Journal of Heredity, 2026, **117**, 329–339
https://doi.org/10.1093/jhered/esaf063
Advance access publication 5 September 2025
Genome Resources

American Genetic Association
OXFORD

Genome Resources

A highly contiguous genome assembly for the wrentit (*Chamaea fasciata*), the sole representative of the babbler radiation in the Americas

Phred M. Benham[1,2,*], Carla Cicero[1], Kevin J. Burns[3], Merly Escalona[4], Eric Beraut[5], Colin W. Fairbairn[5], William E. Seligmann[5], Mohan P.A. Marimuthu[6], Oanh Nguyen[6], Noravit Chumchim[6], Erin Toffelmier[7,8], Michael W. Nachman[1,2], Rauri C.K. Bowie[1,2]

[1]Museum of Vertebrate Zoology, University of California Berkeley, Berkeley, CA 94720, United States, [2]Department of Integrative Biology, University of California Berkeley, Berkeley, CA 94720, United States, [3]Department of Biology, San Diego State University, San Diego, CA 92182-4614, United States, [4]Department of Biomolecular Engineering, University of California Santa Cruz, Santa Cruz, CA 95064, United States, [5]Department of Ecology and Evolutionary Biology, University of California Santa Cruz, Santa Cruz, CA 95064, United States, [6]Deoxyribonucleic Acid Technologies and Expression Analysis Core Laboratory, Genome Center, University of California- Davis, CA 95616, United States, [7]Department of Ecology and Evolutionary Biology, University of California, Los Angeles, CA 90095-7239, United States, [8]La Kretz Center for California Conservation Science, Institute of the Environment and Sustainability, University of California, Los Angeles, CA 90095-7239, United States

*Corresponding author: Museum of Vertebrate Zoology, University of California Berkeley, Berkeley, CA 94720, United States. Email: phbenham@gmail.com

Abstract

The wrentit (*Chamaea fasciata*) is a chaparral and scrub specialist bird found from coastal Oregon to northern Baja California. We generated a draft reference assembly for the species using PacBio HiFi long read and Omni-C chromatin-proximity sequencing data as part of the California Conservation Genomics Project. Sequenced reads were assembled into 1342 scaffolds totaling 1.19 gigabase in length. A contig N50 of 4.5 Mb, scaffold N50 of 73.3 Mb, and Benchmarking Universal Single-Copy Orthologs completeness score of 96.8% indicate that the wrentit genome is a highly contiguous assembly in line with other high quality avian assemblies. An annotation of the assembly identified 16 821 protein-coding genes. We detected a translocation between chromosome 4A of the zebra finch to the Z chromosome of the wrentit. This translocation has previously been identified as a neo-sex chromosome shared across the superfamily Sylvioidea. Finally, we found a negative correlation between transposable element richness and gene density across the genome, but a positive relationship between guanine–cytosine content and gene density. This reference will serve as an essential resource for studies on the biogeography, local adaptation, and conservation genetics of this iconic species of California's chaparral.

Key words: Paradoxornithidae, chaparral, neo-sex chromosome, transposable elements, California conservation genomics project

Introduction

Individual species found far from a clade's geographic center of diversity represent biogeographical enigmas that can nevertheless provide key insights into the evolutionary history of a lineage (Min et al. 2005; Hosner et al. 2015). One such enigma is the wrentit (*Chamaea fasciata*) of western North America. When initially described by western scientists in 1845, the wrentit was placed with titmice and chickadees in the genus *Parus* (Gambel 1845). Gambel subsequently placed it in the monotypic genus *Chamaea*. Since the 1850s, different authors have argued for its placement with babblers (Timaliini), Sylviini warblers, bushtits (Aegithalidae), wrens (Troglodytidae), mockingbirds (Mimidae), or its own family Chamaeidae. Not until Sibley and Ahlquist's pioneering deoxyribonucleic acid (DNA)–DNA hybridization studies was its affiliation with Asiatic babblers (Timaliidae) confirmed (Sibley and Ahlquist 1982). Mitochondrial DNA later placed the wrentit with *Sylvia* warblers, which were in turn recovered as sister to core babblers (Barhoum and Burns 2002). Finally, a multi-locus phylogenetic analysis of core babblers and allies within the superfamily Sylvioidea identified the wrentit as sister to the parrotbills in the genus *Paradoxornis* with a divergence date of ~10 million years ago (Cai et al. 2019). Currently, the wrentit is assigned to the family Paradoxornithidae which is sister to the family Sylviidae (Billerman 2023), and it remains the only North American member of the > 450 species in the 'babbler' radiation.

Wrentits are a characteristic species of scrub and chaparral habitats. Their range extends south from the Columbia River through Oregon and California to northern Baja California, Mexico. Their distribution is limited to the east by the Sierra Nevada and Cascade mountains, and they are largely absent

Received on 12 June 2025; accepted on 3 September 2025

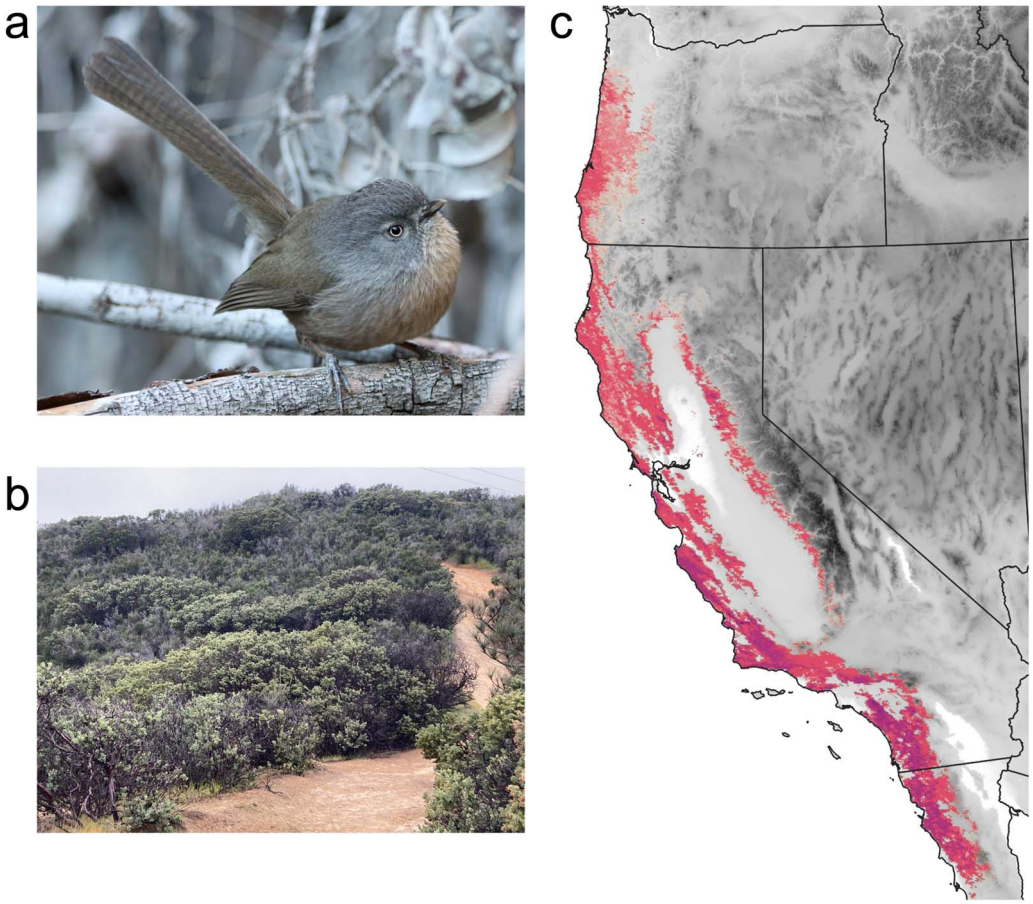

Fig. 1. a) Photograph of a wrentit (*C. fasciata*). Photo provided by Russell Campbell and taken at Ernest E. Debs Regional Park, Los Angeles County, CA. b) Characteristic chaparral habitat of the wrentit at Black Diamond Mines Regional Preserve, Contra Costa County, CA. Photo by Phred M. Benham. c) Wrentit distribution in western North America. Shading of magenta reflects wrentit population density with regions with darker shading indicative of higher densities of birds. Species distribution data accessed January 2024 from eBird (https://science.ebird.org).

from the Central Valley of California (Fig. 1). Five subspecies are recognized (Geupel and Ballard 2023). In Oregon two subspecies are found: *C. f. phaea* is found along the entire Oregon coast and *C. f. margra* was recently described from south-central, interior Oregon (Browning 1992). In coastal California, *C. f. rufula* is found from the Oregon border south to the San Francisco Bay Area and is replaced by *C. f. fasciata* from the San Francisco Bay Area south to San Luis Obispo County. *C. f. henshawi* is the most widespread subspecies and can be found throughout interior California from Oregon to northern Mexico. Subspecies differences relate primarily to morphological and plumage characters with birds becoming darker in more humid areas (Grinnell and Miller 1944; Bowers 1960). Phylogeographic work based on mitochondrial DNA found little agreement between subspecies limits and genetic differentiation, with the authors hypothesizing a post-glacial expansion from a southern California refugia (Burns and Barhoum 2006).

Wrentits are non-migratory and have among the lowest dispersal capabilities of any North American bird species. Juvenile dispersal from the natal site is typically less than 400 m (Baker et al. 1995) and they can be highly reluctant to cross large open spaces between scrub patches. Not surprisingly, the species remains absent from many nearshore islands in California (e.g. Channel Islands; Johnson 1972) and has never been observed on the Farallon Islands (27 miles off the coast of San Francisco) despite long-term banding and monitoring efforts on the islands detecting over 375 species (Pyle and Henderson 1991). Also unusual for many temperate North American songbirds, wrentits form lifelong pair-bonds, co-defend territories, and both sexes share in nest building and incubation (Erickson 1938; Geupel and Ballard 2023). While the wrentit may be unique among temperate North American birds, other members of the Paradoxornithidae also tend to be sedentary (with some elevational movements) and exhibit a high degree of biparental nest building, incubation, and care among the studied species (Billerman 2023).

The wrentit remains a common species throughout much of its distribution and is classified as of least concern (BirdLife International 2021). Breeding bird survey data suggest that wrentits have declined at a rate of 0.5% per year between 1966–2021 (Ziolkowski Jr et al. 2022). Recent trends based on eBird data suggest that coastal populations may be increasing while interior populations are declining (Fink et al. 2023). Much of the species' distribution is centered on chaparral habitats in California. Chaparral is characterized by high levels of endemism and is a dominant vegetation type within the California Floristic Province. Chaparral is under threat from increasingly frequent wildfires, climate change, and urbanization (Underwood et al. 2018). Understanding patterns of genetic diversity within iconic chaparral species like the wrentit will be critical for the future conservation and management of this vital habitat. For these reasons the wrentit has been included in the California Conservation Genomics Project (CCGP; Shaffer et al. 2022). For the CCGP, we generated a highly contiguous genome of the wrentit using PacBio

HiFi long-read and Omni-C chromatin-proximity sequencing data along with ribonucleic acid sequencing (RNA-seq) of a diverse set of tissues to aid in functional annotation of the wrentit genome. Here we present this high-quality genome and annotation for use by the research community.

Methods

Biological materials

We sampled a female wrentit (*C. fasciata rufula*) on 14 October 2020 in the Cow Mountain Recreation Management Area, Mayacamas Mountains, Lake Co., California (39.09208°N; 123.08350°W) at 2200 feet in elevation. The individual was collected with approval of the California Department of Fish and Wildlife (permit #: SCP-458) and the U.S. Fish and Wildlife Service (permit #: MB153526–0). The bird was captured and euthanized using methods approved by the University of California, Berkeley IACUC (AUP-2016-04-8665-1). A voucher specimen has been deposited at the Museum of Vertebrate Zoology, Berkeley, CA (https://arctos.database.museum/guid/MVZ:Bird:193981). We obtained liver tissue from the specimen for PacBio HiFi and Omni-C library preparation and sequencing, and we also collected tissue samples for RNA-seq from the following 13 organs: fat, brain, eye, gizzard, heart, intestine, kidney, liver, lung, muscle, ovary, spleen, and tongue.

PacBio HiFi library preparation

High molecular weight (HMW) genomic DNA (gDNA) was extracted from 28 mg of liver tissue using the Nanobind Tissue Big DNA kit according to the manufacturer's instructions (Pacific BioSciences—PacBio, Menlo Park, CA). The DNA purity was estimated using absorbance ratios (260/280 = 1.81 and 260/230 = 2.26) measured on the NanoDrop ND-1000 spectrophotometer. The final DNA yield (13 μg) was quantified using the Quantus Fluorometer (QuantiFluor ONE dsDNA Dye assay; Promega, Madison, WI). The size distribution of the HMW DNA was estimated using the Femto Pulse system (Agilent, Santa Clara, CA), and we found that 53% of the fragments were 40 kilobases (Kb) or longer.

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 gDNA was sheared to a target DNA size distribution between 15 Kb–20 Kb using a 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, and 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; Cat #BLF7510/HPE7510) to collect fragments greater than 7 Kb–9 Kb. The 15 Kb–20 Kb average HiFi SMRTbell library was sequenced at UC Davis DNA Technologies Core (Davis, CA) using two 8 M SMRT cells (PacBio, Cat #101–389-001), Sequel II sequencing chemistry 2.0, and 30-hour movies each on a PacBio Sequel IIe sequencer.

Omni-C library preparation and sequencing

The Omni-C library was prepared using the Dovetail Omni-C Kit (Dovetail Genomics, Scotts Valley, CA) according to the manufacturer's protocol with slight modifications. First, specimen 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 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 a NEB Ultra II DNA Library Prep kit (New England Biolabs, Ipswich, MA) with an Illumina compatible y-adaptor. Biotin-containing fragments were then captured using streptavidin beads. The post capture product was split into two replicates prior to PCR enrichment to preserve library complexity with each replicate receiving unique dual indices. The library was sequenced at Vincent J. Coates Genomics Sequencing Lab (Berkeley, CA) on an Illumina NovaSeq 6000 platform (Illumina, San Diego, CA) to generate approximately 100 million 2 × 150 bp read pairs per gigabase (Gb) of genome size.

Transcriptome library preparation and sequencing

Total RNA from 13 tissues (lung, gizzard, heart, brain, ovary, spleen, muscle, fat, intestine, eye, tongue, liver, kidney) were extracted using a Qiagen RNeasy Mini Kit (Qiagen, Netherlands), according to the manufacturer's protocol. RNA libraries were then prepared using the KAPA messenger RNA (mRNA) HyperPrep Kit (Roche, Switzerland) according to the manufacturer's protocol. Libraries were sequenced with 100 bp reads on an Illumina NovaSeq 6000 platform (Illumina, San Diego, CA), to generate approximately 50 M reads per library.

Nuclear genome assembly

We assembled the genome of a *C. fasciata* individual following the CCGP assembly pipeline Version 5.0, outlined in Table 1 listing the tools and non-default parameters used in the assembly process. We removed the remnant adapter sequences from the PacBio HiFi dataset using HiFiAdapterFilt (Sim et al. 2022) and generated an initial phased assembly using HiFiasm (Cheng et al. 2021; Cheng et al. 2022) in HiC mode with the filtered PacBio HiFi reads and the OmniC short-reads. We then aligned the Omni-C data to each assembly following the Arima Genomics Mapping Pipeline (https://github.com/ArimaGenomics/mapping_pipeline) and then scaffolded the assemblies with SALSA (Ghurye et al. 2017; Ghurye et al. 2019).

The assemblies were manually curated by 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 (Open2C et al. 2023). Subsequently, we generated multi-resolution Omni-C matrices

Table 1. Assembly pipeline and software used.

Assembly	Software and any non-default options	Version
Filtering PacBio HiFi adapters	HiFiAdapterFilt	Commit 64d1c7b
K-mer counting	Meryl (k = 21)	1
Estimation of genome size and heterozygosity	GenomeScope	2
De novo assembly (contiging)	HiFiasm (Hi-C Mode, –primary, output p_ctg.hap1, p_ctg.hap2)	0.16.1-r375
Scaffolding		
Omni-C data alignment	Arima Genomics Mapping Pipeline	Commit 2e74ea4
Omni-C Scaffolding	SALSA (-DNASE, –i 20, –p yes)	2
Gap closing	YAGCloser (–mins 2 -f 20 -mcc 2 -prt 0.25 -eft 0.2 -pld 0.2)	Commit 0e34c3b
Omni-C contact map generation		
Short-read alignment	BWA-MEM (-5SP)	0.7.17-r1188
SAM/BAM processing	samtools	1.11
SAM/BAM filtering	pairtools	0.3.0
Pairs indexing	pairix	0.3.7
Matrix generation	cooler	0.8.10
Matrix balancing	hicExplorer (hicCorrectmatrix correct –filterThreshold –2 4)	3.6
	HiGlass	2.1.11
	PretextMap	0.1.4
	PretextView	0.1.5
Contact map visualization	PretextSnapshot	0.0.3
Manual curation tools	Rapid curation pipeline (Wellcome Trust Sanger Institute, Genome Reference Informatics Team)	Commit 4ddca450
Genome quality assessment		
Basic assembly metrics	QUAST (–-est-ref-size)	5.0.2
Assembly completeness	BUSCO (–m geno, –l aves)	5.0.0
	Merqury	2020-01-29
Contamination screening		
Local alignment tool	BLAST+ (–db nt, –outfmt '6 qseqid staxids bitscore std', –max_target_seqs 1, –max_hsps 1, –evalue 1e-25)	2.10
General contamination screening	BlobToolKit (PacBIo HiFi Coverage, NCBI Taxa ID = 190 680, BUSCODB = aves)	2.3.3
Mitochondrial assembly		
Mitochondrial genome assembly	MitoHiFi (–r, –p 90, –o 1)	2.2

with cooler (Abdennur and Mirny 2020) and balanced them with hicExplorer (Ramírez et al. 2018). We used HiGlass (Kerpedjiev et al. 2018) and the PretextSuite (https://github.com/wtsi-hpag/PretextView; https://github.com/wtsi-hpag/PretextMap; https://github.com/wtsi-hpag/PretextSnapshot) to visualize the contact maps where we identified misassemblies 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/or 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 assembly assessment and annotation

We generated k-mer counts from the PacBio HiFi reads using meryl (https://github.com/marbl/meryl) and then used the k-mer database in GenomeScope2.0 (Ranallo-Benavidez et al. 2020) to estimate genome features including genome size, heterozygosity, and repeat content. We ran QUAST (Gurevich et al. 2013) to obtain general contiguity metrics, and used BUSCO (Manni et al. 2021) with the Aves ortholog database (aves_odb10, which contains 8338 genes) to evaluate genome quality and functional completeness. 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). 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 = 80 for *C. fasciata* (Bird Chromosome Database, Chromosome number data V3.0/2022, Degrandi et al. 2020). Quality metrics for the notation were calculated on the primary assembly.

Genome annotation

Sequences generated from RNA sequencing were used for functional annotation of the wrentit genome by NCBI using the Eukaryotic Genome Annotation Pipeline (https://www.ncbi.nlm.nih.gov/refseq/annotation_euk/process/; Thibaud-Nissen et al. 2013).

Comparison to zebra finch

We compared higher level synteny between the wrentit scaffolds and chromosomes of the zebra finch genome (*Taeniopygia guttata;* GenBank accession: GCA_003957565.4) using the JupiterPlot pipeline (Chu 2018; https://github.com/JustinChu/JupiterPlot). We mapped the longest scaffolds representing 80% of the draft assembly to zebra finch chromosomes exceeding 1 Mb in length.

Mitochondrial genome assembly

We assembled the mitochondrial genome of *C. fasciata* from the PacBio HiFi reads using the reference-guided pipeline MitoHiFi (Allio et al. 2020; Uliano-Silva et al. 2021). The mitochondrial sequence of *Acrocephalus orientalis* (NCBI:MT081418.1) 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.

Repeat annotation

We performed de novo repeat annotation of the draft wrentit reference assembly using the program RepeatModeler2 with the ltrstruct option selected to improve identification of LTR elements (Flynn et al. 2020). Transposable element (TE) sequences identified using repeat modeler were then compared against a library of avian TE elements downloaded from repbase along with recently published repeat libraries for other songbird species (Benham et al. 2023; Benham et al. 2024). We used cd-hit-est (Li et al. 2006) to merge TE sequences that were considered the same subfamily following the 80–80–80 rule (Wicker et al. 2007). Specifically, this rule considers consensus sequences to be the same subfamily if they are 80 base pairs in length and share > 80% similarity over > 80% of their length. The resulting repeat library was then used to annotate TE diversity in the genome assemblies of the wrentit assembly using RepeatMasker (Smit et al. 2015).

We next compared patterns of gaps, GC content, and repeat content across chromosomes of the wrentit assembly using a Circos plot with code modified from Wu et al. (2024). The 31 largest scaffolds were assigned to chromosomes 1,1A,2,3,4,4A,5–15,17–28, W, and Z of the zebra finch from Jupiter Plot analyses above. Across these genomic regions we estimated GC content and number of Ns in 250 Kb windows using the nuc function in bedtools (Quinlan and Hall 2010). We additionally estimated repeat content and gene density in the same 250 Kb windows using the output gff files from RepeatMasker and the NCBI annotation pipelines. Finally, we performed a series of linear regressions using base R (v4.4.1; R Core Team 2024) to assess the influence of repeat density and GC content on gene density across the genome.

Results

Sequencing data

The Omni-C and PacBio HiFi sequencing libraries generated 314.5 million read pairs and 3.1 million reads, respectively (Table 2). The latter yielded 37.3-fold coverage with N50 read length 13 939 bp; minimum read length 82 bp; mean read length 13 076 bp; maximum read length of 51 989 bp (see SupplementaryFigureS1, for read length distribution). Based on PacBio HiFi reads, Genomescope 2.0 estimated a genome size of 1.08 Gb, and a sequencing error rate of 0.196%. The k-mer spectrum shows a bimodal distribution with two major peaks at ~18- and 37-fold coverage, where lower and higher coverage peaks correspond to heterozygous and homozygous states of a diploid species, respectively (Fig. 2a).

Sequencing of mRNA libraries for lung, gizzard, heart, brain, ovary, spleen, muscle, fat, intestine, eye, tongue, liver, and kidney yielded 44 M, 59 M, 43 M, 55 M, 56 M, 62 M, 47 M, 38 M, 57 M, 48 M, 39 M, 46 M, and 44 M paired end reads, respectively.

Assembly metrics

The primary wrentit assembly comprises 1342 scaffolds totaling 1.19 Gb in length (Table 2). The assembly was highly contiguous with a contig N50 of 4.58 Mb, a scaffold N50 of 73.26 Mb, and an average number of Ns per 100 Kb of less than three (see Table 2 for detailed summary statistics). Graphical representation of quality metrics and the Hi-C contact maps for the primary assembly are shown in Fig. 2b and c, respectively (see Supplementary Figs S2 and S3 for the alternate assembly). BUSCO scores further indicated a high degree of completeness with 96.8% of the 8338 orthologs in the aves_odb10 database present in the wrentit genome, while only 2.8% of avian orthologs were found to be completely missing from the assembly. Jupiter plots confirmed that the largest scaffolds in the wrentit assembly correspond to chromosomes in the zebra finch genome (Fig. 3). However, one major translocation was detected with the terminal end of the wrentit Z chromosome mapped to the zebra finch chromosome 4A.

Repeat and gene annotation

Over 236 Mb (19.77%) of the wrentit genome was spanned by repetitive elements (Table 3). Interspersed repeats were the most common element (18.04%) Long interspersed nuclear elements (LINEs: 4.95%) and endogenous retroviral elements (4.78%) were the two most abundant classes of repeat content in the wrentit genome. Repeats were found on all chromosomes (Fig. 3a) with relatively high repeat density noted on the W chromosome. Repeat density was also positively correlated with the proportion of each chromosome found in gaps ($R^2_{adj.} = 0.51$; $F_{1,28} = 30.7$; $P < 0.001$; Supplemental Fig. S4).

The NCBI annotation pipeline identified 21 645 genes of which 16 821 were protein coding. The full annotation report can be found at NCBI: https://www.ncbi.nlm.nih.gov/refseq/annotation_euk/Chamaea_fasciata/GCF_029207755.1-RS_2024_06/. We analyzed patterns of gene density within 250 Kb windows across the genome (Fig. 3). These analyses revealed a slight but significant positive relationship between GC content and gene density ($R^2_{adj.} = 0.03$; $F_{1,4194} = 123.0$; $P < 0.001$; Fig. 3b). In contrast, a weak and negative relationship existed between repeat content and gene density ($R^2_{adj.} = 0.06$; $F_{1,4194} = 271.0$; $P < 0.001$; Fig. 3c). GC content was also significantly elevated within the smaller micro-chromosomes ($R^2_{adj.} = 0.79$; $F_{2,27} = 57.1$; $P < 0.001$; Supplemental Fig. S4).

Discussion

We present a highly contiguous genome assembly and complete annotation for the enigmatic wrentit of western North America. The contig N50 of the assembly (4.58 Mb) surpasses the 1 Mb mark, placing it among other highly contiguous

Table 2. Genome assembly metrics.

Bio Projects and Vouchers	**CCGP NCBI BioProject**					**PRJNA720569**
	Genera NCBI BioProject					**PRJNA766288**
	Species NCBI BioProject					**PRJNA777153**
	NCBI BioSample					**SAMN33386005**
	Specimen identification					**CC4436**
	NCBI Genome accessions			**Primary**		**Alternate**
	Assembly accession			**JARCOQ000000000**		**JARCOR000000000**
	Genome sequences			**GCA_029207755.1**		**GCA_029207785.1**
Genome Sequence	PacBio HiFi reads		Run	1 PACBIO_SMRT (Sequel II) run: 3.1 M spots, 40.4G bases, 27.1Gb		
			Accession			SRX21210274
	Omni-C Illumina reads		Run	2 ILLUMINA (Illumina NovaSeq 6000) runs: 314.5 M spots, 94.9G bases, 30.1Gb		
			Accession			SRX21210275–6
Genome Assembly Quality Metrics	Assembly identifier (Quality code ∗)					bChaFas1(6.7.P6.Q60.C)
	HiFi Read coverage §					37.33X
				Primary		Alternate
	Number of contigs			1695		1692
	Contig N50 (bp)			4 583 527		4 583 527
	Contig NG50 §			5 789 442		5 789 442
	Longest Contigs			31 700 060		31 700 060
	Number of scaffolds			1342		1350
	Scaffold N50			73 256 386		73 097 583
	Scaffold NG50 §			74 821 068		74 221 735
	Largest scaffold			153 300 038		153 299 962
	Size of final assembly			1 194 545 626		1 194 527 585
	Phased block NG50 §			6 374 890		5 670 717
	Gaps per Gbp (# Gaps)			296(354)		286(342)
	Indel QV (Frame shift)			41.49963142		41.49963142
	Base pair QV			60.7931		60.786 Full assembly = 60.7895
	k-mer completeness			91.8363		91.8347 Full assembly = 91.8363
	BUSCO completeness∗∗ (aves) *n* = 8338		C	S	D	F	M
		P‡	96.80%	96.20%	0.60%	0.60%	2.80%
	A‡	96.70%	96.10%	0.60%	0.60%	2.70%	
	Organelles		1 complete mitochondrial sequence			CM055129.1	

§ Read coverage and NGx statistics have been calculated based on the estimated genome size of 1.08 Gb. ‡ (P)rimary and (A)lternate assembly values. [a] Assembly quality code x.y.P.Q.C derived notation, from (Rhie et al. 2021). x = log10[contig NG50]; y = log10[scaffold NG50]; P = log10 [phased block NG50]; Q = Phred base accuracy QV (quality value); C = % genome represented by the first 'n' scaffolds, following a known karyotype for the species of 2n = 78, an estimated value of the number of chromosomes for this species based on the mode of the number of chromosome from other closely related species (Genome on a Tree—GoaT; tax_name (*C. fasciata*); Challis et al. 2023). Quality code for all the assembly denoted by primary assembly (bChaFas1.0.hap1). [b] BUSCO Scores. Complete BUSCOs(C). Complete and single-copy BUSCOs (S). Complete and duplicated BUSCOs (D). Fragmented BUSCOs (F). Missing BUSCOs (M).

avian genome assemblies (Bravo et al. 2021). The wrentit reference assembly is the third reference genome for the family Paradoxornithidae and assembly metrics suggest that contig N50 is 60× higher and scaffold N50 75× higher than the genome for the vinous-throated parrotbill (*Suthora webbiana;* Table 4). The wrentit genome assembly reported here is more comparable to the highly contiguous genome recently produced for the spectacled fulvetta (Yan et al. 2024) and two warbler species in the sister family Sylviidae. The number of protein-coding genes annotated by NCBI RefSeq also exceeded the number of genes identified through annotations from other close relatives except for the spectacled fulvetta (Table 4).

The 31 longest scaffolds in the wrentit assembly showed strong synteny with zebra finch chromosomes (Fig. 2d). A notable exception was an apparent translocation between the zebra finch chromosome 4A and the Z chromosome of the wrentit. This translocation corresponds to a neo-sex chromosome that has previously been documented in close relatives of the wrentit. The neo-sex chromosome is thought to have occurred at the base of the superfamily Sylvioidea (larks, swallows, bulbuls, leaf warblers, reed warblers, babblers) approximately 42.2 million years ago (Pala et al. 2012a; Shakya et al. 2022). The evolution and composition of this neo-sex chromosome has been well characterized in other members of the Sylvioidea such as *Acrocephalus* and *Sylvia*

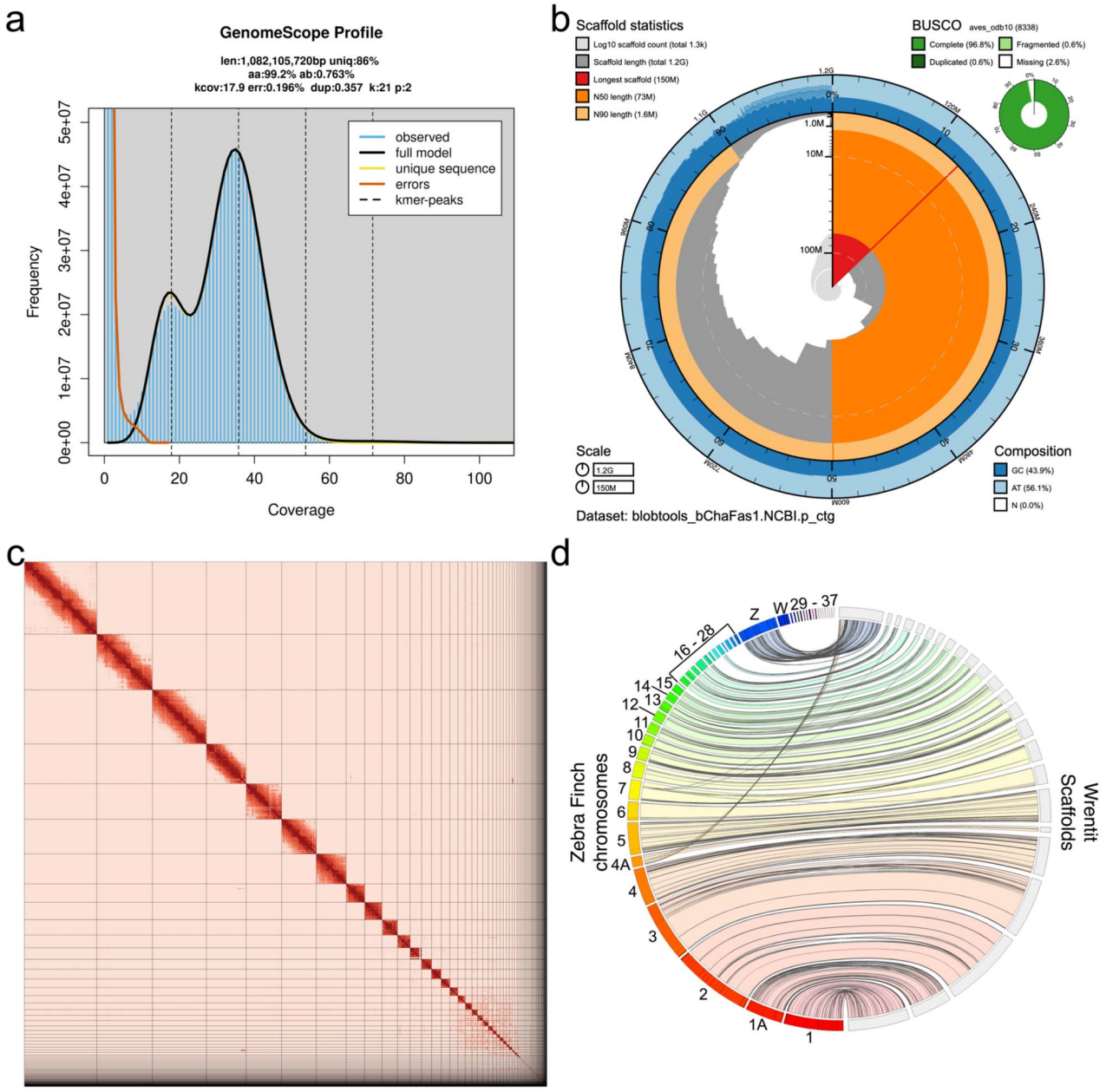

Fig. 2. Visual overview of genome assembly metrics. a) K-mer spectra output generated from PacBio HiFi data without adapters using GenomeScope2.0. The bimodal pattern observed corresponds to a diploid genome and the k-mer profile matches that of low (<1%) heterozygosity. Lower coverage and lower frequency peak reflects heterozygous k-mers, whereas the higher coverage and higher frequency correspond to homozygous k-mers. b) BlobToolKit snail plot showing a graphical representation of the quality metrics presented in Table 2 for the *C. fasciata* primary assembly (bChaFas1). 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). c) Hi-C contact maps for the primary 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. d) Jupiter plot comparing higher-level synteny and completeness between the zebra finch (*T. guttata*) genome and the wrentit draft assembly. Zebra finch chromosomes are on the left in each plot (colored) and wrentit scaffolds are on the right (light gray).

warblers (Pala et al. 2012b; Sigeman et al. 2021). Although expected given the placement of Paradoxornithidae within Sylvioidea, the generation of a highly contiguous wrentit assembly is the first to confirm the presence of this neo-sex chromosome in the parrotbill family. The generation of a wrentit assembly thus provides an important contribution to future studies of karyotype evolution in songbirds.

The levels of repeat content we found for the wrentit (19.77% or 236.1 Mb) were much higher than many previous repeat annotations of songbird genomes (Warren et al. 2010; Feng et al. 2020), but in line with more recent assemblies based on long read sequencing technology (Benham et al. 2024). In general, repeat density was elevated at the ends of scaffolds and especially high on the W chromosome (Fig. 3a).

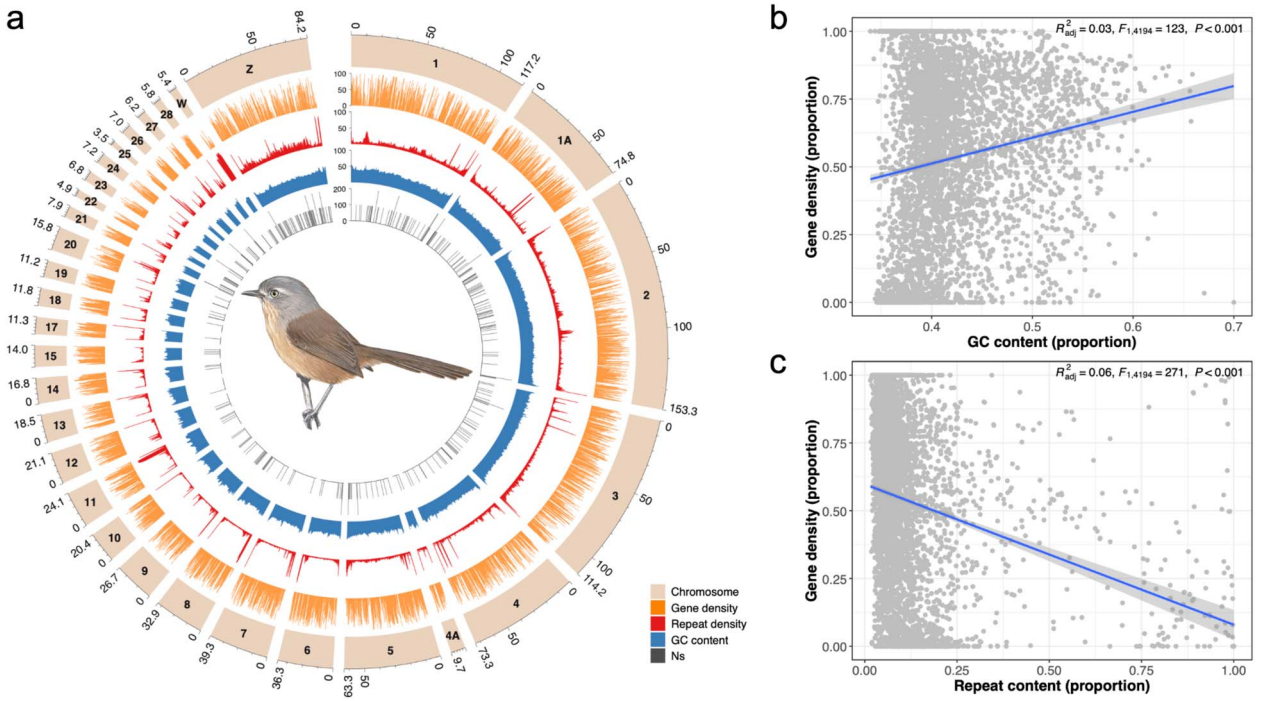

Fig. 3. a) Circos plot showing density of different genomic elements in 250 Kb windows across scaffolds syntenic with zebra finch chromosomes. From outer to inner ring: Chromosome number, gene density, repeat density, GC content, and assembly gaps. Illustration of wrentit in center reproduced with permission of Birds of the World (https://birdsoftheworld.org) and Lynx Edicions (artist: Hilary Burn). b) The relationship between GC content and gene density across the genome. Each gray dot represents a 250Kb window. c) Relationship between repeat content and gene density in 250 Kb windows across the genome.

Table 3. Summary of repeat annotation based on output from RepeatMasker.

	Number of elements	Total element length (bp)	Percentage of genome
Retroelements	400 743	139 321 160	11.66
SINEs:	161 180	23 058 413	1.93
LINEs:	152 184	59 174 346	4.95
L2/CR1/Rex	152 184	59 174 346	4.95
LTR elements:	87 379	57 088 401	4.78
Gypsy/DIRS1	98	29 197	0
Retroviral	79 758	54 145 088	4.53
DNA transposons	1500	198 358	0.02
hobo-activator	218	39 870	0
Unclassified:	85 146	75 956 874	6.36
Total interspersed repeats		215 476 392	18.04
Small RNA:	2336	2 800 568	0.23
Satellites:	3207	1 475 467	0.12
Simple repeats:	273 459	13 617 557	1.14
Low complexity:	51 172	2 832 418	0.24
Total repeat content:		236 134 814	19.77

Both patterns have been reported in other avian genomes (Kapusta and Suh 2016; Peona et al. 2021). Compared to close relatives, the amount of repeat content in the wrentit is nearly 3x the amount recently identified in the European blackcap genome (7.68%; Bours et al. 2023), but consistent with the 17.82% found in the spectacled fulvetta (Yan et al. 2024). The blackcap, fulvetta, and wrentit genomes were all generated using long-read approaches. However, the fulvetta and wrentit genome assemblies were both over 100–150 Mb longer than either the blackcap or garden warbler genomes (Table 4). This points to a possible TE expansion within Paradoxornithidae following divergence from Sylviidae. Prior analyses of avian genomes found a negative correlation between TE-rich and gene-rich regions of the genome (Kapusta and Suh 2016; Bours et al. 2023). Our analyses provide further support for this negative relationship between TE and gene density (Fig. 3c). In contrast, we find a positive correlation between gene density and GC content, a pattern also in line with previous work showing a high density of genes in GC-rich regions of the genome such as the microchromosomes (Smith et al. 2000).

The availability of a high quality wrentit genome assembly will be an important resource for studies of songbird evolution, ecology, and conservation. As the only North American

Table 4. Comparison of the wrentit assembly metrics to genome assemblies for other members of the family Paradoxornithidae and the sister family Sylviidae.

Species	Wrentit	Spectacled fulvetta	Vinous-throated parrotbill	Garden warbler	European blackcap
Scientific name	*C. fasciata*	*Fulvetta ruficapilla*	*Suthora webbiana*	*Sylvia borin*	*Sylvia atricapilla*
Family	Paradoxornithidae	Paradoxornithidae	Paradoxornithidae	Sylviidae	Sylviidae
Genbank assembly	GCA029207755.1	GCA042477295.1	GCA013397395.1	GCA014839755.1	GCA009819655.1
Assembly length (Bp)	1 194 545 626	1 208 308 458	1 050 295 543	1 045 652 180	1 066 786 587
GC %	43.93	43.04	42	42	42
# Contigs	1695	580	68 918	310	602
Contig N50 (Mb)	4.583	18.795	0.079	31.915	7.346
Contig L50	63	14	3723	10	40
# Scaffolds	1342	504	31 009	176	189
Scaffold N50 (Mb)	73.264	75.916	0.971	72.750	72.978
Scaffold L50	6	5	283	5	5
Protein-coding genes	16 821	23 774	14 265	No annotation	15 521

representative of the primarily Asian family Paradoxornithidae (Billerman 2023), the genomic resources developed for this species will be critical for further understanding the biogeographic history within this clade and the unique evolutionary trajectory of the wrentit. Second, the wrentit distribution spans a steep precipitation gradient from the mesic Oregon coast to the arid interior of Baja California. Along this gradient wrentits follow Gloger's rule, an ecogeographical rule that finds species from wetter habitats tend to have darker plumage than more arid habitats (Bowers 1960). Population genomic studies of the wrentit can now take advantage of a high-quality, annotated reference genome to identify the genomic underpinnings of these color differences and other genotypic variants that may be under selection in response to different precipitation regimes. Finally, the wrentit is an iconic species of chaparral habitats in western North America and can serve as an indicator species for other species co-occurring in this critically important habitat. Previous work using microsatellite loci found wrentit populations responded negatively to habitat fragmentation in southern California (Delaney et al. 2010; Thomassen et al. 2018). The genomic resources presented here will enable even finer grained analysis of the effects of habitat fragmentation, climate change, and urbanization on genetic diversity across the wrentit genome.

Acknowledgments

We would like to thank Joshua Ho for assistance with tissue subsampling, and Russell Campbell for contributing the wrentit photo. PacBio Sequel II library prep 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 and the UC Santa Cruz Paleogenomics Laboratory for their diligence and dedication to generating high quality sequence data. RNA extraction, library preparation and sequencing was carried out at the UCLA Technology Center for Genomics and Bioinformatics.

Supplementary material

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

Author contributions

Phred M. Benham (Conceptualization, Data curation, Formal analysis, Writing—original draft, Writing—review & editing), Carla Cicero (Data curation, Resources, Writing—review & editing), Kevin Burns (Resources, Writing—review & editing), Merly Escalona (Data curation, Formal analysis, Methodology, Visualization, Writing—review & editing), Eric Beraut (Methodology, Resources, Visualization, Writing—review & editing), Noravit Chumchim (Methodology, Resources, Writing—review & editing), Michael W Nachman (Conceptualization, Project administration, Writing—review & editing), Colin W. Fairbairn (Methodology, Resources, Writing—review & editing), William E. Seligmann (Methodology, Resources, Writing—review & editing), Mohan P.A. Marimuthu (Methodology, Resources, Writing—review & editing), Oanh Nguyen (Methodology, Resources, Writing—review & editing), Erin Toffelmier (Data curation, Project administration, Resources, Writing—review & editing), and Rauri C.K. Bowie (Conceptualization, Funding acquisition, Writing—review & editing)

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 PRJNA720569 for the CCGP. Raw sequencing data for the wrentit (NCBI BioSample SAMN33386005) are deposited in the NCBI Short Read Archive (SRA) under SRS18467718. Data generated for the annotation in this study are available under NCBI BioProject PRJNA1011365. Raw RNA sequencing data (NCBI BioSample(s) lung: SAMN40863487, gizzard: SAMN40863482, heart: SAMN40863483, brain: SAMN40863480, ovary: SAMN40863489, spleen: SAMN40863490, muscle: SAMN40863488, fat: SAMN40863479, intestine: SAMN40863484, eye: SAMN40863481, tongue: SAMN40863491, liver: SAMN40863486, kidney: SAMN40863485) are deposited in the NCBI SRA [SRR29134370, SRR29134375, SRR29134374, SRR29134380, SRR29134379, SRR29134378, SRR29134369, SRR29134381, SRR29134373, SRR29134376, SRR29134377, SRR29134371, SRR29134372], respectively). The TE library used to mask repeats is included as a supplemental file. Assembly scripts and other data for the analyses presented can be found at the following GitHub repository: www.github.com/ccgproject/ccgp_assembly.

References

Abdennur N, Mirny LA. Cooler: scalable storage for Hi-C data and other genomically labeled arrays. *Bioinformatics*. 2020;**36**:311–316. https://doi.org/10.1093/bioinformatics/btz540

Allio R, Schomaker-Bastos A, Romiguier J, Prosdocimi F, Nabholz B, Delsuc F. MitoFinder: efficient automated large-scale extraction ofmitogenomic data in target enrichment phylogenomics. *Mol Ecol Resour*. 2020;**20**:892–905.

Baker M, Nur N, Geupel GR. Correcting biased estimates of dispersal and survival due to limited study area: theory and an application using wrentits. *Condor*. 1995;**97**:663–674.

Barhoum DN, Burns KJ. Phylogenetic relationships of the wrentit based on mitochondrial cytochrome b sequences. *Condor*. 2002;**104**:740–749.

Benham PM, Cicero C, Escalona M, Beraut E, MPA M, Nguyen O, Nachman MW, RCK B. A highly contiguous genome assembly for the California quail (*Callipepla californica*). *J Hered*. 2023;**114**:418–427.

Benham PM, Cicero C, Escalona M, Beraut E, Fairbairn C, MPA M, 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. *Genome Biol Evol*. 2024;**16**:evae067.

Billerman, S. M. 2023. *Parrotbills (Paradoxornithidae), version 1.0. in birds of the world* (S. M. Billerman, B. K. Keeney, and N. D. Sly, Editors). Cornell Lab of Ornithology, Ithaca, NY, USA. https://doi.org/10.2173/bow.parado1.01

BirdLife International. Species factsheet: wrentit. *Chamaea fasciata*. 2021; Downloaded from https://datazone.birdlife.org/species/factsheet/wrentit-chamaea-fasciata on 15/05/2025

Bours A, Pruisscher P, Bascón-Cardozo K, Odenthal-Hesse L, Liedvogel M. The blackcap (Sylvia atricapilla) genome reveals a recent accumulation of LTR retrotransposons. *Sci Rep*. 2023;**13**:16471.

Bowers DE. Correlation of variation in the wrentit with environmental gradients. *Condor*. 1960;**62**:91–120.

Bravo GA, Schmitt CJ, Edwards SV. What have we learned from the first 500 avian genomes? *Annu Rev Ecol Evol Syst*. 2021;**52**:611–639. https://doi.org/10.1146/annurev-ecolsys-012121-085928

Browning MR. A new subspecies of Chamaea fasciata (wrentit) from Oregon (Aves: Timaliinae). *Proc Biol Soc Washington*. 1992;**105**:414–419.

Burns KJ, Barhoum DN. Population-level history of the wrentit (Chamaea fasciata): implications for comparative phylogeography in the California Floristic Province. *Mol Phylogenetics Evol*. 2006;**38**:117–129.

Cai T, Cibois A, Alström P, Moyle RG, Kennedy JD, Shao S, Zhang R, Irestedt M, PGP E, Gelang M, et al. Near-complete phylogeny and taxonomic revision of the world's babblers (Aves: Passeriformes). *Mol Phylogenetics Evol*. 2019;**130**:346–356.

Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL. BLAST+: architecture and applications. *BMC Bioinformatics*. 2009;**10**:1–9.

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

Challis R, Kumar S, Sotero-Caio C, Brown M, Blaxter M. Genomes on a Tree (GoaT): A versatile, scalable search engine for genomic and sequencing project metadata across the eukaryotic tree of life. *Wellcome Open Research*. 2023;**8**:24. https://doi.org/10.12688/wellcomeopenres.18658.1

Cheng H, Concepcion GT, Feng X, Zhang H, Li H. Haplotype-resolved de novo assembly using phased assembly graphs with hifiasm. *Nat Methods*. 2021;**18**:170–175.

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

Chu J. Jupiter plot: a Circos-based tool to visualize genome assembly consistency (1.0). *Zenodo*. 2018. https://doi.org/10.5281/zenodo.1241235

Degrandi TM, Barcellos SA, Costa AL, Garnero ADV, Hass I, Gunski RJ. Introducing the bird chromosome database: an overview of cytogenetic studies in birds. *Cytogenet Genome Res*. 2020;**160**:199–205. https://doi.org/10.1159/000507768

Delaney KS, Riley SPD, Fisher RN. A rapid, strong, and convergent genetic response to urban habitat fragmentation in four divergent and widespread vertebrates. *PLoS One*. 2010;**5**:e12767. https://doi.org/10.1371/journal.pone.0012767

Erickson M. Territory, annual cycle, and numbers in a population of wren-tits (*Chamaea fasciatia*). *Univ of California Publ Zool*. 1938;**42**:247–334.

Feng S, Stiller J, Deng Y, Armstrong J, Fang Q, Reeve AH, Zhang G. Dense sampling of bird diversity increases power of comparative genomics. *Nature*. 2020;**587**:252–257. https://doi.org/10.1038/s41586-020-2873-9

Fink, D., Auer T, Johnston A, Strimas-Mackey M, Ligocki S, Robinson O, Hochachka W, Jaromczyk L, Crowley C, Dunham K et al. 2023. *eBird status and trends, data version: 2022; released: 2023*. Cornell Lab of Ornithology, Ithaca, New York. https://doi.org/10.2173/ebirdst.2022

Flynn JM, Hubley R, Goubert C, Rosen J, Clark AG, Feschotte C, Smit AF. RepeatModeler2 for automated genomic discovery of transposable element families. *Proc Natl Acad Sci USA*. 2020;**117**:9451–9457. https://doi.org/10.1073/pnas.1921046117

Gambel W. Description of new and little known birds collected in upper California. *Proc Acad Natl Sci Phila*. 1845;**2**:264.

Geupel, G. R. and G. Ballard (2023). *Wrentit* (Chamaea fasciata), version 1.1. In *Birds of the world* (P. G. Rodewald, Editor). Cornell Lab of Ornithology, Ithaca, NY, USA. https://doi.org/10.2173/bow.wrenti.01.1

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, Koren S. Integrating Hi-C links with assembly graphs for chromosome-scale assembly. *PLoS Computation Biology*. 2019;**15**:e1007273. https://doi.org/10.1371/journal.pcbi.1007273

Grinnell J, Miller AH. The distribution of birds of California. *Pacific Coast Avifauna*. 1944;**27**:1–608.

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

Hosner PA, Braun EL, Kimball RT. Land connectivity changes and global cooling shaped the colonization history and diversification of new world quail (Aves: Galliformes: Odontophoridae). *J Biogeogr*. 2015;**42**:1883–1895. https://doi.org/10.1111/jbi.12555

Johnson NK. Origin and differentiation of the avifauna of the Channel Islands, California. *Condor*. 1972;**74**:295–315.

Kapusta A, Suh A. Evolution of bird genomes-a transposon's-eye view. *Ann N Y Acad Sci*. 2016. https://doi.org/10.1111/nyas.13295

Kerpedjiev P, Abdennur N, Lekschas F, McCallum C, Dinkla K, Strobelt H, Gehlenborg N. 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, . . . Jarvis, E. D. 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

Li W, Godzik A. Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics*. 2006;**22**:1658–1659.

Li H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. 2013; arXiv. Retrieved from http://arxiv.org/abs/1303.3997

Manni M, Berkeley MR, Seppey M, Simão FA, Zdobnov EM. BUSCO update: novel and streamlined workflows along with broader and deeper phylogenetic coverage for scoring of eukaryotic, prokaryotic, and viral genomes. *Mol Biol Evol*. 2021;**38**:4647–4654.

Min MS, Yang SY, Bonett RM, Vieites DR, Brandon RA, Wake DB. Discovery of the first Asian plethodontid salamander. *Nature*. 2005;**435**:87–90.

Pala I, Hasselquist D, Bensch S, Hansson B. Patterns of molecular evolution of an avian neo-sex chromosome. *Mol Biol Evol*. 2012b;**29**: 3741–3754

Pala I, Naurin S, Stervander M, Hasselquist D, Bensch S, Hansson B. Evidence of a neo-sex chromosome in birds. *Heredity*. 2012a;**108**: 264–272. https://doi.org/10.1038/hdy.2011.70

Peona V, Palacios-Gimenez OM, Blommaert J, Liu J, Haryoko T, Jønsson KA, Irestedt M, Zhou Q, Jern P, Suh A. The avian W chromosome is a refugium for endogenous retroviruses with likely effects on female-biased mutational load and genetic incompatibilities. *Philos Trans R Soc B*. 2021;**376**:20200186. https://doi.org/10.1098/rstb.2020.0186

Pyle P, Henderson RP. The birds of southeast Farallon island: occurrence and seasonal distribution of migratory species. *Western Birds*. 1991;**22**:41–84.

Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics*. 2010;**26**:841–842. https://doi.org/10.1093/bioinformatics/btq033

R Core Team. *R: a language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing; 2024: <https://www.R-project.org/>.

Ramírez F, Bhardwaj V, Arrigoni L, Lam KC, Grüning BA, Villaveces J, 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. Nature. *Communications*. 2020;**11**:1432. https://doi.org/10.1038/s41467-020-14998-3

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

Rhie A, McCarthy SA, Fedrigo O, Damas J, Formenti G, Koren S, Uliano-Silva M, Chow W, Fungtammasan A, Kim J, et al. Towards complete and error-free genome assemblies of all vertebrate species. *Nature*. 2021;**592**:737–746.

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

Shakya SB, Wang-Claypool CY, Cicero C, Bowie RCK, Mason NA. Neo-sex chromosome evolution and phenotypic differentiation across an elevational gradient in horned larks (*Eremophila alpestris*). *Mol Ecol*. 2022;**31**:1783–1799.

Sibley CG, Ahlquist JE. The relationships of the Wrentit as indicated by DNA-DNA hybridization. *Condor*. 1982;**84**:40–44.

Sigeman H, Strandh M, Proux-Wéra E, Kutschera VE, Ponnikas S, Zhang H, Lundberg M, Soler L, Bunikis I, Tarka M, et al. Avian neo-sex chromosomes reveal dynamics of recombination suppression and W degeneration. *Mol Biol Evol*. 2021;**38**:5275–5291.

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

Smit AFA, Hubley R, Green P. RepeatMasker Open-4.0. 2013–2015; <http://www.repeatmasker.org>

Smith J, Bruley CK, Paton IR, Dunn I, Jones CT, Windsor D, Morrice DR, Law AS, Masabanda J, Sazanov A, et al. Differences in gene density on chicken macrochromosomes and microchromosomes. *Anim Genet*. 2000;**31**:96–103. https://doi.org/10.1046/j.1365-2052.2000.00565.x

Thibaud-Nissen F, Souvorov A, Murphy T, Eukaryotic Genome Annotation Pipeline. The NCBI handbook. In: *Bethesda (MD): National Center for biotechnology information (US); 2013*. 2nd ed. 2013: Available from: https://www.ncbi.nlm.nih.gov/books/NBK169439/.

Thomassen HA, Harrigan RJ, Semple Delaney K, Riley SPD, Serieys LEK, Pease K, Wayne RK, Smith TB. Determining the drivers of population structure in a highly urbanized landscape to inform conservation planning. *Conserv Biol*. 2018;**32**:148–158. https://doi.org/10.1111/cobi.12969

Uliano-Silva M, Nunes JGF, Krasheninnikova K, McCarthy SA. Marcelauliano/MitoHiFi: mitohifi_v2.0 (v2.0). Zenodo. 2021; https://doi.org/10.5281/zenodo.5205678

Underwood EC, Franklin J, Molinari NA, Safford HD. Global change and the vulnerability of chaparral ecosystems. *Bull Ecol Soc Am*. 2018;**99**:1–10.

Warren WC, Clayton DF, Ellegren H, Arnold AP, Hillier LW, Künstner A, Searle S, White S, Vilella AJ, Fairley S, et al. The genome of a songbird. *Nature*. 2010;**464**:757–762. https://doi.org/10.1038/nature08819

Wicker T, Sabot F, Hua-Van A, Bennetzen JL, Capy P, Chalhoub B, Flavell A, Leroy P, Morgante M, Panaud O, et al. A unified classification system for eukaryotic transposable elements. *Nat Rev Genet*. 2007;**8**:973–982.

Wu Z, Miedzinska K, Krause JS, Pérez JH, Wingfield JC, Meddle SL, Smith J. A chromosome-level genome assembly of a free-living white-crowned sparrow (*Zonotrichia leucophrys gambelii*). *Sci Data*. 2024;**11**:86. https://doi.org/10.1038/s41597-024-02929-6

Yan C, Si S, Chen HM, Zhang YT, Liu LM, Wu F, Wang MS. A high-quality genome assembly of the spectacled Fulvetta (*Fulvetta ruficapilla*) endemic to China. *Sci Data*. 2024;**11**:1252. https://doi.org/10.1038/s41597-024-04094-2

Ziolkowski DJ Jr, Lutmerding M, Aponte VI, Hudson M-AR. North American breeding bird survey dataset 1966–2021: U.S. *Geological Survey data release*. 2022. https://doi.org/10.5066/P97WAZE5

Journal of Heredity, 2026, 117, 329–339
https://doi.org/10.1093/jhered/esaf063
Genome Resources