Journal of Heredity, 2026, **117**, 592–599
https://doi.org/10.1093/jhered/esaf102
Advance access publication 15 December 2025
Genome Resources

American Genetic Association
OXFORD

Genome Resources

A genome assembly of Douglas' meadowfoam, *Limnanthes douglasii* R. BR.

Brian L. Dorsey[1,*], Merly Escalona[2], Mohan P.A. Marimuthu[3], Oanh Nguyen[3], Noravit Chumchim[3], Colin W. Fairbairn[4], William Seligmann[4], Courtney Miller[5], Erin Toffelmier[5,6], H. Bradley Shaffer[5,6], Vanessa M. Handley[7,8,*]

[1]The Huntington Library, Art Museum, and Botanical Gardens, San Marino, CA 91108, United States, [2]Department of Biomolecular Engineering, University of California Santa Cruz, Santa Cruz, CA 95064, United States, [3]DNA Technologies and Expression Analysis Core Laboratory, Genome Center, University of California-Davis, CA 95616, United States, [4]Department of Ecology and Evolutionary Biology, University of California, Santa Cruz, Santa Cruz, CA 95064, United States, [5]Department of Ecology and Evolutionary Biology, University of California, Los Angeles, CA 90095, United States, [6]La Kretz Center for California Conservation Science, Institute for Environment and Sustainability, University of California, Los Angeles, CA 90095, United States, [7]California Academy of Sciences, Botany Department, San Francisco, CA 94118, United States, [8]Montgomery Botanical Center, Coral Gables, FL 33156, United States

*Corresponding authors: Brian L. Dorsey, The Huntington Library, Art Museum, and Botanical Gardens, San Marino, CA 91108, USA. Email: bdorsey@huntington.org; Vanessa M. Handley. Email: vhandley@calacademy.org

Abstract

The meadowfoams are annual plants in the genus *Limnanthes*, a group characterized by striking displays of bloom in damp, grassy habitats. The genus is endemic to western North America and, of the seven species and nineteen subspecies currently described, several are listed as rare or endangered due to extensive habitat loss. Species in the genus exhibit diverse reproductive systems and associations with specialist pollinators, both of which influence genetic diversity and gene flow. However, genomic studies across the genus remain limited, particularly for threatened taxa, hindering effective conservation strategies. Morphological variation within *Limnanthes* subspecies is also poorly understood, and phylogenetic relationships among several taxa remain unresolved. As part of the California Conservation Genomics Project, a reference genome for *Limnanthes douglasii* R. BR. has been assembled to help address these gaps. This genomic resource will support research into the evolutionary relationships, ecological interactions, and population structure of *Limnanthes*. Additionally, given that *Limnanthes* seeds yield a unique, high-stability oil, the genome has agricultural relevance, particularly for enhancing oil production and desirable reproductive traits in cultivated taxa.

Key words: Limnanthes, conservation, seed oil, vernal pool, phylogenetics

Introduction

Limnanthes is one of two genera within Limnanthaceae, a small family of flowering plants endemic to temperate North America. *Limnanthes* occurs throughout much of California and into southeastern Oregon, with one disjunct species endemic to British Columbia. Commonly called meadowfoams, these diminutive but showy annual herbs are vibrant elements of the endangered vernal pool habitats of California and southern Oregon (Fig. 1). Approximately 90% of the vernal pool habitat present before Spanish colonization has been lost to land conversion (Holland 2009) and, consequently, multiple taxa in *Limnanthes* are listed as rare or endangered. Species ranges in California vary in size with some taxa occupying over a dozen counties spanning much of the Central Valley (*L. douglasii* R. Br., *L.alba* Hartweg ex Bentham, *L. montana* Jepson), to a few dozen occurrences in a single county (*L bakeri* J.T. Howell). Reproductive systems of *Limnanthes* species vary from strictly selfing through gynodioecy to outcrossing, with several species documented to have specialist bee pollinators. Both pollinators and sexual system have been linked to genetic diversity and gene flow (Arroyo 1973; Thorp 1990; Sloop et al. 2012). While several studies have estimated genetic diversity and structure within a few species (Sloop et al. 2011; Meyers et al. 2012; Sloop et al. 2012), comprehensive evaluation of the population and landscape genomics of the threatened species is lacking. Understanding the dynamics and structure of genomic diversity within threatened species is a prerequisite to successful conservation efforts. Comparisons of these parameters with those of more abundant species would provide targets for successful reintroductions. Moreover, landscape level genomic diversity data would allow testing of hypotheses regarding connectivity of endangered vernal pool habitats.

We report here a high-quality reference genome assembly for *Limnanthes douglasii*, the first genome assembly for the Limnanthaceae family. This genome was generated as part of the California Conservation Genomics Project

Received on 3 June 2025; revised on 27 November 2025; accepted on 6 December 2025

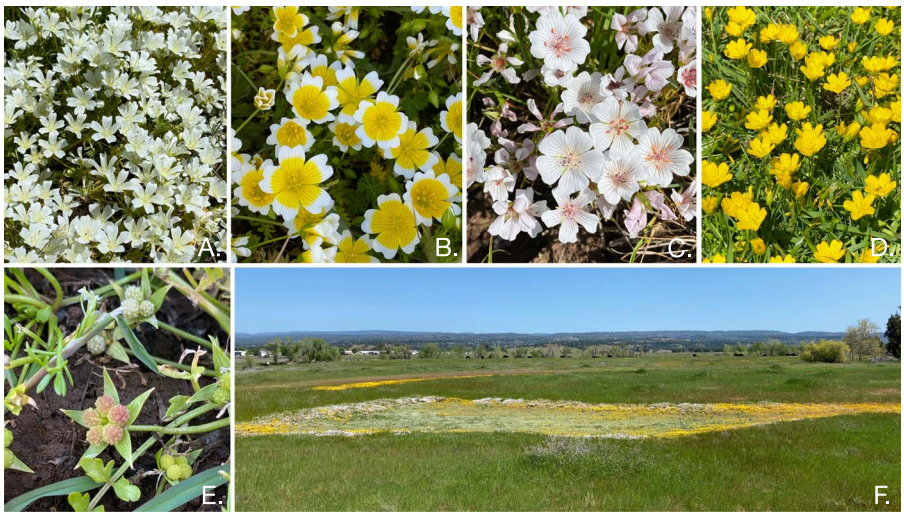

Fig. 1. *Limnanthes douglasii* is an annual plant with a distribution from Central California through southwestern Oregon. Subspecies are primarily characterized by variations in floral morphology, (A) *L. douglasii* subsp. *nivea*, (B) *L. douglasii* subsp. *douglasii*, (C) *L. douglasii* subsp. *rosea*, (D) *L. douglasii* subsp. *sulphurea*, and fruit ornamentation, (E) *L. douglasii* subsp. *rosea*. Plants are typically found in vernal pools (F), seasonally wet meadows or seeps.

(CCGP), a State-wide initiative which aims to produce landscape genomic data for over 200 species from across all California bioregions (Beninde et al. 2022; Fiedler et al. 2022; Shaffer et al. 2022; Toffelmier et al. 2022). *L. douglasii* is a relatively widespread species occurring throughout the central and northern Coast Ranges, the Central Valley, and the Sierra Nevada foothills of California. Of the six recognized subspecies, four, including two narrow endemics, overlap in the western half of the range and two in the eastern half. There is also an enigmatic variant population found in an agricultural field near Half Moon Bay that most resembles *L. macounii* of British Columbia, but which clustered more closely with samples of *L. douglasii* in an initial phylogenetic study (Meyers et al. 2010). This species contains the most distinct subspecies in the genus with multiple flower color patterns, although intermediates have been observed. The relationships among *L. douglasii* and the three other species within section *Limnanthes* (*L. bakeri*, *L. vinculans* Ornduff, *L. macounii* Trelease) are the least understood. These taxa were not well-supported as reciprocally monophyletic in Meyers et al. (2010) nor were the subspecies of *L. douglasii* resolved as distinct. This lack of phylogenetic resolution hampers assessment and recovery efforts for the rare and imperiled taxa.

Limnanthes is also the source of a distinct seed oil whose stability, among other properties, make it useful in applications from cosmetics to biodiesel (Moser et al. 2010; Kling 2021). Varieties developed from *L. alba* are grown commercially on a relatively small scale in Western Oregon (Jolliff et al. 1981; Kling 2021). Studies have investigated the transcriptome of developing seed to characterize the genetic basis of oil production (Slabaugh et al. 2015) and transgenic experiments have produced the novel *Limnanthes* seed oil in *Glycine* and *Brassica* species (Jadhav et al. 2005). Further characterization of the genomic basis for this promising oil could help improve yield and produce varieties with reproductive systems more amenable to cultivation.

The *L. douglasii* genome will provide a foundational resource for investigating the ecology, evolutionary history and genomic basis of key traits in *Limnanthes*. More broadly, it will support studies of the interactions between *Limnanthes* species and their pollinators, as well as illuminate the history and connectivity of remaining critical vernal pool habitats. In addition, the availability of a high-quality reference genome will facilitate the development of crop improvement strategies targeting oil production and reproductive traits, hence enhancing the agricultural potential of *Limnanthes*.

Methods

Biological materials

We collected an entire young plant of *L. douglasii* subsp. *nivea* (C.T. Mason) C.T. Mason on 2021 May 27 from a seasonally wet meadow near Howard Lake, Mendocino NF, Mendocino Co., CA USA (39.87874, −122.98828). A voucher specimen (Handley 716) is deposited in HNT. This plant was grown to maturity in potting soil in Oakland, CA. All above ground tissue, excluding senescing tissue and developing fruit, was harvested for DNA extraction on 21 June 2021.

Nucleic acid library preparation

High-molecular weight DNA extraction:

We extracted high molecular weight (HMW) genomic DNA (gDNA) from 1 g of a whole plant 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 solution; (ii) we repeated the tissue homogenate wash steps until the supernatant turned clear; (iii) we performed 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 = 1.89 and 260/230 = 2.20) measured using the NanoDrop ND-1000 spectrophotometer (Thermo Fisher Scientific, Waltham, MA). The DNA yield (15.2 μg) was quantified using a Quantus Fluorometer (QuantiFluor ONE dsDNA Dye assay; Promega, Madison, WI), and the size distribution of the DNA was estimated using the Femto Pulse system (Genomic DNA 165 kb kit, Agilent, Santa Clara, CA), where 58% of the DNA fragments were found to be 30 kilobases (kb) or longer.

HiFi PacBio library preparation and sequencing

The HiFi SMRTbell library was constructed using the SMRTbell prep kit 3.0 (Pacific Biosciences—PacBio, Menlo Park, CA; Cat. #102–182-700) according to the manufacturer's instructions. HMW gDNA 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 1X of SMRTbell cleanup beads provided in the SMRTbell prep kit 3.0 for the repair and a-tailing incubation at 37 °C for 30 min and 65 °C for 5 min, followed by ligation of overhang adapters at 20 °C for 30 min, clean-up using 1X SMRTbell cleanup beads, and nuclease treatment at 37 °C for 15 min. The SMRTbell library was size selected using 3.1X of 35% v/v diluted AMPure PB beads (PacBio, Cat. #100–265-900) to progressively remove SMRTbell templates < 5 kb. The 15 to 18 kb average HiFi SMRTbell library was sequenced at UC Davis DNA Technologies Core (Davis, CA) using two 8 M SMRT cells (PacBio, Cat #101–389-001), Sequel II sequencing chemistry 2.0, and 30-h movies each on a PacBio Sequel IIe sequencer.

Omni-C library preparation and sequencing

The Omni-C library was prepared using the Dovetail™ Omni-C™ Kit (Dovetail Genomics, Scotts Valley, CA) according to the manufacturer's protocol with slight modifications. First, specimen tissue (leaves, ID: Handley 716) was thoroughly ground with a mortar and pestle while cooled with liquid nitrogen. Nuclear isolation was then performed using published methods (Workman et al. 2019). 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 that was not internal to ligated fragments. An NGS library was generated using an NEB Ultra II DNA Library Prep kit (New England Biolabs, Ipswich, MA) with an Illumina compatible y-adaptor. Biotin-containing fragments were then captured using streptavidin beads. The post capture product was split into two replicates prior to 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 6000 platform (Illumina, CA) to generate approximately 100 million 2 x 150 bp read pairs per gigabase (Gb) of genome size.

Nuclear genome assembly

We assembled the genome of a *L. douglasii* subsp. *nivea* individual following the CCGP assembly pipeline Version 5.0, as outlined in Table 1 listing the tools and non-default 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; Cheng et al. 2022) in HiC mode with the filtered PacBio HiFi reads and the Omni-C short-reads, a process that generates two assemblies, one per haplotype. We then aligned the Omni-C data to each assembly separately 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. 2024). 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. 2019).

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 GenomeScope2.0 (Ranallo-Benavidez et al. 2020) 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 (quality value, 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 is based on the size of the contigs generated by HiFiasm on 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 ; C = % genome represented by the first 'n' scaffolds, following a karyotype of 2n = 10 for this species (Genome on a Tree—GoaT; tax_name [*L. douglasii*]; Challis et al. 2023; Rice et al. 2015; Zuo et al. 2022). Quality metrics for the notation were calculated on haplotype one assembly.

Results

Sequencing data

The Omni-C library generated 627.97 million read pairs and the PacBio HiFi library generated 4.58 million reads. The PacBio HiFi sequences yielded ~56X genome coverage and had an N50 read length of 16 143 bp; a minimum read length

Table 1. Assembly pipeline and software used.

Assembly step	Software and any non-default options	Version
Initial assembly		
Filtering PacBio HiFi adapters	HiFiAdapterFilt	Commit 64d1c7b
K-mer counting	Meryl (k = 21)	1
Estimation of genome size and heterozygosity	GenomeScope (k = 21)	2
De novo assembly (contiging)	HiFiasm (Hi-C Mode, −primary, output hic.hap1.p_ctg, hic.hap2.p_ctg)	0.16.1-r375
Scaffolding		
Omni-C data alignment	Arima Genomics Mapping Pipeline	Commit 2e74ea4
Arima Genomics Mapping Pipeline (AGMP)	BWA-MEM	0.7.17-r1188
	samtools	1.11
	filter_five_end.pl (AGMP)	Commit 2e74ea4
	two_read_bam_combiner.pl (AGMP)	Commit 2e74ea4
	picard	2.27.5
Omni-C Scaffolding	SALSA (-DNASE, −i 20, −p yes)	2
Omni-C Contact map generation		
Short-read alignment	BWA-MEM (-5SP)	0.7.17-r1188
SAM/BAM processing	samtools	1.11
SAM/BAM filtering	pairtools	0.3.0
Pairs indexing	pairix	0.3.7
Matrix generation	cooler	0.8.10
Matrix balancing	hicExplorer (hicCorrectmatrix correct –filterThreshold −2 4)	3.6
Contact map visualization	HiGlass	2.1.11
	PretextMap	0.1.4
	PretextView	0.1.5
	PretextSnapshot	0.0.3
Manual curation tools	Rapid curation pipeline (Wellcome Trust Sanger Institute, Genome Reference Informatics Team)	Commit 7acf220c
Genome quality assessment		
Basic assembly metrics	QUAST (−-est-ref-size)	5.0.2
	BUSCO (−m geno, −l embryophyta)	5.0.0
Assembly completeness	Merqury	1/29/20
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
General contamination screening	BlobToolKit (HiFi coverage, BUSCO = embryophyta, NCBI Taxa ID = 28 973)	2.3.3

of 127 bp; a mean read length of 14 097 bp; and a maximum read length of 55 423 bp (see Supplementary Figure S1 for read length distribution). Based on the PacBio HiFi data, Genomescope 2.0 estimated a genome size of 1.79 Gb, a 0.149% sequencing error rate, and 2.84% heterozygosity. The k-mer spectrum shows a bimodal distribution with a major peak at ∼17-fold coverage and a minor peak at ∼35-fold coverage (Fig. 2A).

Nuclear genome assembly

The final genome assembly (dmLimDoug1) consists of two phased haplotypes. The assemblies are similar in size, and similar in size to the estimated genome assembly size from GenomeScope2.0.

The haplotype one assembly (dmLimDoug1.0.hap1) consists of 659 scaffolds spanning 1.78 Gb with a contig N50 of 2.89 Mb, a scaffold N50 of 333.67 Mb, the largest contig size of 12.42 Mb, and the largest scaffold size of 405.37 Mb. The haplotype two assembly (dmLimDoug1.0.hap2) consists of 655 scaffolds spanning 1.78 Gb with a contig N50 of 2.9 Mb, a scaffold N50 of 333.49 Mb, the largest contig size of 12.42 Mb and the largest scaffold size of 406.34 Mb.

The haplotype one assembly has a BUSCO completeness score for the Embryophyta gene set of 99.0%, a base pair quality value (QV) of 64.37, a k-mer completeness of 69.45%, and a frameshift indel QV of 49.50. The haplotype two assembly has a BUSCO completeness score for the Embryophyta gene set of 99.0%, a base pair QV of 64.38, a k-mer completeness of 69.45%, and a frameshift indel QV of 49.18.

During manual curation we made a total of 728 joins (389 on haplotype one and 339 on haplotype two) and 122 breaks (38 on haplotype one and 74 on haplotype two) based on the Omni-C contact map signal. We closed a total of 196 gaps (99 on haplotype one and 97 on haplotype two) and we filtered out 2 contigs corresponding to contamination from Proteobacteria (*Xanthomonas campestris*, size: 29266 bp; and *Pantoea agglomerans*, size: 19608 bp). The Omni-C contact maps show highly contiguous assemblies, with chromosome-length scaffolds (Fig. 2B). Assembly statistics are reported in Table 2 and represented graphically in Fig. 2C. We have deposited the genome assembly on NCBI GenBank (see Table 2 and Data availability for details).

Discussion

Here we report on the assembly of a reference genome for *L. douglasii* (Limnanthaceae, Brassicales), a winter annual endemic to California vernal pools and allied habitats. The assembled sequence length is 1.789 Gb and the second longest

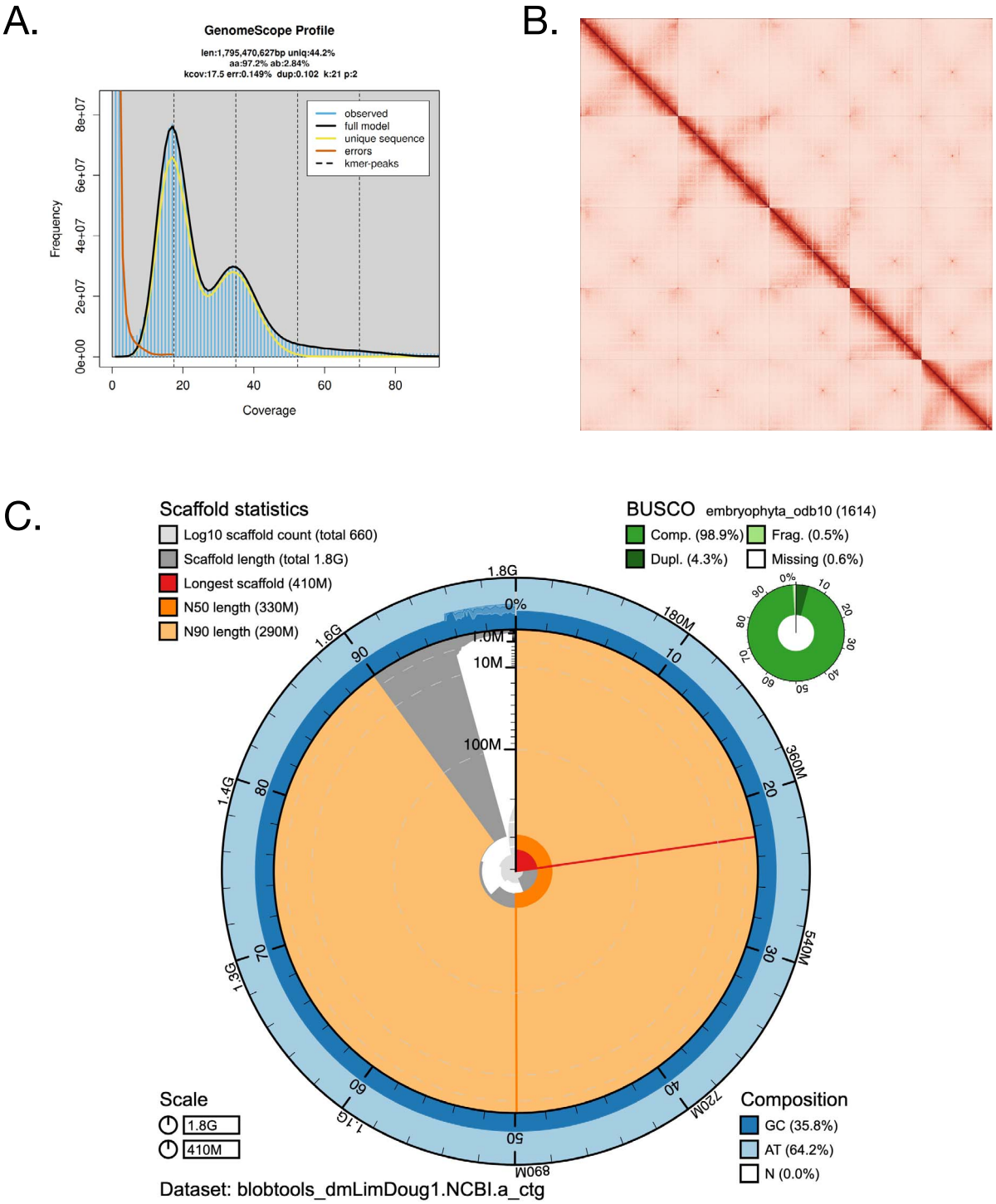

Fig. 2. (A) K-mer distribution generated by PacBio HiFi data using GenomeScope2.0. (B) omni-C contact maps generated with PretexSnapshot. (C) BlotToolKit snail plot showing quality metrics for the *Limnanthes douglasii* subsp. *nivea* haplotype 1.

of the 112 reference genomes from Brassicales available from NCBI. This estimate is shorter than the C-value-based estimate of Zuo et al. (2022) for *L. douglasii*, but within the range they documented for the genus. The contig N50 is 2.89 Gb and the scaffold N50 is 333 Gb. Among the published Brassicales genomes these N50 values are in the 67th and 99th percentiles, respectively.

Limnanthaceae is the first or second lineage to diverge after the At-ß whole genome duplication event in the Brassicales (Edger et al. 2018). Zuo et al. (2022) suggested that this phylogenetic position, coupled with the low chromosome number, among other factors, make *Limnanthes* a prime candidate as a model organism for investigating genome evolution in the Brassicales. Our genome assembly also represents the only lineage between published genomes of Caricaceae and Moringaceae and those of the core Brassicales, filling an important phylogenetic gap in genome coverage for the order.

At the species and genus level, this new reference genome is the first step in developing data sets for landscape genomics of

Table 2. Sequencing and assembly statistics, and accession numbers.

Bio Projects & Vouchers	CCGP NCBI BioProject		PRJNA720569				
	Genera NCBI BioProject		PRJNA765618				
	Species NCBI BioProject		PR				
	NCBI BioSample		SAMN36697170				
	Specimen identification		HANDLEY_CCGP117; Handley 716 HNT				
	NCBI Genome accessions		Haplotype 1	Haplotype 2			
	Assembly accession		JAUPFV000000000	JAUPFW000000000			
	Genome sequences		GCA_036936605.1	GCA_036936615.1			
Genome Sequence	PacBio HiFi reads	Run	1 PACBIO_SMRT (Sequel IIe) run: 4.6 M spots, 64.6G bases, 39.5Gb				
		Accession	SRX25151928				
	Omni-C Illumina reads	Run	2 ILLUMINA (Illumina NovaSeq 6000) runs: 628 M spots, 189.6G bases, 62.9Gb				
		Accession	SRX25151929, SRX25151930				
Genome Assembly Quality Metrics	Assembly identifier (Quality code*)		dmLimDoug1(6.8.P6.Q64.C95)				
	HiFi Read coverage[a]		36X				
			Haplotype 1	Haplotype 2			
	Number of contigs		1564	1566			
	Contig N50 (bp)		2 897 994	2 901 631			
	Contig NG50[a]		2 894 932	2 900 618			
	Longest Contigs		12 421 381	12 421 381			
	Number of scaffolds		659	655			
	Scaffold N50		333 674 471	333 494 407			
	Scaffold NG50[a]		333 674 471	333 494 407			
	Largest scaffold		405 374 481	406 344 990			
	Size of final assembly		1 789 517 456	1 789 518 053			
	Phased block NG50[a]		2 876 516	2 907 125			
	Gaps per Gbp (# Gaps)		368(658)	509(911)			
	Indel QV (Frame shift)		48.77897644	48.78259104			
	Base pair QV		64.3782	64.3843			
			Full assembly = 64.3812				
	k-mer completeness		69.4541	69.4541			
			Full assembly = 69.4541				
	BUSCO completeness[b]		C	S	D	F	M
	(embryophyta) $n = 1614$	H1[c]	99.00%	94.70%	4.30%	0.50%	0.50%
		H2[c]	99.00%	94.70%	4.30%	0.50%	0.50%

[a]Read coverage and NGx statistics have been calculated based on the estimated genome size of 1.7 Gb [b]BUSCO Scores. Complete BUSCOs (C). Complete and single-copy BUSCOs (S). Complete and duplicated BUSCOs (D). Fragmented BUSCOs (F). Missing BUSCOs (M). [c](H1) Haplotype 1 and (H2) Haplotype 2 assembly values. *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 known karyotype for the species of 2n = 10. Quality code for all the assembly denoted by primary assembly (dmLimDoug1.0.hap1)

rare and endangered species of *Limnanthes* via genome resequencing of individuals across their respective ranges. It will also enable more accurate assembly of reduced representation genomic sequencing reads for population-level conservation genomic analysis and phylogenomic study.

As currently circumscribed, *Limnanthes* consists of seven species and nineteen subspecies divided between two well-defined sections (Meyers et al. 2010; Morin 2020). The number of subspecies reflects the morphological variation within named species although the distribution of this variation among individuals and populations has not been comprehensively evaluated. Thus, it is unclear whether these traits reflect genetically structured entities or widespread phenotypic variation. Further, the few molecular phylogenetic studies to date have left unresolved the relationships of taxa within sections (Meyers et al. 2010; Zuo et al. 2022). This lack of phylogenetic resolution among taxa has implications for the conservation of six rare, threatened, or endangered taxa within the genus. Clear circumscription of evolutionarily significant units is necessary for assessment of threatened status and conservation efforts. As part of the California Conservation Genomics Project, the reference genome produced here and pending resequencing data will allow us to infer the phylogeny and taxonomic boundaries of the genus, test hypotheses of landscape genomic patterns, infer population genomic dynamics, and ultimately inform conservation decisions at the state level.

Acknowledgments

PacBio Sequel II/IIe library prep and sequencing was 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. We thank Dr Peggy Fiedler for encouraging our pursuit of a *Limnanthes*-focused project. We also thank Mendocino National Forest for permission to collect plant specimens and to Amie Garcia for assistance with field work for this project.

Author contributions

Brian L Dorsey (Conceptualization, Investigation, Project administration, Resources, Visualization, Writing—original draft, Writing—review & editing), Merly Escalona (Data curation, Formal analysis,

Investigation, Methodology, Project administration, Resources, Software, Validation, Visualization, Writing—original draft, Writing—review & editing), Mohan PA Marimuthu (Data curation, Formal analysis, Investigation, Software, Validation, Visualization), Oanh Nguyen (Formal analysis, Investigation, Resources, Validation, Visualization, Writing—original draft), Noravit Chumchim (Formal analysis, Investigation, Resources, Validation, Visualization, Writing—original draft), Colin W Fairbairn (Data curation, Formal analysis, Investigation, Methodology, Software, Validation), William Seligmann (Data curation, Formal analysis, Investigation, Resources, Software, Validation, Visualization), Courtney Miller (Project administration, Validation, Writing—original draft), Erin Toffelmier (Methodology, Project administration, Supervision, Validation), H Bradley Shaffer (Conceptualization, Funding acquisition, Methodology, Project administration, Supervision, Validation), and Vanessa Handley (Conceptualization, Funding acquisition, Investigation, Project administration, Resources, Visualization, Writing—original draft, Writing—review & editing)

Supplementary material

Supplementary material is available at *Journal of Heredity* online.

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 PRJNA986189. Raw sequencing data for sample HANDLEY_CCGP117 (Voucher: Handley 716 HNT; NCBI BioSample SAMN36697170) are deposited in the NCBI Short Read Archive (SRA) under SRR29647753 for PacBio HiFi sequencing data, and SRR29647751,SRR29647752 for the Omni-C Illumina sequencing data. GenBank accessions for both primary and alternate assemblies are GCA_036936605.1 and GCA_036936615.1; and for genome sequences JAUPFV000000000 and JAUPFW000000000. 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**:311–316. https://doi.org/10.1093/bioinformatics/btz540

Arroyo MTK. A Taximetric study of infraspecific variation in Autogamous Limnanthes floccosa (Limnanthaceae). *Brittonia*. 1973;**25**:177–191. https://doi.org/10.2307/2805936

Beninde J, Toffelmier E, Shaffer HB. A brief history of population genetic research in California and an evaluation of its utility for conservation decision-making. *J Hered*. 2022;**113**:604–614. https://doi.org/10.1093/jhered/esac049

Challis R, Richards E, Rajan J, Cochrane G, Blaxter M. *BlobToolKit – interactive quality assessment of genome assemblies*. G3 Genes|Genomes|Genetics. 2020;**10**: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 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**:170–175. https://doi.org/10.1038/s41592-020-01056-5

Cheng H, Jarvis ED, Fedrigo O, Koepfli K-P, Urban L, Gemmell NJ, Li H. Haplotype-resolved assembly of diploid genomes without parental data. *Nat Biotechnol*. 2022;**40**:1332–1335. https://doi.org/10.1038/s41587-022-01261-x

Edger PP, Hall JC, Harkess A, Tang M, Coombs J, Mohammadin S, Schranz ME, Xiong Z, Leebens-Mack J, Meyers BC, et al. Brassicales phylogeny inferred from 72 plastid genes: a reanalysis of the phylogenetic localization of two paleopolyploid events and origin of novel chemical defenses. *Am J Bot*. 2018;**105**:463–469. https://doi.org/10.1002/ajb2.1040

Fiedler PL, Erickson B, Esgro M, Gold M, Hull JM, Norris JM, Shapiro B, Westphal M, Toffelmier E, Shaffer HB. Seizing the moment: the opportunity and relevance of the California conservation genomics project to state and federal conservation policy. *J. Hered*. 2022;**113**:589–596. https://doi.org/10.1093/jhered/esac046

Ghurye J, Pop M, Koren S, Bickhart D, Chin C-S. Scaffolding of long read assemblies using long range contact information. *BMC Genomics*. 2017;**18**:527. https://doi.org/10.1186/s12864-017-3879-z

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**: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**:1072–1075. https://doi.org/10.1093/bioinformatics/btt086

Holland R. *California's Gerat Valley vernal pool habitat status and loss: Rephotorevised 2005*. Placer Land Trust; 2009.

Inglis P, Pappas M, Resende L, 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**:e0206085. https://doi.org/10.1371/journal.pone.0206085

Jadhav A, Marillia E-F, Babic V, Giblin EM, Cahoon EB, Kinney AJ, Mietkiewska E, Brost JM, Taylor DC. Production of 22:2Δ5,Δ13 and 20:1Δ5 in Brassica carinata and soybean breeding lines via introduction of Limnanthes genes. *Mol Breed*. 2005;**15**:157–167. https://doi.org/10.1007/s11032-004-4730-1

Jolliff GD, Calhoun W, Crane JM. *Meadowfoam (Limnanthes alba) : its research and development as a potential new oilseed crop for the Willamette Valley of Oregon (technical report No. 648), station bulletin*. Oregon State University Agricultural Experiment Station; 1981.

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**:125. https://doi.org/10.1186/s13059-018-1486-1

Kling, J.G, 2021. Meadowfoam Breeding. In: Irwin G, editor. *Plant Breeding Reviews*. John Wiley & Sons, Ltd, pp. 187–245. https://doi.org/10.1002/9781119717003.ch6

Korlach J, Gedman G, Kingan SB, Chin C-S, Howard JT, Audet J-N, 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**:gix085 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**:1729–1731. https://doi.org/10.1093/bioinformatics/btab870

Li H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *ArXiv*. 2013;1301.10.48550/arXiv.1303.3997

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**:4647–4654. https://doi.org/10.1093/molbev/msab199

Meyers SC, Liston A, Meinke R. A molecular phylogeny of Limnanthes (Limnanthaceae) and investigation of an anomalous Limnanthes population from California. *U S A Syst Bot*. 2010;**35**:552–558. https://doi.org/10.1600/036364410792495854

Meyers SC, Liston A, Meinke R. An evaluation of putative sympatric speciation within Limnanthes (Limnanthaceae). *PLoS One*. 2012;**7**:e36480. https://doi.org/10.1371/journal.pone.0036480

Morin NR. Limnanthes. In: Flora of North America Editorial Committee, editors. 1993+. *Flora of North America North of Mexico [Online]*. 25+ vols. New York and Oxford. Vol. 7. URL http://floranorthamerica.org/Limnanthes (accessed 4.23.25).

Moser BR, Knothe G, Cermak SC. Biodiesel from meadowfoam (Limnanthes alba L.) seed oil: oxidative stability and unusual fatty acid composition. *Energy Environ Sci*. 2010;**3**:318–327. https://doi.org/10.1039/B923740M

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**:e1012164. https://doi.org/10.1371/journal.pcbi.1012164

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**:189. https://doi.org/10.1038/s41467-017-02525-w

Ranallo-Benavidez TR, Jaron KS, Schatz MC. GenomeScope 2.0 and Smudgeplot for reference-free profiling of polyploid genomes. *Nat Commun*. 2020;**11**:1432 10.1038/s41467-020-14998-3

Rhie A, Walenz BP, Koren S, Phillippy AM. Merqury: reference-free quality, completeness, and phasing assessment for genome assemblies. *Genome Biol*. 2020;**21**:245. https://doi.org/10.1186/s13059-020-02134-9

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**:737–746. https://doi.org/10.1038/s41586-021-03451-0

Rice A, Glick L, Abadi S, Einhorn M, Kopelman NM, Salman-Minkov A, Mayzel J, Chay O, Mayrose I. The chromosome counts database (CCDB) – a community resource of plant chromosome numbers. *New Phytol*. 2015;**206**:19–26. https://doi.org/10.1111/nph.13191

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**:157. https://doi.org/10.1186/s12864-022-08375-1

Slabaugh MB, Cooper LD, Kishore VK, Knapp SJ, Kling JG. Genes affecting novel seed constituents in Limnanthes alba Benth: transcriptome analysis of developing embryos and a new genetic map of meadowfoam. *PeerJ*. 2015;**3**:e915. https://doi.org/10.7717/peerj.915

Sloop CM, Pickens C, Gordon SP. Conservation genetics of Butte County meadowfoam (Limnanthes floccosa ssp. californica Arroyo), an endangered vernal pool endemic. *Conserv Genet*. 2011;**12**:311–323. https://doi.org/10.1007/s10592-010-0142-4

Sloop CM, Eberl R, Ayres DR. Genetic diversity and structure in the annual vernal pool endemic Limnanthes vinculans Ornduff (Limnanthaceae): implications of breeding system and restoration practices. *Conserv Genet*. 2012;**13**:1365–1379. https://doi.org/10.1007/s10592-012-0380-8

Thorp R. Vernal pool flowers and host-specific bees. *Stud Herb*. 1990;**8**:109–122

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

Workman R, Fedak R, Kilburn D, Hao S, Liu K, Timp W. *High molecular weight DNA extraction from recalcitrant plant species for third generation sequencing*. 2019. https://doi.org/10.17504/protocols.io.4vbgw2n

Zuo S, Mandáková T, Kubová M, Lysak MA. Genomes, repeatomes and interphase chromosome organization in the meadowfoam family (Limnanthaceae, Brassicales). *Plant J*. 2022;**110**:1462–1475. https://doi.org/10.1111/tpj.15750

Journal of Heredity, 2026, 117, 592–599
https://doi.org/10.1093/jhered/esaf102
Genome Resources