Title

A highly contiguous genome assembly for the pocket mouse Perognathus longimembris longimembris

Journal

Journal of Heredity, 115(1)

ISSN

0022-1503

Authors

Kozak, Krzysztof M

Escalona, Merly

Chumchim, Noravit

et al.

Publication Date

2024-02-03

DOI

10.1093/jhered/esad060

Journal of Heredity, 2024, **115**, 130–138
https://doi.org/10.1093/jhered/esad060
Advance access publication 4 October 2023
Genome Resources

American Genetic Association

OXFORD

Genome Resources

A highly contiguous genome assembly for the pocket mouse *Perognathus longimembris longimembris*

Krzysztof M. Kozak[1,*,†], Merly Escalona[2,†], Noravit Chumchim[3], Colin Fairbairn[4], Mohan P.A. Marimuthu[3], Oanh Nguyen[3], Ruta Sahasrabudhe[3], William Seligmann[4], Chris Conroy[1], James L. Patton[1], Rauri C.K. Bowie[1] and Michael W. Nachman[1]

[1]Museum of Vertebrate Zoology and Department of Integrative Biology, University of California, Berkeley, CA 94720, 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

[†]These authors contributed equally.
[*]Corresponding author: Email: evoecogen@gmail.com

Corresponding Editor: Klaus-Peter Koepfli

Abstract

The little pocket mouse, *Perognathus longimembris*, and its nine congeners are small heteromyid rodents found in arid and seasonally arid regions of Western North America. The genus is characterized by behavioral and physiological adaptations to dry and often harsh environments, including nocturnality, seasonal torpor, food caching, enhanced osmoregulation, and a well-developed sense of hearing. Here we present a genome assembly of *Perognathus longimembris longimembris* generated from PacBio HiFi long read and Omni-C chromatin-proximity sequencing as part of the California Conservation Genomics Project. The assembly has a length of 2.35 Gb, contig N50 of 11.6 Mb, scaffold N50 of 73.2 Mb, and includes 93.8% of the BUSCO Glires genes. Interspersed repetitive elements constitute 41.2% of the genome. A comparison with the highly endangered Pacific pocket mouse, *P. l. pacificus*, reveals broad synteny. These new resources will enable studies of local adaptation, genetic diversity, and conservation of threatened taxa.

Key words: California Conservation Genomics Project, comparative genomics, conservation genetics, Heteromyidae, *Perognathus*, repetitive elements

Introduction

To understand how animals will cope with the rapidly changing conditions of the Anthropocene, we must discover the genomic underpinnings of specific morphological, physiological, and behavioral adaptations (Lancaster et al. 2022). Of special interest are organisms adapted to arid environments, since these habitats are expected to expand globally under changing climate (Mirzabaev et al. 2019). Comparing whole genomes of multiple individuals within a species is an effective approach to identify the evolutionary processes that generate such adaptations. Genomic data can also be used to identify important geographic regions for preservation efforts and facilitate the development of data-driven management plans for species of conservation concern.

Pocket mice (Castorimorpha: Heteromyidae) are highly specialized for arid environments and thus serve as useful models for understanding adaptation to these conditions (Altschuler et al. 1979; Marra et al. 2014). The little pocket mouse *P. longimembris* (Coues 1875) (Fig. 1) occurs in dry grasslands and shrub-steppe habitats of the Colorado, Great Basin and Mojave deserts, and coastal sage habitats along the southern California coast (Hafner 2016). There are 16 recognized subspecies (Fig. 2) (Hafner 2016; Patton and Fisher 2023), and they vary greatly in their ranges. *Perognathus longimembris pacificus* is federally endangered, and four subspecies are considered imperiled or critically imperiled (*Perognathus longimembris bangsi*, *Perognathus longimembris brevinasus*, *Perognathus longimembris internationalis*, *Perognathus longimembris salinensis*, and *Perognathus longimembris tularensis*), and thus listed as Species of Special Concern in California (Comrack et al. 2008). The effective population size of *P. l. pacificus* used to be an order of magnitude higher than today, but decreased precipitously in the last glacial period, showing that even robust populations may undergo dramatic reductions (Wilder et al. 2022).

Pocket mice are well adapted to arid habitats through seasonal dormancy and many other physiological, behavioral, and anatomical changes (Webster and Webster 1975; Jenkins and Breck 1998). For example, they have greatly elongated renal medullae for enhanced osmoregulation, and like some other heteromyid rodents, are able to survive indefinitely without access to free water (Altschuler et al. 1979; Marra et al. 2014). This attribute made them well-suited for experiments in space. On December 9th, 1972, five *P.*

Received June 29, 2023; Accepted September 30, 2023

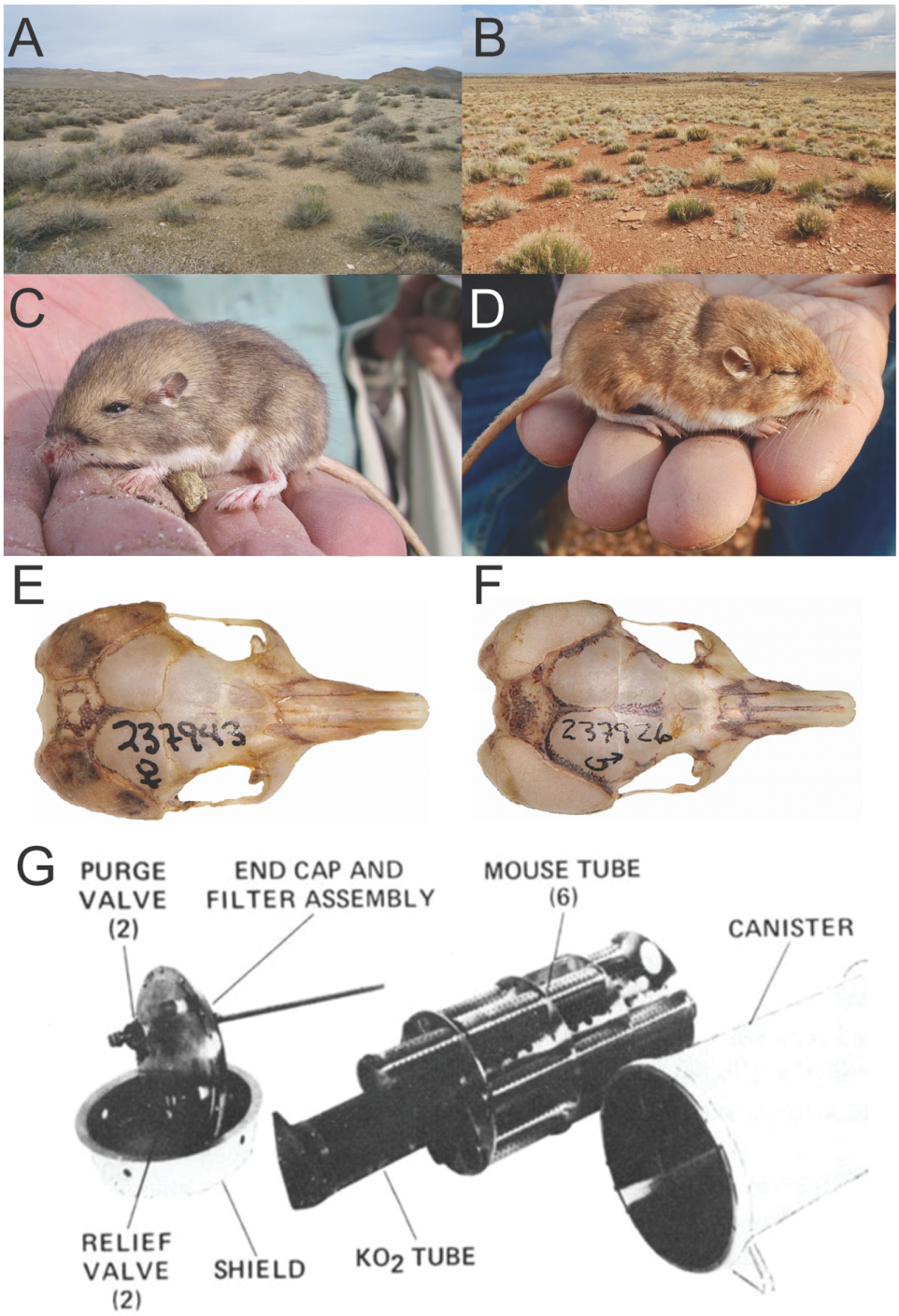

Fig. 1. *Perognathus longimembris*, the little pocket mouse. Arid habitats of *P. longimembris*: A) Harrisburg Flat, Death Valley National Park, Inyo County, California, United States; B) Bullrush Canyon, Mohave County, Arizona, USA. C) *P. l. panamintinus* and (D) torpid *P. l. arizonensi*. Pelage of pocket mice matches the color of the local soil. Pronounced intraspecific differences in cranial morphology can be seen between the topotypes of (E) *P. l. panamintinus* and (F) *P. l. arizonensis* from the collection of the Museum of Vertebrate Zoology, UC Berkeley. G) The rodent orbiting module of the Apollo 17 mission. Image credits: Peggy Moore (C), Carol Patton (D), James Patton (A, B, E, and F), NASA (G).

longimembris entered lunar orbit on board the Apollo 17 mission. Assisted by a primate (Captain Ronald E. Evans Jr), they set a record of 75 revolutions around the Moon. As subjects of the Biological Cosmic Ray Experiment, the rodents were flown in a specialized carrier module (Fig. 1G) and implanted with dosimeters to evaluate the impact of high-energy cosmic Z rays (Haymaker et al. 1975).

Although further space exploration by pocket mice is regrettably not planned, *Perognathus* remains an attractive system for studies of adaptation on Earth. Pocket mice have also served as an important model for understanding the genetic basis of coat color variation, as populations show color matching to the local substrate even over short distances (Fig. 1A–D) (Benson 1933). Their characteristically

enlarged auditory bullae likely help to hear predators in open environments, and variation in this trait across habitat types (Fig. 1E and F) suggests differences in auditory performance between subspecies (Webster and Webster 1975). However, little is known about the genetic basis of these complex phenotypes. An assembly of a reference genome is a useful step to uncover their molecular mechanisms by determining the composition of the genome.

A high-quality reference genome also provides an opportunity to characterize repetitive elements. Recent comparative studies have led to a better appreciation of the role that transposable elements (TEs) play in the evolution of genome size and complexity in vertebrates (Kapusta et al. 2017; Osmanski et al. 2023). The turnover of TEs is especially high in rodents (Kapusta et al. 2017), hence their annotation in pocket mice is helpful for understanding genome evolution over short timescales.

The overall goals of the California Conservation Genomics Project (CCGP) are to evaluate the genetic diversity and future prospects of diverse species across California with the aim of informing management decisions (Shaffer et al. 2022). A key aspect of this approach is a comparison of genetic diversity in common and endangered taxa. As part of CCGP, we describe a highly contiguous genome assembly and associated genomic resources for the little pocket mouse, the Mojave Desert subspecies *Perognathus longimembris longimembris* (Coues 1875) and compare it with that of the critically endangered *P. l. pacificus* (Mearns 1898) of the southern California coast (Fig. 2; see Wilder et al. 2022).

Methods

Sample, DNA preparation, and sequencing

Details of the laboratory procedures are described in the Supplementary Methods. Here we provide a brief outline. One male *P. l. longimembris* was trapped at the mouth of Freeman Canyon, Kern County, California, United States (Fig. 2) under collecting permit D S-192560001 to JLP, and sacrificed following the Guidelines of the American Society of Mammalogists (Sikes et al. 2016). Tissue was flash-frozen in LN_2 in the field, and a voucher specimen (skin and skull) was deposited in the mammal collection of the UC Berkeley Museum of Vertebrate Zoology (MVZ:Mamm:240093). High molecular weight genomic DNA was extracted using the Nanobind Tissue Big DNA kit (Pacific BioSciences—PacBio; Menlo Park, CA) and sheared to fragment sizes between 15 and 18 kb. We constructed a HiFi SMRTbell library (PacBio) with an average fragment size of 15 to 20 kb and sequenced it using six 8M SMRT cells on the PacBio Sequel II platform.

The Omni-C library was prepared from liver tissue according to the manufacturer's protocol with slight modifications (see Supplementary Methods). The NGS library from the Omni-C fragments was generated using an NEB Ultra II DNA Library Prep kit (NEB, Ipswich, MA) and sequenced on an Illumina NovaSeq 6000 platform as 150 paired-end reads.

Genome assembly

We assembled the genome of *P. l. longimembris* using the CCGP assembly pipeline Version 5.0 (Table 1). An initial partially phased diploid assembly was generated from adapter-filtered HiFi reads and the Omni-C data using HiFiasm in Hi-C mode (Cheng et al. 2021). This approach results in two assemblies, with some switches occurring between the parental genomes (Cheng et al. 2022). Next, Omni-C data were aligned to each assembly (https://github.com/ArimaGenomics/mapping_pipeline). The two assemblies were then scaffolded with SALSA (Ghurye et al. 2017, 2019). We have manually curated the resulting genomes using the Rapid Curation toolkit (https://gitlab.com/wtsi-grit/rapid-curation) and closed some of the remaining gaps using the HiFi reads (https://github.com/merlyescalona/yagcloser). Furthermore, we identified and resolved major misassemblies based on Omni-C contact maps and synteny analyses against *P. l. pacificus* (GCF_023159225.1). The mitochondrial genome was assembled with the MitoHiFi pipeline (2022 Uliano-Silva et al. 2022), guided by the *Castor canadensis* sequence (NCBI:NC_033912.1) (Lok et al. 2017). We checked for contamination using the BlobToolKit (Challis et al. 2020). The quality of the nuclear genome was assessed based on the k-mer distribution reported by meryl (https://github.com/marbl/meryl). We assessed the contiguity of the assembly with QUAST (Gurevich et al. 2013), and its functional completeness by checking ortholog gene sets with BUSCO against the Glires database (Manni et al. 2021), and scanning for frameshifts (Korlach et al. 2017, p. 20).

Genome annotation

Repetitive elements were identified using the RepeatModeler2 (Flynn et al. 2020) and RepeatMasker pipeline (Smit et al. 2013), with emphasis on Long Terminal Repeats (Ellinghaus et al. 2008; Ou and Jiang 2018). Homology to elements known from Glires was examined by searches against five databases (Vassetzky and Kramerov 2013; Bao et al. 2015; Penzkofer et al. 2017; Storer et al. 2021; Liao et al. 2022). Gene annotations were transferred from the *P. l. pacificus* assembly (GCF_023159225.1) with Liftoff (Shumate and Salzberg 2021, 2022).

Results

Genome assembly quality and statistics

The Omni-C and PacBio HiFi sequencing libraries generated 217.3 million read pairs and 6.6 million reads, respectively. The latter yielded ~40-fold coverage based on the Genomescope2.0 (Ranallo-Benavidez et al. 2020) genome size estimate of 2.27 Gb (N50 read length 13,870 bp; minimum read length 265 bp; mean read length 13,504 bp; maximum read length 51,878 bp). From PacBio HiFi reads we estimated a 0.214% sequencing error rate. The k-mer spectrum of HiFi reads shows a bimodal distribution with two major peaks at 16 and 28, which correspond to homozygous and heterozygous states of a diploid species (Fig. 3A).

Multiple rounds of curation with different lines of evidence resulted in two publicly available versions of the sequence, mPerLon1.0, and mPerLon1.1, each consisting of two partially phased assemblies (primary and alternate). Ninety percent of the primary assembly of mPerLon1.0 (2.11 Gb) was assigned to chromosomes by alignment to the chromosomal-level assembly of *P. l. pacificus*, with 93.2% average confidence in placement and 97.4% in orientation. This alignment revealed misassemblies in multiple regions (Supplementary Fig. 3), confirmed by further inspection of the Omni-C data. Therefore, we carried out additional curation, producing version mPerLon1.1 (Table 2). In total, we generated 368 joins

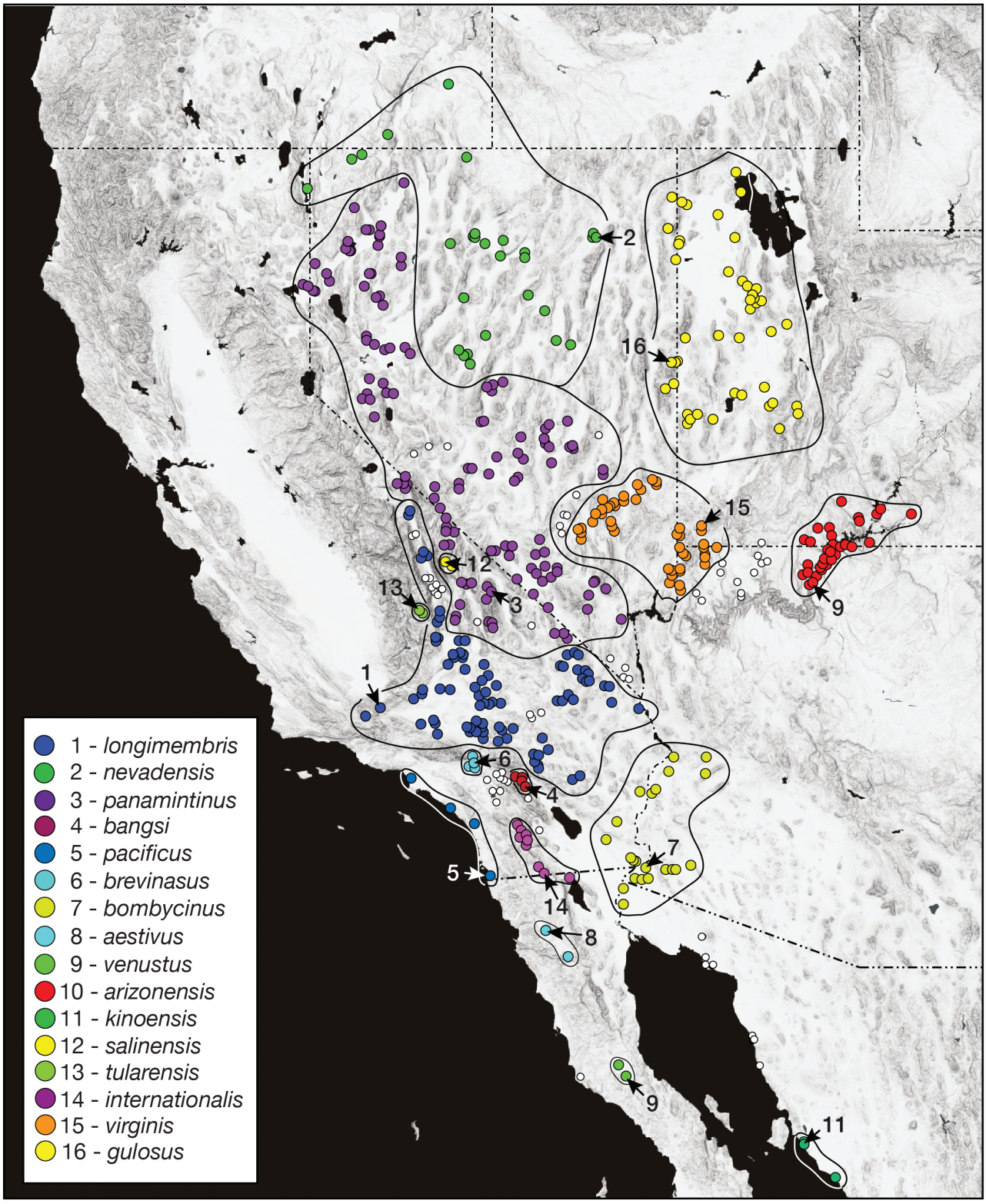

Fig. 2. Distribution of the subspecies of *P. longimembris*, based on a revision of museum collections by JLP. Stars indicate the provenance of the reference genome samples: (A, cyan) *P. longimembris longimembris* described here; (B, yellow) *P. l. pacificus* (GCF_023159225.1; Wilder et al. 2022).

(170 on the primary assembly and 198 on the alternate), 83 breaks (50 primary, 33 alternate), and closed 9 gaps (all on the alternate). Finally, we filtered out 29 contigs (16 from the primary and 13 from the alternate assembly) corresponding to mitochondrial contamination. The scaffold N50 increased from 25.3 Mb to 74.3 Mb (Table 2). We recommend the assembly mPerLon1.1.hap1 be used in future applications.

The final assembly, mPerLon1.1 consists of two partially phased assemblies that we will call primary and alternate throughout this paper, both similar in size to the value

Table 1. Software used in the assembly pipeline.

Assembly	Software and options[a]	Version
Filtering PacBio HiFi adapters	HiFiAdapterFilt	Commit 64d1c7b
K-mer counting	Meryl (k = 21)	1
Estimation of genome size and heterozygosity	GenomeScope	2
De novo *assembly (contiging)*	HiFiasm (Hi-C Mode, –primary, output p_ctg.hap1, p_ctg.hap2)	0.16.1-r375
Scaffolding		
Omni-C data alignment	Arima Genomics Mapping Pipeline	Commit 2e74ea4
Omni-C scaffolding	SALSA (-DNASE, -i 20, -p yes)	2
Gap closing	YAGCloser (-mins 2 -f 20 -mcc 2 -prt 0.25 -eft 0.2 -pld 0.2)	Commit 0e34c3b
Reference-guided scaffolding	RagTag (scaffold mode -m 1e6) minimap2	2.1.0 2.24
Visualization of GW alignment	dotPlotly (-m 1e5 -q 1e6) pafr	Commit 1174484 0.02
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
Genome quality assessment		
Basic assembly metrics	QUAST (--est-ref-size)	5.0.2
Assembly completeness	BUSCO (-m geno, -l glires)	5.0.0
	Merqury	2020-01-29
Contamination screening		
Local alignment tool	BLAST+	2.1
General contaminant screening	BlobToolKit	2.3.3
Annotation	**Software and options[a]**	**Version**
Finding tandem repeats	ULTRA	0.99.17
Modeling novel interspersed repeats	RepeatModeler2 (-LTRStruct)	2.0.3
Classification of interspersed repeats	DeepTE (-m M -prop_thr 0.8)	Commit babd65e950
Identification of protein sequences in repeats	transposonPSI	1.0
Collapsing redundant sequences	cd-hit-est	4.8.1
Masking repeatable sequence	RepeatMasker (-species Glires)	4.1.4
Annotation of repeatable sequence	bedtools maskfasta (-soft)	2.30.0
Lift-over of a gene annotation	Liftoff (-polish -copies -unplaced) LiftoffTools (clusters)	1.6.3 0.4.4

Software citations are listed in the text.
[a]Options detailed for non-default parameters.

estimated by Genomescope2.0 (Fig. 3A) and with comparable contiguity metrics. The two assemblies cannot be interpreted directly as the maternal and paternal sequence, and are expected to contain some switch errors (Cheng et al. 2021). The primary assembly (so designated based on completeness) comprises 982 scaffolds spanning 2.35 Gb with a contig N50 of 11.6 Mb, scaffold N50 of 74.3 Mb, longest contig of 61.9 Mb, and largest scaffold of 167.8 Mb. The alternate assembly consists of 577 scaffolds, spanning 2.26 Gb with a contig N50 of 10.1 Mb, scaffold N50 of 73.2 Mb, largest contig of 52.1 Mb, and largest scaffold of 164.1 Mb. Detailed assembly statistics of mPerLon1.1 are reported in Table 2, and the primary assembly is represented in Fig. 2B. For details of the alternate sequence and the earlier assembly mPerLon1.0 see Supplementary Figs. 1 and 2. The primary assembly has BUSCO completeness score of 93.8% using the *Glires* gene set, per-base quality (QV) of 55.99, k-mer completeness of 90.27% and frameshift indel QV of 39.79; while the alternate assembly has BUSCO completeness score of 91.2%, per-base

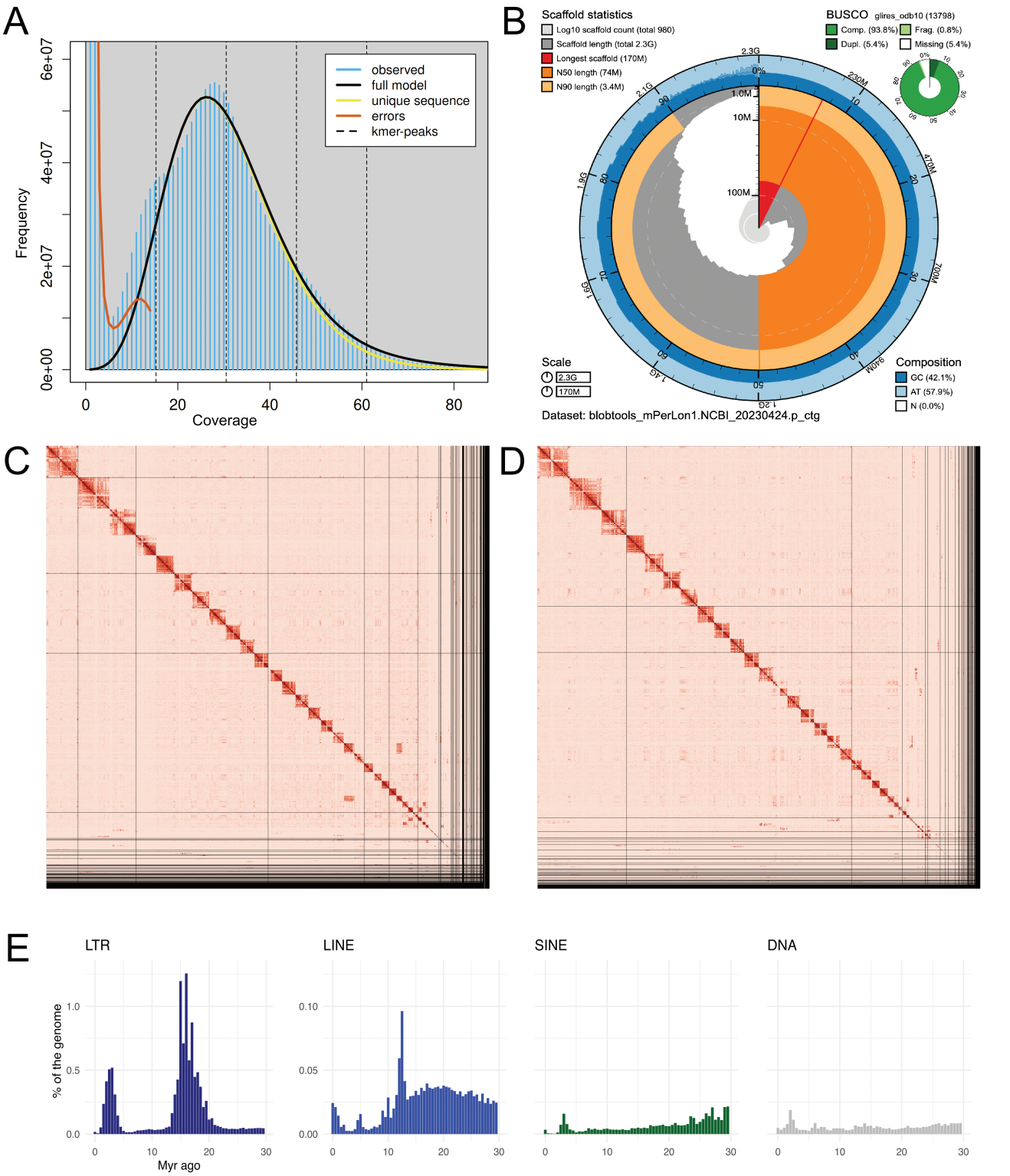

Fig. 3. Quality metrics and composition of the final *P. l. longimembris* primary genome assembly (mPerLon1.1). A) The k-mer spectrum of the adapter-trimmed HiFi data. The observed bimodal distribution (peaks at coverage of 16× and 28×) is expected for a diploid genome. The GenomeScope2.0 statistics are: uniq 67.8%, ab 0.001%, kcov 15.3%, err 0.214%, dup 3.83. B) Snail plot of quality metrics. The circumference represents the length of the assembly: scaffolds are drawn clockwise in order of size, and the red line indicates the longest. The middle arcs represent N50 (dark orange) and N90 (light orange). Completeness of the BUSCO core gene set assembly is shown in the top right panel. C) PretextSnapshot Omni-C contact maps of the primary and alternate (D) assemblies. Every cell represents data supporting linkage between genomic regions found in proximity in 3D space. E) The evolution of transposable element landscapes throughout the diversification history of pocket mice (Perognathinae). The *y* axis represents how much of the genome was derived from a specific TE family (Class I: LTR, LINE and SINE elements; Class II: DNA transposons) in each period of 500 Kyr since the divergence of *Perognathus* from its sister genus *Chaetodipus*. Notice the different scale for the accumulation of LTRs.

Table 2. Quality metrics of the final primary and the alternate assemblies mPerLon1.1.

	Primary	Alternate
Number of contigs	1,349	947
Contig N50 (bp)	11,606,741	10,098,643
Contig NG50[a]	12,053,573	10,098,643
Longest Contigs	61,996,900	52,133,366
Number of scaffolds	982	577
Scaffold N50	74,284,026	73,175,689
Scaffold NG50[a]	28,082,779	28,569,105
Largest scaffold	167,813,103	164,142,689
Size of final assembly	2,346,744,926	2,257,239,605
Phased block NG50[a]	13,333,580	11,084,738
Gaps per Gbp (# Gaps)	92 (215)	101 (229)
Indel QV (Frame shift)	41.56	41.67
Base pair QV	55.99	56.68
k-mer completeness	90.27	87.26
BUSCO		
Complete genes	93.8%	91.2%
Complete: single-copy	88.3%	85.4%
Complete: duplicated	5.5%	5.8%
Fragmented	0.8%	0.9%
Missing	5.4%	7.9%

[a]Read coverage and NGx statistics have been calculated based on the estimated genome size of 2.27 Gb.

quality (QV) of 56.68, k-mer completeness of 87.26% and frameshift indel QV of 39.91.

We assembled a partial mitochondrial genome with a length of 16,929 bp. The final assembly is composed of $A = 30.64\%$, $C = 17.45\%$, $G = 22.08\%$, $T = 29.83\%$, and encodes 22 unique transfer RNAs, two rRNAs and 13 polypeptides. The gene order in the mitochondrial genome is consistent with that found in most rodents.

Genome annotation

Repeat annotation based on the standard RepeatMasker and Dfam pipeline (Osmanski et al. 2023) identified 37.5% of the primary assembly as repetitive sequence (Supplementary Table 2). Additional searches, especially for non-Long Terminal Repeat (non-LTR) retrotransposons, increased the total to 44.9% (Supplementary Table 3). Since the divergence of *Perognathus* from *Chaetodipus* approximately 22 MYA (Upham et al. 2019) LTR-retrotransposon insertions have been the primary source of new sequence (Fig. 3E). The most widespread element is a relatively young (10.5% median divergence between copies) LTR of the Gypsy subtype, constituting 0.08% of the genome.

A large majority of the *P. l. pacificus* annotation was successfully transferred to our assembly by mapping, including 97.4% of the genes and 95.5% of all features. 183 CDS genes have multiple copies in both subspecies (2.70 copies on average). Of those, 82 have known functions and 22 belong to KEGG pathways associated with response to bacterial (Fold Enrichment 12.8) and viral infections (FE 13.9). The 681 genes that could not be transferred are not enriched for any specific pathway, but there is an overrepresentation of noncoding RNAs ($P < 0.0001$, binomial test). This is expected, as homology of RNAs is typically difficult to establish (Handzlik et al. 2020).

Discussion

We generated a high-quality genome assembly for *P. l. longimembris* and compared it to the recently published reference for *P. l. pacificus*. Our new assembly is more contiguous than its counterpart (982 vs. 6,180 scaffolds), although very similar scaffold N50 values show that the difference is merely a result of a broader distribution of scaffold length values for *P. l. pacificus*. Comparisons of high-quality genomes within a species can be useful for identifying recent patterns of genome evolution (Ferraj et al. 2023). The two genomes of the *P. longimembris* subspecies form a foundation for understanding differences between common and endangered subspecies of the little pocket mouse, as well as for discovering mechanisms of adaptation to arid conditions. To this end, our assembly will be annotated de novo through the NCBI Eukaryotic Annotation Pipeline (Thibaut-Nissen et al. 2016), based on RNAseq of four tissue types harvested from an individual caught at the same locality (MVZ:Mamm:240093).

Transposable elements may be a source of functionally important structural variation (Ferraj et al. 2023). Our stepwise annotation procedure demonstrates the utility of homology-based searches using a diverse set of data sources, as reliance on a single database may lead to under-masking (Supplementary Table 3). The increased diversity of reference sequences aided our homology searches because in rodents transposable element insertions appear highly dynamic at short evolutionary scales (Kapusta et al. 2017). This is apparent in *Perognathus*, where both LTRs and SINEs expanded in the late Miocene, after the origin of the genus (Patton and Fisher 2023), and followed by an expansion of LINEs in *P. longimembris* during the last 2 million years (Fig. 3E). The relative content of different transposable element classes in *P. longimembris* follows the general trends seen in other mammals (Osmanski et al. 2023), although the large proportion of recently accumulated LTRs is at odds with the usual expectation of high LINE content. Nevertheless, similarly ubiquitous LTRs have been recently documented in a few other rodents (Osmanski et al. 2023). Genomic hotspots of accumulation for retrotransposons, regulatory elements, and coding genes are conserved in mammals (Buckley et al. 2017). Hence our thorough annotations of these features will aid in understanding the evolution of genome architecture in *Perognathus*, and broadly in Heteromyidae, a group of ~100 species with only three contiguous genomes available at present (Harder et al. 2022; Wilder et al. 2022).

The genome assemblies for *P. longimembris* will be useful for evolutionary studies and for conservation. Pocket mice show many adaptations to living in arid environments, and a comparison of their genomes with those of species found in different environments may help identify the genetic basis of these adaptive traits. The ongoing analysis of genomic variation in multiple individuals from different geographic locations across the range of *P. longimembris* will be the next step in identifying the genetic basis of highly variable traits like pelage color and cranial morphology. These data will also form the basis for assessing the levels of genetic diversity in

common as well as threatened subspecies, and ultimately, for informing management decisions (Shaffer et al. 2022).

Supplementary Material

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

Acknowledgments

This paper is dedicated to the aridity-adapted astronauts: Fe, Fi, Fo, Fum, and Phooey. We thank the staff of the DNA Technologies and Expression Analysis Cores at the UC Davis Genome Center. Annotation of the draft genome was carried out on the Savio HPC at UC Berkeley. Phred Benham, José Cerca, and Daren Card provided advice on repeat annotation.

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). The UCD Cores were funded by the NIH Shared Instrumentation Grant 1S10OD010786-01.

Data Availability

Data generated in this study are available under NCBI BioProject PRJNA777209. Raw sequencing data for individual JLP 29072 (MVZ:Mamm:240093; NCBI BioSample SAMN29044251) are deposited in the NCBI Short Read Archive (SRA) under SRR20722010 for PacBio HiFi sequencing data, and SRR20722008 and SRR20722009 for the Omni-C Illumina sequencing data. GenBank accessions are GCA_024363575.2 (mPerLon1.1, primary) and GCA_024363435.2 (mPerLon1.1, alternate). The GenBank organelle genome assembly for the mitochondrial genome is GCA_024364775.1. Assembly scripts and other data for the analyses presented can be found at www.github.com/ccgproject/ccgp_assembly and https://github.com/evo-eco-gen/CCGP_Perognathus. De novo predicted repeatable element sequences are available on Dfam. Repeatable element masks, gene Liftoff annotation, and RagTag chromosomal assignment files can be found in the Dryad repository DOI: 10.5061/dryad.zs7h44jgc.

References

Altschuler EM, Nagle RB, Braun EJ, Lindstedt SL, Krutzsch PH. Morphological study of the desert heteromyid kidney with emphasis on the genus *Perognathus*. Anat Rec. 1979:194:461–468.

Bao W, Kojima KK, Kohany O. Repbase update, a database of repetitive elements in eukaryotic genomes. Mobile DNA. 2015:6:11.

Benson SB. Concealing coloration among some desert rodents. Univ Calif Publ Zool. 1933:40:1–70.

Buckley RM, Kortschak RD, Raison JM, Adelson DL. Similar evolutionary trajectories for retrotransposon accumulation in mammals. Genome Biol Evol. 2017:9:2336–2353.

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.

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.

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.

Comrack L. Species of special concern: a brief description of an important California department of fish and game designation. Sacramento, CA: California Department of Fish and Game, Wildlife Branch, Nongame Wildlife Program; 2008. Report No.: 2008-03. https://wildlife.ca.gov/Conservation/SSC [accessed 2023 1 April].

Ellinghaus D, Kurtz S, Willhoeft U. LTRharvest, an efficient and flexible software for de novo detection of LTR retrotransposons. BMC Bioinf. 2008:9:18.

Ferraj A, Audano PA, Balachandran P, Czechanski A, Flores JI, Radecki AA, Mosur V, Gordon DS, Walawalkar IA, Eichler EE, et al. Resolution of structural variation in diverse mouse genomes reveals chromatin remodeling due to transposable elements. Cell Genom. 2023:3:100291. [Preprint]. doi:10.1016/j.xgen.2023.100291

Flynn JM, Hubley R, Goubert C, Rosen J, Clark AG, Feschotte C, Smit AF. RepeatModeler2 for automated genomic discovery of transposable element families. Proc Natl Acad Sci USA. 2020:117:9451–9457.

Ghurye J, Pop M, Koren S, Bickhart D, Chin C-S. Scaffolding of long read assemblies using long range contact information. BMC Genom. 2017:18:527.

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.

Gurevich A, Saveliev V, Vyahhi N, Tesler G. QUAST: quality assessment tool for genome assemblies. Bioinformatics. 2013:29:1072–1075.

Hafner J. Heteromyidae. In: Handook of the mammals of the World. Vol. 6. Lagomorphs and rodents I. In: Wilson DE, Lacher Jr TE, Mittermeier RA, editors. Barcelona: Lynx Edicions; 2016.

Handzlik JE, Tastsoglou S, Vlachos IS, Hatzigeorgiou AG. Manatee: detection and quantification of small non-coding RNAs from next-generation sequencing data. Sci Rep. 2020:10.

Harder AM, Walden KKO, Marra NJ, Willoughby JR. High-quality reference genome for an arid-adapted mammal, the banner-tailed kangaroo rat (*Dipodomys spectabilis*). Genome Biol Evol. 2022:14:evac005.

Haymaker W, Look BC, Winter DI, Benton E, Cruty M. Project BIOCORE/M212/, a biological cosmic ray experiment - Procedures, summary, and conclusions. Physics. 1975. [accessed 2022 27 September]. https://www.semanticscholar.org/paper/Project-BIOCORE-%2FM212%2F%2C-a-biological-cosmic-ray-and-Haymaker-Look/e0a5120269c2eeab739f6b9bf192f06ba53da161

Jenkins SH, Breck SW. Differences in food hoarding among six species of heteromyid rodents. J Mammal. 1998:79:1221–1233.

Kapusta A, Suh A, Feschotte C. Dynamics of genome size evolution in birds and mammals. Proc Natl Acad Sci USA. 2017:114:E1460–E1469.

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. 2017:6:10. doi:10.1093/gigascience/gix085

Lancaster LT, Fuller ZL, Berger D, Barbour MA, Jentoft S, Wellenreuther M. Understanding climate change response in the age of genomics. J Anim Ecol. 2022:91:1056–1063.

Liao X, Hu K, Salhi A, Zou Y, Wang J, Gao X. MsRepDB: a comprehensive repetitive sequence database of over 80,000 species. Nucleic Acids Res. 2022:50:D236–D245.

Lok S, Paton TA, Wang Z, Kaur G, Walker S, Yuen RKC, Sung WWL, Whitney J, Buchanan JA, Trost B, et al. De novo genome and transcriptome assembly of the Canadian beaver (*Castor canadensis*). G3 Genes Genomes Genetics. 2017:7:755–773.

Manni M, Berkeley MR, Seppey M, Zdobnov EM. BUSCO: assessing genomic data quality and beyond. Curr Protoc. 2021:1:e323.

Marra NJ, Romero A, DeWoody JA. Natural selection and the genetic basis of osmoregulation in heteromyid rodents as revealed by RNA-seq. Mol Ecol. 2014:23:2699–2711.

Mirzabaev, A. Desertification. In: Climate change and land: an IPCC special report on climate change, desertification, land degradation, sustainable land management, food security, and greenhouse gas fluxes in terrestrial ecosystems. The Intergovernmental Panel on Climate Change; 2019. https://www.ipcc.ch/srccl [accessed 2023 1 April].

Osmanski AB, Paulat NS, Korstian J, Grimshaw JR, Halsey M, Sullivan KAM, Moreno-Santillán DD, Crookshanks C, Roberts J, Garcia C, et al; Zoonomia Consortium†. Insights into mammalian TE diversity through the curation of 248 genome assemblies. Science. 2023:380:eabn1430.

Ou S, Jiang N. Ltr_retriever: a highly accurate and sensitive program for identification of long terminal repeat retrotransposons. Plant Physiol. 2018:176:1410–1422.

Patton JL, Fisher RN. Taxonomic reassessment of the Little pocket mouse, *Perognathus longimembris* (Rodentia, Heteromyidae) of southern California and northern Baja California. Therya. 2023:14:131–160.

Penzkofer T, Jäger M, Figlerowicz M, Badge R, Mundlos S, Robinson PN, Zemojtel T. L1Base 2: more retrotransposition-active LINE-1s, more mammalian genomes. Nucleic Acids Res. 2017:45:D68–D73.

Ranallo-Benavidez TR, Jaron KS, Schatz MC. GenomeScope 2.0 and Smudgeplot for reference-free profiling of polyploid genomes. Nat Commun. 2020:11:1432.

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.

Shumate A, Salzberg SL. Liftoff: accurate mapping of gene annotations. Bioinformatics. 2021:37:1639–1643.

Shumate A, Salzberg S. LiftoffTools: a toolkit for comparing gene annotations mapped between genome assemblies. F1000Research. 2022:11:1230.

Sikes RS; Animal Care and Use Committee of the American Society of Mammalogists. 2016 Guidelines of the American Society of Mammalogists for the use of wild mammals in research and education. J Mammal. 2016:97:663–688.

Smit, AFA, Hubley, R, Green P. RepeatMasker Open-4.0; 2013. [accessed 2023 15 January]. http://www.repeatmasker.org

Storer J, Hubley R, Rosen J, Wheeler TJ, Smit AF. The Dfam community resource of transposable element families, sequence models, and genome annotations. Mobile DNA. 2021:12:2.

Thibaut-Nissen F, DiCuccio M. Hlavina W, Kimchi A, Kitts PA, Murphy TD, Pruitt KD, Souvorov A. The NCBI eukaryotic genome annotation pipeline. J Anim Sci. 2016:94:184.

Uliano-Silva M, Ferreira JGRN, Krasheninnikova K, Darwin Tree of Life Consortium, Formenti G, Abueg L, Torrance J, Myers EW, Durbin R, et al.. MitoHiFi: a python pipeline for mitochondrial genome assembly from PacBio High Fidelity reads. BMC Bioinformatics. 2023:24:288, doi:10.1186/s12859-023-05385-y.

Upham NS, Esselstyn JA, Jetz W. Inferring the mammal tree: species-level sets of phylogenies for questions in ecology, evolution, and conservation. PLoS Biol. 2019:17:e3000494.

Vassetzky NS, Kramerov DA. SINEBase: a database and tool for SINE analysis. Nucleic Acids Res. 2013:41:D83–D89.

Webster DB, Webster M. Auditory systems of Heteromyidae: functional morphology and evolution of the middle ear. J Morphol. 1975:146:343–376.

Wilder AP, Dudchenko O, Curry C, Korody M, Turbek SP, Daly M, Misuraca A, Wang G, Khan R, Weisz D, et al. A chromosome-length reference genome for the endangered Pacific Pocket Mouse reveals recent inbreeding in a historically large population. Genome Biol Evol. 2022:14:evac122.