.1708621114.

Van Ooijen, J.W. (2011). Multipoint maximum likelihood mapping in a full-sib family of an outbreeding species. Genet. Res. **93**:343–349. https://doi.org/10.1017/S0016672311000279.

Vaughn, K.C., and Duke, S.O. (1984). Function of polyphenol oxidase in higher plants. Physiol. Plantarum **60**:106–112. https://doi.org/10.1111/j.1399-3054.1984.tb04258.x.

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

Wang, L., Yu, S., Tong, C., Zhao, Y., Liu, Y., Song, C., Zhang, Y., Zhang, X., Wang, Y., Hua, W., et al. (2014). Genome sequencing of the high oil crop sesame provides insight into oil biosynthesis. Genome Biol. **15**:R39. https://doi.org/10.1186/gb-2014-15-2-r39.

Wang, S., Chen, J., Zhang, W., Hu, Y., Chang, L., Fang, L., Wang, Q., Lv, F., Wu, H., Si, Z., et al. (2015a). Sequence-based ultra-dense genetic and physical maps reveal structural variations of allopolyploid cotton genomes. Genome Biol. **16**:108. https://doi.org/10.1186/s13059-015-0678-1.

Wang, X., Wang, S., Lin, Q., Lu, J., Lv, S., Zhang, Y., Wang, X., Fan, W., Liu, W., Zhang, L., et al. (2022). The wild allotetraploid sesame genome provides novel insights into evolution and lignan biosynthesis. J. Adv. Res. **50**:13–24. https://doi.org/10.1016/j.jare.2022.10.004.

Wang, Y., Coleman-Derr, D., Chen, G., and Gu, Y.Q. (2015b). OrthoVenn: a web server for genome wide comparison and annotation of orthologous clusters across multiple species. Nucleic Acids Res. **43**:W78–W84. https://doi.org/10.1093/nar/gkv487.

Wang, Y., Tang, H., DeBarry, J.D., Tan, X., Li, J., Wang, X., Lee, T.H., 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. https://doi.org/10.1093/nar/gkr1293.

Waterhouse, R.M., Seppey, M., Simão, F.A., Manni, M., Ioannidis, P., Klioutchnikov, G., Kriventseva, E.V., and Zdobnov, E.M. (2018). BUSCO applications from quality assessments to gene prediction and phylogenomics. Mol. Biol. Evol. **35**:543–548. https://doi.org/10.1093/molbev/msx319.

Wei, M., Liu, A., Zhang, Y., Zhou, Y., Li, D., Dossa, K., Zhou, R., Zhang, X., and You, J. (2019). Genome-wide characterization and expression analysis of the HD-Zip gene family in response to drought and salinity stresses in sesame. BMC Genom. **20**:748. https://doi.org/10.1186/s12864-019-6091-5.

Wei, X., Liu, K., Zhang, Y., Feng, Q., Wang, L., Zhao, Y., Li, D., Zhao, Q., Zhu, X., Zhu, X., et al. (2015). Genetic discovery for oil production and quality in sesame. Nat. Commun. **6**:8609. https://doi.org/10.1038/ncomms9609.

Wood, T.E., Takebayashi, N., Barker, M.S., Mayrose, I., Greenspoon, P.B., and Rieseberg, L.H. (2009). The frequency of polyploid speciation in vascular plants. Proc. Natl. Acad. Sci. USA **106**:13875–13879. https://doi.org/10.1073/pnas.0811575106.

Wu, S., Lau, K.H., Cao, Q., Hamilton, J.P., Sun, H., Zhou, C., Eserman, L., Gemenet, D.C., Olukolu, B.A., Wang, H., et al. (2018). Genome sequences of two diploid wild relatives of cultivated sweetpotato reveal targets for genetic improvement. Nat. Commun. **9**:4580. https://doi.org/10.1038/s41467-018-06983-8.

Xu, X., Pan, S., Cheng, S., Zhang, B., Mu, D., Ni, P., Zhang, G., Yang, S., Li, R., Wang, J., et al. (2011). Genome sequence and analysis of the tuber crop potato. Nature **475**:189–195. https://doi.org/10.1038/nature10158.

Xu, Y., Bi, C., Wu, G., Wei, S., Dai, X., Yin, T., and Ye, N. (2016). VGSC: a web-based vector graph toolkit of genome synteny and collinearity. BioMed Res. Int. **2016**, 7823429. https://doi.org/10.1155/2016/7823429.

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

Yang, Q.F., Liu, L., Liu, Y., and Zhou, Z.G. (2017). Telomeric localization of the Arabidopsis-type heptamer repeat, (TTTAGGG)n, at the chromosome ends in *Saccharina japonica* (Phaeophyta). J. Phycol. **53**:235–240. https://doi.org/10.1111/jpy.12497.

Yang, X., Zhong, S., Zhang, Q., Ren, Y., Sun, C., and Chen, F. (2021). A loss-of-function of the dirigent gene *TaDIR-B1* improves resistance to Fusarium crown rot in wheat. Plant Biotechnol. J. **19**:866–868. https://doi.org/10.1111/pbi.13554.

Yang, Z. (1997). PAML: a program package for phylogenetic analysis by maximum likelihood. Bioinformatics **13**:555–556. https://doi.org/10.1093/bioinformatics/13.5.555.

Yano, K., Yamamoto, E., Aya, K., Takeuchi, H., Lo, P.C., Hu, L., Yamasaki, M., Yoshida, S., Kitano, H., Hirano, K., and Matsuoka, M. (2016). Genome-wide association study using whole-genome sequencing rapidly identifies new genes influencing agronomic traits in rice. Nat. Genet. **48**:927–934. https://doi.org/10.1038/ng.3596.

Yue, J.X., Meyers, B.C., Chen, J.Q., Tian, D., and Yang, S. (2012). Tracing the origin and evolutionary history of plant nucleotide-binding site-leucine-rich repeat (NBS-LRR) genes. New Phytol. **193**:1049–1063. https://doi.org/10.1111/j.1469-8137.2011.04006.x.

Zhang, H.Y., Miao, H.M., and Ju, M. (2019). Potential for adaptation to climate change through genomic breeding in sesame. In Book: Genomic Designing of Climate-Smart Oilseed Crops, C. Kole, ed. (Springer), pp. 371–440. https://doi.org/10.1007/978-3-319-93536-2_7.

Zhang, H.Y., Wang, L., and Miao, H.M. (2021). Background of the sesame genome project. In Book: The Sesame Genome. Compendium of Plant Genomes, H.M. Miao, H.Y. Zhang, and C. Kole, eds. (Springer), pp. 199–204. https://doi.org/10.1007/978-3-319-98098-0_10.

Zhang, H., Li, C., Miao, H., and Xiong, S. (2013a). Insights from the complete chloroplast genome into the evolution of *Sesamum indicum* L. PLoS One **8**, e80508. https://doi.org/10.1371/journal.pone.0080508.

Zhang, H.Y., Miao, H.M., Li, C., Wei, L.B., and Ma, Q. (2012). Analysis of sesame karyotype and resemblance-near coefficient. Bull Bot **47**:602–614. https://doi.org/10.3724/SP.J.1259.2012.00602.

Zhang, H., Miao, H., Wang, L., Qu, L., Liu, H., Wang, Q., and Yue, M. (2013b). Genome sequencing of the important oilseed crop

Sesamum indicum L. Genome Biol. **14**:401. https://doi.org/10.1186/gb-2013-14-1-401.

Zhang, H., Miao, H., Li, C., Wei, L., Duan, Y., Ma, Q., Kong, J., Xu, F., and Chang, S. (2016). Ultra-dense SNP genetic map construction and identification of *SiDt* gene controlling the determinate growth habit in *Sesamum indicum* L. Sci. Rep. **6**, 31556. https://doi.org/10.1038/srep31556.

Zhang, L., Chen, F., Zhang, X., Li, Z., Zhao, Y., Lohaus, R., Chang, X., Dong, W., Ho, S.Y.W., Liu, X., et al. (2020). The water lily genome and the early evolution of flowering plants. Nature **577**:79–84. https://doi.org/10.1038/s41586-019-1852-5.

Zhao, R., Miao, H., Song, W., Chen, C., and Zhang, H. (2018). Identification of sesame (*Sesamum indicum* L.) chromosomes using the BAC-FISH system. Plant Biol. **20**:85–92. https://doi.org/10.1111/plb.12647.