The First Highly Contiguous Genome Assembly for the Western bluebird (*Sialia mexicana*)

Carlos W. Esperanza[1*], Haley K. Gee[2], Renée A. Duckworth[2], Andrea D. Schreier[1]

[1]Department of Animal Science, University of California Davis, Davis, CA 95616

[2]Department of Ecology and Evolutionary Biology, University of Arizona, Tucson, AZ 85721

*Corresponding author: cesperanza@ucdavis.edu

Abstract

The western bluebird (*Sialia mexicana*) is a secondary cavity-nesting thrush that has experienced historical population declines, local extirpations, and more recent recoveries associated with nest box programs. Despite these regional successes, recent eBird estimates suggest continued range-wide declines and substantial geographic variation in population trajectories, making this species a useful system for future studies of demographic change, connectivity, and conservation genomics. However, genomic resources for western bluebirds remain limited, and no reference genome currently exists for any species in the genus *Sialia*. Here, we present the first high-quality *de novo* reference genome for *S. mexicana*. Using PacBio HiFi long-read sequencing from an adult female, we generated a highly contiguous, phased 1.3 Gb nuclear assembly with a contig N50 of 24.8 Mb and high BUSCO completeness of 98.3%. We annotated the nuclear genome using transcriptomic and protein evidence, identifying 16,656 protein-coding genes and 26,060 transcripts/protein isoforms. We also assembled a complete

~16 kb mitochondrial genome from Illumina short-read data. This reference genome provides a foundational resource for future studies of population structure, genetic diversity, connectivity, demographic history, and adaptation in western bluebirds and related taxa.

Keywords

Western bluebird; *Sialia mexicana*; reference genome; conservation genomics; population genomics.

Significance

Western bluebirds (*Sialia mexicana*) have experienced historical declines, local extirpations, and human-mediated recolonization through nest box programs, yet the genomic consequences of these demographic changes remain largely unknown. This gap is important because genome-wide data are needed to evaluate population connectivity, genetic diversity, inbreeding, and adaptation across heterogeneous landscapes. Here, we present the first high-quality reference genome for *S. mexicana* and the first highly contiguous reference genome for the genus *Sialia*. This conspecific resource supports read-mapping, variant discovery, gene annotation, comparative analyses, and future population genomic studies.

Introduction

Extrinsic pressures that reduce population size, such as habitat loss, over-exploitation, increased competition from invasive species, and climate change are major drivers of biodiversity loss worldwide (Bellard et al., 2022). Birds are particularly vulnerable to these pressures, with many species experiencing ongoing population declines globally and

across North America (Rosenberg et al, 2019; Lees et al., 2022). These declines provide an important context for investigating how the basic principles of population genetics shape the persistence of natural populations in changing environments. Demography and genetics, for example, interact at a fundamental level in determining extinction risk: small, isolated populations can experience inbreeding, potentially leading to reduced fitness, while genetic drift can randomly reduce the genetic variation that is needed for future adaptation (Lande, 1988). Thus, characterizing patterns of genetic diversity and gene flow among populations undergoing decline and demographic change is necessary for predicting the future of vulnerable species. Reference genome assemblies provide a foundational map of the features and organization of an organism's genome, enabling genome-wide analyses of variation, population structure, and demographic history. Using a high-quality conspecific assembly can improve read mapping, variant discovery, and demographic inference by reducing biases that arise when genomic data are aligned to a more distantly related species (Bentley & Armstrong, 2022).

The Western bluebird (*Sialia mexicana*) provides an ideal system for studying the genomic consequences of demographic shifts and anthropogenic changes (Fig. S1). In western North America, the destruction of grasslands and logging of forests have been major factors contributing to the decline of terrestrial bird populations over the past century (DeSante & George, 1994). Western bluebirds were once widely common in western North America but experienced substantial population declines, near extirpations, and elevational habitat shifts during the mid-20th century due to habitat fragmentation, interspecific competition, and pesticide exposure (Dawson & Bowles, 1909; Bertrand & Scott, 1979; Gilligan et al., 1994; Campbell et al., 1997; Duckworth &

Badyaev, 2007). Subsequent recolonization throughout parts of the species' historical range was largely facilitated by nest box programs (Campbell et al., 1997; Purcell et al., 1997). However, data from eBird citizen science observations suggest that this species is still undergoing declines throughout most of its breeding range, with a median change in estimated abundance of 21.8% from 2007-2022 (Fig. 1a).

Despite the species' history of decline, recolonization, and regionally variable demographic trends, genomic resources for this species remain limited. Previous genetic studies in *S. mexicana* have focused primarily on phylogenetic placement or behavioral traits, and no high-quality reference genome currently exists for any species in the genus *Sialia*. Generating a high-quality conspecific reference genome is therefore an essential first step for future population genomic analyses of genetic diversity, connectivity, hybridization dynamics, demographic history, and adaptation (Klicka et al., 2005; Charmantier et al., 2007; Ferree et al., 2008; Duckworth & Semenov, 2017; Sauve et al., 2021; Moelling & Duckworth, 2024; Veale et al., 2024). Here, we fill this resource gap by presenting the first high-quality *de novo* genome assembly for the Western bluebird generated using PacBio HiFi long-read sequencing and annotated with protein and transcriptomic evidence. A complete mitochondrial genome was generated using whole genomes sequenced via Illumina short-read sequences. This genome provides a foundational resource for future studies of population structure, connectivity, demographic history, adaptation, and conservation genomics in this species and related taxa.

Results & Discussion

Nuclear Genome Assembly

PacBio HiFi sequencing generated 122.41 Gb of sequence data, corresponding to approximately 97.4× coverage based on the *k*-mer-estimated genome size (Fig. 1b; Table 1; Fig. S2, S3). *K*-mer counting predicted a genome size of 1,256,283,126 bp, with 27.2% repetitive content and 1.1% heterozygosity (Fig. 1b). Assembly with Hifiasm produced two partially phased haplotype assemblies, arbitrarily labeled hap1 and hap2. After further post-assembly duplication purging, the final haplotype assemblies were 1.42 Gb and 1.30 Gb, respectively (Table 1; Table S1; Fig. s4). Hap2 was selected as the primary assembly because its total size was closer to the *k*-mer estimated genome size, contained fewer contigs, and had a higher contig N50, higher *k*-mer QV completeness, and higher BUSCO completeness than hap1 (Table 1; Table S1; Fig. s5). The final primary assembly was 1,302,139,236 bp and consisted of 168 contigs, with a contig N50 of 24.78 Mb, a base-pair QV of 77.49, and 98.0% single copy BUSCOs using the aves_odb10 dataset (Table 1, Fig. 1c).

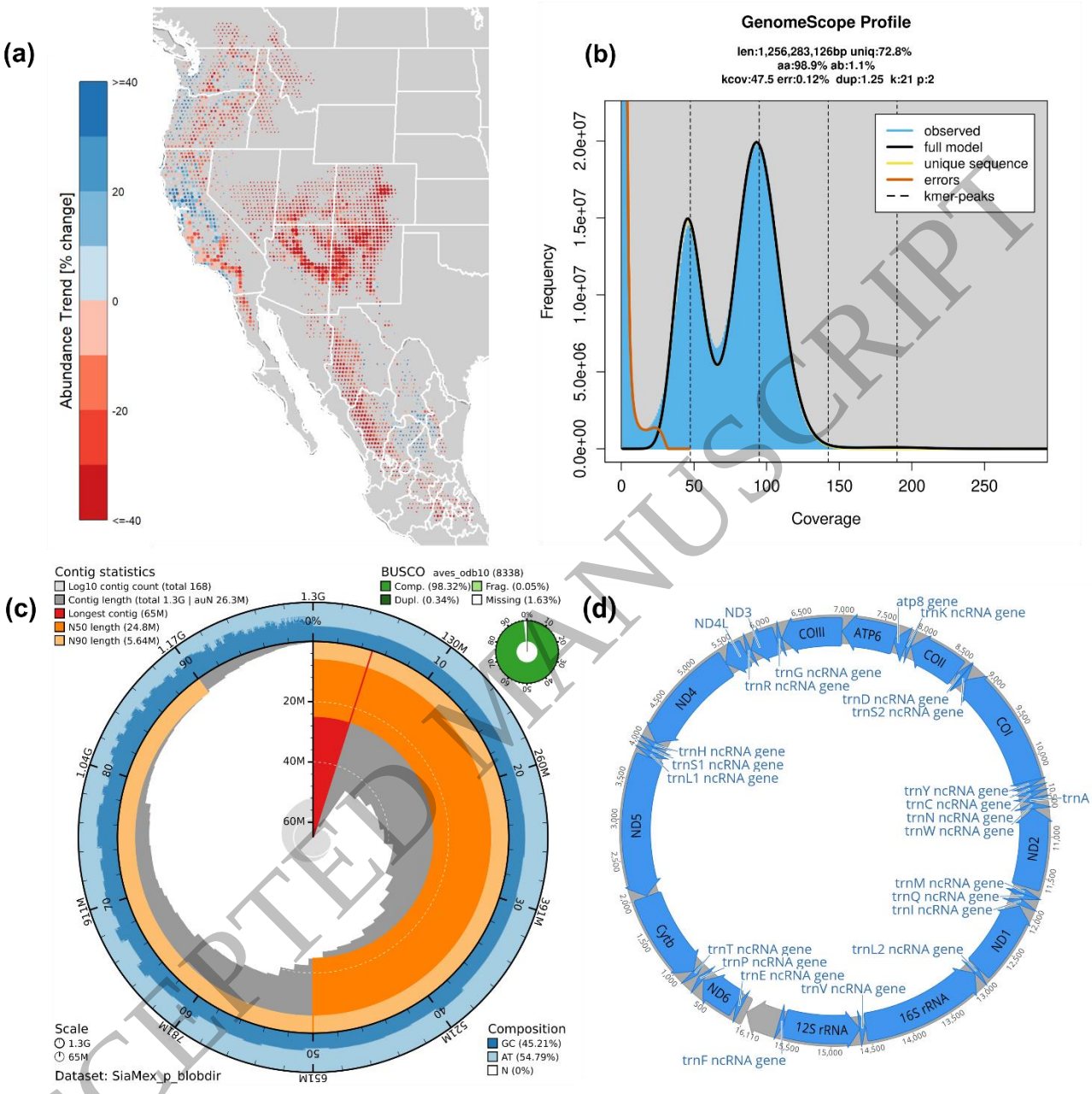

Figure 1: Contig-level genome assembly of *Sialia mexicana*. (a) Estimated changes in Western bluebird abundance from 2007 to 2022 across the species' breeding range. Red areas indicate declining abundance, whereas blue areas indicate increasing abundance. (b) GenomeScope *k*-mer profile (*k*=21) showing the estimated genome size, % heterozygosity, and % repetitive content. The histogram depicts a clear diploid *k*-mer

profile with two distinct peaks representing heterozygous *k*-mers. (c) Snail plot for haplotype 2 showing quality metrics (N50, N90, GC%) and BUSCO completeness. (d) Complete mitochondrial assembly consisting of 16,110 bp, 13 protein-coding genes, 22 tRNAs, and 2 rRNAs.

Alt text: Four-panel summary of the Western bluebird genome resource. Panel a is a map of western North America showing geographic variation in abundance trends, with widespread red declining regions and fewer blue increasing regions. Panel b is a bimodal *k*-mer frequency plot with peaks near approximately 50-fold and 100-fold coverage. Panel c is a circular snail plot showing a highly contiguous primary genome assembly and high BUSCO completeness. Panel d is a circular mitochondrial genome map with labeled protein-coding and ribosomal RNA genes.

Repetitive Content Annotation & Masking

Repeat annotation identified 368,840,933 bp of repetitive sequence in the primary assembly, representing 28.33% of the genome (Table S2). This estimate was broadly consistent with the *k*-mer-based repetitive content estimate of 27.2% (Fig. 1b). The largest fraction of annotated repeats consisted of unclassified elements, which accounted for 254,242,511 bp, or 19.52% of the assembly. Among classified repeats, retroelements represented 5.49% of the genome, including LTR elements (2.78%) and LINEs (2.71%), while SINEs were rare (0.002%) (Table S2). DNA transposons accounted for 3.31% of the assembly, with Tc1-IS630-Pogo elements representing the most abundant classified DNA transposon group. Small RNA-associated repeats made up only 0.10% of the genome. Compared with other published avian genomes, the Western bluebird assembly

falls among the more repeat-rich species included in our comparison and is similar in repetitive content to Black redstart (*Phoenicurus ochruros*; 30.58%) and Bell's sparrow (*Artemisiospiza belli*; 31.2%) (Benham et al., 2024; Ghimire et al., 2025) (Fig. S6). In contrast, the closely related thrushes Bicknell's thrush (*Catharus bicknelli*) and gray-cheeked thrush (*C. minimus*) have substantially lower repetitive content despite belonging to a nearby passerine lineage (Termignoni-Garcia et al., 2022). This elevated repeat content is consistent with the relatively large genome size of Western bluebird among the sampled avian genomes, supporting the broader observation that repeat content tends to increase with genome size in birds (Kapusta et al., 2017).

Structural & Functional Gene Annotation

BRAKER3 predicted 16,656 protein-coding genes, representing 26,060 transcript/protein isoforms (Table 1). The final structural annotation contained 326,452 exon features, 326,452 CDS features, and 300,392 intron features. The protein and CDS FASTA files each contained 26,060 sequences. BUSCO analysis recovered 93.5% of aves_odb10 orthologues as single copy, with 0.6% duplicated, 0.5% fragmented, and 5.4% missing (Table 1).

Functional annotation using DIAMOND against Swiss-Prot assigned best-hit similarity annotations to 23,408 predicted protein isoforms (Table 1). InterProScan identified domain, family, or functional feature matches for 25,339 of 26,060 predicted proteins (97.2%), producing a broad set of InterPro, Gene Ontology, and pathway annotations (Table 1).

Orthology-based functional annotation with eggNOG-mapper assigned annotations to 24,075 predicted protein isoforms. These annotations included eggNOG orthologous groups, seed orthologs, COG functional categories, putative protein descriptions, preferred gene names, Gene Ontology terms, KEGG orthologs, KEGG pathway/module assignments, and Pfam annotations where available (Table S3).

Mitochondrial Genome

The final draft *S. mexicana* mitochondrial genome assembly was 16,110 bp in length and included 13 protein-coding genes, 22 tRNA genes, and two ribosomal RNA genes, consistent with the typical structure of vertebrate mitogenomes (Fig. 1c).

Table 1: Genome assembly and annotation metrics

HiFi read coverage	97.4X
Hifiasm assembly	**Primary (hap2)**
Assembly metrics	–
Number of contigs	168
Contig N50 (bp)	24,780,003
Contig N90 (bp)	5,638,089
Contig NG50	26,963,476
Contig L50	17
Contig L90	66
auN	26,315,297.40
auNG	27,275,842.90
GC (%)	45.21
Largest contig	65,039,750
Size of final assembly	1,302,139,236
Base pair QV	77.49
k-mer completeness (%)	85.84

BUSCO completeness (aves) n=8338	C:98.3% [S:98.0%, D:0.3%], F:0.1%, M:1.6%
Annotation metrics	–
Protein-coding genes	16,656
Transcripts	26,060
Predicted protein sequences	26,060
BUSCO completeness (aves) n=8338	C:94.1% [S:93.5%, D:0.6%], F:0.5%, M:5.4%
CDS/Exon features	326,452
Unique GO terms	9,512
Unique pathway annotations	15,240
Proteins with Swiss-Prot similarity	23,408/26,060 (89.8%)
Proteins with InterProScan matches	25,339/26,060 (97.2%)
Proteins with eggNOG annotations	24,075/26,060 (92.4%)

Materials and Methods

See supplementary file for complete methods.

Sample Collection, Library Prep, and Sequencing

A single adult female western bluebird was captured (USGS banding permit #23555, UC Davis IACUC protocol #24508) at the Putah Creek Nestbox Highway in California, USA, for PacBio HiFi genome sequencing. High molecular weight DNA extracted from blood was used for library preparation and long-read sequencing at the UC Davis DNA Technologies and Expression Analysis Core. Additional Illumina whole-genome resequencing data from 10 male western bluebirds collected between 2001 and 2014 from multiple western Montana populations (see Duckworth & Semenov, 2017 and Duckworth et al., 2017 for detailed site descriptions) were used for mitochondrial genome assembly and validation analyses. For transcriptome-assisted genome annotation, 25

embryonic brain tissue samples were collected from western Montana nestbox populations during May–June 2024. RNA extracted from embryonic brain tissue was used for RNA sequencing and downstream evidence-guided gene annotation.

Nuclear Genome Assembly and Assessment

K-mer frequency spectra were generated using meryl and visualized with GenomeScope2.0 to estimate genome characteristics prior to assembly (Ranallo-Benavidez et al., 2020). PacBio HiFi reads were assembled using Hifiasm v0.24.0 with built-in duplication purging (Cheng et al., 2022), producing a primary assembly and two phased haplotype assemblies. Remaining duplicated haplotigs were removed using purge_dups v1.2.5 (Guan et al., 2020). Assembly contamination was evaluated using Kraken2 and BLASTn (Camacho et al., 2009; Wood et al., 2019). Assembly statistics were calculated using QUAST (Gurevich et al., 2013), and genome completeness was assessed with BUSCO v5.7.1 against the aves_odb10 lineage dataset (Manni et al., 2021). Base-level accuracy and k-mer completeness were evaluated using Merqury (Rhie et al., 2021).

Repeat Annotation and Gene Prediction

Repetitive elements were identified using a combined de novo and structure-based workflow with RepeatModeler v2.0.8 (Flynn et al., 2020) and EDTA (Ou et al., 2019). Repeat libraries generated by both pipelines were combined with RepBase sequences to construct a custom repeat library, and genome masking was performed using RepeatMasker v4.2.3 with the RMBlast search engine (Smit et al., 2013).

Protein-coding genes were predicted using BRAKER3 in ETP mode, integrating RNA-seq evidence, vertebrate protein homology evidence from OrthoDB v11, and ab initio gene prediction (Brůna et al., 2024; Gabriel et al., 2024). Predicted protein sequences were functionally annotated using complementary similarity-, domain-, and orthology-based approaches. DIAMOND v2.1.7 was used to search predicted proteins against Swiss-Prot (UniProt release 2024_07) with an e-value threshold of 1×10^{-5} (Buchfink et al., 2021). Conserved protein domains and associated functional annotations were identified using InterProScan v5.77-108.0 with InterPro accession lookup, Gene Ontology term assignment, and pathway annotation enabled. Orthology-based annotation was performed using eggNOG-mapper v2.1.13 in DIAMOND search mode against eggNOG v5.0.2 to assign orthologous groups, COG categories, Gene Ontology terms, and KEGG annotations (Cantalapiedra et al., 2021).

Mitochondrial Genome Assembly and Annotation

Illumina resequencing reads were mapped to a reference avian mitochondrial genome (NCBI accession OR757050) using BWA v0.7.17 and SAMtools v1.19.2 to identify the individual with the highest mitochondrial coverage (Li & Durbin, 2009; Danecek et al., 2021). The mitochondrial genome was assembled using GetOrganelle v1.7.7.1 (Jin et al., 2020) and annotated using MITOS2 through the Galaxy platform (Donath et al., 2019; The Galaxy Community, 2024). Circular mitochondrial genome visualization was performed in Geneious Prime 2026.1.1.

Supplementary Material

Please see the supplementary file for additional methods, tables, and figures.

Author Contributions

Conceptualization: C.W.E. and A.D.S.; sample collection and preparation: C.W.E., H.G. and R.A.D; data curation: C.W.E., H.G. and R.A.D; formal analysis: C.W.E.; visualization: C.W.E.; funding acquisition: C.W.E., R.A.D., and A.D.S.; investigation: C.W.E.; methodology: C.W.E., A.D.S.; project administration: C.W.E.; supervision: R.A.D., A.D.S.; writing—original draft: C.W.E.; writing—review & editing: C.W.E., H.G., R.A.D., and A.D.S.

Funding

This work was funded by the Eugene Cota-Robles Fellowship and Jastro & Shields Graduate Research Award from the University of California, Davis. NSF grants DEB-918095 and DEB-1350107 to R.A.D. RNA sequencing was funded by grants from the University of Arizona Core Facilities Pilot Program to R.A.D. The University of Arizona Graduate and Professional Student Council and University of Arizona EEB Silliman Memorial award to H.K.G.

Conflict of Interest

The authors declare no conflict of interest.

Data Availability

PacBio sequencing and data are available through NCBI BioProject PRJNA1471246, and RNA-seq data are available under PRJNA1479381. The mitochondrial genome is deposited under accession PZ467643. The final primary and alternate assemblies can be found on GenBank under accession numbers JCAOEU000000000 and JCAXWL000000000, respectively. Annotation files are available on Figshare at https://doi.org/10.6084/m9.figshare.32493435 , and assembly code is available at

https://github.com/CarlosEsperanza/WEBL_genome_assembly.git.

Acknowledgements

The sequencing was carried by the DNA Technologies and Expression Analysis Core at the UC Davis Genome Center (RRID:SCR_017740), supported by NIH Shared Instrumentation Grant 1S10OD010786-01**.** Special thank you to the UC Davis Museum of Wildlife and Fish Biology. Thank you also to the Blackfoot-Clearwater Game Range, Paws Up Montana staff for access to sites, and the Good lab at the University of Montana, and the University of Arizona Genetics core assisted with sample storage and processing the RNAseq data.

Literature Cited

Al Arab M., et al. 2017. Accurate annotation of protein-coding genes in mitochondrial genomes. *Mol. Phylogenet. Evol.* 106:209–216. https://doi.org/10.1016/j.ympev.2016.09.024

Bellard C., Marino C., Courchamp F. 2022. Ranking threats to biodiversity and why it doesn't matter. *Nat. Commun*. 13:2616. https://doi.org/10.1038/s41467-022-30339-y

Benham P. M. et al. Remarkably High Repeat Content in the Genomes of Sparrows: The Importance of Genome Assembly Completeness for Transposable Element Discovery. *Genome Biol. Evol.* **16**, evae067 (2024). https://doi.org/10.1093/gbe/evae067

Bentley B. P., Armstrong E. E. 2022. Good from far, but far from good: The impact of a reference genome on evolutionary inference. *Mol. Ecol. Resour.* 22:12–14. https://doi.org/10.1111/1755-0998.13531

Bertrand G. A., Scott J. M. 1979. *A checklist of the birds of Oregon*. 3rd ed. Eltzroth M. S., Ramsey F. L., editors. Corvallis, OR: Oregon State University Bookstore. https://digitalcommons.usf.edu/western_birds/vol12/iss4/1

Blum M. et al. 2021. The InterPro protein families and domains database: 20 years on. *Nucleic Acids Res.* 49:D344–D354. https://doi.org/10.1093/nar/gkaa977

Brůna T., Lomsadze A., Borodovsky M. 2024. GeneMark-ETP significantly improves the accuracy of automatic annotation of large eukaryotic genomes. *Genome Res.* 34:757–768. https://doi.org/10.1101/gr.278373.123

Buchfink B. J., Reuter K., Drost H. G. 2021. Sensitive protein alignments at tree-of-life scale using DIAMOND. *Nat. Methods* 18:366–368. https://doi.org/10.1038/s41592-021-01101-x

Camacho C. et al. 2009. BLAST+: architecture and applications. *BMC Bioinformatics* 10:421. https://doi.org/10.1186/1471-2105-10-421

Campbell R. W. et al. 1997. *The birds of British Columbia. Volume 3. Passerines: Flycatchers through vireos*. Vancouver, Canada: University of British Columbia Press.

Cantalapiedra C. P., Hernández-Plaza A., Letunic I., Bork P., Huerta-Cepas J. 2021. eggNOG-mapper v2: Functional annotation, orthology assignments, and domain prediction at the metagenomic scale. *Mol. Biol. Evol.* 38:5825–5829. https://doi.org/10.1093/molbev/msab293

Charmantier A., Keyser A. J., Promislow D. E. 2007. First evidence for heritable variation in cooperative breeding behaviour. *Proc. Biol. Sci.* 274:1757–1761. https://doi.org/10.1098/rspb.2007.0012

Cheng H. et al. 2021. Haplotype-resolved de novo assembly using phased assembly graphs with hifiasm. *Nat Methods* 18:170–175. https://doi.org/10.1038/s41592-020-01056-5

Danecek P. et al. 2021. Twelve years of SAMtools and BCFtools. *GigaScience* 10:giab008. https://doi.org/10.1093/gigascience/giab008

Dawson W. L., Bowles J. H. 1909. *The birds of Washington*. Seattle, WA: The Occidental Publishing Company. https://doi.org/10.5962/bhl.title.14519

DeSante D. F., George L. T. 1994. Population trends in the landbirds of western North America. *Stud. Avian Biol.* 15:173–190. https://digitalcommons.usf.edu/sab/vol15/iss1/15

Donath A. et al. 2019. Improved annotation of protein-coding gene boundaries in metazoan mitochondrial genomes. *Nucleic Acids Res.* 47:10543–10552. https://doi.org/10.1093/nar/gkz833

Duckworth R. A., Badyaev A. V. 2007. Coupling of dispersal and aggression facilitates the rapid range expansion of a passerine bird. *Proc Natl Acad Sci U S A* 104:15017–15022. https://doi.org/10.1073/pnas.0706174104

Duckworth R. A., Hallinger K. K., Hall N., Potticary A. L. 2017. Switch to a novel breeding resource influences coexistence of two passerine birds. *Frontiers in Ecology and Evolution* 5:72. https://doi.org/10.3389/fevo.2017.00072

Duckworth R. A., Semenov G. A. 2017. Hybridization associated with cycles of ecological succession in a passerine bird. *The American Naturalist* 190:E94–E105. doi: 10.1086/693160

Ferree E. D. et al. 2008. Development and cross-species testing of western bluebird (*Sialia mexicana*) microsatellite primers. *Mol. Ecol. Resour.* 8:1348–1350. https://doi.org/10.1111/j.1755-0998.2008.02265.x

Flynn J. M. et al. 2020. RepeatModeler2 for automated genomic discovery of transposable element families. *Proc. Natl. Acad. Sci. U.S.A.* 117:9451–9457. https://doi.org/10.1073/pnas.1921046117

Gabriel L. et al. 2024. BRAKER3: Fully automated genome annotation using RNA-seq and protein evidence with GeneMark-ETP, AUGUSTUS, and TSEBRA. *Genome Res.* 34:769–777. https://doi.org/10.1101/gr.278090.123

Galaxy Community. 2024. The Galaxy platform for accessible, reproducible, and collaborative data analyses: 2024 update. *Nucleic Acids Res.* 52:W83–W94. https://doi.org/10.1093/nar/gkae410

Ghimire P. et al. A complete reference genome assembly and annotation of the Black Redstart (*Phoenicurus ochruros*). *Sci Data* 12:1885 (2025). https://doi.org/10.1038/s41597-025-06156-5

Gilligan J. D., Rogers M., Smith A., Contreras A. 1994. *Birds of Oregon: Status and Distribution*. Cinclus Publications, McMinnville, OR, USA.

Guan D. et al. 2020. Identifying and removing haplotypic duplication in primary genome assemblies. *Bioinformatics* 36:2896–2898. https://doi.org/10.1093/bioinformatics/btaa025

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

Jin J. J. et al. 2020. GetOrganelle: a fast and versatile toolkit for accurate de novo assembly of organelle genomes. *Genome Biol.* 21:241. https://doi.org/10.1186/s13059-020-02154-5

Jones P. et al. 2014. InterProScan 5: genome-scale protein function classification. *Bioinformatics* 30:1236–1240. https://doi.org/10.1093/bioinformatics/btu031

Kapusta A., Suh A., Feschotte C. 2017. Dynamics of genome size evolution in birds and mammals. *Proc. Natl. Acad. Sci.* 114:E1460–E1469. https://doi.org/10.1073/pnas.1616702114

Klicka J., Voelker G., Spellman G. M. 2005. A molecular phylogenetic analysis of the "true thrushes" (Aves: Turdinae). *Mol. Phylogenet. Evol.* 34:486–500. https://doi.org/10.1016/j.ympev.2004.10.001

Lande R. 1988. Genetics and demography in biological conservation. *Science* 241:1455–1460. https://doi.org/10.1126/science.3420403

Lees A. C. et al. 2022. State of the world's birds. *Annu. Rev. Environ. Resour.* 47:231–260. https://doi.org/10.1146/annurev-environ-112420-014642

Li H. 2018. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics* 34:3094–3100. https://doi.org/10.1093/bioinformatics/bty191

Li H., Durbin R. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 25:1754–1760. https://doi.org/10.1093/bioinformatics/btp324

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. *Mol. Biol. Evol.* 38:4647–4654. https://doi.org/10.1093/molbev/msab199

Moelling M. H., Duckworth R. A. 2024. Climate change reduces the tension of conflicting selection pressures on breeding date in a passerine bird. *Proceedings of the Royal Society B: Biological Sciences* 291. https://doi.org/10.1098/rspb.2024.0959

Ou S. et al. 2019. Benchmarking transposable element annotation methods for creation of a streamlined, comprehensive pipeline. *Genome Biol.* 20:275. https://doi.org/10.1186/s13059-019-1905-y

Purcell K. L., Verner J., Oring L. W. 1997. A comparison of the breeding ecology of birds nesting in boxes and tree cavities. *Auk* 114:646–656. https://doi.org/10.2307/4089284

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

Rhie A. et al. 2021. Towards complete and error-free genome assemblies of all vertebrate species. *Nature* 592:737–746. https://doi.org/10.1038/s41586-021-03451-0

Rosenberg K. V. et al. 2019. Decline of the North American avifauna. *Science* 366:120–124.

Sauve D., Dale C. A., Tigano A., Ratcliffe L. M., Friesen V. L. 2021. Do candidate genes for migration and behavior explain migratory variation in bluebirds (*Sialia* spp.)? *Wilson J. Ornithol.* 132:820–829. https://doi.org/10.1676/19-00120

Smit A. F. A., Hubley R., Green P. 2013. RepeatMasker Open-4.0. http://www.repeatmasker.org.

Termignoni-Garcia F., Kirchman J. J., Clark J., Edwards S. V. 2022. Comparative Population Genomics of Cryptic Speciation and Adaptive Divergence in Bicknell's and Gray-Cheeked Thrushes (Aves: *Catharus bicknelli* and *Catharus minimus*). *Genome biology and evolution* 14(1):evab255. https://doi.org/10.1093/gbe/evab255

UniProt Consortium. 2015. UniProt: a hub for protein information. *Nucleic Acids Res.* 43:D204–D212. https://doi.org/10.1093/nar/gku989

Veale A., Reudink M. W., Burg T. M. 2024. Neutral markers reveal complex population structure across the range of a widespread songbird. *Ecol. Evol.* 14:e11638. https://doi.org/10.1002/ece3.11638

Wood D. E., Lu J., Langmead B. 2019. Improved metagenomic analysis with Kraken 2. *Genome Biol.* 20:257. https://doi.org/10.1186/s13059-019-1891-0