**Title: A chromosome-level genome assembly of a vernal pool specialist amphibian, the Western Spadefoot, *Spea hammondii***

**Authors: Ben Thompsky[1,2], Eric Beraut[3], Robert D. Cooper[1,2,4], Merly Escalona[5,6], Robert E. Espinoza[7], Robert N. Fisher[8], Courtney Miller[1,2], Oanh Nguyen[9], Samuel Sacco[3], Ruta Sahasrabudhe[9], William E. Seligmann[3], Erin Tofflemier[1,2], Ian J. Wang[10,11], H. Bradley Shaffer[1,2]**

**Author Affiliations:**
**[1] Department of Ecology & Evolutionary Biology, University of California, Los Angeles, CA 90095-7239, USA;**
**[2] La Kretz Center for California Conservation Science, Institute of the Environment and Sustainability, University of California, Los Angeles, CA 90095-7239, USA;**
**[3]Department of Ecology and Evolutionary Biology, University of California, Santa Cruz, Santa Cruz, CA 95064, USA;**
**[4]Department of Fish and Wildlife Conservation, Virginia Polytechnic Institute and State University, Blacksburg, VA 24060, USA;**
**[5] Institute for Systems Genomics, University of Connecticut, Storrs, CT06269, USA;**
**[6]Department of Biomolecular Engineering, University of California Santa Cruz, Santa Cruz, CA 95064, USA;**
**[7]Department of Biology, California State University, Northridge, Northridge, California 91330-8303, USA;**
**[8] U.S. Geological Survey, 4165 Spruance Road Suite 200, San Diego, CA 92101, USA;**
**[9] DNA Technologies and Expression Analysis Core Laboratory, Genome Center, University of California-Davis, CA 95616, USA;**
**[10]Department of Environmental Science, Policy, and Management, University of California, Berkeley, CA 94720, USA;**
**[11]Museum of Vertebrate Zoology, University of California, Berkeley, CA 94720, USA**

Address correspondence to Ben Thompsky by email:
benthompsky@g.ucla.edu

**Running head:** Western spadefoot genome assembly

**Abstract**
We assembled and annotated a chromosome-level reference genome for the Western Spadefoot, *Spea hammondii* (Anura, Scaphiopodidae) representing one of only three amphibians included in the California Conservation Genomics Project (CCGP). *Spea hammondii* is a vernal pool breeding anuran native to California and northwestern Baja California which has undergone both range contractions and local extirpations across its distribution, primarily due to habitat degradation and drought. The species is recognized by the state of California as a Species of Special Concern and is proposed for listing under the United States Endangered Species Act. Using the established CCGP pipeline, this *S. hammondii* genome was produced using Pacific Biosciences HiFi long-reads and Omni-C proximity ligation, resulting in a *de novo* genome assembly 1.14 Gb in length, distributed across 479 scaffolds (scaffold N50 = 120.8 Mb; largest scaffold = 183.6 Mb) with a BUSCO completeness score of 90.9% using a conserved tetrapod ortholog set. Our assembly shows high base accuracy (QV = 63.7) and low frameshift error in coding regions (QV 50.42). Annotation of this genome yielded 20,434 genes with a BUSCO completeness score of 94.7%. This reference genome, in combination with range-wide resequencing data from CCGP, will facilitate statewide population genomic assessments to delineate conservation units, quantify inbreeding and genomic load, and test for adaptive variation associated with vernal pool hydrology and drought tolerance, all of which are important considerations in the proposed federal listing.

**Key Words:** California Conservation Genomics Project, conservation genomics, genome assembly, endangered species, Scaphiopodidae

**Introduction:**
The Western Spadefoot, *Spea hammondii*, is a species of medium-sized fossorial anuran that was historically widespread in the low-elevation, non-desert regions of central and southern California and northern Baja California, Mexico (Figure 1). The species often occurs in discrete and isolated populations tied to small seasonal wetlands, including human-constructed livestock ponds. Adult *S. hammondii* are primarily found in grassland, chaparral and oak woodland habitats with loose soil conducive to burrowing (Morey 2005; Hansen and Shedd 2025). During the non-rainy season, post-metamorphic juveniles and adults live underground in self-constructed burrows, preferring sand and silt over clay soils (Baumberger 2019). They emerge

from burrows during rainy nights, generally in the winter and spring. *Spea hammondii* breeding typically takes place between January and June. Breeding is wholly dependent on the presence of ephemeral water bodies, rendering the species exceptionally vulnerable to drought and climatic changes that lower or change the frequency of precipitation (Thomson, Wright, Shaffer 2016).

Larval development is extremely rapid and has been the focus of much of the published research on *S. hammondii*. In short-hydroperiod vernal pools, *S. hammondii* exhibit rapid development, with smaller sizes at metamorphosis. Larvae also show remarkable life history adaptations and phenotypic plasticity, with the potential to develop into cannibalistic carnivorous tadpoles with increased keratinization of the jaw in response to prey density and pond dry-down (Levis, Isdaner, and Pfennig 2018). While this 'cannibal morph' phenotype is better documented in the other three species in the genus (Levis *et al.* 2015), at least one published observation from northwestern Baja California has documented cannibalistic behaviors in metamorphosing *S. hammondii* (Alvarez *et al.* 2024).

The taxonomic history of the genus *Spea,* including S. *hammondi,* has been a source of confusion at multiple levels. Long placed in the family Pelobatidae, current taxonomy recognizes the New-World spadefoots as members of Scaphiopodidae, a small clade of seven species and two genera (*Spea* and *Scaphiopus*) (see Frost 2025 for a detailed taxonomic history). Scaphiopodids, along with the Old-World families Pelodytidae, Megophryidae and Pelobatidae, form a clade that is the sister group to Neobatrachia, a large clade that includes well over 90% of living anurans (Pyron and Wiens 2011). Recent genetic work (Neal *et al.* 2018) has shown that *S. hammondii* contains two species-level lineages that are divided into northern and southern groups by the Transverse Ranges, a mid-elevation mountain range that often forms a biogeographic boundary between many northern and southern California biotic elements (Rissler *et al.* 2006). Neal *et al.* (2018) suggested that each should be managed as distinct conservation units, based on both genetic data and species distribution modeling. While the single range-wide entity is currently recognized by California as a Species of Special Concern (Thomson *et al.* 2016) the northern and southern clades are each currently under consideration for listing under the United States Endangered Species Act (USFWS 2023).

This genome release is part of the California Conservation Genomics Project (CCGP), which provides genomic resources and their analyses to inform conservation actions at a statewide scale (Shaffer *et al.* 2022; Fiedler *et al.* 2022). This chromosome-level reference genome, the third available for the genus (see also *S. multiplicata* (Seidl *et al.* 2019)*,* GenBank accession GCA_009364415.1, *S.bombifrons*, GenBank accession GCF_027358695.1) provides a key resource to help address outstanding questions in the systematics and conservation of *S. hammondii*. As part of the CCGP, this genome will enable range-wide population-genomic analyses to delineate conservation units, quantify inbreeding and genomic load, and map adaptive variation associated with hydroperiod, temperature, and drought (Shaffer *et al.* 2022).

This information is of particular importance for this species as it exists in increasingly disjointed and likely inbred populations, making a reference genome vital to develop cost-effective marker panels for genetically-informed reintroductions, translocations, and long-term monitoring across California's fragmented seasonal-wetland landscapes. In congeneric *Spea*, reference genomes have provided a framework for understanding and locating the metabolic and endocrine pathways implicated in phenotypic plasticity (Seidl *et al.* 2019); this genome may provide the ability to determine whether these adaptations are also present in *S. hammondii*. Finally, this reference genome adds to a growing set of California vernal pool specialist species, including the freshwater crustaceans *Branchinecta lynchi* (Blair *et al.* 2023a), *Branchinecta lindahli* (Blair *et al.* 2023b), and *Lepidurus packardi* (Blair *et al.* 2022), and the vernal pool plants *Neostapfia colusana* (Pennington *et al.* 2025) and *Tuctoria greenei* (Toews *et al.* 2025), that will allow cross-taxon analyses of the genomic health of California's rich, endemic, and extremely endangered vernal pool ecosystems.

**Methods:**

*Biological Materials:*

We sequenced an adult *S. hammondii* (HBS 135690) that was collected as a tadpole from a drying ephemeral puddle (a tire track) on a dirt trail and reared in captivity for approximately 2.5 years in the California State University Northridge vivarium and euthanized on 19 August 2020. Liver, thigh muscle, tongue, and heart tissues were harvested and flash-frozen in liquid nitrogen within two minutes of euthanasia, and the remaining carcass was formalin-fixed for later museum deposition. This sample belongs to the southern clade/distinct population segment (Neal *et al.* 2018).

*High Molecular Weight Genomic DNA Isolation:*

High molecular weight genomic DNA (HMW gDNA) was isolated from a liver sample. 20 mg of flash frozen liver tissue was homogenized using a mortar and a pestle in the presence of liquid nitrogen. Homogenized tissue was lysed with 2ml of lysis buffer containing 100 mM NaCl, 10 mM Tris-HCL-pH 8.0, 25 mM EDTA, 0.5% SDS and 100 µg/ml proteinase K overnight at room temperature. Lysate was treated with 20 µg/ml RNase at 37°C for 30 minutes and then cleaned with equal volumes of phenol/chloroform using phase-lock gels (Quantabio, Beverly, MA, USA; Cat # 2302830). DNA was precipitated by adding 0.4X volume of 5M ammonium acetate and 3X volume of ice-cold ethanol. The resulting DNA pellet was washed twice with 70% ethanol and resuspended in an elution buffer (10mM Tris, pH 8.0). Purity of DNA was assessed with a NanoDrop ND-1000 spectrophotometer where 260/280 ratios of 1.88 and 260/230 of 2.14 were observed. Total DNA yield of 13.3 µg was obtained as quantified using Qbit 2.0 Fluorometer (Thermo Fisher Scientific, Waltham, MA, USA). DNA fragment size was quantified using a Femto pulse system (Agilent Technologies, Santa Clara, CA, USA); 44% of the DNA was in fragments > 100 kilobases (kb).

*HiFi library preparation:*
The HiFi SMRTbell library was constructed using the SMRTbell Express Template Prep Kit v2.0 (Pacific Biosciences - PacBio, Menlo Park, CA, USA; Cat. #100-938-900) according to the manufacturer's instructions. HMW gDNA was sheared to a target DNA size distribution between 15 kb – 20 kb using Diagenode's Megaruptor 3 system (Diagenode, Belgium; Cat. B06010003). The sheared gDNA was concentrated using 0.45X of AMPure PB beads (PacBio, Cat. #100-265-900) for the removal of single-strand overhangs at 37° C for 15 minutes, followed by further enzymatic steps of DNA damage repair at 37° C for 30 minutes, end repair and A-tailing at 20° C for 10 minutes and 65° C for 30 minutes, ligation of overhang adapters v3 at 20° C for 60 minutes and 65° C for 10 minutes to inactivate the ligase, then nuclease treated at 37° C for 1 hour. The SMRTbell library was purified and concentrated with 0.45X AMPure PB beads for size selection using the BluePippin/PippinHT system (Sage Science, Beverly, MA, USA; Cat #BLF7510/HPE7510) to collect fragments greater than 7-9 kb. The 15 – 20 kb average HiFi SMRTbell library was sequenced at UC Davis DNA Technologies Core (Davis, CA, USA) using two 8M SMRT cell (PacBio, Cat #101-389-001), Sequel II sequencing chemistry 2.0, and 30-hour movies each on a PacBio Sequel IIe sequencer.

*Omni-C Library Preparation:*
The Omni-C library was prepared using the Dovetail™ Omni-C™ Kit (Dovetail Genomics, Scotts Valley, CA, USA) according to the manufacturer's protocol with slight modifications. First, liver tissue 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 µm and 40 µm cell strainers to remove large debris. Fixed chromatin was 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, USA) 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, USA) on an Illumina NovaSeq 6000 platform (Illumina, San Diego, CA, USA) to generate approximately 100 million 2 x 150 bp read pairs per giga base pair (Gb) of genome size.

*Transcriptome RNA Extraction, Library Preparation, and Sequencing:*
Total RNA from 3 tissues (tongue, muscle, heart) was extracted using a Qiagen RNeasy Mini Kit (Qiagen, Netherlands), according to the manufacturer's protocol. RNA libraries were then prepared using the KAPA mRNA HyperPrep Kit (Roche, Switzerland) according to the

manufacturer's protocol. Libraries were sequenced with 100bp reads on an Illumina NovaSeq 6000 platform (Illumina), to generate approximately 50M reads per library at the Technology Center for Genomics and Bioformatics (UCLA, CA, USA) on an Illumina NovaSeq X.

*Nuclear genome assembly:*
We assembled the genome of a *S. hammondii* individual following the CCGP assembly pipeline Version 5.0, as outlined in Table 1, where we list the tools and non-default parameters used in the assembly process. We removed the remnants adapter sequences from the PacBio HiFi dataset using HiFiAdapterFilt (Sim *et al.* 2022) and generated an initial diploid phased assembly using HiFiasm (Cheng *et al.* 2021) in HiC mode with the filtered PacBio HiFi reads and the Omni-C short-reads, a process that generates two assemblies, one per haplotype. We then aligned the Omni-C data to each assembly following the Arima Genomics Mapping Pipeline (https://github.com/ArimaGenomics/mapping_pipeline) and scaffolded them independently 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. To generate the contact maps, we aligned the Omni-C data with BWA-MEM (Li 2013), identified ligation junctions, and generated Omni-C pairs (Lee *et al.* 2022) using pairtools (Open2C *et al.* 2023). Then, we generated multi-resolution Omni-C matrices with cooler (Abdennur and Mirny 2020) and balanced them with hicExplorer (Ramírez *et al.* 2018). We used HiGlass (Kerpedjiev *et al.* 2018) and the PretextSuite (ttps://github.com/wtsi-hpag/PretextView; https://github.com/wtsi-hpag/PretextMap; https://github.com/wtsi-hpag/PretextSnapshot) to visualize the contact maps. We identified misassemblies and misjoins in these contact maps and modified the assemblies using the Rapid Curation pipeline from the Wellcome Trust Sanger Institute, Genome Reference Informatics Team (https://gitlab.com/wtsi-grit/rapid-curation). Some of the remaining gaps (joins generated during scaffolding and/or curation) were closed using the PacBio HiFi reads and YAGCloser (https://github.com/merlyescalona/yagcloser). We checked for contamination using the BlobToolKit Framework (Challis *et al.* 2020).

*Mitochondrial Genome:*
We assembled the mitochondrial genome of *S. hammondii* from the PacBio HiFi reads using the reference-guided pipeline MitoHiFi (Allio *et al.* 2020; Uliano-Silva *et al.* 2023). The mitochondrial sequence of a *Pelobates fuscus* (NCBI:NC_051953.1) was used as the starting sequence. After completion of the nuclear genome, we searched for matches of the resulting mitochondrial assembly sequence in the nuclear genome assembly using BLAST+ (Camacho *et al.* 2009) and filtered out contigs and scaffolds from the nuclear genome with a percentage of sequence identity > 99% and size smaller than the mitochondrial assembly sequence.

*Genome quality assessment and size estimation:*

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* 2020) to estimate genome features including genome size, heterozygosity, and repeat content. To obtain general contiguity metrics, we ran QUAST (Gurevich et*t* 2013). To evaluate genome quality and functional completeness, we used BUSCO (Manni *et al.* 2021) with the Tetrapoda ortholog database (tetrapoda_odb10) which contains 5,310 genes. Assessment of base level accuracy (QV) and k-mer completeness were 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 were 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); and C = % genome represented by the first 'n' scaffolds, following a karyotype of 2n=26 for this species (Wasserman 1970). Quality metrics for the notation were calculated on the assembly for the haplotype one assembly.

*Genome Annotation*

We annotated the genome assembly using the NCBI Eukaryotic Genome Annotation Pipeline v0.4.1-alpha (hereafter, "EGAPx") which is published in the NCBI RefSeq database (O'Leary *et al.* 2016) and accessible through the NCBI GitHub page ("ncbi/egapx"). Annotation features were identified by aligning transcripts and proteins from related taxa in the RefSeq database using BLAST (Camacho *et al.* 2009). Novel, species-specific RNA-Seq reads were also aligned to the assembly using the alignment software STAR (Dobin *et al.* 2013). Additional features are predicted using HMM-based gene models using the NCBI Gnomon software. We evaluated the quality and completeness of our annotation by comparing the longest protein for each annotated coding gene to tetrapods (odb10) using BUSCO v5.7.1 (Manni *et al.* 2021). We report the number of annotation features and the BUSCO results in Table 2.

**Results:**

*Sequencing data:*

The Omni-C library generated 289.96 million read pairs, and the PacBio HiFi library generated 4.65 million reads. The PacBio HiFi sequences yielded ~56X genome coverage and had an N50 read length of 18,307 bp; a minimum read length of 1 bp; a mean read length of 18,118 bp; and a maximum read length of 65,001 bp (see Supplementary Figure S1 for read length distribution). Based on the PacBio HiFi data, Genomescope 2.0 estimated a genome size of 1.09 Gb, a 0.164 % sequencing error rate, and 0.355% heterozygosity. The k-mer spectrum shows a unimodal distribution with a major coverage peak at ~43-fold coverage (Figure 2A). Sequencing

of mRNA libraries for SAMN40863797, SAMN40863796, and SAMN40863795 yielded 28M, 28M, 32M paired end reads, respectively.

*Nuclear genome assembly:*

The final genome assembly (aSpeHam1) consists of two phased haplotypes. Both assemblies are similar in size, but not equal to the estimated genome size from GenomeScope2.0, as has been observed in other taxa (see Pflug *et al.* 2020, for example).

The haplotype one assembly (aSpeHam1.0.hap1) consists of 479 scaffolds spanning 1.14 Gb with a contig N50 of 11.56 Mb, a scaffold N50 of 120.77 Mb, the largest contig size of 61.12 Mb, and the largest scaffold size of 183.6 Mb. The haplotype two assembly (aSpeHam1.0.hap2) consists of 351 scaffolds spanning 1.15 Gb with a contig N50 of 14.33 Mb, a scaffold N50 of 120.84 Mb, the largest contig size of 46.95 Mb, and the largest scaffold size of 188 Mb. The haplotype one assembly has a BUSCO completeness score for the Tetrapoda gene set of 90.9%, a base pair quality value (QV) of 63.7, a k-mer completeness of 95.96%, and a frameshift indel QV of 50.42. The haplotype two assembly has a BUSCO completeness score for the Tetrapoda gene set of 91.5%, a base pair QV of 64.35, a k-mer completeness of 96.1%, and a frameshift indel QV of 50.31.

During manual curation we made 217 joins (103 on haplotype one and 114 on haplotype two) and 28 breaks (12 on haplotype one and 16 on haplotype two) based on the Omni-C contact map signal. We closed 20 gaps (9 on haplotype one and 11 on haplotype two). We filtered out two contigs corresponding to mitochondrial contamination. No further contigs were removed. The Omni-C contact maps show highly contiguous assemblies supporting the genome organization of 13 chromosomes per haplotype (Figure 2B). Classic cytogenetic work reports a diploid number of 2n=26 for *S. hammondii,* historically treated as *Scaphiopus hammondii* (Wasserman 1970). By sorting our haploid assembly scaffolds by length, we found that the largest 13 contain 90.87% of the total genome length. We believe the remaining sequence represents repeat-dense, unresolved regions and is likely contained in the smaller scaffolds. Assembly statistics are reported in Table 2 and represented graphically in Figure 2C. We have deposited the genome assembly on NCBI GenBank (See Table 2 and Data Availability for details).

*Mitochondrial genome assembly:*
We assembled a partial mitochondrial genome for *S. hammondii* with MitoHiFi. The final mitochondrial sequence has a size of 16,682 bp, with a base composition of A=29.85%, C=27.89%, G=, 15.2%, T=27.07%, and consists of 2 rRNAs, 22 unique transfer RNAs, and 13 protein-coding genes.

***Genome Annotation***

Our final genome annotation for *S. hammondii* included 20,434 genes, with a tetrapoda BUSCO completeness of 94.70%. A list of annotation statistics and BUSCO score breakdowns are reported in Table 2.

**Discussion:**

There are presently 63 chromosome-level genomes for anurans in the NCBI database, representing 42 species. To this data set, we add a chromosome-level reference genome for the Western Spadefoot, *S. hammondii*. Assembly quality is reflected in core metrics (N50 = 120.8 Mb) with the largest scaffolds reaching 183.6-188.0 Mb, and completeness (BUSCO = 90.9% against a conserved tetrapod ortholog set). Omni-C contact maps support long-range contiguity, and our scaffold N50 exceeds the expected mean chromosome length of pelobatoid frogs (~88 Mb, using the pelobatoid haploid karyotype of 13), justifying the chromosomal scale of this assembly.

Amphibian genomes are remarkably variable in size, with many anurans showing clade-specific genome expansions or contractions. Compared to other anurans, *Spea* genomes are extremely compact. Of the 495 anuran genomes reported in the Animal Genome Size Database (https://www.genomesize.com), scaphiopodids are almost always among the smallest 10 species; their reported c-values (average = 1.42, n = 17 observations) are much smaller than anurans generally (average = 4.47, n = 495). Zeng *et al.* (2014) analyzed the genome size distribution within pelobatoid frogs (a group including *Spea* and related families) and found a strong statistical association between larval development time and genome size. However, the genome size statistics reported by Zeng *et al.* (2014) are also derived from c-values in the Animal Genome Size Database rather than assembled genomes, and these two estimates can vary. Our assembly spans 1.142 Gb, which aligns with the other two available *Spea* genomes but is on the low end of previously published c-values for *S. hammondii* (c-values of 1.27, 1.34, 1.55, and 2.34; Gregory 2025). We interpret this as reflecting the modest underrepresentation of repetitive sequence in our current assembly and/or variability in older Feulgen c-value estimates. Regardless, as annotated genomes of other anurans continue to accumulate, future studies should explore whether the developmental time/genome size correlation persists across pelobatoids and anurans more generally.

This *S. hammondii* reference genome also offers a comparative framework for studying phenotypic plasticity in spadefoots. Studies in congeners (e.g. *S. multiplicata*) found that rapid development and resource polyphenism involve shifts in lipid and cholesterol metabolism, and peroxisome biology (Levis *et al.* 2020). Our reference genome enables future mapping of these candidate pathways and their regulatory neighborhoods in *S. hammondii*, and testing for selection on these regions along hydroperiod gradients.

This reference genome is directly useful for supporting the conservation goals of the CCGP. Range-wide genome resequencing built on this coordinate system will enable delineation of conservation units across the patchy distribution for this species (including the genetic division north and south of the Transverse Ranges identified by Neal *et al.* [2018] based on RADseq data), estimation of relatedness, runs of homozygosity, and quantification of genomic load on a population basis. Landscape genetics analyses can identify loci linked to hydroperiod, desiccation tolerance, developmental rate, and temperature sensitivity. If inbreeding is detected, this information can be used to plan reintroductions and assisted gene flow to potentially minimize the negative effects of inbreeding while preventing the proliferation of harmful alleles and the degradation of natural diversity.

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

**Acknowledgements:**
PacBio Sequel II/IIe library prep and sequencing were carried out at the DNA Technologies and Expression Analysis Core at the UC Davis Genome Center, supported by NIH Shared Instrumentation Grant 1S10OD010786-01. Deep sequencing of Omni-C libraries used the Novaseq S4 sequencing platforms at the Vincent J. Coates Genomics Sequencing Laboratory at UC Berkeley, supported by NIH S10 OD018174 Instrumentation Grant. We thank the staff at the UC Davis DNA Technologies and Expression Analysis Core and the UC Santa Cruz Paleogenomics Laboratory for their diligence and dedication to generating high quality sequence data. We thank the vivarium staff in the Department of Biology at California State University, Northridge for animal care and financial support from the College of Science and Mathematics. RNA extraction, library preparation and sequencing was carried out at the UCLA Technology Center for Genomics & Bioinformatics.

**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 PRJNA937275. Raw sequencing data for sample HBS135690 (NCBI BioSample SAMN33386125) are deposited in the NCBI Short Read Archive (SRA) under SRR25482847 for PacBio HiFi sequencing data, and SRR25482845-46 for the Omni-C Illumina sequencing data. GenBank accessions for both primary and alternate assemblies are GCA_029215705.1 and GCA_029215755.1; and for

genome sequences JARDYI000000000 and JARDYJ000000000. The GenBank genome assembly for the mitochondrial genome is CM055131. Data generated for the annotation in this study are available under NCBI BioProject PRJNA1013362. Raw RNA sequencing data (NCBI BioSamples SAMN40863797, SAMN40863796 and SAMN40863795) are deposited in the NCBI SRA [SRR29346342, SRR29346343, SRR29346344], respectively). 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., Fudenberg, G., Flyamer, I. M., Galitsyna, A. A., Goloborodko, A., Imakaev, M., & Venev, S. V. (2024). Pairtools: From sequencing data to chromosome contacts. *PLOS Computational Biology, 20*(5), e1012164. https://doi.org/10.1371/journal.pcbi.1012164

Abdennur, N., & Mirny, L. A. (2020) Cooler: Scalable storage for Hi-C Data and other genomically labeled arrays." *Bioinformatics,* 36(1),311–16. https://doi.org/10.1093/bioinformatics/btz540

Allio, R., Schomaker-Bastos, A., Romiguier, J., Prosdocimi, F., Nabholz, B., & Delsuc, F.(2020) MitoFinder: Efficient automated large-scale extraction of mitogenomic data in target enrichment phylogenomics. *Molecular Ecology Resources* 20(4), 892–905. https://doi.org/10.1111/1755-0998.13160

Alvarez, J., Valdez-Villavicencio, J., & Peralta-Garcia, A. (2024) Putative cannibalism in the western spadefoot (*Spea hammondii*) in northern Baja California, México. *Sonoran Herpetologist 37*(June), 96-97.

Baumberger, K. L., Eitzel, M. V., Kirby, M. E., & Horn, M. H. (2019). Movement and habitat selection of the western spadefoot (Spea hammondii) in southern California. *PloS one*, *14*(10), e0222532. https://doi.org/10.1371/journal.pone.0222532

Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., & Madden, T. L. (2009). BLAST+: Architecture and applications. *BMC Bioinformatics, 10*, 421. https://doi.org/10.1186/1471-2105-10-421

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

Cheng, H., Jarvis, E. D., Fedrigo, O., … (2021). Robust haplotype-resolved assembly of diploid individuals without parental data [Preprint]. *arXiv*. https://doi.org/10.48550/arXiv.2109.04785

Danecek P, Bonfield JK, Liddle J, Marshall J, Ohan V, Pollard MO, Whitwham A, Keane T, McCarthy SA, Davies RM, Li H, Twelve years of SAMtools and BCFtools, GigaScience (2021) 10(2) giab008 [33590861]

Dobin, A., Davis, C. A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., & Gingeras, T. R. (2013). STAR: Ultrafast universal RNA-seq aligner. *Bioinformatics, 29*(1), 15–21. https://doi.org/10.1093/bioinformatics/bts635

Fiedler, P. L., Erickson, B., Esgro, M., Gold, M., Hull, J. M., Norris, J., … Shaffer, H. B. (2022). Seizing the moment: The opportunity and relevance of the California Conservation Genomics Project to state and federal conservation policy. *Journal of Heredity, 113*, 589–596. https://doi.org/10.1093/jhered/esac046

Frost, D. R. (2025). Amphibian Species of the World: An online reference (Version 6.1). American Museum of Natural History. *Retrieved October 2025, from* https://amphibiansoftheworld.amnh.org/

Gregory, T. R. 2025. *Animal Genome Size Database*. Accessed October 2025. http://www.genomesize.com

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

Ghurye, J., Rhie, A., Walenz, B. P., *et al.* (2019). Integrating Hi-C links with assembly graphs for chromosome-scale assembly. *PLOS Computational Biology, 15*(8), e1007273. https://doi.org/10.1371/journal.pcbi.1007273

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

Hansen, R. W., & Shedd, J. D. (2025). *California amphibians and reptiles* (Princeton Field Guides). Princeton University Press.

Kerpedjiev, P., Abdennur, N., Lekschas, F., McCallum, C., Dinkla, K., Strobelt, H., Luber, J. M., Ouellette, S. B., Azhir, A., Kumar, N., Hwang, J., Lee, S., Alver, B. H., Pfister, H., Mirny, L. A., Park, P. J., & Gehlenborg, N. (2018). HiGlass: Web-based visual exploration and analysis of genome interaction maps. *Genome Biology, 19*(1), 125. https://doi.org/10.1186/s13059-018-1486-1

Kieran Blair, S. R., Schreier, A., Escalona, M., Finger, A. J., Joslin, S. E. K., Sahasrabudhe, R., … Hull, J. M. (2023). A draft reference genome of the vernal pool fairy shrimp, *Branchinecta lynchi*. *Journal of Heredity, 114*(1), 81–87. https://doi.org/10.1093/jhered/esac056

Kieran Blair, S. R., Schreier, A., Escalona, M., Finger, A. J., Joslin, S. E. K., Sahasrabudhe, R., … Hull, J. M. (2023). A chromosome-level reference genome for the versatile fairy shrimp, *Branchinecta lindahli*. *Journal of Heredity, 114*(1), 74–80. https://doi.org/10.1093/jhered/esac057

Kieran Blair, S. R., Hull, J. M., Escalona, M., Finger, A. J., Joslin, S. E. K., Sahasrabudhe, R., Marimuthu, M. P. A., Nguyen, O. H., Chumchim, N., Morris, E. R., Velazquez, S., & Schreier, A. D. (2022). The reference genome of the vernal pool tadpole shrimp, *Lepidurus packardi*. *Journal of Heredity, 113*(6), 706–711. https://doi.org/10.1093/jhered/esac051

Korlach, J., Gedman, G., Kingan, S. B., Chin, C.-S., Howard, J. T., Audet, J.-N., Cantin, L., & Jarvis, E. D. (2017). De novo PacBio long-read and phased avian genome assemblies correct and add to reference genes generated with intermediate and short reads. *GigaScience, 6*(10), gix085. https://doi.org/10.1093/gigascience/gix085

Lee, S., Bakker, C. R., Vitzthum, C., Alver, B. H., & Park, P. J. (2022). Pairs and Pairix: A file format and a tool for efficient storage and retrieval for Hi-C read pairs. *Bioinformatics, 38*(6), 1729–1731. https://doi.org/10.1093/bioinformatics/btab870

Levis, N. A., de la Serna Buzón, S., & Pfennig, D. W. (2015). An inducible offense: Carnivore morph tadpoles induced by tadpole carnivory. *Ecology and Evolution, 5*(7), 1405–1411. https://doi.org/10.1002/ece3.1448

Levis, N. A., Isdaner, A. J., & Pfennig, D. W. (2018). Morphological novelty emerges from pre-existing phenotypic plasticity. *Nature Ecology & Evolution, 2*(8), 1289–1297. https://doi.org/10.1038/s41559-018-0601-8

Levis, N. A., Reed, E. M. X., Pfennig, D. W., & Burford Reiskind, M. O. (2020). Identification of candidate loci for adaptive phenotypic plasticity in natural populations of spadefoot toads. *Ecology and Evolution, 10*(16), 8976–8988. https://doi.org/10.1002/ece3.6602

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

Manni, M., Berkeley, M. R., Seppey, M., Simão, F. A., & Zdobnov, E. M. (2021). BUSCO update: Novel and streamlined workflows along with broader and deeper phylogenetic coverage for scoring of eukaryotic, prokaryotic, and viral genomes. *Molecular Biology and Evolution, 38*(10), 4647–4654. https://doi.org/10.1093/molbev/msab199

Morey, S. R. (2005). *Spea multiplicata*. In M. J. Lannoo (Ed.), *Amphibian declines: The conservation status of United States species* (pp. 519–522). University of California Press.

Neal, K. M., Johnson, B. B., & Shaffer, H. B. (2018). Genetic structure and environmental niche modeling confirm two evolutionary and conservation units within the western spadefoot (*Spea hammondii*). *Conservation Genetics, 19*(4), 937–946. https://doi.org/10.1007/s10592-018-1066-7

O'Leary, N. A., Wright, M. W., Brister, J. R., Ciufo, S., Haddad, D., McVeigh, R., Rajput, B., Robbertse, B., Smith-White, B., Ako-Adjei, D., Astashyn, A., Badretdin, A., Bao, Y., Blinkova, O., Brover, V., Chetvernin, V., Choi, J., Cunningham, F., DiCuccio, M., … Pruitt, K. D. (2016). Reference sequence (RefSeq) database at NCBI: Current status, taxonomic expansion, and functional annotation. *Nucleic Acids Research, 44*(D1), D733–D745. https://doi.org/10.1093/nar/gkv1189

Pennington, L. K., Escalona, M., Toews, D. J., Chumchim, N., Cooper, R., Fairbairn, C. W., Marimuthu, M. P. A., Miller, C., Nguyen, O. H., Ruiz-Ramos, D., Seligmann, W. E., Stephens, M., Toffelmier, E., Shaffer, H. B., Meyer, R. S., & Sexton, J. P. (2025). A reference genome for Colusa grass, *Neostapfia colusana*, a threatened and endangered California vernal pool plant. *Journal of Heredity*. Advance online publication. https://doi.org/10.1093/jhered/esaf075

Pflug, J. M., Holmes, V. R., Burrus, C., Johnston, J. S., & Maddison, D. R. (2020). Measuring genome sizes using read-depth, k-mers, and flow cytometry: Methodological comparisons in beetles (Coleoptera). *G3: Genes|Genomes|Genetics, 10*(9), 3047–3060. https://doi.org/10.1534/g3.120.401028

Pyron, R. A., & Wiens, J. J. (2011). A large-scale phylogeny of Amphibia including over 2800 species, and a revised classification of extant frogs, salamanders, and caecilians. *Molecular Phylogenetics and Evolution, 61*(2), 543–583. https://doi.org/10.1016/j.ympev.2011.06.012

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

Ranallo-Benavidez, T. R., Jaron, K. S., & Schatz, M. C. (2020). GenomeScope 2.0 and Smudgeplot for reference-free profiling of polyploid genomes. *Nature Communications, 11*(1), 1432. https://doi.org/10.1038/s41467-020-14998-3

Rhie, A., McCarthy, S. A., Fedrigo, O., Damas, J., Formenti, G., Koren, S., Uliano-Silva, M., Chow, W., Fungtammasan, A., Kim, J., Lee, C., Ko, B. J., Chaisson, M., Gedman, G. L., Cantin, L. J., Thibaud-Nissen, F., Haggerty, L., Bista, I., Smith, M., … Jarvis, E. D. (2021).

Towards complete and error-free genome assemblies of all vertebrate species. *Nature, 592*(7856), 737–746. https://doi.org/10.1038/s41586-021-03451-0

Rhie, A., Walenz, B. P., Koren, S., & Phillippy, A. M. (2020). Merqury: Reference-free quality, completeness, and phasing assessment for genome assemblies. *Genome Biology, 21*(1), 245. https://doi.org/10.1186/s13059-020-02134-9

Rissler, L. J., Hijmans, R. J., Graham, C. H., Moritz, C., & Wake, D. B. (2006). Phylogeographic lineages and species comparisons in conservation analyses: A case study of California herpetofauna. *The American Naturalist, 167*(5), 655–666. https://doi.org/10.1086/503332

Shaffer, H. B., Toffelmier, E., Corbett-Detig, R. B., Escalona, M., Erickson, B., Fiedler, P., Gold, M., Harrigan, R. J., Hodges, S., Luckau, T. K., Miller, C., Oliveira, D. R., Shaffer, K. E., Shapiro, B., Sork, V. L., & Wang, I. J. (2022). Landscape genomics to enable conservation actions: The California Conservation Genomics Project. *Journal of Heredity, 113*(6), 577–588. https://doi.org/10.1093/jhered/esac020

Sim, S. B., Corpuz, R. L., Simmonds, T. J., & Geib, S. M. (2022). 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, 23*(1), 157. https://doi.org/10.1186/s12864-022-08375-1

Seidl, F., Levis, N. A., Schell, R., Pfennig, D. W., Pfennig, K. S., & Ehrenreich, I. M. (2019). Genome of *Spea multiplicata*, a rapidly developing, phenotypically plastic, and desert-adapted spadefoot toad. *G3: Genes|Genomes|Genetics, 9*(12), 3909–3919. https://doi.org/10.1534/g3.119.400705

Thomson, R. C., Wright, A. N., & Shaffer, H. B. (2016). *California amphibian and reptile species of special concern* (1st ed.). University of California Press. https://www.jstor.org/stable/10.1525/j.ctv1xxxbx

Toews, D. J., Escalona, M., Pennington, L. K., Chumchim, N., Fairbairn, C. W., Marimuthu, M. P. A., Miller, C., Nguyen, O. H., Ruiz-Ramos, D., Seligmann, W. E., Stephens, M., Cooper, R. D., Toffelmier, E., Shaffer, H. B., Meyer, R. S., & Sexton, J. P. (2025). A genome assembly of Greene's tuctoria, *Tuctoria greenei*, an amphibious endemic and endangered California vernal pool grass. *Journal of Heredity*. Advance online publication. https://doi.org/10.1093/jhered/esaf084

Toffelmier, E., Beninde, J., & Shaffer, H. B. (2022). The phylogeny of California, and how it informs setting multispecies conservation priorities. *Journal of Heredity, 113*(6), 597–603. https://doi.org/10.1093/jhered/esac045

Uliano-Silva, M., Ferreira, J. G. R. N., Krasheninnikova, K., Darwin Tree of Life Consortium, Blaxter, M., Mieszkowska, N., Hall, N., Holland, P., Durbin, R., Richards, T., Kersey, P.,

Hollingsworth, P., Wilson, W., Twyford, A., Gaya, E., Lawniczak, M., Lewis, O., Broad, G., Martin, F., … McCarthy, S. A. (2023). MitoHiFi: A python pipeline for mitochondrial genome assembly from PacBio high fidelity reads. *BMC Bioinformatics, 24*(1), 288. https://doi.org/10.1186/s12859-023-05385-y

U.S. Fish and Wildlife Service. (2023, December 5). Endangered and threatened wildlife and plants: Threatened status with section 4(d) rule for the northern and southern distinct population segments of the western spadefoot. *Federal Register, 88*(232), 84252–84278. https://www.federalregister.gov/documents/2023/12/05/2023-26579/endangered-and-threatened-wildlife-and-plants-threatened-status-with-section-4d-rule-for-the

Wasserman, A. O. (1970). Chromosomal Studies of the Pelobatidae (Salientia) and Some Instances of Ploidy. *The Southwestern Naturalist*, *15*(2), 239–248. https://doi.org/10.2307/3670352

Zeng, C., Gomez-Mestre, I., & Wiens, J. J. (2014). Evolution of rapid development in spadefoot toads is unrelated to arid environments. *PLOS ONE, 9*(5), e96637. https://doi.org/10.1371/journal.pone.0096637

**Table 1.** Assembly pipeline and software used.

| **Assembly step** | **Software and any non-default options** | **Version** | **Reference** |
|---|---|---|---|
| **Initial assembly** | | | |
| Filtering PacBio HiFi adapters | HiFiAdapterFilt | Commit 64d1c7b | Sim *et al.* 2022 |
| K-mer counting | Meryl (k=21) | 1 | https://github.com/marbl/meryl |
| Estimation of genome size and heterozygosity | GenomeScope | 2 | Ranallo-Benavidez *et al.* 2020 |
| *De novo assembly (contiging)* | HiFiasm (Hi-C Mode, –primary, output hic.hap1.p_ctg, hic.hap2.p_ctg) | 0.16.1-r375 | Cheng *et al.* 2021 |

| | | | |
|---|---|---|---|
| **Scaffolding** | | | |
| Omni-C data alignment | Arima Genomics Mapping Pipeline | Commit 2e74ea4 | https://github.com/ArimaGenomics/mapping_pipeline |
| Arima Genomics Mapping Pipeline (AGMP) | BWA-MEM | 0.7.17-r1188 | Li 2013 |
| | samtools | 1.11 | Danecek *et al.* 2021 |
| | filter_five_end.pl (AGMP) | Commit 2e74ea4 | https://github.com/ArimaGenomics/mapping_pipeline |
| | two_read_bam_combiner.pl ( (AGMP)) | Commit 2e74ea4 | https://github.com/ArimaGenomics/mapping_pipeline |
| | picard | 2.27.5 | https://broadinstitute.github.io/picard/ |
| Omni-C Scaffolding | SALSA (-DNASE, -i 20, -p yes) | 2 | Ghurye *et al.* 2017, Ghurye *et al.*. 2019 |
| **Omni-C Contact map generation** | | | |
| Short-read alignment | BWA-MEM (-5SP) | 0.7.17-r1188 | Li 2013 |
| SAM/BAM processing | samtools | 1.11 | Danecek *et al.* 2021 |
| SAM/BAM filtering | pairtools | 0.3.0 | Open2C *et al.* 2024 |
| Pairs indexing | pairix | 0.3.7 | Lee *et al.* 2022 |
| Matrix generation | cooler | 0.8.10 | Abdennur and Mirny 2020 |
| Matrix balancing | hicExplorer (hicCorrectmatrix correct --filterThreshold -2 4) | 3.6 | Ramírez *et al.* 2018 |
| Contact map visualization | HiGlass | 2.1.11 | Kerpedjiev *et al.* 2018 |
| | PretextMap | 0.1.4 | https://github.com/wtsi-hpag/PretextView |

| | | | |
|---|---|---|---|
| | PretextView | 0.1.5 | https://github.com/wtsi-hpag/PretextMap |
| | PretextSnapshot | 0.0.3 | https://github.com/wtsi-hpag/PretextSnapshot |
| Manual curation tools | Rapid curation pipeline (Wellcome Trust Sanger Institute, Genome Reference Informatics Team) | Commit 7acf220c | https://gitlab.com/wtsi-grit/rapid-curation |
| **Genome quality assessment** | | | |
| Basic assembly metrics | QUAST (--est-ref-size) | 5.0.2 | Gurevich *et al.* 2013 |
| Assembly completeness | BUSCO (-m geno, -l tetrapoda) | 5.0.0 | Manni *et al.* 2021 |
| | Merqury | 2020-01-29 | Rhie *et al.* 2020 |
| **Contamination screening** | | | |
| Local alignment tool | BLAST+ (-db nt, -outfmt '6 qseqid staxids bitscore std' , -max_target_seqs 1, -max_hsps 1, -evalue 1e-25 ) | 2.15 | Camacho *et al.* 2009 |
| General contaminatio n screening | BlobToolKit (HiFi coverage, BUSCO=tetrapoda, NCBI Taxa ID=228670) | 2.3.3 | Challis *et al.* 2020 |
| **Mitochondrial assembly** | | | |
| Mitochondria l genome assembly | MitoHiFi (-r, -p 90, -o 1) Reference: Pelobates fuscus (NCBI:NC_051953.1) | 2.2 | Uliano-Silva *et al.* 2023 |

| | | | |
|---|---|---|---|
| MitoHiFi: Mitochondrial genome annotation | MitoFinder (MitoHiFi pipeline parameters) | 1.4 | Allio *et al.* 2020 |

**Table 2:** Sequencing and assembly statistics and accession numbers.

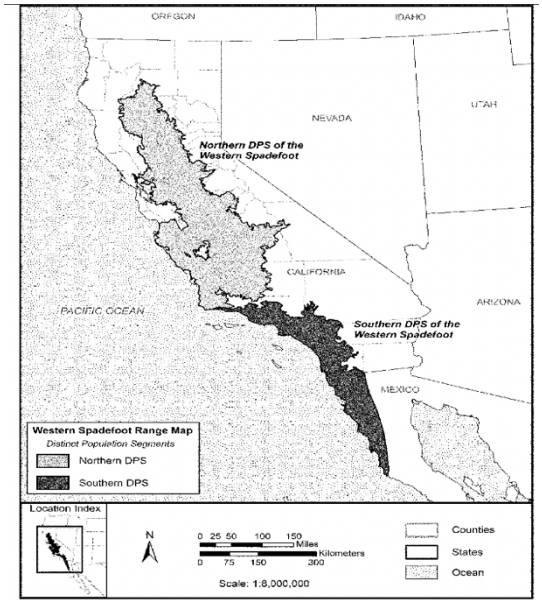
| | | | | |
|---|---|---|---|---|
| Bio Projects & Vouchers | CCGP NCBI BioProject | | PRJNA720569 | |
| | Genera NCBI BioProject | | PRJNA765864 | |
| | Species NCBI BioProject | | PRJNA937275 | |
| | NCBI BioSample | | Assembly: SAMN33386125 | |
| | | | Transcriptome: SAMN40863797, | |
| | | | SAMN40863796, SAMN40863795 | |
| | Specimen identification | | HBS 135690 | |
| | NCBI Genome accessions | | **Haplotype one** | **Haplotype two** |
| | Assembly accession | | JARDYI000000000 | JARDYJ000000000 |
| | Genome sequences | | GCA_029215705.1 | GCA_029215755.1 |
| Genome Sequence | PacBio HiFi reads | Run | 1 PACBIO_SMRT (Sequel II) run: 2.9M spots, 48.5G bases, 35.2Gb | |
| | | Accession | SRR25482847 | |
| | Omni-C Illumina reads | Run | 2 ILLUMINA (Illumina NovaSeq 6000) runs: 101.3M spots, 30.6G bases, 10.4Gb | |
| | | Accession | SRR25482845-46 | |
| Genome Assembly | Assembly identifier (Quality code *) | | aSpeHam1(7.8.P7.Q64.C90) | |
| | HiFi Read coverage § | | 44.26X | |
| | | | **Haplotype one** | **Haplotype two** |

| | | | | | | |
|---|---|---|---|---|---|---|
| Quality Metrics | Number of contigs | | | 677 | | 541 |
| | Contig N50 (bp) | | | 11,561,794 | | 14,335,781 |
| | Contig NG50 § | | | 12,894,986 | | 15,063,856 |
| | Longest Contigs | | | 61,120,978 | | 46,956,032 |
| | Number of scaffolds | | | 479 | | 351 |
| | Scaffold N50 | | | 120,773,319 | | 120,842,655 |
| | Scaffold NG50 § | | | 120,773,319 | | 120,842,655 |
| | Largest scaffold | | | 183,606,164 | | 188,005,797 |
| | Size of final assembly | | | 1,142,014,214 | | 1,155,539,324 |
| | Phased block NG50 § | | | 12,894,986 | | 15,063,856 |
| | Gaps per Gbp (# Gaps) | | | 173(198) | | 164(190) |
| | Indel QV (Frame shift) | | | 50.42043639 | | 50.31166 |
| | Base pair QV | | | 64.3477 | | 64.3477 |
| | | | | Full assembly = 64.0148 | | |
| | k-mer completeness | | | 95.968 | | 96.1908 |
| | | | | Full assembly = 99.1337 | | |
| | BUSCO completeness (tetrapoda) n=5310 | | **C** | **S** | **D** | **F** | **M** |
| | | H1 ‡ | 90.90% | 89.90% | 1.00% | 2.30% | 6.80% |
| | | H2 ‡ | 91.50% | 90.50% | 1.00% | 2.30% | 6.20% |
| | Organelles | | 1 complete mitochondrial sequence | | | CM055131 | |
| Genome Annotation Quality Metrics | | | | **Count of Features** | | |
| | Genes | | | 20,434 | | |
| | mRNA | | | 24,883 | | |
| | lncRNA | | | 887 | | |
| | CDSs | | | 24,883 | | |
| | BUSCO completeness | | **C** | **S** | **D** | **F** | **M** |
| | (tetrapoda_odb10) n= 5,310 | | 94.70% | 93.20% | 1.50% | 0.70% | 4.60% |

* 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 SPECIES of 2n=26 (Wasserman 1970). Quality code for all the assembly denoted by haplotype one assembly (aSpeHap1.0.hap1).

| |
|---|
| § Read coverage and NGx statistics have been calculated based on the estimated genome size of 1.09 Gb |
| ‡ (H1) Haplotype one and (H2) haplotype two assembly values. |

A)




B)

C)

D)

E)

F)

A) Range map of *Spea hammondii* showing the two genetically separate subpopulations north and south of the Transverse Ranges (USFWS 2023). B) Mid-metamorphosis spadefoot toad tadpole. Animals at this stage have well-developed spades on their rear feet and are capable of completing metamorphosis on land (Robert D. Cooper, image). (E) *Spea hammondii* tadpole, displaying their relatively large size (Nora Papian/USFWS, Public Domain, https://www.fws.gov/media/spadefoot-toad-tadpole). C) An adult Western spadefoot from northern Santa Barbara County, California (H. Bradley Shaffer, image). D) Vernal pool used by *S. hammondii* for breeding. This pool is located in San Diego County, California (H. Bradley Shaffer, image). F) Close-up of the keratinized spade on the rear foot of an adult *S. hammondii*. Spadefoots use this spade to dig burrows in loose sandy soil (H. Bradley Shaffer, image)

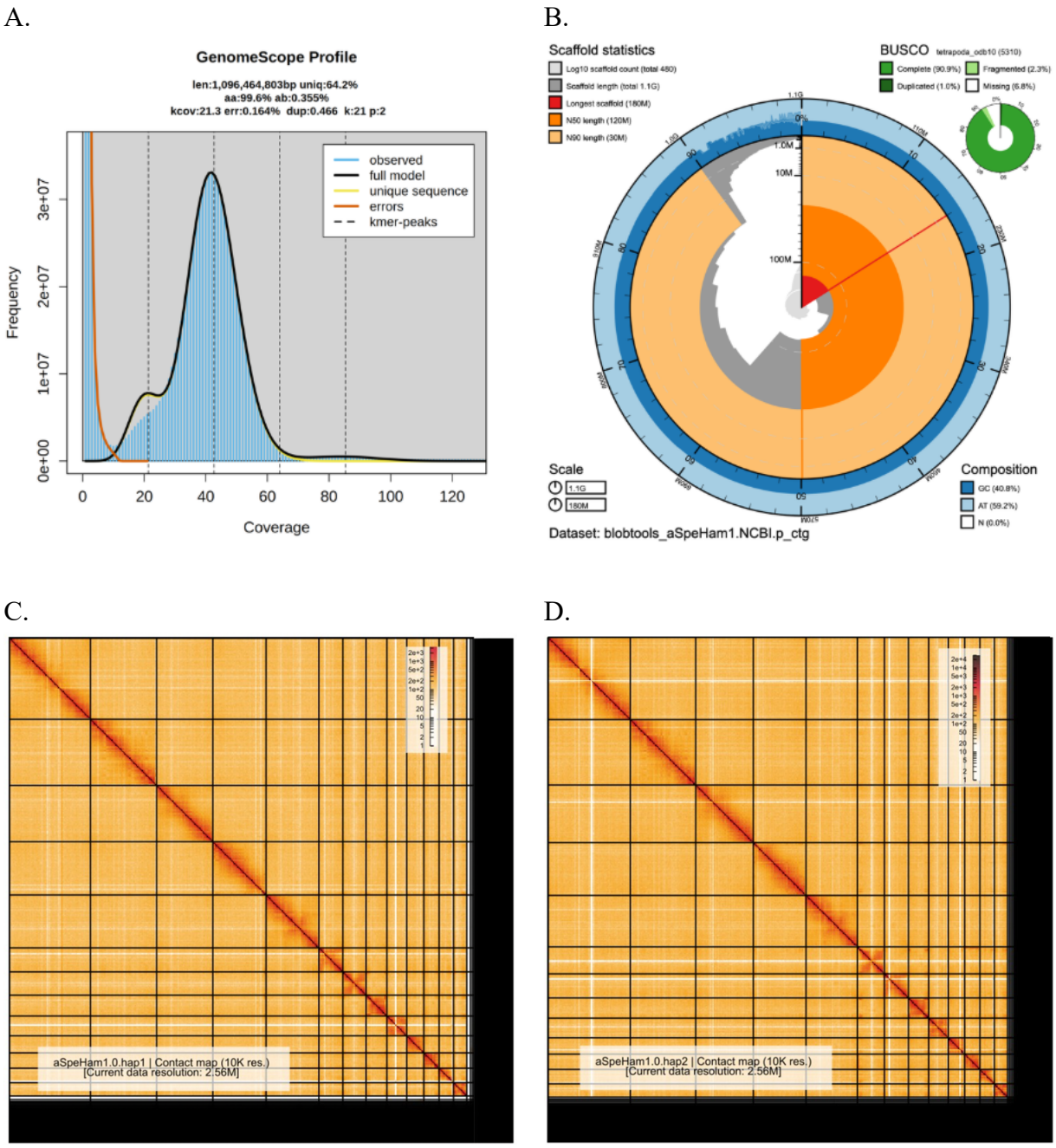



Figure 2: Visual overview of assembly metrics. A) K-mer spectrum output generated from PacBio HiFi data without adapters using GenomeScope2.0. The unimodal pattern observed corresponds to a haploid genome. K-mers covered at lower coverage and low frequency correspond to differences between haplotypes, whereas the higher coverage and high frequency k-mers correspond to the similarities between haplotypes. B) BlobToolkit Snail plot showing a graphical representation of the quality metrics presented in Table 2 for the *Spea hammondii* primary assembly (haplotype one; GenBank accession no. GCA_029215705.1). The plot circle represents the full size of the assembly. From the inside-out, the central plot covers length related

metrics. The red line represents the size of the longest scaffold; all other scaffolds are arranged in size-order moving clockwise around the plot and drawn in grey starting from the outside of the central plot. Dark and light orange arcs show the scaffold N50 and scaffold N90 values. The central light grey spiral shows the cumulative scaffold count with a white line at each order of magnitude. White regions in this area reflect the proportion of Ns in the assembly. The dark vs. light blue area around it shows mean, maximum and minimum GS vs. AT content at 0.1% intervals (Challis *et al.* 2020). For BUSCO summary data, the definitions are 'complete and single copy' (Comp.) or 'complete and duplicated' (Dupl.). C-D) Omni-C Contact maps for the haplotype 1 (2C) and haplotype 2 (2D) assemblies generated with PretextSnapshot. Omni-C contact maps translate proximity of genomic regions in 3-D space to contiguous linear organization. Each line in the contact map corresponds to sequencing data supporting the linkage (or join) between two of such regions. Scaffolds are separated by black lines and higher density corresponds to high levels of fragmentation.