DATA NOTE

The genome sequence of the Andalusian wall lizard, *Podarcis vaucheri* (Boulenger, 1905)

[version 1; peer review: 2 approved]

Nathalie Feiner[1,2], Tobias Uller[2], Javier Abalos[2,3], Tahar Slimani[4], Abderrahim S'Khifa[4], Joana Meier[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, Tree of Life Core Informatics collective

[1]Max Planck Institute for Evolutionary Biology, Plön, Germany
[2]Lund University, Lund, Sweden
[3]Ethology Lab, Cavanilles Institute of Biodiversity and Evolutionary Biology,, University of Valencia, Valencia, Spain
[4]Biodiversity and Ecology of Vertebrates Unit, Water, Biodiversity and Climate Change Laboratory, Faculty of Sciences Semlalia, Cadi Ayyad University, Marrakech, Morocco
[5]Tree of Life, Wellcome Sanger Institute, Hinxton, England, UK

v1 **First published:** 20 May 2025, **10**:265
https://doi.org/10.12688/wellcomeopenres.24168.1
Latest published: 20 May 2025, **10**:265
https://doi.org/10.12688/wellcomeopenres.24168.1

Abstract

We present a genome assembly from a female specimen of *Podarcis vaucheri* (Andalusian wall lizard; Chordata; Lepidosauria; Squamata; Lacertidae). The assembly contains two haplotypes with total lengths of 1,614.91 megabases and 1,510.74 megabases. Most of haplotype 1 (98.26%) is scaffolded into 20 chromosomal pseudomolecules, including the W and Z sex chromosomes. Most of haplotype 2 (98.63%) is scaffolded into 18 chromosomal pseudomolecules. The mitochondrial genome has also been assembled, with a length of 17.24 kilobases.

Keywords

Podarcis vaucheri, Vaucher'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 20 May 2025	✓ view	✓ view

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

2. **Jorge Gutiérrez-Rodríguez**, Universidad Nacional Autonoma de Mexico Instituto de Ecologia (Ringgold ID: 98857), Mexico City, Mexico

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

Corresponding author: 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; **Slimani T**: Resources, Writing – Original Draft Preparation, Writing – Review & Editing; **S'Khifa A**: 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).
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 Andalusian wall lizard, *Podarcis vaucheri* (Boulenger, 1905) [version 1; peer review: 2 approved]** Wellcome Open Research 2025, **10**:265
https://doi.org/10.12688/wellcomeopenres.24168.1

First published: 20 May 2025, **10**:265 https://doi.org/10.12688/wellcomeopenres.24168.1

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 vaucheri* (Boulenger, 1905) (NCBI:txid140051)

Background

The Andalusian wall lizard *Podarcis vaucheri* (named after the Swiss botanist Henri Vaucher) (Boulenger, 1905; Figure 1) is a diurnal lacertid lizard with a distribution area in Andalusia in southern Spain and Morocco in Northern Africa. An introduced population of *P. vaucheri* has been discovered on the outskirts of Athens in Greece (Spilani *et al.*, 2018). Another introduced population has been found in Eastern Spain (Renoult *et al.*, 2010), suggesting potential for further human-mediated expansion of the species. The Andalusian wall lizard is classified as Least Concern (LC) by the IUCN (Mateo *et al.*, 2009).

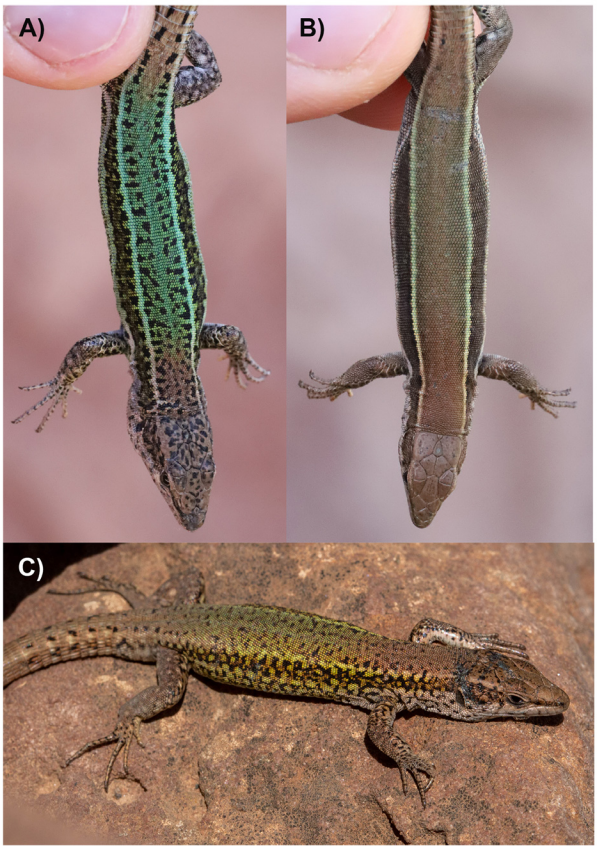

Figure 1. A) Male and **B**) female Andalusian wall lizards (*Podarcis vaucheri*) photographed immediately upon being captured near Oukaimeden (Morocco). **C**) Male *Podarcis vaucheri* photographed while basking on a rock near Tighedouine (Morocco). Note that the specimen selected for genome sequencing is from the same locality as specimens shown in panel **A** and **B**, but not shown on the photographs. Photographs by Javier Abalos.

This species is one of 28 currently described species of the genus *Podarcis*. Within this species, there is a deep split between a Northern lineage, occupying Spain and parts of Morocco, and a Southern lineage restricted to Morocco (Busack *et al.*, 2005). *Podarcis vaucheri* 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). *P. vaucheri* is sister species to *P. hispanicus* and they diverged roughly 4 MYA (Yang *et al.*, 2021).

The Andalusian wall lizard inhabits temperate forest, Mediterranean shrubland, rocky areas and urban environments. It can be found from sea level to >3000 m altitude (Mateo *et al.*, 2009). The activity period is mostly in spring and early summer, when reproduction takes place (Mamou *et al.*, 2017), although lizards can be seen all year round. The species is oviparous and Meiri *et al.* (2020) report an average clutch size of 3.00. The colouration is dominated by brown, with adult males having a green dorsum. This species is one of several *Podarcis* that exhibit ventral colour polymorphism with yellow, orange/red and white morphs (Andrade *et al.*, 2019).

P. vaucheri 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* and *P. liolepis*). This will be a valuable resource for studies of this species (e.g., its complex phylogeography in Africa; Busack *et al.*, 2005) and the Iberian and North African *Podarcis* more generally.

Genome sequence report

Sequencing data

The genome of a specimen of *Podarcis vaucheri* (Figure 1) was sequenced using Pacific Biosciences single-molecule HiFi long reads, generating 52.57 Gb (gigabases) from 6.04 million reads. GenomeScope analysis of the PacBio HiFi data estimated the haploid genome size at 1,526.71 Mb, with a heterozygosity of 0.57% and repeat content of 24.87%. 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 33.0x coverage of the genome. Chromosome conformation Hi-C sequencing produced 227.22 Gb from 1,504.76 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 204 misjoins or missing joins. These interventions decreased the scaffold count by 24.8% and increased the scaffold N50 by 1.49%. The final assembly has a total length of 1,614.91 Mb in 284 scaffolds, with 949 gaps, and a scaffold N50 of 99.22 Mb (Table 2).

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

Project information			
Study title	Podarcis vaucheri (Andalusian wall lizard)		
Umbrella BioProject	PRJEB77310		
Species	*Podarcis vaucheri*		
BioSpecimen	SAMEA115336770		
NCBI taxonomy ID	140051		
Specimen information			
Technology	**ToLID**	**BioSample accession**	**Organism part**
PacBio long read sequencing	rPodVau1	SAMEA115336779	terminal body
Hi-C sequencing	rPodVau1	SAMEA115336779	terminal body
Sequencing information			
Platform	**Run accession**	**Read count**	**Base count (Gb)**
Hi-C Illumina NovaSeq X	ERR13350322	1.50e+09	227.22
PacBio Revio	ERR13362598	6.04e+06	52.57

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

Genome assembly	**Haplotype 1**	**Haplotype 2**
Assembly name	rPodVau1.hap1.1	rPodVau1.hap2.1
Assembly accession	GCA_965113315.1	GCA_965113305.1
Assembly level	chromosome	chromosome
Span (Mb)	1,614.91	1,510.74
Number of contigs	1,233	1,003
Number of scaffolds	284	151
Longest scaffold (Mb)	141.9	144.97
Assembly metrics* (benchmark)	**Haplotype 1**	**Haplotype 2**
Contig N50 length (≥ 1 Mb)	2.97 Mb	2.99 Mb
Scaffold N50 length (= chromosome N50)	99.22 Mb	100.94 Mb
Consensus quality (QV) (≥ 40)	63.3	63.3
k-mer completeness	87.15%	83.32%
Combined *k*-mer completeness (≥ 95%)	99.56%	
BUSCO** (S > 90%; D < 5%)	C:95.1%[S:93.1%,D:2.1%], F:0.8%,M:4.0%,n:7,480	C:92.2%[S:90.5%,D:1.7%], F:0.9%,M:6.9%,n:7,480
Percentage of assembly mapped to chromosomes (≥ 90%)	98.26%	98.63%
Sex chromosomes (localised homologous pairs)	W and Z	-
Organelles (one complete allele)	Mitochondrial genome: 17.24 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.

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 (98.26%) 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, it was noted that The Z and W chromosomes were identified based on their single-copy status within a diploid assembly and synteny analysis with *Podarcis raffonei* (GCA_027172205.1). The exact order and orientation of contigs remain unknown for specific regions on several chromosomes: chromosome 2 (120.7–128.0 Mb), chromosome 6 (7.5–15.0 Mb), chromosome 7 (97.3–99.2 Mb), chromosome 9 (68.4–71.0 Mb), chromosome 13 (39.8–41.3 Mb), and chromosome W (37.4–41.2 Mb).

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

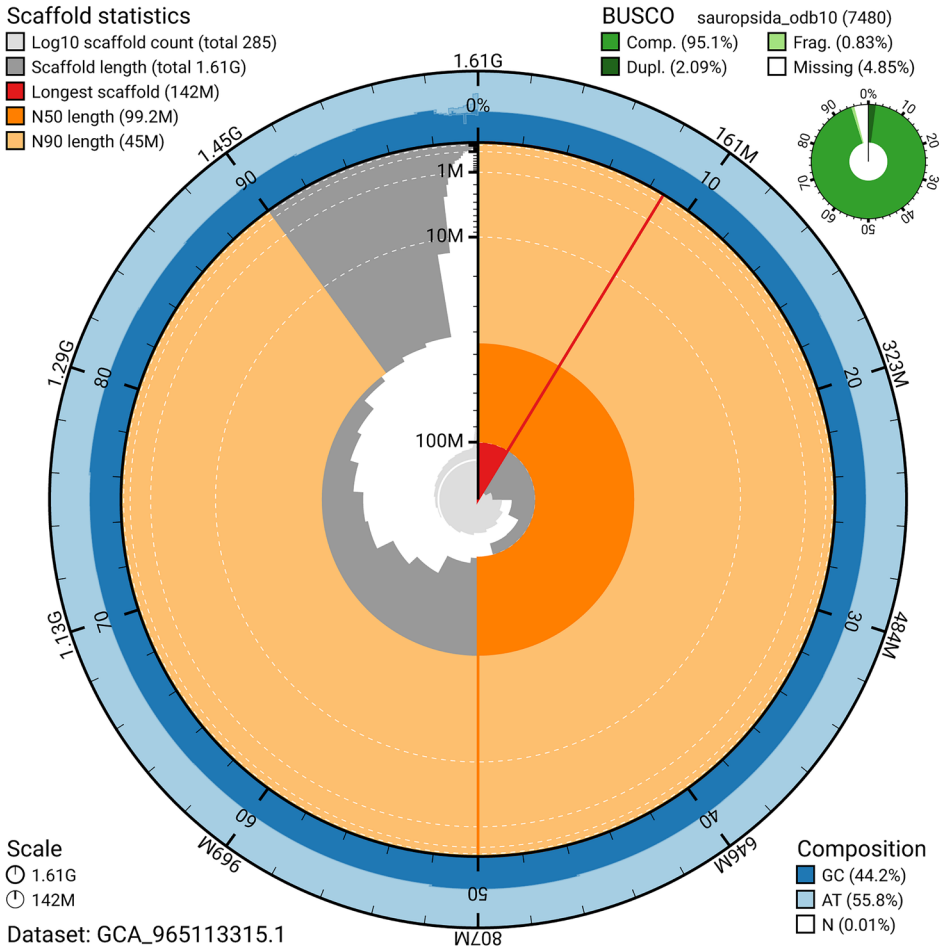

Figure 2. Genome assembly of *Podarcis vaucheri*, rPodVau1.hap1.1: metrics. The BlobToolKit snail plot provides an overview of assembly metrics and BUSCO gene completeness. The circumference represents the length of the whole genome sequence, and the main plot is divided into 1,000 bins around the circumference. 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 sauropsida_odb10 set is presented at the top right. An interactive version of this figure is available at https://blobtoolkit.genomehubs.org/view/GCA_965113315.1/dataset/GCA_965113315.1/snail.

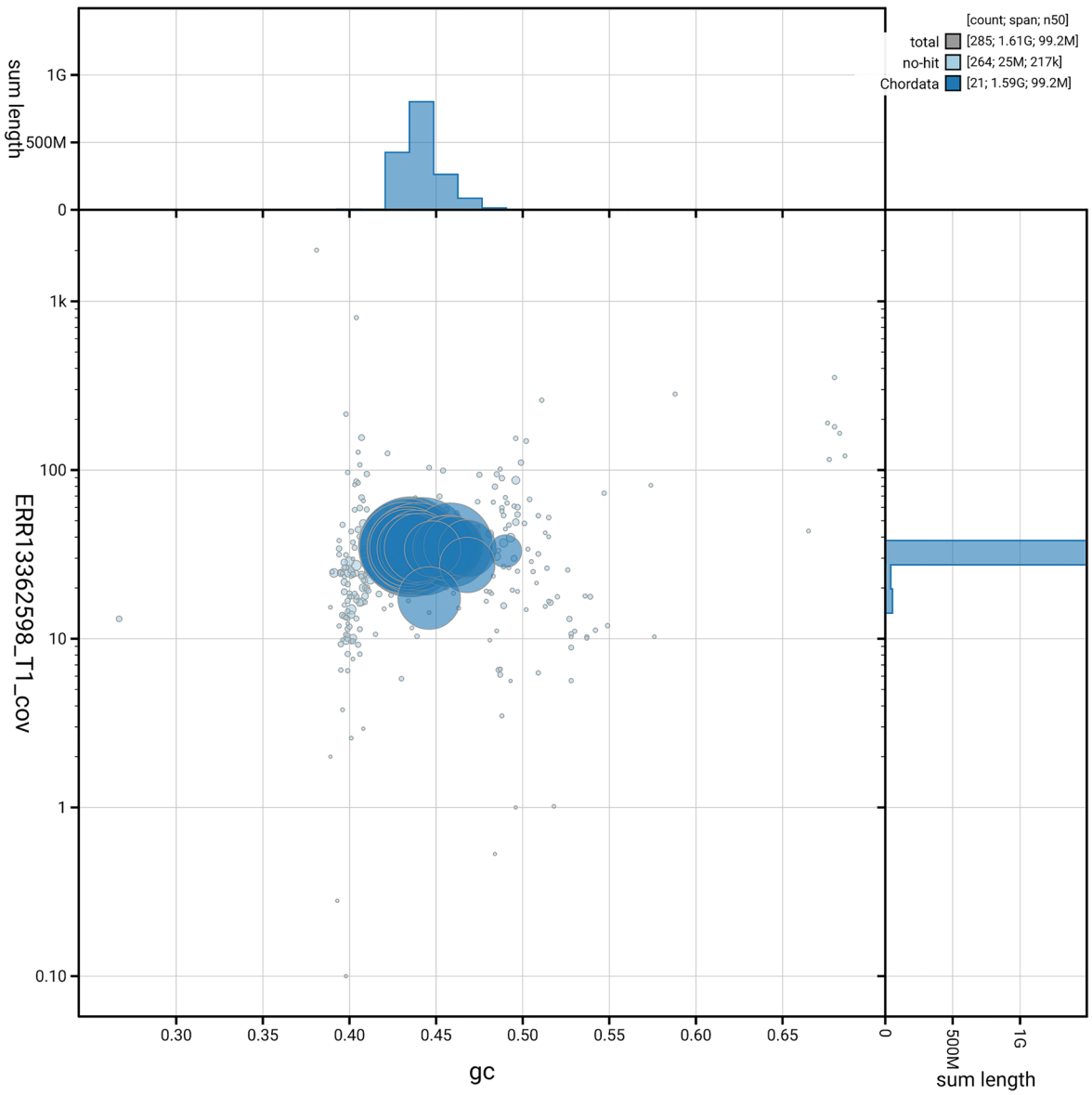

Figure 3. Genome assembly of *Podarcis vaucheri*, rPodVau1.hap1.1: BlobToolKit GC-coverage plot. Blob plot showing sequence coverage (vertical axis) and GC content (horizontal axis). The circles represent scaffolds, with the size proportional to scaffold length and the colour representing phylum membership. 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_965113315.1/dataset/GCA_965113315.1/blob.

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 63.3, and for haplotype 2, 63.3. When the two haplotypes are combined, the assembly achieves an estimated QV of 63.3. The *k*-mer recovery for haplotype 1 is 87.15%, and for haplotype 2 83.32%, while the combined haplotypes have a *k*-mer recovery of 99.56%. BUSCO v.5.5.0 analysis using the sauropsida_odb10 reference set (n = 7,480) identified 95.1% of the expected gene set (single = 93.1%, duplicated = 2.1%) 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 assembly achieves the EBP reference standard of **6.C.Q63.**

Methods

Sample acquisition

The specimen, an adult female *P. vaucheri* lizard (specimen ID SAN25002447, ToLID rPodVau1) was collected on 11 May 2023 from a site close to Oukaimeden in the Atlas Mountains in Morocco (latitude: 31.207837; longitude: –7.85302). The specimen was caught by noosing, standard morphometric

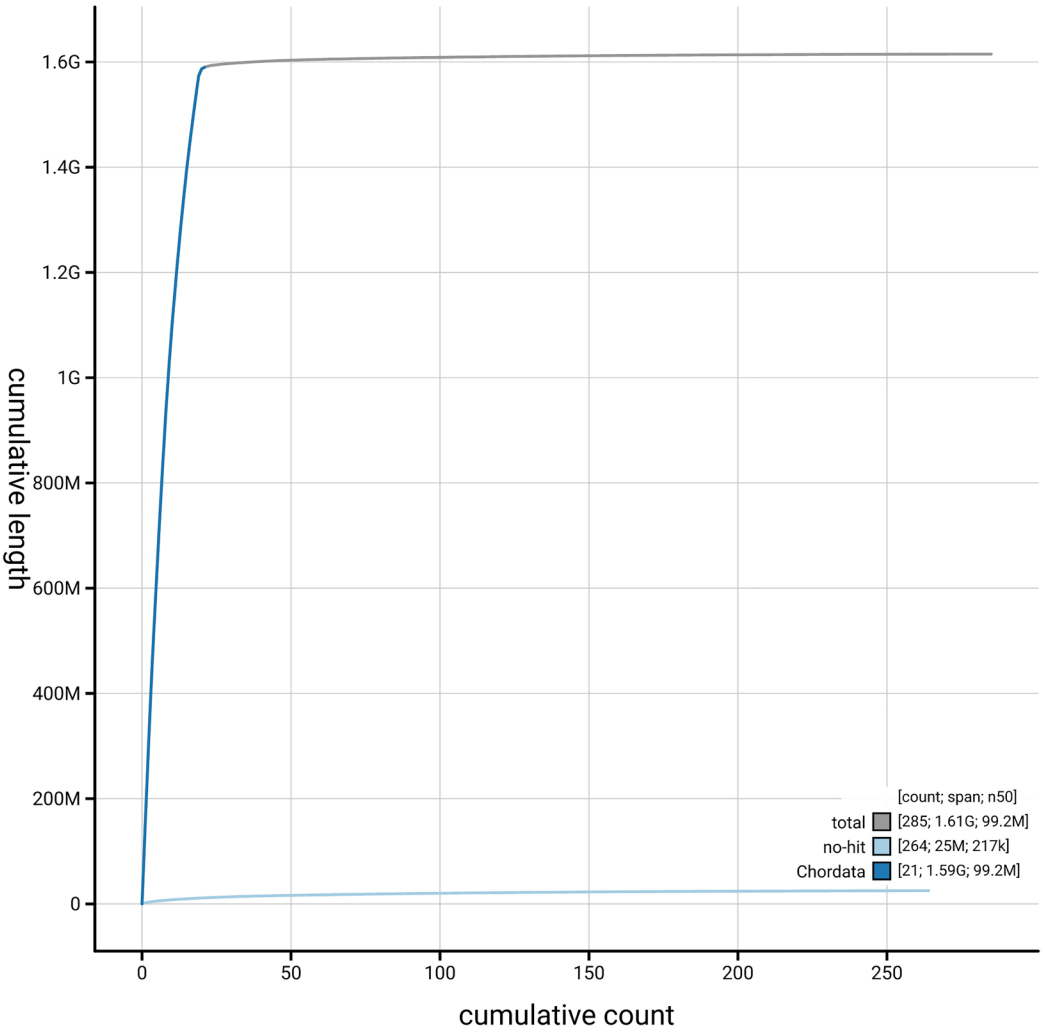

Figure 4. Genome assembly of *Podarcis vaucheri,* rPodVau1.hap1.1: BlobToolKit cumulative sequence plot. The grey line shows cumulative length for all scaffolds. Coloured lines show cumulative lengths of scaffolds assigned to each phylum using the buscogenes taxrule. An interactive version of this figure is available at https://blobtoolkit.genomehubs.org/view/GCA_965113315.1/dataset/GCA_965113315.1/cumulative.

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 Decision N°09/2023 ANEF/DPNAPPN/DPNAP/SECFF provided by the National Water and Forestry Agency-Morocco.

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 rPodVau1 sample was prepared for DNA extraction by weighing and dissecting it on dry ice (Jay *et al.*, 2023). Tissue from the terminal body was homogenised using a PowerMasher II tissue disruptor (Denton *et al.*, 2023a). HMW DNA was extracted using the Manual MagAttract v1 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 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 and crosslinking

Hi-C data were generated from the terminal body of the rPodVau1 sample using the Arima-HiC v2 kit (Arima Genomics) with 20–50 mg of frozen tissue (stored at –80 °C). As per manufacturer's instructions, tissue was fixed, and the DNA crosslinked using a TC buffer with 22% formalde-

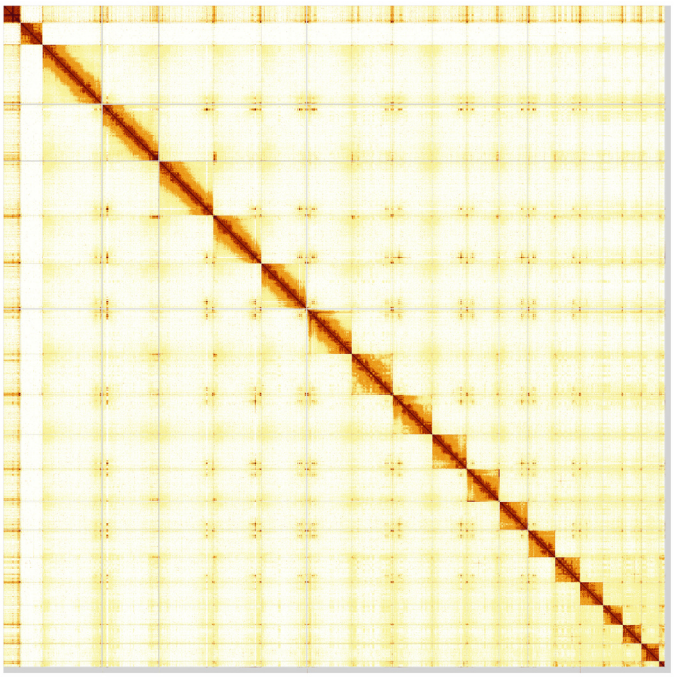

Figure 5. Genome assembly of *Podarcis vaucheri:* Hi-C contact map of the rPodVau1.hap1.1 assembly, visualised using PretextSnapshot. Chromosomes are shown in order of size.

Table 3. Chromosomal pseudomolecules in the genome assembly of *Podarcis vaucheri*, rPodVau1.

Haplotype 1				Haplotype 2			
INSDC accession	**Name**	**Length (Mb)**	**GC%**	**INSDC accession**	**Name**	**Length (Mb)**	**GC%**
OZ220363.1	1	141.9	43.5	OZ220345.1	1	144.97	43.5
OZ220364.1	2	135.18	44	OZ220346.1	2	134.17	44
OZ220365.1	3	130.68	43.5	OZ220347.1	3	129.37	43.5
OZ220366.1	4	116.34	43	OZ220348.1	4	114.44	43
OZ220367.1	5	108.31	43.5	OZ220349.1	5	107.87	43.5
OZ220368.1	6	108.03	44	OZ220350.1	6	106.68	43.5
OZ220369.1	7	99.22	46	OZ220351.1	7	100.94	46
OZ220370.1	8	95.65	43.5	OZ220352.1	8	94.82	43.5
OZ220371.1	9	83.3	43.5	OZ220353.1	9	84.23	43.5
OZ220372.1	10	76.87	44	OZ220354.1	10	76.4	44
OZ220373.1	11	66.83	43.5	OZ220355.1	11	66.38	43.5
OZ220374.1	12	65.34	44	OZ220356.1	12	65.46	44
OZ220375.1	13	60.09	45.5	OZ220357.1	13	59.51	45.5
OZ220376.1	14	55.72	45.5	OZ220358.1	14	55.73	45.5
OZ220377.1	15	47.58	46	OZ220359.1	15	47.08	46
OZ220378.1	16	44.95	47	OZ220360.1	16	45.1	47
OZ220379.1	17	42.97	45	OZ220361.1	17	43.16	45
OZ220380.1	18	13.25	49	OZ220362.1	18	13.74	49
OZ220381.1	W	41.27	47				
OZ220382.1	Z	53.36	44.5				
OZ220383.1	MT	0.02	39				

hyde concentration, and a final formaldehyde concentration of 2%. The tissue was then homogenised using the Diagnocine Power Masher-II. The crosslinked DNA was digested using a restriction enzyme master mix, then biotinylated and ligated. A clean up was performed with SPRIselect beads prior to library preparation. DNA concentration was quantified using the Qubit Fluorometer v4.0 (Thermo Fisher Scientific) and Qubit HS Assay Kit, and sample biotinylation percentage was estimated using the 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), depending on genome size and sequencing depth required. Libraries were prepared using the SMRTbell Prep Kit 3.0 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. Size-selection and clean-up were carried out using diluted AMPure PB beads (Pacific Biosciences). DNA concentration was quantified using the Qubit Fluorometer v4.0 (ThermoFisher 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 the 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 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 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, biotinylated DNA constructs were fragmented using the Covaris E220 sonicator and size selected using SPRISelect beads to 400 to 600 bp. The DNA was then enriched using the Arima-HiC v2 kit Enrichment beads. The NEBNext Ultra II DNA Library Prep Kit (New England Biolabs) was used for end repair, A-tailing, and adapter ligation. This uses a custom protocol which resembles the standard Ultra II DNA Library Prep protocol but where library preparation occurs while DNA is bound to the Enrichment beads. Library PCR amplification was carried out using KAPA HiFi Hot start mix and a custom IDT UDI (Unique Dual Index) 96 barcode plate (Integrated DNA Technologies). Depending on sample concentration and biotinylation percentage determined at the crosslinking stage, samples were run for 10–16 PCR cycles. Post-PCR, samples were cleaned-up using SPRISelect beads and the INTEGRA MINI 96 (INTEGRA). Hi-C sequencing was performed an Illumina NovaSeq X instrument, using paired-end sequencing with a read length of 150 bp.

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 (Rao *et al.*, 2014) were mapped to the primary contigs using bwa-mem2 (Vasimuddin *et al.*, 2019). The contigs were further scaffolded using the provided 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. 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. The curation process 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.

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 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 Darwin Tree of Life Partner. The submission

Table 4. Software tools: versions and sources.

Software tool	Version	Source
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
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	44086069ee7d4d3f6f3f0012569789ec138f42b84 aa44357826c0b6753eb28de	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	0.6.0	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

of materials by a Darwin Tree of Life Partner is subject to the **'Darwin Tree of Life Project Sampling Code of Practice'**, which can be found in full on the Darwin Tree of Life website here. By agreeing with and signing up to the Sampling Code of Practice, the Darwin Tree of Life Partner agrees they will meet the legal and ethical requirements and standards set out within this document in respect of all samples acquired for, and supplied to, the Darwin Tree of Life Project.

Further, the 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 further undertaken according to a Research Collaboration Agreement or Material Transfer Agreement entered into by the Darwin Tree of Life Partner, Genome Research Limited (operating as the Wellcome Sanger Institute), and in some circumstances other Darwin Tree of Life collaborators.

Data availability

European Nucleotide Archive: *Podarcis vaucheri* (Andalusian wall lizard). Accession number PRJEB77310; https://identifiers.org/ena.embl/PRJEB77310. The genome sequence is released openly for reuse. The *Podarcis vaucheri* 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

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): 5633–5642.
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

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

Busack S, Lawson R, Arjo W: **Mitochondrial DNA, allozymes, morphology and historical biogeography in the *Podarcis vaucheri* (Lacertidae) species complex.** *Amphibia-Reptilia.* 2005; **26**(2): 239–256.
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

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

Gaczorek TS, Chechetkin M, Dudek K, *et al.*: **Widespread introgression of MHC genes in Iberian *Podarcis* lizards.** *Mol Ecol.* 2023; **32**(14): 4003–4017.
PubMed Abstract | 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 Lond.* 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

Mamou R, Moudilou E, Amroun M, *et al.*: **Reproductive cycle of male wall lizard, *Podarcis vaucheri* (Reptilia: Sauria: Lacertidae), in Djurdjura, Northern Algeria.** In: Djurdjura, Northern Algeria, *Basic Appl Herpetol.* 2017; **31**: 77–89.
Publisher 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

Mateo JA, Cheylan M, Nouira MS, *et al.*: ***Podarcis vaucheri* (errata version published in 2016).** *The IUCN Red List of Threatened Species. 2009; e.T61556A86439309.* 2009; [Accessed 14 March 2025].
Publisher 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**: 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

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

Renoult J, Geniez P, Beddek M, *et al.*: **An isolated population of *Podarcis vaucheri* (Sauria: Lacertidae) in south-eastern Spain: genetic data suggest human-mediated range expansion.** *Amphibia-Reptilia.* 2010; **31**(2): 287–293.
Publisher 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

Spilani L, Strachinis I, Lampropoulos A, *et al.*: ***Podarcis vaucheri* (Sauria: Lacertidae) far away from home: a new invasive species in Greece.** *Amphibia-Reptilia.* 2018; **39**(3): 363–368.
Publisher 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 26 December 2025

https://doi.org/10.21956/wellcomeopenres.26659.r138095

✔ **Jorge Gutiérrez-Rodríguez**
Universidad Nacional Autonoma de Mexico Instituto de Ecologia (Ringgold ID: 98857), Mexico City, Mexico City, Mexico

The manuscript describes a high-quality genome sequencing and assembly effort that employs cutting-edge methodologies, including PacBio HiFi long reads and Hi-C phasing. The clear documentation of procedures and the thorough reporting of quality metrics align well with the highest standards for contemporary genome projects (e.g., those of the EBP). The generation of a haplotype-resolved, chromosome-level assembly represents a notable scientific contribution.

Minor comments:
In the sentence “During curation, it was noted that The Z and W chromosomes were identified based on their single-copy status within a diploid assembly and synteny analysis with Podarcis raffonei (GCA_027172205.1)”, please change “The” to “the”.
In the sentence: “P. vaucheri 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 and P. liolepis)” — add citations.

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 02 December 2025

https://doi.org/10.21956/wellcomeopenres.26659.r138098

Hong-Xin Xie
University of Glasgow, Glasgow, UK

The authors present a high-quality, haplotype-resolved, chromosome-level genome assembly of the Andalusian wall lizard *Podarcis vaucheri*, including both W and Z sex chromosomes. The Data Note is well written and provides sufficient biological context along with clear descriptions of the sequencing and assembly methods. The use of PacBio HiFi reads combined with Hi-C scaffolding is consistent with current best practices, and the resulting assembly appears to be of comparable quality to recently published squamate genomes. Overall, the manuscript meets the expectations for a genome Data Note and will be a valuable resource for the community. I have only two minor comments:

1. Figure 5 legend: The legend states that “Chromosomes are shown in order of size,” but the first two chromosomes depicted do not appear to be the largest; they seem to correspond to the sex chromosomes. It may be worth double-checking the chromosome ordering or clarifying the legend.
2. Assembly QV values: In the assembly quality metrics, the manuscript reports: “For haplotype 1, the estimated QV is 63.3, and for haplotype 2, 63.3. When the two haplotypes are combined, the assembly achieves an estimated QV of 63.3.” Given that all three values are identical, it may be useful to confirm whether this is correct or if a minor typo has occurred.

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: genomics, evolution, 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.