Check for updates

DATA NOTE

REVISED ERGA-BGE reference genome of the Azores Bullfinch - *Pyrrhula murina* Godman, 1866: an IUCN Vulnerable Species endemic to a single island in the Azores Archipelago (Portugal)

[version 3; peer review: 2 approved, 1 approved with reservations]

Ricardo Jorge Lopes[1,2], Astrid Böhne[3], Thomas Marcussen[4,5], Rebekah A. Oomen[4-8], Torsten Hugo Struck[4], Laura Aguilera[9,10], Marta Gut[9,10], Francisco Câmara Ferreira[9,10], Fernando Cruz[9,10], Jèssica Gómez-Garrido[9,10], Tyler S. Alioto[9,10], Rita Monteiro[3]

[1]CE3C, Center for Ecology, Evolution and Environmental Change & CHANGE, Departamento de Biologia Animal, Faculdade de Ciências, Universidade de Lisboa, Lisboa, Portugal
[2]MUHNAC, Museu Nacional de História Natural e da Ciência, Universidade de Lisboa, Rua da Escola Politécnica, Lisboa, Portugal
[3]Leibniz Institute for the Analysis of Biodiversity Change, Museum Koenig Bonn, Bonn, 53113, Germany
[4]Natural History Museum, University of Oslo, Oslo, Norway
[5]Centre for Ecological & Evolutionary Synthesis, University of Oslo, Oslo, Norway
[6]Department of Biological Sciences, University of New Brunswick Saint John, Saint John, Canada
[7]Tjärnö Marine Laboratory, University of Gothenburg, Gothenburg, Sweden
[8]Centre for Coastal Research, University of Agder, Kristiansand, Norway
[9]Centro Nacional de Análisis Genómico (CNAG), Barcelona, Spain
[10]Universitat de Barcelona (UB), Barcelona, Spain

v3 **First published:** 01 Aug 2025, **5**:210
https://doi.org/10.12688/openreseurope.20666.1
Second version: 11 Feb 2026, **5**:210
https://doi.org/10.12688/openreseurope.20666.2
Latest published: 11 Mar 2026, **5**:210
https://doi.org/10.12688/openreseurope.20666.3

Abstract

The reference genome of Pyrrhula murina's reference genome will substantially enhance the current monitoring of genetic diversity and population viability and will allow us to understand the effective population size trends throughout time and recent bottlenecks and population expansions. A total of 42 contiguous chromosomal pseudomolecules were assembled from the genome sequence, using ONT, HI-C and Illumina data. This chromosome-level assembly encompasses 1.2 Gb, composed of 65 contigs and 60 scaffolds, with contig and scaffold N50 values of 65 Mb and 76 Mb, respectively.

Open Peer Review

Approval Status ✓ ✓ ?

	1	2	3
version 3 (revision) 11 Mar 2026	✓ view	✓ view	
version 2 (revision) 11 Feb 2026	? view	✓ view	? view
version 1 01 Aug 2025	? view	? view	

1. **Luke W Silver**, University of Sydney,

Keywords
Pyrrhula murina, Fringillidae family, Azores Bullfinch, genome assembly, European Reference Genome Atlas, Biodiversity Genomics Europe, Earth Biogenome Project

This article is included in the Horizon Europe gateway.

This article is included in the Genome Reports from the Biodiversity Genomics Europe Project collection.

Sydney, Australia

2. **Annabel Whibley**, University of Auckland, Auckland, New Zealand
Bragato Research Institute, Lincoln, New Zealand

3. **Madeline Chase**, Swiss Ornithological Institute, Sempach, Switzerland

Any reports and responses or comments on the article can be found at the end of the article.

Corresponding author: Rita Monteiro (R.Monteiro@leibniz-lib.de)

Author roles: Lopes RJ: Investigation, Resources, Writing – Original Draft Preparation, Writing – Review & Editing; **Böhne A**: Funding Acquisition, Methodology, Supervision, Writing – Review & Editing; **Marcussen T**: Methodology, Supervision, Validation, Writing – Review & Editing; **Oomen RA**: Funding Acquisition, Methodology, Supervision, Writing – Review & Editing; **Struck TH**: Funding Acquisition, Methodology, Supervision, Writing – Review & Editing; **Aguilera L**: Investigation, Writing – Original Draft Preparation, Writing – Review & Editing; **Gut M**: Investigation, Writing – Review & Editing; **Câmara Ferreira F**: Data Curation, Formal Analysis, Writing – Review & Editing; **Cruz F**: Data Curation, Formal Analysis, Writing – Review & Editing; **Gómez-Garrido J**: Data Curation, Formal Analysis, Writing – Review & Editing; **Alioto TS**: Data Curation, Formal Analysis, Supervision, Writing – Review & Editing; **Monteiro R**: Methodology, Supervision, Validation, Visualization, Writing – Original Draft Preparation, Writing – Review & Editing

Competing interests: No competing interests were disclosed.

Grant information: Biodiversity Genomics Europe (Grant no.101059492) is funded by Horizon Europe under the Biodiversity, Circular Economy and Environment call (REA.B.3); co-funded by the Swiss State Secretariat for Education, Research and Innovation (SERI) under contract numbers 22.00173 and 24.00054; and by the UK Research and Innovation (UKRI) under the Department for Business, Energy and Industrial Strategy's Horizon Europe Guarantee Scheme. RJL was funded by the grant UID/00329/2025 awarded to Centro de Ecologia, Evolução e Alterações Ambientais (CE3C).
The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Copyright: © 2026 Lopes RJ *et al*. This is an open access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

How to cite this article: Lopes RJ, Böhne A, Marcussen T *et al.* **ERGA-BGE reference genome of the Azores Bullfinch - *Pyrrhula murina* Godman, 1866: an IUCN Vulnerable Species endemic to a single island in the Azores Archipelago (Portugal) [version 3; peer review: 2 approved, 1 approved with reservations]** Open Research Europe 2026, **5**:210
https://doi.org/10.12688/openreseurope.20666.3

First published: 01 Aug 2025, **5**:210 https://doi.org/10.12688/openreseurope.20666.1

REVISED **Amendments from Version 2**

In the current version, we have addressed the reviewer's comments, adding some additional details to the Library preparation and sequencing section.

Any further responses from the reviewers can be found at the end of the article

Introduction

The Azores Bullfinch, *Pyrrhula murina* Godman, 1866, is restricted to a small area that includes native laurel forest, in the east of the largest island (São Miguel) of the Azores archipelago (Figure 1). It is thought to have diverged from its nearest sister species more than 1 MYA (Töpfer *et al.*, 2011).

By the late 19th/early 20th century, the Azores Bullfinch was already very rare and restricted to the higher mountain valleys in the east of São Miguel island (Aubrecht, 2000). It was rediscovered only in the last century, and numbers have increased as a result of conservation efforts to protect this species and its main habitat. Its current population size is estimated at around 1000 individuals (Ceia *et al.*, 2011; Costa *et al.*, 2023; Gil *et al.*, 2016), and its conservation status has been down-listed from "Critically Endangered" to "Vulnerable" on the IUCN Red List, as the population size is considered to be stable (BirdLife International, 2021).

It is one of the few endemic bird species that is strongly connected to the last remnants of the Laurel Forest in the Azores, providing key ecological services for the sustainability of these habitats.

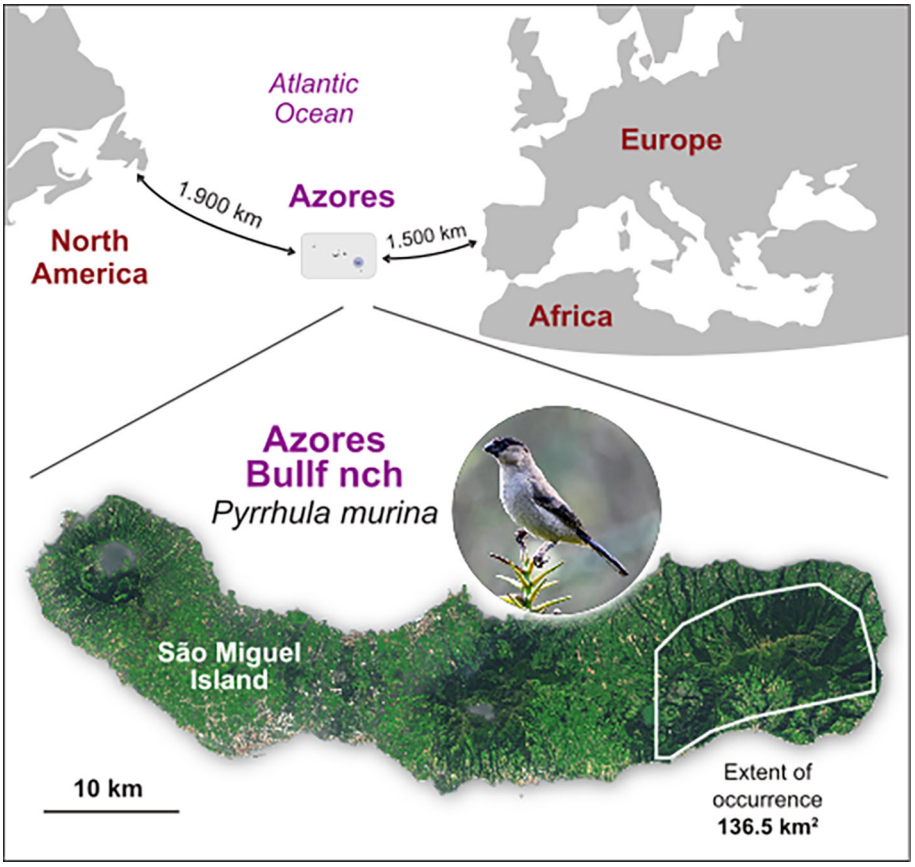


Figure 1. Spatial location of the Azores archipelago and the distribution range of the Azores Bullfinch (*Pyrrhula murina*) in the east part of São Miguel Island. Also shown the latest estimates of the extent of occurrence, from Costa et al. (2023).

A high-quality reference genome will allow us to enhance our knowledge on the long-term viability of this small population, informing us about the impact of the multiple bottlenecks on its genetic diversity and also on the current and future evolution of this population. Of particular importance for the viability of the species, is the fact that *Pyrrhula murina*, along with his sister species, the Eurasian Bullfinch, *Pyrrhula pyrrhula*, have relatively small and possibly neotenous sperm, an ancestral trait that evolved before the two taxa diverged (Lifjeld *et al.*, 2013).

The generation of this reference resource was coordinated by the European Reference Genome Atlas (ERGA) initiative's Biodiversity Genomics Europe (BGE) project, supporting ERGA's aims of promoting transnational cooperation to promote advances in the application of genomics technologies to protect and restore biodiversity (Mazzoni *et al.*, 2023).

Materials & methods

ERGA's sequencing strategy includes Oxford Nanopore Technology (ONT) and/or Pacific Biosciences (PacBio) for long-read sequencing, along with Hi-C sequencing for chromosomal architecture, Illumina Paired-End (PE) for k-mer profiling and polishing (i.e. recommended for ONT-only assemblies), to facilitate genome assembly, as well as Illumina RNA sequencing for transcriptomic profiling and annotation.

Sample and sampling information

On November 27, 2023, a female adult (determined based on genetic sexing) of *Pyrrhula murina* was sampled and identified by Ricardo Jorge Lopes. The specimen was caught using mist-nets in a woodland area in Salto do Cavalo, São Miguel island, in the Azores archipelago, Portugal.

This species is unmistakable, being the sole representative of this Genus in the Azores Archipelago.

Sampling was performed under permit 119/2023/DRAAC, from the Regional Directorate for the Environment and Climate Action, of the Azores Government, and under the Internationally Recognized Compliance certificate CCIR-RAA/2023/65. The specimen's blood samples were snap-frozen immediately after harvesting and stored in liquid nitrogen until DNA extraction.

Vouchering information

An electronic voucher image of the sequenced individual is available from ERGA's EBI BioImageArchive dataset www.ebi.ac.uk/biostudies/bioimages/studies/S-BIAD1012?query=ERGA under accession ID SAMEA115966740.

A frozen reference blood sample from the sequenced individual was deposited at MUHNAC (National Museum of Natural History and Science of the University of Lisbon), under the voucher ID C67972_Blood_05.

Genetic information

The estimated genome size, based on ancestral taxa, is 1.33 Gb, while the estimation based on reads kmer profiling using GenomeScope2 (kmer length = 21) is 1.08 Gb. This is a diploid genome with a haploid number of 39 chromosomes (2 n = 78), including Z and W sex chromosomes in females, based on ancestral data (no direct estimates are available. All information for this species was retrieved from Genomes on a Tree (Challis *et al.*, 2023).

DNA/RNA processing

DNA was extracted from blood using the Blood & Cell Culture DNA Midi Kit (Qiagen) following the manufacturer's instructions. DNA quantification was performed using a Qubit dsDNA BR Assay Kit (Thermo Fisher Scientific), and DNA integrity was assessed using a Genomic DNA 165 Kb Kit (Agilent) on the Femto Pulse system (Agilent). The DNA was stored at +4 °C until sequenced.

RNA was extracted from blood using an RNeasy Mini Kit (Qiagen) according to the manufacturer's instructions. RNA quantification was performed using the Qubit RNA BR kit, and RNA integrity was assessed using a Bioanalyzer 2100 system (Agilent) RNA 6000 Nano Kit (Agilent). The RNA was stored at -80 °C until sequenced.

Library preparation and sequencing

For long-read whole genome sequencing, a library was prepared using the SQK-LSK114 Kit (Oxford Nanopore Technologies, ONT), which was then sequenced using a R10.4.1 flow cell on a PromethION 24 A Series instrument (ONT). Basecalling was performed using Dorado v7.4.14 (Oxford nanopore Technologies) with the super accurate (SUP) basecalling model v4.3.0 at 400 bps, integrated within MinKNOW v24.06.15. A short-read whole-genome sequencing library was prepared using the KAPA Hyper Prep Kit (Roche).

A Hi-C library was prepared from blood using the Dovetail Omni-C Kit (Cantata Bio), followed by the KAPA Hyper Prep Kit for Illumina sequencing (Roche).

The RNA library was prepared using the KAPA mRNA Hyper prep kit (Roche). All short-read libraries were sequenced on a NovaSeq 6000 instrument (2x150 bp, Illumina).

In total, 100x Oxford Nanopore, 99x Illumina WGS shotgun, and 79x Hi-C data were sequenced to generate the assembly. In addition, 91 million reads of Illumina RNA-seq were sequenced to facilitate future annotation efforts.

Genome assembly methods

The genome was assembled using the CNAG CLAWS pipeline v2.3.0 (Gomez-Garrido, 2024). Briefly, ONT reads were preprocessed for quality and length using Filtlong v0.2.1 (--min_length 1000 --min_mean_q 97 -t 80000000000) , and initial contigs were directly assembled from the filtered ONT data using Hifiasm v0.24.0 with the –ont option and Hi-C phasing, followed by scaffolding with YaHS v1.2a using the Hi-C data. Assembled scaffolds were curated via manual inspection using Pretext v0.2.5 with the Rapid Curation Toolkit (https://gitlab.com/wtsi-grit/rapid-curation) to remove any false joins and incorporate any sequences not automatically scaffolded into their respective locations in the chromosomal pseudomolecules (or super-scaffolds). The blobtoolkit nextflow pipeline v0.6.0 (https://pipelines.tol.sanger.ac.uk/blobtoolkit/0.6.0/usage) confirmed the absence of contaminants. The sex chromosomes were identified

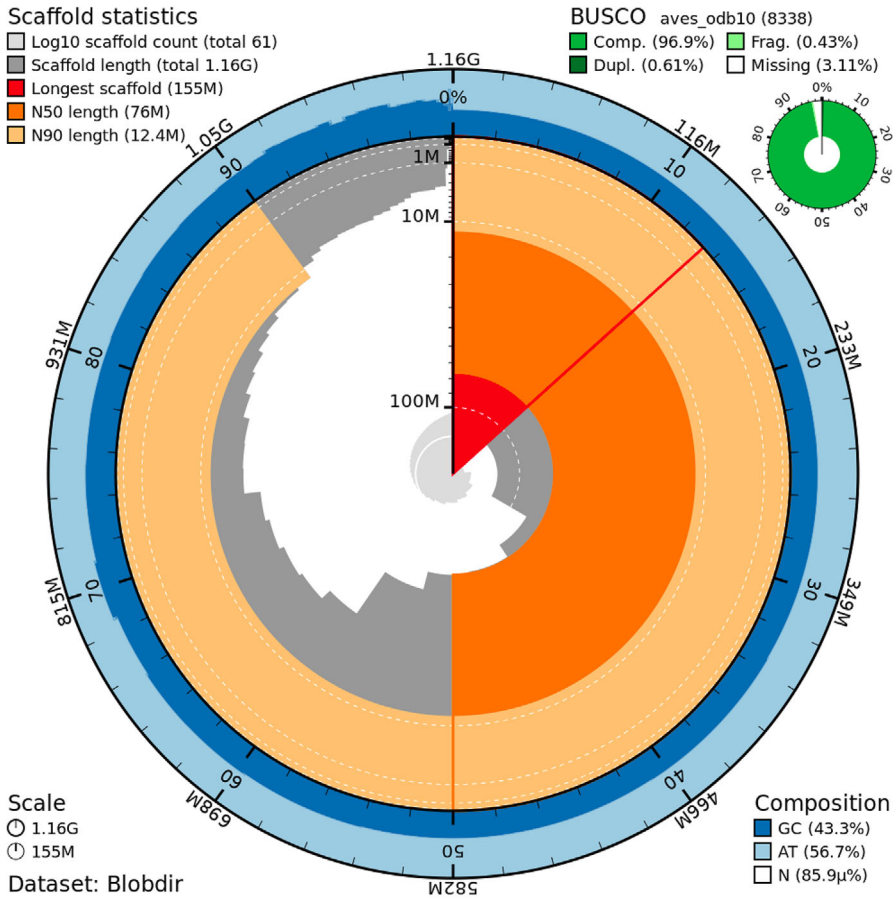


Figure 2. Snail plot summary of assembly statistics. The main plot is divided into 1,000 size-ordered bins around the circumference, with each bin representing 0.1% of the 1,163,884,374bp assembly, including the mitochondrial genome. The distribution of sequence lengths is shown in dark grey, with the plot radius scaled to the longest sequence present in the assembly (155 Mb bp, shown in red). Orange and pale-orange arcs show the scaffold N50 and N90 sequence lengths (76,036,967 and 12,418,366 bp), respectively. The pale grey spiral shows the cumulative sequence count on a log-scale, with white scale lines showing successive orders of magnitude. The blue and pale-blue area around the outside of the plot shows the distribution of GC, AT, and N percentages in the same bins as the inner plot. A summary of complete, fragmented, duplicated, and missing BUSCO genes found in the assembled genome from the avian database (odb10) is shown on the top right.

by sequencing coverage. Assignment of Z and W chromosomes was according to conservation of length: bPyrMur chrZ (OZ238997.1: 86793913 bp) vs. bGalGal chrZ (CM028522.1: 86044486 bp). Whole genome alignment against GRCg7b (GCA_016699485.1) confirmed these assignments. Finally, the mitochondrial genome (OZ239036.1) was assembled as a single circular contig of 16,836 bp using the FOAM pipeline v0.5 (https://github.com/cnag-aat/FOAM) and included in the released assembly (GCA_965183895.1). K-mer counts for estimating genome size, QV and k-mer completeness were performed on the combined set of Illumina WGS and filtered ONT reads. Summary analysis of the released assembly was performed using the ERGA-BGE Genome Report ASM Galaxy workflow (De Panis, 2024).

Results

Genome assembly

The genome assembly has a total length of 1,163,884,374 bp in 61 scaffolds, including 40 autosomes, the Z and W chromosomes, and the mitogenome (Figure 2 and Figure 3), with a GC content of 43.29%. It features a contig N50 of 64,819,875 bp (L50=6) and a scaffold N50 of 76,036,967 bp (L50=6). There are 5 gaps, totaling 1,000 kb in cumulative size. The ten largest autosomes as well as the Z and W scaffolds are longer than 28 Mb and the rest (30 autosomes) are shorter than 23 Mb, the shortest being about 3.5 Mb. The single-copy gene content analysis using the aves_odb10 database with BUSCO v5.5.0 (Manni *et al.*, 2021) resulted in 96.9% completeness (96.3% single and 0.6% duplicated). 98.6% of reads k-mers were present in the assembly, and the assembly has a base accuracy Quality Value (QV) of 52.6 as calculated by Merqury v1.3 (Rhie *et al.*, 2020).

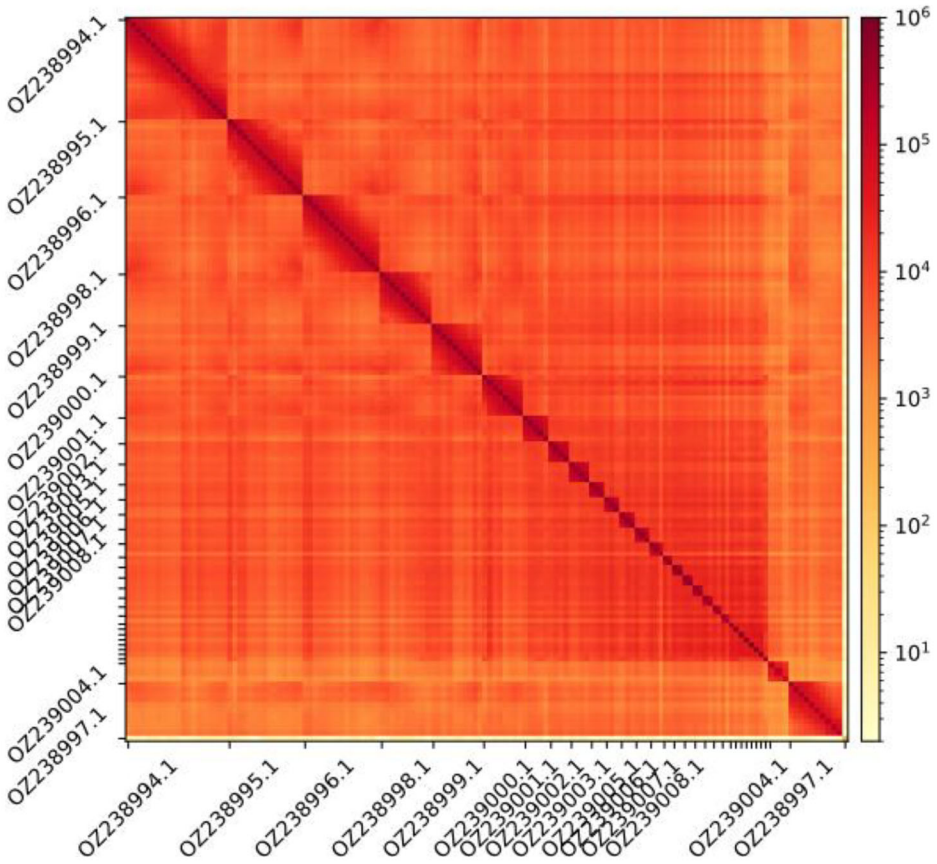


Figure 3. Hi-C contact map showing spatial interactions between regions of the genome. The diagonal corresponds to intra-chromosomal contacts, depicting chromosome boundaries. The frequency of contacts is shown on a logarithmic heatmap scale. Hi-C matrix bins were merged into a 150 kb bin size for plotting. Due to space constraints on the axes, only the GenBank names of the 13th largest autosomes and the sex chromosomes are shown.

Author contributions

RJL coordinated the project, collected the species, identified the species, sampled and preserved biological material and provided metadata, RM, TM, RAO, THS, and AsB provided sampling and metadata support and management, LA and MG extracted DNA, prepared libraries, and performed sequencing, FCF, JGG and FC performed genome assembly and curation under the supervision of TA, RM generated the analysis and report. All authors contributed to the writing, review, and editing of this genome note and read and approved the final version.

Data availability

Pyrrhula murina and the related genomic study were assigned to Tree of Life ID (ToLID) 'bPyrMur1', and all sample, sequence, and assembly information are available under the umbrella BioProject PRJEB86373 from https://www.ebi.ac.uk/ena/browser/view/PRJEB86373. The sample information is available at the following BioSample accessions: SAMEA115966741 and SAMEA115966742. The genome assembly is accessible from ENA under accession number GCA_965183895.

Sequencing data produced as part of this project are available from ENA at the following accessions: ERX14064180, ERX14064181, and ERX14064182. Documentation related to the genome assembly and curation can be found in the ERGA Assembly Report (EAR) document available at https://github.com/ERGA-consortium/EARs/tree/main/Assembly_Reports/Pyrrhula_murina/bPyrMur1. The mitochondrial genome was assembled into a single circular contig using the FOAM pipeline v0.5 (https://github.com/cnag-aat/FOAM). Further details and data about the project are hosted on the ERGA portal at https://portal.erga-biodiversity.eu/data_portal/928672.

Acknowledgements

We acknowledge the support of the Portuguese Society for the Study of Birds (SPEA) for the logistical support during the sampling. We acknowledge the support of the Freiburg Galaxy Team: Saim Momin and Björn Grüning, Bioinformatics, University of Freiburg (Germany), funded by the German Federal Ministry of Education and Research, BMBF grant 031 A538A de. NBI-RBC and the Ministry of Science, Research and the Arts Baden-Württemberg (MWK) within the framework of LIBIS/de. NBI Freiburg.

References

Aubrecht G: **The Azores Bullfinch - *Pyrrhula murina* GODMAN, 1866 The history of a bird species: persecuted - missing - rediscovered - protected (?) (including a list of all known specimens and syntypes).** *Ann Nat Mus Wien.* 2000; **102**: 23–62.
Reference Source

BirdLife International: **Species factsheet: Azores Bullfinch *Pyrrhula murina.*** 2021.
Reference Source

Ceia RS, Ramos JA, Heleno RH, *et al.*: **Status assessment of the critically endangered Azores Bullfinch *Pyrrhula murina.*** *Bird Conserv Int.* 2011; **21**(4): 477–489.
Publisher Full Text

Challis R, Kumar S, Sotero-Caio C, *et al.*: **Genomes on a Tree (GoaT): a versatile, scalable search engine for genomic and sequencing project metadata across the eukaryotic Tree of Life [version 1; peer review: 2 approved].** *Wellcome Open Res.* 2023; **8**:24. 24.
PubMed Abstract | Publisher Full Text | Free Full Text

Costa TMM, Gil A, Timóteo S, *et al.*: **How many Azores Bullfinches (*Pyrrhula murina*) are there in the world? case study of a threatened species.** *Diversity.* 2023; **15**(5): 685.
Publisher Full Text

De Panis D: **ASM analyses (one-asm WGS Illumina PE + HiC).** *WorkflowHub.* 2024.
Reference Source

Gil A, Ceia R, Coelho R, *et al.*: **The Priolo Atlas: a citizen science-based census initiative for supporting *Pyrrhula murina* habitat conservation and restoration policies in São Miguel Island (Azores, Portugal).** *Ecol Eng.* 2016; **86**: 45–52.
Publisher Full Text

Gomez-Garrido J: *CLAWS (CNAG's Long-read Assembly Workflow in Snakemake) [Computer software].* 2024.
Publisher Full Text

Lifjeld JT, Hoenen A, Johannessen LE, *et al.*: **The Azores bullfinch (*Pyrrhula murina*) has the same unusual and size-variable sperm morphology as the Eurasian bullfinch (*Pyrrhula pyrrhula*).** *Biol J Linn Soc.* 2013; **108**(3): 677–687.
Publisher Full Text

Manni M, Berkeley MR, Seppey M, *et al.*: **BUSCO: assessing genomic data quality and beyond.** *Curr Protoc.* 2021; **1**(12): e323.
PubMed Abstract | Publisher Full Text

Mazzoni CJ, Ciofi C, Waterhouse RM: **Biodiversity: an Atlas of European reference genomes.** *Nature.* 2023; **619**(7969): 252.
PubMed Abstract | Publisher Full Text

Rhie A, Walenz BP, Koren S: **Merqury: reference-free quality, completeness, and phasing assessment for genome assemblies.** *Genome Biol.* 2020; **21**(1): 245.
PubMed Abstract | Publisher Full Text | Free Full Text

Töpfer T, Haring E, Birkhead TR, *et al.*: **A molecular phylogeny of bullfinches *Pyrrhula* Brisson, 1760 (Aves: Fringillidae).** *Mol Phylogenet Evol.* 2011; **58**(2): 271–282.
PubMed Abstract | Publisher Full Text

Open Peer Review

Current Peer Review Status: ✔ ✔ ?

Version 3

Reviewer Report 15 April 2026

https://doi.org/10.21956/openreseurope.25295.r71045

© **2026 Whibley A.** This is an open access peer review report distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

✔ **Annabel Whibley**
[1] University of Auckland, Auckland, New Zealand
[2] Grapevine Improvement, Bragato Research Institute, Lincoln, New Zealand

No additional comments

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Genomics, Bioinformatics

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.

Reviewer Report 13 March 2026

https://doi.org/10.21956/openreseurope.25295.r71044

© **2026 Silver L.** This is an open access peer review report distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

✔ **Luke W Silver**
University of Sydney, Sydney, Australia

No further comments

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Genome assembly, genetic diversity

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.

Version 2

Reviewer Report 12 March 2026

https://doi.org/10.21956/openreseurope.24922.r69658

© 2026 Chase M. This is an open access peer review report distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

?

Madeline Chase
[1] Swiss Ornithological Institute, Sempach, Switzerland
[2] Swiss Ornithological Institute, Sempach, Switzerland

The authors generate a high quality, chromosome scale genome assembly for the Azores Bullfinch. This is a vulnerable species with a restricted range and low population size, and the reference genome may aid conservation efforts.

The authors generally provide sufficient details on their methods; however, citations are missing for several of the software used in the genome assembly section that should be provided. Hifiasm and YaHS for instance should include citations.

One other note is that in the introduction, this sentence needs revision: "Of particular importance for the viability of the species, is the fact that Pyrrhula murina, along with his sister species,"

"*Pyrrhula murina*" should be in italics and "his" should be their or its.

Is the rationale for creating the dataset(s) clearly described?
Yes

Are the protocols appropriate and is the work technically sound?
Yes

Are sufficient details of methods and materials provided to allow replication by others?
Partly

Are the datasets clearly presented in a useable and accessible format?
Yes

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Population and speciation genomics

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however I have significant reservations, as outlined above.

Reviewer Report 27 February 2026

https://doi.org/10.21956/openreseurope.24922.r69516

© 2026 Whibley A. This is an open access peer review report distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Annabel Whibley
[1] University of Auckland, Auckland, New Zealand
[2] Grapevine Improvement, Bragato Research Institute, Lincoln, New Zealand
[3] University of Auckland, Auckland, New Zealand
[4] Grapevine Improvement, Bragato Research Institute, Lincoln, New Zealand

The authors have satisfactorily addressed my earlier concerns and I express my thanks for their engagement and efforts.

Minor comments:
Please note that "*Pyrrhula murina*" needs italicizing in the 4th paragraph of the introduction.

I would also like to see the basecalling model and dorado version specified for the ONT sequencing (my apologies for not flagging this earlier).

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Genomics, Bioinformatics

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.

Author Response 05 Mar 2026

Rita Monteiro

We thank the reviewer for these suggestions. 1.Corrected 2. Added to the main text: Basecalling was performed using Dorado v7.4.14 (Oxford nanopore Technologies) with the super accurate (SUP) basecalling model v4.3.0 at 400 bps, integrated within MinKNOW v24.06.15.

Competing Interests: No competing interests were disclosed.

Reviewer Report 27 February 2026

https://doi.org/10.21956/openreseurope.24922.r69517

© **2026 Silver L.** This is an open access peer review report distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

?

Luke W Silver
[1] University of Sydney, Sydney, Australia
[2] University of Sydney, Sydney, Australia

The authors have addressed all of my concerns and the additional methods included have improved the manuscripts. One final comment is that the authors have undertaken repeatmasking of the assembly, could they please include the methods and the results (in the form of a table of repeat types) in the manuscript
Competing Interests: No competing interests were discl

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Genome assembly, genetic diversity

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however I have significant reservations, as outlined above.

Author Response 05 Mar 2026

Rita Monteiro

We thank the reviewer for this suggestion. However, ERGA-BGE genome reports are specifically scoped to present the assembly and its core metrics, rather than provide detailed genomic analyses. Additionally, including a full repeat breakdown could conflict with more in-depth work being carried out by the sample provider for future publications. We have therefore respectfully maintained the current level of detail in this section.

Competing Interests: No competing interests were disclosed.

Version 1

Reviewer Report 12 September 2025

https://doi.org/10.21956/openreseurope.22356.r57666

© 2025 Whibley A. This is an open access peer review report distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

?

Annabel Whibley
[1] University of Auckland, Auckland, New Zealand
[2] Grapevine Improvement, Bragato Research Institute, Lincoln, New Zealand
[3] University of Auckland, Auckland, New Zealand
[4] Grapevine Improvement, Bragato Research Institute, Lincoln, New Zealand
[5] University of Auckland, Auckland, New Zealand
[6] Grapevine Improvement, Bragato Research Institute, Lincoln, New Zealand

Lopes and colleagues describe the generation of a reference genome assembly for the Azores Bullfinch, *Pyrrhula murina*, a species classified as Vulnerable by the IUCN. The quality of the genome is high, with >96% BUSCO completeness and a base-level accuracy of 52.6QV. The assembly has been scaffolded into chromosomal pseudomolecules, although no comment is made on the discrepancy between the 42 pseudochromosomes recovered and the expected 2n=78 nor to whether any of the scaffolds/contigs remain unlinked to chromosomes.

The assembly and supporting data have been publicly released and the accession links are active. The authors describe obtaining RNA sequencing data but an annotation is not provided here.

While the manuscript is generally well-structured and is close to the Data Note template used by the Darwin Tree of Life consortium, there are notable gaps in the reporting of methods and important metadata details. For example, the methods section presents a genome size estimate from kmer analysis without noting how this method was obtained (GenomeScope2 with a kmer length of X?). The versions of software used are not always reported and key parameters are sometimes omitted. What are the parameters used to filter the sequencing reads? What are the properties (beyond coverage) of the ONT libraries? The flow cell type (R10.4.1) should probably be reported for the ONT sequencing along with the library type. Methods reporting manual curation would typically also give an indication of the number and type of interventions.

Some of the language is a little informal. For example, the opening of the abstract with a possessive rather than “The reference genome of…” is unusual. And “This kind of high-quality…” in the introduction is vague.

This sentence needs revision: “Of particular importance for its viability, is the fact that, along with his sister species, the Eurasian Bullfinch, *Pyrrhula pyrrhula*, they have relatively small and possibly neotenous sperm, an ancestral trait that evolved before the two taxa diverged (Lifjeld et al., 2013).” I would suggest : Of particular importance for **the viability of the species**, is the fact that ***Pyrrhula murina*****,** along with **a** sister species, the Eurasian Bullfinch, *Pyrrhula pyrrhula*, have relatively small and possibly neotenous sperm, an ancestral trait that evolved before the two taxa diverged (Lifjeld et al., 2013).”

The introduction of the methods describes RNA sequencing data without noting the sequencing type and states that it is used to facilitate genome assembly and annotation, with scant

methodological evidence that the RNA dataset has informed the assembly generation.

The latitude and longitude co-ordinates for the collection site should be included.

Is the rationale for creating the dataset(s) clearly described?
Yes

Are the protocols appropriate and is the work technically sound?
Yes

Are sufficient details of methods and materials provided to allow replication by others?
Partly

Are the datasets clearly presented in a useable and accessible format?
Yes

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Genomics, Bioinformatics

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however I have significant reservations, as outlined above.

Author Response 29 Jan 2026

Rita Monteiro

We thank the reviewer for their comments. We have addressed all of them as following:

1. The discrepancy between the previously reported of 39 pseudochromosomes (2n=78) and the 42 pseudochromosomes recovered in our assembly can be explained by differences in how the chromosome number was inferred. The number 2n=78 pseudochromosomes was obtained from GOAT, which estimates it based on comparative karyotype data from ancestral taxa (as indicated in the main text) rather than direct cytogenetic observation for this species. We have now modified the text to clarify this point better. In contrast, our assembly provides species specific evidence, revealing 40 pairs of autosomes plus two sex chromosomes. In addition, 11 scaffolds remain unplaced in the current assembly.
2. As noted in our response to the other reviewer, a full genome annotation for this species is not yet available: it is currently being generated and will be released as soon as completed.
3. We thank the reviewer for pointing this out. As suggested we have now included the genome size estimation method - GenomeScope2 with a k-mer length of 21. In addition, we have also included the corresponding software versions BUSCO v5.5.0 and Merqury v1.3, as also suggested by the other reviewer. Furthermore, we now report the number and kind of interventions made during manual curation.
4. We have modified the highlighted examples to improve clarity and formality.

5. Modified accordingly.
6. We have clarified the text, separating the RNA-seq data production from the rest of the data, highlighting that the purpose of its production is to aid future annotation efforts. We include transcriptomic data for the purpose of efficiency. Collecting RNA from the same specimen (or from different specimens) during the same sampling event is important and saves time, effort and costs when it comes to annotation of the reference genomes produced. The annotation step is downstream of the assembly and can take substantially more time, thus we release the assemblies as soon as possible and decouple the annotation step from necessary inclusion in the Genome Note.
7. All relevant metadata, including the collection site and coordinates are are available under the umbrella BioProject PRJEB86373 (https://www.ebi.ac.uk/ena/browser/view/PRJEB86373) and in the ERGA portal (https://portal.erga-biodiversity.eu/data_portal/928672). Both links are included in the main text.

Competing Interests: No competing interests were disclosed.

Reviewer Report 28 August 2025

https://doi.org/10.21956/openreseurope.22356.r57670

© 2025 Silver L. This is an open access peer review report distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

?

Luke W Silver
[1] University of Sydney, Sydney, Australia
[2] University of Sydney, Sydney, Australia
[3] University of Sydney, Sydney, Australia

Reviewer Comments
The authors generate a draft genome for the Azores Bullfinch, a IUCN Vulnerably listed species, using ONT, HiC and Illumina data. Overall the methods of genome assembly and quality checking are sound and will provide a useful resource, however, there are a few details which would improve the data note.

Include sequencing technology used in the Abstract
Where is the Azores archipelago located in wider context – i.e. XX km west of Portugal in the Atlantic Ocean or include a figure of the location, maybe a panel with an image of the Azores Bullfinch
“As a result of” instead of “due to”
“A high quality..” instead of “this kind of...”

"Sequencing" instead of "used"
What type of RNA short read libraries – 150bp paired end?
How many reads were generated by RNA sequencing?
Say which data type was used in each of the steps – i.e. scaffolding using HiC data with YaHS...
Some details of the assembly summary analysis is needed, such as the which version of BUSCO and Merqury are used in this workflow?
Was any repeat identification undertaken?
Which scaffold was identified as the mitochondrial genome?
Method details are needed for how the sex scaffolds were identified
Were the RNA seq reads used for any analysis or just generated for ENA submission?
If there are any other genomes for closely related species is it expected that there are 13 macrochromosomes and then microchromosomes or is the number of macrochromosomes unknown?

Is the rationale for creating the dataset(s) clearly described?

Yes

Are the protocols appropriate and is the work technically sound?

Yes

Are sufficient details of methods and materials provided to allow replication by others?

Partly

Are the datasets clearly presented in a useable and accessible format?

Yes

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Genome assembly, genetic diversity

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however I have significant reservations, as outlined above.

Author Response 29 Jan 2026

Rita Monteiro

We thank the reviewer for their comments. We have addressed all of them as following:

1. Added to the Abstract
2. We included a new figure (Figure 1) that shows the wider spatial context in relation to the nearest continental grounds, and the location of the Azores Bullfinch range inside the São Miguel Island. We also include a photo of the Azores Bullfinch.
3. Modified accordingly.
4. Modified accordingly.
5. Modified accordingly.
6. Were produced 150bp paired end RNA libraries. This information was added to the

main text.

7. 91MR were generated, added to the text.
8. We assembled contigs from ONT data with Hi-C for phasing. We scaffolded with YaHS using Hi-C data. K-mer counts for estimating genome size, QV and k-mer completeness were performed on the combined set of Illumina WGS and filtered ONT reads. This information was added to the text.
9. For the assembly summary analysis we used BUSCO v5.5.0 and Merqury v1.3. This information was added to the main text.
10. Yes, repeats were identified using RepeatMasker (Smit, AFA, Hubley, R & Green, P., 2013-2015), (version 4.0.5 with parameters -nolow -engine "RMBlast"), dustmasker (Morgulis, A., et al., 2006), and tandem repeat finder, TRF (Benson, G., 1999). In addition to the Repbase (Bao, W., et al, 2015) library, where available, a custom repeat library is used with RepeatMasker. Custom libraries are created using RepeatModeler2 (Flynn, J.M., 2020)
11. The mitochondrial genome corresponds to chr OZ239036.1. This was added to the main text.
12. We have added the following text to the methods:
"The sex chromosomes were identified by sequencing coverage. Assignment of Z and W chromosomes was according to conservation of length: bPyrMur chrZ (OZ238997.1: 86793913 bp) vs. bGalGal chrZ (CM028522.1: 86044486 bp). Whole genome alignment against GRCg7b (GCA_016699485.1) confirmed these assignments."
13. As indicated, RNA sequencing was performed for transcriptome profiling with the primary aim of supporting genome assembly and annotation. Although a full genome annotation for this species is not yet available, it is currently being generated and will be released as soon as it becomes available.
14. We did not include any information with respect to macro- vs. microchromosomes as there is some debate in the literature as to the length threshold (35 Mb? 30? 20 Mb?). It seems that the categorization is done by karytyping and varies from one bird species to the next. For P. murina, the clearest size break occurs between about 23 and 28 Mb. It has 10 autosomes and the Z and W that are longer than 28 Mb and the rest (30 autosomes) are shorter than 23 Mb, the shortest being about 3.5 Mb. We have added this information to the text but avoid the macro- micro- labels.

Competing Interests: No competing interests were disclosed.