Reference Genome of the Northwestern Pond Turtle, *Actinemys marmorata*

Brian D. Todd, Thomas S. Jenkinson, Merly Escalona, Eric Beraut, Oanh Nguyen, Ruta Sahasrabudhe, Peter A. Scott, Erin Toffelmier, Ian J. Wang, and H. Bradley Shaffer

From the Department of Wildlife, Fish, and Conservation Biology, University of California, Davis, CA 95616, USA (Todd and Jenkinson); Department of Biomolecular Engineering, University of California–Santa Cruz, Santa Cruz, CA 95064, USA (Esca1ona); Department of Ecology and Evolutionary Biology, University of California–Santa Cruz, Santa Cruz, CA 95064, USA (Beraut); DNA Technologies and Expression Analysis Core Laboratory, Genome Center, University of California, Davis, CA 95616, USA (Nguyen and Sahasrabudhe); Department of Life, Earth, and Environmental Sciences, West Texas A&M University, Canyon, TX 79016, USA (Scott); Department of Ecology and Evolutionary Biology, University of California, Los Angeles, CA 90095-7239, USA (Toffelmier and Shaffer); La Kretz Center for California Conservation Science, Institute of the Environment and Sustainability, University of California, Los Angeles, CA 90095-7239, USA (Toffelmier and Shaffer); Department of Environmental Science, Policy, and Management, University of California, Berkeley, CA 94720, USA (Wang); and Museum of Vertebrate Zoology, University of California, Berkeley, CA 94720, USA (Wang).

Address correspondence to B. D. Todd at the address above, or e-mail: btodd@ucdavis.edu

Corresponding Editor: Beth Shapiro

Abstract

The northwestern pond turtle, *Actinemys marmorata*, and its recently recognized sister species, the southwestern pond turtle, *A. pallida*, are the sole aquatic testudines occurring over most of western North America and the only living representatives of the genus *Actinemys*. Although historically ranged from Washington state through central California, USA, populations of the northwestern pond turtle have been in decline for decades and the species is afforded state-level protection across its range; it is currently being considered for protection under the US Endangered Species Act. Here, we report a new, chromosome-level assembly of *A. marmorata* as part of the California Conservation Genomics Project (CCGP). Consistent with the reference genome strategy of the CCGP, we used Pacific Biosciences HiFi long reads and Hi-C chromatin-proximity sequencing technology to produce a de novo assembled genome. The assembly comprises 198 scaffolds spanning 2,319,339,408 base pairs, has a contig N50 of 75 Mb, a scaffold N50 of 146 Mb, and BUSCO complete score of 96.7%, making it the most complete testudine assembly of the 24 species from 13 families that are currently available. In combination with the *A. pallida* reference genome that is currently under construction through the CCGP, the *A. marmorata* genome will be a powerful tool for documenting landscape genomic diversity, the basis of adaptations to salt tolerance and thermal capacity, and hybridization dynamics between these recently diverged species.

**Key words**: California Conservation Genomics Project, CCGP, conservation genetics, testudine, Emydidae

The northwestern pond turtle (*Actinemys marmorata*) and southwestern pond turtle (*A. pallida*) are endemic to western North American and were together recognized as a single species until recently (Spinks et al. 2014). They comprise the only wide-ranging freshwater turtle species on North America’s Pacific coast (Turtle Taxonomy Working Group 2021) and are the only native turtles in most of the region. As currently understood, the northwestern pond turtle range extends from central Washington south to California’s San Francisco Bay-Delta, in both coastal and more interior drainages, and continues south throughout the San Joaquin Valley, with outlying populations in Nevada (Spinks et al. 2014; Thomson et al. 2016; Bury 2017). *Actinemys marmorata* is replaced in the coast ranges south of San Francisco Bay by *A. pallida*, which extends south to San José, Baja California, Mexico (Valdez-Villavicencio et al. 2016). Limited admixture between the species has been found around the San Francisco Bay-Delta (Spinks et al. 2014), and in scattered localities further south. With few other native freshwater turtles in their range—and none currently extant in California or Northern Baja California—the *A. marmorata/ pallida* complex has historically occupied low-mid elevation habitats below 2000 m, ranging from permanent lakes and rivers to intermittent ponds and streams. They are also one of only a handful of freshwater turtles globally with populations found commonly in brackish water habitats (Agha et al. 2018). Despite their ecological breadth and relatively large range, *A. marmorata* populations have declined precipitously over the last several decades. As a result, they are currently listed as endangered in the state of Washington (WDFW 1993), “sensitive/critical” in Oregon (ODFW 2021), a “Species of Conservation Priority” in Nevada (NDOW 2012), and a “Species of Special Concern” in California (Thomson et al. 2016). The US Fish and Wildlife Service announced in 2015 that protection under the Endangered Species Act (ESA 1973, 2017).
as amended) may be warranted, and its listing status is currently under review (USFWS 2015). Important causes of decline include predation on hatchling turtles, particularly by non-native American bullfrogs (*Rana catesbeiana*), and largemouth bass (*Micropterus salmoides*), pathogens—including a newly described pathogenic shell fungus—and habitat alteration (Woodburn et al. 2019; Nicholson et al. 2020; Manzo et al. 2021).

Here, we report the first chromosome-level genome assembly for the species, produced as part of the California Conservation Genomics Project (CCGP). The overarching goal of the CCGP is to discover patterns of genomic diversity across the state of California by sequencing the complete genomes of approximately 150 carefully selected species (Shaffer et al. 2022). Many of these taxa are threatened or endangered, and the combined reference genome sequences plus the landscape genomics approach of the CCGP will enable the identification of hotspots of diversity across California, providing a unique framework for informed conservation decisions and management plans (Shaffer et al. 2022). This is the first completed animal reference genome constructed for the CCGP and follows our first plant taxon (Huang et al. 2021). This genome assembly will provide a foundational resource for future studies on the unique ecology, biogeography, evolutionary history, and conservation of *A. marmorata*.

**Methods**

**Biological Materials**

We captured an adult male northwestern pond turtle using baited hoop nets in Putah Creek in Solano County, CA (38.517263°N, 121.751965°W) on 16 July 2020 (CDFW permit# SC-11197 to BDT) (Figure 1). Distinguishing *A. marmorata* from *A. pallida* can be difficult; we identified the captured individual based on the geographic location of the collection site and the presence of enlarged, triangular inguinal scutes (Figure 1), which are found in 89% of northwestern pond turtles, but are reduced or absent in 94% of southwestern pond turtles (Seeliger 1945; Nicholson et al. 2020). BDT collected 0.8 mL of fresh whole blood from the subcarapacial sinus, briefly stored it in a Vacutainer tube with EDTA, and delivered it to the UC Davis Genome Center within 30 min of collection. Diagnostic characters for confirming the species were photographed for later accessioning of photo-vouchers and the turtle was released alive at the point of capture.

**High-Molecular-Weight Genomic DNA Isolation**

High-molecular-weight genomic DNA (gDNA) was isolated from whole blood following a method described previously with some modifications (Jain et al. 2018). Briefly, 50 µL of whole blood preserved in EDTA was lysed with 2 mL of lysis buffer containing 10 mM Tris–HCl pH 8.0, 25 mM EDTA, 0.5% (w/v) SDS, and 100 µg/mL Proteinase K until the solution was homogenous. The lysate was treated with 20 µg/ml RNase A at 37°C for 30 min. The lysate was cleaned with equal volumes of phenol/chloroform using phase lock gels (Quantabio, Cat # 2302830). DNA was precipitated by adding 0.4× volume of 5 M ammonium acetate and 3× volume of ice-cold ethanol. The DNA pellet was washed twice with 70% ethanol and resuspended in an elution buffer.

![Figure 1. (A) A northwestern pond turtle, *Actinemys marmorata*. (B) Close-up image of the plastron (lower shell) of the reference genome specimen (photograph voucher MWFB Acc 2021-49-02 from the UC Davis Museum of Wildlife and Fish Biology). The arrow points to the enlarged inguinal scute that is characteristic of northwestern pond turtles and that is typically absent or greatly reduced in southwestern pond turtles, *A. pallida*. (C) Putah Creek, Solano County, California. This is an example of representative stream habitat for northwestern pond turtles and the trapping locality for the reference genome specimen.](image-url)
HiFi Library Preparation and Sequencing

A HiFi SMRTbell library was constructed using the SMRTbell Express Template Prep Kit v2.0 (Pacific Biosciences (PacBio), Menlo Park, CA; Cat. #100-938-900) according to the manufacturer's instructions. Ten micrograms of HMW gDNA were sheared to an average size distribution of ~18 kb using Diagenode's Megaruptor 3 system (Diagenode, Belgium; Cat. #B06010001). Sheared DNA was quantified by Quantus Fluorometer QuantiFluor ONE dsDNA Dye assay (Promega, Madison, WI; Cat. #E6150), and the size distribution was checked by Agilent Femto Pulse (Agilent Technologies, Santa Clara, CA; Cat. #P-0003-0817). The sheared gDNA was concentrated using 0.45× of AMPure PB beads (PacBio, Cat. #100-265-900). Concentrated, sheared gDNA was quantified by Quantus Fluorometer QuantiFluor ONE dsDNA Dye assay (Promega, Madison, WI; Cat. #E6150). Six micrograms of concentrated, sheared gDNA was used as input for the removal of single-strand overhangs at 37 °C for 15 min, followed by further enzymatic steps of DNA damage repair at 37 °C for 30 min, end repair and A-tailing at 20 °C for 10 min and 65 °C for 30 minutes, ligation of overhang adapter v3 at 20 °C for 60 min followed by 65 °C for 10 min to inactivate the ligase, and nucllease treatment of SMRTbell library at 37 °C for 1 h to remove damaged or nonintact SMRTbell templates (SMRTbell Enzyme Cleanup Kit, PacBio, Cat. #107-746-400). The SMRTbell library was purified and concentrated with 0.8x Ampure PB beads (PacBio, Cat. #100-265-900) for size selection using the BluePippin system (Sage Science, Beverly, MA; Cat #BLF7510). The input of 2.2 µg purified SMRTbell library was used to load into the BluePippin 0.75% Agarose Cassette (Sage Science, Cat. #BLF7510) using cassette definition 0.75% DF Marer S1 3-10kb Improved Recovery for the run protocol. Fragments greater than 7.5 kb were collected from the cassette elution well. The size selected SMRTbell library was purified and concentrated with 0.8x AMPure beads (PacBio, Cat. #100-265-900). The 18 kb average HiFi SMRTbell library was sequenced at UC Davis DNA Technologies Core (Davis, CA) using three 8M SMRTCells, Sequel II sequencing chemistry 2.0, and 30-h movies on a PacBio Sequel II sequencer.

Omni-C Library Preparation and Sequencing

The Omni-C library was prepared using Dovetail Omni-C Kit according to the manufacturer's protocol with slight modifications. Briefly, chromatin was fixed in place in the nucleus. Fixed chromatin was digested with DNase I, then extracted. 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 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 (NEB, Ipswich, MA) with an Illumina compatible y-adaptor. Biotin-containing fragments were then captured using streptavidin beads. The postcapture product was split into 2 replicates prior to PCR enrichment to preserve library complexity with each replicate receiving unique dual indices. The library was sequenced at Vincent J. Coates Genomics Sequencing Lab (Berkeley, CA) on an Illumina NovaSeq platform to generate approximately 516 million reads.

Nuclear Genome Assembly

We assembled the genome of the northwestern pond turtle following the CCGP assembly protocol Version 2.0, which is adapted from the initial CCGP assembly efforts (see Huang et al. 2021). The main difference between versions is the use of Omni-C instead of Hi-C data, alongside PacBio HiFi Reads, for the generation of high-quality and highly contiguous nuclear genome assemblies. The final output corresponds to a diploid assembly that consists of 2 pseudo-haplotypes (primary and alternate). The primary assembly is more complete and consists of longer phased blocks. The alternate consists of haplotigs (contigs of clones with the same haplotype) in heterozygous regions and is not as complete and more fragmented. Given the characteristics of the latter, it cannot be considered on its own, but as a complement of the primary assembly (https://nh3.github.io/2021/04/17/concepts-in-phased-assemblies; https://www.ncbi.nlm.nih.gov/grc/help/definitions/).

We removed remnant adapter sequences from the PacBio HiFi dataset using HiFiAdapterFilt (Version 1.0) (Sim 2021) and obtained the initial diploid assembly using HiFiasm (Version 0.15-r327) (Cheng et al. 2021) (see Table 1 for relevant software). Next, we identified sequences corresponding to haplotypic duplications, contig overlaps and repeats on the primary assembly with purge_dups (Version 1.2.5) (Guan et al. 2020) and transferred them to the alternate assembly. We scaffolded both assemblies using the Omni-C data with SALSA (Version 2.2) (Ghurye et al. 2017, 2019) and closed gaps generated during scaffolding with the PacBio HiFi reads and YACcloser (commit 20e2769) (https://github.com/merlyescalona/yaccloser).

The primary assembly was manually curated by iteratively generating and analyzing Omni-C contact maps. To generate the contact maps, we aligned the Omni-C data against the corresponding reference with bwa mem (Version 0.7.17-r1188, options -5SP) (Li 2013), identified ligation junctions, and generated Omni-C pairs using pairtools (Version 0.3.0) (Sim 2021) and the PretextSuite (https://github.com/wtsi-hpag/PretextMap; https://github.com/wtsi-hpag/PretextSnapshot) to visualize the contact maps. We then checked for contamination using the BlobToolKit Framework (Version 2.3.3) (Challis et al. 2018) and the PretextSuite to identify assemblies in phasing-assemblies, https://www.ncbi.nlm.nih.gov/grc/help/definitions/).

Finally, we trimmed remnants of sequence adapters and mitochondrial contamination upon NCBI contamination screening.

Mitochondrial Genome Assembly

We assembled the mitochondrial genome of the northwestern pond turtle from the PacBio HiFi reads using the reference-guided pipeline MitoHiFi (Version 2) (https://github.com/
marcelauliano/MitoHiFi (Allio et al. 2020). The mitochondrial sequence of *Trachemys scripta* (NC_011573.1) was used as the starting reference sequence. After completion of the nuclear genome, we searched for matches of the resulting mitochondrial assembly sequence in the nuclear genome assembly using BLAST + (Version 2.10) (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 Size Estimation and Quality Assessment**

We generated k-mer counts \((k = 21)\) from the PacBio HiFi reads using meryl (Version 1) (https://github.com/sheinasim/HiFiAdapterFilt). The generated k-mer database was then used in GenomeScope2.0 (Version 2.0) (Ranallo-Benavidez et al. 2020) to estimate genome features including genome size, heterozygosity, sequencing error, and repeat content. To obtain general contiguity metrics, we ran QUAST (Version 5.0.2) (Gurevich et al. 2013). To evaluate genome quality and completeness, we used BUSCO (Version 5.0.0) (Simão et al. 2015; Seppey et al. 2019) with the vertebrata ortholog database (vertebrate_odb10), which contains 3354 genes. Assessment of base-level accuracy (QV) and k-mer completeness was performed using the previously generated meryl database and merqury (Rhie et al. 2020). We further estimated genome assembly accuracy via BUSCO gene set frameshift analysis using the pipeline described in Korlach et al. (2017).

**Results**

We generated a de novo nuclear genome assembly of the northwestern pond turtle (rActMar1) using 113.9 million read pairs of Omni-C data and 2.4 million PacBio HiFi reads. The latter yielded 30-fold coverage (N50 read length 14 712 bp; minimum read length 28 bp; mean read length 14 535 bp; maximum read length of 53 168 bp). Calculation of coverage is based on the initially estimated 2.6 Gb genome size. Assembly statistics are reported in tabular and graphical form in Table 2 and Figure 2B, respectively.

---

**Table 1.** Assembly pipeline and software usage. Software citations are listed in the text

| Assembly | Software | Version |
|----------|----------|---------|
| Filtering PacBio HiFi adapters | HiFiAdapterFilt | https://github.com/sheinasim/HiFiAdapterFilt |
| K-mer counting | Meryl | 1 |
| Estimation of genome size and heterozygosity | GenomeScope | 2 |
| De novo assembly (contiging) | HiFiasm | 0.15-c327 |
| Long read, genome-genome alignment | minimap2 | 2.16 |
| Remove low-coverage, duplicated contigs | purge_dups | 1.2.6 |

| Scaffolding | Arima Genomics mapping pipeline | Commit 2e74ea4 |
|-------------|--------------------------------|---------------|
| Hi-C mapping for SALSA | SALSA | 2 |
| Gap closing | YAGCloser | Commit 20e2769 |

| Hi-C Contact map generation | bwa | 0.7.17-r1188 |
|----------------------------|-----|------------|
| SAM/BAM processing | samtools | 1.11 |
| SAM/BAM filtering | pairtools | 0.3.0 |
| Pairs indexing | pairix | 0.3.7 |
| Matrix generation | Cooler | 0.8.10 |
| Matrix balancing | hicExplorer | 3.6 |
| Contact map visualization | HiGlass | 2.1.11 |
| PretextMap | 0.1.4 |
| PretextView | 0.1.5 |
| PretextSnapshot | 0.0.3 |

| Organelle assembly | MitoHiFi | Commit c06ed3e |
|-------------------|----------|---------------|
| Genome quality assessment | QUAST | 5.0.2 |
| Assembly completeness | BUSCO | 5.0.0 |
| Merqury | 1.0.2 |
| Contamination screening | BLAST+ | 2.10 |
| General contamination screening | BlobToolKit | 2.3.3 |
The primary assembly consists of 49 scaffolds spanning 2.3 Gb with contig N50 of 75 Mb, scaffold N50 of 146 Mb, longest contig of 224 Mb, and largest scaffold of 362 Mb. The Omni-C contact map suggests that the primary assembly is chromosome-level (Figure 2C). As expected, the alternate assembly, which consists of sequence from heterozygous regions, is less contiguous (Figure 2D). Because the assembly is not fully phased, we have deposited scaffolds corresponding to the alternate pseudohaplotype in addition to the primary assembly.

The final genome size (2.3 Gb) is very similar to the estimated values from the Genomescope2.0 k-mer spectra. The k-mer spectrum output shows a bimodal distribution with 2 major peaks, at ~15- and ~30-fold coverage, where peaks correspond to homozygous and heterozygous states respectively (Figure 2A).

Based on the PacBio HiFi reads, we estimated a 0.164% sequencing error rate and 0.42% nucleotide heterozygosity rate. The assembly has a BUSCO completeness score of 96.7% using the vertebrata gene set, a per base quality (QV) of 66, a k-mer completeness of 92, and a frameshift indel QV of 46.59.

**Mitochondrial Assembly**

We assembled a mitochondrial genome with MitoHiFi. Final mitochondrial genome size was 17 148 bp. The base composition of the final assembly version is A = 34.62%, C = 26.37%,
G = 12.65%, T = 26.34%, and consists of 22 transfer RNAs and 13 protein coding genes.

Discussion
There are currently 24 species of turtles or tortoises with publicly available reference genomes, including representatives of 13 of the 14 recognized chelonian families; one of these reference genomes is an earlier assembly for *A. marmorata*. The *A. marmorata* genome presented here has the highest contig N50 (range of the other taxa, 21.3 kb–39.4 Mb, current assembly is 75.08 Mb), and one of the highest scaffold N50 values (range of other taxa, 228.1 kb–147.4 Mb; current assembly is 146.2 Mb). Across this diverse set of taxa, which spans a

---

**Figure 2.** Visual overview of genome assembly metrics. (A) K-mer spectra output generated from PacBio HiFi data without adapters using GenomeScope2.0. The bimodal pattern observed corresponds to a diploid genome and the k-mer profile matches that of low (<1%) heterozygosity. K-mers covered at lower coverage and lower frequency correspond to differences between haplotypes, whereas the higher coverage and higher 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 *Actinemys marmorata* primary assembly (rActMar1). 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 gray starting from the outside of the central plot. Dark and light orange arcs show the scaffold N50 and scaffold N90 values. The central light gray spiral shows the cumulative scaffold count with a white line at each order of magnitude. White regions in this area reflect the proportion of Ns in the assembly; the dark versus light blue area around it shows mean, maximum, and minimum GC vs. AT content at 0.1% intervals (Challis et al. 2020). Hi-C Contact maps for the primary (C) and alternate (D) genome assembly generated with PretextSnapshot. Hi-C contact maps translate proximity of genomic regions in 3D space to contiguous linear organization. Each cell in the contact map corresponds to sequencing data supporting the linkage (or join) between two of such regions.
crown group age of roughly 220 million years (Shaffer et al. 2017; Thomson et al. 2021), genome size estimates are remarkably conserved (average of previously published genomes, 2.2 Gb, range is 1.8–2.6 Gb), and the A. marmorata genome (2.3 Gb) falls comfortably within this range. Compared to the other 4 species in its family (Emydidae), the A. marmorata genome is average in size (average is 2.4 Gb for the other emydid species) and GC content (range for 4 species 44–45%; A. marmorata is 45%).

For both evolutionary and conservation studies, the current A. marmorata genome is an invaluable resource. Ongoing work, supported by the California Department of Fish and Wildlife, is nearing completion for a RADseq analysis of 1399 western pond turtle samples from across the combined range of A. marmorata and A. pallida, and both the A. marmorata and the forthcoming A. pallida reference genomes will provide the resources needed to accurately map RAD fragments. The same is true for a set of 60 resequenced individuals currently being analyzed to identify salt tolerance genes from the San Francisco Bay-Delta (Todd, unpublished data).

One of the most controversial aspects of the evolutionary history of the genus Actinemys has been the confused taxonomic history of the 2 contained species and their geographic boundaries. Seeliger (1945) first recognized that there were 2 entities of western pond turtles, but confusion over their geographic ranges and generic/species/subspecies status persisted for another 70 years (Spinks et al. 2014; Thomson et al. 2016; Nicholson et al. 2020). Our A. marmorata reference genome, in combination with a similar resource for its sister species A. pallida, will allow us to dissect, in detail, the genomic differences between these 2 taxa, including the potential identification of genes associated with speciation over their estimated several million years of evolutionary divergence (Thomson et al. 2021).

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), with additional support from the California Department of Water Resources (agreement #4600011531 to UC Davis) and the California Agricultural Experiment Station (CA-D-WFB-2617-H to B.D.T.).

Acknowledgements
PacBio Sequel II library preparation and sequencing was carried out at the DNA Technologies and Expression Analysis Cores 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 Cores, the UC Santa Cruz Paleogenomics Laboratory, and the CCGP Scientific Executive Committee and staff for their diligence and dedication to generating high-quality reference genome sequence data. Partial support was provided by Illumina for Omni-C sequencing. We thank Logan K. Todd for assisting with live capture of reference specimen and blood collection.

Data Availability
Data generated for this study are available under NCBI BioProject PRJNA720569. No museum voucher exists because the animal was released at point of capture, but it is photo voucheder at the UC Davis Museum of Wildlife and Fish Biology (photograph voucher MWFB Acc 2021-49). Raw sequencing data for this sample (NCBI BioSample SAMN21436765) are deposited in the NCBI Short Read Archive (SRA) under SRR17460090 for PacBio HiFi data and SRX13631283 for Omni-C Illumina short-read data. GenBank accession for both primary and alternate assemblies are GCA GCA_022086475.1 and GCA GCA_022086895.1; and for genome sequences JAJLPC000000000 and JAJLPD000000000. The GenBank organelle genome assembly for the mitochondrial genome is CM039065.1. Assembly scripts and other data for the analyses are at www.github.com/ccgproject/ccgp_assembly.

References
Abdennur N, Mirny LA. 2020. Cooler: scalable storage for Hi-C data and other genomically labeled arrays. Bioinformatics. 36:311–316.
Agha M, Ennen JR, Bower D, Nowakowski AJ, Sweat S, Todd BD. 2018. Salinity tolerances and use of saline environments by freshwater turtles: implications of sea level rise. Biol Rev. 93:1634–1648.
Allio R, Schomaker-Bastos A, Romiguier J, Prosdocimi F, Nahbolz B, Deluc F. 2020. MitoFinder: Efficient automated large-scale extraction of mitogenomic data in target enrichment phylogenomics. Mol Ecol Resour. 20:892–905.
Bury RB. 2017. Biogeography of western pond turtles in the western Great Basin: dispersal across a Northwest Passage? Western Wildl. 4:72–80.
Camacho C, Coulouris G, Avagya V, Ma N, Papadopoulos J, Bealer K, Madden TL. 2009. BLAST+: architecture and applications. BMC Bioinf. 10:421.
Challis R, Richards E, Rajan J, Cochran G, Blaxter M. 2020. BlobToolKit—interactive Quality Assessment of Genome Assemblies. G3 Genes Genomes Genet. 10:1361–1374.
Cheng H, Concepcion GT, Feng X, Zhang H, Li H. 2021. Haplotypereolved de novo assembly using phased assembly graphs with hifiasm. Nat Methods. 18:170–175.
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.
Ghurye J, Rhiie A, Walenz BP, Schmitt A, Selvaraj S, Pop M, Phillippy AM, Koren S. 2019. Integrating Hi-C links with assembly graphs for chromosome-scale assembly. PLoS Comput Biol. 8:e1007273.
Goloborodko A, Abdennur N, Venev S, Hbbrandao, Gfendugben. 2018. mirnylab/pairtools: v0.2.0. Available from: https://zenodo.org/rec-

ord/1490831.
Guan D, McCarthy SA, Wood J, Howe K, Wang Y, Durbin R. 2020. Identifying and removing haplotypic duplication in primary genome assemblies. Bioinformatics. 36:2896–2898.
Gurevich A, Saveliev V, Yakhni T, Tesler G. 2013. QUAST: quality assessment tool for genome assemblies. Bioinformatics. 29:1072–1075.
Huang Y, Escalona M, Morrison G, Marimuthu MPA, Nguyen O, Toffelmier E, Shaffer HB, Litt A. 2021. Reference genome assembly of the big berry Manzanita (Arctostaphylos glauca). J Hered. doi:10.1093jhered/esa071.
Jain M, Koren S, Migà KH, Quick J, Rand AC, Sasani TA, Tyson JR, Beggs AD, Dilthey AT, Fiddes IT, et al. 2018. Nanopore sequencing and assembly of a human genome with ultra-long reads. Nat Biotechnol. 36(4):338–345.

Biotechnol. 8:e1007273.

Great Basin: dispersal across a Northwest Passage? Western Wildl. 4:72–80.

Great Basin: dispersal across a Northwest Passage? Western Wildl. 4:72–80.

Great Basin: dispersal across a Northwest Passage? Western Wildl. 4:72–80.
Kerpediev P, Abdennur N, Lekshas F, McCallum C, Dinkla K, Strobelt H, Luber JM, Ouellette SB, Azhir A, Kumar N, et al. 2018. HiGlass: web-based visual exploration and analysis of genome interaction maps. *Genome Biol.* 19:125.

Korlach J, Gedman G, Kingan SB, Chin C-S, Howard JT, Audet J-N, Cantin L, Jarvis ED. 2017. De novo PacBio long-read and phased avian genome assemblies correctly add and read to reference genes generated with intermediate and short reads. *GigaScience.* 6:1–16.

Li H. 2013. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *arXiv [q-bio.GN]* http://arxiv.org/abs/1303.3997.

Manzo S, Nicholson EG, Devereux Z, Fisher RN, Brown CW, Scott PA, Shaffer HB. 2021. Conservation of Northwestern and Southwestern Pond Turtles: threats, population size estimates, and population viability analysis. *J Fish Wildl Manag.* 12(2):485–501.

ODFW. 2012. Nevada wildlife action plan. Available from: https://www.odfw.org/wp-content/uploads/2022/01/2013-NV-WAP-Complete-NOT-ADA.pdf.

Nicholson EG, Manzo S, Devereux Z, Morgan TP, Fisher RN, Brown C, Dagit R, Scott PA, Shaffer HB. 2020. Historical museum collections and contemporary population studies implicate roads and introduced predatory bullfrogs in the decline of western pond turtles. *PeerJ.* 8:e9248.

ODFW. 2021. Sensitive species list. Salem (OR): Oregon Department of Fish and Wildlife. Available from: https://www.dfw.state.or.us/wildlife/diversity/species/docs/Sensitive_Species_List.pdf.

Ramírez F, Bhardwaj V, Arrigoni L, Lam KC, Grüning BA, Villaveces J, Habermann B, Akhtar A, Manke T. 2018. High-resolution TADs reveal DNA sequences underlying genome organization in flies. *Nat Commun.* 9:189.

Ranallo-Benavidez TR, Jaron KS, Schatz MC. 2020. GenomeScope 2.0 reveals sequence reads, clone sequences and assembly contigs with BWA-MEM. *arXiv [q-bio.GN]* http://arxiv.org/abs/1303.3997.

Shaffer HB, McCartney-Melstad E, Near TJ, Mount GG, Spinks PQ. 2017. Phylogenomic analyses of 539 highly informative loci date a fully resolved time tree for the major clades of living turtles (Testudines). *Mol Phylogenet Evol.* 115:7–15.

Simão FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM. 2015. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics.* 31:3210–3212.

Sim S. 2021. shenassin/HiFiAdapterFilt: first release. Available from: https://zenodo.org/record/4716418.

Spinks PQ, Thomson RC, Shaffer HB. 2014. The advantages of going large: genome-wide SNPs clarify the complex population history and systematics of the threatened western pond turtle. *Mol Ecol.* 23:2228–2241.

Thomson RC, Spinks PQ, Shaffer HB. 2021. A global phylogeny of turtles reveals a burst of climate-associated diversification on continental margins. *Proc Natl Acad Sci USA.* 118(7):e2012215118.

Thomson RC, Wright AN, Shaffer HB. 2016. *California amphibian and reptile species of special concern.* Oakland (CA): California Department of Fish and wildlife. University of California Press.