

High-quality *de novo* genome assembly for the Galápagos endemic lava gull using Oxford Nanopore Technologies

Jessica A. Martin,^{1,2} James B. Henderson,² Vera de Ferran,^{1,2} Gabriela Pozo ,³ Alice Skehel,⁴ Athena Lam,² John P. Dumbacher ,^{1,2} Jaime A. Chaves^{1,5,6,*}

¹Department of Biology, San Francisco State University, San Francisco, CA 94132-1722, United States

²Institute for Biodiversity Science and Sustainability, California Academy of Sciences, San Francisco, CA 94118, United States

³Laboratorio de Biotecnología Vegetal, Colegio de Ciencias Biológicas y Ambientales, Universidad San Francisco de Quito USFQ, Quito 170901, Pichincha, Ecuador

⁴Faculty of Science and Engineering, University of the Sunshine Coast, Sippy Downs, Queensland 4556, Australia

⁵Laboratorio de Biología Evolutiva, Instituto Biósfera, Colegio de Ciencias Biológicas y Ambientales, Universidad San Francisco de Quito, Quito 170901, Pichincha, Ecuador

⁶Galapagos Science Center, Universidad San Francisco de Quito USFQ, Islas Galápagos 200101, Ecuador

*Corresponding author: Department of Biology, San Francisco State University, San Francisco, CA 94132-1722, United States. Email: jachaves@sfsu.edu

High-quality reference genomes permit deeper investigation into species' evolution and provide insight into species management and conservation. Next-generation sequencing technologies, such as Oxford Nanopore Technologies, allow researchers to generate high-accuracy long-read genetic data in real-time from anywhere in the world, increasing accessibility to sequence data without exportation of the sample. The lava gull (*Leucophaeus fuliginosus*), an endemic bird species of the Galápagos archipelago, is the world's rarest gull with an estimated population of 300 to 600 individuals. Little genetic research has been done on this species due to its solitary nature and small population size. Here we present a chromosome-level reference genome assembly of an adult female lava gull, generated using ultra-long reads from the Oxford Nanopore Ultra-Long DNA Sequencing Kit and a PromethION 2 Solo device. Initial sequencing generated 1.78 million reads, consisting of 29.6 gigabases (Gbp), with a mean Q-score of 17.8 at an average 22.5x coverage. Our final assembly has a total length of 1.31 Gbp, with 450 scaffolds, and a scaffold N50 of 85.1 Mbp and contig N50 of 42.8 Mbp. The generation of a high-quality whole genome for the lava gull is an important step for investigation into the species' phylogeography and population genetics.

Keywords: Laridae; lava gull; *Leucophaeus fuliginosus*; whole-genome sequencing; genome assembly; Oxford Nanopore sequencing; Galápagos

Introduction

High-quality whole genomes permit deeper investigation of a species' evolution through intraspecific and interspecific comparison (Ryder 2005), including studies in phylogeography and population genetics, which are critical for the assessment, monitoring, and management of biodiversity (Brito and Edwards 2009; Genome 10K Community of Scientists 2009; Romanov et al. 2009; Paez et al. 2022). Publicly accessible reference genomes enable researchers to align newly collected sequences with a species-specific reference (Pozo et al. 2024), which is crucial for a wide range of biological studies key to the conservation of a species (Brandies et al. 2019). This includes genome-wide discovery of single-nucleotide polymorphisms that support investigations into historical demography and population structure (Vlček et al. 2025), examinations into genetic diversity, genomic erosion, inbreeding, and genetic load (Díez-del-Molino et al. 2018; Van Der Valk et al. 2019), as well as disease susceptibility and other conservation-related genetic traits (Fuentes-Pardo and Ruzzante 2017; Díez-del-Molino et al. 2018). As sequencing technologies improve in portability, such as with Oxford Nanopore Technologies (ONT), they enable real-time genomic

data generation in remote, biodiverse regions without sequencing infrastructure (Watsa et al. 2020). This helps overcome challenges in areas where biological sample export is restricted, such as the Galápagos Islands.

The Galápagos Islands, located 1,000 kilometers off the coast of Ecuador, are renowned for their role in shaping Darwin's theory of evolution. Like other oceanic islands, the archipelago hosts many endemic species (Kier et al. 2009) that evolved from a combination of isolation, unoccupied niches, and a small number of founders. The resulting biodiversity makes the Galápagos important for global conservation efforts (Lack 1947; Snow and Grant 1988). For example, of the 199 bird species recorded, 47 are endemic (CDF, 2025), but due to their small geographic distribution and population sizes, these species are highly susceptible to human disturbances (Vitousek et al. 1997; Wiedenfeld and Jiménez-Uzcátegui 2008; Benítez-Capistros et al. 2014; Chaves 2018). Evolving in the absence of invasive predators, native species are primarily threatened by introduced rats and cats (Steadman 1986; Cruz and Cruz 1987; Carrion et al. 2011). In addition, the archipelago has experienced an exponential increase in tourism since the 1960s, which developed land for urban infrastructure and agriculture (Epler 2007; Walsh and Mena 2016).

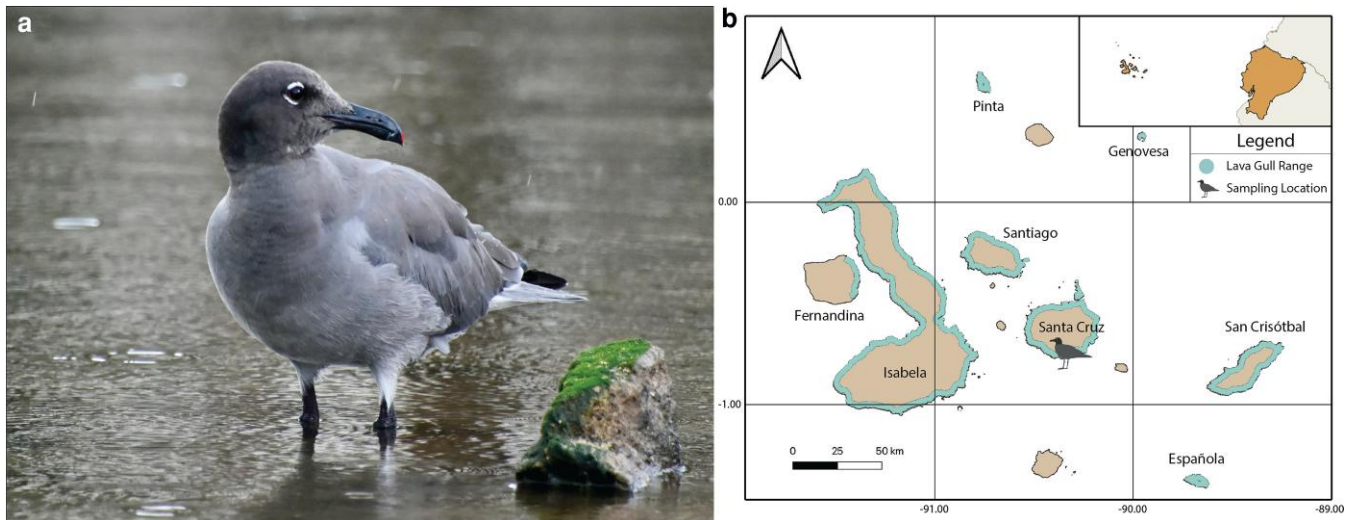


Fig. 1. Lava Gull and Galápagos Islands. a) The lava gull is the rarest gull in the world and is endemic to the Galápagos archipelago (Photo: Alice Skehel). b) The Galápagos archipelago is located 1,000 kilometers off the coast of Ecuador (upper right inset). The sample used to create the lava gull reference genome was collected from an adult female gull at Puerto Ayora on Santa Cruz Island, as indicated by the lava gull icon on the map. Contoured islands indicates the lava gull's current range (BirdLife International 2018).

One Galápagos endemic species, the lava gull (*Leucophaeus fuliginosus*), is of particular concern due to increasing habitat loss and disturbance by urban activity (Jiménez-Uzcátegui et al. 2019). First described by John Gould (Gould 1841), the lava gull is the rarest gull species in the world, with an estimated population of only 300 to 600 individuals across the archipelago (Grant et al. 2014) (Fig. 1). The IUCN currently lists lava gulls as Vulnerable because their population appears to be stable, but they face a number of potential threats (BirdLife International 2018). Lava gulls are adept scavengers and predators attracted to urban centers by human food and waste (Wiedenfeld and Jiménez-Uzcátegui 2008), but their proximity to these areas makes them especially sensitive to anthropogenic disturbance (Snow BK and Snow DW, 1969; Jiménez-Uzcátegui et al. 2019) and pollutants in the environment (Muñoz-Pérez et al. 2023). Despite the lava gull's small population size and classification as Vulnerable by the IUCN (BirdLife International 2018), research on this species remains limited. No genomes are currently available for this species, with the closest being a scaffold-level assembly of the laughing gull (*Leucophaeus atricilla*, GCA_045784845.1) (Petersen 2024), limiting the ability to conduct high-resolution genetic analyses in a taxon that urgently requires it due to its imperiled conservation status.

We present the first high-quality reference genome for the lava gull, a Galápagos-endemic and the world's rarest gull species. All sequence data were generated in situ using nanopore sequencing on the Galápagos Islands from a blood sample collected from a single female individual. We assess the accuracy and performance of two competing *de novo* assemblers, generate a complete mitogenome, and present an assembly with chromosomal resolution, including both avian sex chromosomes (W, Z). Finally, we compare the quality and completeness of our nanopore genome (i.e. assembly metrics and synteny) to those of other gull (Laridae) genomes produced using different sequencing platforms.

Methods and materials

Sampling

A 0.4 ml blood sample was collected from the brachial vein using a 27-gauge needle and a 1 ml syringe from an adult female lava gull

on July 9, 2024, at the Muelle Municipal (Lat: 0°44'54.83"S, Long: 90°18'46.97"W) of Puerto Ayora, Santa Cruz Island, Galápagos (Fig. 1b). The gull was caught using a baited remote-operated bowmen net and released once the blood sample and morphological measurements were taken. Upon collection, the sample was preserved in a two-part NAP buffer, one part blood sample solution (Camacho-Sanchez et al. 2013) at the site. The sample was stored at -20 °C within 30 min of collection and at -80 °C within 12 h upon return to the Galapagos Science Center (GSC) on San Cristóbal Island in the Galápagos.

Sequencing methods and preparation

All lab work was performed on-site at the Galapagos Science Center, Puerto Baquerizo Moreno, San Cristóbal Island. DNA was extracted from nucleated blood using the Monarch HMW DNA Extraction Kit for Cells and Blood (New England Biolabs, Inc., Ipswich, MA), and its concentration was assessed with Qubit Fluorometric Quantitation (Thermo Fisher Scientific). A library was constructed using the Ultra-Long DNA Sequencing Kit SQK-ULK114 (Oxford Nanopore Technologies) and loaded on a PromethION 2 Solo (P2 Solo) device immediately after library completion. The P2 Solo was connected to a DELL Windows laptop with a 13th Gen Intel(R) Core(TM) i9-13950HX 2.20 GHz, 128 GB RAM, a 64-bit operating system, and x64-based processor for computational support and acquisition of the data onto the laptop drive. Sequencing was conducted using two PromethION Flow Cells, with R10.4.1 chemistry, FLO-PRO114M to generate POD5 format files that capture sequence signal data. Live basecalling requires GPU support and was not activated. Lab work and sequencing were performed following the manufacturer's protocols with minor alterations outlined in the supplement (see Supplement: Sequencing Methods and Preparation). Subsequent analyses were performed at the California Academy of Sciences, San Francisco, CA, USA (CAS).

Basecalling and adapter removal

For basecalling, POD5 files were copied to a desktop computer with GPU support. We then used the super-accurate model (SUP) in Dorado v0.7.4 (github.com/nanoporetech/dorado), the

Table 1. Sequencing results from Flye v2.9.5 and HiFiasm v0.24 contig-level assemblies.

Metric	Flye v2.9.5	HiFiasm v0.24
Number of Scaffolds	1,300	671
Longest Scaffold	141 Mbp	120 Mbp
Scaffold N50/L50	38.4 Mbp/10	41.2 Mbp/9
Number of Contigs	1,363	671
Total Contig Length	1.35 Gbp	1.30 Gbp
Longest Contig	100 Mbp	120 Mbp
Contig N50/L50	37.1 Mbp/12	41.2 Mbp/9
Contig N90/L90	1.34 Mbp/74	1.40 Mbp/73
auN50	52.1 Mbp	50.0 Mbp
BUSCO counts, n = 8,338	C: 8,334, D: 17, F: 3, M: 1	C: 8,316, D: 17, F: 7, M: 15
BUSCO percentages, n = 8,338	C: 99.95%, F: 0.04%, M: 0.01%	C: 99.74%, F: 0.08%, M: 0.18%

The Flye assembly has just one missing BUSCO, while the HiFiasm assembly misses 15.

most accurate of the supplied models, and performed basecalling through minKNOW core v6.0.8 to generate 56 gigabytes of fastq files. Basecalling resulted in gzip fastq files with ~4,000 reads each, which we combined into a single fastq file and eliminated reads less than 250 bp. This fastq file was used from the CAS high-performance computing (HPC) environment for all additional work. Basecalled reads were analyzed with NanoPlot v1.43.0 (De Coster et al. 2018) and a custom script (seq_summary_qscore_lens.sh, see Code Availability) to determine initial sequencing results, including number of reads, read length, read N50, and mean quality. Library adapters were removed using the program Porechop v0.2.4 (Wick et al. 2017) to generate trimmed fastq reads, which were assessed with the same programs.

Genome assembly

To generate initial information, including genome size, coverage, and heterozygosity levels, GeneScopeFK (github.com/thegenemyers/GENESCOPE.FK) was used with kmer size 21. GeneScopeFK is a command-line implementation of GenomeScope2 (Ranallo-Benavidez et al. 2020), with the same results, that creates a kmer histogram and GenomeScope2 plot figures.

Two genome assemblers were used to determine which produced the best *de novo* assembly. The first was generated with Flye v2.9.5 (Kolmogorov et al. 2019) using the `-scaffold` option and trimmed read input. The second was constructed with HiFiasm v0.24 (Cheng et al. 2021, 2025) using trimmed read input and the `-ont` option to enable its ONT read error correction. Assembly quality and completeness were assessed by standard metrics, including N50, L50, and the longest contig using in-house modified script `asmstats.pl` (see Code Availability), `seqtk telo` (github.com/lh3/seqtk), `compleasm` v0.2.6 (Huang and Li, 2023) with `miniprot` (Li, 2023), and BUSCO v5.4.7 (Simão et al. 2015) with `metaEuk` (Levy Karin et al. 2020) using the `aves_odb10` lineage containing 8,338 orthologs. The two assemblies were compared using these results. Though both assemblies were of a high quality, the Flye assembly had marginally better sequencing metrics, including BUSCO scores and longest record length, which led to the selection of the Flye result for contig-level assembly used for follow-on input (Table 1).

To screen for contamination and leftover adapters, the Flye assembly was run through NCBI's Foreign Contamination Screen (github.com/ncbi/fcs) programs FCS-adaptor and FCS-GX against taxonomy ID 328047. Mitochondrial contaminants were identified and removed by blasting against the mitochondrial assembly (see

Mitochondrial Assembly) to eliminate any mitochondrial reads present in the nuclear genome assembly. A minimum 80% read length mapping was selected to keep nuclear-integrated mitochondrial sequences (NUMTs) from being excluded.

Chromosomal-level scaffolding by reference

The *de novo* assembly was scaffolded with references from two species of Laridae, the lesser black-backed gull (*Larus fuscus*, GCA_963932225.1) (Lopez Colom et al. 2024) and the yellow-legged gull (*L. michahellis*, GCA_964199755.1) (Ramos et al. 2025), to create a mapped reference using RagTag scaffold v2.28 (Alonge et al. 2022). These two reference genomes were the highest quality chromosomal-level gulls available on NCBI at the time of assembly. The lesser black-backed gull lacked a W chromosome, so the yellow-legged gull was used to scaffold both the Z and W chromosomes. The purge-dups v1.2.5 (Guan et al. 2020) pipeline was run with the scaffolded assembly, identifying Haplotig, Repeat, and Junk elements in the assembly and removing them to create a new assembly fasta. The purged reference provided contig groupings, orientations, locations, and confidence intervals. Following this, gap filling between scaffolded contigs was attempted with quarTeT GapFiller v1.2.1 (Lin et al. 2023) using the trimmed reads input for the Flye assembly. Records shorter than 5,000 bases were removed and the assembly named `bLeucoFulig_1.0`.

Repeat analysis

Repeat models were identified by RepeatModeler v2.0.1 (Flynn et al. 2020) using RepeatScout (github.com/Dfam-consortium/RepeatScout), RECON (www.repeatmasker.org), LtrHarvest (Ellinghaus et al. 2008), and Ltr_retriever (github.com/oushujun/LTR_retriever) to identify transposable elements, DNA sequences that move locations within a genome, and other repeat motifs. Identified models were combined with curated Aves models from Dfam 3.8 (dfam.org/releases/Dfam_3.8). The combined models and the assembly file were inputs to Repeatmasker v4.1.5 (Smit et al. 2015), which outputs a table categorizing repeats and their lengths. Repeatmasker also outputs a soft-masked repeat version of the assembly used for subsequent analysis. This soft-masked version identifies the repeats with lowercase letters as opposed to a hard-masked version, which replaces repeats with Ns. Identification and masking of repeats prior to conducting gene annotation preclude their inclusion in analyzed sequences.

Genome annotation

The soft-masked genome was structurally annotated using *ab initio* gene modeling program BRAKER3 v3.0.8 (Gabriel et al. 2024) in EP mode since we had no specimen RNA-seq data. This uses cross-species protein homology to determine intron and exon structure and predicts genes with GeneMark-EP+ and AUGUSTUS (Brůna et al. 2021). We used OrthoDB v11 vertebrate proteins for the homology set (Kuznetsov et al. 2023). BRAKER3 was run with `-busco_lineage=aves` added to invoke TSEBRA (Gabielli et al. 2024), which uses compleasm to enhance recovery of BUSCO proteins. This resulted in gff3 annotation, coding sequence DNA, and protein sequence amino acid files. From these BRAKER3 output files, gene models lacking a start and a stop codon were removed, as were gene models nested within another gene model.

Functional annotation began with InterProScan v5.72-103.0 (Jones et al. 2014), identifying protein domains within the amino acid sequences. Then DNA and amino acid sequences were queried in several GenBank databases, using `blastn` v2.15.0 with the `nt` database (built Dec 22, 2023), `blastp` v2.15.0 with the `uniprot_sprot` database (built Dec 31, 2023), and `diamond` `blastp`

v2.1.10 (Buchfink et al. 2021) with TrEMBL and nr databases (both built Dec 20, 2023). Additionally, we searched the eggNOG v5.0.2 database through eggNOG-mapper v2.1.11 (Cantalapiedra et al. 2021). InterProScan protein domains and gene ontology were included in the gff3 annotation and sequences files for each gene model, as was the functional annotation description from the lowest eValue, maximum $1e-10$, for each gene from blast or eggNOG results. BUSCO was used to evaluate the resulting retained gene models with the *aves_odb10* lineage in its protein mode. The gene models were further evaluated with OMArk v2.0.3 (Nevers et al. 2025) and the associated OMA LUCA.h5 database. OMArk employs Hierarchical Orthologous Groups (HOGs) to evaluate proteome completeness.

Genome comparison

The lava gull genome assembly was compared with three other gull species currently available on GenBank: lesser black-backed gull, yellow-legged gull, and laughing gull. The comparison of the completeness of these assemblies was assessed using the date they were published to GenBank, the assembly level, the synteny, sex chromosomes availability, their size, the number of complete and missing BUSCOs, and assembly metrics including: number of contigs and scaffolds, longest contig and scaffolds, contig and scaffold N50 and L50, contig and scaffold N90 and L90, auN50, and synteny visualized with Circos plots generated from two of the genomes' fasta files and BUSCO results.

Mitochondrial assembly

Mitochondrial sequences were pulled from the trimmed reads used with Flye and assembled using HiFiMiTie v0.08 (github.com/calacademy-research/HiFiMiTie), a long-read mitochondrial assembler. Assembly of the mitochondrial genome was performed concurrently with the Flye nuclear assembly to allow for the identification and removal of potential mitochondrial contaminants from the nuclear genome. Briefly, sequencing reads were compared against avian entries in the NCBI mitochondrial database using BLAST. Reads with $\geq 50\%$ sequence coverage were retained for mitochondrial genome assembly, annotation, and subsequent analysis of heteroplasmy and tandem repeat structure within the control region. The Aves taxonomic identification was used to determine a canonical tRNA starting point for the sequence and the mitochondrial genetic code. MITOS2 (Bernt et al. 2013; Donath et al. 2019) was used to create a consensus annotation to confirm the annotation produced in HiFiMiTie. A visual representation of the circular genome was generated in Geneious Prime v2025.1.3 (www.geneious.com) using the mitochondrial fasta file and the annotation gff file.

Results and discussion

Sequencing

Sequencing generated 1.78 million basecalled reads comprising 29.6 Gbp. Reads had a mean length of 16.6 kbp and 17.8 mean quality score. This sequencing output represents 22.5 $\times$ average coverage for the species based on an estimated genome size of 1.31 Gbp (Supplementary Table 1). This 22.5 $\times$ coverage obtained from the lava gull basecalled nanopore sequenced reads is lower than other avian genome assemblies sequenced using differing sequencing platforms, generated coverages of 36.07 $\times$, 37 $\times$, 79 $\times$, and 117 $\times$ coverage (Supplementary Table 2). Though the coverage of our resulting lava gull assembly had two-thirds the coverage of the lowest other assembly, the quality and completeness were comparable to other assemblies sequenced at a higher coverage.

Genome assembly

The Flye assembly from 29.4 Gbp of trimmed reads (Supplementary Table 3) has a total size of 1.35 Gbp, consisting of 1,363 contigs and 1,300 scaffolds. The assembly has a contig N50 of 37.1 Mbp and a contig L50 of 12. The longest contig was 100 Mbp, and the longest scaffold was 141 Mbp (Supplementary Table 4). The HiFiasm assembly of trimmed reads has a total size of 1.30 Gbp, comprising fewer contigs ($n = 671$) (Supplementary Table 5). The Flye assembly was chosen over HiFiasm based on marginally better metrics, including the Flye assembly was missing only 1 BUSCO as opposed to the 15 missing in the HiFiasm assembly, and the Flye generated the longest record length, 141 Mbp, as opposed to 120 Mbp (Table 1). The assembly had a contig N50 of 41.2 Mbp and a contig L50 of 9 (Table 1). The choice of the Flye assembly over HiFiasm for the lava gull contig-level assembly is contrary to the results from the assembly of the Galápagos petrel (*Pterodroma phaeopygia*), which was sequenced using identical laboratory and sequencing protocols. However, with the Galápagos petrel data, HiFiasm performed better or the same in each of the sequencing metrics; in particular, both assemblies had only one missing BUSCO (Sessi et al. in prep). The difference in HiFiasm result between the two taxa may lie in the Galápagos petrel sequencing, at 36 $\times$, having 14 $\times$ increased coverage over the lava gull input of 22 $\times$. This was due to the lava gull single sequencing library, which generated 1.78 million reads, as opposed to the Galápagos petrel sequencing, which used two libraries and generated 4.10 million reads (Supplementary Table 2).

The Flye assembly was scaffolded using both the lesser black-backed gull and the yellow-legged gull assemblies, generating a 1.35 Gbp genome made up of 1,187 scaffolds, with the longest scaffold increasing to 218 Mbp and scaffold N50 increasing from 38.4 Mbp to 84.8 Mbp (Supplementary Table 6). Purge_dups yielded an assembly of 1.31 Gbp with 628 contigs, removing 737 contigs, 532 of which were identified as either Highcov ($n = 41$), Junk ($n = 72$), or Repeat ($n = 419$) (Supplementary Table 7). Gap filling closed 68 gaps, adding 1.07 Mbp bases to our genome assembly. This left 84 remaining gaps, all of which had flanking regions too short to use (Supplementary Table 8). After gap filling, 26 contig records with fewer than 5,000 bases were removed.

Final scaffolded assembly

The final chromosomal-scale scaffolded assembly resulted in a genome size of 1.31 Gbp (Fig. 2a) and was named bLeucoFulg 1.0. It consists of 450 scaffolds with a mean length of 2.92 Mbp. Scaffold N50 length is 85.1 Mbp with L50 count 5. The assembly is also made up of 533 contigs with a mean length of 2.46 Mbp. Contig N50 length is 42.8 Mbp at L50 count 9 (Table 2, column 2 lava gull, Supplementary Table 9). Comparing long contigs and scaffolds to the chromosome-level assembly of the lesser black-backed gull and yellow-legged gull, there were 34 pseudo-chromosomes in the lava gull genome. Of these 34 pseudo-chromosomes, nine have one telomere, including the Z chromosome, which was the fifth-longest chromosome in the genome with a length of 85.1 Mbp. This length is comparable to other avian Z chromosomes (Supplementary Fig. 1). We named autosomal sequences to match the numbers assigned to chromosomal records in the reference scaffolds. These results are similar to other publicly available chromosomal gull assemblies, ranging from 32 to 35 chromosomes. The use of reference-based scaffolding, as used here with the lesser black-backed gull and the yellow-legged gull, is an effective way of producing chromosomal-level scaffolds in birds (Jorquera et al. 2025; Shin et al. 2025) and mammals (Alvarez-Costes et al. 2025), assuming conservation of

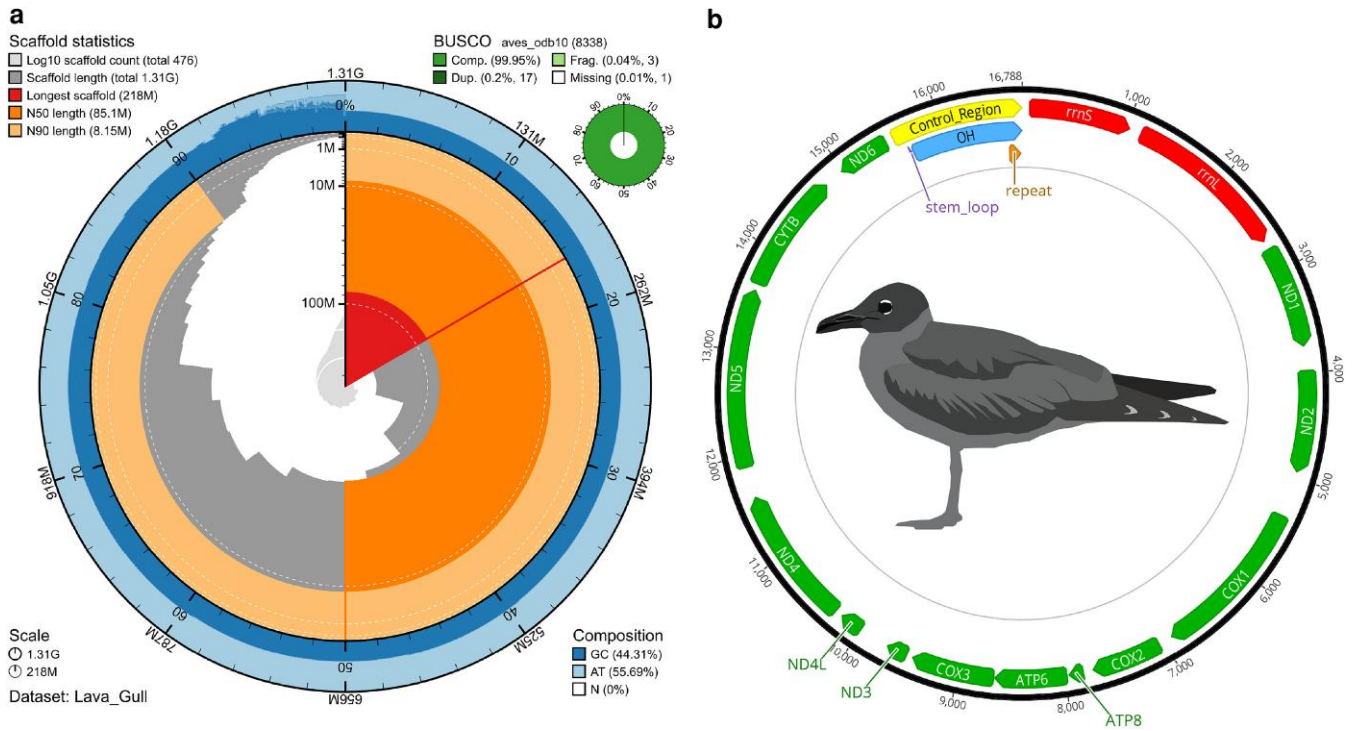


Fig. 2. Snail and mitogenome plots for the Lava Gull. a) BlobToolKit snail plot depicting the assembly metrics for the lava gull genome. The circumference of the circle represents a genome size of 1.31 Gbp. The pale orange indicates scaffold N90 length of 8.15 Mbp, the deeper orange indicates scaffold N50 of 85.1 Mbp, red indicates the longest scaffold of 218 Mbp, and the light blue and dark blue on the perimeter indicate the AT and GC composition, respectively. The top right circle depicts the results from the BUSCO analysis, illustrating that the assembly was highly complete, containing 99.95% complete BUSCOs and only one missing and three fragmented BUSCOs. b) The mitochondrial genome assembly of the lava gull is made up of 16,788 bases and contains 13 protein-coding genes, illustrated in green, and two rRNAs, denoted in red.

Table 2. Genome assembly comparison between the lava gull, laughing gull, lesser black-backed gull, and yellow-legged gull using assembly metrics including: number of contigs and scaffolds, longest contig and scaffold, contig and scaffold N50 and L50, which sex chromosomes are present and their lengths.

Metric	Lava gull (<i>Leucophaeus fuliginosus</i>)	Laughing gull (<i>Leucophaeus atricilla</i>)	Lesser black-backed gull (<i>Larus fuscus</i>)	Yellow-legged gull (<i>Larus michahellis</i>)
Date	This genome	Dec 6, 2024	Feb 7, 2024	Jul 14, 2024
Genome Size	1.31 Gbp	1.18 Gbp	1.32 Gbp	1.41 Gbp
Genome Coverage	22.5x	160x	30x	68x
Number of Scaffolds	450	125,582	1,367	618
Longest Scaffold	218 Mbp	269 kbp	218 Mbp	224 Mbp
Scaffold N50/L50	85.1 Mbp/5	40.4 kbp/10,320	86.5 Mbp/5	87.7 Mbp/5
Scaffold N90/L90	8.15 Mbp/21	4.99 kbp/42,689	3.51 Mbp/25	2.90 Mbp/29
Number of Contigs	533	131,549	1,904	1,118
Longest Contig	141 Mbp	221 kbp	17.2 Mbp	17.7 Mbp
Contig N50/L50	42.8 Mbp/9	37.3 kbp/10,756	3.57 Mbp/106	4.22 Mbp/98
Contig N90/L90	3.25 Mbp/41	4.82 kbp/45,749	421 kbp/436	901 kbp/369
Contig auN50	101 Mbp	39.4 kbp	129 Mbp	99.8 Mbp
BUSCO $n = 8,338$	C:8,334 , D:17, F:3, M:1	C: 6,828, D:7, F:981, M: 529	C: 8,332, D: 7, F:5, M: 1	C: 8,331, D: 22, F:4, M: 3
BUSCO percentage	C: 99.95% , F:0.04%, M: 0.01%	C: 81.89%, F:11.77%, M: 6.34%	C: 99.93%, F:0.06%, M: 0.01%	C: 99.92%, F:0.05%, M: 0.04%
Assembly Level	Chromosome	Scaffold	Chromosome	Chromosome
Sex Chromosomes	Z, W (female)	N/A	Z (male)	Z, W (female)
Sex Chromosome Sizes	85.1 Mbp/25.6Mbp	N/A	214 Mbp/N/A	87.7 Mbp/26.9 Mbp
Sequencing Platform	22.5x ONT Ultra-long	DIP-SEQ	30x PacBio, Arima2 High-C data	68x PacBio, Arima2 High-C data

Bolded values indicate the best value in each metric. BUSCO (Benchmarking Universal Single-Copy Orthologs) provides measures for quantitative assessment of genome assembly: C: complete; D: duplicated; F: fragmented; M: missing.

synteny (Takagi and Sasaki 1974; Waters et al. 2021). From the 8,338 BUSCO genes, BUSCO identified 99.95% complete genes (single-copy: 99.75%, duplicated: 0.20%), 0.04% fragmented, and

0.01% missing (Supplementary Table 10). Of the 17 duplicated BUSCOs, 11 of these duplications occur as a single copy on each of the Z and W chromosomes (i.e. they are gametologs), indicating

there is a high degree of paralogy between the two sex chromosomes and suggesting that they were appropriately assembled and counted (Supplementary Table 11). This is similarly reported in the yellow-legged gull, which has 22 duplicated BUSCOs, 14 of which occur on the Z and W chromosomes. With 99.95% complete BUSCO genes, this assembly contains a high number of conserved genes, indicating a highly complete assembly (Huang and Li, 2023).

Repeat analysis and annotation

RepeatMasker masked 209 Mbp or 15.89% of the 1.31 Gbp genome, comparable to other bird genomes available. Of the 209 Mbp of repeats, long interspersed nuclear elements (LINE) comprised 5.67%, long terminal repeat (LTR) elements 1.86%, and DNA transposons 0.56% (Supplementary Fig. 2, Supplementary Table 12, Supplementary Table 13). Using kmer length 21, GeneScopeFK, a reimplement of GenomeScope2, estimated a genome size of 1.15 Gbp, smaller than the assembled genome size of 1.31 Gbp. GeneScopeFK also identified 83.3 Mbp of repeats and estimated a genome heterozygosity of 0.391% (Supplementary Fig. 3). This underestimates the number of repeats by 125 Mbp, an underreporting commonly observed with this method; however, the estimated non-repeat length of 1.08 Gbp closely aligns with the 1.10 Gbp of non-repeats identified by RepeatMaster, suggesting general concordance between the two approaches for non-repeat content (Supplementary Table 14).

BRAKER3 yielded 24,814 gene models and 26,902 mRNA models. From the initial BRAKER3 models, 21 nested genes and 25 mRNAs that lacked either a start or stop codon were removed to generate 24,768 candidate gene models. Functional annotation and refinement yielded 21,808 gene models and 23,928 mRNAs (Supplementary Table 15). Of these 21,808 gene models, the functional annotation identified 20,027 with names and 16,432 containing one or more protein domains detected by InterProScan (Supplementary Table 16). From these 23,928 mRNA sequences, OMAR identified the input as belonging to Neognathae. Of the 11,061 conserved Hierarchical Orthologous Groups (HOGs) expected for this clade, 10,083 HOGs were found in these mRNAs. The proteome consisted of 21,808 proteins, of which 17,954 were consistent, 2,472 were inconsistent, and 1,382 were unknown (Supplementary Fig. 4, Supplementary Table 17).

Mitochondrial genome

The mitochondrial genome was assembled from 37 adapter-trimmed reads using the HiFiMitie pipeline. The mitogenome contained 22 tRNAs, 2 rRNAs, and 13 protein-coding genes (Fig. 2b), consistent with most vertebrate mitogenomes (Formenti et al. 2021). The assembled circular mitochondrial genome is 16,788 bp, similar in size to other gull species: 16,701 bp in the southern black-backed gull (*Larus dominicanus*) (Slack et al. 2007) and 16,788 bp in the black-tailed gull (*L. crassirostris*) (Kim and Park 2016). The single control region consisted of 1,242 bp and contained a 7 bp consensus repeat motif repeated 16.6 times. Given that the Oxford Nanopore Technologies (ONT) sequencing platform is designed to target long reads during sequencing (store.nanoporetech.com/us/ultra-long-dna-sequencing-kit-v14.html) and with a sequencing mean read length of 16,586 bp in the present study, reads of mitochondrial origin are likely to be under-represented or missed. However, with the 37 mitochondrial reads, we were able to attain an average read length of 10,022 bp, 30x coverage of the 16,788 bp mitochondrial genome, and the read annotations have high contiguity (Supplementary Table 18).

Genome comparison

The species used for the comparison include the laughing gull, which belongs to the same genus (*Leucophaeus*) as the lava gull, as well as the lesser black-backed gull and the yellow-legged gull, both in the genus *Larus*. The lesser black-backed gull and the yellow-legged gull were the most complete gull genomes available on GenBank, and the laughing gull was the only available genome in the same genus as the lava gull at the time of our lava gull assembly. Of the four, the lava gull genome assembly presented here has the fewest number of contigs at 533 and scaffolds at 450, the longest contig at 141 Mbp, the longest contig N50 at 42.8 Mbp, and the smallest contig L50 of 9 (Table 2).

The lava gull assembly also had the most complete BUSCOs at 8,334, and both the lava gull and the yellow-legged gull were missing only one BUSCO. This same BUSCO (28097at8782) was not present in this lava gull assembly, nor the three gull species used in this comparison, as well as two other species, the black-headed gull (*Chroicocephalus ridibundus*, GCA_963924245.1) (Colom et al. 2025) and the European herring gull (*Larus argentatus*, GCA_964417175.1) (www.genomeark.org/vgp-all/Larus_argentatus.html) (Supplementary Table 19). This could indicate that this BUSCO is absent from the lineage; however, this requires further investigation into other Laridae genera. The large percentage of complete BUSCOs (99.95%) indicates that the lava gull assembly is highly complete, surpassing the quality of the other available *Leucophaeus* genome.

Though the initial sequence coverage of the lava gull, at 22.5x, was the lowest compared to other gull species (Table 2), we were able to generate a highly complete and contiguous reference genome, missing just one aves lineage BUSCO. This illustrates the utility and performance of the longer and more accurate nanopore portable sequencing technology for genome assembly, even at a relatively low coverage and sequenced on-site in the Galápagos archipelago.

The contig synteny of the lava gull assembly was compared with the chromosomal-level assemblies for the lesser black-backed gull and the yellow-legged gull, which were both used for the reference-based scaffolding of the lava gull genome. The lava gull assembly had a high degree of synteny in comparison with both reference genomes (Fig. 3). In the synteny assessment with the lesser black-backed gull, several BUSCOs located on the W chromosome of the lava gull were mapped to the Z chromosome of the lesser black-backed gull. This discrepancy resulted from the lack of a W chromosome from the male lesser black-backed gull; therefore, sequences from the lava gull's W chromosome map to the Z chromosome in the absence of a corresponding W scaffold. The Z chromosome in the lesser black-backed gull, at length 214 Mbp and with the absence of BUSCO mapping in about 60% of the scaffold, is likely a misassembly. Also, an in-house script we use to complement seqtk telo, telomere_report.sh (see Code Availability), finds 26 telomere sequences mapping in this BUSCO desert. These ranged from 1,332 bp to 25,308 bp in length, starting at position 88.14 Mbp (Supplementary Table 20).

Conclusions

We present the first highly complete genome of a Galápagos endemic Charadriiformes species, assembled using Oxford Nanopore sequencing performed in the Galápagos Islands, and the first chromosomal-level assembly for the genus *Leucophaeus*. With a haploid genome of 1.31 Gbp, our assembly is similar to

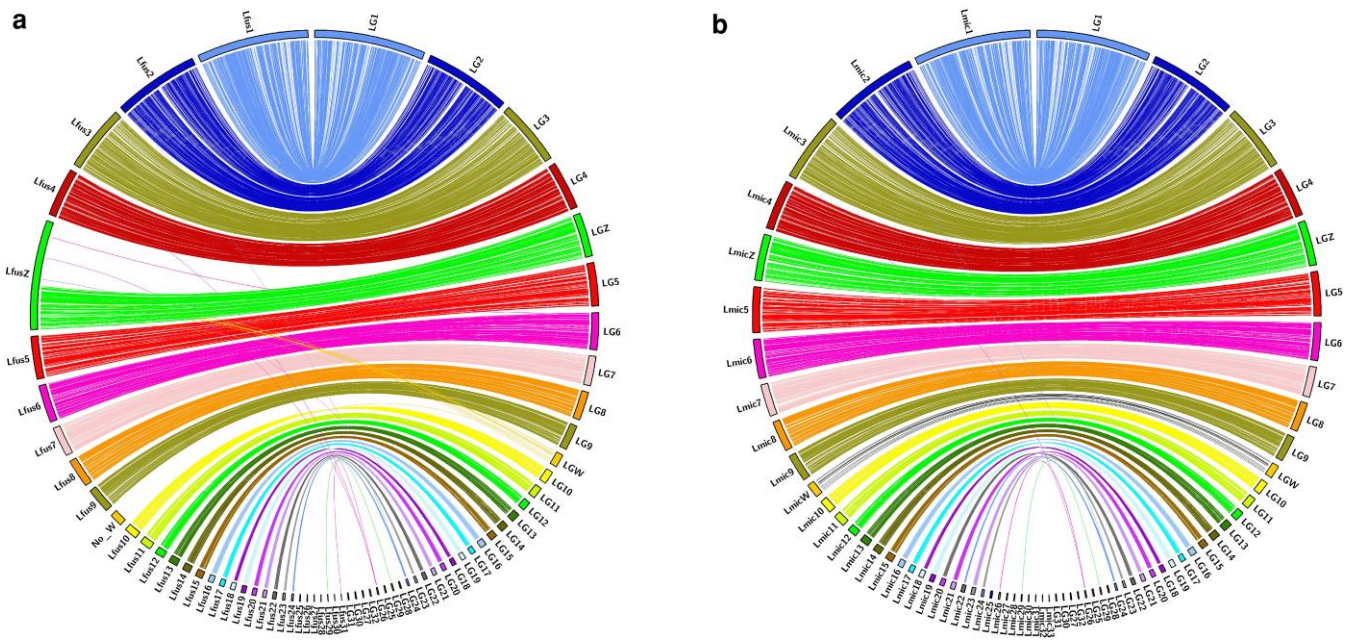


Fig. 3. Synteny between the lava gull genome and the Lesser Black-Backed and Yellow-Legged Gulls. Circos plots illustrating synteny between the lava gull genome and a) the lesser black-backed gull and b) the yellow-legged gull. In each plot, the lava gull assembly is shown on the right and the comparative genome on the left.

other gulls available on GenBank (range 1.2 Gbp to 1.4 Gbp), surpassing all previously published genomes for Laridae based on quality and completeness. Furthermore, our lava gull genome has the highest BUSCO completeness of 99.95% compared to other available Laridae genomes and makes available both sex chromosomes for this genus. This newly assembled genome will be immediately accessible to a wide range of biological studies that inform science-based conservation and management strategies. More broadly, our study demonstrates the feasibility of generating high-quality, chromosomal-level genome assemblies using portable sequencing technologies in remote field settings, opening new possibilities for real-time biodiversity genomics in regions where rapid, *in situ* data generation can have the greatest impact.

Data availability

All raw sequencing data are deposited at NCBI under PRJNA1275045, BioProject SAMN49010278, and BioSample SAMN49010278 : LVGU_060. The genome assembly is deposited under the same BioProject on NCBI and will be made public at the time of publication. <https://dataview.ncbi.nlm.nih.gov/object/PRJNA1275045?reviewer=3bu7kcf60sgf0f7k39oj81j4u1>. GFF files have been submitted to GSA Figshare following the website submission instructions. The publicly accessible datasets used in this study (lesser black-backed gull, yellow-legged gull, laughing gull, European herring gull, black-headed gull, little vermillion flycatcher, dark-eyed junco, and Galápagos petrel) can be found on GenBank and are cited within this paper.

Supplemental material available at [G3](#) online.

Code availability

Custom scripts are in the github.com/jaimechaves76/GalapaGenomes-Lava-Gull-G3 GitHub repository. Scripts and programs starting from the adapter removal of the raw fastq files to the assembly annotation are located there, including

seq_summary_qscore_lens.sh for read statistics, asmstats.pl for assembly statistics, basic_gff_stats.sh, telomere_report.sh, and script dependencies. The bioawk_cas program (<https://github.com/calacademy-research/bioawk.CAS>), an enhanced version of Heng Li's bioawk (github.com/lh3/bioawk), is used by several scripts.

Acknowledgments

We thank the Center for Comparative Genomics at the California Academy of Sciences for their assistance and training, the Galapagos Science Center for allowing us to conduct this work in their facility located on San Cristóbal-Galápagos, including Paúl Yepéz, Cristina Vintimilla, Gabriela Bautista, Jessenia Sotamba, and Sylvia Sotamba, and the Agencia de Regulación y Control de la Bioseguridad y Cuarentena para Galápagos (ABG), in particular, to Alberto Vélez. We would also like to thank the many field assistants who participated in fieldwork, including rangers from the Galápagos National Park, Andrea Loyola, veterinary students from NC-State University, and community members Andrea Villacres and Desire Sobreno. All samples were collected following approved protocols by the Directorate of the Galapagos National Park (permits PC 74-17 and PC-27-19). Fieldwork was funded by the Galápagos Conservation Trust and the University of Sunshine Coast. This work is part of J.A.M.'s research and was supported by the Genentech Foundation Scholars Program FR-015977-P13419 and the National Science Foundation under Grant Number HRD-1826490. All material for the reference genome was accessed under MAE-DNB-CM-2016-0041 and MAATE-DBI-CM-2022-0249 to JAC. This genome is part of the *GalapaGenomes* project funded by NSF (BRC-BIO) to J.A.C. under San Francisco State University IACUC A2021-27.

Funding

This project is based upon work supported by the National Science Foundation under Grant Number 2233210 (BRC-BIO: *GalapaGenomes*) to J.A.C.

Conflicts of interest

All none declared.

Author contribution

J.A.C. and J.P.D. designed the project; J.A.C. acquired funding for the project; A.L. provided critical training and guidance in the lab; J.A.M. and G.P. conducted the lab work and produced all the genetic data; J.B.H. and J.M. performed the bioinformatic analysis; J.A.M. wrote the manuscript with input from J.B.H., J.P.D., J.A.C., and V.F.; A.S. acquired funding for fieldwork and the biological sample. All authors checked and approved the final version of the manuscript.

Note added in proof

<https://doi.org/10.1093/g3journal/jkag004> in this issue for a related work.

Literature cited

- Alonge M et al. 2022. Automated assembly scaffolding using RagTag elevates a new tomato system for high-throughput genome editing. *Genome Biol.* 23:258. <https://doi.org/10.1186/s13059-022-02823-7>.
- Alvarez-Costes S et al. 2025. Leveraging synteny to generate reference genomes for conservation: assembling the genomes of Hector's and Maui dolphins. *Mol Ecol Resour.* 25:e14109. <https://doi.org/10.1111/1755-0998.14109>.
- Benitez-Capistros F, Hugé J, Koedam N. 2014. Environmental impacts on the Galapagos Islands: identification of interactions, perceptions and steps ahead. *Ecol Indic.* 38:113–123. <https://doi.org/10.1016/j.ecolind.2013.10.019>.
- Bernt M et al. 2013. MITOS: improved de novo metazoan mitochondrial genome annotation. *Mol Phylogenet Evol.* 69:313–319. <https://doi.org/10.1016/j.ympev.2012.08.023>.
- BirdLife International. 2018. *Larus fuliginosus*. The IUCN Red List of Threatened Species 2018: e.T22694452A132552501. <https://doi.org/10.2305/IUCN.UK.2018-2.RLTS.T22694452A132552501.en>. Accessed on 30 June 2025.
- Brandies P, Peel E, Hogg CJ, Belov K. 2019. The value of reference genomes in the conservation of threatened species. *Genes (Basel).* 10: 1–17. <https://doi.org/10.3390/genes10110846>.
- Brito PH, Edwards SV. 2009. Multilocus phylogeography and phylogenetics using sequence-based markers. *Genetica.* 135:439–455. <https://doi.org/10.1007/s10709-008-9293-3>.
- Brůna T, Hoff KJ, Lomsadze A, Stanke M, Borodovsky M. 2021. BRAKER2: automatic eukaryotic genome annotation with GeneMark-EP+ and AUGUSTUS supported by a protein database. *NAR Genom Bioinform.* 3:lqaa108. <https://doi.org/10.1093/nargab/lqaa108>.
- Buchfink B, 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-Sanchez M, Burraco P, Gomez-Mestre I, Leonard JA. 2013. Preservation of RNA and DNA from mammal samples under field conditions. *Mol Ecol Resour.* 13:663–673. <https://doi.org/10.1111/1755-0998.12108>.
- Cantalapiedra CP, 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>.
- Carrion V, Donlan CJ, Campbell KJ, Lavoie C, Cruz F. 2011. Archipelago-wide island restoration in the Galápagos Islands: reducing costs of invasive mammal eradication programs and re-invasion risk. *PLoS One.* 6:e18835. <https://doi.org/10.1371/journal.pone.0018835>.
- Charles Darwin Foundation. Galapagos Species Database, dataZone. Charles Darwin Foundation. 2025. <https://datazone.darwininfoundation.org/en/checklist/>. Date accessed 25 April 2025.
- Chaves JA. 2018. Genetic consequences of invasive species in the Galapagos Islands. In: de Lourdes Torres M, Mena CF, editors. *Understanding invasive species in the Galapagos Islands: from the molecular to the landscape*. Springer International Publishing. p. 19–32.
- Cheng H et al. (2025). Efficient near telomere-to-telomere assembly of Nanopore Simplex reads [preprint]. *bioRxiv* 2025.04.14.648685.
- Cheng H, Concepcion GT, Feng X, Zhang H, Li H. 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>.
- Colom RL et al. 2025. The genome sequence of the black-headed gull, *Chroicocephalus ridibundus* (Linnaeus, 1766). *Wellcome Open Res.* 9:399. <https://doi.org/10.12688/wellcomeopenres.22741.2>.
- Cruz JB, Cruz F. 1987. Conservation of the dark-rumped petrel *Pterodroma phaeopygia* in the Galapagos Islands, Ecuador. *Biol Conserv.* 42:303–311. [https://doi.org/10.1016/0006-3207\(87\)90074-7](https://doi.org/10.1016/0006-3207(87)90074-7).
- De Coster W, D'hert S, Schultz DT, Cruts M, Van Broeckhoven C. 2018. NanoPack: visualizing and processing long-read sequencing data. *Bioinformatics.* 34:2666–2669. <https://doi.org/10.1093/bioinformatics/bty149>.
- Díez-del-Molino D, Sánchez-Barreiro F, Barnes I, Gilbert MTP, Dalén L. 2018. Quantifying temporal genomic erosion in endangered species. *Trends Ecol Evol.* 33:176–185. <https://doi.org/10.1016/j.tree.2017.12.002>.
- Donath A et al. 2019. Improved annotation of protein-coding genes boundaries in metazoan mitochondrial genomes. *Nucleic Acids Res.* 47:10543–10552. <https://doi.org/10.1093/nar/gkz833>.
- Ellinghaus D, Kurtz S, Willhoeft U. 2008. LTRharvest, an efficient and flexible software for de novo detection of LTR retrotransposons. *BMC Bioinform.* 9:1–14. <https://doi.org/10.1186/1471-2105-9-18>.
- Epler B. 2007. Tourism, the economy, population growth, and conservation in Galapagos. Charles Darwin Foundation. 55:103–132.
- Flynn JM et al. 2020. RepeatModeler2 for automated genomic discovery of transposable element families. *Proc Natl Acad Sci.* 117: 9451–9457. <https://doi.org/10.1073/pnas.1921046117>.
- Formenti G et al. 2021. Complete vertebrate mitogenomes reveal widespread repeats and gene duplications. *Genome Biol.* 22:120. <https://doi.org/10.1186/s13059-021-02336-9>.
- Fuentes-Pardo AP, Ruzzante DE. 2017. Whole-genome sequencing approaches for conservation biology: advantages, limitations and practical recommendations. *Mol Ecol.* 26:5369–5406. <https://doi.org/10.1111/mec.14264>.
- 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>.
- Gabrielli M et al. 2024. Demographic responses of oceanic island birds to local and regional ecological disruptions revealed by whole-genome sequencing. *Mol Ecol.* 33:e17243. <https://doi.org/10.1111/mec.17243>.
- Genome 10K Community of Scientists. 2009. Genome 10K: a proposal to obtain whole-genome sequence for 10 000 vertebrate species. *J Hered.* 100:659–674. <https://doi.org/10.1093/jhered/esp086>.

- Gould J. 1841. The zoology of the voyage of the H.M.S. Beagle, Part III: birds (Edited and superintended by Charles Darwin). Smith, Elder and Co.
- Grant KT, Estes OH, Estes GB. 2014. Observations on the breeding and distribution of Lava Gull. *Cotinga*. 37:22–35.
- 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>.
- Huang N, Li H. 2023. Compleasm: a faster and more accurate reimplement of BUSCO. *Bioinformatics*. 39:btad595. <https://doi.org/10.1093/bioinformatics/btad595>.
- Jiménez-Uzcátegui G et al. 2019. Threats and vision for the conservation of Galápagos birds. *Open Ornithol J*. 12:1–15. <https://doi.org/10.2174/1874453201912010001>.
- Jones P et al. 2014. InterProScan 5: genome-scale protein function classification. *Bioinformatics*. 30:1236–1240. <https://doi.org/10.1093/bioinformatics/btu031>.
- Jorquera, J et al. (2025). Genomic introgression and adaptation of southern seabird species facilitate recent polar colonization. *Mol Biol Evol.*, 42, msaf053.
- Kier G et al. 2009. A global assessment of endemism and species richness across island and mainland regions. *Proc Natl Acad Sci*. 106: 9322–9327. <https://doi.org/10.1073/pnas.0810306106>.
- Kim JY, Park YC. 2016. The complete mitogenome of the Black-tailed gull *Larus crassirostris* (Charadriiformes: Laridae). *Mitochondrial DNA Part A*. 27:1885–1886. <https://doi.org/10.3109/19401736.2014.971271>.
- Kolmogorov M, Yuan J, Lin Y, Pevzner PA. 2019. Assembly of long, error-prone reads using repeat graphs. *Nat Biotechnol*. 37: 540–546. <https://doi.org/10.1038/s41587-019-0072-8>.
- Kuznetsov D et al. 2023. OrthoDB v11: annotation of orthologs in the widest sampling of organismal diversity. *Nucleic Acids Res*. 51: D445–D451. <https://doi.org/10.1093/nar/gkac998>.
- Lack D. 1947. Darwin's finches. Cambridge Univ.
- Levy Karin E, Mirdita M, Söding J. 2020. MetaEuk—sensitive, high-throughput gene discovery, and annotation for large-scale eukaryotic metagenomics. *Microbiome*. 8:1–15. <https://doi.org/10.1186/s40168-020-00808-x>.
- Li H. 2023. Protein-to-genome alignment with miniprot. *Bioinformatics*. 39:btad014. <https://doi.org/10.1093/bioinformatics/btad014>.
- Lin Y et al. 2023. Quartet: a telomere-to-telomere toolkit for gap-free genome assembly and centromeric repeat identification. *Hortic Res*. 10:uhad127. <https://doi.org/10.1093/hr/uhad127>.
- Lopez Colom R et al. 2024. The genome sequence of the lesser black-backed gull, *Larus fuscus* Linnaeus, 1758. *Wellcome Open Res*. 9: 444. <https://doi.org/10.12688/wellcomeopenres.22766.2>.
- Muñoz-Pérez JP et al. 2023. Galápagos and the plastic problem. *Front Sustain*. 4:1091516. <https://doi.org/10.3389/frsus.2023.1091516>.
- Nevers Y et al. 2025. Quality assessment of gene repertoire annotations with OMArk. *Nat Biotechnol*. 43:124–133. <https://doi.org/10.1038/s41587-024-02147-w>.
- Paez S et al. 2022. Reference genomes for conservation. *Science*. 377: 364–366. <https://doi.org/10.1126/science.abm8127>.
- Petersen B. 2024. *Leucophaeus atricilla* isolate B10K-D-NMHD-67, whole genome shotgun sequencing project. Nucleotide-NCBI [cited 2025 June 13]. Available from <https://www.ncbi.nlm.nih.gov/nuccore/JBJGIG000000000.1>.
- Pozo G et al. 2024. First whole-genome sequence and assembly of the Ecuadorian brown-headed spider monkey (*Ateles fusciceps fusciceps*), a critically endangered species, using Oxford Nanopore Technologies. G3. 14:jkae014. <https://doi.org/10.1093/g3journal/jkae014>.
- Ramos E et al. 2025. The genome sequence of the Yellow-legged Gull, *Larus michahellis* Naumann, 1840. *Wellcome Open Res*. 10:129. <https://doi.org/10.12688/wellcomeopenres.23849.1>.
- Ranallo-Benavidez TR, Jaron KS, Schatz MC. 2020. GenomeScope 2.0 and Smudgeplot for reference-free profiling of polyploid genomes. *Nat Commun*. 11:1432. <https://doi.org/10.1038/s41467-020-14998-3>.
- Romanov MN et al. 2009. The value of avian genomics to the conservation of wildlife. *BMC Genomics*. 10:S10. <https://doi.org/10.1186/1471-2164-10-S2-S10>.
- Ryder OA. 2005. Conservation genomics: applying whole genome studies to species conservation efforts. *Cytogenet Genome Res*. 108:6–15. <https://doi.org/10.1159/000080796>.
- Shin HD et al. 2025. Chromosome-level genome assembly of Korean long-tailed chicken and pangenome of 40 *Gallus gallus* assemblies. *Sci Data*. 12:51. <https://doi.org/10.1038/s41597-024-04287-9>.
- 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. <https://doi.org/10.1093/bioinformatics/btv351>.
- Slack KE, Delsuc F, Mclenachan PA, Arnason U, Penny D. 2007. Resolving the root of the avian mitogenomic tree by breaking up long branches. *Mol Phylogenet Evol*. 42:1–13. <https://doi.org/10.1016/j.ympev.2006.06.002>.
- Smit AFA, Hubley R, Green P. 2013–2015. RepeatMasker Open-4.0. <https://www.repeatmasker.org/faq.html#faq3>.
- Snow D, Grant P. 1988. Ecology and evolution of Darwin's finches. *J Anim Ecol*. 57:320. <https://doi.org/10.2307/4785>.
- Snow BK, Snow DW. 1969. Observations on the Lava Gull *Larus fuliginosus*. *Ibis (Lond 1859)*. 111:30–35. <https://doi.org/10.1111/j.1474-919X.1969.tb01602.x>.
- Steadman DW. 1986. Holocene vertebrate fossils from Isla Floreana, Galapagos. *Smithson Contrib Zool*. 1–104.
- Takagi N, Sasaki M. 1974. A phylogenetic study of bird karyotypes. *Chromosoma*. 46:91–120. <https://doi.org/10.1007/BF00332341>.
- Van Der Valk T, Díez-del-Molino D, Marques-Bonet T, Guschanski K, Dalén L. 2019. Historical genomes reveal the genomic consequences of recent population decline in eastern gorillas. *Curr Biol*. 29:165–170.e6. <https://doi.org/10.1016/j.cub.2018.11.055>.
- Vitousek PM, Mooney HA, Lubchenco J, Melillo JM. 1997. Human domination of Earth's ecosystems. *Science*. 277:494–499. <https://doi.org/10.1126/science.277.5325.494>.
- Vlček J et al. 2025. Genomes of Galápagos mockingbirds reveal the impact of island size and past demography on inbreeding and genetic load in contemporary populations. *Mol Ecol*. 34:e17665. <https://doi.org/10.1111/mec.17665>.
- Walsh SJ, Mena CF. 2016. Interactions of social, terrestrial, and marine sub-systems in the Galapagos Islands, Ecuador. *Proc Natl Acad Sci*. 113:14536–14543. <https://doi.org/10.1073/pnas.1604990113>.
- Waters PD et al. 2021. Microchromosomes are building blocks of bird, reptile, and mammal chromosomes. *Proc Natl Acad Sci*. 118: e2112494118. <https://doi.org/10.1073/pnas.2112494118>.
- Watsa M, Erkenswick GA, Pomerantz A, Prost S. 2020. Portable sequencing as a teaching tool in conservation and biodiversity research. *PLoS Biol*. 18:e3000667. <https://doi.org/10.1371/journal.pbio.3000667>.
- Wick RR, Judd LM, Gorrie CL, Holt KE. 2017. Completing bacterial genome assemblies with multiplex MinION sequencing. *Microb Genom*. 3:e000132. <https://doi.org/10.1099/mgen.0.000132>.
- Wiedenfeld DA, Jiménez-Uzcátegui G. 2008. Critical problems for bird conservation in the Galápagos Islands. *Cotinga*. 29:22–27.