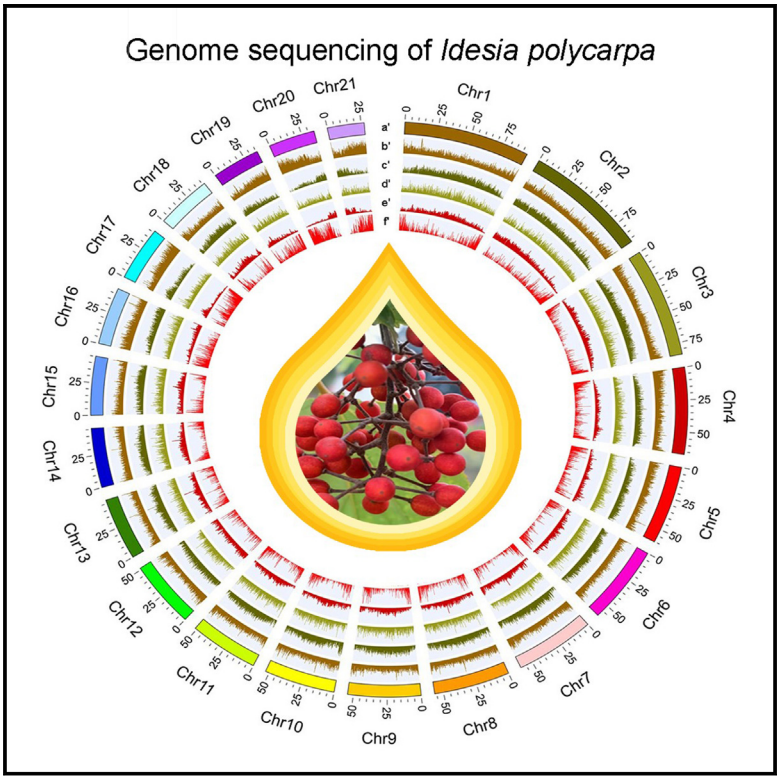
Cell Reports

Article

The *Idesia polycarpa* genome provides insights into its evolution and oil biosynthesis

Graphical abstract



Highlights

- A chromosome-level genome of *Idesia polycarpa* is presented
- *I. polycarpa* and *P. trichocarpa* share a common ancestor in *Salicaceae* family
- A potential oil-content-related gene, *SUGAR TRANSPORTER* 5 (*IpSTP5*), is identified

Authors

Yi Zuo, Hongbing Liu, Bin Li, ..., Minxian Wang, Chengzhi Liang, Lei Wang

Correspondence

wanglei@ibcas.ac.cn

In brief

Zuo et al. report a chromosome-level genome of *Idesia polycarpa*, a Chinese native deciduous tree, whose fruits can be used for extracting high-quality edible oil. Further, they identify a potential oil-content-related gene, *SUGAR TRANSPORTER* 5 (*IpSTP5*), by population analysis of 42 wild accessions.

Zuo et al., 2024, Cell Reports *43*, 113909
March 26, 2024
https://doi.org/10.1016/j.celrep.2024.113909

CellPress

Cell Reports

CellPress
OPEN ACCESS

Article

The *Idesia polycarpa* genome provides insights into its evolution and oil biosynthesis

Yi Zuo,[1,4,8] Hongbing Liu,[2,8] Bin Li,[1,4,8] Hang Zhao,[2] Xiuli Li,[2] Jiating Chen,[1,5] Lu Wang,[1] Qingbo Zheng,[1,4] Yuqing He,[1,3] Jiashuo Zhang,[1,5] Minxian Wang,[6] Chengzhi Liang,[7] and Lei Wang[1,3,4,5,9,*]

[1]Key Laboratory of Plant Molecular Physiology, Institute of Botany, Chinese Academy of Science, Beijing 100093, China
[2]Shenzhen Branch, Guangdong Laboratory for Lingnan Modern Agriculture, Genome Analysis Laboratory of the Ministry of Agriculture, Agricultural Genomics Institute at Shenzhen, Chinese Academy of Agricultural Sciences, Shenzhen 518124, China
[3]Academician Workstation of Agricultural High-Tech Industrial Area of the Yellow River Delta, National Center of Technology Innovation for Comprehensive Utilization of Saline-Alkali Land, Dongying 257300, China
[4]China National Botanical Garden, Beijing 100093, China
[5]University of Chinese Academy of Sciences, Beijing 100049, China
[6]CAS Key Laboratory of Genome Sciences and Information, Beijing Institute of Genomics, Chinese Academy of Sciences and China National Center for Bioinformation, Beijing 100101, China
[7]State Key Laboratory of Plant Genomics, Institute of Genetics and Developmental Biology, Innovation Academy for Seed Design, Chinese Academy of Sciences, Beijing 100101, China
[8]These authors contributed equally
[9]Lead contact
*Correspondence: wanglei@ibcas.ac.cn
https://doi.org/10.1016/j.celrep.2024.113909

SUMMARY

The deciduous tree *Idesia polycarpa* can provide premium edible oil with high polyunsaturated fatty acid contents. Here, we generate its high-quality reference genome, which is ~1.21 Gb, comprising 21 pseudochromosomes and 42,086 protein-coding genes. Phylogenetic and genomic synteny analyses show that it diverged with *Populus trichocarpa* about 16.28 million years ago. Notably, most fatty acid biosynthesis genes are not only increased in number in its genome but are also highly expressed in the fruits. Moreover, we identify, through genome-wide association analysis and RNA sequencing, the *I. polycarpa SUGAR TRANSPORTER 5* (*IpSTP5*) gene as a positive regulator of high oil accumulation in the fruits. Silencing of *IpSTP5* by virus-induced gene silencing causes a significant reduction of oil content in the fruits, suggesting it has the potential to be used as a molecular marker to breed the high-oil-content cultivars. Our results collectively lay the foundation for breeding the elite cultivars of *I. polycarpa*.

INTRODUCTION

The latest projections by the United Nations indicate that the global human population will reach 9.7 billion by 2050.[1] The food demands of these individuals and the popularity of fried foods will likely lead to a significant increase in the demand for edible oils. Edible oil consumption per capita is expected to be ~30 kg/year. However, climatic deterioration, biodiversity declination, carbon emissions, and asynchronous production are expected to threaten the stability of edible oil production and cause the price to sharply accelerate, at a compound annual growth rate of 7.6% between 2022 and 2029.[2] Meeting the rapidly growing demands for vegetable oil in an environmentally and economically sustainable manner is a worldwide challenge and has led the countries with the highest consumption rates to actively promote self-sufficiency. For example, India used agricultural policies to expand domestic oil palm plantations by 30-fold in 1991–2015 while maintaining biodiversity.[3] It is necessary to develop new "orphan" oil crops with unusual properties, especially deciduous trees, which can simultaneously increase plant biodiversity and carbon fixation.

Idesia polycarpa Maxim. var. *vestita* Diels (referred to here simply as *I. polycarpa*) belongs to the *Salicaceae* family, which includes the genera *Populus*, *Salix*, *Idesia*, and so on. In *Populus* and *Salix*, there are numerous species around the world and at various altitudes.[4,5] In contrast, *I. polycarpa* is the only species of monotypic genus *Idesia*, distributed in East Asia, such as in Korea, Japan, and southern China.[6] *I. polycarpa* is an ideal candidate woody oil species (Figure 1A), as it is commonly known as the "oil depot on trees" or "oil grape," and oil content in the dried fruits ranges from 20% to 45%[7] (Figures 1B and 1C). The ripe fruits are red and resemble bunches of grapes. They can produce 1.5–2.5 kg of oil per tree and 2.25–3.75 tons of oil per hectare.[7] Cold-pressed *I. polycarpa* oil has been used in foods in Sichuan Province since the Qing dynasty.[7] The chemical composition of the total crude oil is desirable, containing high content of linoleic acid (LA) (C18:2). This is a natural advantage of *I. polycarpa* compared to other woody oil plants, such as oil palm, olive, and oil Camellia[8]; LA consumption can not only

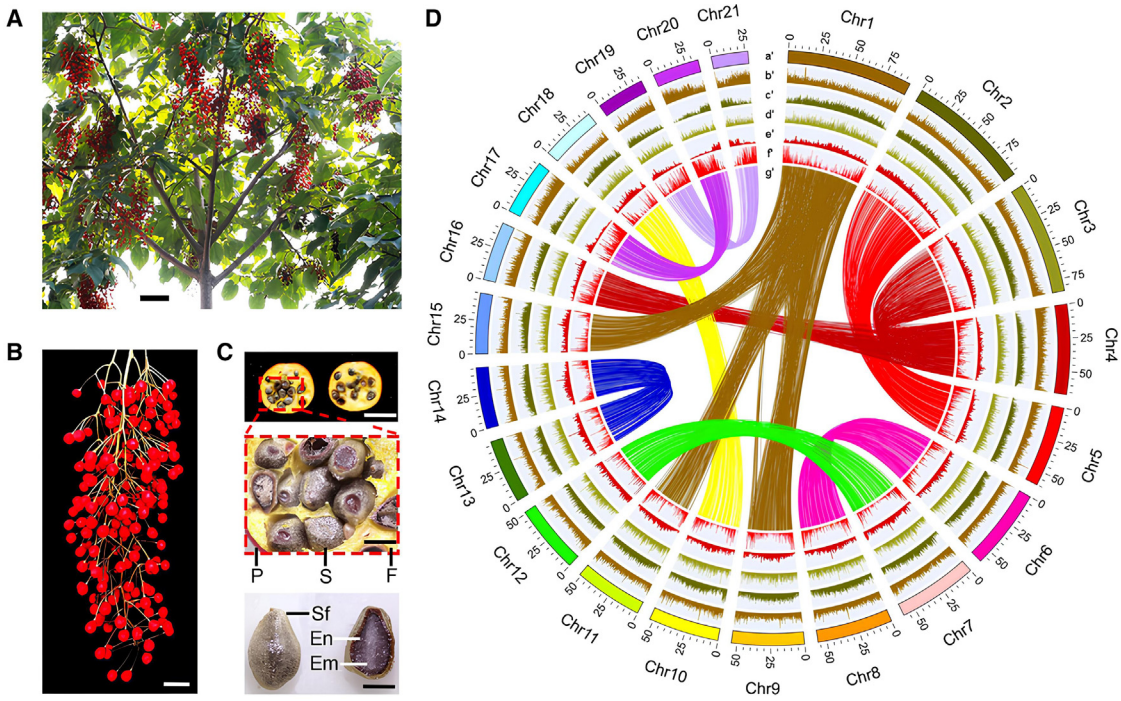


Figure 1. *I. polycarpa* morphology and genomic landscape

(A) General morphology of a representative *I. polycarpa* tree at the fruit stage. Scale bar: 40 cm.

(B) A ripe *I. polycarpa* spike. Scale bar: 3 cm.

(C) Cross-sections of representative *I. polycarpa* fruit and seeds. Top, fruit cross-section. The region outlined in dashed red is enlarged in the middle. Scale bar: 6 mm. Middle, enlarged image of the pericarp and seeds. Scale bar: 2 mm. Bottom, a coated seed intact (left) and in cross-section (right). Scale bar: 1 mm. P, pericarp; S, seed; F, flesh; Sf, seed film; En, endosperm; Em, embryo.

(D) Summary of the *I. polycarpa* genomic landscape. (a′) Length of each pesudochromosome in megabases (Mb). (b′) CG content (28%–35%). (c′) Number of repeat element (100–500). (d′) Number of *Copia* elements (0–72). (e′) Number of *Gypsy* elements (0–221). (f′) Number of genes (30–500). (g′) Syntenic blocks in homologous chromosomes. Each line represents a block longer than 2 kb. All data were based on 200-kb windows.

decrease low-density lipoprotein and increase high-density lipoprotein levels in the blood but can also improve glucose-insulin homeostasis and promote maintenance of a healthy weight.[9–12] *I. polycarpa* oil also contains squalene, volatile oils, vitamin E, and many phenolic substances, which are beneficial to human health. Due to these desirable characteristics, *I. polycarpa* edible oil was approved by the National Health Commission of China as a new edible woody oil in 2020.

Although oil biosynthesis pathways have been extensively characterized in many plant species,[13–15] the mechanisms by which high oil yield and premium oil content are achieved have yet to be fully elucidated. Comparative analyses of olive and sesame plants have uncovered that decreased expression of *FATTY ACID DESATURASE 2* (*FAD2*) and increased expression of *STEAROYL-ACYL CARRIER PROTEIN DESATURASE* (*SAD*) genes may promote exceptionally high oleic acid levels in olive.[16] In oil Camellia, elite alleles of genes such as *SUGAR-DEPENDENT TRIACYLGLYCEROL LIPASE 1*, *β-KETOACYL-ACYL CARRIER PROTEIN SYNTHASE III*, and *SAD*s have important roles in enhancing the production and quality of seed oil.[17] Although yield is usually positively related to oil content,[18] it is still a challenge to unravel the intrinsic factors determining plant organ size in relation to yield and quality. Thus far, more than 100 genes related to fruit size have been cloned, and some genetic loci were found to be involved in cell size and number through various pathways.[6,19] Importantly, fruit size is not only closely related to oil yield but is also related to photosynthetic rates and sugar transporter performance.[20,21] In *Glycine max* (soybean), GmSWEET10a can simultaneously increase soybean seed size and oil content.[22,23] In *Oryza sativa* (rice), the MATE transporter GFD1 can improve grain size through interactions with OsSUT2 and OsSWEET4a to optimize carbohydrate partitioning.[24] However, the genetic basis of fruit size and oil yield in *I. polycarpa* has remained unknown due to the lack of a high-quality reference genome.

We here generated a high-quality, chromosome-level reference genome for *I. polycarpa* by combining Pacific Biosciences (PacBio) and Hi-C data. The assembled genome was ~1.21 Gb in size; it included 42,086 predicted protein-coding genes, most of which were located on the 21 assembled chromosomes. *I. polycarpa* was determined to have diverged with *Populus trichocarpa* ~16.28 million years ago (Mya) and retained a 2-to-2 syntenic relationship with *P. trichocarpa* due to their shared salicoid duplication and re-diploiding events. Importantly, we identified an oil-content-related gene *SUGAR TRANSPORTER 5* (*IpSTP5*) by genome-wide association study (GWAS), RNA sequencing analysis, and virus-induced gene silencing (VIGS) assay. Together, the genome resources and the identified key

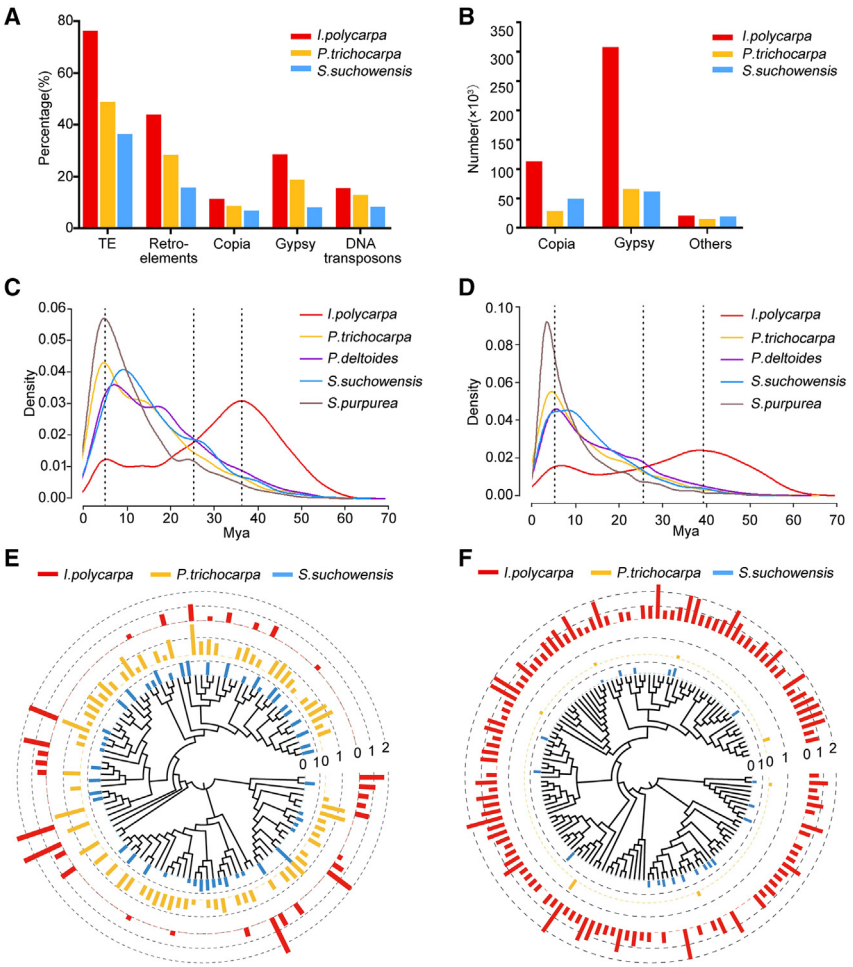


Figure 2. Characteristics of *Copia* and *Gypsy* retroelements in *I. polycarpa* and other woody plants in the family *Salicaceae*

(A) Distribution of repeat element types in *I. polycarpa* (red), *Populus trichocarpa* (yellow), and *Salix suchowensis* (blue).

(B) The number of intact retroelements in *I. polycarpa*, *P. trichocarpa*, and *S. suchowensis*.

(C and D) Expansion of *Copia* (C) and *Gypia* (D) retroelements in *I. polycarpa* and other species in the family *Salicaceae* (*P. trichocarpa*, *S. suchowensis*, *P. deltoides*, and *S. purpurea*).

(E and F) Phylogenetic trees showing expanded *Copia* (E) and *Gypsy* (F) retroelement families in *I. polycarpa*, *P. trichocarpa*, and *S. suchowensis* and the number of activated *Copia* and *Gypsy* elements in each. Bars are $\log_{10}$ (numbers).

genes could be used for future whole-genome selection breeding of the elite cultivars of *I. polycarpa*.

RESULTS

Genome sequencing and assembly

Based on flow cytometry results, the nuclear genome of female *I. polycarpa* was estimated to be ~1,030 Mb (Figure S1A). From the ~178.30 Gb of short-read sequencing data (representing ~144.95× coverage), *K*-mer analysis revealed the genome size and the heterozygosity ratio to be ~1.23 Gb and ~0.90%, respectively (Figure S1B). To obtain a high-quality genome, we generated an additional ~111.83 Gb of PacBio continuous long read (CLR) data (~90.92× coverage; N50 = 31.52 kb) and ~119.00 Gb of 150-bp paired-end Hi-C reads (~96.75× coverage) (Table S1A). The PacBio CLR reads were trimmed, corrected, and assembled with Canu (v.2.0) to form a genome of ~1.21 Gb, constituting 1,692 contigs (N50 = 5.82 Mb) (Table S1B). The longest contig was 61.48 Mb. Assessment with Benchmarking Universal Single-Copy Orthologs[25] revealed that the completeness was 97.5% (Figure S1C). To obtain a high-quality chromosome-level genome, these contigs were further anchored, ordered, and oriented with Juicer and 3D-DNA[26] using the Hi-C data. Approximately 1.14 Gb of the contig sequences (~94.5%) were anchored to 21 pseudochromosomes, consistent with the haploid chromosome number of *I. polycarpa* (2n = 42) (Figures 1D, S1D, and S1E).[27] Pseudochromosome lengths ranged from ~27.4 to 93.6 Mb and had an N50 value of 56.54 Mb. Chromosome numbers were assigned based on length in descending order (Figure 1D; Table S1C). The average GC content of *I. polycarpa* was 31.19%, slightly lower than that of *P. trichocarpa* (33.70%),[28] *Populus alba* (33.76%),[29] *Salix brachista* (34.15%),[5] and *Salix suchowensis* (34.9%).[30] The chloroplast genome was assembled into circular DNA molecules totaling 156,200 bp, which is similar to other species in *Salicaceae* (Figure S2). Collectively, these results suggested that the chromosome-scale assembly of the *I. polycarpa* genome generated here was of high quality (Figure 1D).

Repetitive sequence and gene annotation

Extensive *de novo* TE annotator, RepeatModeler, and RepeatMasker were used to identify transposable elements (TEs) and repeats in the genome.[31,32] The assembly contained 75.46% repeated regions, which was much higher than the proportion of repeated regions in the genomes of *Populus* and *Salix* species (41%–48% and 26%–42%, respectively).[5] The most abundant repeat type was long terminal repeats (LTRs), which accounted for 43.39% of the genome; *Copia* and *Gypsy* elements comprised 10.79% and 27.95%, respectively (Figure 2A; Table S2). Notably, *I. polycarpa* contained five times more *Copia* and *Gypsy* elements than *Populus* and *Salix*. Moreover, there were 305,157 *Gypsy* elements in *I. polycarpa*, nearly 10 times more than the number of other LTR types (Figure 2B). These results suggested that LTR expansion was the primary contributor to the enlarged *I. polycarpa* genome size.

To further explain the history of *Gypsy* and *Copia* elements in *I. polycarpa*, we calculated and compared their insertion times in

the *I. polycarpa* genome with those in four other important woody plants: *P. trichocarpa*,[28] *Populus deltoides*,[33] *S. suchowensis*,[30] and *Salix purpurea*.[8] There were 2,011 full-length *Copia* and 6,889 full-length *Gypsy* elements in *I. polycarpa*. As expected, the molecular paleontology of *Copia* and *Gypsy* elements revealed great differences between *Salicaceae* species. The period of most abundant expansion (the burst time) for *Copia* and *Gypsy* elements in *Populus* and *Salix* began ~25 Mya and reached a peak ~5–10 Mya. In contrast, *Copia* and *Gypsy* elements became active in *I. polycarpa* ~60 Mya and peaked at both ~35–40 and ~5 Mya. These results suggested that the *Copia* and *Gypsy* burst times were much earlier in *I. polycarpa* than in the other four species (Figures 2C and 2D) and that the burst process occurred slowly over a long period of time in this family. To further investigate expansion of the *Copia* and *Gypsy* repeat families in *I. polycarpa*, *P. trichocarpa*, and *S. suchowensis*, we clustered intact copies into subfamilies and reconstructed their phylogeny based on internal protein-coding sequences. *S. suchowensis* contained the most *Copia* families, whereas *I. polycarpa* contained the fewest *Copia* families but the largest number of subfamily members (Figure 2E). There were huge disparities between *I. polycarpa*, *P. trichocarpa*, and *S. suchowensis* in the number of *Gypsy* repeat subfamilies (119, 13, and 24, respectively) and in the number of subfamily members (7,064, 35, and 140, respectively) (Figure 2F). These results suggested that differences in LTR activity and accumulation were major contributors to dynamic changes in *Salicaceae* genome sizes over the course of evolutionary time.

To accurately annotate the protein-coding genes present in *I. polycarpa*, we performed RNA sequencing assay. Approximately 214.92 Gb of data corresponding to the major tissues were generated for this study and downloaded from publicly available datasets found on the National Center for Biotechnology Information (NCBI) website (Table S3A). We also obtained protein sequences for *P. trichocarpa*, *S. suchowensis*, *Vernicia fordii* (Tung tree), *Sesamum indicum* (sesame), and *Olea europaea* (olive tree) to generate homology-based predictions for *I. polycarpa* genes. Using an integrated MAKER annotation pipeline,[34] we annotated a total of 42,086 high-confidence protein-coding genes, 41,285 (98%) of which were present on the 21 pseudochromosomes. Gene density was greatly increased toward the distal ends of chromosome arms, whereas the opposite was true of LTR density (Figure 1D). The median transcript and protein lengths were 3,077 bp and 285 amino acids, respectively (Table S3B).

To functionally annotate the protein-coding genes in *I. polycarpa*, we first searched a representative protein for each gene (the longest putative protein) against the NCBI non-redundant database. Of the 42,086 genes, 38,341 (91.10%) could be annotated using this method. Of those annotated genes, 78.73% and 13.68% contained homologs in the genera *Populus* and *Salix,* respectively; 7.59% of the genes had homologs in other families, such as *Hevea* and *Manihot* (e.g., *Hevea brasiliensis* and *Manihot esculenta*) (Figure S3A). The representative proteins were also analyzed using the InterProScan database to identify gene families, domains, and functional sites; 38,124 (90.59%) of the predicted proteins could be annotated. Proteins were further annotated using the Panther, Pfam, Gene3d, and Superfamily databases, with which 35,833, 29,753, 25,531, and 23,543 proteins were annotated, respectively (Figure S3B; Table S3C). Moreover, 24,813 (~58.96%) of the proteins had Gene Ontology (GO) annotations. GO annotations are classified using three categories: biological processes, molecular functions, and cellular components; 43.94%, 86.09%, and 15.33% of the protein-coding genes had annotations in these categories, respectively (Figure S3C). Analysis of the Kyoto Encyclopedia of Genes and Genomes database revealed biochemical pathway annotations for 14,555 (~34.58%) of the genes. Data from the Plant Transcription Factor Database indicated that 2,473 of the genes encoded transcription factors (TFs) (Table S10A); this was comparable to the *P. trichocarpa* genome, in which 2,347 of the 34,699 protein-coding genes are TFs. The most abundant TF families were basic helix-loop-helix, MYB, ERF (Ethylene responsive factor), and NAC, accounting for 207, 206, 175, and 159 TFs, respectively; the HRT (Hordeum repressor transcription)-like, LFY (LEAFY), and NZZ/SPL (NOZZLE/SPOROCYTELESS) families were the least abundant. Compared to *P. trichocarpa,* the most expanded TF families in *I. polycarpa* were the M-type MADS family (39 vs. 68, respectively) and the GRAS (Gibberellic acid insensitive [GAI], Repressor of GAI, and SCARECROW) family (86 vs. 117, respectively), whereas the most contracted family was ARR-B (Arabidopsis response regulator type B) (14 vs. 7, respectively). In total, using multiple databases, 39,431 protein-coding genes were sufficiently annotated in *I. polycarpa*.

Gene family analysis

To accurately identify the evolutionary position of *I. polycarpa*, we clustered the genes encoding its 42,086 nuclear proteins with those of eight other species: *P. trichocarpa*, *S. suchowensis*, *V. fordii*, *Arabidopsis thaliana*, *O. sativa*, *S. indicum*, *O. europaea*, and *Medicago truncatula* (Figure 3A). Gene family analysis showed that the genes could be classified into 26,668 orthogroups (Table S4A). A phylogenetic tree constructed using these orthogroups revealed a much closer relationship between *I. polycarpa* and species in *Populus* or *Salix* than those from other genera. *Populus* and *Salix* diverged ~11.30 Mya[5]; we therefore estimated that *I. polycarpa* diverged from *Populus* and *Salix* ~16.28 Mya. These results were consistent with those from the gene functional annotation and repeat annotation analyses.

To identify the unique and shared gene families in *I. polycarpa*, we compared gene families between *I. polycarpa*, *P. trichocarpa*, *S. suchowensis*, and *V. fordii.* In this analysis, 120,414 genes were clustered into 18,494 gene families. Of those gene families, 12,046 were shared by the four species, whereas 1,244 gene families (comprising 3,872 genes) were unique to *I. polycarpa* (Figure 3B). GO enrichment analysis of the genes unique to *I. polycarpa* revealed that they were primarily enriched in GO terms related to source-phosphate synthase (SPS) activity (q value = 7.83e−06) and nucleosome assembly (q value = 4.28e−15) (Table S10B). SPS activity may enhance the ratio of sucrose to starch to increase the photosynthetic rate.[35] Nucleosome assembly may contribute to the basal chromatin dynamic of tissue-specific transcription regulation, apoptosis, histone shutting, and regulation.[36] Thus, these specific genes may contribute to fruit production and the comparatively large genome size of *I. polycarpa*.

Gene family expansions and contractions usually contribute to the dynamic evolution of metabolic, regulatory, and signaling

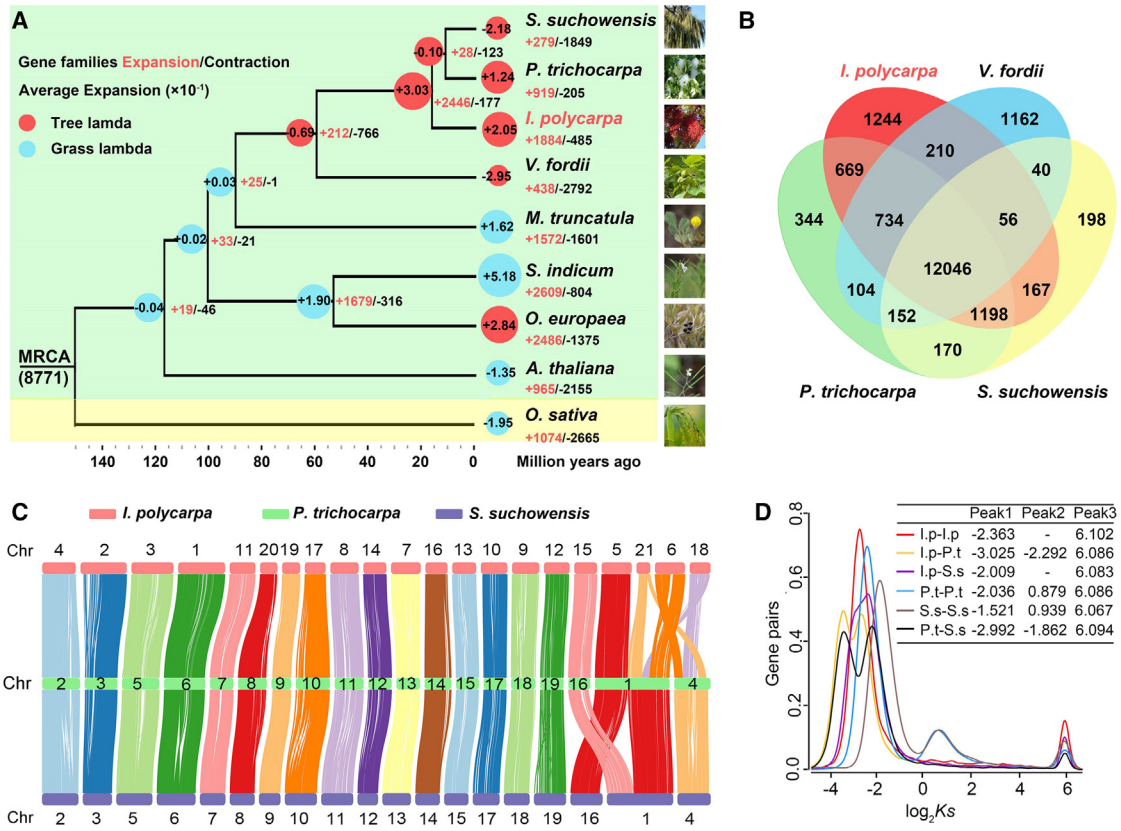


Figure 3. Genome evolution of *I. polycarpa*

(A) Phylogenetic tree showing relationships between *I. polycarpa*, *P. trichocarpa*, *S. suchowensis*, *Vernicia fordii*, *Arabidopsis thaliana*, *Oryza sativa*, *Sesamum indicum*, *Olea europaea*, and *Medicago truncatula*. The red and blue circles represent tree and grass lambda, respectively. The number in each circle represents the average expansion rate (×e−1).

(B) Shared and unique gene families in *I. polycarpa*, *P. trichocarpa*, *S. suchowensis*, and *V. fordii*.

(C) Syntenic comparison of homologous chromosomes in *I. polycarpa, S. suchowensis*, and *P. trichocarpa*.

(D) Distribution of the synonymous substitution rate (*Ks*) among homologous gene pairs in the family *Salicaceae*.

networks, and their functional diversification can further contribute to the adaptations required for life in terrestrial habitats. The most recent common ancestor of these nine species contained 8,771 gene families. After the divergence from *Populus*, 1,884 gene families expanded and 485 gene families contracted in *I. polycarpa*, representing nearly twice as many expanded and contracted gene families as there were in *Populus* after this divergence. Of the 1,884 expanded gene families, 116 (comprising 774 genes) were statistically significantly expanded compared to *Populus* ($p < 0.05$). GO analysis of the statistically significant expanded genes indicated that they were mainly enriched in the diterpenoid biosynthetic process (q value = 1.05e−32) and anther well tapetum development (q value = 1.01e−10) (Table S10C). In contrast, among the 485 contracted gene families, only 13 (comprising 70 genes) were significantly contracted; these were enriched in functions related to the glutathione metabolic process (q value = 4.39e−25) (Table S4B).

Salicaceae phylogeny

The ancestral genome of *Salicaceae* family was reported to have only nine chromosomes and to have undergone a whole-genome duplication (WGD) event in *Populus*, *Salix*, and *Idesia* genera.[27] However, the evolutionary trajectory leading to *Populus*, *Salix*, and especially *Idesia* remains a mystery. To clarify the evolutionary history, we first investigated syntenic relationships between *I. polycarpa* chromosomes and found clear signs of a WGD event in the genome (Figures 1D; Table S4C). Additionally, chromosome 9 (chr9), chr11, and chr15 all showed collinearity with chr1, and chr16 and chr3 showed collinearity with chr4. This indicated that the ancestral genome underwent a partial chromosomal replication event after the WGD.

To determine why the *I. polycarpa* genome had two more chromosomes (n = 21) than *P. trichocarpa* and *S. suchowensis* (n = 19) after speciation, we aligned the *I. polycarpa*, *P. trichocarpa*, and *S. suchowensis* genomes to identify syntenic regions (Figures 3C, S3D, and S3E). Despite the different chromosome numbers, there were strong 2-to-2 syntenic relationships between *I. polycarpa* and *P. trichocarpa* and between *I. polycarpa* and *S. suchowensis* that could be explained by the ancestral WGD event. There was also a clear 1-to-1 homologous relationship between the chromosomes of each genome. Interestingly, we found evidence of three pairs of chromosome

rearrangement events. *I. polycarpa* chr15 corresponded to *P. trichocarpa* chr16 and *S. suchowensis* chr1; *I. polycarpa* chr5 corresponded to *P. trichocarpa* chr1 and *S. suchowensis* chr16; and *I. polycarpa* chr6, chr18, and chr21 corresponded to *P. trichocarpa* and *S. suchowensis* chr1 and chr4. This indicated that chromosome rearrangements after the WGD event caused speciation of *I. polycarpa*, *P. trichocarpa*, and *S. suchowensis*. The largest chromosome fusion was located between *P. trichocarpa* chr1 (which was ~49.7 Mb) and *I. polycarpa* chr5, chr6, and chr21 (which was ~147.7 Mb in total). *P. trichocarpa* and *S. suchowensis* showed a chromosomal rearrangement event between chr1 and chr16, consistent with previous findings.[30]

We next estimated the timing of the WGD event by calculating the synonymous substitution rate (*Ks*) between paralogous genes. The *Ks* distributions were highly similar between *I. polycarpa* and *P. trichocarpa* and between *I. polycarpa* and *S. suchowensis*, suggesting that the *I. polycarpa* genome had similar levels of divergence from the *P. trichocarpa* and *S. suchowensis* genomes (Figure 3D). There was a sharp peak in the *P. trichocarpa Ks* distribution (−2.036) corresponding to ~32.3 Mya that was similar to peaks in *I. polycarpa* (−2.363) at ~25.67 Mya and in *S. suchowensis* (−1.521) at ~46.10 Mya. This indicated that the *Salicoid* duplication event was the major contributor to shared lineages in *Salicaceae*.[28] The different characteristic in evolutionary pace was caused by differences in the molecular clock ticking rate.[37] The last small common peak identified in the *Ks* distributions in these three species corresponded to the Eurosid duplication event, which is shared by all eudicots.[38]

Analysis of fatty acid biosynthesis pathway

As with olives, oil can be extracted directly from whole *I. polycarpa* fruits by physical squeezing. The maximum oil content in dried *I. polycarpa* fruits can reach 40%, much higher than the average oil content of olive fruits (25%).[39] Moreover, the ratios of fatty acid components differ between olive and *I. polycarpa* fruits; the former are enriched in oleic acid (~75% of the oil content), whereas the latter are enriched in LA (>70%).[16] To understand the molecular mechanisms underlying these characteristics of *I. polycarpa* oil, we analyzed key genes in the fatty acid biosynthesis pathway. Using OrthoFinder orthogroups analysis, we identified 133 genes in 15 key gene families that are involved in fatty acid biosynthesis. *I. polycarpa* contained 11 copies of SAD genes, which are responsible for synthesis of C18:1-ACP from C18:0-ACP. *I. polycarpa* thus had more SAD genes than that were found in olive (seven) or sesame (six) (Figure 4A; Table S5). Next, a gene regulation network was constructed to depict associations among the 133 genes. Highly connected hub genes were identified using the maximal clique centrality method in Cytoscape. The 10 most connected hub genes comprised two biotin carboxyl carriers (BCCPs), two β-ketoacyl-ACP reductases (FabGs), three β-ketoacyl-ACP synthases (KASs), one FabD, one FabZ1, and one FAB1 (Figures 4B and S4A). To investigate the expression patterns of genes involved in oil biosynthesis, we collected eight distinct *I. polycarpa* tissues and performed RNA sequencing analysis (Figure S4B; Table S3A). Notably, *IpFAD2_2* was expressed 1,000-fold higher in fruit-related tissues (namely the seed, pericarp, unripe fruit, and ripe fruit) compared to the vegetative tissues. In contrast, there were no significant differences in expression levels of the five *FAD3* genes between any of the tissues (Figures 4A, S4C, and S4D). This may partially explain the high LA and low α-linolenic acid levels measured in *I. polycarpa* fruits. Moreover, we found that *IpPDCT_1* and *IpPDAT_2*, which are involved in the lipid biosynthesis pathway, were highly expressed in fruits (Figure S4E).

I. polycarpa population structure and GWAS for oil content

Oil content is the most important agronomic trait in *I. polycarpa* but is highly variable among wild *I. polycarpa* trees. To identify the key genes that may control oil content in *I. polycarpa*, we collected 42 accessions (varying in oil content) from different geographic regions, including Sichuan, Tibet, Shanxi, and Beijing, and performed genome sequencing in each (Figure S5A). We generated ~471.5 Gb of 150-bp paired-end data, corresponding to an average sequencing depth of ~9.3× for each tree (Table S6A). After aligning all clean reads to our reference genome, we identified ~22.68 million single-nucleotide polymorphisms (SNPs) and ~2.47 million insertion/deletion (indel) mutations. The SNP distribution was unequal across each chromosome (Figure S5B; Table S6B). Gene structure annotation revealed that ~48.5% and ~40.3% of SNPs and indels, respectively, were in intergenic regions (Table S6C). Of the ~2.05% SNPs (n = 892,970) that located in coding regions, there were 490,971 missense, 11,524 nonsense, and 396,996 synonymous mutations.

The fruit oil content in the 42 trees ranged from 23% to 40% and showed a normal distribution (p = 0.225) (Figure 5A). Principal-component analysis and population structure results showed a clear division of these accessions into two groups (group 1 and group 2) based on all SNPs (Figures 5B and S5C). Interestingly, there was a significant difference in oil content between the two groups (25.1% in group 1 and 34.4% in group 2) (Figure 5C); this suggested that the oil content phenotype was closely correlated with the genotype and thus largely controlled by genetics rather than the environment. There was also a significant difference in genetic diversity (π) between the two groups (π_{group1} = 6.6e−04; π_{group2} = 4.4e−03), and based on a 2.5% significant selection, the cutoff ($\pi_{group2}/\pi_{group1}$) was 46, revealing 29.5 Mb of selective sweeps. This indicated that a much stronger genetic bottleneck had taken place in group 2 than in group 1 (Figure S5D). To further identify the genetic loci that may contribute most to oil content, we calculated the linkage disequilibrium (LD) decay and conducted a GWAS. The maximum LD decay rate was 0.42, and at a 150-kb decay distance, the LD decay rate was 0.10 (Figure S5E). A GWAS for oil content was conducted using the fixed and random model circulating probability unification method and clearly revealed strong signals ($-\log_{10}$(p value) > 5) in several chromosomes (Figure 5D; Table S7A). And the most significant locus *Qfoc4.1* was on chr4 (13.99–14.35 Mb), the 306 kb including LD decay distance on both sides, which contained 20 protein-coding genes (Figure 5E; Table S7B). This locus was therefore considered as a candidate for association with oil content in *I. polycarpa*.

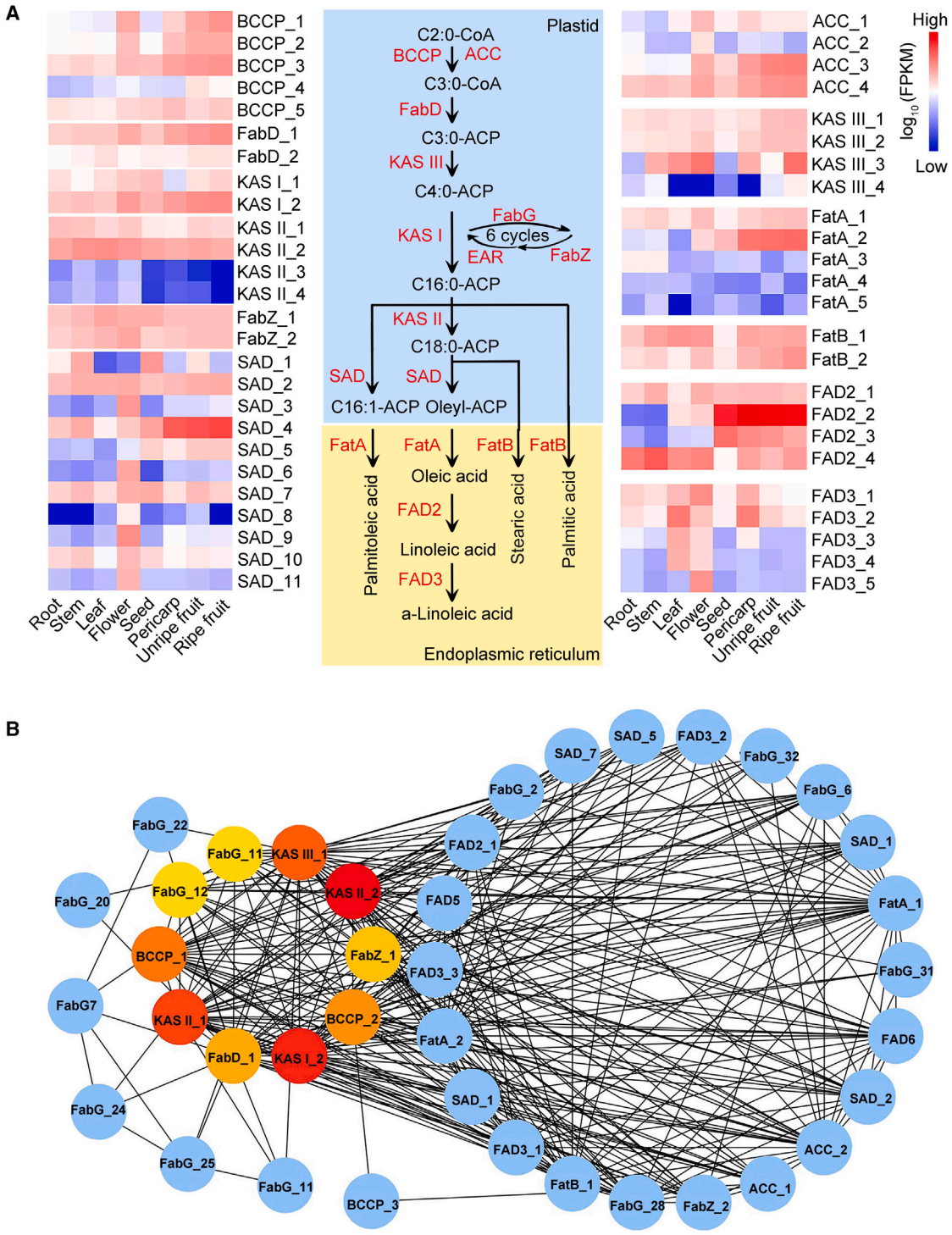


Figure 4. Expression patterns of putative fatty acid biosynthesis-related genes in *I. polycarpa*

(A) Center, diagram of the metabolic pathway for plant fatty acid biosynthesis. Blue indicates the plastid, and yellow indicates the endoplasmic reticulum. Left and right, heatmaps showing expression profiles of fatty acid biosynthesis genes in distinct tissues of *I. polycarpa*. Expression levels are shown in fragments per kilobase of transcript per million mapped reads (FPKM).

(B) Predicted interaction network of fatty acid metabolic genes in *I. polycarpa*. The 10 most connected hub genes are shaded from red to yellow to indicate the degrees of connectedness, which were calculated using the maximum clique centrality (MCC) method.

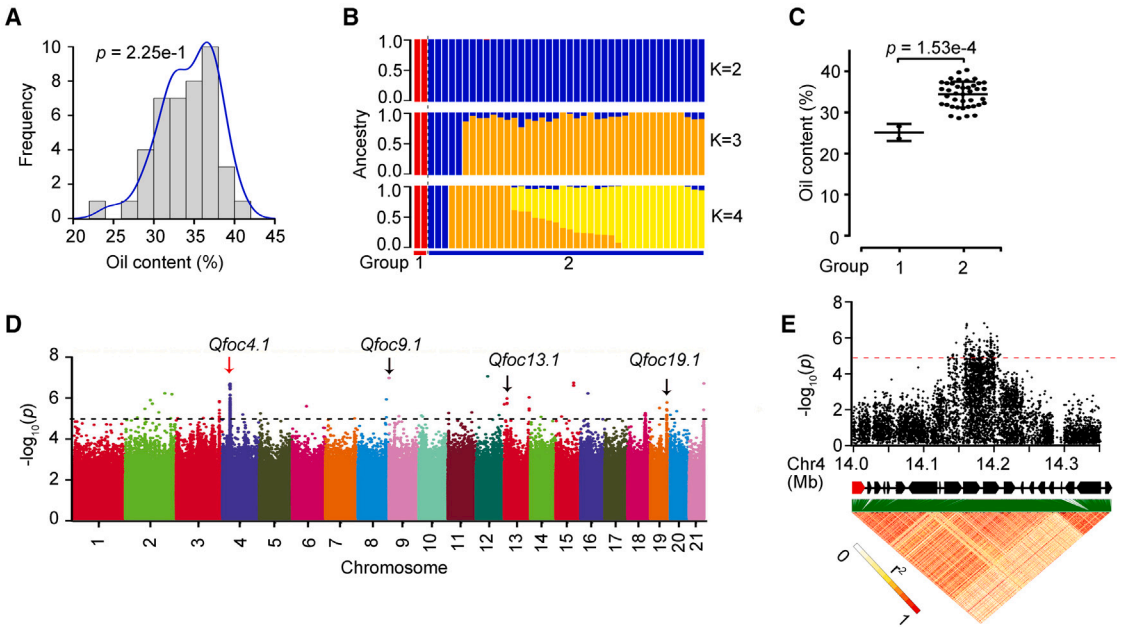


Figure 5. Genome-wide association study (GWAS) of oil content in *I. polycarpa*

(A) Distribution of oil content among 42 *I. polycarpa* accessions. The data were normally distributed (p = 2.25e−1, Shapiro-Wilk test).

(B) Genetic structure of the *I. polycarpa* population (at $K = 2–4$). All accessions are represented along the x axis, and the proportion of ancestry at each K value is shown along the y axis.

(C) Comparison of oil content between two groups of *I. polycarpa* accessions. p values were derived from a two-tailed analysis of variance (ANOVA).

(D) Manhattan plot showing results of the GWAS for oil content. The red arrow indicates the location of peak.

(E) Local Manhattan plot (top) and linkage disequilibrium plot (bottom) for single-nucleotide polymorphisms (SNPs) surrounding the peak on chr4. The location of *IpSTP5* is indicated in red.

Identification of candidate genes associated with high oil content

Within a GWAS candidate region, genes that are differentially expressed between individuals with variations in the phenotype of interest are likely to be causal factors for that phenotype. We therefore selected two representative *I. polycarpa* accessions, *I. polycarpa* cv. "China National Botanical Garden 01" (CNBG01) and *I. polycarpa* cv. "China National Botanical Garden 02" (CNBG02) for RNA sequencing. CNBG02 fruits were much darker green than CNBG01 fruits at the unripe stage (75 days after pollination [DAP]) (Figure 6A, top). The ripe fruits (127 DAP) also showed significant differences in width and length (Figures 6A, bottom, and 6B). Dried CNBG02 and CNBG01 fruits contained 40.2% and 22.0% oil content, respectively (Figure 6C). The fatty acid profiles were further analyzed by gas chromatography-mass spectrometry (GC-MS) method. Evidently, the contents of individual fatty acids were significant different between CNBG01 and CNBG02 (Figure 6D). In addition, the Person correlation coefficient between volume and oil content of fruit is 0.39 for the 42 accessions, indicating that fruit size is positively correlated with oil content in *I. polycarpa* (Figure 6E).

Further, we collected developing fruits from CNBG01 and CNBG02 at 25 DAP and analyzed them via RNA sequencing. Using a threshold of >2-fold change with a false discovery rate < 0.05, we identified 5,252 differentially expressed genes (DEGs) between accessions that were shared between the samples harvested in the morning and at night, including 2,406 up-regulated DEGs and 2,846 down-regulated DEGs in CNBG02 compared to CNBG01 (Figure S6A; Table S10D). GO analysis indicated that up-regulated DEGs in CNBG02 were enriched in photosynthesis, defense responses, transmembrane transport, and lipid metabolic processes, whereas down-regulated DEGs were enriched in peptide transport, response to wounding, and regulation of hormone levels (Figure S6B). Among the 20 candidate genes identified via GWAS, *Ip097893*/*IpSTP5* showed significantly higher expression in CNBG02 than in CNBG01 (Figure 6F). In addition, the correlation between expression of *IpSTP5* and oil content is positive (Figure 6G). Based on the polymorphisms in the two groups, *IpSTP5* could be classified into two major haplotypes in the studied accessions (Figures S7A and S7B). Moreover, the contents of LA, the most enriched fatty acid in oil composition of *I. polycarpa*, displayed the significant difference between two *IpSTP5* haplotypes among 42 accessions (Figure S7C; Table S8A). Functional analysis revealed that *IpSTP5* encoded a homolog of *SUGAR TRANSPORTER PROTEIN* (*STP*) in *Arabidopsis* and was classified into the transmembrane transport term (GO: 0055085).

To further investigate the function of *IpSTP5*, we performed VIGS assay in unripe fruits. Two different fragments of *IpSTP5* were constructed into pTRV2 vectors (Figure 6H). Evidently, the expression of *IpSTP5* was decreased in *IpSTP5* silencing fruits by VIGS compared to those of mock group (Figure 6I). Accordingly, the oil contents were also reduced in *IpSTP5* silencing fruits (Figure 6J). Further, the fatty acid profiles were analyzed by GC-MS in the *IpSTP5* silencing fruits. It was found

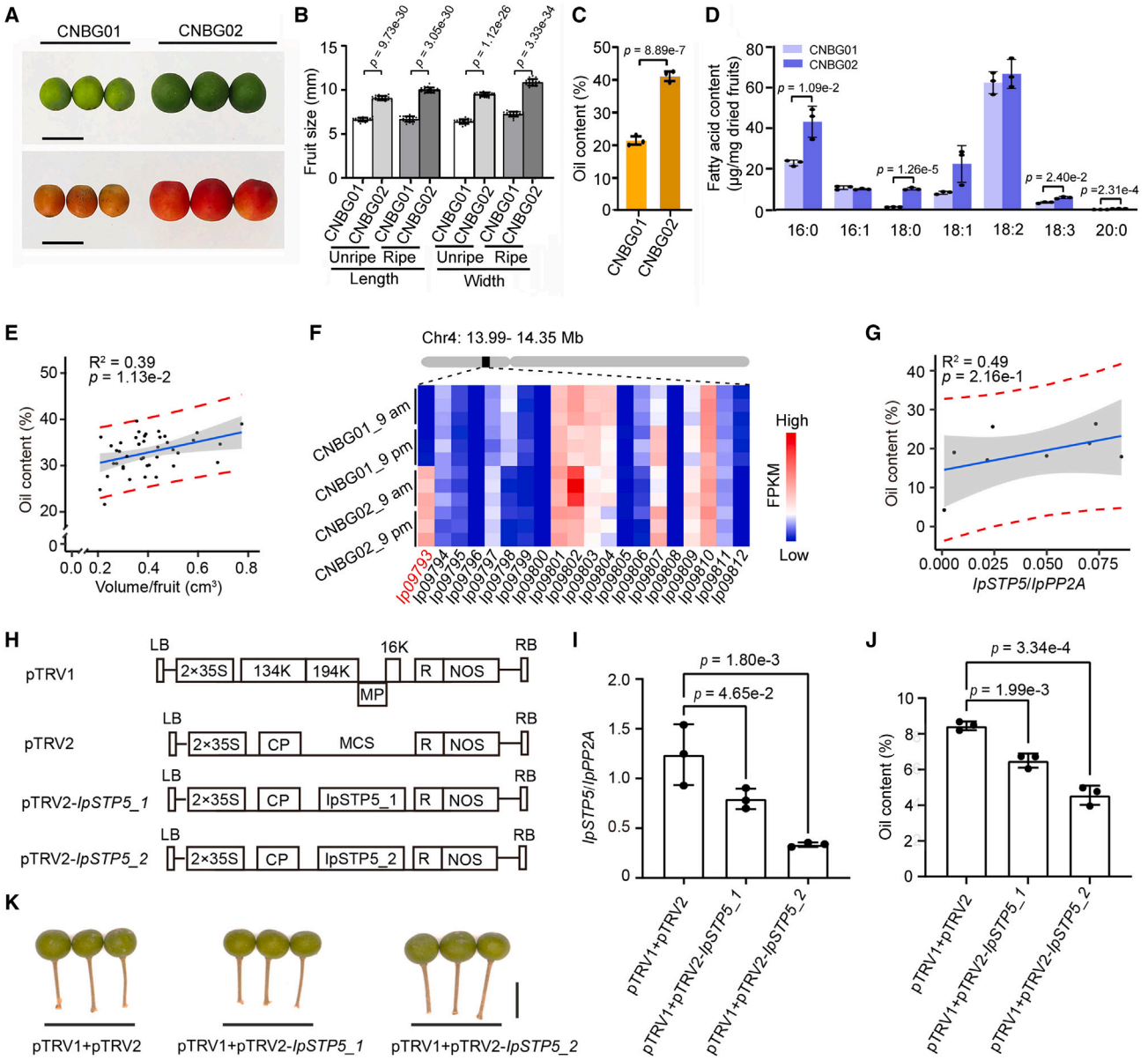


Figure 6. The genetic basis of oil content in *I. polycarpa*

(A) Fruit phenotypes of two accessions with varying oil content: China National Botanical Garden (CNBG) 01 and CNBG02. Top, unripe fruits at 75 days after pollination (DAP). Bottom, ripe fruits at 127 DAP. Scale bars: 1 cm.

(B) Statistical analysis of CNBG01 and CNBG02 fruit size at the unripe and ripe stages. Values are mean ± SEM (n = 20, p values were derived from ANOVA).

(C) Total oil content of dried fruits measured by Soxhlet extraction as expressed as a percentage (w/w) of total fruit weight. Values are mean ± SEM (n = 3, p value was derived from ANOVA).

(D) Content of individual fatty acids in dried fruits measured by GC-MS. Values are mean ± SEM (n = 3, p value was derived from ANOVA). Individual species of fatty acids (number of carbons:number of double bonds; 16:0, palmitic acid; 16:1, palmitoleic acid; 18:0, stearic acid; 18:1, oleic acid; 18:2, linoleic acid; 18:3, linolenic acid; 20:0, arachidic acid) are listed on the x axis.

(E) Correlation analysis between fruit volume and oil content in *I. polycarpa* (R^2 = 0.39).

(F) Heatmap showing expression of genes located in the candidate region of chr4 identified with a GWAS. *Ip09793*/*IpSTP5*, which was significantly differentially expressed between CNBG01 and CNBG02, is indicated in red text.

(G) Correlation analysis between *IpSTP5* expression levels and oil content in *I. polycarpa* (R^2 = 0.49).

(H) Map of pTRV1, pTRV2, pTRV2-*IpSTP5_1*, and pTRV2-*IpSTP5_2* vectors used in this study. pTRV2-*IpSTP5_1* was cloned from 596 to 1,026 bp, and pTRV2-*IpSTP5_2* was cloned from 1,123 to 1,521 bp.

(I) RT-qPCR assay of *IpSTP5* expression in unripe fruits after 14 days of inoculation. Values are mean ± SEM (n = 3, p values were derived from ANOVA).

(J) Statistic analysis of the oil content of unripe fruits after 14 days of inoculation. Values are mean ± SEM (n = 3, p values were derived from ANOVA).

(K) The unripe fruit phenotypes after 14 days of inoculation in *IpSTP5* VIGS assay. Scale bars: 1 cm.

the content of LA was decreased in the *IpSTP5* silencing fruits, indicating that *IpSTP5* also affected the fatty acid profile, likely in an indirect manner (Figure S7D; Table S8B). Nevertheless, the size of *IpSTP5* silencing fruits was comparable to those mock-treated fruits, suggesting that *IpSTP5* specifically affects the oil content and fatty acid profile in the developing fruits rather than the fruit size (Figures 6K and S7E).

Identification of an optimal control on fruit in VIGS assay

In order to generate a visualizable control in research fruit characterization, fruit size is one of the most intuitive phenotypes among numerously agronomic traits. To explore genes regulating fruit size in *I. polycarpa*, we conducted a literature search to identify genes known to be related to fruit size in other species (e.g., rice and tomato). Of the 164 genes identified from the literature, 136 corresponding genes were found in *I. polycarpa* via homologous alignment (Table S10E). The indole-3-acetic acid (IAA) glucose hydrolase gene *IpTGW6*, which has positive effects on increase IAA content,[40–42] was expressed at higher levels in CNBG02 than in CNBG01 at both 9 a.m. and 9 p.m. (Figure S8A). Meanwhile, we found that IAA levels were significantly higher in CNBG02 than in CNBG01 (Figure S8B). As IAA was a positive hormone at the fruit expanding stage,[43] we proposed that the larger fruit size of CNBG02 may be related to the higher expression of *IpTGW6*. To further understand the function of *IpTGW6* in *I. polycarpa*, we performed VIGS assay in unripe fruits. We found the fruit size of silenced *IpTGW6* was obviously smaller than that of mock (Figures S8C and S8D). Consistently, the transcript level of *IpTGW6* was significantly decreased (Figure S8E), while the oil content was not affected (Figure S8F), indicating that *IpTGW6* could be used as a marker for induced visible symptom on fruit in VIGS assay. Moreover, these data also suggested that *IpTGW6* might act positive regulator of fruit size.

DISCUSSION

I. polycarpa is an important member of the *Salicaceae* family that has high economic, biological, and ornamental value. Despite this importance, its genome size, chromosome number, and evolutionary trajectory have remained unknown, limiting its domestication and molecular design breeding. High-quality genome sequences and annotations build a solid foundation for various genetic techniques and analysis methods, comparative genomics, evolutionary studies, and dissection of the genomic architecture associated with traits of interest. In the present study, we report the first high-quality reference genome assembly for *I. polycarpa*. The quality was high, comparable to that of several other recently completed crop genome assemblies.[17,44] Gene family evolution analysis revealed a close relationship between *I. polycarpa* and both *P. trichocarpa* and *S. suchowensis*, with a divergence time of ~16.28 Mya. Interestingly, *I. polycarpa* had a much larger genome size (~1.21 Gb) than other species in the family *Salicaceae* (e.g., 390 Mb for *P. trichocarpa* and 356 Mb for *S. suchowensis*). This larger genome size was mainly due to drastic TE expansions, especially *Gypsy* elements. Syntenic analysis of homologous chromosomes indicated that there had been frequent fissure and fusion events between several chromosomes in all three species, explaining the origin of the additional *I. polycarpa* chromosomes. In addition, inner genome synteny analysis revealed the evolutionary history of *I. polycarpa* from the ancestral *Salicaceae* genome, which had nine chromosomes and underwent both WGD and partial chromosome duplication. Future assembly and analysis of related species with nine chromosomes, such as those in the genus *Scyphostegia*,[27] will shed additional light on the evolution of *Salicaceae* family.

We identified 1,244 gene families that were unique to *I. polycarpa*, and these were found to be significantly enriched in SPS activity (q value = 7.83e−06). This activity may enhance the ratio of sucrose to starch, increasing photosynthesis and contributing to fruit production. Moreover, 1,884 gene families were expanded in *I. polycarpa* compared to *P. trichocarpa* and *S. suchowensis*, and these were enriched in the diterpenoid biosynthetic process (q value = 1.05e−32); this process is closely related to hormone biosynthesis and may regulate fruit oil content. Collectively, these results not only reveal valuable information about the evolutionary history of the family *Salicaceae* but will also expedite future genetic studies of woody plants in this family and beyond.

Recently, core fatty acid biosynthesis genes and pathways have been extensively studied in major oil crops such as sesame, olive, and soybean. However, there is still a large knowledge gap in the genetic factors that finely regulate oil content, especially in woody plants such as the deciduous tree *I. polycarpa*. Using the newly generated and fully annotated reference genome, we comprehensively identified expansions and contractions in oil biosynthesis-related gene families in *I. polycarpa*. Furthermore, by combining GWASs for oil content using 42 accessions and the transcriptomes of two representative materials with large differences in fruit oil content, we identified several key candidate genes that may contribute to the high oil content in *I. polycarpa* fruits. One of these genes was *IpSTP5*, which encodes a sugar transport protein. RNA sequencing data suggested that coregulation of *IpFAD2* and *IpSAD* expression may improve oil quality. The genes identified here are strong candidates for future use in genomic selective breeding of *I. polycarpa*.

In woody oil plants, oil content and quality are complex agronomic traits in practical production. They are affected by genome structure, regulation, and modification and by environmental signals, including light, water, humidity, and temperature.[45] Here, we generated re-sequencing data for representative natural *I. polycarpa* accessions, which will serve as an important resource; such haplotype data have potential applications in genome manipulation, trait discovery and allele mining, and plant genetic improvement. The genomic data presented here for *I. polycarpa* enhance our understanding of genetic and genomic characteristics, allowing researchers to dissect the trait domestication that underlies different selection programs in closely related plant species. Insights regarding the genetic diversity and population structure of *I. polycarpa* accessions, as well as the genes and chromosomal regions that have been subject to human selection, will shape future efforts in genetic research and breeding.

Limitations of the study

I. polycarpa can provide premium edible oil, but it is still not domesticated and not fully utilized by human due to the lack of genome and key target genes for genome-wide selection

breeding. Here, we report its chromosome-level genome and identity a fruit-oil-related gene, *IpSTP5*, through population analysis and functional investigation. Nevertheless, the genetic and biochemical characterization of *IpSTP5* requires further exploration, especially investigating *IpSTP5* function genetically. To address this, the gene-edited alleles or over-expressing lines of *IpSTP5* in *I. polycarpa* are required. However, the transgenic system of *I. polycarpa* is still not available, which needs to be resolved in the near future. Moreover, more *I. polycarpa* accessions with rich genetic and phenotype diversities are required to be collected worldwide for systematically mapping the quantitative trait loci of oil content and accumulation.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead contact
  - Materials availability
  - Data and code availability
- EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS
  - Plant materials and growing conditions
- METHOD DETAILS
  - Plant materials
  - DNB, Pacbio, and Hi-C sequencing
  - Genome size evaluation
  - Genome assembly and pseudomolecule construction by Hi-C
  - Genome repeat element identification
  - Gene annotation and gene family analysis
  - Phylogeny and synteny analysis of *I. polycarpa*
  - Population analysis
  - RNA extraction and RT-qPCR
  - Transcriptome analysis
  - Oil content determination
  - Fatty acid composition determination
  - Auxin content determination
  - VIGS assay

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.celrep.2024.113909.

ACKNOWLEDGMENTS

We thank Dr. Suhua Yang and Dr. Bin Han from the Public Technology Center of the Institute of Botany, Chinese Academy of Sciences, for excellent technical assistance on flow cytometry determination, respectively. We also thank Dr. Chenxu Liu from China Agriculture University for his excellent technical assistance on oil content by low-field nuclear magnetic resonance. This work was supported by Science & Technology Specific Projects in Agricultural High-tech Industrial Demonstration Area of the Yellow River Delta (2022SZX13), the Rural Revitalization Projects of the CAS (KFJ-XCZX-202301), and the National Natural Science Foundation of China (nos. 32002066 and 32370307).

AUTHOR CONTRIBUTIONS

Y.Z. and H.L. analyzed the data and wrote the article. B.L. performed material collection and RNA sequencing. H.Z. and X.L. analyzed the data. J.C. performed VIGS assay. Lu Wang performed fatty acid determination. Q.Z., Y.H., and J.Z. performed DNA extraction. M.W., C.L., and Lei Wang revised the manuscript. Lei Wang agrees to serve as the author responsible for contact and ensures communication.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: June 29, 2023
Revised: January 26, 2024
Accepted: February 20, 2024

REFERENCES

1. Modolo, L.V., da-Silva, C.J., Brandão, D.S., and Chaves, I.S. (2018). A minireview on what we have learned about urease inhibitors of agricultural interest since mid-2000s. J. Adv. Res. *13*, 29–37.
2. Zhou, Y., Zhao, W., Lai, Y., Zhang, B., and Zhang, D. (2020). Edible plant oil: global status, health issues, and perspectives. Front. Plant Sci. *11*, 1315.
3. Srinivasan, U., Velho, N., Lee, J.S.H., Chiarelli, D.D., Davis, K.F., and Wilcove, D.S. (2021). Oil palm cultivation can be expanded while sparing biodiversity in India. Nat. Food *2*, 442–447.
4. Lin, Y.C., Wang, J., Delhomme, N., Schiffthaler, B., Sundström, G., Zuccolo, A., Nystedt, B., Hvidsten, T.R., de la Torre, A., Cossu, R.M., et al. (2018). Functional and evolutionary genomic inferences in *Populus* through genome and population sequencing of American and European aspen. Proc. Natl. Acad. Sci. USA *115*, E10970–E10978.
5. Chen, J.H., Huang, Y., Brachi, B., Yun, Q.Z., Zhang, W., Lu, W., Li, H.N., Li, W.Q., Sun, X.D., Wang, G.Y., et al. (2019). Genome-wide analysis of Cushion willow provides insights into alpine plant divergence in a biodiversity hotspot. Nat. Commun. *10*, 5230.
6. Li, N., and Li, Y. (2016). Signaling pathways of seed size control in plants. Curr. Opin. Plant Biol. *33*, 23–32.
7. Yang, F.X., Su, Y.Q., Li, X.H., Zhang, Q., and Sun, R.C. (2009). Preparation of biodiesel from *Idesia polycarpa* var. vestita fruit oil. Ind. Crops Prod. *29*, 622–628.
8. Zhou, R., Macaya Sanz, D., Carlson, C.H., Schmutz, J., Jenkins, J.W., Kudrna, D., Sharma, A., Sandor, L., Shu, S., Barry, K., et al. (2020). A willow sex chromosome reveals convergent evolution of complex palindromic repeats. Genome Biol. *21*, 38.
9. Katan, M.B., Zock, P.L., and Mensink, R.P. (1994). Effects of fats and fatty acids on blood lipids in humans: an overview. Am. J. Clin. Nutr. *60*, 1017s–1022s.
10. Klok, A.J., Lamers, P.P., Martens, D.E., Draaisma, R.B., and Wijffels, R.H. (2014). Edible oils from microalgae: insights in TAG accumulation. Trends Biotechnol. *32*, 521–528.
11. Imamura, F., Micha, R., Wu, J.H.Y., de Oliveira Otto, M.C., Otite, F.O., Abioye, A.I., and Mozaffarian, D. (2016). Effects of saturated fat, polyunsaturated fat, monounsaturated fat, and carbohydrate on glucose-insulin homeostasis: a systematic review and meta-analysis of randomised controlled feeding trials. PLoS Med. *13*, e1002087.
12. Marklund, M., Wu, J.H.Y., Imamura, F., Del Gobbo, L.C., Fretts, A., de Goede, J., Shi, P., Tintle, N., Wennberg, M., Aslibekyan, S., et al. (2019). Biomarkers of dietary Omega-6 fatty acids and incident cardiovascular disease and mortality. Circulation *139*, 2422–2436.

13. 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.

14. Chen, X., Li, H., Pandey, M.K., Yang, Q., Wang, X., Garg, V., Li, H., Chi, X., Doddamani, D., Hong, Y., et al. (2016). Draft genome of the peanut A-genome progenitor (*Arachis duranensis*) provides insights into geocarpy, oil biosynthesis, and allergens. Proc. Natl. Acad. Sci. USA *113*, 6785–6790.

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

16. 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.

17. Lin, P., Wang, K., Wang, Y., Hu, Z., Yan, C., Huang, H., Ma, X., Cao, Y., Long, W., Liu, W., et al. (2022). The genome of oil-Camellia and population genomics analysis provide insights into seed oil domestication. Genome Biol. *23*, 14.

18. Yang, X., Ma, H., Zhang, P., Yan, J., Guo, Y., Song, T., and Li, J. (2012). Characterization of QTL for oil content in maize kernel. Theor. Appl. Genet. *125*, 1169–1179.

19. Mauxion, J.P., Chevalier, C., and Gonzalez, N. (2021). Complex cellular and molecular events determining fruit size. Trends Plant Sci. *26*, 1023–1038.

20. Long, S.P., Taylor, S.H., Burgess, S.J., Carmo-Silva, E., Lawson, T., De Souza, A.P., Leonelli, L., and Wang, Y. (2022). Into the shadows and back into sunlight: photosynthesis in fluctuating light. Annu. Rev. Plant Biol. *73*, 617–648.

21. Wen, S., Neuhaus, H.E., Cheng, J., and Bie, Z. (2022). Contributions of sugar transporters to crop yield and fruit quality. J. Exp. Bot. *73*, 2275–2289.

22. Wang, S., Liu, S., Wang, J., Yokosho, K., Zhou, B., Yu, Y.C., Liu, Z., Frommer, W.B., Ma, J.F., Chen, L.Q., et al. (2020). Simultaneous changes in seed size, oil content and protein content driven by selection of homologues during soybean domestication. Natl. Sci. Rev. *7*, 1776–1786.

23. Duan, Z., Zhang, M., Zhang, Z., Liang, S., Fan, L., Yang, X., Yuan, Y., Pan, Y., Zhou, G., Liu, S., and Tian, Z. (2022). Natural allelic variation of *GmST05* controlling seed size and quality in soybean. Plant Biotechnol. J. *20*, 1807–1818.

24. Sun, C., Wang, Y., Yang, X., Tang, L., Wan, C., Liu, J., Chen, C., Zhang, H., He, C., Liu, C., et al. (2023). MATE transporter GFD1 cooperates with sugar transporters, mediates carbohydrate partitioning and controls grain-filling duration, grain size and number in rice. Plant Biotechnol. J. *21*, 621–634.

25. Manni, M., Berkeley, M.R., Seppey, M., Simão, F.A., and Zdobnov, E.M. (2021). BUSCO update: novel and streamlined workflows along with broader and deeper phylogenetic coverage for scoring of eukaryotic, prokaryotic, and viral Genomes. Mol. Biol. Evol. *38*, 4647–4654.

26. 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.

27. Zhang, Z.S., Zeng, Q.Y., and Liu, Y.J. (2021). Frequent ploidy changes in Salicaceae indicates widespread sharing of the salicoid whole genome duplication by the relatives of *Populus* L. and *Salix* L. BMC Plant Biol. *21*, 535.

28. Tuskan, G.A., Difazio, S., Jansson, S., Bohlmann, J., Grigoriev, I., Hellsten, U., Putnam, N., Ralph, S., Rombauts, S., Salamov, A., et al. (2006). The genome of black cottonwood, *Populus trichocarpa* (Torr. & Gray). Science *313*, 1596–1604.

29. Liu, Y.J., Wang, X.R., and Zeng, Q.Y. (2019). *De novo* assembly of white poplar genome and genetic diversity of white poplar population in Irtysh River basin in China. Sci. China Life Sci. *62*, 609–618.

30. Wei, S., Yang, Y., and Yin, T. (2020). The chromosome-scale assembly of the willow genome provides insight into Salicaceae genome evolution. Hortic. Res. *7*, 45.

31. Wheeler, T.J., Clements, J., Eddy, S.R., Hubley, R., Jones, T.A., Jurka, J., Smit, A.F.A., and Finn, R.D. (2013). Dfam: a database of repetitive DNA based on profile hidden Markov models. Nucleic Acids Res. *41*, D70–D82.

32. Ou, S., Su, W., Liao, Y., Chougule, K., Agda, J.R.A., Hellinga, A.J., Lugo, C.S.B., Elliott, T.A., Ware, D., Peterson, T., et al. (2019). Benchmarking transposable element annotation methods for creation of a streamlined, comprehensive pipeline. Genome Biol. *20*, 275.

33. Xue, L., Wu, H., Chen, Y., Li, X., Hou, J., Lu, J., Wei, S., Dai, X., Olson, M.S., Liu, J., et al. (2020). Evidences for a role of two Y-specific genes in sex determination in *Populus deltoides*. Nat. Commun. *11*, 5893.

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

35. Babb, V.M., and Haigler, C.H. (2001). Sucrose phosphate synthase activity rises in correlation with high-rate cellulose synthesis in three heterotrophic systems. Plant Physiol. *127*, 1234–1242.

36. Ramirez-Parra, E., and Gutierrez, C. (2007). The many faces of chromatin assembly factor 1. Trends Plant Sci. *12*, 570–576.

37. Dai, X., Hu, Q., Cai, Q., Feng, K., Ye, N., Tuskan, G.A., Milne, R., Chen, Y., Wan, Z., Wang, Z., et al. (2014). The willow genome and divergent evolution from poplar after the common genome duplication. Cell Res. *24*, 1274–1277.

38. Jiao, Y., Wickett, N.J., Ayyampalayam, S., Chanderbali, A.S., Landherr, L., Ralph, P.E., Tomsho, L.P., Hu, Y., Liang, H., Soltis, P.S., et al. (2011). Ancestral polyploidy in seed plants and angiosperms. Nature *473*, 97–100.

39. Žanetić, M., Jukić Špika, M., Ožić, M.M., and Brkić Bubola, K. (2021). Comparative study of volatile compounds and sensory characteristics of dalmatian monovarietal virgin olive oils. Plants *10*, 1995.

40. Ishimaru, K., Hirotsu, N., Madoka, Y., Murakami, N., Hara, N., Onodera, H., Kashiwagi, T., Ujiie, K., Shimizu, B.I., Onishi, A., et al. (2013). Loss of function of the IAA-glucose hydrolase gene *TGW6* enhances rice grain weight and increases yield. Nat. Genet. *45*, 707–711.

41. Shen, H., Zhong, X., Zhao, F., Wang, Y., Yan, B., Li, Q., Chen, G., Mao, B., Wang, J., Li, Y., et al. (2015). Overexpression of receptor-like kinase ERECTA improves thermotolerance in rice and tomato. Nat. Biotechnol. *33*, 996–1003.

42. Leng, Y.J., Yao, Y.S., Yang, K.Z., Wu, P.X., Xia, Y.X., Zuo, C.R., Luo, J.H., Wang, P., Liu, Y.Y., Zhang, X.Q., et al. (2022). *Arabidopsis* ERdj3B coordinates with ERECTA-family receptor kinases to regulate ovule development and the heat stress response. Plant Cell *34*, 3665–3684.

43. Liao, X., Li, M., Liu, B., Yan, M., Yu, X., Zi, H., Liu, R., and Yamamuro, C. (2018). Interlinked regulatory loops of ABA catabolism and biosynthesis coordinate fruit growth and ripening in woodland strawberry. Proc. Natl. Acad. Sci. USA *115*, E11542–E11550.

44. Wang, L., Lee, M., Sun, F., Song, Z., Yang, Z., and Yue, G.H. (2022). A chromosome-level genome assembly of chia provides insights into high omega-3 content and coat color variation of its seeds. Plant Commun. *3*, 100326.

45. Yang, Y., Kong, Q., Lim, A.R.Q., Lu, S., Zhao, H., Guo, L., Yuan, L., and Ma, W. (2022). Transcriptional regulation of oil biosynthesis in seed plants: current understanding, applications, and perspectives. Plant Commun. *3*, 100328.

46. 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.

47. Li, H. (2013). Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. Preprint at arXiv 1301.3997. https://doi.org/10.48550/arXiv.1303.3997.

48. Mendes, F.K., Vanderpool, D., Fulton, B., and Hahn, M.W. (2021). CAFE 5 models variation in evolutionary rates among gene families. Bioinformatics *36*, 5516–5518.

49. Koren, S., Walenz, B.P., Berlin, K., Miller, J.R., Bergman, N.H., and Phillippy, A.M. (2017). Canu: scalable and accurate long-read assembly via adaptive -mer weighting and repeat separation. Genome Res. *27*, 722–736.

50. Buchfink, B., Xie, C., and Huson, D.H. (2015). Fast and sensitive protein alignment using DIAMOND. Nat. Methods *12*, 59–60.

51. Robinson, M.D., McCarthy, D.J., and Smyth, G.K. (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics *26*, 139–140.

52. Wang, J., and Zhang, Z. (2021). GAPIT version 3: boosting power and accuracy for genomic association and prediction. Dev. Reprod. Biol. *19*, 629–640.

53. Zhang, R., Jia, G., and Diao, X. (2023). geneHapR: an R package for gene haplotypic statistics and visualization. BMC Bioinf. *24*, 199.

54. Feng, C., Wang, X., Wu, S., Ning, W., Song, B., Yan, J., and Cheng, S. (2022). HAPPE: a tool for population haplotype analysis and visualization in editable excel tables. Front. Plant Sci. *13*, 927407.

55. Pertea, M., Kim, D., Pertea, G.M., Leek, J.T., and Salzberg, S.L. (2016). Transcript-level expression analysis of RNA-seq experiments with HISAT, StringTie and Ballgown. Nat. Protoc. *11*, 1650–1667.

56. Zdobnov, E.M., and Apweiler, R. (2001). InterProScan an integration platform for the signature-recognition methods in InterPro. Bioinformatics *17*, 847–848.

57. Campbell, M.S., Holt, C., Moore, B., and Yandell, M. (2014). Genome annotation and curation using MAKER and MAKER-P. Curr. Protoc. Bioinformatics *48*, 4.11.1–4.11.39.

58. 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.

59. Kumar, S., Stecher, G., Li, M., Knyaz, C., and Tamura, K. (2018). MEGA X: Molecular evolutionary genetics analysis across computing platforms. Mol. Biol. Evol. *35*, 1547–1549.

60. He, W., Yang, J., Jing, Y., Xu, L., Yu, K., and Fang, X. (2023). NGenomeSyn: an easy-to-use and flexible tool for publication-ready visualization of syntenic relationships across multiple genomes. Bioinformatics *39*. btad.121.

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

62. Flynn, J.M., Hubley, R., Goubert, C., Rosen, J., Clark, A.G., Feschotte, C., and Smit, A.F. (2020). RepeatModeler2 for automated genomic discovery of transposable element families. Proc. Natl. Acad. Sci. USA *117*, 9451–9457.

63. Miele, V., Penel, S., and Duret, L. (2011). Ultra-fast sequence clustering from similarity networks with SiLiX. BMC Bioinf. *12*, 116.

64. Hedges, S.B., Dudley, J., and Kumar, S. (2006). TimeTree: a public knowledge-base of divergence times among organisms. Bioinformatics *22*, 2971–2972.

65. Lu, S., Yao, S., Wang, G., Guo, L., Zhou, Y., Hong, Y., and Wang, X. (2016). Phospholipase Dε enhances Braasca napus growth and seed production in response to nitrogen availability. Plant Biotechnol. J. *14*, 926–937.

66. Kim, S.C., Edgeworth, K.N., Nusinow, D.A., and Wang, X. (2023). Circadian clock factors regulate the first condensation reaction of fatty acid synthesis in *Arabidopsis*. Cell Rep. *42*, 113483.

67. Guo, Z., Yue, X., Cui, X., Song, L., and Cheng, Y. (2020). *AtMOB1* genes regulate jasmonate accumulation and plant development. Plant Physiol. *182*, 1481–1493.

68. Liu, Y., Schiff, M., Marathe, R., and Dinesh-Kumar, S.P. (2002). Tobacco Rar1, EDS1 and NPR1/NIM1 like genes are required for N-mediated resistance to tobacco mosaic virus. Plant J. *30*, 415–429.

69. Liu, C., Li, J., Chen, M., Li, W., Zhong, Y., Dong, X., Xu, X., Chen, C., Tian, X., and Chen, S. (2022). Development of high-oil maize haploid inducer with a novel phenotyping strategy. Crop J. *10*, 524–531.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Bacterial and virus strains		
Escherichia coli DH5α	Lab owned	N/A
Agrobacterium GV3101	Beijing Genes and Biotech	SAC21
Biological samples		
I. ploycarpa: root	This study	N/A
I. ploycarpa: stem	This study	N/A
I. ploycarpa: leave	This study	N/A
I. ploycarpa: flower	This study	N/A
I. polycarpa: unripe fruit	This study	N/A
I. polycarpa: ripe fruit	This study	N/A
Chemicals, peptides, and recombinant proteins		
*Bam*HI	NEB	R0136V
Xho I	NEB	R0146V
TRIzol Reagent	Thermo Fisher Scientific	Cat#15596018
Critical commercial assays		
KOD FX	TOYOBO	KFX-101
Clone Express® II One Step Cloning Kit	Vazyme	C112-02
SYBR® Green Real-time PCR Master Mix	TOYOBO	QPK-201
PrimeScript RT reagent kit	Takara	#RR047A
Palmitic acid	Sigma-Aldrich	P5585
Palmitoleic acid	Sigma-Aldrich	P9417
Stearic acid	Sigma-Aldrich	S4751
Oleic acid	Sigma-Aldrich	O1008
Linoleic acid	Sigma-Aldrich	L1376
Linolenic acid	Sigma-Aldrich	L2376
Arachidic acid	Sigma-Aldrich	10930
Nonadecanoic Acid	Sigma-Aldrich	72332
Butylated hydroxytoluene	Sigma-Aldrich	660337
Deposited data		
Genome of *I. polycarpa*	National Genomics Data Center	GWHBQLL00000000
Re-sequence data	National Genomics Data Center	CRA009397
RNA-seq	National Genomics Data Center	CRA009483
GC-MS	National Genomics Data Center	OMIX005749
Experimental models: Organisms/strains		
I. polycarpa: unripe fruit	This study	N/A
Oligonucleotides		
IpSTP5-F	GTGTCTACGGGTATGGTTGA	N/A
IpSTP5-R	TAATATGCCGTAGCCCTTGG	N/A
IpTGW6-F	GTAGGGCCAGAGAGCTTGGT	N/A
IpTGW6-R	AAGACCCAAGTAGGCATCAGC	N/A
Other primers see Table S9	This study	N/A
Recombinant DNA		
pTRV2-*IpSTP5_1*	This study	N/A
pTRV2-*IpSTP5_2*	This study	N/A
pTRV2-*IpTGW6*	This study	N/A

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Software and algorithms		
3D-DNA	Dudchenko et al.[46]	https://github.com/aidenlab/3d-dna
BUSCO	Manni et al.[25]	https://busco.ezlab.org/
BWA-MEM	Li[47]	https://github.com/bwa-mem2/bwa-mem2
CAFE	Mendes et al.[48]	https://github.com/hahnlab/CAFE5
Canu	Koren et al.[49]	https://github.com/marbl/canu
Diamond blast	Buchfink et al.[50]	https://github.com/bbuchfink/diamond
EDTA	Ou et al.[32]	https://github.com/oushujun/EDTA
EdgeR	Robinson et al.[51]	https://bioconductor.org/
GAPIT	Wang and Zhang[52]	https://www.zzlab.net/GAPIT/
geneHapR	Zhang et al.[53]	https://gitee.com/zhangrenl/genehapr
HAPPE	Feng et al.[54]	https://github.com/fengcong3/HAPPE
HISAT2	Pertea et al.[55]	http://daehwankimlab.github.io/hisat2/
InterProScan	Zdobnov and Apweiler[56]	http://www.ebi.ac.uk/interpro/
MAKER	Campbell et al.[57]	http://www.yandell-lab.org/software/maker.html
MCScanX	Wang et al.[58]	https://github.com/wyp1125/MCScanX
MEGA 7	Kumar et al.[59]	https://www.megasoftware.net/
NGenomeSyn	He et al.[60]	https://github.com/hewm2008/NGenomeSyn
OrthoFinder2	Emms and Kelly[61]	https://github.com/davidemms/OrthoFinder
R4.0	The R Foundation	https://www.rproject.orgl
RepeatModeler2	Flynn et al.[62]	https://github.com/Dfam-consortium/RepeatModeler
SiLiX	Miele et al.[63]	http://lbbe.univ-lyon1.fr/SiLiX.
StringTie	Pertea et al.[55]	https://ccb.jhu.edu/software/stringtie/
TimeTree	Hedges et al.[64]	http://timetree.org/
VCFtools	Danecek et al.[46]	https://vcftools.github.io/

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact, Lei Wang (wanglei@ibcas.ac.cn).

Materials availability

All unique/stable reagents generated in this study are available from the lead contact with a completed Materials Transfer Agreement.

Data and code availability

- The data supporting this work are available within the paper and supplementary files. The genome and anotation data were archived in the Genome Warehouse in the National Genomics Data Center, Beijing Institute of Genomics (BIG), Chinese Academy of Sciences, under accession number GWHBQLL00000000. And the re-sequence data were archived under the accession number CRA009397. RNA-sequencing data were archived under accession number CRA009483. GC-MS data were archived under accession number OMIX005749.
- This paper does not report original code.
- Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS

Plant materials and growing conditions

Eight-year-old of *I. polycarpa* CNBG01, CNBG02 and CAS07 trees were grown at Institute of botany, the Chinese Academy of Science. The 42 accessions of *I. polycarp* trees were grown in natural condition without chemical fertilizer, at the indicated places.

METHOD DETAILS

Plant materials

The leaves of female *I. polycarpa* (CAS07) was used as the material for extracting high-quality DNA. The unripe fruits of CNBG01 and CNBG02 were harvested at 9 a.m. and 9 p.m. on 25 DAP. All the samples were immediately frozen in liquid nitrogen, and then used for gDNA/RNA extraction.

DNB, Pacbio, and Hi-C sequencing

High-quality DNA were extracted by using cetyltrimethylammonium bromide (CTAB) method to construct DNA Nanoball (DNB) library for short read sequencing, and PacBio library for single-molecule real-time (SMRT) sequencing. To make DNB libraries, 500 ng high-quality DNA was selected, fragmented, end repaired and ligated to the paired-end adaptors. Fragments in 300–500 bp size were selected and purified by using PCR, and then digested by cyclase cleavage to obtain circular DNA. The DNA was amplified to nanoball by using the rolling circle amplification (RCA), and the nanoball was then loaded onto Pattern Array chip and sequenced by DNBseq platform with paired-end 150 bp (PE150) sequencing strategy, and totally generated ~178.30 Gb data.

The high-quality DNA was also used to fragmented, filtered for 30 kb in size, and constructed the libraries for single molecule real-time (SMRT) based on the standard protocols from the Pacific Bioscience (Pacbio, https://www.nature.com/protocolexchange/). These libraries were sequenced on Pacbio RS II platforms, which generated ~111.83 Gb data, with the average length 17.52 kb and subreads N50 31.52 kb.

The method of library construction for Hi-C can be seen from https://www.nature.com/protocolexchange/. Totally, we generated ~119.00 Gb data with PE150 sequencing strategy.

Genome size evaluation

The genome size evaluated firstly by flow cytometry measurements was 1030 Mb with *Arabidopsis* as a standard. And then, the genome size and heterozygosity ratio of *I. polycarpa* were estimated with short-read data by GCE (ftp://ftp.genomics.org.cn/pub/gce, v1.0.0) with following parameters, -k 21 -a 0 -d 0 for kmer_freq_hash and -m 1 -b 1 -H 1 for gce. and the equation genome size = K-mer coverage/Mean K-mer depth.

Genome assembly and pseudomolecule construction by Hi-C

The raw SMRT reads were self-corrected, trimmed and then assembled by Canu (v2.0) with following parameters, genomeSize = 1.23g, minOverlapLength = 700, minReadLength = 1000.[47] Then the assembled contigs were used to remove the bubble sequences by minimap2 (v2.1) with -x asm5 and purge_dups (v1.2.3) with -T 2 parameters. After that, they were used to construct the pseudomolecules with Hi-C data by using Juicer (v1.6) with -s MboI parameter and 3D-DNA (v190716).[26] In detail, the cutting site of restriction enzyme was firstly detected in contigs by juicer script, and then the Hi-C data were aligned to the digested contigs, while based on the interaction intensity of these contigs, they could be anchored, sorted, and finally connected to the pseudomolecule. Then the generated hic file was loaded into Juicebox (v1.8.8) software to generate the interaction density heatmap.

Genome repeat element identification

The EDTA (v1.9.6), RepeatModeler (v2.0.1) and RepeatMasker (v4.1.2) softwares were used to identify the repeat sequences in *I. polycarpa*. In detail, the pseudomolecule was firstly trained by EDTA and RepeatModeler to generate each library file of transposon elements, respectively. For EDTA, the parameters were –anno 1 –force 1 –debug 1 –sensitive 1 –evaluate 1, and for RepeatModeler, the -engine ncbi and other default parameters were used. The library files were then merged as the input file for RepeatMasker with parameters -a -html -gff.

While to identify LTR retrotransposon in the *I. polycarpa*, *P. trichocarpa*, *P. deltoides*, *S. suchowensis and S. purpurea* genomes, the LTR_harvest and LTR_retriever in genometools (v1.6.2) were performed for analysis. The *de novo* identified repeat sequences were then named by searching the homologous to the REXdb database using blast. The results were combined with the plant repeat sequence that was extracted from Repbase, and removed the redundant sequence with over 80% identity by using CD-Hit. After that, this repeat sequence library was fully annotated by RepeatMasker. The clustered *Copia* and *Gypsy* superfamilies was used to calculate their activities with the EMBOSS (v 6.6.0) based on the formula $T = K/(2\times r)$, where r refers to a general substitution rate of 1.28×10^{-9} per site per year in *Salicaceae* family.[65,66] At last, SiLiX with default parameters was used to cluster the *Copia* and *Gypsy* LTRs of *I. polycarpa* and *P. trichocarpa,* and count their numbers in different families. The MEGA7 and EvolView were used to construct the phylogeny of them.

Gene annotation and gene family analysis

The MAKER pipeline (v3.0)[57] consisted of three approaches, *de novo* assembled transcripts from RNA-sequencing, ab initio predictions and protein homologous predictions, was used to annotate the protein-coding genes in *I. polycarpa*. The RNA-sequencing data were sequenced from eight tissues at the fruit stage, which included the root, stem, leaf, flower, seed, pericarp, unripe fruit and ripe fruit, and generated ~88.07 Gb data totally. The RNA-sequencing data was *de novo* assembled by Trinity (v2.4) with the parameter –KMER_SIZE 32. The assembled transcripts were then merged with downloaded publically available evidence, such as cDNA or

CDS of *I. polycarpa* from NCBI, to enlarge the evidence library and improve the accuracy of gene prediction. For the protein homology-based prediction, we downloaded the protein sequences of *P. trichocarpa*, *S. suchowensis, Vernicia fordii, Sesamum indicum* L *and Olea europaea* from Phytozome (http://www.phytozome.net). As for the ab initio prediction, we used Fgenesh (version 3.1.1) with the parameters -pmrna -scip_prom -scip_term to annotate the repeat masked sequences that are masked by RepeatMasker. And then we integrated the predicted mRNA and proteins with the former as input data for the MAKE-R pipeline.

The longest protein sequences for each gene were used to compare with the no-redundant (Nr) database of NCBI to find their homologous with the diamond blast using the followed parameters, -a matches -k 1 -e 1e-5 -p 100 -f 6.[50] While the InterProScan[56] was used to annotate the functional domain and the possible GO terms with the parameters -goterms -t p -f GFF3 -pa -cpu 100. The online tool of jvenn (http://jvenn.toulouse.inra.fr/app/example.html) was used to calculate the number of overlapped genes in the databases of InterProScan. Orthofinder2 (v2.2.7) was used to cluster the genes families for the nine species with the parameter -M msa -A mafft -T fasttree.[61] CAFÉ was used to calculate the expanded and contracted gene families.[48] The SEA method in online AgriGo (v2) was used to analysis the genes enrichment.

Phylogeny and synteny analysis of *I. polycarpa*

The phylogeny tree for the above species generated by Orthofinder2 was then loaded to MEGA7 to further estimate their divergence time. With the referenced divergence time for *Populus* and *Salix* ~11.30 Mya based on TimeTree,[64] we estimate the divergence time for *I. polycarpa* and the two former was ~16.28 Mya. To analyze the synteny of *I. polycarpa*, *P. trichocarpa* and *S. suchowensis*, we aligned their genomes by nucmer and dnadiff in Mummer (v4.0) with the parameter, nucmer -L 1000, and drawn the synteny map with NGenomeSyn (v1.0)[60] and MCScanX (v2.0).[58] The *K*s value for the orthologous gene pairs between *I. polycarpa* and *P. trichocarpa* or *S. suchowensis* that calculated by the online tool Coge (https://genomevolution.org/coge/) were counted for further analysis.

Population analysis

Resequencing libraries of 42 *I. polycarpa* accession were constructed and sequenced as above described DNB-seq and generated ~471.5 Gb of data. The reads were quality controlled with Trimmomatic using default parameters. And then aligned against the reference genome using BWA-MEM (v 0.7.17).[47] High quality reads were processed with Picard v2.0.1 (v2.0.1) for PCR deduplication. The HaplotypeCaller function of GATK (v4.1.2.0) was then used to generate GVCF files for each accession with parameters '-genotyping_mode DISCOVERY -max_alternate_alleles 2 -read_filter OverclippedRead', followed by population variant calling using the function GenotypeGVCFs of GATK with default parameters. Hard filtering was applied to the raw variant set using GATK, with parameters 'QD < 2.0 || FS > 60.0 || MQ < 40.0 || MQRankSum <−12.5 || ReadPosRankSum <−8.0' applied to SNPs, and 'QD < 2.0 || FS > 200.0 ||ReadPosRankSum <−20.0' applied to small indels. And then SNPs were filtered with MAF ≥0.05 and missing rate ≤0.1 by VCFtools (v0.1.16).[46] At last, 22,685,688 SNPs and 2,469,008 INDELs were retained for next analysis.

Further functional annotation of SNPs and Indels was performed with snpEff (v4.3) using a self-built database. PCA (Principal component analysis) was performed using GAPIT with the entire set of SNPs. Population structure was used to estimate the ancestry with a predefined K range of 2–4 using ADMIXTURE. LD decay was calculated for all SNPs within 300 kb using PopLDdecay (v3.40) with default parameters. GWAS analysis was performed using the Farm-CPU model in GAPIT.[52] The most obvious SNP peak (the criterion $-\log_{10}$ (p value) > 5) in Manhattan plot was chosen as the candidate SNP. Haplotype analysis for *IpSTP5* was conducted using the R package 'geneHapR' and HAPPE software with default parameters.[53,54]

RNA extraction and RT-qPCR

Total RNA was extracted from unripe fruit by using TRIzol reagent (Invitrogen). The first-strand cDNA was synthesized from 1 μg Dnase 1-Treated total RNA using PrimeScript RT reagent kit (Takara). *I. polycarpa PP2A* gene (*Ip32098*) was used as internal control to normalize samples. Quantitative reverse transcription PCR was performed on Mx3000P instrument (Stratagene, La Jolla, CA, USA) with a SYBR Green Real-time PCR Master Mix (Toyobo, Osaka, Japan). 15 μL qPCR mixture contained 3 μL 10-fold diluted cDNA samples, 7.5 μL SYBR solution and 0.25 μM of forward and reverse primers for each gene. Each cDNA sample was measured at least three times forming technical repeats. The primers used in this study are listed in Table S9.

Transcriptome analysis

To identify the gene's expression in different tissues, we performed RNA-sequencing. The methods for the library construct can be found in https://www.nature.com/protocolexchange/. And the libraries were sequenced on the Illumina platforms and totally generated ~175.65 Gb RNA-sequencing data with PE150. While the RNA-sequencing data for *I. polycarpa* for different tissues were downloaded from the NCBI with the SRA number. The data was mapped to the referenced genome by HISAT2 (v2.2.0), and the gene expression level was calculated by StringTie (v2.2.0).[55] The DEGs analysis were executed by edgeR.[51] The GO analysis was executed on the OmicShare platform (https://www.omicshare.com/tools/). The top 10 hub genes were filtered by MCC methods in Cytoscape. All the heatmap figures produced by R4.0.

Oil content determination

The fruits were dried at 120°C for 6 h and pulverized (Foss Tecator Cyclotec 1093). The samples (m represented the weight) were packaged into a filter paper and soaked into petroleum ether solvent for oil separation based on Soxhlet extraction method. The

oil content was calculated by the formula: the oil content= $(m_1-m_2)/m\times100\%$, where the m_1 and m_2 referred to filter paper pack dry weight at before and after extraction. The experiment was performed in triplicate.

Fatty acid composition determination

The fatty acid components were determined by GC-MS following previously published protocol with minor modifications.[65,66] For reference substance treatment, 100 μL of each fatty acids (1 mg/mL in hexane) including palmitic acid (Sigma-Aldrich, P5585), palmitoleic acid (Sigma-Aldrich, P9417), stearic acid (Sigma-Aldrich, S4751), oleic acid (Sigma-Aldrich, O1008), linoleic acid (Sigma-Aldrich, L1376), linolenic acid (Sigma-Aldrich, L2376) and arachidic acid (Sigma-Aldrich, 10930), and 50 μg C19:0 as well as 0.01% (w/v) butylated hydroxytoluene (BHT) were mixed, and condensed under a stream of nitrogen to facilitate the low content fatty acids of *I. polycarpa* oil. Then, this mixed standards were treated together with the tested samples. The fruits at mature stage were dried and then pulverized to powder. 10 mg dried powder of each sample and 50 μg C19:0 were incubated in a glass tube with 1.5 mL methanol containing 5% H_2SO_4 and 0.01% BHT to prevent lipid oxidation at 95°C for 1 h for transmethylation. Fatty acid methyl esters were extracted with 1.5 mL of hexane and 1 mL of water. The mixture was vortexed and centrifuged, and then 800 μL of 1.5 mL upper phase was transferred into a 2 mL glass vials with silicone septa lined screw caps, and condensed by a stream of nitrogen, and then dissolved in 40 μL hexane to a GC-QQQ/MS detection system (Agilent Technologies 7890A-7000, Agilent). 1 μL samples were injected under split ratio (10:1) for GC analysis. The GC system was supplied with MRM (multiple reaction monitoring, MRM) model and a capillary column Agilent CP7419 Select FAME (50 m × 0.25 mm × 0.25 μm) with helium carrier at a flow rate of 1 mL/min. The oven temperature was maintained at 60°C for 3 min, and then increased to 180 °C at a ramp rate of 5°C/min holding for 11 min, at last, increased to 250 °C at a ramp rate of 30°C/min holding for 10 min. FAMEs from TAG were identified by comparing their retention times with known standards. Each peak area was obtained, converted to lipid amount based on the internal standard amount, and calculated as ‰ (μg/mg) of the fruit weight.

Auxin content determination

The 0.05g fresh fruits were grinded into flour in liquid nitrogen, and mixed with 10 μL of methanolic internal standard solution ([$^2H_5$]IAA, 10 μg/mL in methanol), and then extracted with 1 mL 80% methanol overnight at 4°C. After centrifugation at 5976*g* for 5 min, the samples were dissolved in 1 mL acetic acid/ethyl acetate (5: 95, v: v), and the supernatant were condensed by a stream of nitrogen. Following then added 30 μL methanol and 100 μL deionized water, and incubated for 2 h at −20°C. After centrifugation at 10625*g* for 7 min, the supernatant was evaporated by a stream of nitrogen. Samples were dissolved in 30 μL bis (trimethylsilyl) trifluoroacetamide with 3 μL pyridine and incubated for 30 min at 80°C, then were determined by GC-MS (7890A-7000, Agilent) as previous described by Guo.[67] The data were analyzed by GraphPad Prism9.0.

VIGS assay

pTRV1 and pTRV2 vectors have been described by Liu.[68] To generate pTRV2-gene, a cDNA fragment was PCR amplified and cloned into *Bam*HI- *Xho*I- cut pTRV2 vector. The pTRV1, pTRV2 and its derivatives were introduced into agrobacterium strain GV3101. The detailed method has been described as follows. A 5mL culture was grown overnight at 28°C in kanamycin and rifampin LB medium. Then the culture was inoculated in a 50mL LB medium, containing kanamycin, rifampin, 10mM MES and 20μm AS. Thereafter, agrobacterium cells were harvest and resuspended into infiltration medium (10mM $MgCl_2$, 10mM MES, 100μm AS), adjusted to an OD 0.5 and stabilized at room temperature for 3h. Agroinfiltration was performed with a needleless 1mL syringe into unripe fruits. Two weeks later, *IpSTP5* groups were determined oil content by low field nuclear magnetic resonance,[69] which was based on the standard line of soxhlet extraction. *IpTGW6* groups were measured width by vernier caliper. The expression of *IpSTP5 and IpTGW6* were determined by RT-qPCR as above description.

Cell Reports, Volume *43*

Supplemental information

The *Idesia polycarpa* genome provides insights into its evolution and oil biosynthesis

Yi Zuo, Hongbing Liu, Bin Li, Hang Zhao, Xiuli Li, Jiating Chen, Lu Wang, Qingbo Zheng, Yuqing He, Jiashuo Zhang, Minxian Wang, Chengzhi Liang, and Lei Wang

Supplementary figures

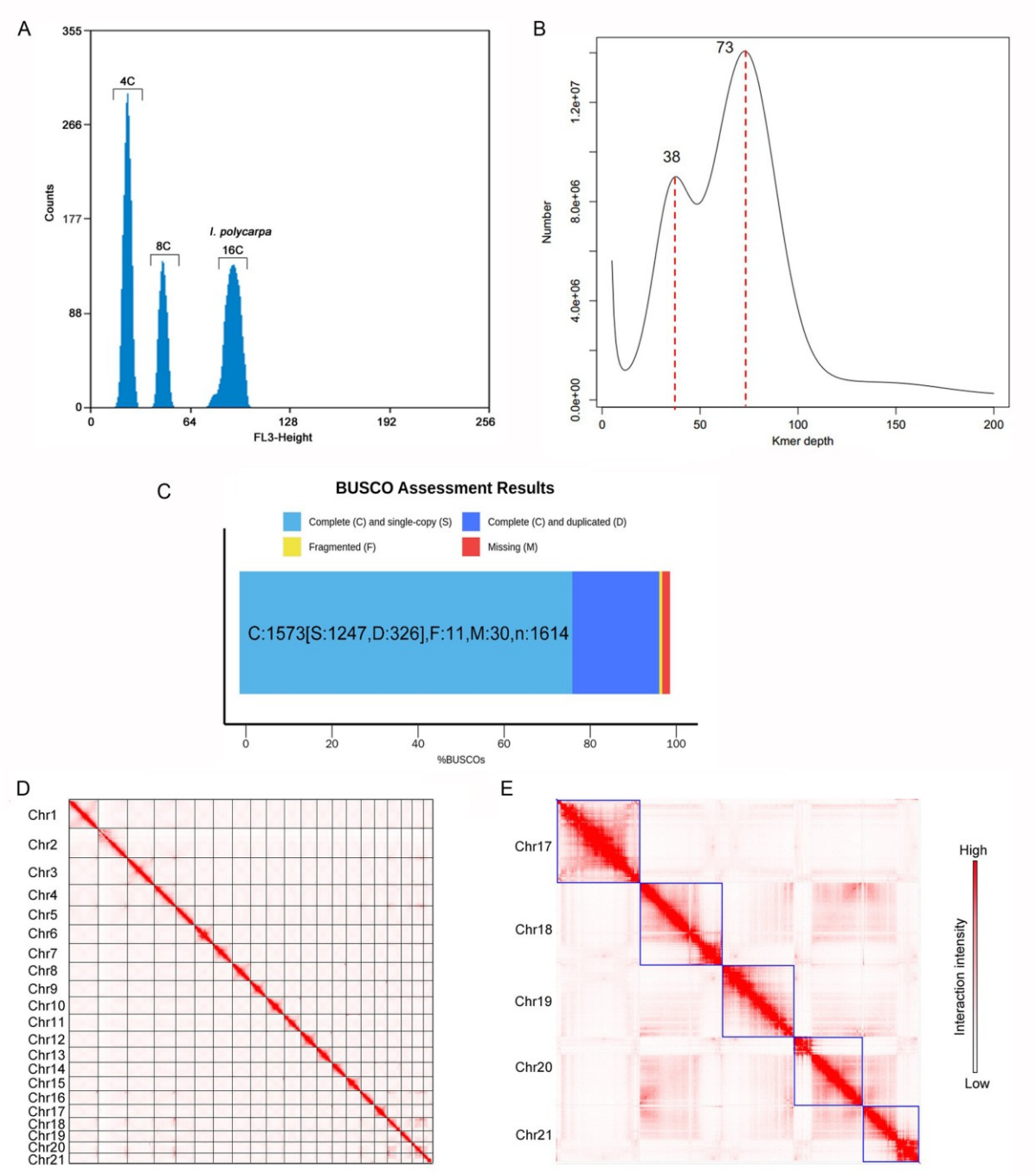


Figure S1. Genome size estimation and Hi-C assisted genome assembly of *I.polycarpa*. Related to Figure 1.

(A) Flow cytometry estimation of *I. polycarpa* genome size for *de novo* sequence. The flow cytometric histograms of relative fluorescence intensities of propidium iodide-stained nuclei were simultaneously isolated from *I. polycarpa* and *Arabidopsis,* as the reference standard. *I. polycarpa* genome size was estimated to be ~1.03Gb. (B) K-mer distribution of Illumina reads for *I. polycarpa*. The genome size was estimated

to be ~1.23 Gb based on the 17-mer distribution curve. And the heterozygosity was ~0.90%. (C) BUSCO assessment of the assembled scaffolds. Approximately 98.1% of the 1,614 BUSCO genes were annotated, about 97.5% of the genes were complete. (D) and (E) The Hi-C-assisted genome assembly of *I. polycarpa***.** (D) Heatmap showing Hi-C intrachromosomal interactions on the 21 pseudochromosomes based on 500-kb resolution. (E) The enlarged Hi-C heatmp of the last five pseudochromosomes.

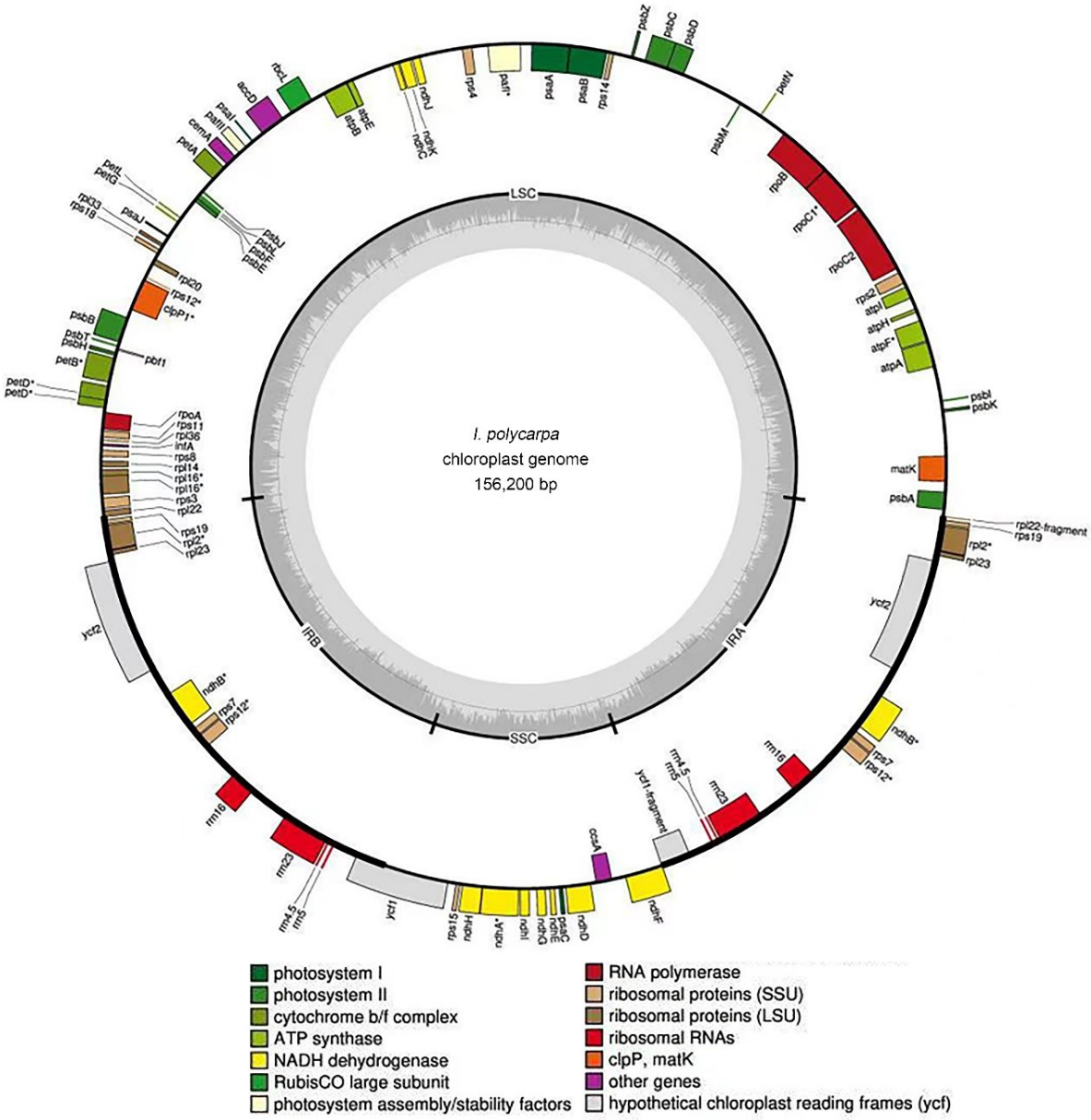


Figure S2. Chloroplast reference genome of *I. polycarpa*. Related to Figure 1.

The localization and function of the annotated genes are illustrated in the outer circle. GC content is graphed around the inner circle with the line indicating 36.79% GC content. The map was generated using OrganellarGenomeDraw.

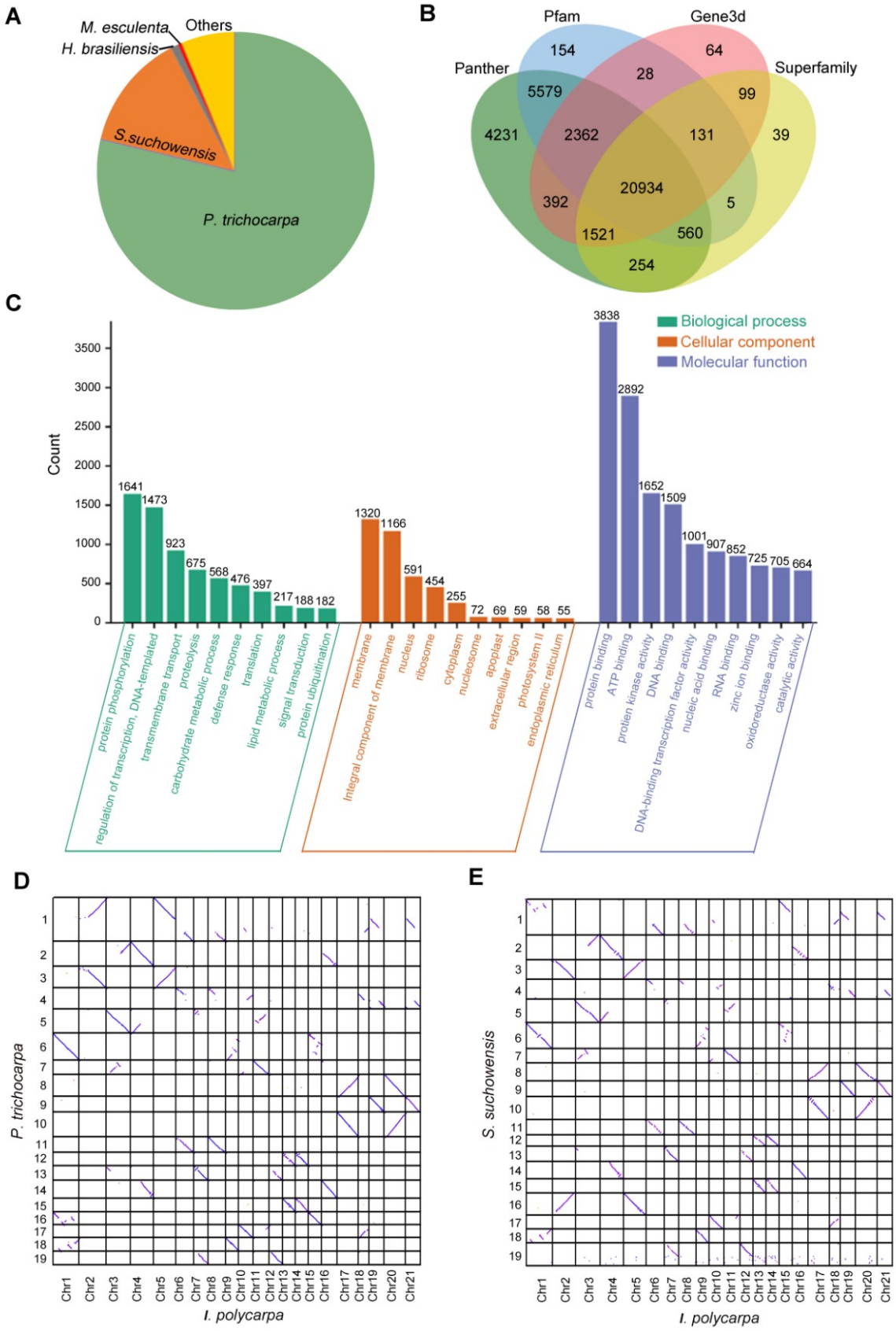


Figure S3. Gene annotation, functional analysis and comparative analysis of the genome. Related to Figure 3.

The functional annotation results of *I. polycarpa* genes by NR (A) and Interproscan (B) databases, respectively. Most of the annotated genes were homologous to genus *P. trichocarpa*, and then *S. suchowensis*. (C) The GO terms and the number of genes for each in *I. polycarpa* genomes. (D) Pattern of synteny between *I. polycarpa* and *P. trichocarpa.*. The results displayed 2-to-2 collinearity relationship with minimum length 2 kb. And with minimum length 5 kb, they have a 1-to-1 homologous chromosomes relationship, which displayed the chromosome rearrangement event of their ancestors. (E) Pattern of synteny between *I. polycarpa* and *S. suchowensis*

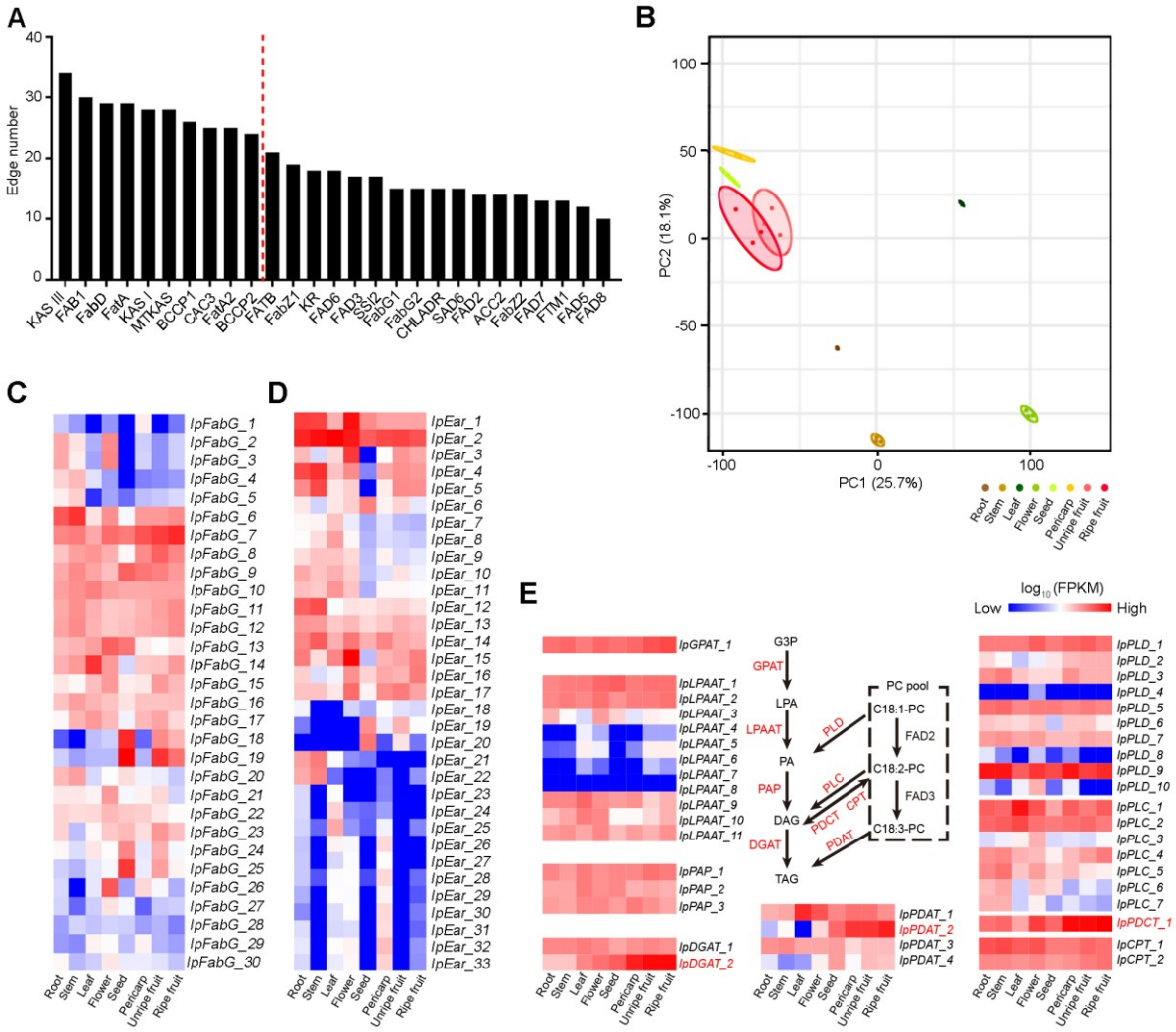


Figure S4. RNA-seq analysis of different tissues of *I.polycarpa.* Related to Figure 4.

(A) The edge number of key fatty acid metabolism genes in the network. The red dash line splits the top 10 hub genes and the rest genes that edge number was more than 10. (B) PCA result of the transcriptome of eight tissues with three biological replicates. Samples from the same tissues are indicated by the same color. (C and D) The expression patterns of 30 FabG (C) and 33 Ear (D) genes in different tissues. (E) The expression patterns of key genes associated with lipid biosynthesis in different tissues.

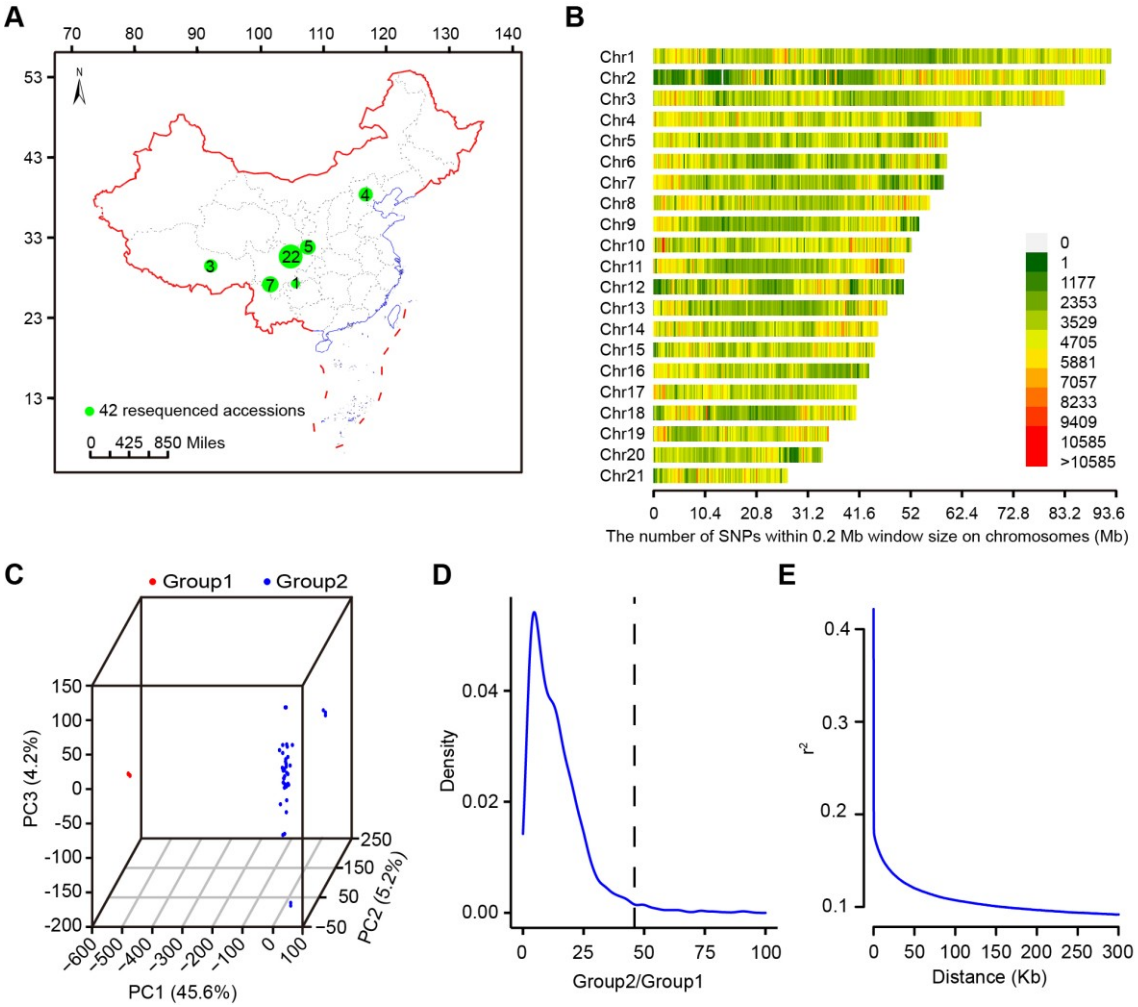


Figure S5. Genome wide association study (GWAS) of 42 *I. polycarpa* accessions. Related to Figure 5.

(A) Geographical distribution of the re-sequenced accessions. (B) The density profile of SNPs on each chromosome of *I. polycarpa*. (C) PCA showed clear separation in *I. polycarpa* accessions. (D) The genetic diversity ($\pi_{group2}/\pi_{group1}$) pattern of re-sequenced accessions. Using a cutoff of $\pi_{group2}/\pi_{group1}$ >46 as significant (2.5%) 29.5 Mb under domestication selection were identified in *I. polycarpa* genome. (E) Linkage disequilibrium (LD) decay pattern of *I. polycarpa*.

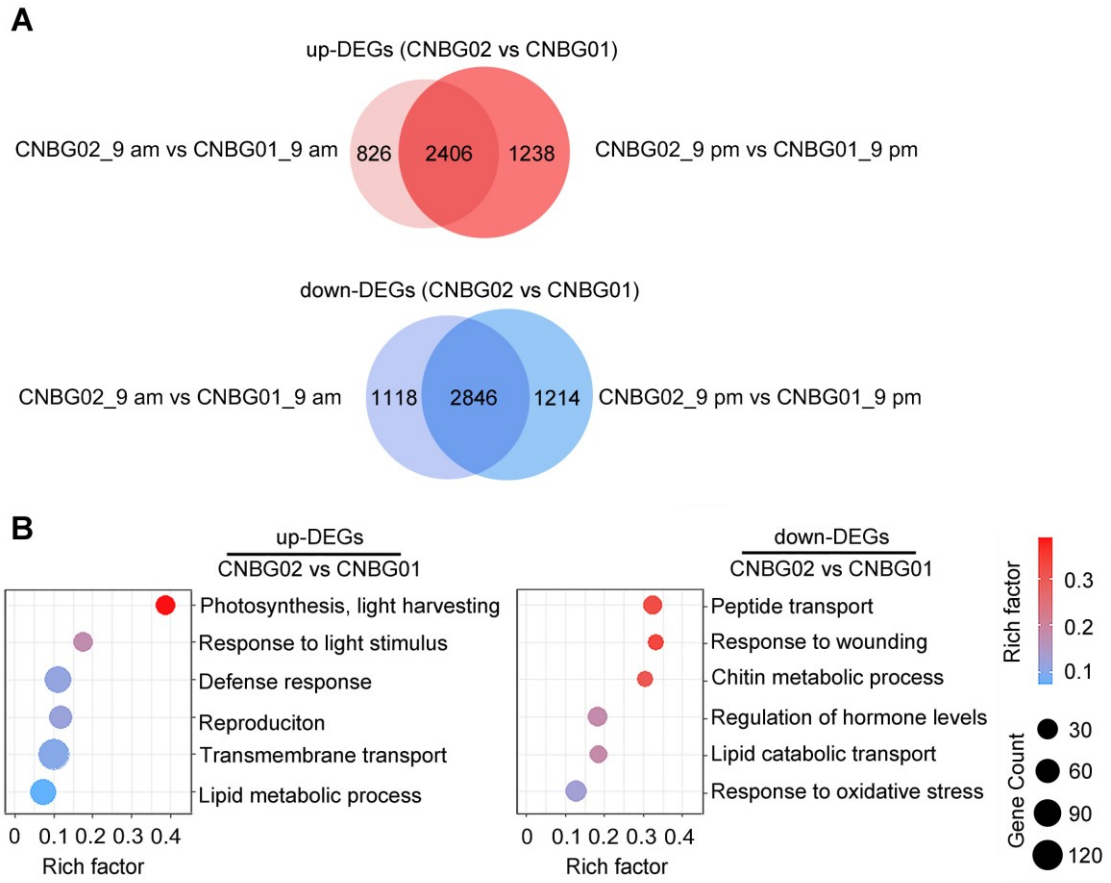


Figure S6. RNA-sequencing analysis of CNBG01 and CNBG02. Related to Figure 6.

(A) Bioinformatics analysis of DEGs (FDR < 0.05 & Fold Change > 2). Overlapped number were the shared DEGs, and then were used for GO analysis. (B) Gene Ontology (GO) analysis showing enriched biological processes among genes up-regulated in CNBG02 compared to CNBG01.

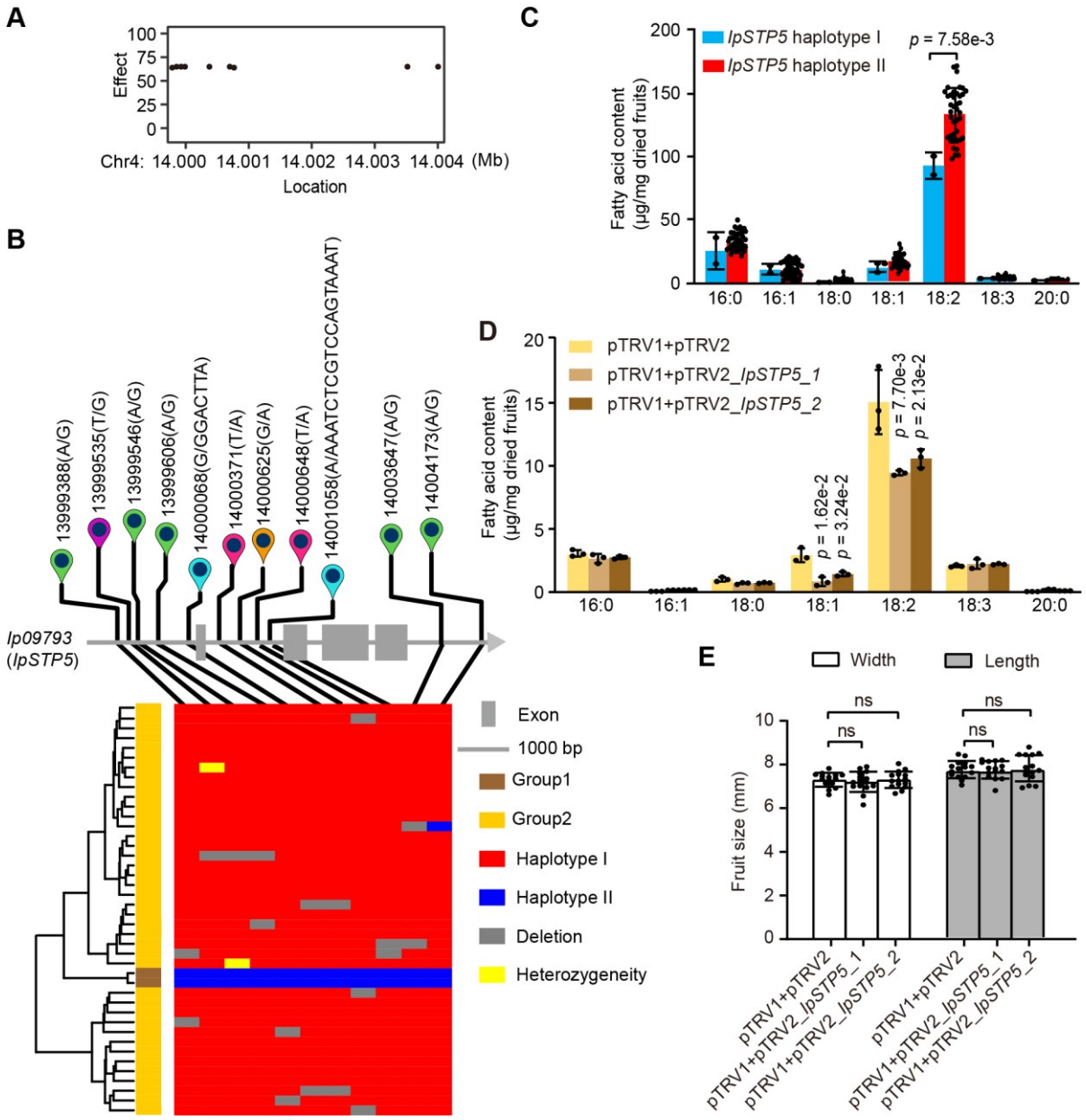


Figure S7. The population analysis and VIGS assay of *IpSTP5*. Related to Figure 6.

(A) The estimation of effect at *IpSTP5* SNPs sites. (B) The haplotypes analysis of *IpSTP5* in 42 accessions including SNPs and INDELs. (C) Comparison of individual fatty acids content between two distinct *IpSTP5* haplotypes among 42 accessions of *I. Polycarpa*. The mature fruits were used for fatty acids determination by GC-MS. Values are mean ± SEM with individual data points, which stands for each of *I. polycarpa* accessions, and *p*-value was derived from ANOVA. (D) Comparison of individual fatty acids content in *IpSTP5* VIGS treated fruits. The unripe fruits after 14

days of inoculation, harvesting at 30 DAP (days after pollination), were taken for fatty acids determination by GC-MS. Values are mean ± SEM (n = 3), and *p*-values were derived from ANOVA. Individual species of fatty acids (number of carbons: number of double bonds; 16:0, palmitic acid; 16:1, palmitoleic acid; 18:0, stearic acid; 18:1, oleic acid; 18:2, linoleic acid; 18:3, linolenic acid; 20:0, arachidic acid) were listed on abscissa axis. (E) Statistical analysis of the width and length of unripe fruits after 14 days of inoculation in VIGS assay. Values are mean ± SEM (n = 15, *p*-values were not significant (ns) by ANOVA).

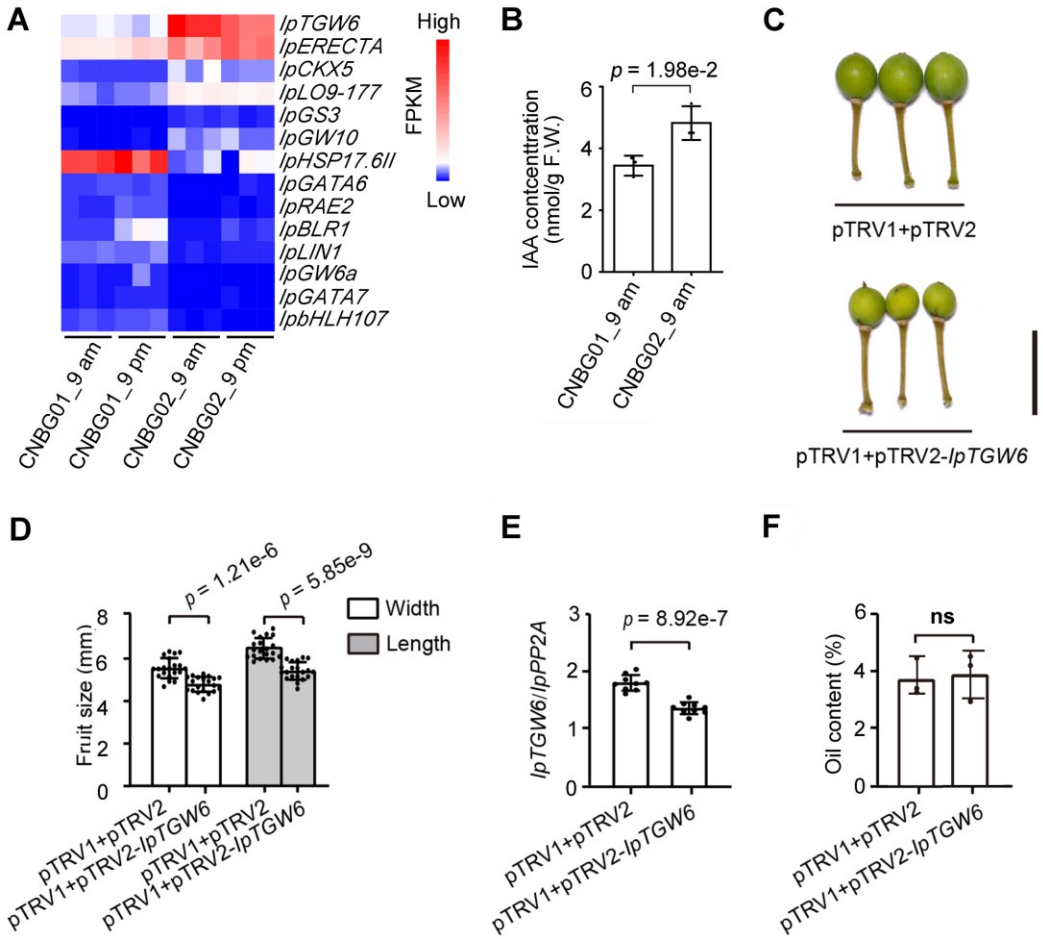


Figure S8. The probable genetic basis of fruit size in *I. polycarpa*. Related to Figure 6.

(A) The heatmap analysis showing 14 DEGs likely associated with fruit size. (B) Indole-3-acetic acid (IAA) concentrations in unripe CNBG01 and CNBG02 fruits. Values are mean ± SEM (n = 3, *p*-values were derived from ANOVA). (C) The unripe fruit phenotypes after 14 days of inoculation. Scale bars = 1 cm. (D) Statistic analysis of the width of unripe fruits after 14 days of inoculation. Values are mean ± SEM (n = 20, *p*-values were derived from ANOVA). (E) RT-qPCR assay of *IpTGW6* expression in unripe fruits after 14 days of inoculation. Values are mean ± SEM (n = 9, *p*-values were derived from ANOVA). (F) Oil content of unripe fruits after 14 days of inoculation in *IpTGW6* VIGS assay. Values are mean ± SEM (n = 3, *p*-values were not significant (ns) by ANOVA).

Supplementary tables

Table S1. Sequence and assembly statistics. Related to Figure 1.

(A) Summary of sequenced data to assemble the genome of *I. polycarpa*. (B) Quality values of the assembled contigs and scaffolds. (C) The assembly and gene annotation of each pseudo-chromosomes for *I. polycarpa* genomes.

(A)

Platform	DNB-seq	Hi-C	PacBio (CLR)
Data volume (Gb)	178.30	119.00	111.83
Reads length (bp)	PE150	PE150	17516
Coverage depth (×)	~144.95	~96.75	~90.92

(B)

Type	Contigs (bp)	Scaffolds (bp)
N50	5,822,192	54,293,521
N60	4,060,266	51,298,384
N70	2,554,638	45,882,993
N80	1,465,942	41,560,487
N90	448,746	34,576,920
Average length	716,567	822,081

(C)

Pseudochromosome	Length(bp)	Gene number	Pseudochromosome	Length(bp)	Gene number
Chr1	93,632,804	3,098	Chr13	47,754,307	1,553
Chr2	92,423,757	3,037	Chr14	45,882,993	1,445
Chr3	84,067,912	2,680	Chr15	45,297,513	1,470
Chr4	67,021,858	2,827	Chr16	44,085,287	1,956
Chr5	60,221,121	2,388	Chr17	41,560,487	2,500
Chr6	59,988,279	1,740	Chr18	41,448,802	1,313
Chr7	59,352,314	1,598	Chr19	35,831,779	1,717
Chr8	56,540,163	1,791	Chr20	34,576,920	2,345
Chr9	54,293,521	1,528	Chr21	27,407,486	1,652
Chr10	52,804,925	1,510	Scaffold	66,697,536	801
Chr11	51,298,384	1,746	Total	1,213,391,197	42,086
Chr12	51,203,049	1,391			

Note: PE, Pair-end.

Table S2. Repeat sequence content in *I. polycarpa* genome. Related to Figure 2.

Type	Number of elements	Length occupied (bp)	Percentage of the genome (%)
Class I: Retroelements	501,757	526,447,486	43.39
SINEs	4,137	948,095	0.08
LINEs	13,640	11,713,489	0.97
LTR/Copia	110,298	130,873,371	10.79
LTR/Gypsy	305,157	339,154,986	27.95
Others	183	524201	0.04
Class II: DNA transposons	507,520	179,988,777	14.83
hobo-Activator	19,865	9,632,147	0.79
Tc1-IS630-Pogo	1,136	366,763	0.03
Harbinger	8,602	3,875,169	0.32
RC	57,910	26,192,107	2.16
Unclassified	495,783	138,907,198	11.45
Small RNA	9,953	7,621,447	0.63
Simple repeats	264,951	37,595,353	3.10
Low complexity	41,751	2,039,818	0.17
Total	--	917,844,091	75.64

Table S3. The RNA-sequencing and annotated genes' information. Related to Figure 3.

(A) The RNA-sequencing data and mapping statistics in this study. (B) The feature of protein-coding genes for *I. polycarpa* genome. (C) The list of genes database and annotated genes number by InterProScan, respectively.

(A)

SRA accession	Data volume (G)	Mapping ratio (%)	GC(%)	Q30(%)	Q20(%)	Tissue
Root_rep1	3.45	87.01	43.55	92.75	97.41	
Root_rep2	3.28	87.76	43.43	93.22	97.26	Root
Root_rep3	3.42	87.56	43.47	92.29	96.90	
Stem_rep1	3.57	87.91	43.46	93.00	97.16	
Stem_rep2	3.37	87.95	43.38	94.18	97.97	Stem
Stem_rep3	3.05	88.10	43.48	94.11	97.96	
Leaf_rep1	4.30	87.81	43.66	93.40	97.79	
Leaf_rep2	3.69	89.94	43.06	91.71	96.71	leaf
Leaf_rep3	3.20	90.16	43.18	92.56	97.03	
Flower_rep1	2.97	89.98	43.03	92.32	97.06	
Flower_rep2	3.33	90.16	42.93	93.04	97.63	Flowers
Flower_rep3	3.33	90.07	43.06	93.19	97.67	
SRR3993520 [1]	3.69	86.29	43.86	92.67	97.09	
SRR3993521 [1]	5.61	86.43	43.28	93.15	97.28	Pericarp
SRR3218560 [1]	5.24	86.45	43.29	93.24	97.32	
SRR3993522 [1]	3.69	86.21	43.97	93.23	97.34	
SRR3993523 [1]	4.27	86.41	44.08	93.47	97.45	Seed
SRR3218547 [1]	4.97	85.83	43.97	93.25	97.36	
SRS8107446 [2]	3.80	87.82	43.17	93.31	97.74	
SRS8107447 [2]	2.99	87.01	43.37	93.11	97.65	Unripe fruit
SRS8107448 [2]	3.25	88.89	43.14	93.76	97.92	
SRS8107449 [2]	3.28	89.41	43.33	93.84	97.97	
SRS8107450 [2]	2.96	87.49	43.26	93.27	97.71	Ripe fruit
SRS8107451 [2]	3.36	87.59	43.30	93.64	97.87	
CNBG01_9 am_rep1	7.60	95.59	43.06	93.59	97.87	
CNBG01_9 am_rep2	9.30	94.92	43.20	91.35	96.87	CNBG01 _9 am
CNBG01_9 am_rep3	7.60	95.82	43.18	94.19	98.11	
CNBG01_9 pm_rep1	8.12	95.38	43.07	92.84	97.55	CNBG01 _9 pm
CNBG01_9 pm_rep2	5.65	94.33	43.11	91.14	96.78	

CNBG01_9 pm_rep3	7.78	94.68	43.12	92.47	97.35	
CNBG02_9 am_rep1	6.14	92.02	43.19	93.43	97.80	
CNBG02_9 am_rep2	6.90	90.96	43.36	90.69	96.53	CNBG02_9 am
CNBG02_9 am_rep3	6.37	91.31	43.19	91.59	97.00	
CNBG02_9 pm_rep1	7.98	90.52	43.10	91.07	96.74	
CNBG02_9 pm_rep2	7.04	91.83	43.11	92.84	97.56	CNBG02_9 pm
CNBG02_9 pm_rep3	7.10	91.33	42.94	92.14	97.24	
Total/average data	175.65	89.69	43.21	92.86	97.41	

(B)

Features of protein-coding genes	Number or ratio
Number of genes	42,086
Median gene length (bp)	2,240
Total gene length (Mb)	132.99
Gene space	10.96%
Number of transcripts	98,901
Median transcripts length (bp)	1,165
Median longest protein length (aa)	285

(C)

Genes database	Annotated genes number	Genes database	Annotated genes number
PANTHER	35,833	*Coils*	5,584
Pfam	29,753	*PIRSR*	5,041
Gene3D	25,531	*PRINTS*	4,990
SUPERFAMILY	23,543	*TIGRFAM*	3,639
ProSiteProfiles	13,875	*PIRSF*	1,716
MobiDBLite	13,648	*Hamap*	1,073
CDD	11,947	*SFLD*	280
SMART	10,765	*AntiFam*	2
ProSitePatterns	6,902		

Note: [1,2] represent the data were download from NCBI.

CNBG01 represents *I. polycarpa* cv. ‘China National Botanical Garden 01’.

CNBG02 represents *I. polycarpa* cv. ‘China National Botanical Garden 02’.

Table S4. Annotated genes analysis. Related to Figure 3.

(A) The number of genes for orthogroups. (B) Enriched GO terms of significant 70 contracted genes. (C) Contribution of duplication events to gene families.

(A)

Species	Total number of genes	Number of gene families in orthogroups	Number of genes in orthogroups	Number of unassigned genes	Average homologous copies of each gene
I. polycarpa	42,086	16,324	37,182	4,904	1.98
P. trichocarpa	34,699	15,417	32,292	2,407	1.95
S. suchowensis	26,599	14,027	25,222	1,377	1.73
V. fordiie	28,422	14,504	25,718	2,704	1.65
A. thaliana	27,540	13,327	24,994	2,546	1.74
O. sativa	55,803	14,422	46,154	9,649	2.32
S. indicum	35,410	13,788	34,449	961	2.40
O. europaea	50,684	14,689	45,442	5,242	2.54
M. truncatula	50,894	15,758	43,616	7,278	2.21

(B)

GO term	Ontology	Description	FDR
GO:0006749	P	glutathione metabolic process	4.40E-25
GO:0006575	P	cellular modified amino acid metabolic process	4.90E-23
GO:0006790	P	sulfur compound metabolic process	1.10E-19
GO:0006518	P	peptide metabolic process	2.20E-09
GO:0043603	P	cellular amide metabolic process	2.70E-09
GO:1901564	P	organonitrogen compound metabolic process	3.10E-05
GO:0004364	F	glutathione transferase activity	3.80E-30
GO:0016765	F	transferase activity, transferring alkyl or aryl (other than methyl) groups	9.00E-23
GO:0008194	F	UDP-glycosyltransferase activity	1.10E-18
GO:0016757	F	transferase activity, transferring glycosyl groups	1.10E-12
GO:0004497	F	monooxygenase activity	5.60E-08
GO:0016740	F	transferase activity	6.20E-08
GO:0016705	F	oxidoreductase activity, acting on paired donors, with incorporation or reduction of molecular oxygen	7.90E-08
GO:0005506	F	iron ion binding	7.90E-08
GO:0020037	F	heme binding	7.40E-07
GO:0046906	F	tetrapyrrole binding	9.10E-07
GO:0003824	F	catalytic activity	0.0012
GO:0016491	F	oxidoreductase activity	0.0018
GO:0046914	F	transition metal ion binding	0.0023
GO:0046872	F	metal ion binding	0.0092
GO:0043169	F	cation binding	0.011
GO:0043167	F	ion binding	0.019

(C)		
Type	Number of families with WGD percentages (%)	Total number of families
Total gene families	9,116 (55.84%)	16,324
Unique gene families	333 (26.77%)	1,244
Expanded gene families	1,509 (80.10%)	1,884
Contracted gene families	255 (52.58%)	485

Table S5. Expansion of key genes working on PUFA pathway in different plant species. Related to Figure 4.

Type\Num	*A. thaliana*	*O. sativa*	*V. fordii*	*M. sativa*	*O. europaea*	*S. indicum*	*P. trichocarpa*	*S. suchowensis*	*I. polycarpa*
ACC	1	0	2	1	3	4	3	3	4
BCCP	4	0	3	4	5	5	5	4	5
FabD	1	1	1	1	2	1	2	2	2
KASIII	1	1	1	2	4	3	1	1	4
KASI	2	3	3	4	6	12	5	3	2
FabG	31	18	29	41	34	40	32	31	36
EAR	23	26	23	41	52	41	17	12	44
FabZ	2	3	1	2	1	1	2	1	2
SAD	11	5	6	4	7	2	10	8	11
FatB	1	3	1	1	2	4	3	2	2
KASII	2	3	2	3	7	7	3	2	4
FAD2	1	1	3	2	5	2	4	3	4
FAD3	3	4	4	4	4	2	5	5	5
FatA	3	5	2	2	9	3	4	3	5
FAD	3	2	2	4	3	5	6	5	3
Total	89	75	83	116	144	132	102	85	133

Table S6. The re-sequenced data and SNPs/InDels statistics. Related to Figure 5.

(A) Summary of the sequencing information of 41 additional re-sequenced accessions. (B) Numbers of SNPs and INDELs in *I. polycarpa*. (C) Distribution of annotated SNPs and INDELs in *I. polycarpa*.

(A)

Sample	Q20 (%)	Q30 (%)	GC Content (%)	Total Reads	Depth (×)
01XZ9	97.58	92.79	32.49	68,634,818	8.37
02XZ5	97.60	92.83	32.66	70,065,870	8.54
03XZ3	97.48	92.50	32.64	71,134,514	8.67
05LZ	97.48	92.30	32.81	114,171,172	13.92
06YY6	97.53	92.64	32.80	66,437,864	8.10
08YY9	97.81	93.46	33.00	79,199,164	9.66
09YY2	97.94	93.84	33.60	83,400,418	10.17
10CAS1	97.84	93.54	33.80	83,228,982	10.15
11YY7	97.76	93.31	33.26	67,817,450	8.27
12BHL	97.96	93.90	33.95	88,921,958	10.84
13YY3	97.92	93.76	33.54	79,893,086	9.74
16ST3	97.88	93.70	34.33	84,038,538	10.25
17ST1	97.53	92.71	34.58	62,350,598	7.60
18ST2	98.15	94.46	33.93	85,635,912	10.44
22CT52	97.65	93.02	33.27	74,191,006	9.05
23QC71	97.52	92.65	33.55	68,318,824	8.33
24CT2	97.75	93.30	33.20	74,734,606	9.11
25YY8	97.75	93.30	33.36	75,447,640	9.20
26YY5	97.58	92.80	34.01	72,072,938	8.79
29ST7	97.64	92.97	33.22	65,512,786	7.99
32ST8	98.07	94.21	33.45	96,010,086	11.71
33CAS2	97.79	93.40	33.30	70,154,102	8.56
35QC9	97.70	93.13	33.04	81,956,598	9.99
36QC83	97.78	93.41	34.12	62,844,540	7.66
37QC82	97.80	93.43	32.90	82,901,074	10.11
38QC81	97.70	93.14	33.09	80,308,616	9.79
39QC51	97.61	92.87	32.79	66,845,572	8.15
41CT11	97.70	93.14	33.48	76,515,042	9.33
42CT12	97.75	93.28	33.10	81,094,714	9.89
43CT13	97.63	92.96	32.98	75,047,044	9.15
44NQ1	97.69	93.08	33.25	74,981,262	9.14
45NQ2	97.60	92.84	32.72	69,513,358	8.48
46NQ3	97.76	93.30	32.79	75,208,766	9.17
48NQ5	97.47	92.32	32.59	93,874,390	11.45
49NQ6	97.74	93.22	32.45	63,938,200	7.80
50CT6	97.73	93.25	33.60	72,485,988	8.84

51CT7	97.79	93.41	33.67	77,769,160	9.48
52CT8	97.70	93.16	33.05	69,632,690	8.49
55CT3	97.63	92.95	33.43	86,078,266	10.50
57CT51	97.64	92.95	33.61	65,285,036	7.96
58QC73	97.85	93.60	35.59	85,729,190	10.45

(B)

Chromosome	SNP	INDELs	Chromosome	SNP	INDELs
Chr1	1,809,852	202,014	Chr13	859,725	96,051
Chr2	1,708,152	164,473	Chr14	1,002,126	105,373
Chr3	1,730,093	182,872	Chr15	947,900	101,800
Chr4	1,462,392	162,404	Chr16	829,028	96,414
Chr5	1,179,841	137,010	Chr17	873,453	107,443
Chr6	1,129,401	120,473	Chr18	779,251	84,469
Chr7	1,048,885	112,084	Chr19	779,343	88,127
Chr8	1,221,927	125,693	Chr20	630,446	82,307
Chr9	958,387	107,099	Chr21	602,517	69,098
Chr10	1,157,240	113,886	Scaffold	77,133	6,680
Chr11	996,775	108,633	Total	22,685,688	2,469,008
Chr12	901,821	94,605			

(C)

Type	SNPs		INDELs	
	Count	Percent (%)	Count	Percent (%)
Intergenic	21,126,895	48.48	2,256,154	40.29
Upstream	9,787,878	22.46	1,511,052	26.99
Downstream	9,214,953	21.15	1,371,746	24.50
Exon	892,970	2.05	41,188	0.74
Intron	2,097,512	4.81	344,292	6.15
3' UTR	229,134	0.53	38,526	0.688
5' UTR	15,6452	0.36	24,691	0.44
Other	69,209	0.16	11,412	0.20
Total	43,575,003	100	5,599,061	100

Note: the upstream and the downstream were defined 2kb range on the upstream and downstream of gene.

Table S7. The candidate genes in GWAS. Related to Figure 5.

(A) The list of candidate locus associated with oil content in GWAS. (B) The list of 20 genes associated *Qfoc4.1* in GWAS.

(A)

Locus	Gene ID	Gene name	Annotation
Qfoc4.1			Shown in Table S7
Qfoc9.1	*Ip19174*	*IpRAF22*	Serine/threonine-protein kinase HT1-like [1]
Qfoc13.1	*Ip25692*	*IpLTPG3*	Lipid transfer-like protein [2]
Qfoc19.1	*Ip36111*	*IpDALL2*	DAD1-like lipase 2

(B)

Orthogroups	Gene ID	Gene name	Annotation
OG0000032	*Ip09793*	*IpSTP5*	Sugar transport protein 5
OG0005043	*Ip09794*	*IpINP1*	Hypothetical protein H0E87_004465
Unassigned	*Ip09795*	-	-
OG0021099	*Ip09796*	*IpSKU5*	Eukaryotic translation initiation factor 5B isoform X2
OG0011262	*Ip09797*	*IpLTPG30*	Hypothetical protein POTOM_007872
OG0004070	*Ip09798*	*IpBTL3*	Probable mitochondrial adenine nucleotide transporter BTL3
OG0008281	*Ip09799*	*IpHIP1*	Low quality protein: synaptonemal complex protein 1
Unassigned	*Ip09800*	*IpLTPG7*	Uncharacterized protein At5g01610
OG0010382	*Ip09801*	*IpA/N-InvF*	Hypothetical protein H0E87_004468
OG0007229	*Ip09802*	-	-
OG0002762	*Ip09803*	*IpPP2C*	Probable protein phosphatase 2C 9
OG0001101	*Ip09804*	*Ip GSM2*	Hypothetical protein H0E87_004475
Unassigned	*Ip09805*	*Ip TPPG*	Coiled-coil domain-containing protein 18 isoform X1
OG0011412	*Ip09806*	*Ip GAL2*	Signal peptidase complex protein
Unassigned	*Ip09807*	*IpCBL3*	Hypothetical protein POTOM_009973
OG0015845	*Ip09808*	*IpLPPbeta*	Hypothetical protein SADUNF_Sadunf02G0081400
OG0005154	*Ip09809*	-	Hypothetical protein DKX38_002491
OG0005896	*Ip09810*	*Ip WRKY21*	Hypothetical protein POTOM_007883
OG0002540	*Ip09811*	*Ip FMO*	Uncharacterized protein LOC7461780 isoform X1
OG0009126	*Ip09812*	-	-

Table S8. Individual fatty acids content of various fruits in *I. polycarpa*. Related to Figure 6.

(A) Individual fatty acids content between two distinct *IpSTP5* haplotypes among 42 accessions of *I. polycarpa.* (B) Individual fatty acids content in *IpSTP5* VIGS treated fruit of *I. polycarpa*.

(A)

Type	accession	16:0(μg/mg)			16:1 (μg/mg)			18:0 (μg/mg)			18:1(μg/mg)			18:2 (μg/mg)			18:3 (μg/mg)			20:0 (μg/mg)		
		rep1	rep2	rep3	rep1	rep2	rep3	rep1	rep2	rep3	rep1	rep2	rep3	rep1	rep2	rep3	rep1	rep2	rep3	rep1	rep2	rep3
IpSTP5 haplotype I	12BHL	16.71	16.16	14.23	8.56	8.73	8.38	1.26	0.78	0.73	7.07	7.54	7.06	83.12	86.17	85.88	3.65	2.74	2.64	0.14	0.10	0.10
IpSTP5 haplotype I	CAS07	34.69	37.46	37.58	13.24	15.45	14.76	1.07	1.10	1.10	13.29	15.50	14.44	91.59	104.79	103.30	2.70	3.08	2.99	0.15	0.16	0.16
IpSTP5 haplotype II	01XZ9	38.69	41.86	34.93	14.90	15.35	15.36	3.68	4.33	3.33	19.01	20.73	18.44	143.23	160.28	148.74	4.32	5.05	3.96	0.26	0.30	0.25
IpSTP5 haplotype II	02XZ5	26.77	23.36	22.60	2.76	1.88	2.10	2.97	2.46	3.20	9.06	8.45	10.70	120.12	95.98	99.52	2.97	2.45	3.25	0.21	0.18	0.21
IpSTP5 haplotype II	03XZ3	25.88	26.20	26.50	19.11	18.05	18.29	1.61	1.62	1.58	11.89	12.09	12.34	112.74	112.12	110.78	3.21	3.33	3.28	0.15	0.16	0.16
IpSTP5 haplotype II	05LZ	31.25	25.11	27.42	8.61	7.23	7.86	3.20	2.67	2.87	14.30	12.29	12.66	143.51	133.77	126.19	4.64	4.04	4.22	0.25	0.22	0.22
IpSTP5 haplotype II	06YY6	33.98	35.98	26.49	20.24	19.38	18.16	2.82	2.85	2.70	19.82	16.05	18.33	137.12	140.15	112.27	2.99	2.86	3.28	0.29	0.30	0.37
IpSTP5 haplotype II	08YY9	22.83	30.79	27.22	14.53	14.16	15.62	1.41	1.70	1.61	10.85	13.67	11.62	88.44	102.41	102.37	2.76	3.61	3.29	0.14	0.18	0.16
IpSTP5 haplotype II	09YY2	40.30	44.19	44.14	2.84	3.83	3.68	9.12	9.63	9.21	15.36	17.04	17.17	110.70	139.93	133.89	5.73	5.87	5.51	0.64	0.73	0.69
IpSTP5 haplotype II	10CAS1	29.62	31.91	29.91	15.98	16.55	14.88	2.52	2.55	2.63	12.05	12.36	9.42	129.85	148.11	127.35	3.97	4.07	4.29	0.33	0.34	0.23
IpSTP5 haplotype II	11YY7	35.42	36.15	40.34	13.25	14.69	14.32	2.49	2.54	2.68	15.93	15.60	16.85	143.85	141.62	157.16	3.94	3.83	3.92	0.34	0.35	0.35
IpSTP5 haplotype II	13YY3	32.99	34.62	30.30	9.38	11.85	9.70	1.99	2.74	2.27	10.82	15.28	13.73	137.81	164.11	109.88	2.33	3.30	2.58	0.17	0.21	0.16
IpSTP5 haplotype II	16ST3	23.75	24.49	22.77	3.97	3.44	4.29	1.37	1.05	1.85	16.40	15.11	17.18	115.25	99.35	103.41	2.69	2.21	2.86	0.15	0.12	0.27
IpSTP5 haplotype II	17ST1	30.66	28.15	23.67	12.07	10.96	10.08	1.68	1.55	1.30	15.65	14.22	12.93	130.40	114.96	98.68	2.88	2.51	2.24	0.14	0.13	0.11
IpSTP5 haplotype II	18ST2	35.02	33.31	34.08	17.41	16.24	17.13	2.58	2.44	2.61	21.14	18.70	21.82	152.16	143.51	151.28	3.00	3.01	3.64	0.30	0.28	0.33
IpSTP5 haplotype II	22CT52	19.87	18.89	24.96	8.08	7.00	9.43	0.93	0.94	1.49	9.11	9.19	9.00	95.66	98.65	108.13	2.26	2.25	3.00	0.13	0.13	0.25
IpSTP5 haplotype II	23QC71	22.40	27.69	30.20	16.12	17.34	17.43	1.35	1.43	1.81	11.31	11.87	14.01	102.03	105.39	130.47	2.75	2.97	3.22	0.14	0.14	0.22
IpSTP5 haplotype II	24CT2	47.94	50.88	46.38	20.18	22.89	20.05	4.88	4.13	4.01	29.74	30.82	28.56	156.94	157.41	149.05	3.92	3.86	3.48	0.23	0.28	0.26
IpSTP5 haplotype II	25YY8	33.55	34.89	37.47	12.15	14.29	12.62	1.41	1.64	1.56	14.03	14.69	15.28	107.31	124.54	119.73	2.99	3.37	3.15	0.25	0.27	0.27
IpSTP5 haplotype II	26YY5	30.47	25.81	27.27	10.68	9.16	10.21	2.54	1.08	1.16	12.06	11.90	12.63	121.36	103.30	117.85	1.95	2.44	2.61	0.17	0.15	0.15
IpSTP5 haplotype II	29ST7	35.43	32.03	34.17	9.72	11.02	9.06	3.44	2.99	3.33	21.20	23.29	23.51	135.95	126.79	130.18	4.02	3.82	3.80	0.25	0.26	0.26
IpSTP5 haplotype II	32ST8	31.48	26.40	30.92	6.37	6.11	6.63	5.36	5.38	5.67	21.22	19.74	22.02	151.78	143.45	159.45	4.18	4.13	4.82	0.37	0.36	0.39

IpSTP5 haplotype II	33CAS2	34.36	33.91	30.31	8.58	6.64	7.33	3.29	3.24	2.86	22.60	21.06	19.56	139.96	151.74	130.21	2.47	2.47	2.10	0.26	0.25	0.22
IpSTP5 haplotype II	35QC9	43.67	39.59	45.60	19.96	20.58	20.50	2.41	2.46	2.49	15.70	16.93	17.22	150.27	148.46	146.49	3.87	4.09	4.13	0.24	0.25	0.25
IpSTP5 haplotype II	36QC83	39.53	35.00	30.81	21.63	20.02	16.80	3.45	3.01	2.73	18.23	16.58	16.37	187.26	175.32	149.33	4.45	4.07	3.82	0.31	0.28	0.27
IpSTP5 haplotype II	37QC82	29.15	38.25	43.39	14.01	14.88	15.49	1.74	2.36	2.48	21.81	21.44	24.26	143.17	147.93	148.73	5.31	5.01	5.42	0.18	0.54	0.28
IpSTP5 haplotype II	38QC81	25.87	24.91	24.36	12.41	10.16	11.78	3.96	3.25	3.84	18.41	14.20	17.63	154.42	151.63	147.25	4.41	3.62	4.21	0.30	0.25	0.29
IpSTP5 haplotype II	39QC51	30.69	24.97	28.43	19.63	20.77	18.82	1.62	1.49	1.52	10.79	11.07	10.26	106.96	97.39	97.28	2.49	2.25	2.25	0.19	0.18	0.17
IpSTP5 haplotype II	41CT11	24.18	22.84	23.35	2.31	1.46	1.67	5.55	4.15	3.34	14.25	13.33	14.15	123.61	106.28	106.10	3.63	2.25	2.76	0.33	0.23	0.24
IpSTP5 haplotype II	42CT12	26.04	28.49	23.29	4.21	3.41	4.89	2.91	2.87	3.75	23.10	21.65	20.95	108.92	120.28	106.27	2.48	2.45	3.19	0.27	0.27	0.23
IpSTP5 haplotype II	43CT13	37.09	39.14	42.83	16.29	18.27	18.36	2.15	2.37	2.45	20.33	20.75	23.39	165.03	172.00	177.29	5.39	5.82	6.25	0.25	0.29	0.29
IpSTP5 haplotype II	44NQ1	22.56	24.11	23.83	1.04	1.43	1.27	2.84	3.88	3.54	19.30	24.75	22.91	141.67	159.17	122.99	2.30	3.07	3.01	0.16	0.22	0.20
IpSTP5 haplotype II	45NQ2	20.78	19.04	19.88	7.11	6.84	7.73	2.82	2.56	2.42	6.19	6.75	6.35	140.39	138.51	138.67	2.38	2.50	2.66	0.22	0.25	0.25
IpSTP5 haplotype II	46NQ3	37.97	38.71	40.85	12.66	14.91	13.07	2.31	2.46	2.46	19.50	18.61	20.96	143.04	150.58	148.90	6.28	6.85	6.74	0.28	0.32	0.30
IpSTP5 haplotype II	48NQ5	39.01	36.80	38.76	16.29	17.12	15.26	1.66	1.64	1.80	14.33	14.33	14.17	144.43	143.36	138.87	5.49	5.36	5.16	0.22	0.22	0.30
IpSTP5 haplotype II	49NQ6	25.26	26.91	29.72	16.49	17.15	17.86	1.63	1.66	1.75	12.55	12.83	13.24	114.62	142.77	124.23	3.29	3.32	3.29	0.16	0.16	0.21
IpSTP5 haplotype II	50CT6	42.13	39.37	43.38	7.88	8.36	7.69	2.33	3.34	2.07	14.95	18.60	16.51	164.34	169.41	167.38	3.07	4.35	2.43	0.19	0.26	0.15
IpSTP5 haplotype II	51CT7	31.87	31.65	31.73	3.40	3.23	3.35	4.33	4.16	4.26	26.42	26.50	25.82	155.55	150.36	157.68	3.44	3.29	3.31	0.28	0.28	0.28
IpSTP5 haplotype II	52CT8	34.87	29.62	30.13	10.04	8.25	8.05	2.70	2.06	1.96	17.12	14.46	14.37	154.86	145.07	148.68	4.91	3.74	3.61	0.22	0.18	0.17
IpSTP5 haplotype II	55CT3	27.14	24.16	26.29	7.40	7.61	6.52	2.32	1.89	1.91	15.07	13.98	11.41	123.03	109.90	114.68	3.29	2.87	2.51	0.20	0.15	0.25
IpSTP5 haplotype II	57CT51	24.46	23.83	23.20	3.96	3.63	3.29	2.18	2.77	2.24	10.17	10.61	10.63	122.73	119.35	116.65	7.29	8.81	6.39	0.40	0.34	0.36
IpSTP5 haplotype II	58QC73	29.28	28.84	29.97	8.65	8.81	6.30	2.17	2.54	1.78	12.62	12.17	13.36	168.00	152.77	141.11	2.06	2.45	1.58	0.17	0.19	0.12

(B)

Type of *IpSTP5* VIGS treated fruit	16:0(μg/mg)			16:1 (μg/mg)			18:0 (μg/mg)			18:1(μg/mg)			18:2 (μg/mg)			18:3 (μg/mg)			20:0 (μg/mg)		
	rep1	rep2	rep3	rep1	rep2	rep3	rep1	rep2	rep3	rep1	rep2	rep3	rep1	rep2	rep3	rep1	rep2	rep3	rep1	rep2	rep3
pTRV1+pTRV2	3.19	2.68	2.84	0.02	0.02	0.02	1.20	0.84	0.99	3.47	2.44	2.66	17.89	12.93	14.37	2.02	1.98	2.17	0.14	0.14	0.14
pTRV1+pTRV2_*IpSTP5_1*	2.17	2.88	2.66	0.02	0.02	0.02	0.56	0.69	0.65	0.47	1.16	0.69	9.28	9.67	9.46	1.88	2.64	2.16	0.15	0.27	0.26
pTRV1+pTRV2_*IpSTP5_2*	2.78	2.67	2.55	0.02	0.02	0.02	0.71	0.66	0.63	1.56	1.33	1.21	11.28	10.72	9.82	2.18	2.28	2.20	0.18	0.17	0.18

Note: μg/mg was calculated with dried fruit; 16:0, palmitic acid; 16:1, palmitoleic acid; 18:0, stearic acid; 18:1, oleic acid; 18:2, linoleic acid; 18:3, linolenic acid; 20:0, arachidic acid.

Table S9. The list of primers in this study. Related to Figure 5 and 6.

Primer name	Primer sequence	Purpose
IpPP2A-F	TACCCCTTACAGCCCTCATT	RT-qPCR
IpPP2A-R	GTCCATGTTCTCTCCAATCTCT	
TRV-*IpSTP5_1*-F	GAAGGCCTCCATGGGGATCCTGTACCCGCAGCTCTAATGAC	pTRV2-*IpSTP5_1*
TRV-*IpSTP5_1*-R	GGACATGCCCGGGCCTCGAGAACCATACCCGTAGACACAAGG	vector
TRV-*IpSTP5_2*-F	GAAGGCCTCCATGGGGATCCATTTCAGGCACGAAGCAGG	pTRV2-*IpSTP5_2*
TRV-*IpSTP5_2*-R	GGACATGCCCGGGCCTCGAGGACCAAGATTTGCCTACCAC	vector
TRV-*IpTGW6*-F	GAAGGCCTCCATGGGGATCCTTAGCACAAGGCCTCGCTTT	pTRV2-*IpTGW6*
TRV-*IpTGW6*-R	GGACATGCCCGGGCCTCGAGAACCAATCCAGAGCTTACCAT	vector

REFERENCES

1. Ramachandiran, I., Vijayakumar, A., Ramya, V. Rajasekharan, R. (2018). *Arabidopsis* serine/threonine/tyrosine protein kinase phosphorylates oil body proteins that regulate oil content in the seeds. Sci. Rep. *8*: 1154.
2. Li-Beisson, Y., Shorrosh, B., Beisson, F., Andersson, M. X., Arondel, V., Bates, P. D., Baud, S., Bird, D., Debono, A., Durrett, T. P., et al. (2013). Acyl-lipid metabolism. The *Arabidopsis* Book *11*, e0161.