Check for updates

DATA NOTE

ERGA-BGE genome of the ironclad beetle *Tarphius canariensis* Wollaston, 1862

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

Brent C Emerson[1], María Conejero[2], Ana Riesgo[2], Víctor Noguerales[1,3], Rita Monteiro[4], Astrid Böhne[4], Javier Palma Guerrero[5], Rosa Fernández[5], Wellcome Sanger Institute Tree of Life Management, Samples and Laboratory team, Wellcome Sanger Institute Scientific Operations: Sequencing Operations, Wellcome Sanger Institute Tree of Life Core Informatics team, Leanne Haggerty[6], Fergal Martin[6], Tom Brown[7,8]

[1]Island Ecology and Evolution Research Group, Institute of Natural Products and Agrobiology (IPNA-CSIC), San Cristóbal de La Laguna, Santa Cruz de Tenerife, Spain
[2]Museo Nacional de Ciencias Naturales- CSIC, Madrid, Spain
[3]Current address: Universidad de La Laguna (ULL), Departamento de Biología Animal, Edafología y Geología, San Cristóbal de La Laguna, Canary Islands, Spain
[4]Leibniz Institute for the Analysis of Biodiversity Change - Museum Koenig Bonn, Bonn, Germany
[5]Institute of Evolutionary Biology (UPF-CSIC), Barcelona, Spain
[6]European Molecular Biology Laboratory, European Bioinformatics Institute, Cambridge, UK
[7]Leibniz Institute for Zoo and Wildlife Research, Berlin, Germany
[8]Berlin Center for Genomics in Biodiversity Research (BeGenDiv), Koenigin-Luise-Str 6-8, Berlin, 14195, Germany

v1 **First published:** 05 Aug 2025, **5**:219
https://doi.org/10.12688/openreseurope.20864.1
Latest published: 05 Aug 2025, **5**:219
https://doi.org/10.12688/openreseurope.20864.1

Abstract

We present a genome assembly from a specimen of *Tarphius canariensis* (Ironclad beetle; Arthropoda; Insecta; Coleoptera; Zopheridae), a representative species from one of the more notable arthropod radiations within the Canary Islands. The entirety of the genome sequence was assembled into 10 contiguous chromosomal pseudomolecules. This chromosome-level assembly encompasses 311 Mb, composed of 50 contigs and 13 scaffolds, with contig and scaffold N50 values of 14.5 Mb and 31.1 Mb, respectively.

Keywords

Tarphius canariensis, genome assembly, European Reference Genome Atlas, Biodiversity Genomics Europe, Earth Biogenome Project, Ironclad beetle

Open Peer Review

Approval Status ✓ ?

	1	2
version 1 05 Aug 2025	✓ view	? view

1. **Merly Escalona**, University of Connecticut Department of Molecular and Cell Biology, Storrs, USA

2. **Valentina Peona**, Uppsala University, uppsala, Sweden
Naturhistoriska riksmuseet (Ringgold ID: 42763), Stockholm, Sweden

Any reports and responses or comments on the

article can be found at the end of the article.

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: Tom Brown (brown@izw-berlin.de)

Author roles: Emerson BC: Investigation, Resources, Writing – Original Draft Preparation, Writing – Review & Editing; **Conejero M**: Investigation, Resources, Writing – Review & Editing; **Riesgo A**: Investigation, Resources, Writing – Review & Editing; **Noguerales V**: Investigation, Resources, Writing – Review & Editing; **Monteiro R**: Methodology, Project Administration, Supervision, Writing – Review & Editing; **Böhne A**: Funding Acquisition, Methodology, Project Administration, Supervision, Writing – Review & Editing; **Guerrero JP**: Methodology, Project Administration, Supervision, Writing – Review & Editing; **Fernández R**: Funding Acquisition, Methodology, Project Administration, Supervision, Writing – Review & Editing; **Haggerty L**: Data Curation, Formal Analysis, Writing – Review & Editing; **Martin F**: Data Curation, Formal Analysis, Supervision, Writing – Review & Editing; **Brown T**: Visualization, Writing – Original Draft Preparation

Competing interests: No competing interests were disclosed.

Grant information: This project received funding from Horizon Europe under the Biodiversity, Circular Economy and Environment (REA.B.3) No 101059492; 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.
The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Copyright: © 2025 Emerson BC *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: Emerson BC, Conejero M, Riesgo A *et al.* **ERGA-BGE genome of the ironclad beetle *Tarphius canariensis* Wollaston, 1862 [version 1; peer review: 1 approved, 1 approved with reservations]** Open Research Europe 2025, **5**:219 https://doi.org/10.12688/openreseurope.20864.1

First published: 05 Aug 2025, **5**:219 https://doi.org/10.12688/openreseurope.20864.1

Introduction

The family Zopheridae Solier, 1834, commonly referred to as ironclad beetles, is represented in the Canary Islands by thirty-one endemic species from the genus *Tarphius* (Emerson & Oromí, 2005). This makes *Tarphius* the seventh most species-rich arthropod genus within the archipelago, and a paradigm for diversification within forest habitats of the macaronesian region. Species within the family Zopheridae feed on rotting wood, or fungus associated with rotten wood. It seems likely that high species sympatry within Tenerife and other islands is facilitated by microhabitat and/or dietary partitioning, as appears to be the case between *Tarphius canariensis* and *Tarphius simplex*. On the island of Tenerife alone, the genus *Tarphius* is represented by 14 species that appear to have descended from a single common Tenerifean ancestor (Emerson & Oromí, 2005). The genus is considered an enigmatic radiation, as nearly all species occur almost exclusively within the geographically restricted cloud forest habitat of the islands, with extensive sympatry among species. Ecological data is limited for most species, but subtle ecological differentiation can be detected. Notably, the closely related sympatric sister species *Tarphius canariensis* and *T. simplex* differ in their vertical stratification, with *T. canariensis* being more commonly found in above-ground vegetation, while *T. simplex* is more soil associated.

Limited phylogenetic work to date based on mtDNA sequence data (Emerson *et al.*, 2000; Emerson & Oromí, 2005) has revealed extensive mtDNA paraphyly coupled with limited mtDNA divergences among species representing extreme phenotypic variation within the genus. A recent analysis of the *T. canariensis* species complex within the island of Tenerife using genome-wide nuclear data has revealed hybridisation within the speciation history of the complex (Noguerales & Emerson, 2025), with evidence for genomic admixture between parental species as a potential driver of speciation.

The almost ubiquitous association between Canary Island *Tarphius* species and cloud forest habitat highlights strong niche constraints on speciation within the genus. It thus stands out as striking that a recent speciation event on the island of La Palma has involved a transition from lower elevation cloud forest habitat (600-1100 masl) to higher elevation alpine habitat (>2200 masl; Emerson *et al.*, 2000). The *Tarphius canariensis* complex is distributed across three islands, and has its origins on the island of Tenerife (TF), from where the neighbouring islands of Gran Canaria (GC) and La Palma (LP) were colonised (Emerson & Oromí, 2005). There populations established into cloud forest habitat analogous to that of the founding population from Tenerife. The colonisation and establishment of *T. canariensis* on La Palma was then followed by the evolution of *T. supranubius*. Thus, divergence by founding events into similar habitat (cloud forest) among the three morphologically consistent island populations of *T. canariensis* is contrasted with peripatric or founder event divergence into a novel habitat (alpine), giving rise to a new ecologically and phenotypically distinct species. The *T. canariensis* complex thus provides an ideal opportunity to contrast (i) genomic change associated with a speciation event implicating divergent selection and adaptation (*T. canariensis* LP – *T. supranubius* LP), with (ii) genomic change where divergent selection is expected to play a much more limited role (*T. canariensis* TF – *T. canariensis* GC, and *T. canariensis* TF – *T. canariensis* LP).

Insular arthropod species radiations stand out when placed within the broader taxonomic context of insular diversification, because the majority of lineages within insular systems typically do not diversify, or do so to only a very limited extent (Patiño *et al.*, 2017). Put another way, within a common environmental setting, some lineages have responded with extensive diversification, while others have not. Insular systems are thus fertile ground to explore intrinsic variables that may underpin why some lineages diversify, while others do not (Patiño *et al.*, 2017). As such, the generation of a reference genome for *T. canariensis* will provide a valuable resource to advance our understanding of the genomic underpinnings of arthropod diversification.

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

Sample and sampling information

Brent C Emerson and Víctor Noguerales sampled 10 specimens of *Tarphius canariensis* (Figure 1) from the Canary Islands, Tenerife, Spain on 20th February 2023. Specimens were identified by Víctor Noguerales based on morphology, and barcoding from one individual from the same location and sampling event. Sampling was performed under permission



Figure 1. Image of a *Tarphius canariensis* specimen. An example of a specimen of *Tarphius canariensis*. Photo credit: Mario Aponte.

Sigma 133-23 AFF 17-23 issued by Servicio Técnico de Gestión Ambiental, Cabildo Insular de Tenerife. Sampling was performed at night via visual sampling and beating vegetation.

Specimens were euthanized by placement on cold Petri dish and subsequently preserved in liquid nitrogen. Until DNA extraction, samples were preserved at -80°C.

Vouchering information

Physical reference material for a proxy individual has been deposited in Museo Nacional de Ciencias Naturales de Madrid (https://www.mncn.csic.es/es/colecciones/cientificas/entomologia-col) under the accession number MNCN-Voucher: 110998-110999.

Frozen reference tissue material of a whole individual is available from a proxy voucher at the Biobank of the Museo Nacional de Ciencias Naturales de Madrid (https://www.mncn.csic.es/es/colecciones/cientificas/tejidos-y-adn-biobanco, accessible through https://gbif.es/ and https://www.ggbn.org/) under the voucher ID MNCN-ADN-110999.

Genetic information

The estimated genome size, based on ancestral taxa, is 450 Mb. This is a diploid genome with a haploid number of 10 chromosomes (2n=20) and unknown sex chromosomes. All information for this species was retrieved from Genomes on a Tree (Challis *et al.*, 2023).

DNA/RNA processing

The workflow for high molecular weight (HMW) DNA extraction at the Wellcome Sanger Institute (WSI) Tree of Life Core Laboratory includes a sequence of procedures: sample preparation and homogenisation, DNA extraction, fragmentation and purification. Detailed protocols are available on protocols.io (Denton *et al.*, 2023b). The icTarCana1 sample was prepared for DNA extraction by weighing and dissecting it on dry ice (Jay *et al.*, 2023). Tissue from the whole organism was homogenised using a PowerMasher II tissue disruptor (Denton *et al.*, 2023a). HMW DNA was extracted using version 3 of the Manual MagAttract protocol (Strickland *et al.*, 2023). DNA was sheared into an average fragment size of 12–20 kb in a Megaruptor 3 system (Bates *et al.*, 2023). Sheared DNA was purified by solid-phase reversible immobilisation, using AMPure PB beads to eliminate shorter fragments and concentrate the DNA (Oatley *et al.*, 2023). The concentration of the sheared and purified DNA was assessed using a Nanodrop spectrophotometer and Qubit Fluorometer using the Qubit dsDNA High Sensitivity Assay kit. Fragment size distribution was evaluated by running the sample on the FemtoPulse system.

RNA was extracted from whole organism tissue of icTarCana3 in the Tree of Life Laboratory at the WSI using the RNA Extraction: Automated MagMax™ mirVana protocol (do Amaral *et al.*, 2023). The RNA concentration was assessed using a Nanodrop spectrophotometer and a Qubit Fluorometer using the Qubit RNA Broad-Range Assay kit. Analysis of the integrity of the RNA was done using the Agilent RNA 6000 Pico Kit and Eukaryotic Total RNA assay.

Whole organism tissue from the icTarCana2 sample was processed at the WSI Scientific Operations core, using the Arima-HiC v2 kit. Tissue (stored at –80 °C) was fixed, and the DNA crosslinked using a TC buffer with 22% formaldehyde. After crosslinking, the tissue was homogenised using the Diagnocine Power Masher-II and BioMasher-II tubes and pestles. Following the kit manufacturer's instructions, crosslinked DNA was digested using a restriction enzyme master mix. The 5'-overhangs were then filled in and labelled with biotinylated nucleotides and proximally ligated. An overnight incubation was carried out for enzymes to digest remaining proteins and for crosslinks to reverse. A clean up was performed with SPRIselect beads prior to library preparation.

Library preparation and sequencing

Library preparation and sequencing were performed at the WSI Scientific Operations core.

At the minimum, samples were required to have an average fragment size exceeding 8 kb and a total mass over 400 ng to proceed to the low input SMRTbell Prep Kit 3.0 protocol (Pacific Biosciences, California, USA), depending on genome size and sequencing depth required. Libraries were prepared using the SMRTbell Prep Kit 3.0 (Pacific Biosciences, California, USA) as per the manufacturer's instructions. The kit includes the reagents required for end repair/A-tailing, adapter ligation, post-ligation SMRTbell bead cleanup, and nuclease treatment. Following the manufacturer's instructions, size selection and clean up was carried out using diluted AMPure PB beads (Pacific Biosciences, California, USA). DNA concentration was quantified using the Qubit Fluorometer v4.0 (Thermo Fisher Scientific) with Qubit 1X dsDNA HS assay kit and the final library fragment size analysis was carried out using the Agilent Femto Pulse Automated Pulsed Field CE Instrument (Agilent Technologies) and gDNA 55kb BAC analysis kit. Prepared libraries were normalised to 2nM and 15µL used for making complexes. For libraries below 2nM all 10uL was used for making complexes. Primers were annealed and polymerases were hybridised to create circularised complexes according to manufacturer's instructions. The complexes were purified with the 1.2X clean up with SMRTbell beads. The purified complexes were then diluted to the Revio loading concentration, in the range of 200–300 pM, and spiked with a Revio sequencing internal control.

Samples were sequenced using the Revio system on Revio 25M SMRT cells (Pacific Biosciences, California, USA). The SMRT link software, a PacBio web-based end-to-end workflow manager, was used to set-up and monitor the run, as well as perform primary and secondary analysis of the data upon completion.

For Hi-C library preparation, DNA was fragmented to a size of 400 to 600 bp using a Covaris E220 sonicator. The DNA was then enriched, barcoded, and amplified using the NEBNext Ultra II DNA Library Prep Kit following manufacturers' instructions. The Hi-C sequencing was performed using paired-end sequencing with a read length of 150 bp on an Illumina NovaSeq X instrument.

Poly(A) RNA-Seq libraries were constructed using the NEB Ultra II RNA Library Prep kit, following the manufacturer's instructions. RNA sequencing was performed on the Illumina NovaSeq X instrument.

Genome assembly methods

The HiFi reads were first assembled using Hifiasm v0.19.8-r603 (Cheng *et al.*, 2021) with the --primary option. The Hi-C reads were mapped to the primary contigs using bwa-mem2 (Vasimuddin *et al.*, 2019). The contigs were further scaffolded using the provided Hi-C data (Rao *et al.*, 2014) in YaHS v1.2a.2 (Zhou *et al.*, 2023) using the --break option for handling potential misassemblies. The scaffolded assemblies were evaluated using Gfastats (Formenti *et al.*, 2022), BUSCO (Manni *et al.*, 2021) and MERQURY.FK (Rhie *et al.*, 2020).

The mitochondrial genome was assembled using MitoHiFi v3 (Uliano-Silva *et al.*, 2023), which runs MitoFinder (Allio *et al.*, 2020) and uses these annotations to select the final mitochondrial contig and to ensure the general quality of the sequence.

The assembly was decontaminated using the Assembly Screen for Cobionts and Contaminants (ASCC) pipeline. Flat files and maps used in curation were generated in TreeVal (Pointon *et al.*, 2023). Manual curation was primarily conducted using PretextView (Harry, 2022), with additional insights provided by JBrowse2 (Diesh *et al.*, 2023) and HiGlass (Kerpedjiev *et al.*, 2018). Scaffolds were visually inspected and corrected as described by Howe *et al.* (2021). Any identified contamination, missed joins, and mis-joins were corrected, and duplicate sequences were tagged and removed. The curation process is documented at https://gitlab.com/wtsi-grit/rapid-curation.

Summary analysis of the released assembly was performed using the ERGA-BGE Genome Report ASM Galaxy workflow (https://workflowhub.eu/workflows/1104?version=2).

Genome annotation methods

A gene set was generated using the Ensembl Gene Annotation system (Aken *et al.*, 2016), primarily by aligning publicly available short-read RNA-seq data from BioSample: SAMEA113595922 to the genome. Gaps in the annotation were filled via protein-to-genome alignments of a select set of clade-specific proteins from UniProt (The UniProt Consortium, 2019), which had experimental evidence at the protein or transcript level. At each locus, data were aggregated and consolidated, prioritising models derived from RNA-seq data, resulting in a final set of gene models and associated non-redundant transcript sets. To distinguish true isoforms from fragments, the likelihood of each open reading frame (ORF) was evaluated against known metazoan proteins. Low-quality transcript models, such as those showing evidence of fragmented ORFs, were removed. In cases where RNA-seq data were fragmented or absent, homology data were prioritised, favouring longer transcripts with strong intron support from short-read data. The resulting gene models were classified into two categories: protein-coding, and long non-coding. Models that did not overlap protein-coding genes, and were constructed from transcriptomic data were considered potential lncRNAs. Potential lncRNAs were further filtered to remove single-exon loci due to their unreliability. Putative miRNAs were predicted by performing a BLAST search of miRBase (Kozomara *et al.*, 2019) against the genome, followed by RNAfold analysis (Gruber *et al.*, 2008). Other small non-coding loci were identified by scanning the genome with Rfam (Kalvari *et al.*, 2018) and passing the results through Infernal (Nawrocki & Eddy, 2013). Summary analysis of the released annotation was carried out using the ERGA-BGE Genome Report ANNOT Galaxy workflow (https://doi.org/10.48546/workflowhub.workflow.1096.1).

Results

Genome assembly

The genome assembly has a total length of 310,741,076 bp in 13 scaffolds including the mitogenome (Figure 2 & Figure 3), with a GC content of 34.73%. The assembly has a contig N50 of 14,492,433 bp and L50 of 8 and a scaffold N50 of 31,130,618 bp and L50 of 5. 99.97% of the assembled sequences are localised to the 10 pseudochromosomal sequences. The assembly has a total of 37 gaps, totaling 7.4 kb in cumulative size. The single-copy gene content analysis using the Endopterygota_odb10 database with BUSCO (Manni *et al.*, 2021) resulted in 99.6% completeness (98.9% single and 0.7% duplicated). 57.01% of read k-mers were present in the assembly and the assembly has an average base accuracy Quality Value (QV) of 67.5 as calculated by Merqury (Rhie *et al.*, 2020).

Genome annotation

The genome annotation consists of 14,119 protein-coding genes with an associated 21,504 transcripts, in addition to 2,278 non-coding genes (Table 1). Using the longest isoform per transcript, the single-copy gene content analysis using the Endopterygota_odb10 database with BUSCO resulted in 96.7% completeness. Using the OMAmer Endopterygota database for OMArk (Nevers *et al.*, 2025) resulted in 93.95% completeness and 79.84% consistency (Table 2).

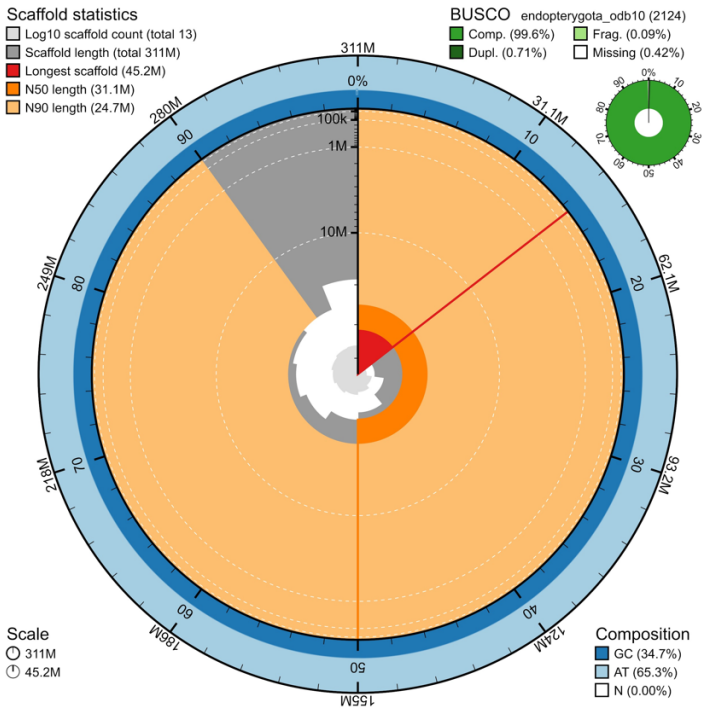

Figure 2. Snail plot summary of assembly statistics. The main plot is divided into 1,000 size-ordered bins around the circumference, with each bin representing 0.1% of the 310,741,076 bp assembly including the mitochondrial genome. The distribution of sequence lengths is shown in dark grey, with the plot radius scaled to the longest sequence present in the assembly (45,241,133 bp, shown in red). Orange and pale-orange arcs show the scaffold N50 and N90 sequence lengths (31,130,618 and 24,717,953 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 endopterygota database (odb10) is shown in the top right.

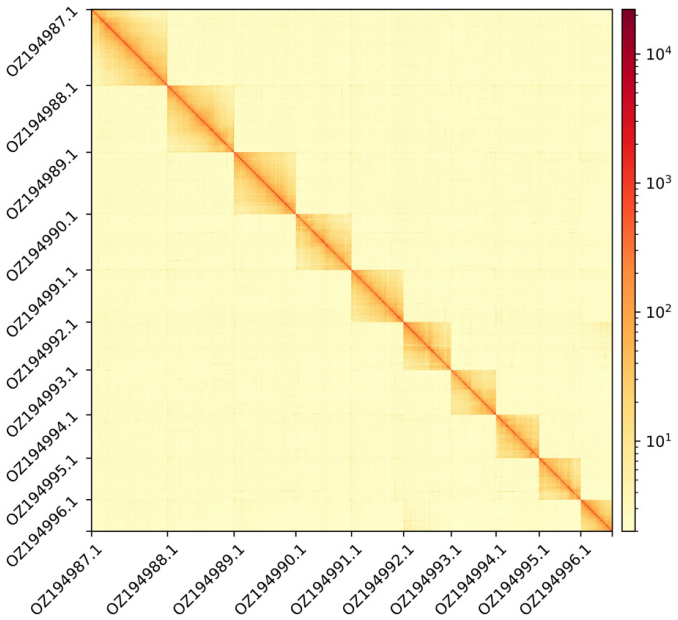

Figure 3. Hi-C contact map showing spatial interactions between regions of the genome. The diagonal corresponds to intra-chromosomal contacts, depicting chromosome boundaries. The frequency of contacts is shown on a logarithmic heatmap scale. Hi-C matrix bins were merged into a 100 kb bin size for plotting.

Table 1. Statistics from assembled gene models.

	No. genes	No. transcripts	Mean gene length (bp)	No. single-exon genes	Mean exons per transcript
mRNA	14,119	21,504	10,688	380	5.7
snoRNA	20	20	120	20	1
lncRNA	1,950	2,123	6,255	3	2.2
snRNA	31	31	142	31	1
rRNA	39	39	1,333	39	1
scRNA	1	1	130	1	1
tRNA	235	235	74	235	1

Table 2. Annotation completeness and consistency scores calculated by BUSCO run in protein mode (endopterygota_odb10) and OMArk (Endopterygota).

	Complete	Singular	Duplicated	Fragmented	Missing
BUSCO	2,054 (96.7%)	2,036 (95.9%)	18 (0.8%)	8 (0.4%)	62 (2.9%)
OMArk	4,475 (93.95%)	4,258 (89.40%)	217 (4.56%)	-	288 (6.05%)
	Consistent	**Inconsistent**	**Contaminants**	**Unknown**	
OMArk	11,272 (79.84%)	484 (3.43%)	0 (0.00%)	2,363 (16.74%)	

Data availability

Tarphius canariensis and the related genomic study were assigned to Tree of Life ID (ToLID) icTarCana1 and all sample, sequence, and assembly information are available in INSDC databases under the umbrella BioProject PRJEB76560. The sample information is available at the following BioSample accessions: SAMEA113595920, SAMEA113595921, SAMEA113595922, and SAMEA113595923. The genome assembly is accessible from INSDC databases under accession number GCA_964277315.1 and the annotated genome is available through the Ensembl ERGA-BGE project page (https://projects.ensembl.org/erga-bge/) and and Ensembl website (https://beta.ensembl.org). Sequencing data produced as part of this project are available at the following accessions: ERX13553711, ERX13020688, ERX12671954, and ERX12671953. 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/Tarphius_canariensis/icTarCana1. Further details and data about the project are hosted on the ERGA portal at https://portal.erga-biodiversity.eu/data_portal/12480

Author contributions

BCE and VN collected the species, VN identified the species, BCE and VN sampled and preserved biological material and provided metadata, RM, AB, JPG and RF provided sampling and metadata support and management, MC and AR carried out barcoding and deposited specimens for vouchering biobanking, the WSI ToL Samples and Laboratory teams and WSI Scientific Operations teams extracted DNA, prepared libraries, and performed sequencing, the WSI ToL Informatics teams performed genome assembly and curation, LH performed genome annotation under the supervision of FM. TB 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.

Author information

Members of the Wellcome Sanger Institute Tree of Life Management, Samples and Laboratory team are listed here: https://doi.org/10.5281/zenodo.13960213.

Members of Wellcome Sanger Institute Scientific Operations: Sequencing Operations are listed here: https://doi.org/10.5281/zenodo.14870789.

Members of the Wellcome Sanger Institute Tree of Life Core Informatics team are listed here: https://doi.org/10.5281/zenodo.15440189.

Acknowledgements

We would like to acknowledge the assembly reviewer, Jean-Marc Aury, from Genoscope.

References

Aken BL, Ayling S, Barrell D, *et al.*: **The ensembl gene annotation system.** *Database (Oxford).* 2016; **2016**: baw093.
PubMed Abstract | Publisher Full Text | Free Full Text

Allio R, Schomaker-Bastos A, Romiguier J, *et al.*: **MitoFinder: efficient automated large-scale extraction of mitogenomic data in target enrichment phylogenomics.** *Mol Ecol Resour.* 2020; **20**(4): 892–905.
PubMed Abstract | Publisher Full Text | Free Full Text

Bates A, Clayton-Lucey I, Howard C: **Sanger Tree of Life HMW DNA fragmentation: diagenode Megaruptor®3 for LI PacBio.** *protocols.io.* 2023.
Publisher Full Text

Challis R, Kumar S, Sotero-Caio C: **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

Cheng H, Concepcion GT, Feng X, *et al.*: **Haplotype-resolved *de novo* assembly using phased assembly graphs with hifiasm.** *Nat Methods.* 2021; **18**(2): 170–175.
PubMed Abstract | Publisher Full Text | Free Full Text

Denton A, Oatley G, Cornwell C, *et al.*: **Sanger Tree of Life sample homogenisation: PowerMash.** *protocols.io.* 2023a.
Publisher Full Text

Denton A, Yatsenko H, Jay J, *et al.*: **Sanger Tree of Life wet laboratory protocol collection V.1.** *protocols.io.* 2023b.
Publisher Full Text

Diesh C, Stevens GJ, Xie P, *et al.*: **JBrowse 2: a modular genome browser with views of synteny and structural variation.** *Genome Biol.* 2023; **24**(1): 74.
PubMed Abstract | Publisher Full Text | Free Full Text

do Amaral RJV, Bates A, Denton A, *et al.*: **Sanger Tree of Life RNA extraction: automated MagMaxTM mirVana.** *protocols.io.* 2023.
Publisher Full Text

Emerson BC, Oromí P, Hewitt GM: **Tracking colonization and diversification of insect lineages on islands: mitochondrial DNA phylogeography of *Tarphius canariensis* (Coleoptera: Colydiidae) on the Canary Islands.** *Proc Biol Sci.* 2000; **267**(1458): 2199–2205.
PubMed Abstract | Publisher Full Text | Free Full Text

Emerson BC, Oromí P: **Diversification of the forest beetle genus *Tarphius* on the Canary Islands, and the evolutionary origins of island endemics.** *Evolution.* 2005; **59**(3): 586–598.
PubMed Abstract | Publisher Full Text

Formenti G, Abueg L, Brajuka A, *et al.*: **Gfastats: conversion, evaluation and manipulation of genome sequences using assembly graphs.** *Bioinformatics.* 2022; **38**(17): 4214–4216.
PubMed Abstract | Publisher Full Text | Free Full Text

Gruber AR, Lorenz R, Bernhart SH, *et al.*: **The Vienna RNA Websuite.** *Nucleic Acids Res.* 2008; **36**(suppl_ 2): W70–W74.
PubMed Abstract | Publisher Full Text | Free Full Text

Harry E: **PretextView (Paired REad TEXTure Viewer): a desktop application for viewing pretext contact maps**. 2022.
Reference Source

Howe K, Chow W, Collins J, *et al.*: **Significantly improving the quality of genome assemblies through curation.** *GigaScience.* 2021; **10**(1): giaa153.
PubMed Abstract | Publisher Full Text | Free Full Text

Jay J, Yatsenko H, Narváez-Gómez JP, *et al.*: **Sanger Tree of Life sample preparation: triage and dissection.** *protocols.io.* 2023.
Publisher Full Text

Kalvari I, Nawrocki EP, Argasinska J, *et al.*: **Non-coding RNA analysis using the Rfam database.** *Curr Protoc Bioinformatics.* 2018; **62**(1): e51.
PubMed Abstract | Publisher Full Text | Free Full Text

Kerpedjiev P, Abdennur N, Lekschas F, *et al.*: **HiGlass: web-based visual exploration and analysis of genome interaction maps.** *Genome Biol.* 2018; **19**(1): 125.
PubMed Abstract | Publisher Full Text | Free Full Text

Kozomara A, Birgaoanu M, Griffiths-Jones S: **miRBase: from microRNA sequences to function.** *Nucleic Acids Res.* 2019; **47**(D1): D155–D162.
PubMed Abstract | Publisher Full Text | Free Full Text

Manni M, Berkeley MR, Seppey M, *et al.*: **BUSCO update: novel and streamlined workflows along with broader and deeper phylogenetic coverage for scoring of eukaryotic, prokaryotic, and viral genomes.** *Mol Biol Evol.* 2021; **38**(10): 4647–4654.
PubMed Abstract | Publisher Full Text | Free 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

Nawrocki EP, Eddy SR: **Infernal 1.1: 100-fold faster RNA homology searches.** *Bioinformatics.* 2013; **29**(22): 2935–5.
PubMed Abstract | Publisher Full Text | Free Full Text

Nevers Y, Warwick Vesztrocy A, Rossier V, *et al.*: **Quality assessment of gene repertoire annotations with OMArk.** *Nat Biotechnol.* 2025; **43**(1): 124–133.
PubMed Abstract | Publisher Full Text | Free Full Text

Noguerales V, Emerson BC: **Arthropod mtDNA paraphyly: a case study of introgressive origin.** *J Evol Biol.* 2025; **38**(2): 272–283.
PubMed Abstract | Publisher Full Text

Oatley G, Sampaio F, Howard C: **Sanger Tree of Life fragmented DNA clean up: automated SPRI.** *protocols.io.* 2023.
Publisher Full Text

Patiño J, Whittaker RJ, Borges PAV, *et al.*: **A roadmap for island biology: 50 fundamental questions after 50 years of *the theory of Island biogeography*.** *J Biogeogr.* 2017; **44**(5): 963–983.
Publisher Full Text

Pointon DL, Eagles W, Sims Y, *et al.*: **sanger-tol/treeval v1.0.0 – Ancient Atlantis.** 2023.
Publisher Full Text

Rao SSP, Huntley MH, Durand NC, *et al.*: **A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping.** *Cell.* 2014; **159**(7): 1665–1680.
PubMed Abstract | Publisher Full Text | Free Full Text

Rhie A, Walenz BP, Koren S, *et al.*: **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

Strickland M, Moll R, Cornwell C, *et al.*: **Sanger Tree of Life HMW DNA extraction: manual MagAttract.** *protocols.io.* 2023.
Publisher Full Text

The UniProt Consortium: **UniProt: a worldwide hub of protein knowledge.** *Nucleic Acids Res.* 2019; **47**(1): D506–D515.
PubMed Abstract | Publisher Full Text | Free Full Text

Uliano-Silva M, Ferreira JGRN, Krasheninnikova K, *et al.*: **MitoHiFi: a python pipeline for mitochondrial genome assembly from PacBio high fidelity reads.** *BMC Bioinformatics.* 2023; **24**(1): 288.
PubMed Abstract | Publisher Full Text | Free Full Text

Vasimuddin M, Misra S, Li H, *et al.*: **Efficient architecture-aware acceleration of BWA-MEM for multicore systems.** In: *2019 IEEE International Parallel and Distributed Processing Symposium (IPDPS).* IEEE, 2019; 314–324.
Publisher Full Text

Zhou C, McCarthy SA, Durbin R: **YaHS: Yet another Hi-C Scaffolding tool.** *Bioinformatics.* 2023; **39**(1): btac808.
PubMed Abstract | Publisher Full Text | Free Full Text

Open Peer Review

Current Peer Review Status: ✔ ?

Version 1

Reviewer Report 15 September 2025

https://doi.org/10.21956/openreseurope.22572.r58533

© **2025 Peona V.** 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.

?

Valentina Peona
[1] Uppsala University, uppsala, Sweden
[2] Naturhistoriska riksmuseet (Ringgold ID: 42763), Stockholm, Stockholm County, Sweden

Emerson et al present the genome assembly of the beetle Tarphius canariensis. The manuscript reads well and the analysis and results are clear.

I have just one minor point to raise.

In the Methods section, under Genetic information, the authors state that the estimated genome size for this organism is 450 Mb. However, when looking at the assembly report on Github (https://github.com/ERGA-consortium/EARs/blob/main/Assembly_Reports/Tarphius_canariensis/icTarCana1/icTarCana1_EAR.pdf), another expected genome size is given (310 Mb which is perfect accordance with the assembly size).
Please clarify how the expected sizes of 310 Mb has been estimated (e.g., kmer-based, etc)

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?
Yes

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

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Repetitive elements, centromere evolution, genome assembly

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 08 September 2025

https://doi.org/10.21956/openreseurope.22572.r57831

© 2025 Escalona 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.

Merly Escalona
University of Connecticut Department of Molecular and Cell Biology, Storrs, Connecticut, USA

The manuscript reports the generation of genomic resources for the ironclad beetle (Tarphius canariensis) endemic of the Canary islands. The resources include: a new chromosome-level haploid genome assembly, a mitochondrial genome and the corresponding genome annotation for both. All resources generated satisfy the minimum metrics set as benchmark for the Vertebrate Genome Project (Rhie et al. 2021) and the Earth BioGenome Project (https://www.earthbiogenome.org/report-on-assembly-standards).

It is a good manuscript in general terms, and aligns with the requirements of the Journal to which it has been submitted. However, I think there are some aspects that should be examined. My notes are going to mix minor and major comments and are ordered more or less according to appearance in the manuscript.

Comments:

While both the nuclear and mitochondrial genomes have been assembled, annotated, and made publicly available, there are no metrics for quality description of the mitochondrial genome, and it would be good to share those as well.

The information for the karyotype of this species was retrieved from the Genome on a Tree database/website and it is based on ancestral data. Upon inspection, the available ancestral species based of theis karyotype information is Zopherus nodulosus. The karyiotype fo Z. nodulosus is n=9, which doesn't match the information described in the manuscript. I'd recommend to be more specific with the information that's being used to actually estimate genome sizes and karyotypes.

Please be consistent with the format of the company names. If you say: Pacific Biosciences, California, USA; then also say: Thermo Fisher Scientific, Massachusets, USA. Same for the rest of the company names.

Please be consistent with the format of the unots for base pairs: "8 kb" vs. "8kb".

Please be consistent with the format when mentioning tools: tool version (reference). There's a full format for tools likes Hifiasm or YaHS, but not for bwa-mem2; Gfastats; BUSCO...

There's information that's available in the genome report (https://github.com/ERGA-consortium/EARs/tree/main/Assembly_Reports/Tarphius_canariensis/) that should be shown/described in the main manuscript. For instance, description about the contamination screen, might be helpful for further processing of beetles like this one, are the species found during the screen expected for the species or sampling process?

There's some information about the karyotype estimation in the repository that should be in the main manuscript.

There are some discrepancies between the report (Github) and the data in the manuscript, when referring to the k-mer completeness. Is this a typo? Is the discrepancy explain by the contamination of the species?

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

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

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

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

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: genome assembly, 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.