Journal of Heredity, 2026, **117**, 347–356
https://doi.org/10.1093/jhered/esaf067
Advance access publication 17 September 2025
Genome Resources

American Genetic Association

OXFORD

Genome Resources

A highly contiguous genome assembly for the California vole, *Microtus californicus,* provides insight into phylogenetic relationships and patterns of synteny among voles

Krzysztof M. Kozak[1], Merly Escalona[2], Isaac J. Linn[1], Noravit Chumchim[3], Colin Fairbairn[4], Mohan P. A. Marimuthu[3], Oanh Nguyen[3], William Seligmann[4], Chris J. Conroy[1], James L. Patton[1], Rauri C. K. Bowie[1], 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, 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, CA 95064, United States

*Corresponding author: Museum of Vertebrate Zoology and Department of Integrative Biology, 3101 Valley Life Sciences Bldg., University of California, Berkeley, CA 94720, USA. Email: mnachman@berkeley.edu

Abstract

The California vole (*Microtus californicus*) is a small cricetid rodent and one of the 20 species of *Microtus* in North America and 60 worldwide. Several subspecies are listed as being of conservation concern in California, and one is federally protected. Here we present the first de novo genome assembly for the California vole, generated as a part of the California Conservation Genomics Project. The *M. californicus* genome was generated using a combination of PacBio HiFi long reads and Omni-C chromatin-proximity sequencing technology. Our high-quality genome is one of the most complete vole assemblies available, with a contig N50 of 49.8 Mb, scaffold N50 of 83.7 Mb, and BUSCO completeness score of 96.4%. Analysis of this genome together with genomes of closely related species revealed phylogenetic relationships and high levels of synteny among voles. The California vole genome provides an important new resource for comparative work across cricetid and muroid genomes. It will also serve as a reference for the analysis of within-species genetic diversity across widespread subspecies as well as more restricted populations of conservation concern.

Key words: California Conservation Genomics Project, CCGP, genomics, long-read assembly, reference genome, rodents

Introduction

Voles of the genus *Microtus* (Cricetidae, Arvicolinae) are a diverse group of rodents restricted to the Northern hemisphere with 60 species worldwide and 20 species in North America (ASM Mammal Diversity Database 2025), characterized by recent and rapid periods of speciation (Conroy and Cook 2000). The paleontological history of *Microtus* is extremely rich, and their distinctive tooth morphology has long been used as a tool to date stratigraphic layers (e.g. Krokhmal, Rekovets and Kovalchuk 2023). Voles are small herbivorous rodents typified by life histories with rapid onset of reproduction and large litter sizes (Hinton 1926; Gromov and Polyakov 1992; Tamarin 1985). Vole ecological variability has been key to numerous studies of basic mammalian biology such as population regulation and cycling (e.g. Krebs 1970; Lidicker 1985; Stenseth 1999; Johnsen, Devineau and Andreassen 2019), with significant practical consequences for agriculture because many species are considered pests of crops ranging from cereals to trees (Byers 1985). Species of *Microtus* are known to differ in their level of sociality, and these differences have served as the basis for studies on the genetic and neurological basis for complex social behaviors including the molecular mechanisms of affiliative behavior, parent-offspring interactions, and pair bonding (e.g. Young et al. 1999, 2014; Lim et al. 2004; Burkett et al. 2016; Tripp et al. 2021). The practical importance and rich natural history of the genus *Microtus* has inspired several efforts to assemble the genomes of multiple species, recently resulting in nearly chromosomal-level assemblies (Table 1).

The California vole (*Microtus californicus*), comprised of 17 subspecies, ranges from Oregon to Baja California, Mexico, and includes four California subspecies of special concern (California Department of Fish and Wildlife 2019), as well as the federally endangered subspecies *M. c. scirpensis* (the

Received on 7 May 2025; revised on 13 August 2025; accepted on 15 September 2025

Table 1. Whole-genome assemblies of different species of *Microtus*[a].

Species	Common name	Length (Gbp)	Scaffold N50 (Mb)	Number of scaffolds	Scaffold L50	Technology	Citation and Genbank ID
M. agrestis	Short-tailed field vole	2.03	13.3	1366	45	Illumina	Unpublished GCA_902806755.1
M. arvalis	Common vole	2.00	68.5	3552	5	Illumina	Gouy et al. 2024 GCA_044665705.1
M. californicus	California vole	2.3	83.7	350	10	PacBio, OmniC	This paper GCA_028537955.1
M. montanus	Montane vole	2.34	2.7	13 074	138	Illumina	Duckett et al. 2021 GCA_020392405.1
M. obscurus	Altai vole	2.29	19.9	2537	30	Illumina, ONT	Unpublished GCA_963919715.1
M. ochrogaster	Prairie vole	2.29	17.3	6450	42	Illumina	Zerbino et al. 2018 GCA_000317375.1
M. oregoni	Creeping vole	3.02	34.8	2147	28	PacBio, ONT	Couger et al. 2021 GCF_018167655.1
M. pennsylvanicus	Eastern meadow vole	2.37	125.3	180	8	PacBio, HiC	Rhie et al. 2021 GCA_037038515.1
M. richardsoni	Water vole	2.27	59.1	2586	14	Illumina	Duckett et al. 2021 GCA_019206885.2

[a]Taxonomy according to the ASM Mammal Diversity Database (2025).

Amargosa vole) (Fig. 1A) (Hall 1981; Cudworth and Koprowski 2010). Subspecies of conservation concern are threatened by habitat degradation in areas of water scarcity. The history of these subspecies would be informed by a better understanding of the history and phylogeography of the species as a whole. Although the phylogeography of *M. californicus* has been investigated with mtDNA (Conroy and Neuwald 2008) and regional ddRADseq data (Krohn et al. 2018; Lin et al. 2018), better genomic tools are necessary for a fuller understanding of the history of the species and for informing management decisions for conservation.

We report a high-quality genome assembly, produced as part of the California Conservation Genomics Project (CCGP, https://www.ccgproject.org/; Shaffer et al. 2022). This genome assembly will be a foundational resource for future population genetic studies on the unique ecology, biogeography, evolutionary history, phylogeny, and conservation of *M. californicus*.

Methods

Biological materials

The specimen used for the de novo genome was a male *M. c. kernensis* captured at Canebrake Creek, on the east end of the Kern River Valley in Kern County, California (35.7507, −118.11022) at an elevation of 1,016 m on 13 August 2020 by JLP. The specimen was preserved as a standard museum study skin and skull with tissues flash frozen in liquid nitrogen for genomic and transcriptomic use as part of the CCGP (catalog numbers MVZ:Mamm:240107, JLP29080). For more details, visit https://arctos.database.museum/guid/MVZ:Mamm:240107.

High molecular weight genomic DNA isolation

High molecular weight (HMW) genomic DNA was extracted from 72 mg of liver tissue using the Nanobind Tissue Big DNA kit as per the manufacturer's instructions [Pacific BioSciences (PacBio), Menlo Park, California]. The DNA purity was estimated using absorbance ratios (260/280 = 1.79 and 260/230 = 2.18) on the NanoDrop ND-1000 spectrophotometer. The final DNA yield (25 μg) was quantified using the Quantus Fluorometer (QuantiFluor ONE dsDNA Dye assay; Promega, Madison, Wisconsin). The size distribution of the HMW DNA was estimated using the Femto Pulse system (Agilent, Santa Clara, California), revealing that 85% of the fragments were 100 kilobases (kb) or longer.

HiFi library preparation and sequencing

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 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 DNA was concentrated using 0.45X of AMPure PB beads (PacBio, Cat. #100-265-900), followed by further enzymatic steps: single-strand overhang removal at 37°C for 15 min, 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 1 h. The SMRTbell library was purified and concentrated with 1X AMPure PB beads for nuclease treatment at 37°C for 30 min and size selection using the BluePippin/PippinHT system (Sage Science, Beverly, Massachusetts; Cat #BLF7510/HPE7510) to select 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, California) using five 8 M SMRT cells, 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, CA) according to the manufacturer's protocol with slight modifications. First, specimen tissue (liver, ID: JLP 29080) was thoroughly ground with a mortar and pestle while cooled with liquid nitrogen. Subsequently, chromatin was fixed in place in the nucleus. The suspended chromatin solution was then passed through 100 and 40 μm cell strainers to remove large debris. Fixed chromatin was digested

Fig. 1. (A) The range of *Microtus californicus* based on specimen records in the online Arctos museum database (https://arctos.database.museum/). The star indicates the geographic location of the individual from which the reference genome was created. Subspecies of conservation concern are indicated in yellow (*M. c. sanpabloensis*), dark blue (*M. c. vallicola*), black (*M. c. Scirpensis*), white (*M. c. mohavensis*), and light blue (*M. c. stephensi*); (B) *M. californicus* (photo credit: Ron Wolf); (C) habitat of the endangered Amargosa vole (*M. c. scirpensis*) in Shoshone, California (photo: Amargosa conservancy).

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, Massachusetts) 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, California) on an Illumina NovaSeq 6000 platform (Illumina, California) to generate approximately 100 million 2 × 150 bp read pairs per gigabase (Gb) of genome size. Accession numbers and NCBI BioProject numbers for all data are provided in Supplementary Table 1.

Nuclear genome assembly

We assembled the genome of *M. californicus* following the CCGP assembly pipeline Version 5.0, as outlined in Supplementary Table 2, which lists the tools and non-default parameters used in the assembly. First, we removed remnant adapter sequences from the PacBio HiFi dataset using HiFiAdapterFilt (Sim et al. 2022) and obtained the initial phased diploid assembly using HiFiasm (Cheng et al. 2022) on HiC mode with the filtered PacBio HiFi reads and the Omni-C dataset. We then aligned the Omni-C data to each haplotype following the Arima Genomics Mapping Pipeline (https://github.com/ArimaGenomics/mapping_pipeline) and then scaffolded them using SALSA (Ghurye et al. 2017; Ghurye et al. 2019).

Both haplotypes were manually curated by iteratively generating and analyzing their corresponding Omni-C contact maps. To generate the contact maps we aligned the Omni-C data with BWA-MEM (Li 2013), identified ligation junctions, and generated Omni-C pairs using pairtools (Golobordko et al. 2018). We generated a multi-resolution Omni-C matrix with cooler (Abdennur and Mirny 2020) and balanced it 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 where we identified misassembles and misjoins and finally 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 curation) were closed using the PacBio HiFi reads and YAGCloser (https://github.com/merlyescalona/yagcloser). Finally, we checked for contamination using the BlobToolKit Framework (Challis et al. 2020).

Genome size estimation and quality assessment

We trimmed bridge sequences from the Omni-C sequencing data with cutadapt (Martin 2011) to get k-mer counts using meryl (https://github.com/marbl/meryl). We also generated k-mer counts from the PacBio HiFi reads. Both k-mer databases

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 completeness we used BUSCO (Manni et al. 2021) with the Glires ortholog database (glires_odb10), which contains 13,798 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 the 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 (quality value); C = % genome represented by the first 'n' scaffolds, following a known karyotype of $2n$ =54 for this species (The Animal Chromosome Count database—V1.0.0; https://cromanpa94.github.io/ACC/) . Quality metrics for the notation were calculated on the primary haplotype sequence.

Mitochondrial genome assembly

The protocol used to produce the HiFi library used the TPK2 enzyme, which digests mitochondrial DNA (Carmen Guarco, *pers. comm.*). As a result, no mitochondrial sequence was recovered from the PacBio reads. To generate a mitochondrial genome sequence, we generated short-read Illumina sequence data from an additional individual. Specifically, DNA was extracted from one *M. californicus kernensis* (MVZ 224987) collected from the China Lake Naval Air Weapons Station in Kern County, California. This sample will be part of future population genomic studies. Extraction, library preparation, and paired-end sequencing on an Illumina HiSeq 6000 instrument were carried out using the standard CCGP/MVZ protocol (Benham et al. 2024). We used NOVOPlasty (Dierckxsens et al. 2016) for assembly using a cyt-b seed from one specimen (NCBI:MG916867; MVZ 233046; Lin et al. 2018) and used MITOS2 in Galaxy for annotation (Al Arab et al. 2017; Donath et al. 2019; The Galaxy Community 2024).

Demographic estimate

Historic population sizes were estimated with the standard PSMC pipeline (https://github.com/lh3/psmc): briefly, we mapped reads from the individual MVZ 224987 to our reference genome with BWA (Li, 2013) and genotyped with bcftools mpileup (https://samtools.github.io/bcftools/bcftools.html). After masking the repetitive regions, we ran PSMC with 30 iterations and 100 bootstraps. To avoid a common artefactual peak close to the present (Hilgers et al., 2025), we treated the last four intervals of time as independent ($-p$ "$1+1+1+1+25*2+4+6$"). Six generations per year and a mutation rate of 5×10^{-9} were assumed (Ullrich and Tautz, 2020).

Phylogenomics of Microtini

The phylogeny and taxonomy of *Microtus* and the tribe Microtini have been revised extensively yet remain unresolved (Conroy and Cook 2000; Wang et al. 2022; Withnell and Scarpetta 2024). We took advantage of our newly assembled *M. californicus* genome, together with other recently generated whole-genome sequences, to revisit the phylogeny of *Microtus* and closely related species. We focused on genomes sequenced with paired-end Illumina reads to at least 4× mean depth of coverage. We incorporated 21 individuals across 15 species of *Microtus* sensu lato (including *Alexandromys, Lasiopodomys, Lemmiscus,* and *Neodon*), and nine individuals in five genera of other Arvicolinae rodents and used *Rattus norvegicus* and *Mus musculus* as outgroups (Supplementary Table 3). Orthology groups expected to be found in at least 80% of rodents were selected with OMA (Altenhoff et al. 2024) and used as references to recover and align 500 protein-coding genes from raw reads, following the Read2Tree pipeline (Dylus et al. 2024). Gene trees were inferred in IQTREE v2 (Minh et al., 2020) with 1,000 bootstrap replicates, and the species tree was estimated under the multispecies coalescent in weighted ASTRAL (Zhang and Mirarab 2022).

Genomic synteny

To explore variation in chromosome organization, we studied patterns of synteny across a subset of arvicoline rodents for which there were contiguous whole-genome assemblies using *M. musculus* (GRCm39) as an outgroup (Supplementary Table 4). We also conducted a separate analysis comparing only the *M. musculus* and *M. californicus* assemblies. We used Mash (Ondov et al. 2016) and estimated the mean sequence divergence between *M. californicus* and *M. musculus* to be 14.78%. Synteny blocks were estimated using ntSynt (Coombe et al. 2024) with the default parameters for taxa, which are more than 10% diverged. Synteny blocks were then plotted using ntSynt-viz (Coombe et al. 2025) after removing scaffolds less than 5 Mb in length.

Results

Sequencing data

The PacBio HiFi sequencing generated 7.39 million reads yielding ∼36-fold coverage based on the Genomescope 2.0 genome size estimation of 2.34 Gb (N50 read length 11,598 bp; minimum read length 114 bp; mean read length 11,521 bp; maximum read length 46,854 bp). The Omni-C sequencing generated 640.23 million read pairs, with an estimated sequencing error rate of 0.331% (Fig. 2A). Because the model fitting of GenomeScope with the HiFi data failed to converge (Fig. 2B), we report the k-mer spectrum based on the Omni-C short reads, which show a bimodal distribution with two major peaks at ∼28- and ∼56-fold coverage (Fig. 2A), where the peaks correspond to heterozygous and homozygous states of a diploid species, respectively.

Assembly metrics

The final assembly (mMicCal1) consists of two partially phased haplotypes that vary slightly in size compared with the estimated value from GenomeScope2.0 (Fig. 2A), as has been observed in other taxa (e.g. Pflug et al. 2020). The primary assembly (mMicCal1.0.hap1) consists of 401 scaffolds spanning 2.29 Gb with a contig N50 of 49.8 megabases (Mb), a scaffold N50 of 83.7 Mb and with a largest scaffold length of 196 Mb. Metrics for the alternate assembly were comparable (Table 2).

During manual curation, we generated a total of seven breaks and 62 joins; three breaks were made on the primary assembly, and four breaks were made on the alternate assembly. Of the 62 joins, 42 were made on the primary assembly and 20 on the alternate assembly. We were able to close a total of six gaps, three on each assembly.

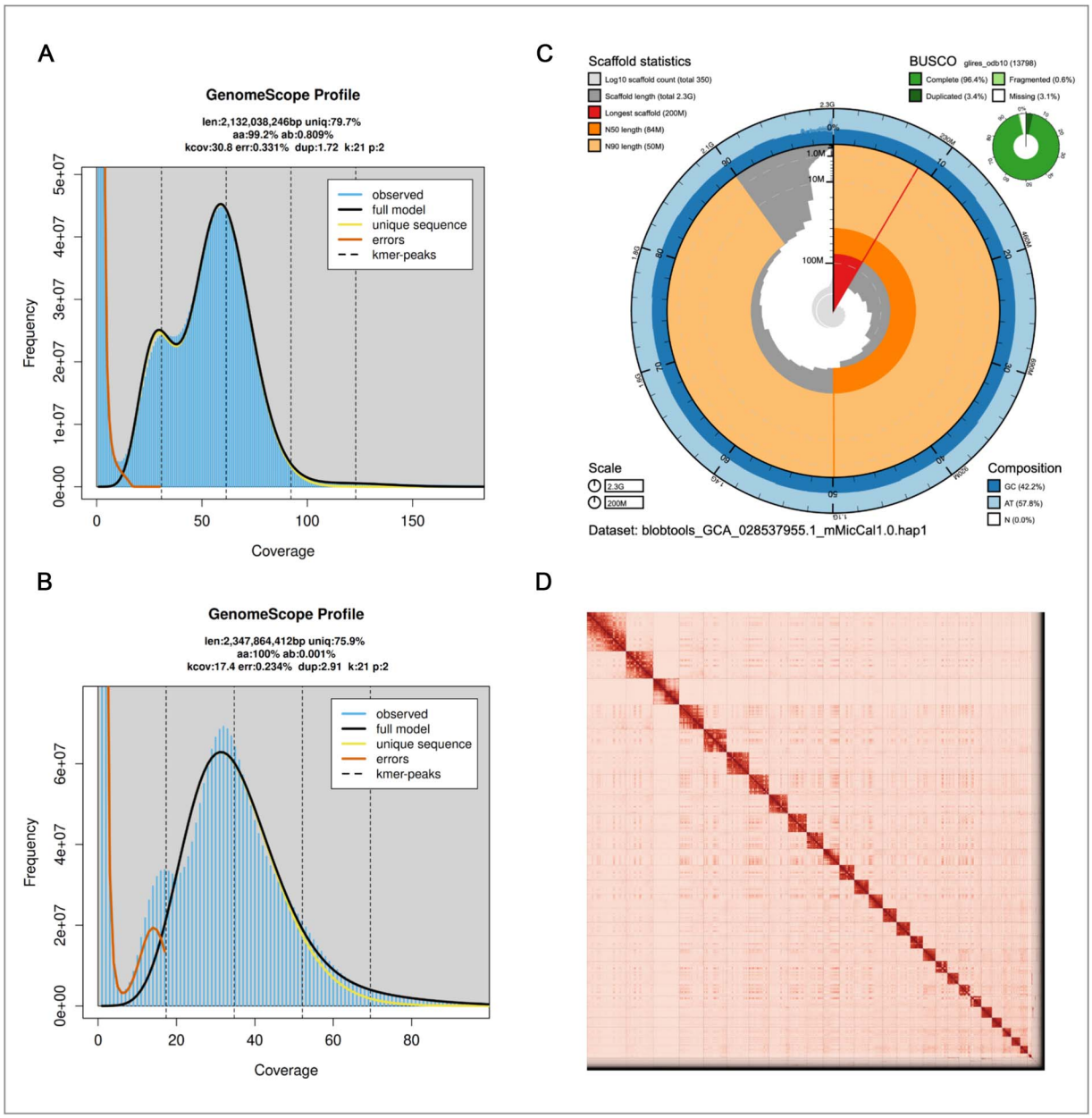

Fig. 2. Quality metrics of the raw data and the final *Microtus californicus* primary genome assembly (mMicCal1). A) The k-mer spectrum of the Omni-C Illumina data used in scaffolding. The bimodal distribution is expected for a diploid genome. B) The k-mer spectrum of the adapter-trimmed HiFi data. C) BlobToolKit Snail plot genome assembly summary. The central circle represents the length and number of scaffolds: the circumference of the circle represents the entire length of the genome and scaffolds are added in order of length starting from the longest scaffold to the shortest. The outer circle represents GC and AT content. BUSCO score is given in the upper right corner and length-related statistics are at the top left. D) The PretextSnapshot Omni-C contact map shows support for linkage between genomic regions of the primary assembly.

The primary assembly had a BUSCO completeness score of 96.4% using the Glires gene set, indicating a highly complete genome. The Omni-C contact maps shows that both assemblies are highly contiguous with some chromosome-length scaffolds (Figs 2C and 2D). Table 2 provides additional metrics for both assemblies.

Mitochondrial genome

The mitochondrial genome size was 16,296 bp, with 22 unique transfer RNAs and 13 protein-coding genes, consistent with expectations. The sequence had more than 14 000x coverage and BLASTs with 99.3% ID to an existing *M. californicus* mitochondrial genome (SAMN13341630).

Phylogenomics of Microtini

A phylogenetic tree depicting relationships of eight *Microtus* species, as well as closely related species for which whole-genome sequences were available, is shown in Fig. 3A. This tree has strong support, with greater than 80% local posterior probabilities on all but one of the nodes. The microtines (*Alexandromys, Lasiopodomys, Microtus*, *Neodon*, and *Stenocranius*) form a well-supported clade, and the genera within this clade are monophyletic, in agreement with earlier studies (e.g. Withnell and Scarpetta 2024). *Lemmiscus* is the sister-group to the microtine clade, and *Arvicola* is the sister group to the clade containing *Lemmiscus* and the microtines. The representatives of the Clethriomyini form a monophyletic

Table 2. Metrics for the primary and alternate assemblies of the California vole (*Microtus californicus*) genome.

Metric	Primary	Alternate
Number of contigs	401	353
Contig N50 (bp)	49,752,295	45,668,671
Contig NG50§	49,254,385	43,814,567
Longest contig	105,859,453	94,920,441
Number of Scaffolds	350	310
Largest scaffold	195,964,630	195,394,907
Scaffold N50	83,674,487	76,329,214
Scaffold NG50§	83,674,487	72,442,157
Phased block NG50 §	49,254,385	43,814,567
Indel QV (Frame shift)	42.09974342	42.83602717
Size of final assembly	2,299,490,594	2,142,917,481
Gaps per Gbp (#Gaps)	22(51)	20(43)
BUSCO		
Complete genes	96.40%	93.80%
Complete: single copy	93.00%	91.10%
Complete: duplicated	3.40%	2.70%
Fragmented	0.60%	0.60%
Missing	3.00%	5.60%
Base pair QV	62.65	62.65
k-mer completeness	91.71	86.95

§ Read coverage and NGx statistics have been calculated based on the estimated genome size of 2.34 Gb.

sister clade to the Microtine-*Arvicola-Lemmiscus* clade. *Ellobius* is the sister group to the other arvicolines. Most relationships agreed with previous work, but some were at odds with earlier mtDNA and some nuclear sequence studies. For example, Abramson et al. (2021), Withnell and Scarpetta (2024), and Abramson et al. (2025) found *Ellobius* to group with the Microtini, whereas we found *Ellobius* to be the sister lineage to the other arvicolines.

Demography

We estimated changes in historical effective population size using PSMC and observed two episodes of population growth and decline. The first episode started approximately 150 Kya and peaked 40 Kya at N_e of 160,000 to 190,000 followed by a steady decline almost to the original level at 15 Kya. A second episode of growth had a slightly lower peak (Ne of 130,000 to 180 000) 7 Kya and subsequently declined to the current population size of approximately 35,000 (Fig. 3B).

Patterns of synteny

We found considerable synteny in comparisons involving *M. musculus* and four representatives of Arvicolinae (Fig. 3C). Large portions of whole chromosomes were collinear in all samples, such as half of *M. musculus* chromosome 2. Synteny was also observed within the Arvicolinae for blocks that are rearranged relative to *M. musculus*. Although some whole-arm translocations distinguish *Microtus arvalis* and *M. californicus*, most of the chromosome organization appears conserved between the two species. Sex chromosomes were not present in all assemblies, but the third largest scaffold of *M. californicus* had conserved synteny with the *M. musculus* X chromosome (Supplementary Table 5, Supplementary Fig. 1).

Discussion

Here we present a highly contiguous genome assembly for the California vole, *M. californicus*. The primary assembly has a total length of 2.30 Gb, a scaffold N50 of 83.7 Mb, 350 scaffolds, and a BUSCO score of 96.4%. The assembly contains several scaffolds of length similar to that of full chromosomes in the most complete rodent assemblies (maximum length 105.85 Mb). The *M. californicus* genome joins seven other species of *Microtus* with complete genome sequences (Table 1). The length of the *M. californicus* genome is not remarkably large or small in relation to the other genomes. The quality is among the highest in terms of both the scaffold N50 and the number of scaffolds.

The presence of multiple whole-genome sequences of arvicoline rodents enabled us to revisit their phylogenetic relationships. One persistent problem with arvicoline systematics has been the rapid pulses of speciation, which have led to short branch lengths in phylogenies making it difficult to find clear branching order among genera and species (Conroy and Cook 1999). The whole-genome sequences analyzed here are inherently strong in the depth of data, but weak in terms of taxonomic coverage. Thus, higher order relationships may be in disagreement due to insufficient taxonomic sampling. Nonetheless, our phylogeny is consistent with some previously estimated relationships. For example, North American *Microtus* appear monophyletic, as supported elsewhere (Conroy and Cook 2000; Jaarola et al. 2004; Steppan and Schenk 2017). *Microtus* sensu lato (including *Lasiopodomys*, *Neodon*, *Stenocranius*, and *Alexandromys*) are also monophyletic (Steppan and Schenk 2017; Withnell and Scarpetta 2024). In contrast, *Lemmiscus* is a sister lineage to *Microtus* sensu lato (Fig. 3), but this is not a universally held relationship. Using two mitochondrial and three nuclear genes, Withnell and Scarpetta (2024) found *Lemmiscus* to be weakly supported in a more basal position in the arvicoline tree. Using mitochondrial sequences, Abramson et al. (2021) found that *Lemmiscus* was placed as the sister group to the snow voles, *Chionomys*. The representatives of the Clethryomyini are monophyletic and somewhat basal in the tree, agreeing with most other large, recent studies (Steppan & Schenk 2017; Abramson et al. 2021, 2025; Withnell and Scarpetta 2024). *Ellobius lutescens*, the mole vole, is the earliest branching lineage in our tree. This conflicts with other recent broad studies, which suggest it is more closely related to *Arvicola* or the *Microtus* group (Steppan & Schenk 2017; Abramson et al. 2021, 2025, 2025; Withnell and Scarpetta 2024). Whole-genome sequences from additional taxa will likely help resolve some of these issues.

The genome presented here, together with other vole genomes, will be helpful for addressing many aspects of vole biology. For example, most arvicoline rodents possess hypsodont (high crowned) continuously growing molars. These have allowed them to exploit primarily grass diets and persist where rodents with omnivorous diets and brachydont (low crowned) molars were at a disadvantage. Some lineages of arvicolines have an intermediate condition where molars are high crowned but rooted. Full-genome sequences will enable studies of the genomic changes that occurred to allow for changes in the rooted condition (e.g. Calamari et al. 2024), and possibly concomitant adaptations in the gut (Wagner et al. 2022) or related aspects of morphology or physiology.

The genus *Microtus* is also characterized by extensive chromosomal variability, both within species (e.g. Gill 1982) and between species (Modi 1987). Comparisons across arvicolines have been used to estimate rates of karyotypic evolution and to identify the order of structural rearrangements

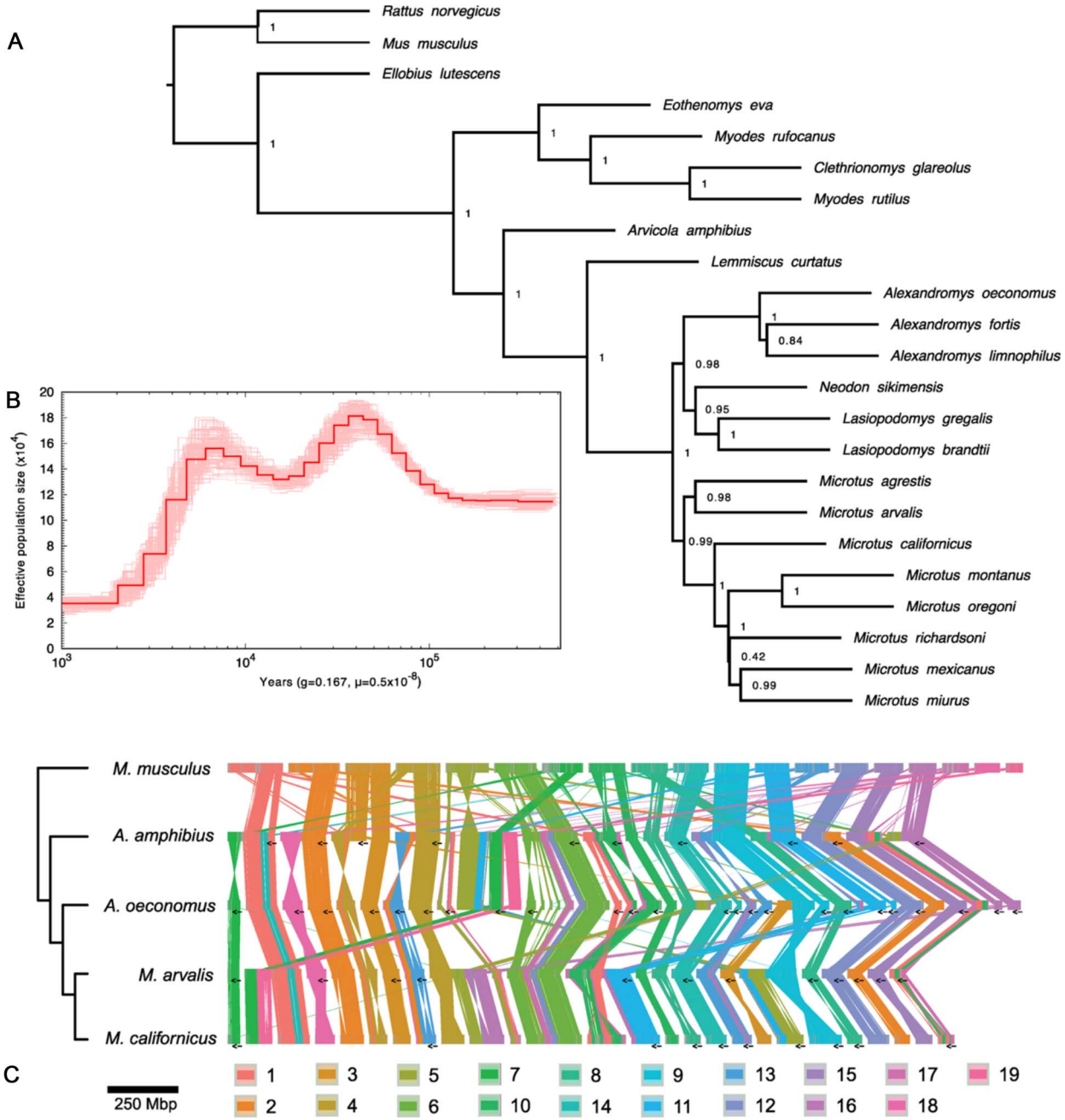

Fig. 3. A. Multispecies coalescent phylogeny of Microtini based on 500 CDS genes. The node labels indicate local posterior probabilities. B. Demographic curve for one individual with 100 bootstraps, estimated using PSMC. C. Blocks of conserved synteny across Arvicolinae using *M. musculus* as an outgroup. Ribbons and synteny blocks across scaffolds are colored based on the chromosome of the corresponding synteny block in the GRCm39 *M. musculus* assembly. The arrows indicate scaffolds where the reverse complement has been plotted, which is normalized to the + strand of the *M. musculus* assembly. Details of assemblies used can be found in Supplementary Table 4.

and ancestral karyotypes (Sitnikova et al. 2007, Lemskaya et al. 2010). We used highly contiguous assemblies to identify blocks of synteny across Arvicolinae genomes, recovering some of the homologous relationships identified by previous genome assemblies and by cytogenetic techniques. For instance, a segment homologous to *M. musculus* chromosomes 1 and 14, which we find as a collinear block in four arvicoline rodents, has been previously identified in both *A. oeconomus* and *M. arvalis* (Sitnikova et al. 2007; Gouy et al. 2024). A growing body of highly contiguous whole-genome assemblies will be useful for investigating the molecular details of karyotypic evolution over deeper timescales.

M. californicus consists of two well-differentiated groups isolated by hybrid male sterility (Gill 1980; Conroy and Neuwald 2008; Lin et al. 2018). The genome presented here will serve as an essential reference for subsequent population genomic studies of differentiation and potential gene flow between these groups.

Although *M. californicus* is widespread and often abundant, there are several populations of conservation concern. The Amargosa vole (*M. c. scirpensis*) is an isolated population at a desert spring and appears to be the result of ancient contact between two divergent lineages (Conroy et al. 2016; Patton and Conroy 2017; Krohn et al. 2018). Population

genomic studies of this and other threatened populations may help inform management decisions. Recent conservation work with the Amargosa vole has included restoring habitat as well as temporarily rescuing individuals in a captive population at UC Davis (Foley et al. 2020). Genomic tools would be useful for monitoring inbreeding and the potential effects of translocating individuals. Other populations of concern are the Mojave river vole (*M. c. mohavensis*), the Owens Valley vole (*M. c. vallicola*), the San Pablo vole (*M. c. sanpabloensis*), and the south coast marsh vole (*M. c. stephensi*). These populations are relatively unstudied but are of concern due to possible loss of habitat and genetic diversity, leading to poor health and low numbers. Genomic data will be useful for improved taxonomic assessments as well as for guiding management decisions.

For *M. californicus*, a genomic toolkit may allow better understanding of genes under selection in bottlenecked populations in remote habitats. As a foundation for this work, we estimated historic effective population sizes (Fig. 3B). The observed periods of growth coincide with deglacial periods when wet grasslands expanded in California (Cowart, 2014), whereas the recent decline may be related to the mid-Holocene altithermal period of drying (Palmer et al., 2023). However, the estimated dates should be interpreted with caution because they are sensitive to estimates of mutation rate and generation length, both of which are not well known. More generally, the reference genome provided here will enable the analysis of whole-genome sequences from across the range of *M. californicus*, including populations in diverse habitats, elevation gradients, and climate conditions.

Acknowledgments

We thank the staff of the DNA Technologies and Expression Analysis Cores at the UC Davis Genome Center. We also thank Lydia Smith and Michelle Davila for lab work. Annotation of the draft genome was carried out on the Savio HPC at UC Berkeley.

Author contributions

Krzysztof M. Kozak (Data curation, Formal analysis, Writing—review & editing), Merly Escalona (Data curation, Formal analysis, Investigation, Methodology, Project administration, Writing—original draft, Writing—review & editing), Isaac J. Linn (Formal analysis, Investigation, Methodology, Writing—review & editing), William Seligmann (Investigation, Methodology, Writing—review & editing), Colin W. Fairbairn (Investigation, Methodology, Writing—review & editing), Oanh Nguyen (Investigation, Methodology, Writing—review & editing), Mohan P.A. Marimuthu (Investigation, Methodology), Christopher J. Conroy (Investigation, Resources, Writing—review & editing), James L. Patton (Resources, Writing—review & editing), Rauri C.K. Bowie (Conceptualization, Funding acquisition, Resources, Writing—review & editing), and Michael W. Nachman (Conceptualization, Funding acquisition, Resources, Supervision, 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); and a National Institutes of Health Shared Instrumentation Grant 1S10OD010786-01 supported the work of the UCD laboratories.

Data availability

The museum voucher of the sequenced animal is MVZ:Mamm:240107. Data generated for this study are available through NCBI under BioProject PRJNA765631, BioSample SAMN31837211. Raw sequencing reads are SRX19356638 (PacBio HiFi), and SRX19356639 and SRX19356640 (Omni-C Illumina). GenBank accessions for the assemblies are GCA_028537755.1 (primary) and GCA_028537955.1 (alternate); mitochondrial genome accession is (pending acceptance by JoH). Assembly scripts: https://github.com/ccgproject/ccgp_assembly. Scripts for post-assembly analyses: https://github.com/evo-eco-gen/Microtus.

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

Abramson NI, Bodrov SY, Bondareva OV, Genelt-Yanovskiy EA, Petrova TV. A mitochondrial genome phylogeny of voles and lemmings (Rodentia: Arvicolinae): evolutionary and taxonomic implications. *PLoS One*. 2021;**16**:e0248198. https://doi.org/10.1371/journal.pone.0248198

Abramson N, Skalon E, Bondareva O, Bodrov S, Petrova T, Dvoyashov I. Resolving rapid radiation of voles and lemmings (Arvicolinae: Cricetinae, Rodentia) with QuaddRAD sequencing and transcriptome analysis. *Diversity*. 2025;**17**:61. https://doi.org/10.3390/d17010061

Al Arab M, C H z S, Tout K, Sahyoun AH, Stadler PF, Bernt M. Accurate annotation of protein-coding genes in mitochondrial genomes. *Mol Phylogenetics Evol*. 2017;**106**:209–216. https://doi.org/10.1016/j.ympev.2016.09.024

Altenhoff AM, Warwick Vesztrocy A, Bernard C, Train CM, Nicheperovich A, Prieto Baños S, Julca I, Moi D, Nevers Y, Majidian S, et al. OMA orthology in 2024: improved prokaryote coverage, ancestral and extant GO enrichment, a revamped synteny viewer and more in the OMA ecosystem. *Nucleic Acids Res*. 2024;**52**:D513–D521. https://doi.org/10.1093/nar/gkad1020

Benham PM, Cicero C, Escalona M, Beraut E, Fairbairn C, Marimuthu MP, Nguyen O, Sahasrabudhe R, King BL, Thomas WK, et al. Remarkably high repeat content in the genomes of sparrows: the importance of genome assembly completeness for transposable element discovery. *GBE*. 2024;**16**:evae067

Burkett JP, Andari E, Johnson ZV, Curry DC, de Waal FB, Young LJ. Oxytocin-dependent consolation behavior in rodents. *Science*. 2016;**351**:375–378. https://doi.org/10.1126/science.aac4785

Byers RE. Management and control. In: Tamarin RH, ed. *Biology of New World Microtus*. Special Publication #8, American Society of Mammalogists; Stillwater, Oklahoma. 1985.

California Department of Fish and Wildlife (CDFW). Natural Diversity Database. *Special Animals List Periodic publication*. 2019;67

Calamari ZT, Song A, Cohen E, Akter M, Roy RD, Hallikas O, Christensen MM, Li P, Marangoni P, Jernvall J, et al. Vole genomics links determinate and indeterminate growth of teeth. *BMC Genomics* 2024;**25**:1000. https://doi.org/10.1186/s12864-024-10901-2

Cowart A. *Paleoenvironmental change in Central California in the late Pleistocene and Holocene: impacts of climate change and human land use on vegetation and fire regimes*. Berkeley, CA: University of California Berkeley; 2014.

Challis R, Richards E, Rajan J, Cochrane G, Blaxter M. BlobToolKit—Interactive quality assessment of genome assemblies. *G3* 2020;**10**:1361–1374. https://doi.org/10.1534/g3.119.400908

Cheng H, Jarvis ED, Fedrigo O, Koepfli KP, Urban L, Gemmell NJ, Li H. Haplotype-resolved assembly of diploid genomes without parental

data. *Nature Biotech*. 2022;**40**:1332–1335. https://doi.org/10.1038/s41587-022-01261-x

Conroy CJ, Cook JA. MtDNA evidence for repeated pulses of speciation within arvicoline and murid rodents. *J Mamm Evol*. 1999;**6**:221–245. https://doi.org/10.1023/a:1020561623890

Conroy CJ, Cook JA. Molecular systematics of a holarctic rodent (*Microtus*: Muridae). *J Mammal*. 2000;**81**:344–359. https://doi.org/10.1093/jmammal/81.2.344

Conroy CJ, Neuwald JL. Phylogeographic study of the California vole, Microtus californicus. *J Mammal*. 2008;**89**:755–767. https://doi.org/10.1644/07-mamm-a-189r1.1

Conroy CJ, Patton JL, Lim MCW, Phuong MA, Parmenter B, Höhna S. Following the rivers: historical reconstruction of California voles *Microtus californicus* (Rodentia: Cricetidae) in the deserts of eastern California. *Biol J Linn Soc*. 2016;**119**:80–98. https://doi.org/10.1111/bij.12808

Coombe L, Kazemi P, Wong K, Birol I, Warren RL. Multi-genome synteny detection using minimizer graph mappings. *bioRxiv*. 2024; https://doi.org/10.1101/2024.02.07.579356

Coombe L, Warren RL, Birol I. ntSynt-viz: visualizing synteny patterns across multiple genomes. *bioRxiv*. 2025; https://doi.org/10.1101/2025.01.15.633221

Couger MB, Roy SW, Anderson N, Gozashti L, Pirro S, Millward LS, Kim M, Kilburn D, Liu KJ, Wilson TM, et al. Sex chromosome transformation and the origin of a male-specific X chromosome in the creeping vole. *Science*. 2021;**372**:592–600. https://doi.org/10.1126/science.abg7019

Cudworth NL, Koprowski JL. *Microtus californicus* (Rodentia: Cricetidae). *Mamm Species*. 2010;**42**:230–243. https://doi.org/10.1644/868.1

Dierckxsens N, Mardulyn P, Smits G. NOVOPlasty: de novo assembly of organelle genomes from whole genome data. *Nucleic acids research* 2017; **45**:e18-e18.

Donath A, Jühling F, Al-Arab M, Bernhart SH, Reinhardt F, Stadler PF, Middendorf M, Bernt M. Improved annotation of protein-coding genes boundaries in metazoan mitochondrial genomes. *Nucleic Acids Res*. 2019;**47**:10543–10552. https://doi.org/10.1093/nar/gkz833

Duckett DJ, Sullivan J, Pirro S, Carstens BC. Genomic resources for the North American water vole (Microtus richardsoni) and the montane vole (Microtus montanus). *Gigabyte*. 2021; **2021**:gigabyte19. https://doi.org/10.46471/gigabyte.19

Dylus D, Altenhoff A, Majidian S, Sedlazeck FJ, Dessimoz C. Inference of phylogenetic trees directly from raw sequencing reads using Read2Tree. *Nat Biotechnol*. 2024;**42**:139–147. https://doi.org/10.1038/s41587-023-01753-4

Foley J, Allan N, Pesapane R, Johnson A, Woods L, Brignolo L, Luce A, Clifford DL, Imai DM. Disease and pathological conditions of an endangered rodent, Microtus californicus scirpensis, in a captive-rearing facility and in the wild. *J Zoo Wildl Med*. 2020;**50**:758–768. https://doi.org/10.1638/2018-0117

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

Gill AE. Partial reproductive isolation of subspecies of the California vole, Microtus californicus. *Genetica*. 1980;**52**:105–117. https://doi.org/10.1007/bf00121821

Gill AE. Variability in the karyotype of the California vole, *Microtus californicus*. *Mammalian Chromosomes Newsletter*. 1982;**23**:18

Goloborodko A, Abdennur N, Venev S, Brandao HB, Fudenberg G. *mirnylab/pairtools: v0.2.0*. 2018. https://doi.org/10.5281/zenodo.1490831

Gouy A, Wang X, Kapopoulou A, Neuenschwander S, Schmid E, Excoffier L, Heckel G. Genomes of Microtus rodents highlight the importance of olfactory and immune systems in their fast radiation. *Genome Biol Evol*. 2024;**16**:evae233. https://doi.org/10.1093/gbe/evae233

Gromov IM, Polyakov IY. Fauna of the USSR. In: Skarlato OA. (ed). *Mammals. Voles (Microtinae)*. Leningrad: Nauka. [English translation].; 1992.

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

Hall ER. *The Mammals of North America*. 2nd ed. New York (NY): John Wiley & Sons, Inc; 1981.

Hilgers L, Liu S, Jensen A, Brown T, Cousins T, Schweiger R, Guschanski K, Hiller M. Avoidable false PSMC population size peaks occur across numerous studies. *Curr Biol*. 2025;**35**:927–930.e3. https://doi.org/10.1016/j.cub.2024.09.028

Hinton MA. *Monograph of the voles & lemmings (Microtinae) living and extinct*. British Museum, London; 1926:1–488.

Jaarola M, Martínková N, Gündüz İ, Brunhoff C, Zima J, Nadachowski A, Amori G, Bulatova NS, Chondropoulos B, Fraguedakis-Tsolis S, et al. Molecular phylogeny of the speciose vole genus *Microtus* (Arvicolinae, Rodentia) inferred from mitochondrial DNA sequences. *Mol Phylogenet Evol*. 2004;**33**:647–663. https://doi.org/10.1016/j.ympev.2004.07.015

Johnsen K, Devineau O, Andreassen HP. Phase-and season-dependent changes in social behaviour in cyclic vole populations. *BMC Ecol*. 2019;**19**:1–9. https://doi.org/10.1186/s12898-019-0222-3

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

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**:1–16. https://doi.org/10.1093/gigascience/gix085

Krebs CJ. *Microtus* population biology: behavioral changes associated with the population cycle in *M. ochrogaster* and *M. pennsylvanicus*. *Ecology*. 1970;**51**:34–52. https://doi.org/10.2307/1933598

Krohn AR, Conroy CJ, Pesapane R, Bi K, Foley JE, Rosenblum EB. Conservation genomics of desert dwelling California voles (*Microtus californicus*) and implications for management of endangered Amargosa voles (*Microtus californicus scirpensis*). *Conserv Genet*. 2018;**19**:383–395. https://doi.org/10.1007/s10592-017-1010-2

Krokhmal O, Rekovets L, Kovalchuk O. Biochronological scheme of the quaternary of the south of Eastern Europe and its substantiation based on arvicoline teeth morphometrics. *Quat Int*. 2023;**674-675**:5–17. https://doi.org/10.1016/j.quaint.2022.12.003

Lemskaya NA, Romanenko SA, Golenishchev FN, Rubtsova NV, Sablina OV, Serdukova NA, O'Brien PCM, Fu B, Yiğit N, Ferguson-Smith MA, et al. Chromosomal evolution of Arvicolinae (Cricetidae, Rodentia). III. Karyotype relationships of ten Microtus species. *Chromosome Res*. 2010;**18**:459–471. https://doi.org/10.1007/s10577-010-9124-0

Li H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *arXiv:13033997 [q-Bio]*. 2013. https://doi.org/10.48550/arXiv.1303.3997

Lidicker WZ. Dispersal. In: Tamarin RH, ed. *Biology of New World Microtus*. Special Publication #8, American Society of Mammalogists; Stillwater, Oklahoma, 1985.

Lim MM, Murphy AZ, Young LJ. Ventral striatopallidal oxytocin and vasopressin V1a receptors in the monogamous prairie vole (*Microtus ochrogaster*). *J Comp Neurol*. 2004;**468**:555–570. https://doi.org/10.1002/cne.10973

Lin D, Bi K, Conroy CJ, Lacey EA, Schraiber JG, Bowie RCK. Mito-nuclear discordance across a recent contact zone for California voles. *Ecol Evol*. 2018;**8**:6226–6241. https://doi.org/10.1002/ece3.4129

Mammal Diversity Database. Mammal diversity database (version 2.0). *Zenodo*. 2025. https://doi.org/10.5281/zenodo.4139818

Manni M, Berkeley MR, Seppey M, FA S, 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

Martin M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal*. 2011;**17**:10–12. https://doi.org/10.14806/ej.17.1.200

Minh BQ, Schmidt HA, Chernomor O, Schrempf D, Woodhams MD, Von Haeseler A, Lanfear R. IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era. *Mol Biol Evol*. 2020;**37**:1530–1534. https://doi.org/10.1093/molbev/msaa015

Modi W. Phylogenetic analyses of chromosomal banding patterns among the nearctic Arvicolidae (Mammalia: Rodentia). *Syst Zool*. 1987;**36**:109–136. https://doi.org/10.2307/2413264

Ondov BD, Treangen TJ, Melsted P, Mallonee AB, Bergman NH, Koren S, Phillippy AM. Mash: fast genome and metagenome distance estimation using MinHash. *Genome Biol*. 2016;**17**:132. https://doi.org/10.1186/s13059-016-0997-x

Palmer HM, Vriesman VP, Livsey CM, Fish CR, Hill TM. Holocene climate and oceanography of the coastal Western United States and California current system. *Clim Past*. 2023;**19**:199–232. https://doi.org/10.5194/cp-19-199-2023

Patton JL, Conroy CJ. The conundrum of subspecies: morphological diversity among desert populations of the California vole (*Microtus californicus*, Cricetidae). *J Mammal*. 2017;**98**:1010–1026. https://doi.org/10.1093/jmammal/gyx074

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**. 2020: 3047–3060. https://doi.org/10.1534/g3.120.401028

Ramirez 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. https://doi.org/10.1038/s41467-020-14998-3

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

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

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

Sitnikova NA, Romanenko SA, O'Brien PCM, Perelman PL, Fu B, Rubtsova NV, Serdukova NA, Golenishchev FN, Trifonov VA, Ferguson-Smith MA, et al. Chromosomal evolution of Arvicolinae (Cricetidae, Rodentia). I. The genome homology of tundra vole, field vole, mouse and golden hamster revealed by comparative chromosome painting. *Chromosome Res*. 2007;**15**:447–456. https://doi.org/10.1007/s10577-007-1137-y

Stenseth NC. Population cycles in voles and lemmings: density dependence and phase dependence in a stochastic world. *Oikos*. 1999;**87**: 427–461. https://doi.org/10.2307/3546809

Steppan SJ, Schenk JJ. Muroid rodent phylogenetics: 900-species tree reveals increasing diversification rates. *PLoS One*. 2017;**12**:e0183070. https://doi.org/10.1371/journal.pone.0183070

Tamarin RH, ed. *Biology of New World Microtus*Special Publication #8,. American Society of Mammalogists; Stillwater, Oklahoma,1985.

The Galaxy Community. The Galaxy platform for accessible, reproducible, and collaborative data analyses: 2024 update. *Nucleic Acids Research*. 2024;**52**:W83–W94.

Tripp JA, Berrio A, McGraw LA, Matz MV, Davis JK, Inoue K, Thomas JW, Young LJ, Phelps SM. Comparative neurotranscriptomics reveal widespread species differences associated with bonding. *BMC Genomics*. 2021;**22**:399. https://doi.org/10.1186/s12864-021-07720-0

Ullrich KK, Tautz D. Population genomics of the house mouse and the Brown rat. In: Dutheil JY. (ed). *Statistical population genomics, methods in molecular biology*. New York: Springer Nature; 2020.

Wagner F, Ruf I, Lehmann T, Hofmann R, Ortmann S, Schiffmann C, Hiller M, Stefen C, Stuckas H. Reconstruction of evolutionary changes in fat and toxin consumption reveals associations with gene losses in mammals: a case study for the lipase inhibitor PNLIPRP1 and the xenobiotic receptor NR1I3. *J Evol Biol*. 2022;**35**:225–239. https://doi.org/10.1111/jeb.13970

Wang X, Liang D, Wang X, Tang M, Liu Y, Liu S, Zhang P. Phylogenomics reveals the evolution, biogeography, and diversification history of voles in the Hengduan Mountains. *Commun Biol*. 2022;**5**:1124. https://doi.org/10.1038/s42003-022-04108-y

Withnell CB, Scarpetta SG. A new perspective on the taxonomy and systematics of Arvicolinae (Gray, 1821) and a new time-calibrated phylogeny for the clade. *PeerJ*. 2024;**12**:p.e16693. https://doi.org/10.7717/peerj.16693

Young LJ, Nilsen R, Waymire KG, MacGregor GR, Insel TR. Increased affiliative response to vasopressin in mice expressing the vasopressin receptor from a monogamous vole. *Nature*. 1999;**400**:766–768. https://doi.org/10.1038/23475

Young KA, Liu Y, Gobrogge KL, Wang H, Wang Z. Oxytocin reverses amphetamine-induced deficits in social bonding: evidence for an interaction with nucleus accumbens dopamine. *J Neurosci*. 2014;**34**: 8499–8506. https://doi.org/10.1523/JNEUROSCI.4275-13.2014

Zerbino DR, Achuthan P, Akanni W, Amode MR, Barrell D, Bhai J, Billis K, Cummins C, Gall A, Girón CG, et al. Ensembl 2018. *Nucleic Acids Res*. 2018;**46**:D754–D761. https://doi.org/10.1093/nar/gkx1098

Zhang C, Mirarab S. Weighting by gene tree uncertainty improves accuracy of quartet-based species trees. *Mol Biol Evol*. 2022;**39**:msac215. https://doi.org/10.1093/molbev/msac215

Journal of Heredity, 2026, 117, 347–356
https://doi.org/10.1093/jhered/esaf067
Genome Resources