Journal of Heredity, 2026, **117**, 600–610
https://doi.org/10.1093/jhered/esaf105
Advance access publication 15 December 2025
Genome Resources

American Genetic Association
OXFORD

Genome Resources

"A genome assembly of the California flannelbush, *Fremontodendron californicum*"

William T. McMahan[1], Merly Escalona[2], Reed Kenny[1], Mohan P.A. Marimuthu[3], Oanh Nguyen[3], Colin W. Fairbairn[4], William Seligmann[4], Courtney Miller[5], Howard Bradley Shaffer[5], Shannon Still[1], Daniel Potter[1,*]

[1]Department of Plant Sciences, University of California, Davis, CA, United States, [2]Department of Biomolecular Engineering, University of California, Santa Cruz, CA, United States, [3]DNA Technologies and Expression Analysis Core Laboratory, Genome Center, University of California, Davis, CA, United States, [4]Department of Ecology and Evolutionary Biology, University of California, Santa Cruz, CA, United States, [5]Department of Ecology and Evolutionary Biology, University of California, Los Angeles, CA, United States

*Corresponding author: Department of Plant Sciences, 1 Shields Avenue, University of California, Davis, CA 95616, United States. Email: dpotter@ucdavis.edu

Abstract

Fremontodendron (Malvaceae) is a genus of shrubs native to the California Floristic Province (CFP) with dense, stellate trichomes on their leaves (hence the common name "flannelbush"). The current treatment of the genus includes the widespread and morphologically variable species *Fremontodendron californicum,* along with the rare *F. decumbens* and *F. mexicanum*. While *F. californicum* is spread across several ecoregions of the CFP, *F. decumbens* and *F. mexicanum* are highly restricted. Here, we introduce the first genome-scale resource with which to study this important genus, a de novo, scaffold scale assembly of an individual of *F. californicum*. Following the overall strategy of the California Conservation Genomics Project (CCGP), we used Pacific Biosciences HiFi long reads and Omni-C chromatin mapping to produce an assembly of the nuclear and plastid (chloroplast) genomes. The nuclear assembly consists of two, phased haplotypes, haplotypes one and two, with similar sizes of around 1.2 Gb, similar scaffold N50s around 26 Mb, and each with a Benchmarking Universal Single-Copy Ortholog (BUSCO) completeness score of 99.4%. This assembly will be a valuable resource for understanding the distribution, genetic variation, and species delimitation of this California genus of conservation value, as well as a tool to further investigate the complex evolutionary history of Malvaceae s.l.

Key words: California Conservation Genomics Project, CCGPCalifornia floristic province, conservation, Malvaceae

Introduction

We present the nuclear and chloroplast genome assemblies of *Fremontodendron californicum* (Torr.) Coville (Malvaceae), "California flannelbush." These assemblies were produced as part of the California Conservation Genomics Project (CCGP), an initiative to document the landscape genomics of the animals and plants which span the range of ecological and phylogenetic diversity of the state of California (Shaffer et al. 2022; Toffelmier et al. 2022). In addition to compiling a genome assembly for each taxon, the project will resequence up to 150 different individuals of each taxon across their geographic range to generate a profile of their genetic diversity which can be used to inform conservation and management decisions. The taxa, some of which are key species and some of which are endangered or threatened, are found in marine, freshwater, and terrestrial habitats. Compiling their genomic information will reveal trends not only at the intraspecific level but also at the community and ecological level, allowing insights into biodiversity hotspots, potential blockages to gene flow, and challenges regarding their preservation and adaptation to climate change.

In general, species of *Fremontodendron* are shrubs which bear lobed leaves with tawny stellate pubescence on the abaxial surfaces (see Table 1). Nine taxa have been recognized, at the specific or infraspecific level, within *Fremontodendron* (Table 1). In the current treatments for the Jepson e-flora (Preston et al. 2012) and the Flora of North America (Hanes 2015), three species are recognized, and no formal taxonomic recognition is given to morphological variants within the most widespread species, *Fremontodendon californicum* (Torr.) Coville. The other two species both have highly restricted distributions and are listed at the state (rare) and federal (endangered) levels (CNDDB 2020), with California Rare Plant Ranks of 1B ("rare, threatened, or endangered in California and elsewhere"; CNPS 2020). *F. mexicanum* Davidson, "Mexican flannelbush" (rank 1B.1), is found only in southern California and Baja California, Mexico, with only one confirmed natural occurrence in each area, the former in San Diego County (Kelman 1991). *F. decumbens* R.M. Lloyd, "Pine Hill flannelbush" (rank 1B.2), is known only from a few populations in the Pine Hill area (El Dorado County) of the Sierra Nevada foothills (Kelman et al. 2006).

Received on 18 April 2025; revised on 22 October 2025; accepted on 8 December 2025

Table 1. Taxa which have been recognized within the genus *Fremontodendron*. Those marked with * were originally described under the generic name *Fremontia*, but this name is considered illegitimate due to having been previously applied (Torrey 1843) to a species later determined to be a synonym of *Batis vermiculates* Hook. (Chenopodiaceae; Torrey 1854), now treated as *Sacrobatus vermiculatus* (Hook.) Torr. (Sarcobataceae).

Taxon	Also treated as	Distinguishing features	Distribution	Currently treated as
F. californicum (Torr.) Coville, 1893*	*F. californicum* var. *californicum*, *F. californicum* subsp. *californicum*	Sepals yellow; sepal pits with hairs; seeds hairy, with appendage	California, Oregon, Arizona, Baja California	*F. californicum*
F. californicum subsp. *crassifolium* (Eastw.) J.H. Thomas, 1955	*Fremontia crassifolia* Eastw., 1934*	Thicker leaves and larger flowers than typical *F. californicum*	Santa Cruz Mountains, California	*F. californicum*
F. californicum subsp. *napense* (Eastw.) Munz, 1964	*Fremontia napensis* Eastw., 1934*	Smaller, thinner leaves and smaller flowers than typical *F. californicum*	Napa and Counties, California	*F. californicum*
F. californicum subsp. *obispoense* (Eastw.) Munz, 1964	*Fremontia obispoensis* Eastw., 1934*, *F. californicum* var. *obispoense* (Eastw.) Hoover, 1970	Small, generally unlobed leaves	San Luis Obispo County, California	*F. californicum*
Fremontia californica var. diegensis M. Harv., 1943*		Leaves unlobed, petioles > blade length	San Diego County, California.	*F. californicum*
Fremontia californica var. integra M. Harv., 1943*		Leaves unlobed, petioles < or = ½ blade length	Tulare and Kern Counties, California.	*F. californicum*
Fremontia californica var. viridis M. Harv., 1943*		Leaf hairs whitish, not tawny.	Tehama County, California	*F. californicum*
F. decumbens R.M. Lloyd, 1964	*F. californicum* subsp. *decumbens* (R.M. Lloyd) A.E. Murray, 1982	Decumbent habit; sepals orange; sepal pits with hairs; seeds hairy, with appendage	Northern Sierra Nevada foothills	*F. decumbens*
F. mexicanum Davidson, 1917	*F. californicum* subsp. *mexicanum* (Davidson) Munz, 1968	Erect habit; sepals orange; sepal pits lacking hairs; seed glabrous, lacking appendage	Southern California, Baja California	*F. mexicanum*

In contrast, *F. californicum* is found throughout the California floristic province, occurring naturally in 12 of the 19 United States Department of Agriculture (USDA) ecoregions within California and extending into Oregon, Arizona, and northern Baja California, Mexico. It grows in a range of habitats including coastal scrub, oak/pine woodlands, chaparral, and arid high desert. It occurs at elevations ranging from 180 to 2320 m, spanning the Coast Ranges to the Western Sierra Nevada foothills (Preston et al. 2012). It is valuable to animals and insects as a source of both nectar and seed and is considered to be a fire recruiter and fire-dependent given its ability to survive and regenerate as well as sprout from seed in post-fire environments. Its drought resistance and showy yellow calyces have made it a favorite of California native plant horticulture (Abrahamson 2021). Individuals of *F. californicum* are generally erect shrubs, 2 to 5 m in height, and are highly morphologically variable, both between and within populations (Preston et al. 2012).

Despite the well-studied variability in *F. californicum* and the conservation status of *F. mexicanum* and *F. decumbens*, there have been few studies of their genetic diversity, and few genomic resources are available to help evaluate the status of ambiguous individuals or to guide the management of populations of concern. The chromosome number for *F. californicum* is reported at $n = 20$ (Lenz 1950), and the only study of genetic diversity in the genus relied on amplified fragment length polymorphisms (Kelman et al. 2006). Here, we present both the first scaffold-scale nuclear genome assembly and a chloroplast genome assembly for *F. californicum*. In collaboration with and under the guidance of the CCGP, these assemblies are based on Pacific Biosciences (PacBio) HiFi long read sequencing, and the nuclear assembly was scaffolded de novo using Omni-C data. These genomic resources for *F. californicum* will form the basis of studies to determine species evolution, delimitation, and diversity within this important California native plant.

Methods

Biological materials

We selected a naturally occurring individual of *F. californicum* as the basis for the genome assemblies. The individual is located in Napa County, California, along the Knoxville-Berryessa Road, approximately 0.5 miles south of the border with Lake County (38.85135°N, 122.39593°W, datum WGS84). Leaves for extraction and enough plant material suitable for a herbarium voucher were collected on 31 October 2020. The corresponding herbarium voucher is deposited at the UC Davis Center for Plant Diversity (Acc. # DAV 238751, Fig. 1A).

Nucleic acid library preparation and sequencing

Pacific Biosciences HiFi Library

We extracted high-molecular-weight (HMW) genomic DNA from 500 mg of young leaves (Sample ID: 1011) using the cetyltrimethylammonium bromide (CTAB) method as described in Inglis et al. (2018), with the following modifications: (i) we used sodium metabisulfite (1% w/v) instead of 2-mercaptoethanol (1% v/v) in the sorbitol wash buffer and CTAB solutions; (ii) we repeated the tissue homogenate wash steps until the supernatant turned clear; (iii) we performed

Fig. 1. Images of the California flannelbush (*Fremontodendron californicum*). A: Image: https://cch2.org/imglib/cch2/DAV/DAV238/DAV238751_lg.jpg. (A) The herbarium specimen of the individual used to make the assemblies presented here, collected in Napa County, CA. B: Image: SMS_90568_2021-04-19.jpeg. (B) An adult individual of *F. californicum*, exhibiting the characteristic upright, shrubby growth habit. C: Image: IMG_6965.jpeg. (C) Leaves of *F. californicum* with stellate pubescence visible. Photo credit for (B, C): Shannon Still.

the CTAB lysis step at 45 °C and (iv) the chloroform extraction step twice using ice-cold chloroform. The DNA purity was estimated by absorbance ratios (260/280 = 1.80 and 260/230 = 2.16) measured using the NanoDrop ND-1000 spectrophotometer (Thermo Fisher Scientific, MA). The DNA yield (13.2 μg) was quantified using a Quantus Fluorometer (QuantiFluor ONE dsDNA Dye assay; Promega, WI), and the size distribution of the DNA was estimated using the Femto Pulse system (Genomic DNA 165 kb kit, Agilent, CA), where 71% of the DNA fragments were found to be 30 kb or longer.

The HiFi SMRTbell library was constructed using the SMRTbell Express Template Prep Kit v2.0 (PacBio Cat. #100-938-900) according to the manufacturer's instructions. HMW genomic DNA was sheared to a target DNA size distribution between 15 and 18 kb using Diagenode's Megaruptor 3 system (Diagenode, Belgium; cat. B06010003). The sheared gDNA was concentrated using 0.45× of AMPure PB beads (PacBio CA; Cat. #100-265-900) for the removal of single-strand overhangs at 37 °C for 15 min, followed by further enzymatic steps of DNA damage repair at 37 °C for 30 min, end repair and A-tailing at 20 °C for 10 min and 65 °C for 30 min, and ligation of overhang adapters v3 at 20 °C for 60 min. The SMRTbell library was purified and concentrated with 1× AMPure PB beads for nuclease treatment at 37 °C for 30 min and size selection using the PippinHT system (Sage Science, MA; Cat #HPE7510) to collect fragments greater than 7 to 9 kb. The 15 to 20 kb average HiFi SMRTbell library was sequenced at UC Davis DNA Technologies Core (Davis, CA) using one 8 M SMRT cell, Sequel II sequencing chemistry 2.0, and 30-h movies each on a PacBio Sequel II sequencer.

Omni-C library

The Omni-C library was prepared using the Dovetail Omni-C Kit (Dovetail Genomics, CA) according to the manufacturer's protocol with slight modifications. First, specimen tissue (leaves, ID: 1101C) was thoroughly ground with a mortar and pestle while cooled with liquid nitrogen. Nuclear isolation was then performed using published methods (Workman et al. 2018). Subsequently, chromatin was fixed in place in the nucleus and digested under various conditions of DNase I until a suitable fragment length distribution of DNA molecules was obtained. Chromatin ends were repaired and ligated to a biotinylated bridge adapter, followed by proximity ligation of adapter containing ends. After proximity ligation, crosslinks were reversed and the DNA was purified from

proteins. Purified DNA was treated to remove biotin which was not internal to ligated fragments. A Next Generation Sequencing (NGS) library was generated using an NEB Ultra II DNA Library Prep kit (New England Biolabs, MA) with an Illumina-compatible y-adaptor. Biotin-containing fragments were then captured using streptavidin beads. The postcapture product was split into two replicates prior to Polymerase Chain Reaction (PCR) enrichment to preserve library complexity with each replicate receiving unique dual indices. The library was sequenced at Vincent J. Coates Genomics Sequencing Lab (Berkeley, CA) on an Illumina NovaSeq platform (Illumina, CA) to generate approximately 100 million 2 × 150 bp read pairs per gigabase of genome size.

Nuclear genome assembly

We assembled the genome of the *F. californicum* individual following the CCGP assembly pipeline Version 5.0, as outlined in Table 2, listing the tools and nondefault parameters used in the assembly process. We removed the remnant adapter sequences from the PacBio HiFi dataset using HiFiAdapterFilt (Sim et al. 2022) and generated an initial diploid phased assembly using HiFiasm (Cheng et al. 2021) in HiC mode with the filtered PacBio HiFi reads and the Omni-C short-reads, a process which generates two assemblies, one per haplotype. We then aligned the Omni-C data to both assemblies following the Arima Genomics Mapping Pipeline (https://github.com/ArimaGenomics/mapping_pipeline) and then scaffolded both assemblies with SALSA (Ghurye et al. 2017; Ghurye et al. 2019).

The assemblies for both haplotypes were manually curated by iteratively generating and analyzing their corresponding Omni-C contact maps. Briefly, to generate the contact maps we aligned the Omni-C data with BWA-MEM (Li 2013), identified ligation junctions, and generated Omni-C pairs (Lee et al. 2022) using pairtools (Open2C et al. 2023). Then, we generated multi-resolution Omni-C matrices with cooler (Abdennur and Mirny 2020) and balanced them with hicExplorer (Ramírez et al. 2018). We used HiGlass (Kerpedjiev et al. 2018) and the PretextSuite (https://github.com/wtsi-hpag/PretextView; https://github.com/wtsi-hpag/PretextMap; https://github.com/wtsi-hpag/PretextSnapshot) to visualize the contact maps. We identified misassemblies and misjoins in these contact maps and modified the assemblies using the Rapid Curation pipeline from the Wellcome Trust Sanger Institute, Genome Reference Informatics Team (https://gitlab.com/wtsi-grit/rapid-curation). Some of the remaining gaps (joins generated during scaffolding and/or curation) were closed using the PacBio HiFi reads and YAGCloser (https://github.com/merlyescalona/yagcloser). We checked for contamination using the BlobToolKit Framework (Challis et al. 2020).

Genome quality assessment

We generated *k*-mer counts from the PacBio HiFi reads using meryl (https://github.com/marbl/meryl). The k-mer counts were then used in GenomeScope (Vurture et al. 2017) to estimate genome features including genome size, heterozygosity, and repeat content. To obtain general contiguity metrics, we ran QUAST (Gurevich et al. 2013). To evaluate genome quality and functional completeness, we used BUSCO (Manni et al. 2021) with the Embryophyta ortholog database (Embryophyta_odb10) which contains 1614 genes. Assessment of base level accuracy (QV) and k-mer completeness was performed using the previously generated meryl database and merqury (Rhie et al. 2020). We further estimated genome assembly accuracy via BUSCO gene set frameshift analysis using the pipeline described in Korlach et al. (2017). Measurements of the size of the phased blocks are based on the size of the contigs generated by HiFiasm in HiC mode. We follow the quality metric nomenclature established by Rhie et al. (2021), with the genome quality code x.y.P.Q.C, where, x = log10[contig NG50]; y = log10[scaffold NG50]; P = log10 [phased block NG50]; Q = Phred base accuracy QV (quality value); C = % genome represented by the first "n" scaffolds, following a karyotype of 2n = 92 for this species, estimated as a mode from ancestral species number of chromosomes (Genome on a Tree—GoaT; tax_name(*F. californicum*); Challis et al. 2023). Quality metrics for the notation were calculated using the haplotype one assembly. We further assessed the size and classification of repeat elements within the nuclear assembly by using RepeatModeler and RepeatMasker (Smit and Hubley 2015; Smit et al. 2015).

Chloroplast genome assembly

We used the Oatk pipeline (https://github.com/c-zhou/oatk; Zhou et al. 2024) to generate a chloroplast assembly from the PacBio HiFi reads, with coverage set to 50. We visualized the Oatk output in Bandage (Wick et al. 2015) to verify the completeness of the assembly. Bandage generated an expected chloroplast graph pattern with one small single copy, one large single copy, and two inverted repeat regions. We annotated the chloroplast assembly using GeSeq (Tillich et al. 2017).

Results

Sequencing data

The Omni-C library generated 146.99 million read pairs, and the PacBio HiFi library generated 2.45 million reads. The PacBio HiFi sequences yielded ∼57× genome coverage and had an N50 read length of 13,916 bp; a minimum read length of 65 bp; a mean read length of 13,063 bp; and a maximum read length of 54,523 bp. Based on the PacBio HiFi data, Genomescope estimated a genome size of 883.75 Mb, a 0.132% sequencing error rate, and 1.01% heterozygosity. The *k*-mer spectrum shows a bimodal distribution with a major peak at ∼25-fold coverage and a minor peak ∼13-fold coverage (Fig. 2A).

Nuclear genome assembly

The final genome assembly (ddFreCali1) consists of two phased haplotypes. Both assemblies are similar in size, and they are also similar in size but not equal to the estimated genome size from GenomeScope, as has been observed in other taxa (see Pflug et al. 2020, for example).

The haplotype one assembly (ddFreCali1.0.hap1) consists of 389 scaffolds spanning 1.23 Gb with a contig N50 of 11.83 Mb, a scaffold N50 of 23.3 Mb, the largest contig size of 42.29 Mb, and the largest scaffold size of 44.02 Mb. The haplotype two assembly (ddFreCali1.0.hap2) consists of 202 scaffolds spanning 1.22 Gb with a contig N50 of 13.98 Mb, a scaffold N50 of 26.7 Mb, the largest contig size of 28.83 Mb, and the largest scaffold size of 44.01 Mb.

The haplotype one assembly has a BUSCO completeness score for the Embryophyta gene set of 99.4%, a base pair quality value (QV) of 66.07, a kmer completeness of 83.96%, and

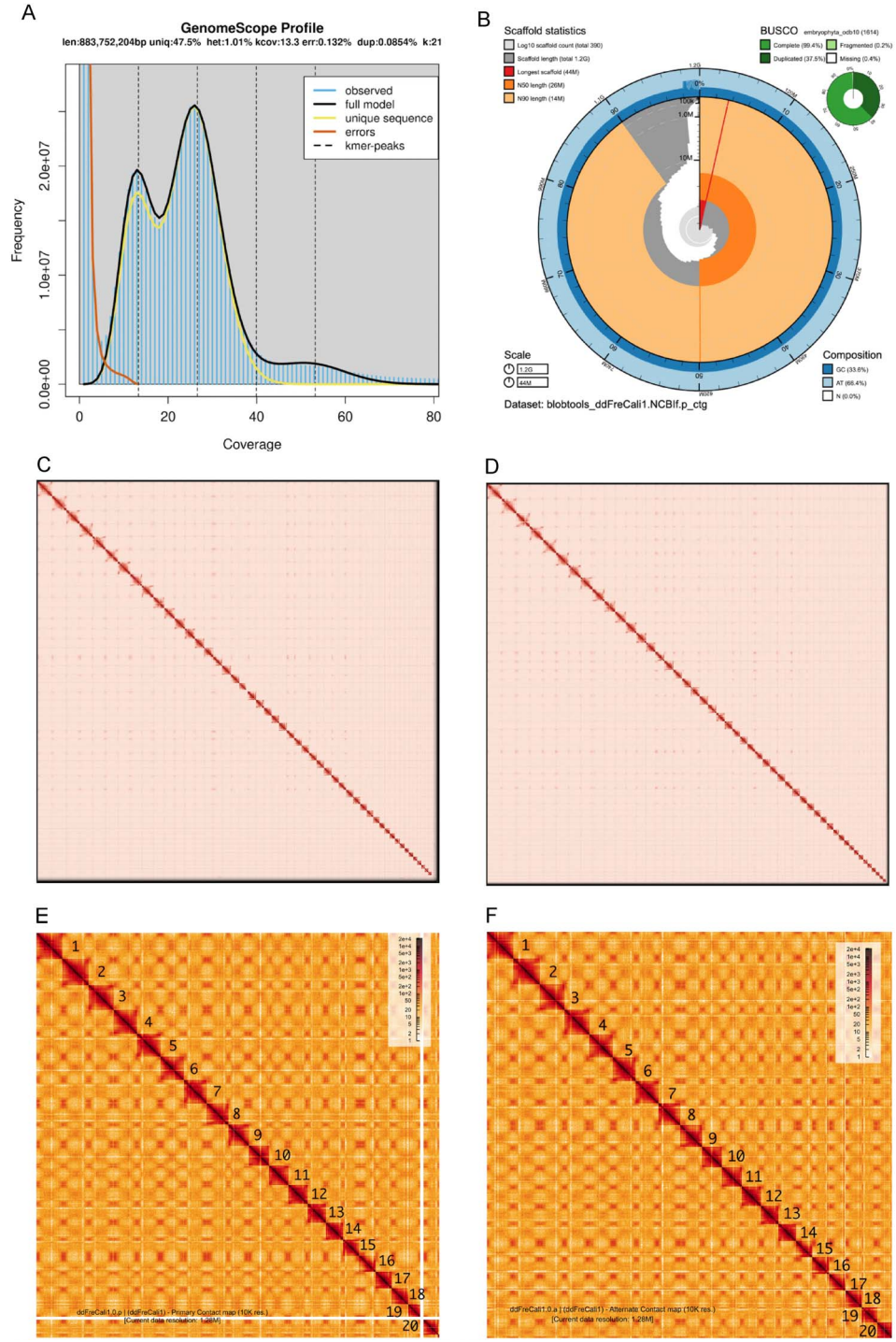

Fig. 2. A: Image: GenomeScopeProfile. (A) K-mer spectrum output of the PacBio HiFi data, trimmed of adapters, using GenomeScope2.0. The bimodal distribution indicates a diploid genome. *K*-mers with lower coverage and low frequency represent differences between haplotypes, whereas higher coverage and higher frequency k-mers represent similarities between the two haplotypes. B: Image: Haplotype 1 BlobToolKitSnailplot. (B) A graphical representation of the quality metrics of haplotype 1 (rendered numerically in Table 3) represented as a BlobToolKit Snail plot. The plot circle represents the full size of the assembly. From the inside out, the central plot covers length-related metrics. The longest line represents the size of the longest scaffold; all other scaffolds are arranged in size order, moving clockwise around the plot, and start from the outside of the central plot. The arcs show the scaffold N50 and scaffold N90 values. The central light-gray spiral shows the cumulative scaffold count with a white line at each order of magnitude. White regions in this area reflect the proportion of Ns in the assembly. The dark vs. light-blue area around it shows mean, maximum, and minimum GC vs. AT content at 0.1% intervals (Challis et al. 2020). C: Image: Haplotype1contactmap. D: Image: Haplotype2contactmap. E: Image: Haplotype1contactmapofthe20largestscaffolds. F: Image: Haplotype2contactmapofthe20largestscaffolds. Caption: (C–F) Omni-C contact maps for the haplotype 1 (C) and haplotype 2 (D) nuclear assembly generated with PretextSnapshot. The same contact maps but for the 20 largest scaffolds are presented in (E) and (F), respectively. Contact maps translate proximity of genomic regions in 3D space to contiguous linear organization. Each cell in the contact map corresponds to sequencing data supporting the linkage (or join) between two regions of the genome assembly. Scaffolds are separated by black lines and higher density corresponds to higher levels of fragmentation.

Table 2. Assembly pipeline and software used.

Assembly step	Software and any nondefault options	Version	Reference
Initial assembly			
Filtering PacBio HiFi adapters	HiFiAdapterFilt	Commit 64d1c7b	Sim et al. 2022
K-mer counting	Meryl (k = 21)	1	https://github.com/marbl/meryl
Estimation of genome size and heterozygosity	GenomeScope	1	Vurture et al. 2017
De novo *assembly (contiging)*	HiFiasm (Hi-C Mode, −primary, output hic.hap1.p_ctg, hic.hap2.p_ctg)	0.16.1-r375	Cheng et al. 2021
Scaffolding			
Omni-C data alignment	Arima Genomics Mapping Pipeline	Commit 2e74ea4	https://github.com/ArimaGenomics/mapping_pipeline
Arima Genomics Mapping Pipeline (AGMP)	BWA-MEM	0.7.17-r1188	Li 2013
	samtools	1.11	Danecek et al. 2021
	filter_five_end.pl (AGMP)	Commit 2e74ea4	https://github.com/ArimaGenomics/mapping_pipeline
	two_read_bam_combiner.pl (AGMP)	Commit 2e74ea4	https://github.com/ArimaGenomics/mapping_pipeline
	picard	2.27.5	https://broadinstitute.github.io/picard/
Omni-C Scaffolding	SALSA (-DNASE, -i 20, -p yes)	2	Ghurye et al. 2017, Ghurye et al. 2019
Omni-C Contact map generation			
Short-read alignment	BWA-MEM (-5SP)	0.7.17-r1188	Li 2013
SAM/BAM processing	samtools	1.11	Danecek et al. 2021
SAM/BAM filtering	pairtools	0.3.0	Open2C et al. 2024
Pairs indexing	pairix	0.3.7	Lee et al. 2022
Matrix generation	cooler	0.8.10	Abdennur and Mirny 2020
Matrix balancing	hicExplorer (hicCorrectmatrix correct –filterThreshold -2 4)	3.6	Ramírez et al. 2018
Contact map visualization	HiGlass	2.1.11	Kerpedjiev et al. 2018
	PretextMap	0.1.4	https://github.com/wtsi-hpag/PretextView
	PretextView	0.1.5	https://github.com/wtsi-hpag/PretextMap
	PretextSnapshot	0.0.3	https://github.com/wtsi-hpag/PretextSnapshot
Manual curation tools	Rapid curation pipeline (Wellcome Trust Sanger Institute, Genome Reference Informatics Team)	Commit 7acf220c	https://gitlab.com/wtsi-grit/rapid-curation
Genome quality assessment			
Basic assembly metrics	QUAST (−-est-ref-size)	5.0.2	Gurevich et al. 2013
Assembly completeness	BUSCO (−m geno, –l embryophyta)	5.0.0	Manni et al. 2021
	Merqury	2020 January 29	Rhie et al. 2020
Repeat content analysis	RepeatModeler	2.0.5	Smit and Hubley 2015
	RepeatMasker	4.1.5	Smit et al. 2015
Contamination screening			
Local alignment tool	BLAST+ (-db nt, -outfmt '6 qseqid staxids bitscore std', -max_target_seqs 1, -max_hsps 1, -evalue 1e-25)	2.15	Camacho et al. 2009
General contamination screening	BlobToolKit (HiFi coverage, BUSCO = embryophyta, NCBI Taxa ID = 93 767)	2.3.3	Challis et al. 2020
Chloroplast genome assembly			
de novo genome assembler	Oatk (−c 50)	1.0	Zhou et al. 2024
Graph visualizer	Bandage	2022.8-dev	Wick et al. 2015
Annotation	GeSeq (https://chlorobox.mpimp-golm.mpg.de/geseq.html)	2021	Tillich et al. 2017

a frameshift indel QV of 45.37. The haplotype two assembly has a BUSCO completeness score for the Embryophyta gene set of 99.4%, a base pair QV of 66.17, a kmer completeness of 84.06%, and a frameshift indel QV of 46.31.

During manual curation, we made a total of 154 joins (77 per haplotype) and 45 breaks (24 on haplotype one and 21 on haplotype two) based on the Omni-C contact map signal. We closed a total of 40 gaps (22 on haplotype one and 18 on

Table 3. Assembly statistics.

Bio Projects & vouchers	**CCGP NCBI BioProject**		**PRJNA720569**			
	Genera NCBI BioProject		**PRJNA765826**			
	Species NCBI BioProject		**PRJNA765826**			
	NCBI BioSample		**SAMN35735723**			
	Specimen identification		**D. Potter 1011**			
	NCBI Genome accessions		**Haplotype 1**		**Haplotype 2**	
	Assembly accession		**JAULJK000000000**		**JAULJL000000000**	
	Genome sequences		**GCA_030761205.1**		**GCA_030761185.1**	
Genome Sequence	PacBio HiFi reads	Run	1 PACBIO_SMRT (Sequel IIe) run: 2.5 M spots, 32G bases, 18.4 Gbytes			
		Accession	SRX25150516			
	Omni-C Illumina reads	Run	2 ILLUMINA (Illumina NovaSeq 6000) runs: 147 M spots, 44.4G bases, 15.9 Gbytes			
		Accession	SRX25150517-8			
Genome Assembly Quality Metrics	Assembly identifier (quality code[a])		ddFreCali1(7.7.P7.Q66.C95)			
	HiFi read coverage[b]		57.39X			
			Haplotype 1		**Haplotype 2**	
	Number of contigs		505		320	
	Contig N50 (bp)		11 831 472		13 982 549	
	Contig NG50[b]		22 154 149		20 612 997	
	Longest contigs		42 296 852		28 838 739	
	Number of scaffolds		389		202	
	Scaffold N50		26 303 202		26 782 959	
	Scaffold NG50[b]		37 899 299		36 980 028	
	Largest scaffold		44 025 563		44 017 524	
	Size of final assembly		1 233 188 417		1 222 422 046	
	Phased block NG50[b]		21 039 688		20 612 997	
	Gaps per Gbp (# Gaps)		94(116)		97(118)	
	Indel QV (Frame shift)		45.37399284		46.31611688	
	Base pair QV		66.1746		66.1746	
		Full assembly = 66.1283				
	k-mer completeness		83.966		84.069	
			Full assembly = 98.8237			
	BUSCO completeness (embryophyta) $n = 1614$	**C**[c]	**S**[c]	**D**[c]	**F**[c]	**M**[c]
	H1[d]	99.40%	61.90%	37.50%	0.20%	0.40%
	H2[d]	99.40%	62.50%	36.90%	0.20%	0.40%
	Organelles	Complete chloroplast sequence			PV236216	

[a] Assembly quality code x.y.P.Q.C derived notation, from (Rhie et al. 2021). x = log10[contig NG50]; y = log10[scaffold NG50]; P = log10 [phased block NG50]; Q = Phred base accuracy QV (quality value); C = % genome represented by the first "n" scaffolds, following a karyotype of 2n = 92 for this species, estimated as a mode from ancestral species number of chromosomes (Genome on a Tree—GoaT; tax_name(*Fremontodendron californicum*); Challis et al. 2023). Quality code for all the assembly denoted by haplotype one assembly (ddFreCali1.0.hap1). [b] Read coverage and NGx statistics have been calculated based on the estimated genome size of 883.75 Mb. [c] BUSCO Scores. Complete BUSCOs (C). Complete and single-copy BUSCOs (S). Complete and duplicated BUSCOs (D). Fragmented BUSCOs (F). Missing BUSCOs (M). [d] (H1) Haplotype 1 and (H2) Haplotype 2 assembly values.

haplotype two). No other contigs were modified or removed. The Omni-C contact maps show highly contiguous assemblies, with chromosome-length scaffolds (Fig. 2B). Assembly statistics are reported in Table 3 and represented graphically in Fig. 2C. We have deposited the genome assembly on NCBI GenBank (see Table 3 and Data Availability for details).

De novo repeat libraries proposed by RepeatModeler were used to mask repeat elements in each assembly with RepeatMasker. Haplotype 1 and haplotype 2 are similar in terms of repeat element content with 71.87% and 72.06% masked, respectively (Table 4). Retroelements comprise most of the previously described repeat element content (about 32% for each haplotype) followed by DNA transposons (2% for each haplotype). Long terminal repeats and especially the Gypsy/DIRS-1 group of LTRs make up most of the retroelements (roughly 26% per haplotype), while the MULE-MuDR family makes up most of the DNA transposons (at 1.5%). Almost half of the masked repeat elements are designated as unclassified by RepeatMasker at around 33% per haplotype.

Chloroplast genome assembly

The chloroplast assembly size is 160,827 bp, with 36.75% GC content. The large single copy (LSC) is 89,841 bp long, the small single copy (SSC) is 20,317 bp long, and the inverted repeat (IR) is 25,333 bp long. This plastome contains a total of 290 genes including rRNAs and tRNAs (not counting the duplication of the IR).

Table 4. Results of RepeatMasker for both the haplotype 1 and haplotype 2 assemblies. Custom repeat libraries for RepeatMasker for each assembly were built de novo using RepeatModeler.

	Haplotype 1 (ddFreCali1.0.hap1)			Haplotype 2 (ddFreCali1.0.hap2)		
	Number of elements	Length occupied (bp)	% of sequence	Number of elements	Length occupied (bp)	% of sequence
Retroelements	**237,048**	**403,922,529**	**32.75**	**217,571**	**391,630,342**	**32.03**
SINEs:	91	21 994	0.00	676	107,715	0.01
Penelope:	340	313,297	0.03	0	0	0.00
LINEs:	15,953	14,778,940	1.20	14,006	14,599,861	1.19
CRE/SLACS	0	0	0.00	0	0	0.00
L2/CR1/Rex	0	0	0.00	0	0	0.00
R1/LOA/Jockey	420	57,348	0.00	0	0	0.00
R2/R4/NeSL	0	0	0.00	0	0	0.00
RTE/Bov-B	433	238,308	0.02	954	348,927	0.03
L1/CIN4	15,100	14,483,284	1.17	13,052	14,250,934	1.17
LTR elements:	221,004	389,121,595	31.55	202,889	376,922,766	30.83
BEL/Pao	13,129	14,428,258	1.17	4468	3,200,323	0.26
Ty1/Copia	39,916	50,787,200	4.12	42,885	50,177,934	4.10
Gypsy/DIRS1	153,121	314,257,337	25.48	151,285	321,818,303	26.32
Retroviral	12,613	8,390,902	0.68	239	191,896	0.02
DNA transposons	**48,246**	**30,040,759**	**2.44**	**39,739**	**24,818,635**	**2.03**
hobo-Activator	8385	2,254,647	0.18	3587	1,271,713	0.10
Tc1-IS630-Pogo	548	192,830	0.02	0	0	0.00
En-Spm	0	0	0.00	0	0	0.00
MULE-MuDR	16,873	18,016,929	1.46	21,361	19,085,571	1.56
PiggyBac	0	0	0.00	0	0	0.00
Tourist/Harbinger	705	349,504	0.03	647	389,942	0.03
Other (Mirage, P-element, Transib)	0	0	0.00	0	0	0.00
Rolling-circles	**6004**	**2,601,743**	**0.21**	**9851**	**3,898,411**	**0.32**
Unclassified:	**1,056,958**	**417,975,707**	**33.89**	**1,093,401**	**436,438,733**	**35.70**
Total interspersed repeats:		**852,252,292**	**69.11**		**852,887,710**	**69.76**
Small RNA:	23 915	13,830,187	1.12	1584	8,880,462	0.73
Satellites:	322	156,719	0.01	0	0	0.00
Simple repeats:	300,908	14,429,618	1.17	287,131	12,493,773	1.02
Low complexity:	56,915	3,016,923	0.24	54,726	2,797,318	0.23

Discussion

The nuclear genome assembly of *F. californicum* is the first of its kind for this genus and will be an invaluable resource to evaluate the intraspecific and interspecific diversity of this important constituent of the California flora. It will also expand our genomic knowledge of the family Malvaceae, which contains several taxa of tremendous economic importance and has also been the subject of much evolutionary research. The precise phylogenetic position of *Fremontodendron* within the family has not been established with certainty, but the most recent evidence indicates it is a basal member of the Malvatheca, a clade which encompasses the subfamilies Bombacoideae and Malvoideae (Nyffeler et al. 2005). Thus, this genome assembly should be of considerable interest beyond its significance for conservation efforts within California.

The number of chromosomes as indicated by the Omni-C contact map ($n = 46$, Fig. 2) does not match the karyotype as reported from a camera lucida study where $n = 20$ (Lenz 1950), which was based on an individual of *F. californicum* from Tehama County. It's unlikely that the individual of *F. californicum* used in our study is an auto or allopolyploid as the GenomeScope Profile's bimodal distribution indicates a diploid genome (Fig. 2). The taxonomic treatment of *F. decumbens* estimated the chromosome number of that species to be $n = 49$ (Lloyd 1965), which more closely aligns with our finding. It's possible that the camera lucida study vastly underestimated the number of chromosomes, given the nature of that technique; it's also possible that there is variation in chromosome number within *F. californicum*, as it's currently described, which warrants further investigation.

Other published assemblies for genera in Malvaceae (s.l.) include *Gossypium*, *Hibiscus*, and *Abelmoschus* in subfamily Malovideae; *Bombax* and *Adansonia* in Bombacoideae; *Durio* in Helicteroideae; *Firmiana* in Sterculioideae; *Theobroma* in Byttnerioideae; and *Microcos* and *Corchorus* in Grewioideae (following the circumscriptions as proposed by Cvetković et al. 2021). The *Fremontodendron* assembly has a similar chromosome number but about two-thirds the size of another basal member of the Malvatheca, *Ochroma pyramidale*, where $n = 42$ with a genome size of 1884 Mb (Sahu et al. 2023). The *Fremontodendron* assembly, while not at the chromosome scale of the previously discussed assemblies, demonstrates a comparable scaffold N50, suggesting similar contiguity with other published genomes in Malvaceae such as *Theobroma cacao* = 36.4 Mb and *Hibiscus sabdariffa* = 26.25 Mb (Kim et al. 2024).

The total percentage of repeat elements is higher than similar analyses of *T. cacao* genomes (52.3% to 61.5%), with the same study suggesting that domesticated genotypes have lower repeat content than wild counterparts (Nousias et al. 2024). *F. californicum*'s repeat content is similar to that of a

recent assembly of *H. sabdariffa* (74.48% masked), a domesticated tetraploid species (Kim et al. 2024). These results indicate that *F. californicum*'s genome has undergone a complex evolutionary history, maintaining a large amount of repetitive elements, especially a relatively high percentage of as-of-yet unclassified elements (33.89% and 35.7% in haplotype 1 and haplotype 2, respectively). Whole genome multiplications (WGMs) have been documented throughout Malvaceae s.l. with a notable allopolyploid multiplication occurring in the common ancestor of the Malvatheca (Conover et al. 2019). This multiplication may be best described as a decaploidization of the Malvaceae ancestral karyotype ($n = 11$) resulting from an allotetraploid hybridizing with an allohexaploid, which was also the common ancestor of the Helicteroideae. In addition to WGMs, chromosomal structural changes, including reciprocal translocations and chromosomal fusions, have played a considerable role in shaping the genomic architecture of the Malvatheca (Sun et al. 2024). These findings may help to explain why *F. californicum*'s diploid genome has a relatively high chromosome count ($n = 46$), a large volume of repetitive elements, and observed duplications among the BUSCO genes (Table 3). This assembly will be a tool to help further clarify Malvaceae s.l.'s complex evolutionary history.

The genome assembly of *F. californicum* will also become the basis for future studies to understand the genomic and morphological variation within *Fremontodendron*, which will have consequences for the taxonomic identity and status as rare of different populations throughout California. Establishing the taxonomic identities of morphologically ambiguous populations will clarify the precise distribution of the widespread and morphologically variable *F. californicum* and its rare and endangered sister taxa, *F. decumbens* and *F. mexicanum*. This, in turn, will have important management and conservation implications. This genomic resource furthers the goals of the CCGP to generate genomic data across multiple ecoregions, and between widespread and restricted/rare taxa, within California. This genome assembly, in combination with subsequent resequencing data funded through the CCGP, will provide a rich source of information which can guide current and future conservation strategies in the state.

Acknowledgments

PacBio Sequel II/IIe library prep and sequencing were carried out at the DNA Technologies and Expression Analysis Core at the UC Davis Genome Center, supported by NIH Shared Instrumentation Grant 1S10OD010786-01. Deep sequencing of Omni-C libraries used the Novaseq S4 sequencing platforms at the Vincent J. Coates Genomics Sequencing Laboratory at UC Berkeley, supported by NIH S10 OD018174 Instrumentation Grant. We thank the staff at the UC Davis DNA Technologies and Expression Analysis Core and the UC Santa Cruz Paleogenomics Laboratory for their diligence and dedication to generating high-quality sequence data. Support from the UC Davis Agricultural Experiment Station (project number CA-D-PLS-6273-H) is gratefully acknowledged.

Author contributions

William T. McMahan (Formal analysis, Project administration, Writing—original draft, Writing—review & editing), Merly Escalona (Data curation, Formal analysis, Methodology, Project administration, Software, Visualization, Writing—original draft, Writing—review & editing), Reed Kenny (Data curation, Formal analysis, Methodology), Mohan P.A. Marimuthu (Data curation, Methodology, Writing—original draft), Oanh Nguyen (Data curation, Methodology, Writing—original draft), Colin W. Fairbairn (Data curation, Methodology), William Seligmann (Data curation, Methodology, Writing—original draft), Courtney Miller (Project administration, Resources), H. Bradley Shaffer (Funding acquisition, Project administration, Resources, Supervision), Shannon Still (Conceptualization, Data curation, Funding acquisition, Writing—original draft), and Daniel Potter (Conceptualization, Data curation, Funding acquisition, Project administration, Supervision, Writing—original draft)

Funding

This work was supported by the California Conservation Genomics Project, with funding provided to the University of California by the State of California, State Budget Act of 2019 [UC Award ID RSI-19-690224].

Data availability

Data generated for this study are available under NCBI BioProject PRJNA1126700. Raw sequencing data for sample Potter1011 (NCBI BioSample SAMN35735723) are deposited in the NCBI Short Read Archive (SRA) under SRR29646341 for PacBio HiFi sequencing data, and SRR29646339, SRR29646340 for the Omni-C Illumina sequencing data. GenBank accessions for both primary and alternate assemblies are GCA_030761205.1 and GCA_030761185.1; and for genome sequences JAULJK000000000 and JAULJL000000000. Assembly scripts and other data for the analyses presented can be found at the following GitHub repository: www.github.com/ccgproject/ccgp_assembly

References

Abdennur N, Mirny LA. Cooler: scalable storage for hi-C data and other genomically labeled arrays. *Bioinformatics*. 2020;**36**(1):311–316. https://doi.org/10.1093/bioinformatics/btz540

Abrahamson IL. Fire effects information system, [online]. In: *Fremontodendron californicum, California flannelbush*. U.S.: Department of Agriculture, Forest Service, Rocky Mountain Research Station, Missoula Fire Sciences Laboratory (Producer); 2021.

Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL. BLAST+: architecture and applications. *BMC Bioinformatics*. 2009;**10**:1–9. https://doi.org/10.1186/1471-2105-10-421

Challis R, Richards E, Rajan J, Cochrane G, Blaxter M. BlobToolKit–interactive quality assessment of genome assemblies. G3: genes, genomes. *Genetics*. 2020;**10**(4):1361–1374. https://doi.org/10.1534/g3.119.400908

Challis R, Kumar S, Sotero-Caio C, Brown M, Blaxter M. Genomes on a tree (GoaT): a versatile, scalable search engine for genomic and sequencing project metadata across the eukaryotic tree of life. *Wellcome Open Res*. 2023;**8**:24. https://doi.org/10.12688/wellcomeopenres.18658.1

Cheng H, Concepcion GT, Feng X, Zhang H, Li H. Haplotype-resolved de novo assembly using phased assembly graphs with hifiasm. *Nat Methods*. 2021;**18**(2):170–175. https://doi.org/10.1038/s41592-020-01056-5

CNDDB. State of California Natural Resources Agency. Department of Fish and Wildlife. Biogeographic data branch. California natural diversity database (CNDDB). State and federally listed endangered, threatened, and rare plants of California. January 2, 2020. 2020: accessed 2020 March 8.

CNPS. *(California native plant society, rare plant program). 2020. Inventory of rare and endangered plants of California (online edition, v8–03 0.39)*Website. accessed 2020 March 7.

Conover JL, Karimi N, Stenz N, Ané C, Grover CE, Skema C, Tate JA, Wolff K, Logan SA, Wendel JF, et al. A Malvaceae mystery: A mallow maelstrom of genome multiplications and maybe misleading methods? *J Integr Plant Biol*. 2019;**61**:12–31. https://doi.org/10.1111/jipb.12746

Cvetković T, Areces-Berazain F, Hinsinger DD, Thomas DC, Wieringa JJ, Ganesan SK, Strijk JS. Phylogenomics resolves deep subfamilial relationships in Malvaceae sl. *G3*. 2021;**11**(7):jkab136. https://doi.org/10.1093/g3journal/jkab136

Danecek P, Bonfield JK, Liddle J, Marshall J, Ohan V, Pollard MO, Whitwham A, Keane T, McCarthy SA, Davies RM, et al. Twelve years of SAMtools and BCFtools. *Gigascience*. 2021;**10**(2):giab008

Ghurye J, Pop M, Koren S, Bickhart D, Chin CS. Scaffolding of long read assemblies using long range contact information. *BMC Genomics*. 2017;**18**:1–11.

Ghurye J, Rhie A, Walenz BP, Schmitt A, Selvaraj S, Pop M, Phillippy AM, Koren S. Integrating hi-C links with assembly graphs for chromosome-scale assembly. *PLoS Comput Biol*. 2019;**15**(8):e1007273. https://doi.org/10.1371/journal.pcbi.1007273

Gurevich A, Saveliev V, Vyahhi N, Tesler G. QUAST: quality assessment tool for genome assemblies. *Bioinformatics*. 2013;**29**(8):1072–1075. https://doi.org/10.1093/bioinformatics/btt086

Hanes MM. "*Malvaceae Juss.*" *pp. 187–37 5 in: Flora of North America editorial committee, eds. Flora of North America north of Mexico. Volume 6: Magnoliophyta: Cucurbitaceae to Droseraceae. Flora of North America north of Mexico. Volume 6: Magnoliophyta: Cucurbitaceae to Droseraceae*. New York: Oxford University Press; 2015.

Inglis PW, Pappas MCR, Resende LV, Grattapaglia D. Fast and inexpensive protocols for consistent extraction of high quality DNA and RNA from challenging plant and fungal samples for high-throughput SNP genotyping and sequencing applications. *PLoS One*. 2018;**13**(10):e0206085. https://doi.org/10.1371/journal.pone.0206085

Kelman WM. A revision of *Fremontodendron* (Sterculiaceae). *Syst Bot*. 1991;**16**:3–20. https://doi.org/10.2307/2418969

Kelman WM, Broadhurst L, Brubaker C, A. and Franklin. Genetic relationships among *Fremontodendron* (Sterculiaceae) populations of the central sierra Nevada foothills of California. *Madrono*. 2006;**53**:380–387. https://doi.org/10.3120/0024-9637(2006)53[380:GRAFSP]2.0.CO;2

Kerpedjiev P, Abdennur N, Lekschas F, McCallum C, Dinkla K, Strobelt H, Luber JM, Ouellette SB, Azhir A, Kumar N, et al. HiGlass: web-based visual exploration and analysis of genome interaction maps. *Genome Biol*. 2018;**19**:1–12.

Kim T, Lee JH, Seo HH, Moh SH, Choi SS, Kim J, S.G. & Kim. Genome assembly of *Hibiscus sabdariffa* L. provides insights into metabolisms of medicinal natural products. G3: genes, genomes. *Genetics*. 2024;**14**(8):jkae134.

Korlach J, Gedman G, Kingan SB, Chin CS, Howard JT, Audet JN, Cantin L, Jarvis ED. De novo PacBio long-read and phased avian genome assemblies correct and add to reference genes generated with intermediate and short reads. *GigaScience*. 2017;**6**(10):gix085. https://doi.org/10.1093/gigascience/gix085

Lee S, Bakker CR, Vitzthum C, Alver BH, Park PJ. Pairs and Pairix: a file format and a tool for efficient storage and retrieval for hi-C read pairs. *Bioinformatics*. 2022;**38**(6):1729–1731. https://doi.org/10.1093/bioinformatics/btab870

Lenz LW. Chromosome numbers of some western American plants. *I El Aliso*. 1950;**2**(3):317–318.

Li H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. 2013; arXiv preprint arXiv:1303.3997.

Lloyd RM. A new species of *Fremontodendron* (Sterculiaceae) from the sierra Nevada foothills. *California Brittonia*. 1965;**17**(4):382–384. https://doi.org/10.2307/2805030

Manni M, Berkeley MR, Seppey M, Simão FA, Zdobnov EM. BUSCO update: novel and streamlined workflows along with broader and deeper phylogenetic coverage for scoring of eukaryotic, prokaryotic, and viral genomes. *Mol Biol Evol*. 2021;**38**(10):4647–4654. https://doi.org/10.1093/molbev/msab199

Nousias O, Zheng J, Li T, Meinhardt LW, Bailey B, Gutierrez O, Baruah IK, Cohen SP, Zhang D, Y. and Yin. Three de novo assembled wild cacao genomes from the upper Amazon. *Sci Data*. 2024;**11**(1):369. https://doi.org/10.1038/s41597-024-03215-1

Nyffeler R, Bayer C, Alverson WS, Yen A, Whitlock BA, Chase MW, Baum DA. Phylogenetic analysis of the Malvadendrina clade (Malvaceae sl) based on plastid DNA sequences. *Org Divers Evol*. 2005;**5**:109–123. https://doi.org/10.1016/j.ode.2004.08.001

Open2C, Abdennur N, Fudenberg G, Flyamer IM, Galitsyna AA, Goloborodko A, Imakaev M, Venev SV. Pairtools: from sequencing data to chromosome contacts. *PLoS Comput Biol*. 2024;**20**(5):e1012164.

Pflug JM, Holmes VR, Burrus C, Johnston JS, Maddison DR. Measuring genome sizes using read-depth, k-mers, and flow cytometry: methodological comparisons in beetles (Coleoptera). G3: genes, genomes. *Genetics*. 2020;**10**(9):3047–3060.

Preston RE, Whetstone RD, Atkinson TA. *Fremontodendron in Jepson Flora project*Jepson eFlora. 2012: [accessed 2023 July 22].

Ramírez F, Bhardwaj V, Arrigoni L, Lam KC, Grüning BA, Villaveces J, Habermann B, Akhtar A, Manke T. High-resolution TADs reveal DNA sequences underlying genome organization in flies. *Nat Commun*. 2018;**9**(1):189. https://doi.org/10.1038/s41467-017-02525-w

Rhie A, Walenz BP, Koren S, Phillippy AM. Merqury: reference-free quality, completeness, and phasing assessment for genome assemblies. *Genome Biol*. 2020;**21**:1–27.

Rhie A, McCarthy SA, Fedrigo O, Damas J, Formenti G, Koren S, Uliano-Silva M, Chow W, Fungtammasan A, Kim J, et al. Towards complete and error-free genome assemblies of all vertebrate species. *Nature*. 2021;**592**(7856):737–746. https://doi.org/10.1038/s41586-021-03451-0

Sahu SK, Liu M, Chen Y, Gui J, Fang D, Chen X, Yang T, He C, Cheng L, Yang J, et al. Chromosome-scale genomes of commercial timber trees (*Ochroma pyramidale*, *Mesua ferrea*, and *Tectona grandis*). *Sci Data*. 2023;**10**(1):512. https://doi.org/10.1038/s41597-023-02420-8

Shaffer HB, Toffelmier E, Corbett-Detig RB, Escalona M, Erickson B, Fiedler P, Gold M, Harrigan RJ, Hodges S, Luckau TK, et al. Landscape genomics to enable conservation actions: the California conservation genomics project. *J Hered*. 2022;**113**:577–588. https://doi.org/10.1093/jhered/esac020

Sim SB, Corpuz RL, Simmonds TJ, Geib SM. HiFiAdapterFilt, a memory efficient read processing pipeline, prevents occurrence of adapter sequence in PacBio HiFi reads and their negative impacts on genome assembly. *BMC Genomics*. 2022;**23**(1):157. https://doi.org/10.1186/s12864-022-08375-1

Smit AFA, Hubley R, Green P. RepeatMasker Open-4.0. 2013-2015. http://www.repeatmasker.org

Smit AFA, Hubley R. RepeatModeler Open-1.0. 2008-2015. http://www.repeatmasker.org

Sun P, Lu Z, Wang Z, Wang S, Zhao K, Mei D, Yang J, Yang Y, Renner SS, Liu J. Subgenome-aware analyses reveal the genomic consequences of ancient allopolyploid hybridizations throughout the cotton family. *Proc Natl Acad Sci*. 2024;**121**(15):e2313921121. https://doi.org/10.1073/pnas.2313921121

Tillich M, Lehwark P, Pellizzer T, Ulbricht-Jones ES, Fischer A, Bock R, Greiner S. GeSeq - versatile and accurate annotation of organelle genomes. *Nucleic Acids Res*. 2017;**45**(W1):W6–W11. https://doi.org/10.1093/nar/gkx391

Toffelmier E, Beninde J, Shaffer HB. The phylogeny of California, and how it informs setting multi-species conservation priorities. *J Hered*. 2022;**113**:597–603. doi.org/10.1093/jhered/esac045

Vurture GW, Sedlazeck FJ, Nattestad M, Underwood CJ, Fang H, Gurtowski J, Schatz MC. GenomeScope: fast reference-free genome profiling from short reads. *Bioinformatics*. 2017;**33**(14):2202–2204. https://doi.org/10.1093/bioinformatics/btx153

Wick RR, Schultz MB, Zobel J, Holt KE. Bandage: interactive visualization of de novo genome assemblies. *Bioinformatics*. 2015;**31**(20): 3350–3352. https://doi.org/10.1093/bioinformatics/btv383

Workman R, Timp W, Fedak R, Kilburn D, Hao S, Liu K. High molecular weight DNA extraction from recalcitrant plant species for third generation sequencing. PROTOCOL (version 1) available at protocol exchange. 2018. https://doi.org/10.1038/protex.2018.059

Zhou C, Brown M, Blaxter M, Darwin Tree of Life Project Consortium, McCarthy SA. *Oatk: a de novo assembly tool for complex plant organelle genomes*bioRxiv. 2024:2024–2010.

Journal of Heredity, 2026, 117, 600–610
https://doi.org/10.1093/jhered/esaf105
Genome Resources