INTRODUCTION

Jojoba (Simmondsia chinensis) plants are dioecious desert shrubs that are native to the Sonoran and Baja California regions of North America (Fig. 1A and fig. S1, A to D; (1, 2)). Jojoba is taxonomically classified as the single member of its family Simmondsiaceae (3). Jojoba is widely regarded for its unusual seed oil that consists primarily of liquid wax esters [WEs; Fig. 1A (4–6)]. The seeds of jojoba are one of the world’s only known sustainable sources of liquid WEs and have been used as an eco-friendly replacement for the similar oils that were once harvested from the spermatocar organ of the sperm whale (Physeter macrocephalus), which nearly drove this species to extinction (7, 8). The WEs from jojoba seeds are esters of a monounsaturated long-chain fatty acid (C20:C24) and a fatty alcohol (C20:C24) and can accumulate up to 60% of the seed weight (5, 9, 10). Jojoba seed oils have been demonstrated to have high compatibility with human sebum and promote retention of skin moisture, and because of this, they are highly valued for their use in a wide range of cosmetic products (11–14). The cosmetics industry is currently the largest market for jojoba oil, and consumer demand continues to rise for natural skin care products [e.g., moisturizers, makeup, shampoos, and conditioners (15)]. In addition to its importance to the cosmetics industry, jojoba oils are also widely regarded for their excellent mechanical lubricity properties including stability at high temperatures and pressures (16), anti-foaming, antiwear, and antirust properties (17, 18), and oxidative stability (19). Because of the high oil content of its seeds, economic value, and the capacity of the plants to grow in hot, arid climates (35° to 48°C), jojoba has garnered considerable attention for domestication in some of the world’s most unfavorable environments. Currently, jojoba is commercially grown in the United States, Israel, Peru, Argentina, Australia (www.iiec.net), and India (20) on these countries’ most marginal lands.

Despite the commercial cultivation of jojoba on marginal lands, crop yields can be reduced markedly by unexpected rainfall, frost, or high temperatures during flowering (2). Over the last decade, there have been considerable efforts to engineer a transgenic row crop with WEs in the seeds including canola (Brassica sp.), camelina (Camelina sativa), crambe (Crambe abyssinica), and Lepidum (Lepidum campestre) (21–24). These efforts were, in part, made possible by the cloning of the jojoba fatty acyl-CoA reductase (ScFAR; (25)) and the wax synthase (ScWS; (26)), which catalyze terminal steps in the synthesis of WEs. However, expression of these jojoba sequences and other allied proteins has resulted in only modest production of WEs in transformed oilseeds (10 to 20%), and nearly all engineered lines were reported to have markedly decreased seed germination rates, especially in seeds with higher WE content (21, 23). As jojoba seeds can accumulate up to 60% seed oil with >95% WE with no observable germination effects, there is likely much to be learned about how jojoba seeds synthesize, package, and mobilize WEs. These challenges and the current lack of genetic resources for jojoba have, in part, motivated our efforts to sequence and annotate the jojoba genome.

RESULTS AND DISCUSSION

Compared to other eukaryotes, plants have considerably complex genomes, which can range in size from 82 Mb [Utricularia gibba,
The jojoba genome contained 614.7 Mb (69.33%) transposable elements, with LTR/Gypsy and LTR/Copia occupying 21.5 and 21.4% of the genome, respectively (table S4). Overall activity of LTRs was much lower than that of Spinacia oleracea and Beta vulgaris, which are in the same order, Caryophyllales (fig. S3). Through a combination of ab initio prediction, homology search, and RNA-seq–aided prediction, 23,490 protein-coding genes were annotated in the jojoba genome, with a mean coding length of 1231 base pairs (bp) and an average of six exons per gene (tables S1 and S5). The protein-coding genes were supported by the Illumina RNA-seq reads, and 91.0% of these genes had significant functional annotation matches to the InterPro and Pfam databases (table S6). In addition, we identified sequences for 24,178 non-coding RNAs consisting of ribosomal RNAs, transfer RNAs, microRNAs, and small nuclear RNAs (table S7). Gene region completeness was evaluated by RNA-seq data. On average, more than 95.6% of the RNA-seq reads could be mapped to the jojoba genome assembly (table S8). Further assessments of genome completeness were evaluated with CEGMA (conserved core eukaryotic gene mapping approach) and BUSCO [benchmarking universal single-copy orthologs; (33, 34)], revealing that 98.8% of conserved core eukaryotic genes from CEGMA and 93.5% from BUSCO were captured in our assembly (table S9).

Evolutionary analysis was performed on the assembled jojoba genome (Fig. 2). First, an evolutionary history was reconstructed by comparing...
the jojoba chromosome assembly to the pre-γ and post-γ ancestral eudicot karyotype (AEK) and three of the evolutionarily least rearranged plant genomes from the Rosid clade (Fig. 2A), *Vitis vinifera* [grape; (35)], *Theobroma cacao* [chocolate; (36)], and *Prunus persica* [peach; (37)]. This evolutionary scenario suggests that jojoba underwent a whole-genome triplication shared among all eudicots and that jojoba diverged from *Vitis vinifera* (Fig. 3, A to D), where WEs were localized primarily to the cotyledonary tissues (WE 42:2 shown in red, Fig. 3, B and D), and TAGs were primarily restricted to the embryonic axis (TAG 62:3 shown in green, Fig. 3, C and D). Total ion counts for both lipid classes were normalized to tissue areas on a per-pixel basis (39, 43, 44), which indicated that cotyledons had a significant (Student’s two-tailed *t* test; *P* < 0.001) 4× greater total WE signal intensity than the axis, whereas the embryonic axis had a significant (Student’s two-tailed *t* test; *P* < 0.001) 21× higher level in total TAG signal intensity (Fig. 3, E and F). To complement the MALDI-MSI distribution data, total lipids from an intact mature jojoba seed were imaged by magnetic resonance imaging [MRI; (43, 46–48)]. The MRI images revealed that total lipids, presumably WEs, were highly enriched in the cotyledon tissues, whereas the signal from the axis was noticeably less (fig. S6). In addition, the WE and TAG composition of lipid extracts from the cotyledon and embryonic axis tissues of mature jojoba seeds were determined by ultra-performance liquid chromatography (UPLC)–nanoESI (nano-electrospray ionization)–tandem mass spectrometry (MS/MS) and largely supported the compositions obtained by MALDI-MSI (figs. S6 and S7). Compositions here were consistent with previous findings of jojoba WE and TAG compositions (10, 49).

In consideration of the heterogeneous lipid metabolite distributions observed in jojoba cotyledons and embryonic axis tissues, transcriptome analysis was performed on these two tissues from seeds collected during the peak of oil accumulation (~75 days after pollination (DAP); Fig. 3, G and H, and tables S14 and S15), from whole developing seeds from early, mid, late developmental time points, and from mature, desiccated seeds (Fig. 3H and table S16). From the jojoba genome annotation, we identified genes known to be involved in acyl-lipid metabolism (Fig. 3, G and H) and the previously described jojoba fatty acyl-CoA elongase 1 (*ScFAE1*; (50), *ScFAE2* (25), and *ScWS* (26) and used them for expression analysis. In Fig. 3G, the acyl-lipid metabolic pathway is diagrammed with the corresponding genes from the jojoba genome, and a magenta-to-blue colored box indicates the differential transcript abundances (log2 fold scale) between the cotyledons (magenta) and embryonic axis (blue). Genes known to be involved in WE synthesis including the *ScFAE1*, *ScFAE2*, and *ScWS* were all highly expressed and significantly biased toward the cotyledonary tissues (Fig. 3G and table S14). In addition, nearly all genes involved in fatty acid synthesis and export from the plastid were similarly significantly highly expressed and biased toward the cotyledons (Fig. 3G and tables S14 and S17). The expression pattern of these genes likely reflects the cotyledons’ involvement in WE synthesis, glycerolipid synthesis, and lipid droplet (LD) packaging (Figs. 3 and 4).

Application of tissue-specific RNA-seq and matrix-assisted laser desorption/ionization–mass spectrometry imaging (MALDI-MSI) has previously revealed heterogeneous distributions of lipid metabolites and gene transcripts involved in lipid metabolism across different seed tissues, reflecting an underappreciated spatial regulation to oilseed lipid metabolism (21, 38–45). Here, jojoba seeds were demonstrated to compartmentalize WEs and express genes involved in storage lipid synthesis in a tissue-specific manner. Longitudinal seed sections from mature jojoba seeds were analyzed by MALDI-MSI, and these data were searched for major molecular species of WEs and triacylglycerols (TAGs). Unexpectedly, the MALDI-MSI analysis revealed that these two major seed storage lipids were differentially enriched in the seed tissues (Fig. 3, A to D), where WEs were localized primarily to the cotyledonary tissues (WE 42:2 shown in red, Fig. 3, B and D), and TAGs were primarily restricted to the embryonic axis (TAG 62:3 shown in green, Fig. 3, C and D). Total ion counts for both lipid classes were normalized to tissue areas on a per-pixel basis (39, 43, 44), which indicated that cotyledons had a significant (Student’s two-tailed *t* test; *P* < 0.001) 4× greater total WE signal intensity than the axis, whereas the embryonic axis had a significant (Student’s two-tailed *t* test; *P* < 0.001) 21× higher level in total TAG signal intensity (Fig. 3, E and F). To complement the MALDI-MSI distribution data, total lipids from an intact mature jojoba seed were imaged by magnetic resonance imaging [MRI; (43, 46–48)]. The MRI images revealed that total lipids, presumably WEs, were highly enriched in the cotyledon tissues, whereas the signal from the axis was noticeably less (fig. S6). In addition, the WE and TAG composition of lipid extracts from the cotyledon and embryonic axis tissues of mature jojoba seeds were determined by ultra-performance liquid chromatography (UPLC)–nanoESI (nano-electrospray ionization)–tandem mass spectrometry (MS/MS) and largely supported the compositions obtained by MALDI-MSI (figs. S6 and S7). Compositions here were consistent with previous findings of jojoba WE and TAG compositions (10, 49).

In consideration of the heterogeneous lipid metabolite distributions observed in jojoba cotyledons and embryonic axis tissues, transcriptome analysis was performed on these two tissues from seeds collected during the peak of oil accumulation (~75 days after pollination (DAP); Fig. 3, G and H, and tables S14 and S15), from whole developing seeds from early, mid, late developmental time points, and from mature, desiccated seeds (Fig. 3H and table S16). From the jojoba genome annotation, we identified genes known to be involved in acyl-lipid metabolism (Fig. 3, G and H) and the previously described jojoba fatty acyl-CoA elongase 1 (*ScFAE1*; (50), *ScFAE2* (25), and *ScWS* (26) and used them for expression analysis. In Fig. 3G, the acyl-lipid metabolic pathway is diagrammed with the corresponding genes from the jojoba genome, and a magenta-to-blue colored box indicates the differential transcript abundances (log2 fold scale) between the cotyledons (magenta) and embryonic axis (blue). Genes known to be involved in WE synthesis including the *ScFAE1*, *ScFAE2*, and *ScWS* were all highly expressed and significantly biased toward the cotyledonary tissues (Fig. 3G and table S14). In addition, nearly all genes involved in fatty acid synthesis and export from the plastid were similarly significantly highly expressed and biased toward the cotyledons (Fig. 3G and tables S14 and S17). The expression pattern of these genes likely reflects the cotyledons’
capacity to serve as the major seed storage organs for WEs. By contrast, the jojoba acyl-CoA:diacylglycerol O-acyltransferase 1 (ScDGAT1), which is a terminal enzyme in TAG synthesis, was significantly expressed and biased toward the embryonic axis tissues. The differential expression of the genes encoding the jojoba ScWS and ScDGAT1 offers a molecular explanation for the observed WE and TAG distributions. In addition to differences in LD gene expression profiles, the morphology of LDs was also different between the cotyledons and embryonic axis, where LDs in cotyledons were significantly larger on average, ~1.0 μm in diameter (61), than those in the embryonic axis, ~0.65 μm in diameter (Fig. 4, C to E). Since the LDs of these two tissues contain different types of hydrophobic molecules in their cores (WEs and TAGs), differences in LD surface protein composition may be required for the proper packaging of WE-containing LDs versus TAG-containing LDs and also may influence LD size. The prevalence of LDAP1 (Fig. 4, A, B, and F), in particular, may have functional significance, since this isoform is not substantially expressed in other oilseeds during TAG accumulation (57, 58). LDAPs are homologous to small rubber particle proteins, which localize to the surface of and stabilize rubber particles, compartments specific for the packaging of the hydrophobic polyisoprenoids [reviewed in (62)]. Perhaps jojoba LDAP1 is part of the process specifically required for the proper packaging of WEs.

In the endoplasm reticulum membrane, SEIPIN proteins are involved in the biogenesis of LDs and are not LD surface proteins per se (59, 60); however, one isoform appeared to be significantly differentially expressed in the embryonic axis. Because different lipids and pathways predominated in different seed parts, we examined the cellular LD packaging machinery for potential differences, both in terms of transcripts for major LD proteins and in terms of proteins on isolated LDs (Fig. 4, A to F). Differential transcript abundances were noted for transcripts for several major LD proteins including oleosins (51, 52), stereoelosins (53), caeleosins (54) and LD-associated proteins [LDAPs; (55–58)]. Several oleosins and one LDAP especially seemed to be significant differentially expressed in cotyledonary tissues (Fig. 4, A and B), and this appeared to be reflected in the relative abundances of these proteins on the isolated LDs from cotyledons (Fig. 4F and tables S18 to S21).
Fig. 3. Heterogeneous metabolite distributions and gene expression levels in jojoba seeds and tissues. (A) Medial longitude jojoba seed section with labeled seed tissues and structures. Scale bar, 1.0 mm. (B) MALDI-MS image of the major molecular species of WE 42:2 (m/z 655.579, WE 22:1:20:1), which was enriched in the cotyledonary tissues. (C) MALDI-MS image of the major molecular TAG species TAG 62:3 (m/z 1035.872, TAG 20:1:20:1:22:1) highly enriched in the embryonic axis. (D) Composite MALDI-MS image of WE 42:2 (red) and TAG 62:3 (green) demonstrating the differential enrichment of WEs and TAGs in jojoba seeds. In images above, color scale (ion intensity) for each lipid was adjusted to emphasize distribution and does not represent absolute amounts of lipid. Mass tolerance for MALDI-MS images was set to 4 parts per million. (E and F) Total ion signal from each of the cotyledons and embryonic axis was acquired for these tissue areas of the MALDI-MS image. Signal was normalized to the total ion count and then to the number of pixels in the embryonic axis or cotyledons, respectively. The overall WE (E) and TAG (F) content of virtually “dissected” cotyledon and embryonic axis tissues was calculated by summing the intensities of all detected WE and TAG molecular species. Student’s t-test was used to calculate significance where *** is P < 0.001. (G) Depiction of metabolic pathways and genes leading to WE and TAG biosynthesis in jojoba. Colored boxes (blue to magenta color scale) adjacent to gene names indicate gene expression bias toward embryonic axis tissues (blue) or toward cotyledon tissues (magenta) and represent a log2 fold change of expression levels (n = 5, *q < 0.05, **q < 0.01, ***q < 0.001). Acyl-CoA, acyl coenzyme A; GPAT, glycerol-3-phosphate acyltransferase; PAP, phosphatic acid phosphatase. (H) Heat map of gene expression levels of genes involved in glycerolipid synthesis and storage in jojoba across different jojoba seed tissues and developmental stages. FATB, fatty acyl-ACP thioesterase B; FATA, fatty acyl-ACP thioesterase A; SAD, stearoyl-ACP desaturase; LPAT, lysophosphatidic acid acyltransferase; PLDo, fatty acyl-CoA, acyl coenzyme A; GPAT, glycerol-3-phosphate acyltransferase; PAP, phosphatic acid phosphatase; FAR, fatty acyl-CoA reductase; KCS, ketoacyl-CoA synthase; FAE, fatty acyl elongation; PC, phosphatidylcholine; LPC, lysophosphatidylcholine; CO, cotyledons; EA, embryonic axis; SC, seed coat; WS, whole seed; EDS, early developing seed; DS, dry seed; L, leaf.

Overall, our results provide a reference-quality genome for jojoba that facilitates the placement of this minimally studied plant species into evolutionary context. In addition, these genomic resources enabled insights into a previously unrecognized heterogeneity of neutral lipid synthesis and storage in jojoba seeds. These resources will be valuable to others interested in the utilization of this species as an oilseed crop and will also fill important knowledge gaps for this taxonomically isolated dioecious species and its life history and adaptations.

MATERIALS AND METHODS
Plant material
S. chinensis (accession no. PARL 940) developing seeds were collected from plants grown under rainfed conditions at the U.S. Department of Agriculture, U.S. Arid Land Agricultural Research Center, Maricopa, AZ (latitude, 33.077514; longitude –111.974276; elevation, 375 m). Developing seeds were collected from wind-pollinated plants. Developing seeds for the RNA-seq developmental time series were collected in 2018 and staged developmentally by their seed weight, with average seed weights of 0.2 g (early), 0.45 g (mid), and 0.7 g (late). Seeds were frozen in liquid nitrogen immediately after harvesting and stored at −80°C. Developing seeds for the tissue-specific RNA-seq experiments were collected in 2017, and the average seed weight of these tissues was ~0.45 g and were at the mid to mid-late oil accumulation seed stage. Seeds were vacuum-infiltrated with RNAlater solution immediately after harvesting and removal from the capsule and stored at −20°C. Mature seeds from these plants were harvested after seeds were desiccated and used here for additional analyses.

Vegetative plant tissues used for gene annotation were collected from 1-year-old jojoba plants grown at the University of North
Fig. 4. Gene expression and protein levels of associated with jojoba LDs. (A) Illustration representing an LD budding from the surface of the endoplasmic reticulum (ER) adorned with various proteins. Colored boxes (blue to magenta scale) adjacent to gene names indicate gene expression bias toward embryonic axis tissues (blue) or toward cotyledon tissues (magenta) and represent a log, fold change of expression levels (n = 5, *q < 0.05, **q < 0.01, ***q < 0.001). (B) Heat map of gene expression levels of genes involved in LD storage and packaging across different jojoba seed tissues and developmental stages. LDAP, lipid droplet-associated protein; LDIP, LDAP-interacting protein. (C and D) Confocal micrographs of BODIPY-stained LDs from tissues of jojoba (C) cotyledon and (D) embryonic axis tissues. (E) Size comparison of LDs from jojoba cotyledons and embryonic axis tissues. Student’s t test was used to calculate significance where *** is P < 0.001. (F) Bar graph (of 100%) representing an estimation of the percentage of oleosin, steroleosin, caleosin, and LDAP proteins on LDs from jojoba cotyledons and embryonic axis tissues. Proportions of known LD proteins were calculated from normalized MS peptide counts.

Texas greenhouses at 30°C under long-day (~16 hours) conditions, with day length supplemented by Na-vapor lamps. Tissues were immediately frozen in liquid nitrogen after harvest and stored at ~80°C.

Jojoba seed germination and plant growth
Mature jojoba seeds were sterilized in 50% bleach solution for 1 hour at room temperature, and then, seeds were washed/imbibed for 16 hours under running water. Seeds were germinated on wet vermiculite in the dark. Vermiculite moisture content was checked every 2 days and adjusted accordingly. After germination, seedlings were moved from the dark and transplanted to soil. Seed germination and seedling growth conditions were 100 μmol m −2 s −1 16-hour light/8-hour dark at 28°C.

DNA sequencing library construction and sequencing
High-quality genomic DNA was extracted from leaves of 3-month-old plant using the DNAsecure Plant Kit (DP320, www.tiangen.com). DNA samples and tissues from a single plant were sent to Novogene (www. novogene.com) for genome sequencing and Hi-C assembly, respectively.

Libraries for SMRT PacBio genome sequencing were constructed according to the released protocol from PacBio. Approximately 20 μg of high-quality genomic DNA was sheared to approximately 20 kb and evaluated by an Agilent 2100 Bioanalyzer. After shearing, DNA samples were subjected to damage and end repair, blunt-end adaptor ligation, and size selection. The final libraries were sequenced by the PacBio Sequel platform (Pacific Biosciences).

Libraries for Illumina sequencing were generated using the Truseq Nano DNA HT Sample preparation Kit (Illumina, USA) according to the manufacturer’s protocol. The DNA sample was fragmented to generate 350 bp in size by sonication. Resulting DNA fragments were end-polished, A-tailed, and ligated with the full-length adapter for Illumina sequencing with further polymerase chain reaction (PCR) amplification. PCR products were purified by the AMPure XP system and quantified by real-time PCR. The Illumina libraries were sequenced by the Illumina HiSeq X Ten platform.

For Hi-C libraries, jojoba leaves from a single plant were fixed with 1% formaldehyde solution in MS buffer and were used for the preparation of two independent libraries. Subsequently, the DNA libraries were subjected to nuclei extraction, nuclei permeabilization, chromatin digestion (Hind III), and proximity ligation treatments. The constructed libraries were sequenced on the Illumina HiSeq X Ten platform.

Estimation of the genome size and heterozygosity
The genome size was estimated by K-mer frequency analysis. The distribution of K-mer depends on the characteristic of the genome and follows a Poisson’s distribution. Before assembly, the K-mer distribution of 100G Illumina short reads was generated using Jellyfish [v2.1.3; (63)], and we obtained an estimated haploid genome size of 1003.02 Mb with a 0.76% heterozygous rate.

Hi-C–based genome assembly
The S. chinensis genome was de novo assembled using FALCON (https://github.com/PacificBiosciences/FALCON) based on PacBio long reads. Contigs were polished using raw PacBio long reads and corrected using Illumina short reads with Pilon [v1.22; (64)]. The raw assembled genome consisted of 994 contigs with a N50 length of 5.21 Mb. On the basis of Hi-C chromatin interactions information, using 3d-dna [v180419; -r 0 -l 5000; (65)] software and manual correction, scaffolds were assembled into 26 chromosomes. As a result, we obtained a high-quality reference genome of 887 Mb with a N50 length of 38.97 Mb and anchored 99.78% scaffolds into chromosome scale.

Genome assessment
CEGMA (34) pipeline analysis was used to validate the genome completeness, and 93% completeness was obtained of Core Eukaryotic
Genes database. In addition, the BUSCO (v3.0.1) (33) database was also used to assess the genome assembly.

**Repeat annotation**

Transposable elements in the genome assembly were identified both at the DNA and protein levels. RepeatModeler (www.repeatmasker.org/RepeatModeler/) software was used to develop a de novo transposable element library. RepeatMasker (www.repeatmasker.org) was applied for identifying transposable element from Repbase (66) and de novo library. At the protein level, RepeatProteinMask (www.repeatmasker.org) was used to conduct WU-BLASTX searches against the transposable element protein database. Overlapping transposable elements belonging to the same type of repeats were combined together.

**Gene prediction**

Gene annotation of S. chinensis genome was conducted by combining de novo prediction, homology information, and RNA-seq data. First, Augustus [v3.2.1; (67)], GlimmerHMM [v3.0.4; (68)], SNAP (http://korflab.ucdavis.edu/software.html), Geneid [v1.4.4; (69)], and Genscan [v1.0; (70)] were used on the repeat masked genome with trained parameters. Then, the nonredundant proteins from six sequenced species, *Amaranthus hypochondriacus*, *Oryza sativa*, *Dianthus caryophyllus*, *B. vulgaris*, and *Fagopyrum tataricum* were mapped onto the *S. chinensis* genome by using TBLASTN [v2.7.1+(71)] with an E-value cutoff of 1 × 10⁻5.

For each BLAST hit, Genewise [v2.4.1; (72)] was used to predict the exact gene structure in the corresponding genomic regions. Furthermore, RNA-seq data were mapped to the genome using TopHat [v2.1.1; (73)], and Cufflinks [v2.2.1; (74)] was used to assemble transcripts to gene models. Last, all predictions were combined with EVidenceModeler [EVM v1.1.1; (75)] to get a nonredundant gene set, and PASA [v2.3.0; (76)] was used to correct the predicted result and annotate alternatively spliced isoforms to finalize the gene set.

Functional annotation of protein-coding genes was evaluated by BLASTP [v2.7.1+(71)] with an E-value cutoff of 1 × 10⁻5 using two integrated protein sequence databases—SwissProt and TrEMBL (77). Protein domains were annotated by searching InterPro and Pfam databases, using InterProScan [v5.28; (78)] and Hmmer [v3.1b2; (79)], respectively. Gene ontology (GO) terms for each gene were obtained from the corresponding InterPro or Pfam entry. The pathways, in which the gene might be involved, were assigned by BLAST against the Kyoto Encyclopedia of Genes and Genomes (KEGG) (80) database.

**Gene family analysis**

Five sequenced plant genomes (*S. chinensis*, *A. thaliana*, *B. vulgaris*, *J. curcas*, and *R. communis*) were selected for gene family analysis. Proteins from these genomes were aligned to each other using BLASTP. OrthoMCL [v2.0.9; (88)] was used to cluster proteins and identify paralogs and orthologs.

**Phylogenetic reconstruction and gene family expansion/contraction**

A total of 501 single-copy orthologous genes were identified using OrthoMCL for *S. chinensis* and 15 other flowering plants (*A. thaliana*, *B. vulgaris*, *Brassica oleracea*, *Glycine max*, *Populus trichocarpa*, *J. curcas*, *O. sativa*, *R. communis*, *Solanum tuberosum*, *Solanum lycopersicum*, *T. cacao*, *Nelumbo nucifera*, *V. vinifera*, *Zeas mays*, and *S. oleracea*). Amino acid sequences encoded by single-copy genes from these 16 species were aligned using MUSCLE [v3.8.31; (89)], and RaxML [v8.2.12; (90)] was used to construct a phylogenetic tree based on multiple sequence alignments. Last, the iTOL (91) tools were used to visualize the phylogenetic tree. To estimate divergence time of all the 16 species in the phylogenetic tree, we used the MCMCTree program in the PAML package [v4.8a; (92)], and *A. thaliana*–*B. oleracea* split time (mean, 25 Ma ago) and monocot–dicot split time (mean, 150 Ma ago) were chosen. The MCMCTree runs 505,000 iterations to calculate divergence time. CAFÉ [v3.1; (93)] was used to calculate the expansion and contraction of gene family numbers based on the phylogenetic tree and gene family statistics.

**Total RNA extraction and RNA-seq library preparation**

Developing seeds were removed from RNAlater, and the seed coat, cotyledons, and embryonic axis tissues were hand-dissected. Tissues of the whole developing seeds, cotyledons, embryonic axis, and seed coats were pulverized in liquid nitrogen into a fine powder. Tissues of three seeds were combined per replicate, with five replicates per tissue (n = 5 per tissue). For the total RNA extraction, 100 mg of tissue (whole developing seeds, seed coat, and cotyledons) per replicate was used. Because of the small size of the embryonic axis tissues, only 25 mg of tissue was used per replicate.

Total RNA was isolated using a combination of hot borate buffer and the Qiagen RNeasy Plant Mini Kit (94). Assessment of total RNA quality and quantity was accomplished using a Qubit analyzer and an Agilent 2100 Bioanalyzer.

A sequencing library was prepared from the total RNA for 2 × 150 bp sequencing (n = 5 per tissue, n = 1 for mature seed) using an Illumina NextSeq. The sequencing reactions were prepared at the University of North Texas (UNT) BioDiscovery Institute Genomics Core Facility from 1.0 μg of total RNA using the Illumina TruSeq Stranded mRNA prep kit with minor modifications. The Frag-Prime enzymatic digestion step was reduced from 8 to 4 min. After the adapter ligation of the sequencing library, the average fragment length was approximately 420 bp (2 × 150 bp library), measured by an Agilent 4200 TapeStation. The 2× 150 paired-end sequencing reads were performed using an Illumina NextSeq instrument with a high-output cassette (120 Gb).

**Read preparation for gene expression analysis**

For gene expression analysis, the 2 × 150 bp raw reads from Illumina RNA-seq were trimmed to remove any remaining adaptor sequences using Trimmomatic [v0.32; (95)]. Read quality after trimming was assessed using FastQC toolkit.
Processed reads were used for gene expression analyses using STAR using the default parameters. Quantification of differential gene expression analysis was completed using DESeq2 [v3.7; (96, 97)] software using the default settings. The significance of differential gene expression was calculated as \( q \) values \((q < 0.05, **q < 0.01, ***q < 0.001)\), which are more stringent than \( P \) values, in that they also account for the false discovery rate rather than the false-positive rate (98). Principal components analysis and gene expression analysis plots were generated using the DESeq2 package pcaExplorer (96, 97). Venn diagrams were assembled using an online Venn diagram builder from http://bioinformatics.psb.ugent.be/webtools/Venn/.

**Seed embedding and cryosectioning**

A 10% (w/v) porcine gelatin solution was prepared and equilibrated in a 40°C water bath with shaking for 2 hours. Mature desiccated jojoba seeds were embedded in the prepared gelatin solution, frozen at −80°C overnight, and then transferred to −20°C for 48 hours before sectioning. Embedded seeds were cryosectioned at a tissue thickness of 30 μm using a cryo-microtome (Leica Microsystems). Cryosections were collected using Cryo-Tape (Leica Microsystems) and then taped to glass slides. Slides containing the thaw-mounted sections were lyophilized for 6 hours and then stored in a desiccator until processed for MALDI-MSI. All MALDI-MSI occurred within 36 hours of cryosectioning. Bright-field images were taken for all sections used for MSI.

**Matrix application and MALDI-MS imaging analysis**

For MALDI-MSI, the matrix 2,5-dihydroxybenzoic acid (DHB) was used for analysis of WEs and TAGs. DHB was applied by sublimation using an adapted method developed by Hankin et al. (99). MALDI-MSI data were collected on a hybrid MALDI-LTQ-Orbitrap XL mass spectrometer (Thermo Fisher Scientific). The instrument was equipped with a Nd laser with a spot size of ~40 μm. MALDI-MSI data acquisition conditions were as follows: Laser energy was 12 μJ per pulse, a raster step size of 40 μm, and 10 laser shots per raster step with one sweepshot; data were acquired using the Orbitrap mass analyzer with a resolution of 60,000, and data were collected with a mass/charge ratio (m/z) scan range of 500 to 1200. Raw mass spectra were processed into MALDI-MS images using ImageQuest software (Thermo Fisher Scientific). Images were generated by searching the exact mass of WE or TAG molecular species using a 4–parts per million mass tolerance. Processed images were normalized to the total ion count for each pixel and then presented on a red (WE) or green (TAG) scale, where the brightest color represents the highest intensity. After processing, the background of the images was removed, and the images were placed on a black background. The virtual dissection intensities were achieved by selecting the area of each cotyledon and embryonic axis tissues with open source software Metabolite Imager (100), averaging the intensities for all WE and TAG molecular species across all cotyledon or embryonic axis tissues, and lastly normalizing the intensities to the number of pixels per tissue type (32,650 cotyledon pixels and 431 embryonic axis pixels). This method is described in more detail elsewhere (43, 44).

**Nuclear MRI of jojoba seeds**

The measurement of the lipid distribution of the jojoba seed was measured on a Bruker Advance II AMX spectrometer (Bruker BioSpin, Rheinstetten, Germany) equipped with a 660 mT/m imaging gradient system. The seed was placed inside an in-house built birdcage coil (inner diameter, 17 mm), and the lipid-selective custom spin echo sequence was adjusted with the following parameters: repetition time, 850 ms; echo time, 7.8 ms; number of averages, 2; matrix size, 140 × 100 × 90; isotropic resolution, 100 μm; RF (radio-frequency) pulse bandwidth, 2000 Hz; and pulse shape, hermite. The data were further analyzed in MATLAB (MathWorks, Natick, MA, USA) with inhouse written algorithms and were exported to AMIRA (Mercury Computer Systems, Chelmsford, USA) for visualization.

**Lipid extractions and thin-layer chromatography**

Tissues from the cotyledons and embryonic axes of jojoba seeds were dissected by hand using a scalpel. Dissected tissues were homogenized in hot (70°C) isopropanol (IPA) with glass beads using a Beadbeater (BioSpec Mini–Beadbeater-16). Total lipids were extracted using hot (70°C) IPA with 0.01% butylated hydroxytoluene and incubated at 70°C for 30 min to inactivate endogenous phospholipases (101). After cooling to room temperature, chloroform (CHCl₃) and water (H₂O) were added to the extracts at a ratio of 2:1:0.45 (IPA:CHCl₃:H₂O; v/v/v). Lipids were extracted overnight at 4°C. Total lipids were partitioned with additional chloroform and further purified by washing with 1.0 M KCl three times. To examine the neutral lipid content of cotyledon and axis tissues, total lipids were separated using thin-layer chromatography (TLC). Aliquots containing 80 μg of total lipids were spotted on TLC Silica gel 60 plates (Merck) and resolved using 80:20:1 (hexane:diethyl ether:acetic acid; v/v/v). Neutral lipids from extracts were compared to a TLC Neutral standard mix 18-5C and the WE standard, heptadecanoyl stearate (Nu-Chek Prep), for identification. Three TLC plates (per tissue) containing nine spots of 80 μg of total lipids were prepared for the embryonic axis and cotyledon tissues. One lane was cut from the TLC plate and stained with iodine vapors to determine the Rf (retention factor) values of the spots. Spots corresponding to WEs and TAGs were scraped, re-extracted (2× with hexane), and dried under nitrogen. WE extracts were suspended in methanol:chloroform (2:1, v/v) containing 5 mM ammonium acetate, whereas TAGs were suspended in tetrahydrofuran:methanol:water (4:4:1, v/v/v).

**NanoESI-MS/MS analysis of wax esters and UPLC-nanoESI-MS/MS analysis of TAGs**

WE extracts from Jojoba cotyledons and embryonic axis tissues \((n = 3\) per tissue) were analyzed on an Applied Biosystems 4000 triple-quadrupole mass spectrometer (ABSciex) equipped with a direct inject nanoESI chip ion source (TriVersa NanoMate, Advion BioSciences). Before analysis, 5 nmol of internal standard, heptadecanoyl heptadecanoate was added to extracts. Samples (10 μl) of each WE extract were analyzed using the following ionization parameters: positive ionization mode, voltage of 1.5 kV, back pressure of 0.5 psi, source temperature of 40°C, and curtain gas set to 10 (arbitrary units). Using multiple reaction monitoring (MRM), 785 WE molecular species were monitored. For the MRM precursor ion, the Q1 mass analyzer was set to the ammonium adduct of the WE, and the Q3 mass analyzer was set for the \([\text{RCO}_2\text{H}_3]^+\) or \([\text{RCO}^-]^+\) ion fragment. Each WE molecular species was measured using 150-ms dwell time with 6 cycles (121.6 s/cycle), and total analysis time per sample was 12.1 min. Data were collected using Analyst software (v 1.5.1), and signal intensities were analyzed using LipidView software (ABSciex). WE signals were selected using a 0.5–atomic mass unit window, a minimal signal-to-noise ratio of 1.0, and a maximum intensity of 0%. Data were normalized to type 13C isotope correction and were corrected by using a WE calibration response factor (10).
UPLC-nanoESI-MS/MS analysis was used to determine the TAG composition of jojoba cotyledon, embryonic axis, and whole seed tissues. TAGs from lipid extracts, prepared as described above, were separated by UPLC (ACQUITY UPLC, Waters) equipped with an ACQUITY UPLC HSS T3 column (100 mm by 1 mm, 1.8 μm; Waters). UPLC instrument settings were as follows: A total of 2 μl was injected using a needle overfill mode with a flow rate of 0.1 ml min⁻¹, temperature was set to 35°C, solvent B was tetrahydrofuran/methanol/20 mM ammonium acetate (6:3:1, v/v/v) with 0.1% (v/v) acetic acid, and solvent A was methanol/20 mM ammonium acetate (3:7 v/v) containing 0.1% (v/v) acetic acid. TAG species were separated with the following linear binary gradient: 90% solvent B held for 2 min, linear increase to 100% solvent B for 2 min, and 100% solvent B held for 4 min and re-equilibration to start conditions in 4 min (102). Separated TAGs were analyzed on an Applied Biosystems 6500 QTRAP triple-quadrupole mass spectrometer (ABSciex) equipped with a direct inject nanoESI chip ion source (TriVersa NanoMate, Advion BioSciences). Ionization parameters were as follows: positive ionization mode, voltage of 1.3 kV, source temperature of 40°C, and curtain gas set to 20 (arbitrary units). MRM transitions were measured with a dwell time of 5 ms per transition and used to detect TAG molecular species and acyl composition by the fatty acid–associated neutral loss from [M + NH₄]⁺ molecular ions. Analyst and LipidView software packages were used to analyze the TAG composition, and identifications were made using a custom TAG fragment database specific for the acyl composition of jojoba.

**Staining and confocal imaging of LDs in seed sections**

Hand sections of developing (65 to 80 days after flowering) jojoba cotyledons and embryonic axis seed tissues were prepared and then fixed in a 4% paraformaldehyde 50 mM Pipes NaOH (pH 7.2) solution overnight. Sections were washed two times (15 min per wash) with 50 mM Pipes NaOH (pH 7.2) buffer and then stained with BODIPY 493/503 [2 μg/ml in 50 mM Pipes NaOH (pH 7.2)] for 45 min under house vacuum. The stained sections were washed two times (15 min per wash under house vacuum) with 50 mM Pipes NaOH (pH 7.2 buffer) and one time in ultrapure water (18.2 megohms) for 10 min under house vacuum. Stained sections were placed on slides with a coverslip and sealed with clear nail polish. Sections were immediately imaged by confocal microscopy. A Zeiss LSM710 confocal laser scanning microscope was used to image BODIPY-stained LDs with the following settings: 63× objective, image resolution of 1024 × 1024, pinhole size of 62.2 Airy Units (AU), and master gain set between 650 and 750. The fluorophore BODIPY 493/503 was excited by a 488-nm laser, and emission signal was collected between 500 and 540 nm. Images were processed using Zeiss Zen imaging software (v8.1).

For the jojoba seed sections, the LD diameter of LDs of the cotyledon and embryonic axis tissues were measured using the ImageJ software package (103). LDs were measured only if the entire circumference of the LD could be distinguished. The diameters of 600 and 400 LDs were measured for three separate seed sections of cotyledon and embryonic axis tissues, respectively.

**LD flotation and protein isolation**

LDs were isolated from ~80-DAP developing jojoba seed cotyledon and embryonic axis tissues by flotation centrifugation using a protocol modified by Chapman and Trelease (104). Two solutions were prepared, solution A, which was 100 mM potassium phosphate (pH 7.5), 400 mM sucrose, 1 mM EDTA, and 10 mM KCl, and solution B, which was 100 mM potassium phosphate (pH 7.5), 600 mM sucrose, 1 mM EDTA, and 10 mM KCl, and stored on wet ice until/ during use. Cotyledon and embryonic axis tissues were each finely chopped in solution B (2:1 solution:tissue, v/w) in a glass petri dish on ice using a razor blade. Resulting homogenates were filtered through Miracloth dampened with solution B into a high-speed centrifuge tube. Tubes were centrifuged at 500g at 4°C for 5 min to pellet large tissue debris. The supernatant, or crude homogenate, was transferred to a new tube, and an aliquot was reserved for protein precipitation. Solution A was carefully layered on top of the crude homogenate and centrifuged in a Sorvall HB-6 swinging bucket rotor at 10,500g at 4°C for 60 min. After centrifugation, the fat pad was removed and placed into a tube with 2.0 ml of buffer B and placed on wet ice. The supernatant, containing the cytosol and microsomes, was placed in a separate tube on wet ice, and the pellet, enriched in plastids and mitochondria, was suspended in 1 ml of ice-cold tris-HCl [100 mM (pH 7.5)] and reserved on wet ice. Solution A was carefully layered on top of the fat pad resuspended in solution B and centrifuged again at 10,500g at 4°C for 60 min. After centrifugation, the fat pad at the top of the tube was carefully removed and suspended in 2.0 ml of solution B, and then, 4.0 ml of buffer A was carefully layered on top. This process was repeated once more to achieve purified LDs. An aliquot (2 μl) of purified LDs in solution B was examined by confocal microscopy, whereas the remaining LD fraction was suspended in 1 ml of ice cold tris-HCl [100 mM (pH 7.5)].

The 10,500g supernatant fraction was centrifuged at 100,000g at 4°C for 60 min (fixed angle rotor, Sorvall Discovery 90) to obtain a microsomal pellet and cytosolic fraction. The supernatant, containing the cytosol fraction, was removed and kept on wet ice, and the pellet, containing the microsomes, was suspended in 1 ml of ice-cold tris-HCl [100 mM (pH 7.5)]. The isolated fractions of the crude homogenate, mitochondria/plastids, purified LDs, cytosol, and microsomes were each combined with four volumes of ~20°C acetone and placed at ~20°C overnight for protein precipitation. Proteins were pelleted and resuspended in 13,000g at 4°C for 15 min, after which the acetone was removed and the pellet was briefly dried. The protein pellets were suspended in 4× Laemmli buffer (Bio-Rad) and heated at 70°C for 15 min. Proteins were loaded on a 4 to 12% bis-tris gel and run at 160 V for ~15 min. The gel was fixed in a solution of 50:40:10 (water:ethanol:acetic acid; v/v/v), stained with QC Colloidal Coomassie Blue Stain (Bio-Rad) for 1 hour, and then destained in ultrapure water (18.2 megohms). For visualizing the proteins in each fraction, the samples were electrophoresed until the dye front was just at the bottom of the gel. For samples to be analyzed for proteomics, the proteins were concentrated by electrophoresis in the stacking gel and continued until the protein samples had just entered the separating gel. Bands in lanes corresponding to each fraction were excised from the gel and stored in a 5% acetic acid solution until prepared for proteomic analysis.

The aliquots of purified LDs were stained with BODIPY 493/503 (0.4 μg/ml) and imaged using a Zeiss confocal scanning microscope. A Zeiss LSM710 confocal laser scanning microscope was used to image BODIPY-stained LDs with the following settings: 63× objective, image resolution of 1024 × 1024, pinhole size of 62.2 AU, and master gain set between 650 and 750. The fluorophore BODIPY 493/503 was excited by a 488-nm laser, and emission signal was collected between 500 and 540 nm. Images were processed using Zeiss Zen imaging software (v8.1).
Proteomics analysis
Proteomics analysis was performed at the Michigan State University Proteomics Facility. Protein gel bands were dehydrated using 100% acetonitrile and incubated with 10 mM dithiothreitol in 100 mM ammonium bicarbonate (pH 8) at 56°C for 45 min, dehydrated again, and incubated in the dark with 50 mM iodoacetamide in 100 mM ammonium bicarbonate for 20 min. Dehydrated gel bands were then washed with ammonium bicarbonate and dehydrated again. Sequencing grade modified trypsin was prepared to 0.01 μg/μl in 50 mM ammonium bicarbonate, and ~50 μl of this was added to each gel band so that the gel was completely submerged. Bands were then incubated at 37°C overnight.

Peptides were extracted from the gel into a solution of 60% acetonitrile/1% trichloroacetic acid by water bath sonication and vacuum-dried to ~2 μl. Extracted peptides were then resuspended in 2% acetonitrile/0.1% trifluoroacetic acid to 20 μl. From this, 5 μl was automatically injected by a Thermo Fisher (www.thermofisher.com) EASYnLC 1000 onto a Thermo Acclaim PepMap RSLC 0.075 mm × 500 mm C18 column over washed for ~5 min. Bound peptides were eluted onto a Thermo Fisher (www.thermofisher.com) EASYnLC 1000 onto trifluoroacetic acid to 20 μl. The resulting MS/MS spectra were converted to peak lists using Mascot Distiller, v2.7 (www.matrixscience.com) and searched against a database of jojoba protein sequences acquired from the de novo transcriptome assembly appended with common laboratory contaminants [downloaded from www.thegpm.org, cRAP (common Repository of Adventitious Proteins) project] using the Mascot searching algorithm, v 2.6.0. The Mascot output was then analyzed using Scaffold, v4.8.7 (www.proteomesciences.com) to probabilistically validate protein identifications. Assignments validated using the Scaffold 1% false discovery rate confidence filter were considered true.

Protein counts taken from Scaffold were used to make relative quantifications of the proteins enriched in each protein fraction collected from the cotyledon and embryonic axis seed tissues. The LD protein fraction often contains protein contaminants from other cellular fractions. Thus, protein levels of the other fractions are needed as controls to eliminate contaminant proteins. Four criteria were used to develop a candidate list of proteins that were highly enriched in the LD protein fraction. Proteins were counted as an LD protein candidate if (i) peptide counts were the highest in the LD fraction, (ii) peptide counts for the protein in the LD fraction were two times greater than mitochondria/plastid fraction, (iii) peptide counts for the candidate LD protein in the cytosolic fraction must be less than 10, and (iv) the peptide counts in the LD fraction must be greater than 5.

**REFERENCES AND NOTES**

lipid droplet-associated proteins (LDAPs) are involved in plant lipid metabolism and have been shown to modulate triacylglycerol accumulation and storage wax production. These proteins are essential for the regulation of neutral lipid compartmentation in plant cells.

Large-scale and comprehensive microarray, next-generation sequencing (NGS), and proteomic analyses have been used to identify thousands of candidates for lipid metabolism-related genes. However, none of these candidates have been experimentally validated. A recent study identified over 1000 candidate genes from jojoba transcriptomes with the aid of a new computational pipeline.

This study utilized the CAP3 de novo assembly algorithm to assemble transcriptomes from different tissues and conditions, followed by the identification of lipid metabolism-related genes using functional annotations and expression profiles. A total of 1041 candidate genes were identified, representing a wide range of functional categories related to lipid metabolism.

The most highly expressed genes in the lipid droplet fraction were those encoding proteins involved in lipid transport and storage. These genes were also highly expressed in the neutral lipid fraction, suggesting that they play a role in lipid storage dynamics.

Furthermore, the study identified several candidate genes that are specific to the lipid droplet fraction, suggesting that they play a role in the regulation of lipid droplet size and distribution.

In conclusion, this study provides a comprehensive list of candidate genes involved in lipid metabolism in jojoba, which can be used as a starting point for further experimental validation and functional studies.


S. Anders, W. Huber, Differential Expression of RNA-Seq Data at the Gene Level–The DESeq Package (European Molecular Biology Laboratory (EMBL), 2012).


Acknowledgments: Genome sequencing and assembly were performed by Novogene Corporation. Sequencing and preliminary analysis of transcriptome data were performed by the UNITE BioDiscovery Institute Genomics Core Facility, and the advice of T. Kim is gratefully acknowledged. The draft genome for this study is available as accession no. PARR 940 from the National Agricultural Genomics Information Network (NAGIN). NAGIN was supported by the USDA National Plant Genetic Resources Program (NPGRP), the USDA-Agricultural Research Service, the USDA, and the National Institutes of Health. J. M. Love, W. Huber, S. Anders, and K. K. Seo assisted in computational methods and data analysis. E. G. W. Chapman, J. W. Hahn, and K. K. Seo assisted in bioinformatics and genetics analysis. C. A. Schneider, W. S. Rasband, and W. K. Eliceiri assisted in the design and implementation of the BioImage analysis pipeline. The following agencies and institutions provided support for this study: the National Science Foundation (grant EPS-0539931), the National Institutes of Health (grant GM071063), the National Institutes of Health (grant GM071063), and the National Institutes of Health (grant GM071063).
L.B. and E.M. performed the whole seed nuclear magnetic resonance analyses. D.S., S.L., Z.-W.Z., Y.S., K.D.C., L.-L.C., and L.G. drafted the manuscript with input from other authors. All authors read and approved the manuscript. Competing interests: The authors declare that they have no competing interests. Data and materials availability: All data needed to evaluate the conclusions in the paper are present in the paper and/or the Supplementary Materials. Additional data related to this paper may be requested from the authors. The jojoba genome assembly and meta-data were deposited in the genome warehouse in Beijing Institute of Genomics (BIG) data center, BIG under accession no. GWHAASQ00000000 that is publicly accessible at http://bigd.big.ac.cn/gwh. RNA-seq transcriptomic data were deposited to NCBI GEO repository and can be accessed using accession no. GSE130603.

Submitted 8 June 2019
Accepted 16 December 2019
Published 11 March 2020
10.1126/sciadv.aay3240

The genome of jojoba (Simmondsia chinensis): A taxonomically isolated species that directs wax ester accumulation in its seeds


Sci Adv 6 (11), eaay3240.
DOI: 10.1126/sciadv.aay3240

ARTICLE TOOLS http://advances.sciencemag.org/content/6/11/eaay3240

SUPPLEMENTARY MATERIALS http://advances.sciencemag.org/content/suppl/2020/03/09/6.11.eaay3240.DC1

REFERENCES This article cites 99 articles, 20 of which you can access for free http://advances.sciencemag.org/content/6/11/eaay3240#BIBL

PERMISSIONS http://www.sciencemag.org/help/reprints-and-permissions

Use of this article is subject to the Terms of Service