DATA NOTE

The genome sequence of Bocage's wall lizard, *Podarcis bocagei* (Lopez-Seoane, 1885)

[version 1; peer review: 2 approved]

Nathalie Feiner[1,2], Tobias Uller[2], Miguel A. Carretero[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]CIBIO-InBIO, University of Porto, Vairão, Portugal
[4]Wellcome Sanger Institute, Hinxton, England, UK

v1 **First published:** 19 May 2025, **10**:246
https://doi.org/10.12688/wellcomeopenres.24032.1
Latest published: 19 May 2025, **10**:246
https://doi.org/10.12688/wellcomeopenres.24032.1

Abstract

We present a genome assembly from a female specimen of *Podarcis bocagei* (Bocage's wall lizard; Chordata; Lepidosauria; Squamata; Lacertidae). The assembly contains two haplotypes with total lengths of 1,615.19 megabases and 1,473.58 megabases. Most of the haplotype 1 assembly (97.72%) is scaffolded into 20 chromosomal pseudomolecules, including the W and Z sex chromosomes. Most of haplotype 2 (97.4%) is scaffolded into 18 chromosomal pseudomolecules. The mitochondrial genome has also been assembled, with a length of 17.4 kilobases.

Keywords

Podarcis bocagei, Bocage's 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 19 May 2025	✓ view	✓ view

1. **Jessica Gomez-Garrido**, Centre Nacional d'Anàlisi Genòmica (CNAG), Barcelona, Spain

2. **Sunandan Das**, University of Helsinki, Helsinki, Finland

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

Corresponding author: Wellcome Sanger Institute Tree of Life Management, Samples and Laboratory team (Mark.Blaxter@sanger.ac.uk)

Author roles: Feiner N: Investigation, Resources, Writing – Original Draft Preparation, Writing – Review & Editing; **Uller T**: Investigation, Resources, Writing – Original Draft Preparation; **Carretero MA**: 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). M.A.C. was supported by the Fundação para a Ciência e Tecnologia (FCT, Portugal; project 2022.03391.PTDC).
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, 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 Research 2025, **10**:246 https://doi.org/10.12688/wellcomeopenres.24032.1

First published: 19 May 2025, **10**:246 https://doi.org/10.12688/wellcomeopenres.24032.1

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 bocagei* (Lopez-Seoane, 1885) (NCBI:txid90520)

Background

Bocage's wall lizard *Podarcis bocagei*[1] (Lopez-Seoane, 1885; Figure 1) is a diurnal lacertid lizard with a distribution area in the Northeast of the Iberian Peninsula. 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*. *P. bocagei* 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, with several hybrid zones (Caeiro-Dias *et al.*, 2018; Caeiro-Dias *et al.*, 2021; Caeiro-Dias *et al.*, 2023; Pinho *et al.*, 2008). *P. bocagei* is sister species to *P. lusitanicus* and they diverged roughly 2 MYA (Yang *et al.*, 2021).

Bocage's wall lizard inhabits a range of habitats within the Atlantic region of Iberia, including open deciduous woodland and scrubland, coastal sand dunes, agricultural walls, and urban environments (Böhme, 1986). Lizards can be active all year round, but breeding is restricted to spring and early summer (Galán, 1997). Females lay between one and three clutches of eggs: Galán (1997) and Meiri *et al.* (2020) provide an average clutch size of 3.4.

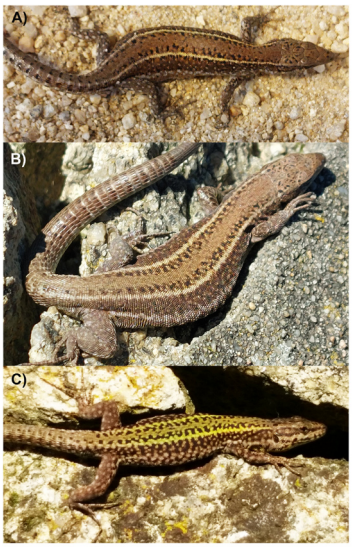

Figure 1. A) Photograph of the female *P. bocagei* (rPodBoc1) from Vila do Conde, Portugal, that was used for genome sequencing. **B**) Another female and **C**) a male from the same location in their natural habitat. Photographs by Miguel A. Carretero

[1] The species is named after the Portuguese naturalist José Vicente Barboza du Bocage.

Bocage's wall lizard is less variable in coloration than many other *Podarcis* lizards. Females are usually brown, while males develop bright green dorsal colouration in spring (Galán, 2008). This species is one of several *Podarcis* that exhibit ventral colour polymorphism with yellow, orange/red and white morphs (Andrade *et al.*, 2019).

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

Genome sequence report

Sequencing data

The genome of a specimen of *Podarcis bocagei* (Figure 1) was sequenced using Pacific Biosciences single-molecule HiFi long reads, generating 48.43 Gb (gigabases) from 6.28 million reads. GenomeScope analysis of the PacBio HiFi data estimated the haploid genome size at 1,549.83 Mb, with a heterozygosity of 0.86% and repeat content of 26.67%. These values provide an initial assessment of genome complexity and the challenges anticipated during assembly. Based on this estimated genome size, the sequencing data provided approximately 30.0x coverage of the genome. Chromosome conformation Hi-C sequencing produced 226.10 Gb from 1,497.37 million reads. Table 1 summarises the specimen and sequencing information.

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 assembly was improved by manual curation, which corrected 179 misjoins or missing joins. These interventions decreased the scaffold count by 15.08%. The final assembly has a total length of 1,615.19 Mb in 506 scaffolds, with 1,005 gaps, and a scaffold N50 of 94.95 Mb (Table 2).

The snail plot in Figure 2 provides a summary of the assembly statistics, indicating the distribution of scaffold lengths and other assembly metrics. Figure 3 shows the distribution of scaffolds by GC proportion and coverage. Figure 4 presents a cumulative assembly plot, with separate curves representing different scaffold subsets assigned to various phyla, illustrating the completeness of the assembly.

Most of the assembly sequence (97.72%) 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 5; Table 3). During curation, chromosomes Z and W were assigned based on Hi-C data and read coverage statistics. The order and orientation of the contigs making up

Table 1. Specimen and sequencing data for *Podarcis bocagei*.

Project information			
Study title	Podarcis bocagei (Bocage's wall lizard)		
Umbrella BioProject	PRJEB76351		
Species	*Podarcis bocagei*		
BioSpecimen	SAMEA115336769		
NCBI taxonomy ID	90520		
Specimen information			
Technology	**ToLID**	**BioSample accession**	**Organism part**
PacBio long read sequencing	rPodBoc1	SAMEA115336778	terminal body
Hi-C sequencing	rPodBoc1	SAMEA115336778	terminal body
Sequencing information			
Platform	**Run accession**	**Read count**	**Base count (Gb)**
Hi-C Illumina NovaSeq X	ERR13248954	1.50e+09	226.1
PacBio Revio	ERR13245283	6.28e+06	48.43

Chromosome W are uncertain, and some may remain in the unplaced scaffolds.

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

Assembly quality metrics

The estimated Quality Value (QV) and *k*-mer completeness metrics, along with BUSCO completeness scores, were calculated for each haplotype and the combined assembly. The QV reflects the base-level accuracy of the assembly, while *k*-mer completeness indicates the proportion of expected *k*-mers identified in the assembly. BUSCO scores provide a measure of completeness based on benchmarking universal single-copy orthologues.

For haplotype 1, the estimated QV is 61.8, and for haplotype 2, 62.2. When the two haplotypes are combined, the assembly achieves an estimated QV of 62.0. The *k*-mer recovery for haplotype 1 is 80.08%, and for haplotype 2 77.14%, while the combined haplotypes have a *k*-mer recovery of 98.71%. BUSCO analysis using the sauropsida_odb10 reference set (n = 7,480) identified 93.8% of the expected gene set (single = 91.7%, duplicated = 2.2%) for haplotype 1.

Table 2 provides assembly metric benchmarks adapted from Rhie *et al.* (2021) and the Earth BioGenome Project Report on Assembly Standards September 2024. The haplotype 1 assembly achieves the EBP reference standard of **6.C.Q61**.

Methods

Sample acquisition

The specimen, an adult female *P. bocagei* lizard (specimen ID SAN25002446, ToLID rPodBoc1) was collected on 2023-07-04 from a site close to Vila do Conde (latitude 41.353, longitude –8.7521; altitude: 11 m a.s.l.), where one of us (M.A.C.) was able to count up to 60 individual lizards during a transect of 10 min along 400 m of wall conducted in favorable weather conditions. 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 kept in captivity for the duration of one day and released at the collecting site. Field work was conducted under the permits no. 724-726/2023/CAPT issued by Instituto de Consevação da Natureza e das Florestas (Portugal).

Nucleic acid extraction

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 rPodBoc1 sample was prepared for DNA extraction by weighing and dissecting it on dry ice (Jay *et al.*, 2023). Tissue from the tail tip was homogenised using a PowerMasher II tissue disruptor (Denton *et al.*, 2023a). HMW DNA was extracted using the Manual MagAttract protocol (Strickland *et al.*, 2023b). DNA was sheared into an average fragment size of 12–20 kb in a Megaruptor 3 system (Todorovic *et al.*, 2023). Sheared DNA

Table 2. Genome assembly data for *Podarcis bocagei*.

Genome assembly	**Haplotype 1**	**Haplotype 2**
Assembly name	rPodBoc1.hap1.1	rPodBoc1.hap2.1
Assembly accession	GCA_964188305.1	GCA_964188325.1
Assembly level	chromosome	chromosome
Span (Mb)	1,615.19	1,473.58
Number of contigs	1,511	1,444
Number of scaffolds	506	627
Longest scaffold (Mb)	142.67	142.85
Assembly metrics* (benchmark)	**Haplotype 1**	**Haplotype 2**
Contig N50 length (≥ 1 Mb)	2.68 Mb	2.84 Mb
Scaffold N50 length *(= chromosome N50)*	94.95 Mb	94.91 Mb
Consensus quality (QV) (≥ 40)	61.8	62.2
k-mer completeness	80.08%	77.14%
Combined ***k***-mer completeness (≥ 95%)	98.71%	
BUSCO** (S > 90%; D < 5%)	C:93.8%[S:91.7%,D:2.2%], F:0.9%,M:5.2%,n:7,480	C:90.9%[S:89.2%,D:1.8%], F:1.0%,M:8.1%,n:7,480
Percentage of assembly mapped to chromosomes (≥ 90%)	97.72%	97.4%
Sex chromosomes (localised homologous pairs)	W and Z	-
Organelles (one complete allele)	Mitochondrial genome: 17.4 kb	-

* BUSCO scores based on the sauropsida_odb10 BUSCO set using version 5.5.0. C = complete [S = single copy, D = duplicated], F = fragmented, M = missing, n = number of orthologues in comparison.

was purified by solid-phase reversible immobilisation, using AMPure PB beads to eliminate shorter fragments and concentrate the DNA (Strickland *et al.*, 2023a). 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.

Hi-C sample preparation

Tissue from the terminal body of the rPodBoc1 sample was processed for Hi-C sequencing at the WSI Scientific Operations core, using the Arima-HiC v2 kit. In brief, 20–50 mg of frozentissue (stored at –80 °C) was fixed, and the DNA crosslinked using a TC buffer. After crosslinking, the tissue was homogenised using the Diagnocine Power Masher-II and BioMasher-II tubes and pestles. Following the Arima-HiC v2 kit manufacturer's instructions, crosslinked DNA was digested using a restriction enzyme master mix. The 5'-overhangs were 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. Additionally, the biotinylation percentage was estimated using the Qubit Fluorometer v4.0 (Thermo Fisher Scientific) and Qubit HS Assay Kit and Arima-HiC v2 QC beads.

Library preparation and sequencing

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

PacBio HiFi

At a 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

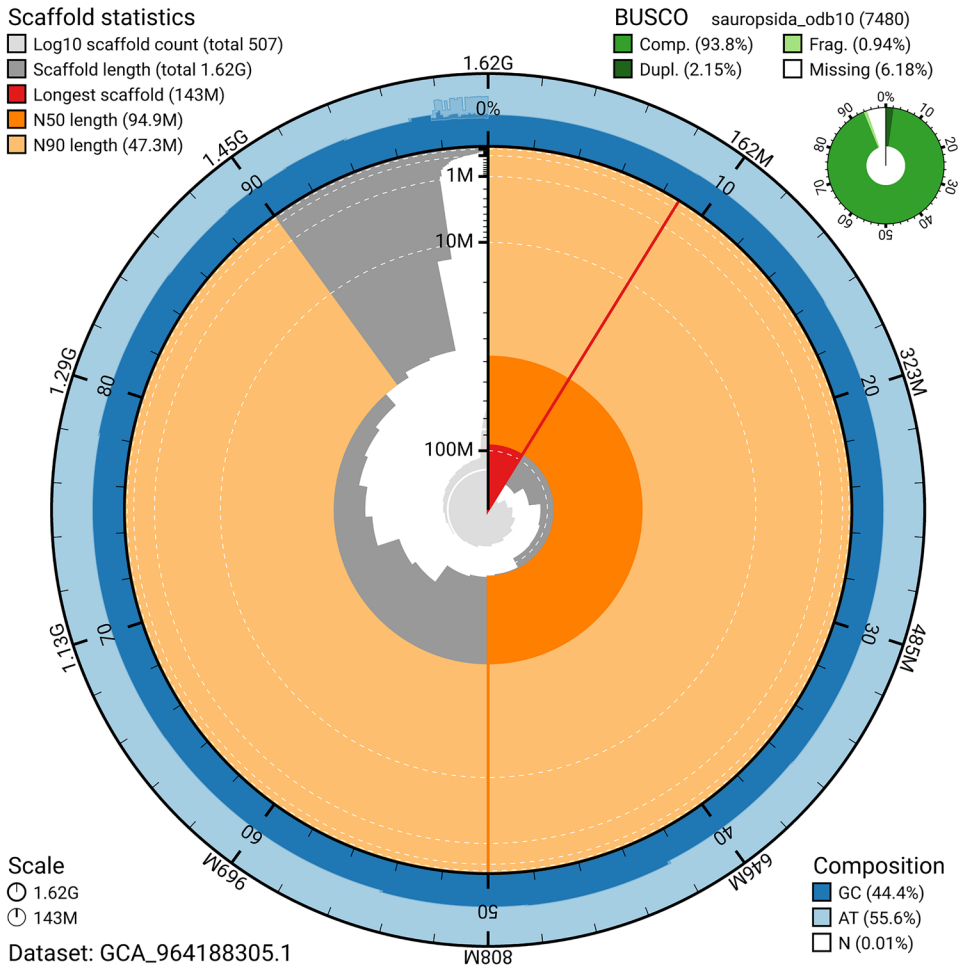

Figure 2. Genome assembly of *Podarcis bocagei*, rPodBoc1.hap1.1: metrics. The BlobToolKit snail plot provides an overview of assembly metrics and BUSCO gene completeness. The circumference represents the length of the whole genome sequence, and the main plot is divided into 1,000 bins around the circumference. 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 sauropside_odb10 set is presented at the top right. An interactive version of this figure is available at https://blobtoolkit.genomehubs.org/view/GCA_964188305.1/dataset/GCA_964188305.1/snail.

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.

Samples were sequenced on a Revio instrument (Pacific Biosciences, California, USA). Prepared libraries were normalised to 2 nM, and 15 µL was used for making complexes. Primers were annealed and polymerases were bound 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 200–300 pM), and spiked with a Revio sequencing internal control. Samples were sequenced 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.

Hi-C

For Hi-C library preparation, DNA was fragmented using the Covaris E220 sonicator (Covaris) and size selected using SPRISelect beads to 400 to 600 bp. The DNA was then

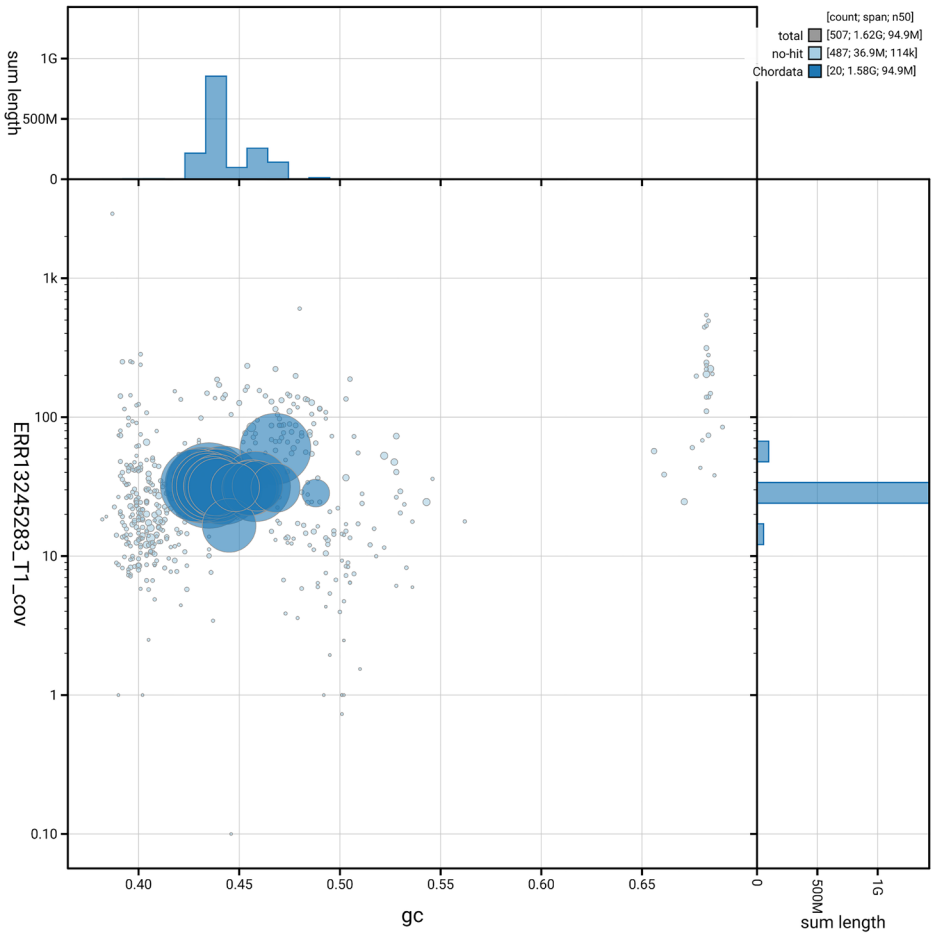

Figure 3. Genome assembly of *Podarcis bocagei*, rPodBoc1.hap1.1: BlobToolKit GC-coverage plot. Blob plot showing sequence coverage (vertical axis) and GC content (horizontal axis). The circles represent scaffolds, with the size proportional to scaffold length and the colour representing phylum membership. 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 at https://blobtoolkit.genomehubs.org/view/GCA_964188305.1/dataset/GCA_964188305.1/blob.

enriched using the Arima-HiC v2 kit Enrichment beads. Using the NEBNext Ultra II DNA Library Prep Kit (New England Biolabs) for end repair, A-tailing, and adapter ligation. This uses a custom protocol which resembles the standard NEBNext Ultra II DNA Library Prep protocol but where library preparation occurs while DNA is bound to the Enrichment beads. For library amplification, 10 to 16 PCR cycles were required, determined by the sample biotinylation percentage. The Hi-C sequencing was performed using paired-end sequencing with a read length of 150 bp on an Illumina NovaSeq X instrument.

Genome assembly, curation and evaluation

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), resulting in a pair of haplotype-resolved assemblies. 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 (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

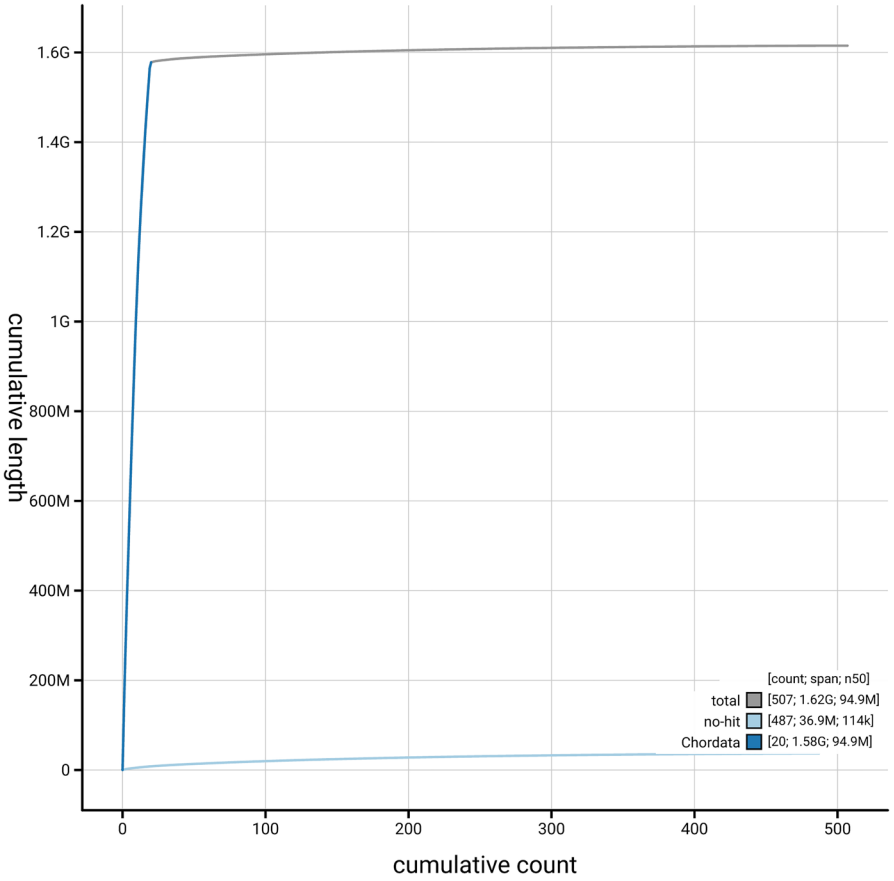

Figure 4. Genome assembly of *Podarcis bocagei*, rPodBoc1.hap1.1: BlobToolKit cumulative sequence plot. The grey line shows cumulative length for all scaffolds. Coloured lines show cumulative lengths of scaffolds assigned to each phylum using the buscogenes taxrule. An interactive version of this figure is available at https://blobtoolkit.genomehubs.org/view/GCA_964188305.1/dataset/GCA_964188305.1/cumulative.

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. Flat files and maps used in curation were generated via the TreeVal pipeline (Pointon *et al.*, 2023). Manual curation was conducted primarily in PretextView (Harry, 2022) and HiGlass (Kerpedjiev *et al.*, 2018), with additional insights provided by JBrowse2 (Diesh *et al.*, 2023). Scaffolds were visually inspected and corrected as described by Howe *et al.* (2021). Any identified contamination, missed joins, and mis-joins were amended, and duplicate sequences were tagged and removed. Sex chromosomes were identified by Hi-C and PacBio read coverage analysis. The curation method is documented at https://gitlab.com/wtsi-grit/rapid-curation.

Assembly quality assessment

The Merqury.FK tool (Rhie *et al.*, 2020), run in a Singularity container (Kurtzer *et al.*, 2017), was used 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.

A Hi-C contact map was produced for the final version of the assembly. The Hi-C reads were aligned using bwa-mem2 (Vasimuddin *et al.*, 2019) and the alignment files were combined using SAMtools (Danecek *et al.*, 2021). The Hi-C alignments were converted into a contact map using BEDTools (Quinlan & Hall, 2010) and the Cooler tool suite (Abdennur & Mirny, 2020). The contact map was visualised in HiGlass (Kerpedjiev *et al.*, 2018).

The blobtoolkit pipeline is a Nextflow (Di Tommaso *et al.*, 2017) port of the previous Snakemake Blobtoolkit pipeline (Challis *et al.*, 2020). It aligns the PacBio reads in SAMtools and minimap2 (Li, 2018) and generates coverage tracks for regions of fixed size. In parallel, it queries the GoaT database (Challis *et al.*, 2023) to identify all matching BUSCO lineages to run BUSCO (Manni *et al.*, 2021). For the three

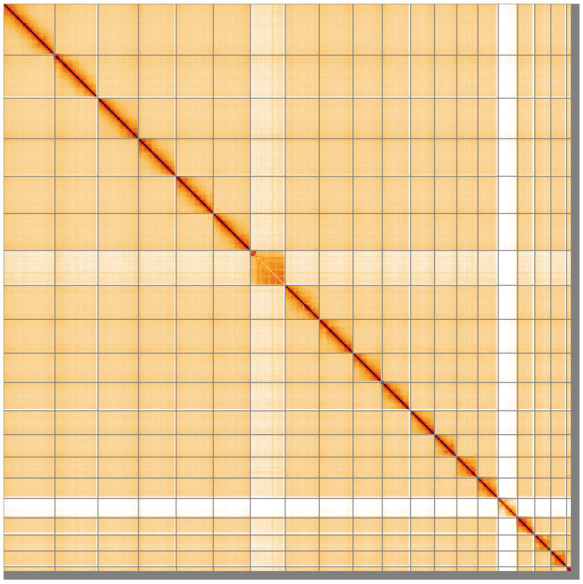

Figure 5. Genome assembly of *Podarcis bocagei*: Hi-C contact map of the rPodBoc1.hap1.1 assembly, visualised using HiGlass. Chromosomes are shown in order of size from left to right and top to bottom. An interactive version of this figure may be viewed at https://genome-note-higlass.tol.sanger.ac.uk/l/?d=QOrcy5OFRcGcohYoIzzG2w.

Table 3. Chromosomal pseudomolecules in the genome assembly of *Podarcis bocagei*, rPodBoc1.

Haplotype 1				Haplotype 2			
INSDC accession	**Name**	**Length (Mb)**	**GC%**	**INSDC accession**	**Name**	**Length (Mb)**	**GC%**
OZ076885.1	1	142.67	43.5	OZ077017.1	1	142.85	43.5
OZ076886.1	2	119.68	44	OZ077018.1	2	133.17	44.5
OZ076887.1	3	112.38	43	OZ077019.1	3	127.84	43.5
OZ076888.1	4	104.57	43.5	OZ077020.1	4	105.96	43.5
OZ076889.1	5	103.51	44	OZ077021.1	5	104.0	43
OZ076890.1	6	103.19	43	OZ077022.1	6	103.43	44
OZ076892.1	7	94.95	46	OZ077023.1	7	94.91	46
OZ076893.1	8	93.79	43.5	OZ077024.1	8	94.25	43.5
OZ076894.1	9	81.34	43.5	OZ077025.1	9	81.11	43.5
OZ076895.1	10	78.82	44	OZ077026.1	10	78.26	44
OZ076896.1	11	66.27	43.5	OZ077027.1	11	62.56	44
OZ076897.1	12	62.63	44	OZ077028.1	12	55.85	45.5
OZ076898.1	13	58.29	45.5	OZ077029.1	13	54.53	43.5
OZ076899.1	14	56.56	45.5	OZ077030.1	14	47.52	46
OZ076901.1	15	47.3	46	OZ077031.1	15	47.44	45.5
OZ076902.1	16	44.62	47	OZ077032.1	16	44.61	47
OZ076903.1	17	43.69	45	OZ077033.1	17	43.45	44.5
OZ076904.1	18	13.17	49	OZ077034.1	18	13.54	49
OZ076891.1	W	96.53	47				
OZ076900.1	Z	54.27	44.5				
OZ076905.1	MT	0.02	39				

domain-level BUSCO lineages, the pipeline aligns the BUSCO genes to the UniProt Reference Proteomes database (Bateman *et al.*, 2023) with DIAMOND blastp (Buchfink *et al.*, 2021). The genome is also divided into chunks according to the density of the BUSCO genes from the closest taxonomic lineage, and each chunk is aligned to the UniProt Reference Proteomes database using DIAMOND blastx. Genome sequences without a hit are chunked using seqtk and aligned to the NT database with blastn (Altschul *et al.*, 1990). The blobtools suite combines all these 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), relying on the Conda package manager, the Bioconda initiative (Grüning *et al.*, 2018), the Biocontainers infrastructure (da Veiga Leprevost *et al.*, 2017), as well as the Docker (Merkel, 2014) and Singularity (Kurtzer *et al.*, 2017) containerisation solutions.

Table 4 contains a list of relevant software tool versions and sources.

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 Wellcome Sanger Institute employs a process whereby due

Table 4. Software tools: versions and sources.

Software tool	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	666652151335353eef2fcd58880bcef5bc2928e1	https://github.com/thegenemyers/FASTK
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	44086069ee7d4d3f6f3f0012569789ec138f42b84a a44357826c0b6753eb28de	https://github.com/higlass/higlass
MerquryFK	d00d98157618f4e8d1a9190026b19b471055b22e	https://github.com/thegenemyers/MERQURY.FK
Minimap2	2.24-r1122	https://github.com/lh3/minimap2
MitoHiFi	3	https://github.com/marcelauliano/MitoHiFi
MultiQC	1.14, 1.17, and 1.18	https://github.com/MultiQC/MultiQC
Nextflow	23.10.0	https://github.com/nextflow-io/nextflow
PretextView	0.2.5	https://github.com/sanger-tol/PretextView
samtools	1.19.2	https://github.com/samtools/samtools
sanger-tol/ascc	-	https://github.com/sanger-tol/ascc
sanger-tol/blobtoolkit	0.5.1	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

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 bocagei (Bocage's wall lizard). Accession number PRJEB76351; https://identifiers.org/ena.embl/PRJEB76351. The genome sequence is released openly for reuse. The *Podarcis bocagei* genome assembly is provided by the Wellcome Sanger Institute Tree of Life Programme (https://www.sanger.ac.uk/programme/tree-of-life/). 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.

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.12162482.

Members of Wellcome Sanger Institute Scientific Operations: Sequencing Operations are listed here: https://doi.org/10.5281/zenodo.12165051.

Members of the Wellcome Sanger Institute Tree of Life Core Informatics team are listed here: https://doi.org/10.5281/zenodo.12160324.

Members of the Tree of Life Core Informatics collective are listed here: https://doi.org/10.5281/zenodo.12205391.

References

Abdennur N, Mirny LA: **Cooler: scalable storage for Hi-C data and other genomically labeled arrays.** *Bioinformatics.* 2020; **36**(1): 311–316.
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

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, de Lanuza GPI, *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): 5633–5642.
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

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

Bowles P: **Podarcis bocagei**. 2024; [Accessed 13 February 2025].
Reference Source

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, Brelsford A, Meneses-Ribeiro M, *et al.*: **Hybridization in late stages of speciation: strong but incomplete genome-wide reproductive isolation and 'large Z-effect' in a moving hybrid zone.** *Mol Ecol.* 2023; **32**(15): 4362–4380.
PubMed Abstract | Publisher 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): 479–492.
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

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

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

da Veiga Leprevost F, Grüning BA, Alves Aflitos S, *et al.*: **BioContainers: an open-source and community-driven framework for software standardization.** *Bioinformatics.* 2017; **33**(16): 2580–2582.
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

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

Di Tommaso P, Chatzou M, Floden EW, *et al.*: **Nextflow enables reproducible computational workflows.** *Nat Biotechnol.* 2017; **35**(4): 316–319.
PubMed Abstract | 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

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

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

Galán P: **Reproductive ecology of the lacertid lizard *Podarcis bocagei*.** *Ecography.* 1997; **20**(2): 197–209.
Publisher Full Text

Galán P: **Ontogenetic and sexual variation in the coloration of the lacertid lizards *Iberolacerta monticola* and *Podarcis bocagei*. Do the females prefer the greener males?** *Anim Biol.* 2008; **58**(2): 173–198.
Publisher Full Text

Grüning B, Dale R, Sjödin A, *et al.*: **Bioconda: sustainable and comprehensive software distribution for the life sciences.** *Nat Methods.* 2018; **15**(7): 475–476.
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

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): 1515–1530.
Publisher Full Text

Merkel D: **Docker: lightweight Linux containers for consistent development and deployment.** *Linux J.* 2014; **2014**(239): 2, [Accessed 2 April 2024].
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

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

Quinlan AR, Hall IM: **BEDTools: a flexible suite of utilities for comparing genomic features.** *Bioinformatics.* 2010; **26**(6): 841–842.
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

Strickland M, Cornwell C, Howard C: **Sanger Tree of Life fragmented DNA clean up: manual SPRI.** *protocols.io.* 2023a.
Publisher Full Text

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

Todorovic M, Sampaio F, Howard C: **Sanger Tree of Life HMW DNA fragmentation: diagenode Megaruptor®3 for PacBio HiFi.** *protocols.io.* 2023.
Publisher 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 14 July 2025

https://doi.org/10.21956/wellcomeopenres.26514.r125439

✔ **Sunandan Das**
University of Helsinki, Helsinki, Finland

In this report, authors report the reference quality genome for *Podarcis bocagei*. Using PacBio HiFi long read sequencing and Hi-C conformation capture sequencing, authors have generated 20 chromosome pseudomolecules. The sex chromosomes have also been identified. The data report is well written. The sequencing method used are generally the state-of-the-art for non-model wildlife. Assembly quality parameters are similar to values reported in contemporary non-model animal genome sequence reports/papers. The method is very detailed, including the in-silico steps, and should allow complete reproduction of the result with similar quality tissue sample. I regard this report to be a good contribution in *Podarcis* population and comparative genomics that should be approved.

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: Systematics, Phylogenomics, Speciation Genomics, Herpetology

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

Reviewer Report 13 June 2025

https://doi.org/10.21956/wellcomeopenres.26514.r124760

Jessica Gomez-Garrido
Centre Nacional d'Anàlisi Genòmica (CNAG), Barcelona, Barcelona, Spain

In this genome note, Feiner et al. present the genome assembly of *Podarcis bocagei*, a species of the genus podarcis that inabits the NortEast of the Iberian Peninsula. 28 different species have been described in the podarcis genus and some of these species also have available reference genomes. This new resource will allow comparative and evolutionary studies within the Podarcis genus and the Lacertidae family.

The sequencing and assembly strategies that the authors follow here are standard for generation of reference genome assemblies and give good results. Two chromosome-level haplotypes are reported here, with good percentage of chromosomal assignments. Both sex chromosomes have been identified and included in haplotype 1, although due to the repetitive nature of chromosome W, it has not been possible to fully reconstruct it. The mitogenome has also been assembled and included in the release.

Data has been correctly deposited and released in the INSDC databases.

Overall, I think this is a good new resource correctly described in this genome note.

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: Genome Assembly and Annotation

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.