DATA NOTE

The genome sequence of the Catalonian Wall Lizard, *Podarcis liolepis* (Boulenger, 1905) (Squamata: Lacertidae)

[version 1; peer review: 2 approved]

Nathalie Feiner[1,2], Tobias Uller[2], Javier Abalos[2,3], Joana Meier[4],
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,
Tree of Life Core Informatics collective

[1]Max Planck Institute for Evolutionary Biology, Plön, Germany
[2]Department of Biology, Lund University, Lund, Sweden
[3]Ethology Lab, Cavanilles Institute of Biodiversity and Evolutionary Biology,, University of Valencia, Valencia, Spain
[4]Tree of Life, Wellcome Sanger Institute, Hinxton, England, UK

v1 **First published:** 10 Sep 2025, **10**:505
https://doi.org/10.12688/wellcomeopenres.24854.1
Latest published: 10 Sep 2025, **10**:505
https://doi.org/10.12688/wellcomeopenres.24854.1

Abstract

We present a genome assembly from an individual female *Podarcis liolepis* (Catalonian Wall Lizard; Chordata; Lepidosauria; Squamata; Lacertidae). The assembly contains two haplotypes with total lengths of 1 574.00 megabases and 1 478.08 megabases. Most of haplotype 1 (97.05%) is scaffolded into 20 chromosomal pseudomolecules, including the W and Z sex chromosomes. Haplotype 2 was assembled to scaffold level. The mitochondrial genome has also been assembled, with a length of 17.32 kilobases.

Keywords

Podarcis liolepis; Catalonian Wall Lizard; genome sequence; chromosomal; Squamata

This article is included in the Tree of Life gateway.

Open Peer Review

Approval Status ✓ ✓

	1	2
version 1 10 Sep 2025	✓ view	✓ view

1. **Hong-Xin Xie**, University of Glasgow, Glasgow, UK

2. **Kathryn R. Elmer**, University of Glasgow, Glasgow, UK

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

Corresponding authors: Nathalie Feiner (feiner@evolbio.mpg.de), Joana Meier (jm66@sanger.ac.uk)

Author roles: Feiner N: Investigation, Resources, Writing – Original Draft Preparation, Writing – Review & Editing; **Uller T**: Investigation, Resources, Writing – Original Draft Preparation; **Abalos J**: Resources, Writing – Review & Editing; **Meier J**: Investigation, Resources, Writing – Review & Editing;

Competing interests: No competing interests were disclosed.

Grant information: This work was supported by Wellcome through core funding to the Wellcome Sanger Institute (220540). N.F. was funded by Starting Grants from the European Research Council (948126) and the Swedish Research Council (2020-03650). J.A. was funded by a Margarita Salas contract from the Spanish Ministry of Science (MS21-053) and an ASAB research grant.
The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

How to cite this article: Feiner N, Uller T, Abalos J *et al.* **The genome sequence of the Catalonian Wall Lizard, *Podarcis liolepis* (Boulenger, 1905) (Squamata: Lacertidae) [version 1; peer review: 2 approved]** Wellcome Open Research 2025, **10**:505 https://doi.org/10.12688/wellcomeopenres.24854.1

First published: 10 Sep 2025, **10**:505 https://doi.org/10.12688/wellcomeopenres.24854.1

Species taxonomy

Eukaryota; Opisthokonta; Metazoa; Eumetazoa; Bilateria; Deuterostomia; Chordata; Craniata; Vertebrata; Gnathostomata; Teleostomi; Euteleostomi; Sarcopterygii; Dipnotetrapodomorpha; Tetrapoda; Amniota; Sauropsida; Sauria; Lepidosauria; Squamata; Bifurcata; Unidentata; Episquamata; Laterata; Lacertibaenia; Lacertidae; Lacertinae; *Podarcis*; *Podarcis liolepis* (Boulenger, 1905) (NCBI:txid662736)

Background

The Catalan wall lizard *Podarcis liolepis* (Boulenger 1905; Figure 1). is a diurnal lacertid lizard native to northeast Spain and southern France. It is classified as Least Concern (LC) by the IUCN (Bowles, 2024).

This species is one of 28 currently described species of the genus *Podarcis*. *Podarcis liolepis* is one of eight species in a monophyletic clade of Iberian and North African wall lizards (Kaliontzopoulou *et al.*, 2011; Yang *et al.*, 2021). The distribution of these species is complex and with several hybrid zones (Caeiro-Dias *et al.*, 2018; Caeiro-Dias *et al.*, 2021; Gaczorek *et al.*, 2023; Pinho *et al.*, 2008). *Podarcis liolepis* is also found on several islands in the Columbretes archipelago east of the Iberian Peninsula (Böhme, 1986).

The Catalan wall lizard inhabits a range of different habitats, from sea level to >1700 m altitude. The species can be common in urban environments. Lizards can be active all year round, except at high altitude, but breeding is restricted to spring and early summer. The species is oviparous and Meiri *et al.* (2020) report an average clutch size of 2.80.

Figure 1. A) Male *Podarcis liolepis* photographed in the Botánic Garden of the University of Valencia. **B**) Female *P. liolepis* photographed in Godella, Spain. Note that the specimen selected for genome sequencing is from the same locality as specimen shown in panel **B**, but not shown on the photographs. Photographs by Javier Abalos.

The Catalan wall lizard is a relatively small-bodied wall lizard. In continental populations, males range 44.9–65.7 mm in snout-to-vent length (SVL) whereas females range 43.3–61.3 mm in SVL (Carazo *et al.*, 2011; Carretero & Salvador, 2016). Lizards from the Columbretes archipelago are larger with an average SVL of 65.3–68 in males and 61.7-61.9 in females (Castilla & Bauwens, 1991a; Castilla & Bauwens, 1991b). The dorsum of this species is mostly showing a reticulated pattern with brown or green hues on a darker background. As in most species of *Podarcis*, the sexes differ in appearance with females being generally smaller and showing a pattern that is more dominated by longitudinal stripes. This species is one of several *Podarcis* that exhibit ventral color polymorphism with yellow, orange/red and white morphs (Andrade *et al.*, 2019).

P. liolepis is one of three species in the Ibero-Maghrebian group of *Podarcis* for which a reference genome is constructed – the other two being *P. bocagei* (Feiner *et al.*, 2025b) and *P. vaucheri* (Feiner *et al.*, 2025a). This will be a valuable resource for studies of this species and of the Iberian and North African *Podarcis* more generally.

We present a chromosome-level genome sequence for *Podarcis liolepis*, the Catalonian Wall Lizard. The assembly was produced using the Tree of Life pipeline from a specimen collected in Godella, Valencia, Spain This assembly was generated as part of the Darwin Tree of Life Project, which aims to generate high-quality reference genomes for all named eukaryotic species in Britain and Ireland to support research, conservation, and the sustainable use of biodiversity (Blaxter *et al.*, 2022).

Methods

Sample acquisition

The specimen, an adult female *P. liolepis* lizard (specimen ID SAN25001762, ToLID rPodLio1; Figure 1), was collected in April 2022 from a site near Godella in Spain (latitude: 39.5258438; longitude: –0.4228414). The specimen was caught by noosing, standard morphometric measurements were taken and the tip of the tail (ca. 2 cm) was collected and preserved in ethanol. The specimen was released again at the site of capture. Field work was conducted under the permit ID 2022 (FAU22_020).

Nucleic acid extraction

Protocols for high molecular weight (HMW) DNA extraction developed at the Wellcome Sanger Institute (WSI) Tree of Life Core Laboratory are available on protocols.io (Howard *et al.*, 2025). The rPodLio1 sample was weighed and triaged to determine the appropriate extraction protocol. Tissue from the terminal body was homogenised by powermashing using a PowerMasher II tissue disruptor. HMW DNA was extracted using the Automated MagAttract v2 protocol. DNA was sheared into an average fragment size of 12–20 kb following the Megaruptor®3 for LI PacBio protocol. Sheared DNA was purified by automated SPRI (solid-phase reversible immobilisation). 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. HMW DNA was extracted in the WSI Scientific Operations core using the Automated MagAttract v2 protocol. DNA was sheared into an average fragment size of 12–20 kb following the Megaruptor®3 for LI PacBio protocol. Sheared DNA was purified by manual SPRI (solid-phase reversible immobilisation). 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. For this sample, the final post-shearing DNA had a Qubit concentration of 13.9 ng/μL and a yield of 1 807.00 ng.

PacBio HiFi library preparation and sequencing

Library preparation and sequencing were performed at the WSI Scientific Operations core. Libraries were prepared using the SMRTbell Prep Kit 3.0 (Pacific Biosciences, California, USA), following the manufacturer's instructions. The kit includes reagents for end repair/A-tailing, adapter ligation, post-ligation SMRTbell bead clean-up, and nuclease treatment. Size selection and clean-up were performed using diluted AMPure PB beads (Pacific Biosciences). DNA concentration was quantified using a Qubit Fluorometer v4.0 (ThermoFisher Scientific) and the Qubit 1X dsDNA HS assay kit. Final library fragment size was assessed with the Agilent Femto Pulse Automated Pulsed Field CE Instrument (Agilent Technologies) using the gDNA 55 kb BAC analysis kit.

The sample was sequenced on a Revio instrument (Pacific Biosciences). The prepared library was normalised to 2 nM, and 15 μL was used for making complexes. Primers were annealed and polymerases bound to generate circularised complexes, following the manufacturer's instructions. Complexes were purified using 1.2X SMRTbell beads, then diluted to the Revio loading concentration (200–300 pM) and spiked with a Revio sequencing internal control. The sample was sequenced on a Revio 25M SMRT cell. The SMRT Link software (Pacific Biosciences), a web-based workflow manager, was used to configure and monitor the run and to carry out primary and secondary data analysis.

Hi-C

Sample preparation and crosslinking

The Hi-C sample was prepared from 20–50 mg of frozen tissue from the terminal body of the rPodLio1 sample using the Arima-HiC v2 kit (Arima Genomics). Following the manufacturer's instructions, tissue was fixed and DNA crosslinked using TC buffer to a final formaldehyde concentration of 2%. The tissue was homogenised using the Diagnocine Power Masher-II. Crosslinked DNA was digested with a restriction enzyme master mix, biotinylated, and ligated. Clean-up was performed with SPRISelect beads before library preparation. DNA concentration was measured with the Qubit Fluorometer (Thermo Fisher Scientific) and Qubit HS Assay Kit. The biotinylation percentage was estimated using the Arima-HiC v2 QC beads.

Hi-C library preparation and sequencing

Biotinylated DNA constructs were fragmented using a Covaris E220 sonicator and size selected to 400–600 bp using SPRISelect beads. DNA was enriched with Arima-HiC v2 kit Enrichment beads. End repair, A-tailing, and adapter ligation were carried out with the NEBNext Ultra II DNA Library Prep Kit (New England Biolabs), following a modified protocol where library preparation occurs while DNA remains bound to the Enrichment beads. Library amplification was performed using KAPA HiFi HotStart mix and a custom Unique Dual Index (UDI) barcode set (Integrated DNA Technologies). Depending on sample concentration and biotinylation percentage determined at the crosslinking stage, libraries were amplified with 10 to 16 PCR cycles. Post-PCR clean-up was performed with SPRISelect beads. Libraries were quantified using the AccuClear Ultra High Sensitivity dsDNA Standards Assay Kit (Biotium) and a FLUOstar Omega plate reader (BMG Labtech).

Prior to sequencing, libraries were normalised to 10 ng/μL. Normalised libraries were quantified again and equimolar and/or weighted 2.8 nM pools. Pool concentrations were checked using the Agilent 4200 TapeStation (Agilent) with High Sensitivity D500 reagents before sequencing. Sequencing was performed using paired-end 150 bp reads on the Illumina NovaSeq X.

Genome assembly

Prior to assembly of the PacBio HiFi reads, a database of *k*-mer counts ($k = 31$) was generated from the filtered reads using FastK. GenomeScope2 (Ranallo-Benavidez *et al.*, 2020) was used to analyse the *k*-mer frequency distributions, providing estimates of genome size, heterozygosity, and repeat content.

The HiFi reads were assembled using Hifiasm in Hi-C phasing mode (Cheng *et al.*, 2021; Cheng *et al.*, 2022), producing two haplotypes. Hi-C reads (Rao *et al.*, 2014) were mapped to the primary contigs using bwa-mem2 (Vasimuddin *et al.*, 2019). Contigs were further scaffolded with Hi-C data in YaHS (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 (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.

Assembly curation

The assembly was decontaminated using the Assembly Screen for Cobionts and Contaminants (ASCC) pipeline. TreeVal was used to generate the flat files and maps for use in curation. Manual curation was conducted primarily in PretextView and HiGlass (Kerpedjiev *et al.*, 2018). Scaffolds were visually inspected and corrected as described by Howe *et al.* (2021). Manual corrections included 254 breaks and 283 joins. The curation process is documented at https://gitlab.com/wtsi-grit/rapid-curation.

PretextSnapshot was used to generate a Hi-C contact map of the final assembly.

Assembly quality assessment

The Merqury.FK tool (Rhie *et al.*, 2020) was run in a Singularity container (Kurtzer *et al.*, 2017) to evaluate *k*-mer completeness and assembly quality for both haplotypes using the *k*-mer databases (k = 31) computed prior to genome assembly. The analysis outputs included assembly QV scores and completeness statistics.

The genome was analysed using the BlobToolKit pipeline, a Nextflow implementation of the earlier Snakemake version (Challis *et al.*, 2020). The pipeline aligns PacBio reads using minimap2 (Li, 2018) and SAMtools (Danecek *et al.*, 2021) to generate coverage tracks. It runs BUSCO (Manni *et al.*, 2021) using lineages identified from the NCBI Taxonomy (Schoch *et al.*, 2020). For the three domain-level lineages, BUSCO genes are aligned to the UniProt Reference Proteomes database (Bateman *et al.*, 2023) using DIAMOND blastp (Buchfink *et al.*, 2021). The genome is divided into chunks based on the density of BUSCO genes from the closest taxonomic lineage, and each chunk is aligned to the UniProt Reference Proteomes database with DIAMOND blastx. Sequences without hits are chunked using seqtk and aligned to the NT database with blastn (Altschul *et al.*, 1990). The BlobToolKit suite consolidates all outputs into a blobdir for visualisation. The BlobToolKit pipeline was developed using nf-core tooling (Ewels *et al.*, 2020) and MultiQC (Ewels *et al.*, 2016), with containerisation through Docker (Merkel, 2014) and Singularity (Kurtzer *et al.*, 2017).

Genome sequence report

Sequence data

PacBio sequencing of the *Podarcis liolepis* specimen generated 55.32 Gb (gigabases) from 8.06 million reads, which were used to assemble the genome. GenomeScope2.0 analysis estimated the haploid genome size at 1 456.62 Mb, with a heterozygosity of 1.52% and repeat content of 24.01% (Figure 2). These estimates guided expectations for the assembly. Based on the estimated genome size, the sequencing data provided approximately 36× coverage. Hi-C sequencing produced 238.19 Gb from 1 577.40 million reads, which were used to scaffold the assembly. Table 1 summarises the specimen and sequencing details.

Assembly statistics

The genome was assembled into two haplotypes using Hi-C phasing. Haplotype 1 was curated to chromosome level, while haplotype 2 was assembled to scaffold level. The final assembly has a total length of 1 574.00 Mb in 289 scaffolds, with 921 gaps, and a scaffold N50 of 95.42 Mb (Table 2).

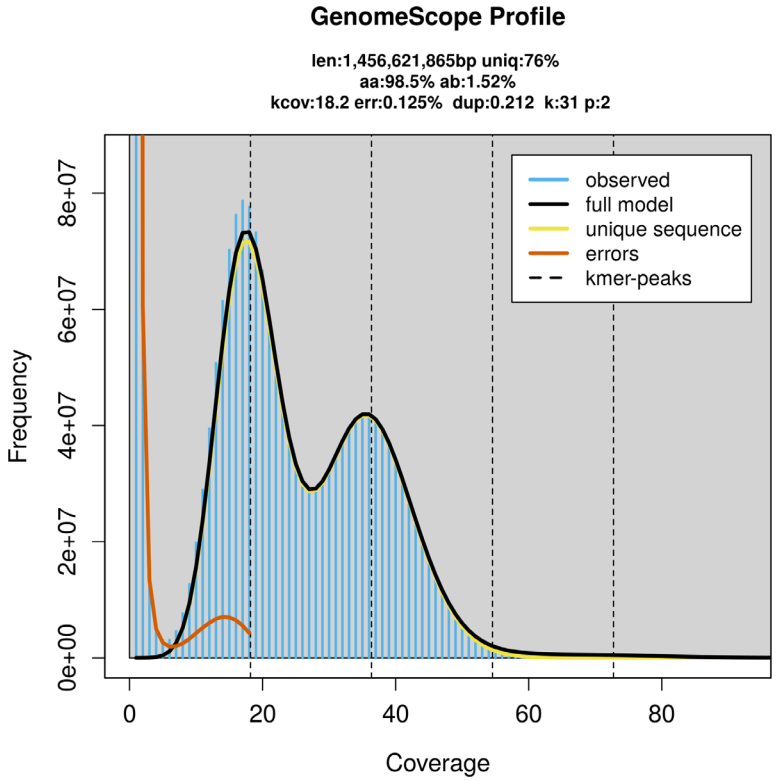

Figure 2. Frequency distribution of *k*-mers generated using GenomeScope2. The plot shows observed and modelled *k*-mer spectra, providing estimates of genome size, heterozygosity, and repeat content based on unassembled sequencing reads.

Most of the assembly sequence (97.05%) was assigned to 20 chromosomal-level scaffolds, representing 18 autosomes and the W and Z sex chromosomes. These chromosome-level scaffolds, confirmed by Hi-C data, are named according to size (Figure 3; Table 3).

The mitochondrial genome was also assembled. This sequence is included as a contig in the multi-fasta file of the genome submission and as a standalone record.

For haplotype 1, the estimated QV is 64.5, and for haplotype 2, 65.1. When the two haplotypes are combined, the assembly achieves an estimated QV of 64.8. The *k*-mer completeness is 73.88% for haplotype 1, 70.62% for haplotype 2, and 99.69% for the combined haplotypes (Figure 4).

BUSCO analysis using the sauropsida_odb10 reference set (n = 7 480) identified 94.8% of the expected gene set (single = 92.6%, duplicated = 2.2%) for haplotype 1. The snail plot in Figure 5 summarises the scaffold length distribution and other assembly statistics for haplotype 1. The blob plot in Figure 6 shows the distribution of scaffolds by GC proportion and coverage for haplotype 1.

Table 4 lists the assembly metric benchmarks adapted from Rhie *et al.* (2021) the Earth BioGenome Project Report on Assembly Standards September 2024. The EBP metric, calculated for the haplotype 1, is **6.C.Q64**, meeting the recommended reference standard.

Wellcome Sanger Institute – Legal and Governance

The materials that have contributed to this genome note have been supplied by a Tree of Life collaborator. The

Table 1. Specimen and sequencing data for BioProject PRJEB73727.

Platform	PacBio HiFi	Hi-C
ToLID	rPodLio1	rPodLio1
Specimen ID	SAN25001762	SAN25001762
BioSample (source individual)	SAMEA114217798	SAMEA114217798
BioSample (tissue)	SAMEA114217802	SAMEA114217802
Tissue	terminal body	terminal body
Instrument	Revio	Illumina NovaSeq X
Run accessions	ERR12736922	ERR12743676
Read count total	8.06 million	1 577.40 million
Base count total	55.32 Gb	238.19 Gb

Table 2. Genome assembly statistics.

Metric	Haplotype 1	Haplotype 2
Assembly name	rPodLio1.hap1.1	rPodLio1.hap2.1
Assembly accession	GCA_965112155.1	GCA_965178375.1
Assembly level	chromosome	scaffold
Span (Mb)	1 574.00	1 478.08
Number of chromosomes	20	N/A
Number of contigs	1 210	1 011
Contig N50	2.88 Mb	2.89 Mb
Number of scaffolds	289	144
Scaffold N50	95.42 Mb	94.04 Mb
Longest scaffold length (Mb)	139.44	N/A
Sex chromosomes	W and Z	N/A
Organelles	Mitochondrion: 17.32 kb	N/A

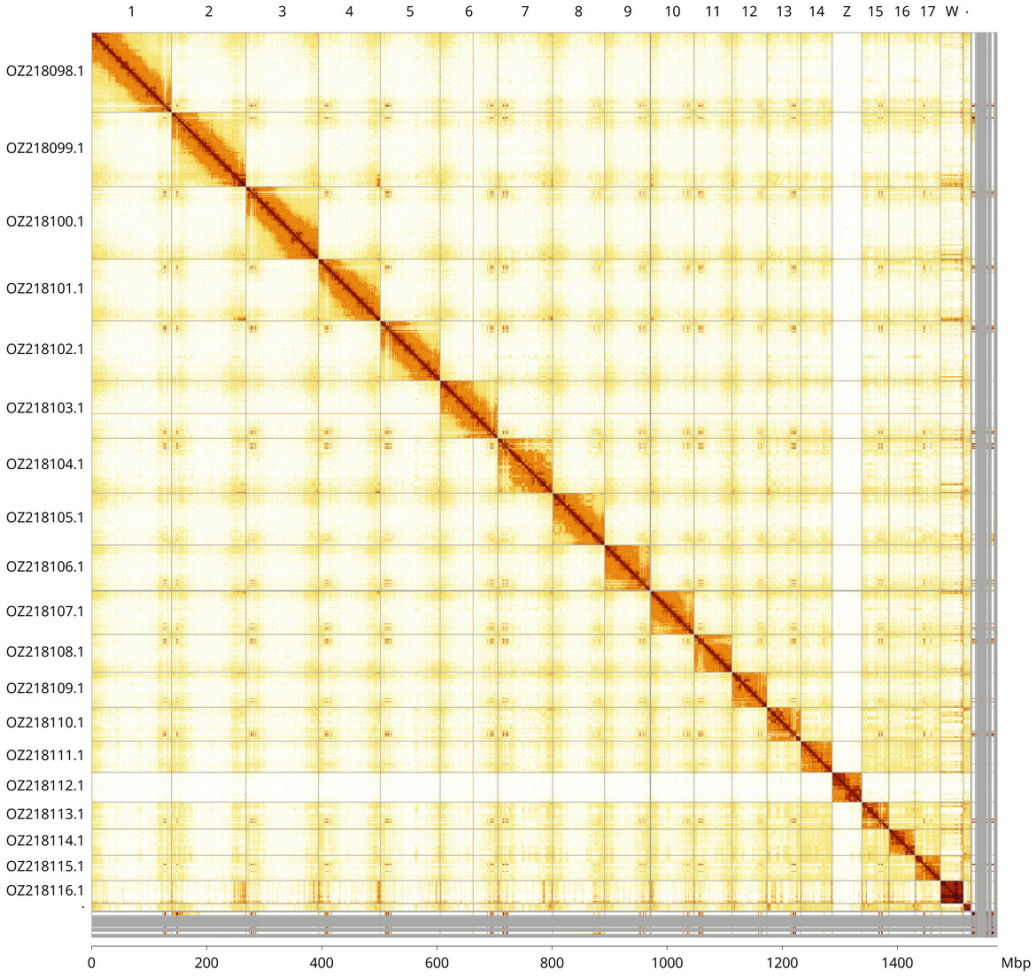

Figure 3. Hi-C contact map of the *Podarcis liolepis* genome assembly. Assembled chromosomes are shown in order of size and labelled along the axes. The plot was generated using PretextSnapshot.

Table 3. Chromosomal pseudomolecules in the haplotype 1 genome assembly of *Podarcis liolepis* rPodLio1.

INSDC accession	Molecule	Length (Mb)	GC%
OZ218098.1	1	139.44	43.50
OZ218099.1	2	129.02	44
OZ218100.1	3	125.94	43
OZ218101.1	4	107.33	43
OZ218102.1	5	103.94	43.50
OZ218103.1	6	100.11	43.50
OZ218104.1	7	95.42	45.50
OZ218105.1	8	90.28	43
OZ218106.1	9	80.31	43.50
OZ218107.1	10	75.16	43.50
OZ218108.1	11	65.82	43.50
OZ218109.1	12	60.73	44
OZ218110.1	13	58.94	45.50
OZ218111.1	14	54.24	45.50
OZ218113.1	15	46.99	46.50
OZ218114.1	16	45.33	46
OZ218115.1	17	44.08	44.50
OZ218117.1	18	12.69	49
OZ218116.1	W	40.11	46
OZ218112.1	Z	51.66	44.50

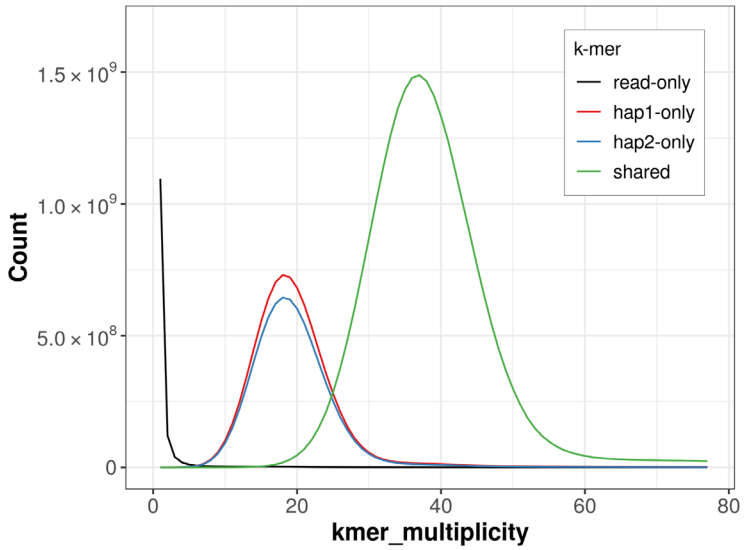

Figure 4. Evaluation of *k*-mer completeness using MerquryFK. This plot illustrates the recovery of *k*-mers from the original read data in the final assemblies. The horizontal axis represents *k*-mer multiplicity, and the vertical axis shows the number of *k*-mers. The black curve represents *k*-mers that appear in the reads but are not assembled. The green curve corresponds to *k*-mers shared by both haplotypes, and the red and blue curves show *k*-mers found only in one of the haplotypes.

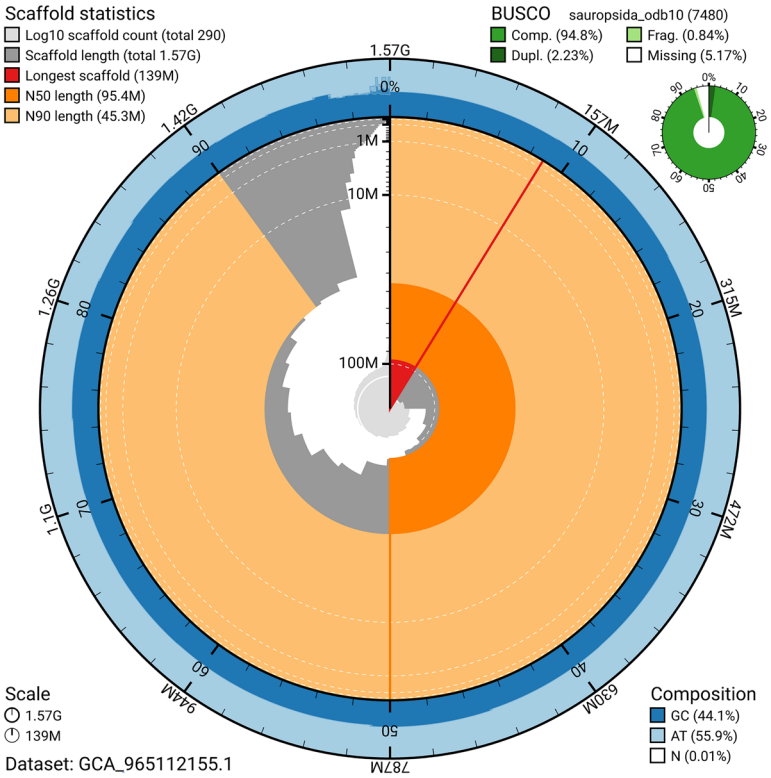

Figure 5. Assembly metrics for rPodLio1.hap1.1. The BlobToolKit snail plot provides an overview of assembly metrics and BUSCO gene completeness. The circumference represents the length of the whole genome sequence, and the main plot is divided into 1 000 bins around the circumference. The outermost blue tracks display the distribution of GC, AT, and N percentages across the bins. Scaffolds are arranged clockwise from longest to shortest and are depicted in dark grey. The longest scaffold is indicated by the red arc, and the deeper orange and pale orange arcs represent the N50 and N90 lengths. A light grey spiral at the centre shows the cumulative scaffold count on a logarithmic scale. A summary of complete, fragmented, duplicated, and missing BUSCO genes in the set is presented at the top right. An interactive version of this figure can be accessed on the BlobToolKit viewer.

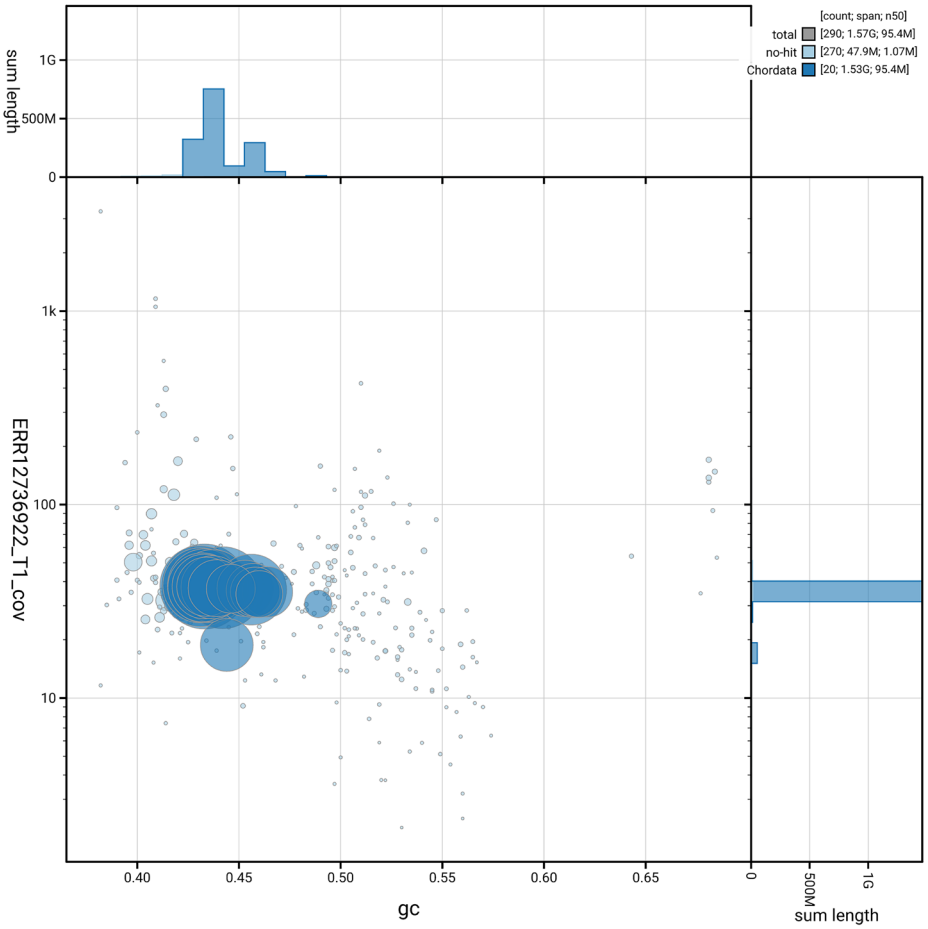

Figure 6. BlobToolKit GC-coverage plot for rPodLio1.hap1.1. Blob plot showing sequence coverage (vertical axis) and GC content (horizontal axis). The circles represent scaffolds, with the size proportional to scaffold length and the colour representing phylum membership. The histograms along the axes display the total length of sequences distributed across different levels of coverage and GC content. An interactive version of this figure is available on the BlobToolKit viewer.

Table 4. Earth Biogenome Project summary metrics for the *Podarcis liolepis* assembly.

Measure	Value	Benchmark
EBP summary (haplotype 1)	6.C.Q64	6.C.Q40
Contig N50 length	2.88 Mb	≥ 1 Mb
Scaffold N50 length	95.42 Mb	= chromosome N50
Consensus quality (QV)	Haplotype 1: 64.5; haplotype 2: 65.1; combined: 64.8	≥ 40
k-mer completeness	Haplotype 1: 73.88%; Haplotype 2: 70.62%; combined: 99.69%	≥ 95%
BUSCO	C:94.8% [S:92.6%; D:2.2%]; F:0.8%; M:4.3%; n:7 480	S > 90%; D < 5%
Percentage of assembly assigned to chromosomes	97.05%	≥ 90%

Wellcome Sanger Institute employs a process whereby due diligence is carried out proportionate to the nature of the materials themselves, and the circumstances under which they have been/are to be collected and provided for use. The purpose of this is to address and mitigate any potential legal and/or ethical implications of receipt and use of the materials as part of the research project, and to ensure that in doing so we align with best practice wherever possible.

The overarching areas of consideration are:

- Ethical review of provenance and sourcing of the material
- Legality of collection, transfer and use (national and international)

Each transfer of samples is undertaken according to a Research Collaboration Agreement or Material Transfer Agreement entered into by the Tree of Life collaborator, Genome Research Limited (operating as the Wellcome Sanger Institute) and in some circumstances other Tree of Life collaborators.

Data availability

European Nucleotide Archive: Podarcis liolepis (Columbretes wall lizard). Accession number PRJEB73727. The genome sequence is released openly for reuse. The *Podarcis liolepis* genome sequencing initiative is part of the Sanger Institute Tree of Life Programme (PRJEB43745) and Vertebrate Genomes Project (PRJNA489243). All raw sequence data and the assembly have been deposited in INSDC databases. The genome will be annotated using available RNA-Seq data and presented through the Ensembl pipeline at the European Bioinformatics Institute. Raw data and assembly accession identifiers are reported in Table 1 and Table 2.

Production code used in genome assembly at the WSI Tree of Life is available at https://github.com/sanger-tol. Table 5 lists software versions used in this study.

Author information

Contributors are listed at the following links:

- Members of the Wellcome Sanger Institute Tree of Life Management, Samples and Laboratory team
- Members of Wellcome Sanger Institute Scientific Operations – Sequencing Operations
- Members of the Wellcome Sanger Institute Tree of Life Core Informatics team
- Members of the Tree of Life Core Informatics collective

Table 5. Software versions and sources.

Software	Version	Source
BEDTools	2.30.0	https://github.com/arq5x/bedtools2
BLAST	2.14.0	ftp://ftp.ncbi.nlm.nih.gov/blast/executables/blast+/
BlobToolKit	4.3.9	https://github.com/blobtoolkit/blobtoolkit
BUSCO	5.5.0	https://gitlab.com/ezlab/busco
bwa-mem2	2.2.1	https://github.com/bwa-mem2/bwa-mem2
Cooler	0.8.11	https://github.com/open2c/cooler
DIAMOND	2.1.8	https://github.com/bbuchfink/diamond
fasta_windows	0.2.4	https://github.com/tolkit/fasta_windows
FastK	1.1	https://github.com/thegenemyers/FASTK
GenomeScope2.0	2.0.1	https://github.com/tbenavi1/genomescope2.0
Gfastats	1.3.6	https://github.com/vgl-hub/gfastats
GoaT CLI	0.2.5	https://github.com/genomehubs/goat-cli
Hifiasm	0.19.8-r603	https://github.com/chhylp123/hifiasm
HiGlass	1.13.4	https://github.com/higlass/higlass
MerquryFK	1.1.2	https://github.com/thegenemyers/MERQURY.FK
Minimap2	2.24-r1122	https://github.com/lh3/minimap2
MitoHiFi	3	https://github.com/marcelauliano/MitoHiFi

Software	Version	Source
MultiQC	1.14; 1.17 and 1.18	https://github.com/MultiQC/MultiQC
Nextflow	23.10.0	https://github.com/nextflow-io/nextflow
PretextSnapshot	N/A	https://github.com/sanger-tol/PretextSnapshot
PretextView	0.2.5	https://github.com/sanger-tol/PretextView
samtools	1.19.2	https://github.com/samtools/samtools
sanger-tol/ascc	0.1.0	https://github.com/sanger-tol/ascc
sanger-tol/ blobtoolkit	0.6.0	https://github.com/sanger-tol/blobtoolkit
Seqtk	1.3	https://github.com/lh3/seqtk
Singularity	3.9.0	https://github.com/sylabs/singularity
TreeVal	1.2.0	https://github.com/sanger-tol/treeval
YaHS	1.2a.2	https://github.com/c-zhou/yahs

References

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

Altschul SF, Gish W, Miller W, *et al.*: **Basic Local Alignment Search Tool.** *J Mol Biol.* 1990; **215**(3): 403–410.
PubMed Abstract | Publisher Full Text

Andrade P, Pinho C, Pérez I de Lanuza G, *et al.*: **Regulatory changes in pterin and carotenoid genes underlie balanced color polymorphisms in the wall lizard.** *Proc Natl Acad Sci U S A.* 2019; **116**(12): 5633–42.
PubMed Abstract | Publisher Full Text | Free Full Text

Bateman A, Martin MJ, Orchard S, *et al.*: **UniProt: the Universal Protein Knowledgebase in 2023.** *Nucleic Acids Res.* 2023; **51**(D1): D523–D531.
PubMed Abstract | Publisher Full Text | Free Full Text

Blaxter M, Mieszkowska N, Di Palma F, *et al.*: **Sequence locally, think globally: the Darwin Tree of Life Project.** *Proc Natl Acad Sci U S A.* 2022; **119**(4): e2115642118.
PubMed Abstract | Publisher Full Text | Free Full Text

Böhme W: **Handbuch der Reptilien und Amphibien Europas. Vol. 2/II.** Aula-Verlag, 1986.

Bowles P: ***Podarcis liolepis*.** The IUCN red list of threatened species 2024: e.T204900953A204901543. 2024.
Publisher Full Text

Buchfink B, Reuter K, Drost HG: **Sensitive protein alignments at Tree-of-Life scale using DIAMOND.** *Nat Methods.* 2021; **18**(4): 366–368.
PubMed Abstract | Publisher Full Text | Free Full Text

Caeiro-Dias G, Luís C, Pinho C, *et al.*: **Lack of congruence of genetic and niche divergence in *Podarcis hispanicus* complex.** *J Zool Syst Evol Res.* 2018; **56**(4): 479–92.
Publisher Full Text

Caeiro-Dias G, Rocha S, Couto A, *et al.*: **Nuclear phylogenies and genomics of a contact zone establish the species rank of *Podarcis lusitanicus* (squamata, lacertidae).** *Mol Phylogenet Evol.* 2021; **164**: 107270.
PubMed Abstract | Publisher Full Text

Carazo P, Font E, Desfilis E: **The role of scent marks in female choice of territories and refuges in a lizard (*Podarcis hispanica*).** *J Comp Psychol.* 2011; **125**(3): 362–65.
PubMed Abstract | Publisher Full Text

Carretero MA, Salvador A: **Lagartija parda – *Podarcis liolepis*.** 2016.
Reference Source

Castilla AM, Bauwens D: **Observations on the natural history, present status, and conservation of the insular lizard *Podarcis hispanica atrata* on the Columbretes archipelago, Spain.** *Biol Conserv.* 1991a; **58**(1): 69–84.
Publisher Full Text

Castilla AM, Bauwens D: **Thermal biology, microhabitat selection, and conservation of the insular lizard *Podarcis hispanica atrata*.** *Oecologia.* 1991b; **85**(3): 366–74.
PubMed Abstract | Publisher Full Text

Challis R, Richards E, Rajan J, *et al.*: **BlobToolKit – interactive quality assessment of genome assemblies.** *G3 (Bethesda).* 2020; **10**(4): 1361–1374.
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

Cheng H, Jarvis ED, Fedrigo O, *et al.*: **Haplotype-resolved assembly of diploid genomes without parental data.** *Nat Biotechnol.* 2022; **40**(9): 1332–1335.
PubMed Abstract | Publisher Full Text | Free Full Text

Danecek P, Bonfield JK, Liddle J, *et al.*: **Twelve years of SAMtools and BCFtools.** *GigaScience.* 2021; **10**(2): giab008.
PubMed Abstract | Publisher Full Text | Free Full Text

Ewels P, Magnusson M, Lundin S, *et al.*: **MultiQC: summarize analysis results for multiple tools and samples in a single report.** *Bioinformatics.* 2016; **32**(19): 3047–3048.
PubMed Abstract | Publisher Full Text | Free Full Text

Ewels PA, Peltzer A, Fillinger S, *et al.*: **The nf-core framework for community-curated bioinformatics pipelines.** *Nat Biotechnol.* 2020; **38**(3): 276–278.
PubMed Abstract | Publisher Full Text

Feiner N, Uller T, Abalos J, *et al.*: **The genome sequence of the Andalusian wall lizard, *Podarcis vaucheri* (Boulenger, 1905) [version 1; peer review: awaiting peer review].** *Wellcome Open Res.* 2025a; **10**: 265.
Publisher Full Text

Feiner N, Uller T, Carretero MA, *et al.*: **The genome sequence of Bocage's wall lizard, *Podarcis bocagei* (Lopez-Seoane, 1885) [version 1; peer review: 2 approved].** *Wellcome Open Res.* 2025b; **10**: 246.
PubMed Abstract | Publisher Full Text | Free 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

Gaczorek TS, Chechetkin M, Dudek K, *et al.*: **Widespread introgression of MHC genes in Iberian *Podarcis* lizards.** *Mol Ecol.* 2023; **32**(14): 4003–17.
PubMed Abstract | Publisher Full Text

Howard C, Denton A, Jackson B, *et al.*: **On the path to reference genomes for all biodiversity: lessons learned and laboratory protocols created in the Sanger Tree of Life core laboratory over the first 2000 species.** *bioRxiv.* 2025.
Publisher Full Text

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

Kaliontzopoulou A, Pinho C, Harris DJ, *et al.*: **When cryptic diversity blurs the picture: a cautionary tale from Iberian and North African *Podarcis* wall lizards.** *Biol J Linn Soc.* 2011; **103**(4): 779–800.
Publisher 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

Kurtzer GM, Sochat V, Bauer MW: **Singularity: scientific containers for mobility of compute.** *PLoS One.* 2017; **12**(5): e0177459.
PubMed Abstract | Publisher Full Text | Free Full Text

Li H: **Minimap2: pairwise alignment for nucleotide sequences.** *Bioinformatics.* 2018; **34**(18): 3094–3100.
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

Meiri S, Avila L, Bauer AM, *et al.*: **The global diversity and distribution of lizard clutch sizes.** *Glob Ecol Biogeogr.* 2020; **29**(9): 1515–30.
Publisher Full Text

Merkel D: **Docker: lightweight Linux containers for consistent development and deployment.** *Linux J.* 2014; **2014**(239): 2.
Reference Source

Pinho C, Harris DJ, Ferrand N: **Non-equilibrium estimates of gene flow inferred from nuclear genealogies suggest that Iberian and North African wall lizards (*Podarcis* spp.) are an assemblage of incipient species.** *BMC Evol Biol.* 2008; **8**(1): 63.
PubMed Abstract | Publisher Full Text | Free Full Text

Ranallo-Benavidez TR, Jaron KS, Schatz MC: **GenomeScope 2.0 and Smudgeplot for reference-free profiling of polyploid genomes.** *Nat Commun.* 2020; **11**(1): 1432.
PubMed Abstract | Publisher Full Text | Free 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, McCarthy SA, Fedrigo O, *et al.*: **Towards complete and error-free genome assemblies of all vertebrate species.** *Nature.* 2021; **592**(7856): 737–746.
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

Schoch CL, Ciufo S, Domrachev M, *et al.*: **NCBI taxonomy: a comprehensive update on curation, resources and tools.** *Database (Oxford).* 2020; **2020**: baaa062.
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

Yang W, Feiner N, Pinho C, *et al.*: **Extensive introgression and mosaic genomes of Mediterranean endemic lizards.** *Nat Commun.* 2021; **12**(1): 2762.
PubMed Abstract | Publisher Full Text | Free 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 29 October 2025

https://doi.org/10.21956/wellcomeopenres.27378.r133063

✔ **Kathryn R. Elmer**
University of Glasgow, Glasgow, UK

This communication describes the sequencing and assembly of a new genome for the wall lizard. The approach follows the well established Tree of Life pipeline for both primary material and analysis of the sequence data. One haplotype is generated that includes the anticipated number of chromosomes including the Z and W and the other Haplotype to comparable scaffold quality. Made available, the data will be useful for a range of evolutionary and comparative biology applications. The details are sufficient for recreating the methods.

minor:

There seems to be a duplication of text related to the shearing method and DNA extraction

Is it correct that the HiC done on ethanol preserved tissue that was then frozen?

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: evolutionary biology

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 29 September 2025

https://doi.org/10.21956/wellcomeopenres.27378.r134271

Hong-Xin Xie
University of Glasgow, Glasgow, UK

The data note provides a de novo haplotype-resolved assembly of the Catalonian Wall Lizard, *Podarcis liolepis.* The presented methods are appropriate and of sufficient detail.

The presented assembly is highly continuous and of high quality as assessed by multiple methods, and especially, fully meets the EBP proposed reference genome standard.

The presented tables and figure plots are also clear and well organized. This is a piece of excellent-written work that I enjoyed.

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: evolutionary genomics, especially focusing on squamate reptiles.

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.