Open Research Europe

Open Research Europe 2025, 5:210 Last updated: 01 AUG 2025

DATA NOTE

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 1; peer review: awaiting peer review]

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

v1 **First published:** 01 Aug 2025, **5**:210
https://doi.org/10.12688/openreseurope.20666.1

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

Open Peer Review

Approval Status *AWAITING PEER REVIEW*

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

Abstract

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. This chromosome-level assembly encompasses 1.2 Gb, composed of 65 contigs and 60 scaffolds, with contig and scaffold N50 values of 64.8 Mb and 76 Mb, respectively.

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.

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: © 2025 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 1; peer review: awaiting peer review]** Open Research Europe 2025, **5**:210 https://doi.org/10.12688/openreseurope.20666.1

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

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. 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 due to 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.

This kind of 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 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).

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 polishing (i.e. recommended for ONT-only assemblies), and RNA sequencing for transcriptomic profiling, to facilitate genome assembly 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 is 1.08 Gb. This is a diploid genome with a haploid number of 39 chromosomes ($2n = 78$), including Z and W sex chromosomes in females. 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 used.

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 used.

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 on a PromethION 24 A Series instrument (ONT). 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). The short-read libraries were sequenced on a NovaSeq 6000 instrument (Illumina).

In total, 100x Oxford Nanopore, 99x Illumina WGS shotgun, and 79x HiC data were sequenced to generate the assembly.

Genome assembly methods

The genome was assembled using the CNAG CLAWS pipeline v2.2.0 (Gomez-Garrido, 2024). Briefly, reads were preprocessed for quality and length using Trim Galore v0.6.7 and Filtlong v0.2.1, and initial contigs were assembled using NextDenovo v2.5.0, followed by polishing of the assembled contigs using HyPo v1.0.3, removal of retained haplotigs using purge-dups v1.2.6 and scaffolding with YaHS v1.2a. Finally, 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. Finally, the mitochondrial genome 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). Summary analysis of the released assembly was performed using the ERGA-BGE Genome Report ASM Galaxy workflow (De Panis, 2024a).

Results

Genome assembly

The genome assembly has a total length of 1,163,884,374 bp in 61 scaffolds, including the mitogenome (Figure 1 and Figure 2), 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 single-copy gene content analysis using the aves_odb10 database with BUSCO (Manni *et al.*, 2021) resulted in 96.9% completeness (96.3% single and 0.6%

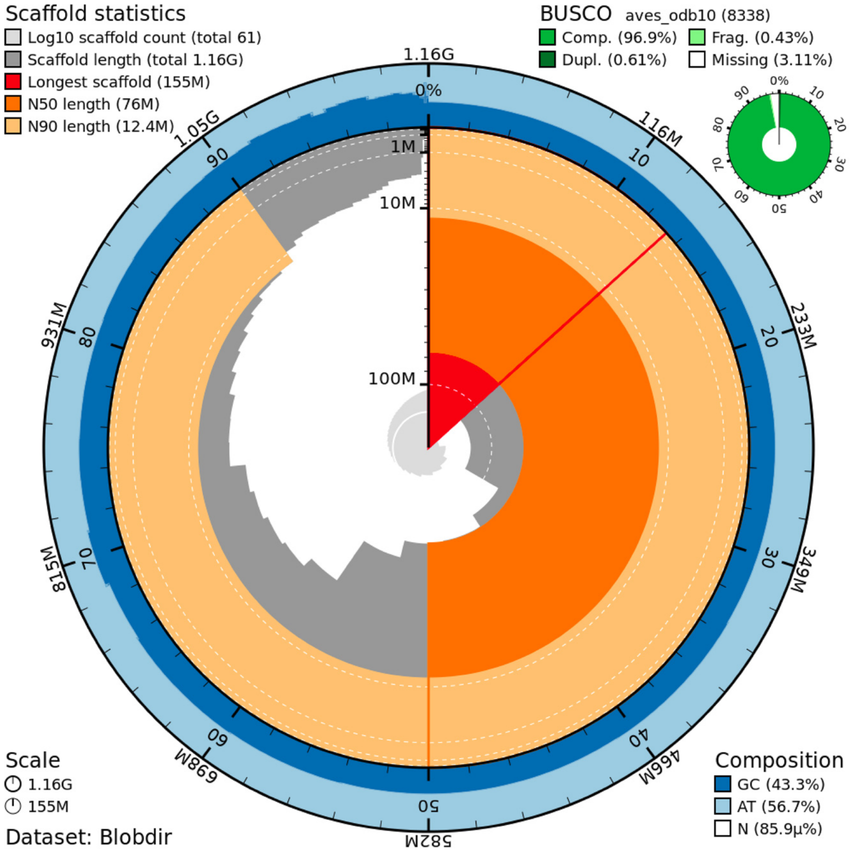

Figure 1. Snail plot summary of assembly statistics. The main plot is divided into 1,000 size-ordered bins around the circumference, with each bin representing 0.1% of the 1,163,884,374bp assembly, including the mitochondrial genome. The distribution of sequence lengths is shown in dark grey, with the 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.

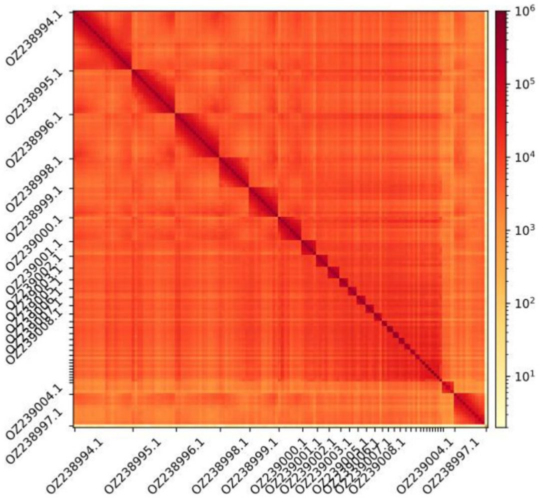

Figure 2. Hi-C contact map showing spatial interactions between regions of the genome. The diagonal corresponds to intra-chromosomal contacts, depicting chromosome boundaries. The frequency of contacts is shown on a logarithmic heatmap scale. Hi-C matrix bins were merged into a 150 kb bin size for plotting. Due to space constraints on the axes, only the GenBank names of the 13th largest autosomes and the sex chromosomes are shown.

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 (Rhie *et al.*, 2020).

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.

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.

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.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

Costa TdMM, 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: **ERGA-BGE Genome Report ANNOT analyses.** WorkflowHub, 2024a.
Publisher Full Text

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**