DATA NOTE

The genome sequence of the Dalmatian wall lizard, *Podarcis melisellensis* (Braun, 1877)

[version 1; peer review: 3 approved]

Nathalie Feiner[1,2], Tobias Uller[2], Lisa Van Linden[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]Lund University, Lund, Sweden
[3]Department of Biology, University of Antwerp, Antwerp, Flanders, Belgium
[4]Tree of Life, Wellcome Sanger Institute, Hinxton, England, UK

v1 **First published:** 19 May 2025, **10**:247
https://doi.org/10.12688/wellcomeopenres.24140.1
Latest published: 19 May 2025, **10**:247
https://doi.org/10.12688/wellcomeopenres.24140.1

Abstract

We present a genome assembly from a female specimen of *Podarcis melisellensis* (Dalmatian wall lizard; Chordata; Lepidosauria; Squamata; Lacertidae). The assembly contains two haplotypes with total lengths of 1,574.91 megabases and 1,437.04 megabases. Most of haplotype 1 (94.94%) is scaffolded into 20 chromosomal pseudomolecules, including the W and Z sex chromosomes, while most of haplotype 2 (97.63%) is scaffolded into 18 chromosomal pseudomolecules. The mitochondrial genome has also been assembled, with a length of 18.2 kilobases.

Keywords

Podarcis melisellensis, Dalmatian wall lizard, genome sequence, chromosomal, Squamata

This article is included in the Tree of Life gateway.

Open Peer Review

Approval Status ✓ ✓ ✓

	1	2	3
version 1 19 May 2025	✓ view	✓ view	✓ view

1. **Mark Wilkinson**, The Natural History Museum, London, UK
2. **Hong-Xin Xie**, University of Glasgow, Glasgow, UK
3. **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: Joana Meier (jm66@sanger.ac.uk)

Author roles: Feiner N: Investigation, Resources, Writing – Original Draft Preparation; **Uller T**: Resources, Writing – Original Draft Preparation; **Van Linden L**: 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). L.V.L. was supported by a PhD fellowship (ID: 1102623N) from the Research Foundation Flanders (FWO).
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, Van Linden L *et al.* **The genome sequence of the Dalmatian wall lizard, *Podarcis melisellensis* (Braun, 1877) [version 1; peer review: 3 approved]** Wellcome Open Research 2025, **10**:247 https://doi.org/10.12688/wellcomeopenres.24140.1

First published: 19 May 2025, **10**:247 https://doi.org/10.12688/wellcomeopenres.24140.1

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 melisellensis* (Braun, 1877) (NCBI:txid65482)

Background

The Dalmatian wall lizard *Podarcis melisellensis* (Braun, 1877; Figure 1) is a diurnal lacertid lizard with a distribution along the eastern coast of the Adriatic Sea and numerous islands from northern Italy to Albania (Böhme, 1986). The Dalmatian wall lizard is classified as Least Concerned (LC) by the IUCN (Bowles, 2024).

This species is one of 28 currently described species of the genus *Podarcis*. *P. melisellensis* is most closely related to the group of *P. ionicus*, *P. tauricus*, *P. gaigeae* and *P. milensis*, which are also native to the Balkan region. The divergence of *P. melisellensis* and these species is estimated to have occurred ca. 7 MYA (Yang *et al.*, 2021).

The Dalmatian wall lizard inhabits a variety of habitats, including rocks, cliffs, open woodland, meadows and Mediterranean shrubland. Individuals can be active all year, but reproduction is (Yang *et al.*, 2021) generally restricted to spring and early summer. *Podarcis melisellensis* is oviparous and Meiri *et al.* (2020) reports an average clutch size of 2.53. Lizards from island populations can vary substantially in body size and colouration (Böhme, 1986), the latter dominated by brown, grey and green colours, with a variety of dorsal patterns, including melanistic populations. This species is one of several *Podarcis* that exhibit ventral colour polymorphism with yellow, orange/red and white morphs (Huyghe *et al.*, 2007).

Figure 1. A) Female and **B**) male *Podarcis melisellensis* photographed on the island of Vis, Croatia. Note that the specimen selected for genome sequencing is from the same island, but not shown on the photographs. Photographs by Tim De Ridder (**A**) and Annelies Jacobs (**B**).

Podarcis melisellensis belongs to the Balkan group of the genus *Podarcis*, which consists of 10 currently described species. Within this group, reference genomes are publicly available for *P. cretensis* (Poulakakis *et al.*, 2024), *P. gaigeae*, and *P. erhardii*. The reference genome of *P. melisellensis* will be a valuable resource, both for comparative studies above the species level, as well as studies addressing the population genetic history and phenotypic differentiation within *P. melisellensis.*

Genome sequence report

Sequencing data

The genome of a specimen of *Podarcis melisellensis* was sequenced using Pacific Biosciences single-molecule HiFi long reads, generating 60.50 Gb (gigabases) from 5.68 million reads. GenomeScope analysis of the PacBio HiFi data estimated the haploid genome size at 1,435.94 Mb, with a heterozygosity of 0.57% and repeat content of 20.99%. 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 41.0x coverage of the genome. Chromosome conformation Hi-C sequencing produced 360.88 Gb from 2,389.96 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 80 misjoins or missing joins. These interventions decreased the scaffold count by 6.76%. The final assembly has a total length of 1,574.91 Mb in 385 scaffolds, with 823 gaps, and a scaffold N50 of 91.07 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 (94.94%) 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).

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

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

Project information			
Study title	Podarcis melisellensis (Dalmatian wall lizard)		
Umbrella BioProject	PRJEB76775		
Species	*Podarcis melisellensis*		
BioSpecimen	SAMEA115336775		
NCBI taxonomy ID	65482		
Specimen information			
Technology	**ToLID**	**BioSample accession**	**Organism part**
PacBio long read sequencing	rPodMel1	SAMEA115336785	terminal body
Hi-C sequencing	rPodMel1	SAMEA115336785	terminal body
Sequencing information			
Platform	**Run accession**	**Read count**	**Base count (Gb)**
Hi-C Illumina NovaSeq X	ERR13317838	1.16e+09	175.15
Hi-C Illumina NovaSeq X	ERR13317837	1.23e+09	185.73
PacBio Revio	ERR13304158	5.68e+06	60.5

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

Genome assembly	**Haplotype 1**	**Haplotype 2**
Assembly name	rPodMel1.hap1.1	rPodMel1.hap2.1
Assembly accession	GCA_964234715.1	GCA_964234825.1
Assembly level	chromosome	chromosome
Span (Mb)	1,574.91	1,437.04
Number of contigs	1,208	954
Number of scaffolds	385	200
Longest scaffold (Mb)	136.76	135.53
Assembly metrics* (benchmark)	**Haplotype 1**	**Haplotype 2**
Contig N50 length (≥ 1 Mb)	3.03 Mb	3.11 Mb
Scaffold N50 length (= chromosome N50)	91.07 Mb	92.09 Mb
Consensus quality (QV) (≥ 40)	62.3	62.5
k-mer completeness	88.41%	84.66%
Combined *k*-mer completeness (≥ 95%)	99.63%	
BUSCO** (S > 90%; D < 5%)	C:95.3%[S:93.3%,D:2.0%], F:0.8%,M:4.0%,n:7,480	C:92.1%[S:90.4%,D:1.7%], F:1.0%,M:6.9%,n:7,480
Percentage of assembly mapped to chromosomes (≥ 90%)	94.94%	97.63%
Sex chromosomes (localised homologous pairs)	W and Z	-
Organelles (one complete allele)	Mitochondrial genome: 18.2 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.

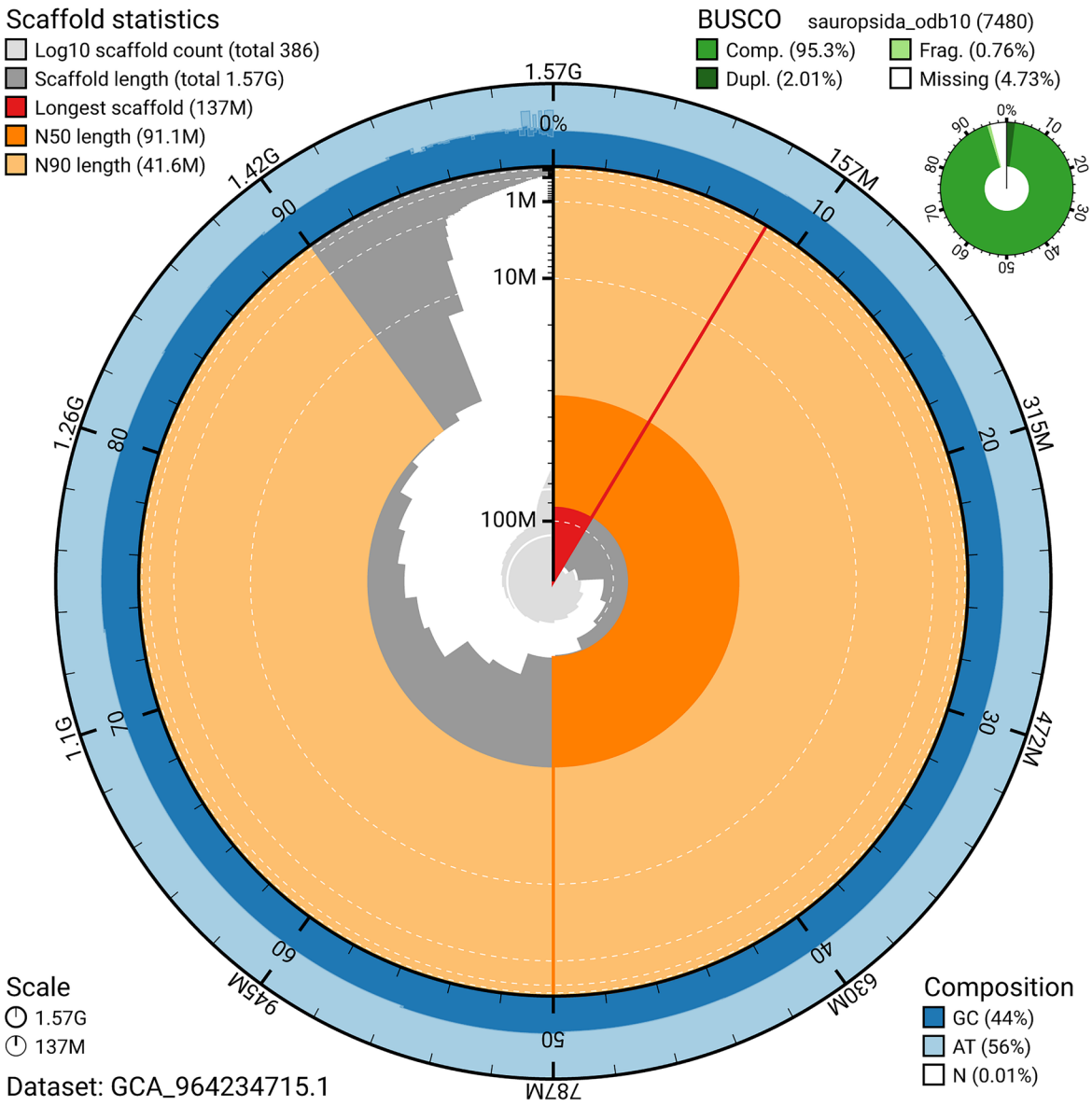

Figure 2. Genome assembly of *Podarcis melisellensis*, rPodMel1.hap1.1: metrics. The BlobToolKit snail plot provides an overview of assembly metrics and BUSCO gene completeness. The circumference represents the length of the whole genome sequence, and the main plot is divided into 1,000 bins around the 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_964234715.1/dataset/GCA_964234715.1/snail.

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 62.3, and for haplotype 2, 62.5. When the two haplotypes are combined, the assembly achieves an estimated QV of 62.4. The *k*-mer recovery for haplotype 1 is 88.41%, and for haplotype 2 84.66%, while the combined haplotypes have a *k*-mer recovery of 99.63%. BUSCO 5.5.0 analysis using the sauropsida_odb10

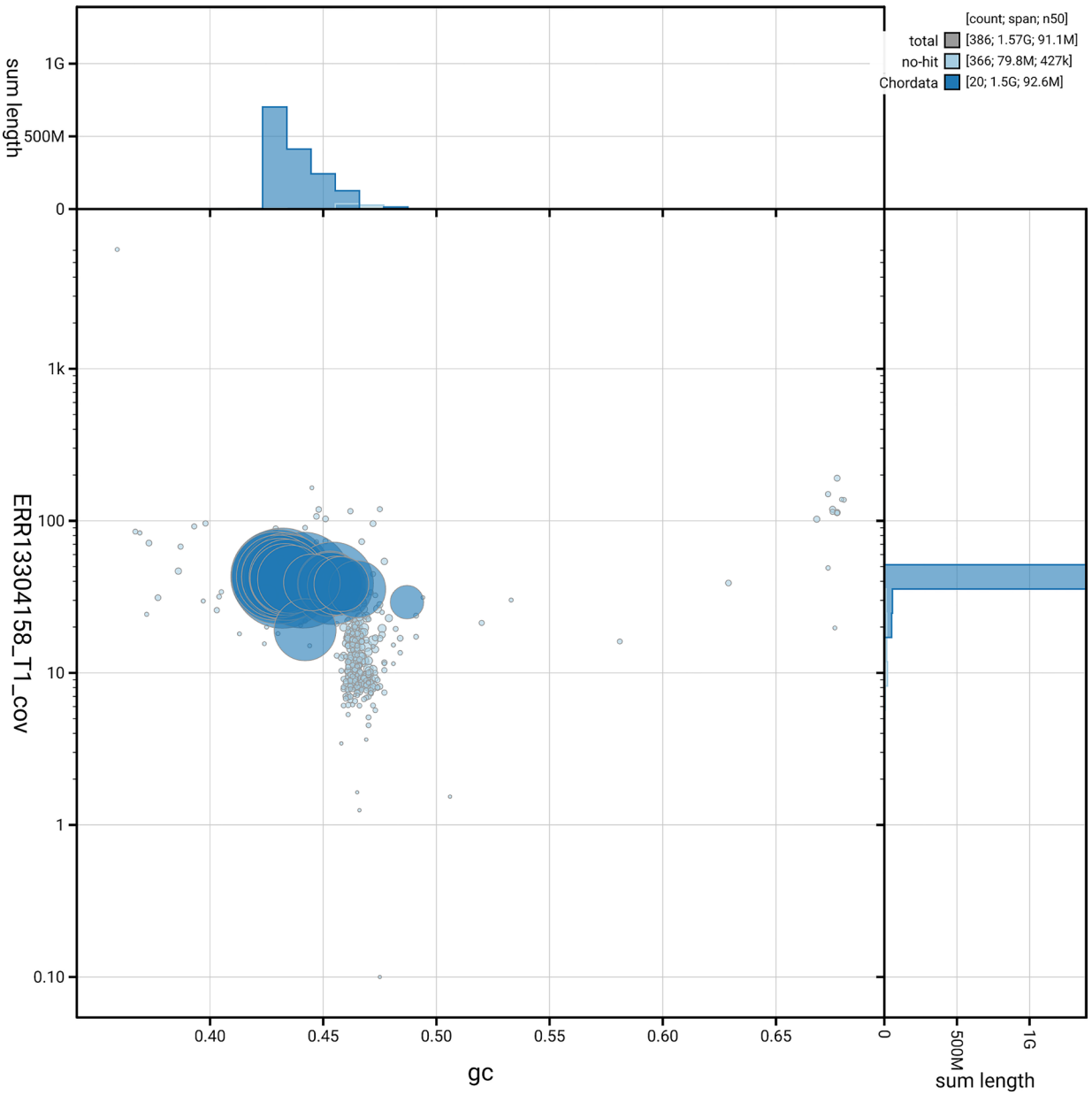

Figure 3. Genome assembly of *Podarcis melisellensis*, rPodMel1.hap1.1: BlobToolKit GC-coverage plot. Blob plot showing sequence coverage (vertical axis) and GC content (horizontal axis). The circles represent scaffolds, with the size proportional to scaffold length and the colour representing phylum 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_964234715.1/dataset/GCA_964234715.1/blob.

reference set (n = 7,480) identified 95.3% of the expected gene set (single = 93.3%, duplicated = 2.0%) for haplotype 1.

Table 2 provides assembly metric benchmarks adapted from Rhie *et al.* (2021) and the Earth BioGenome Project (EBP) Report on Assembly Standards September 2024. The assembly achieves the EBP reference standard of **6.C.Q62**.

Methods

Sample acquisition

The specimen, an adult female *P. melisellensis* lizard (specimen ID SAN25002445, ToLID rPodMel1) was collected 2023-05-01 from a site on the island of Vis (latitude: 43.048141; longitude: 16.081184). The specimen belongs to the nominal subspecies *P. m. melisellensis*. The specimen was caught by

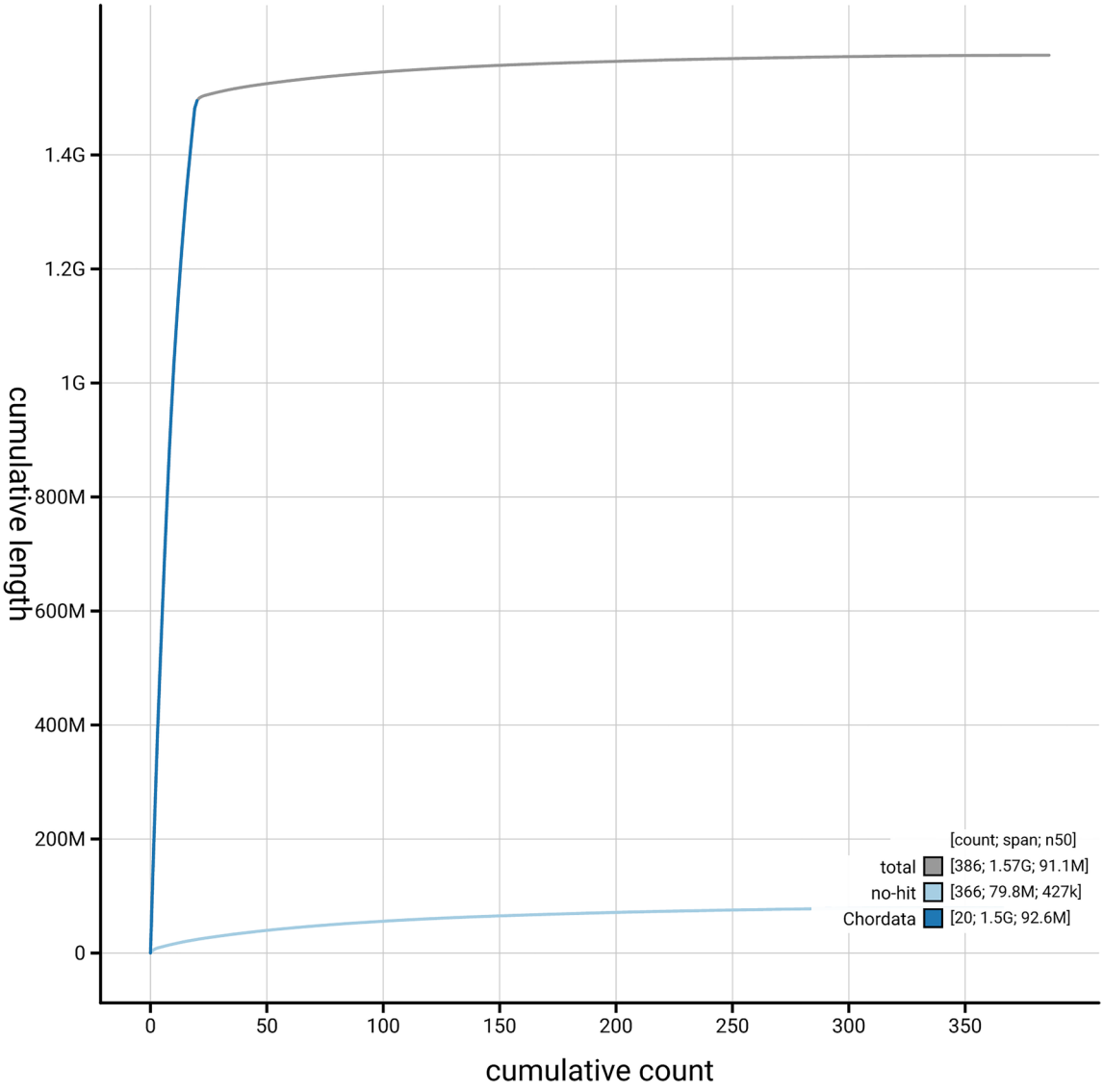

Figure 4. Genome assembly of *Podarcis melisellensis*, rPodMel1.hap1.1: BlobToolKit cumulative sequence plot. The grey line shows cumulative length for all scaffolds. Coloured lines show cumulative lengths of scaffolds assigned to each phylum using the buscogenes taxrule. An interactive version of this figure is available at https://blobtoolkit.genomehubs.org/view/GCA_964234715.1/dataset/GCA_964234715.1/cumulative.

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 517-10-1-2-23-4. The specimen was collected and identified by Lisa Van Linden (University of Antwerp, Belgium).

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 rPodMel1 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

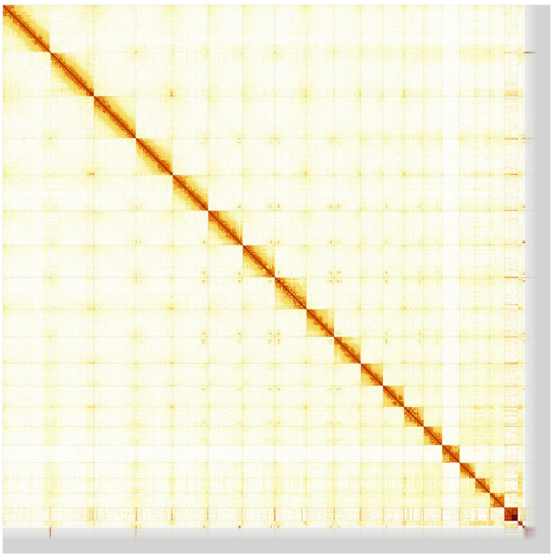

Figure 5. Genome assembly of *Podarcis melisellensis*, rPodMel1.hap1.1: Hi-C contact map of the rPodMel1.hap1.1 assembly, produced in PretextView. Chromosomes are shown in order of size from left to right and top to bottom.

Table 3. Chromosomal pseudomolecules in the genome assembly of *Podarcis melisellensis*, rPodMel1.

Haplotype 1				Haplotype 2			
INSDC accession	Name	Length (Mb)	GC%	INSDC accession	Name	Length (Mb)	GC%
OZ174275.1	1	136.76	43	OZ174246.1	1	135.53	43
OZ174276.1	2	125.27	44	OZ174247.1	2	123.94	44
OZ174277.1	3	120.99	43	OZ174248.1	3	119.86	43
OZ174278.1	4	105.59	43	OZ174249.1	4	106.02	43
OZ174279.1	5	102.05	43.5	OZ174250.1	5	103.24	43.5
OZ174280.1	6	99.26	43.5	OZ174251.1	6	99.69	43.5
OZ174281.1	7	92.56	43	OZ174252.1	7	92.09	43
OZ174282.1	8	91.07	45.5	OZ174253.1	8	90.46	45.5
OZ174283.1	9	79.6	43	OZ174254.1	9	79.3	43
OZ174284.1	10	76.59	43.5	OZ174255.1	10	77.4	43.5
OZ174285.1	11	63.5	43.5	OZ174256.1	11	63.51	43.5
OZ174286.1	12	61.17	43.5	OZ174257.1	12	61.11	43.5
OZ174287.1	13	56.24	45.5	OZ174258.1	13	56.11	45.5
OZ174288.1	14	53.09	45.5	OZ174259.1	14	52.75	45.5
OZ174290.1	15	45.01	45.5	OZ174260.1	15	44.86	45.5
OZ174291.1	16	42.19	46.5	OZ174261.1	16	42.62	46.5
OZ174292.1	17	41.62	44.5	OZ174262.1	17	41.35	44.5
OZ174294.1	18	13.43	48.5	OZ174263.1	18	13.21	48.5
OZ174293.1	W	38.85	46				
OZ174289.1	Z	50.28	44				
OZ174295.1	MT	0.02	36.5				

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 rPodMel1 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 frozen tissue (stored at –80 °C) was fixed, and the DNA crosslinked using a TC buffer with 22% formaldehyde concentration. 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 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 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 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.

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 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 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

Table 4. Software tools: versions and sources.

Software tool	Version	Source
BLAST	2.14.0	http://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	44086069ee7d4d3f6f3f0012569789ec138f42b8 4aa44357826c0b6753eb28de	https://github.com/higlass/higlass
MerquryFK	d00d98157618f4e8d1a9190026b19b471055b 22e	https://github.com/thegenemyers/MERQURY.FK
Minimap2	2.24-r1122	https://github.com/lh3/minimap2
MitoHiFi	3	https://github.com/marcelauliano/MitoHiFi

Software tool	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
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	0.5.1	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

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 melisellensis (Dalmatian wall lizard). Accession number PRJEB76775; https://identifiers.org/ena.embl/PRJEB76775. The genome sequence is released openly for reuse. The *Podarcis melisellensis* 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

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 (Vol. 2/II.).** Aula-Verlag, 1986.

Bowles P: **Podarcis melisellensis.** *The IUCN Red List of Threatened Species.* 2024; **2024**: e.T61549A137855479.
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

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

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

Huyghe K, Vanhooydonck B, Herrel A, *et al.*: **Morphology, performance, behavior and ecology of three color morphs in males of the lizard *Podarcis melisellensis*.** *Integr Comp Biol.* 2007; **47**(2): 211–20.
PubMed Abstract | Publisher 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

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

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

Poulakakis N, Lymberakis P, Stratakis M, *et al.*: **The genome sequence of the Cretan wall lizard, *Podarcis cretensis* (Wettstein, 1952) [version 1; peer review: 2 approved with reservations].** *Wellcome Open Res.* 2024; **9**: 161.
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, Cornwell 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 25 June 2025

https://doi.org/10.21956/wellcomeopenres.26631.r124747

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

The article is a thorough data note reporting the reference genome for *Podarcis melisellensis*. Authors have used long read and chromatin conformation capture sequencing to generate the genome sequence assembled at chromosome level. The methodology is detailed and exact wet lab and *in silico* protocols (including the software name) for everything reported in the Genome sequence report section can be traced in the methods without ambiguity. Methods and protocols are sufficiently detailed to allow the reproduction of this work. *Podarcis* spp. usually possess 38 chromosomes (2n), including the sex chromosomes. So, the number of chromosome pseudomolecules generated with conformation capture seq correctly adds up. The genome size is also around the value known for this genus. The assembly parameters and statistics are of sufficient quality (it achieves the general standard in recent reptile reference genome papers). *Podarcis* spp. is an interesting, predominantly mediterranean lacertid radiation from speciation, population genomic, systematic and comparative genomic points of view and this data note reports a reference genome that will be valuable in all such studies. The only change I would make is to change 'Least Concerned' in the Background section to Least Concern as that is the more standard usage. The usage in the paper, however, is nothing wrong and authors can ignore this if they so wish.

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: Herpetology, Phylogenomics, Speciation genomics, Systematics

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 23 June 2025

https://doi.org/10.21956/wellcomeopenres.26631.r124748

Hong-Xin Xie
University of Glasgow, Glasgow, Scotland, UK

The authors present new genome assemblies for a wall lizard *Podarcis melisellensis*. The genome is chromosome level, curated and quality assessed in a well established pipeline. The reference genome generated will provide a valuable resources for genetic study of the species and comparative genomic studies in wall lizards and other squamate reptiles as well. The only concern I have about the reference genome is the relatively "low" quality of the assembly (contig N50 ~3 mb and 94.94% of sequences assigned to chromosomes). These statistics are generally good, however, considering the study used 41.0x coverage of Pacbio HiFi reads (> 30x is generally considered good) and more than 100x Hi-C data for scaffolding, the assembly matrix seem not good enough. Recent genomes using HiFi +Hi-C can easily reach contigs spanning almost the whole chromosome and a scaffold assignment rate more than 95%. A relatively low quality of the assembly could be caused by species specific factors or the quality of the raw sequencing reads. The first one is less likely because other wall lizards using similar strategy can reach much higher assembly matrix (like Podarcis raffonei GCA_027172095.1). Adding a report about the raw HiFi read quality may help explain the assembly matrix. Using tail tip stored in ethanol may not yield ideal high molecular weight DNA but is totally understandable because this is non-lethal sampling. Secondly, from figure 5 the Hi-C interaction plot, there seems to be some mis-assemblies, like 9th to 11th chromosome that have clear boundaries within a chromosome. Maybe a slightly more detailed error correction report can help explain the result.

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, evolutionary genomic, population genomics

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 23 June 2025

https://doi.org/10.21956/wellcomeopenres.26631.r124755

Mark Wilkinson
The Natural History Museum, London, UK

This is a report of the genome sequence of a Dalmatian wall lizard, a widely distributed lacertid with many island populations, numerous named subspecies and is thus of interest for studies of speciation and local adaptation. The study is of high quality. I was pleased to see the precise locality and subspecific identity of the voucher included in the report. As a taxonomist I would recommend some very minor changes to how taxon names are presented. On first use in Abstract and Mani text I would use the full name "Podarcis melisellensis (Braun, 1877)" as in the title. It doesn't matter if the parenthesis in the authority part of the species name is immediately followed by another '(Dalmation wall lizard...). Secondly on first use in isolation in the text 'The species is one of....' I would add the authority for the genus name Podarcis Wagler, 1830.

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: Herpetology, Phylogenetics

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.