DATA NOTE

The genome sequence of the Tyrrhenian Wall Lizard, *Podarcis tiliguerta* (Gmelin, 1789) (Squamata: Lacertidae)

[version 1; peer review: 2 approved, 1 approved with reservations]

Nathalie Feiner[1,2], Tobias Uller[2], Daniele Salvi[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, Skåne County, Sweden
[3]University of L'Aquila, L'Aquila, Italy
[4]Tree of Life, Wellcome Sanger Institute, Hinxton, England, UK

v1 **First published:** 08 Sep 2025, **10**:486
https://doi.org/10.12688/wellcomeopenres.24853.1
Latest published: 08 Sep 2025, **10**:486
https://doi.org/10.12688/wellcomeopenres.24853.1

Abstract

We present a genome assembly from an individual female *Podarcis tiliguerta* (Tyrrhenian Wall Lizard; Chordata; Lepidosauria; Squamata; Lacertidae). The assembly contains two haplotypes with total lengths of 1 462.31 megabases and 1 394.94 megabases. Most of haplotype 1 (99.26%) is scaffolded into 20 chromosomal pseudomolecules, including the W and Z sex chromosomes. Most of haplotype 2 (99.2%) is scaffolded into 18 chromosomal pseudomolecules. The mitochondrial genome has also been assembled, with a length of 17.19 kilobases.

Keywords

Podarcis tiliguerta; Tyrrhenian 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 08 Sep 2025	✓ view	✓ view	? view

1. **Anthony Snead**, New York University, NY, USA
2. **Andreas Kautt**, Washington University in St. Louis, St. Louis, USA
3. **Raul Araya Donoso**, Yale University, New Haven, USA

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

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

Author roles: Feiner N: Resources, Writing – Original Draft Preparation, Writing – Review & Editing; **Uller T**: Investigation, Writing – Original Draft Preparation; **Salvi D**: 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). D.S. was supported by the Italian Ministry for Research (PRIN project 'Hybrind'-2017KLZ3MA).
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, Salvi D *et al.* **The genome sequence of the Tyrrhenian Wall Lizard, *Podarcis tiliguerta* (Gmelin, 1789) (Squamata: Lacertidae) [version 1; peer review: 2 approved, 1 approved with reservations]** Wellcome Open Research 2025, **10**:486 https://doi.org/10.12688/wellcomeopenres.24853.1

First published: 08 Sep 2025, **10**:486 https://doi.org/10.12688/wellcomeopenres.24853.1

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 tiliguerta* (Gmelin, 1789) (NCBI:txid65485)

Background

The Tyrrhenian wall lizard *Podarcis tiliguerta* (Gmelin, 1789; Figure 1) is a diurnal lacertid lizard endemic to Corsica, Sardinia and surrounding islets. It is common throughout the main islands and is only absent in agricultural areas and at high altitudes on Corsica, although it occurs at least up to 1800 m (Mertens, 1957). More than 10 subspecies have been described, reflecting the large phenotypic diversity found in this species (Böhme, 1986). *Podarcis tiliguerta* 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. tiliguerta* is the sister species to the group of *P. lilfordi* and *P. pityusensis* (Yang *et al.*, 2021), which are endemic to the Balearic islands. The divergence between *P. tiliguerta* and the Balearic species is estimated to have occurred over 10 MYA (Yang *et al.*, 2021). The lineages of *P. tiliguerta* that inhabit Sardinia and Corsica are highly divergent and further genetic differentiation exist within each of these lineages (Harris *et al.*, 2005; Rodríguez *et al.*, 2017; Salvi *et al.*, 2017; Senczuk *et al.*, 2019).

The species can be found in a broad variety of habitats, including Mediterranean shrubland, bushy vegetation, grassland, sandy or rocky shores as well as anthropogenic environments. Lizards can be seen active all year round, except at high altitude, but the main activity period is in spring and summer. *P. tiliguerta* is oviparous and reported to lay multiple clutches per year (Böhme, 1986); clutch size is uncertain.

Figure 1. Photograph of the *Podarcis tiliguerta* (rPodTil1) specimen used for genome sequencing.

Tyrrhenian wall lizards exhibit high diversity in coloration and body size and some island populations can differ substantially from populations on the mainland in these characters (Böhme, 1986). As in most species of *Podarcis*, the sexes differ in appearance with females being generally smaller and showing a pattern that is more dominated by longitudinal stripes.

P. tiliguerta is one of three species of the Western Islands group of *Podarcis*. A reference genome of the species *P. lilfordi* is currently available (Gomez-Garrido *et al.*, 2023), and a reference for the third species of this group, *P. pityusensis* has recently been generated (Meier *et al.*, 2025). Reference genomes for every species in this group will be a valuable resource for comparative studies above the species level. In addition, we anticipate that the reference genome of *P. tiliguerta* will support ongoing research on the evolutionary history of this species and help to resolve taxonomic uncertainty regarding the status of different lineages.

We present a chromosome-level genome sequence for *Podarcis tiliguerta*, the Tyrrhenian Wall Lizard. The assembly was produced using the Tree of Life pipeline from a specimen collected in Punta Falcone, Sardinia, Italy (Figure 1). This assembly was generated as part of the Darwin Tree of Life Project, which aims to generate high-quality reference genomes for all named eukaryotic species in Britain and Ireland to support research, conservation, and the sustainable use of biodiversity (Blaxter *et al.*, 2022).

Methods

Sample acquisition

The specimen, an adult female *P. tiliguerta* lizard (specimen ID SAN25001764, ToLID rPodTil1; Figure 1), was collected on September 10th 2019 from a site near Punta Falcone on northern Sardinia (latitude: 41.25152; longitude: 9.22911). The specimen was caught by noosing, standard morphometric measurements were taken and the tip of the tail (ca. 2 cm) was collected and preserved in ethanol. The specimen was released again at the site of capture. Field work was conducted under a permit from the Italian Ministry of Environment: PNM-2019-0008130 (MATTM, Ministero dell'Ambiente e della Tutela del Territorio e del Mare, Italy).

Nucleic acid extraction

Protocols for high molecular weight (HMW) DNA extraction developed at the Wellcome Sanger Institute (WSI) Tree of Life Core Laboratory are available on protocols.io (Howard *et al.*, 2025). The rPodTil1 sample was weighed and triaged to determine the appropriate extraction protocol. Tissue from the terminal body was homogenised by powermashing using a PowerMasher II tissue disruptor. HMW DNA was extracted using the Automated MagAttract v2 protocol. DNA was sheared into an average fragment size of 12–20 kb following the Megaruptor®3 for LI PacBio protocol. Sheared DNA was purified by manual SPRI (solid-phase reversible immobilisation). The concentration of the sheared and purified DNA was assessed using a Nanodrop spectrophotometer and Qubit Fluorometer

using the Qubit dsDNA High Sensitivity Assay kit. Fragment size distribution was evaluated by running the sample on the FemtoPulse system. For this sample, the final post-shearing DNA had a Qubit concentration of 21.0 ng/μL and a yield of 2 730.00 ng.

PacBio HiFi library preparation and sequencing

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

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

Hi-C

Sample preparation and crosslinking

The Hi-C sample was prepared from 20–50 mg of frozen tissue from the terminal body of the rPodTil1 sample using the Arima-HiC v2 kit (Arima Genomics). Following the manufacturer's instructions, tissue was fixed and DNA crosslinked using TC buffer to a final formaldehyde concentration of 2%. The tissue was homogenised using the Diagnocine Power Masher-II. Crosslinked DNA was digested with a restriction enzyme master mix, biotinylated, and ligated. Clean-up was performed with SPRISelect beads before library preparation. DNA concentration was measured with the Qubit Fluorometer (Thermo Fisher Scientific) and Qubit HS Assay Kit. The biotinylation percentage was estimated using the Arima-HiC v2 QC beads.

Hi-C library preparation and sequencing

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

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

Genome assembly

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

The HiFi reads were assembled using Hifiasm in Hi-C phasing mode (Cheng *et al.*, 2021; Cheng *et al.*, 2022), producing two haplotypes. Hi-C reads (Rao *et al.*, 2014) were mapped to the primary contigs using bwa-mem2 (Vasimuddin *et al.*, 2019). Contigs were further scaffolded with Hi-C data in YaHS (Zhou *et al.*, 2023), using the --break option for handling potential misassemblies. The scaffolded assemblies were evaluated using Gfastats (Formenti *et al.*, 2022), BUSCO (Manni *et al.*, 2021) and MERQURY.FK (Rhie *et al.*, 2020).

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

Assembly curation

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

Assembly quality assessment

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

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

Genome sequence report

Sequence data

PacBio sequencing of the *Podarcis tiliguerta* specimen generated 50.67 Gb (gigabases) from 8.30 million reads, which were used to assemble the genome. GenomeScope2.0 analysis estimated the haploid genome size at 1 405.36 Mb, with a heterozygosity of 1.92% and repeat content of 20.45% (Figure 2). These estimates guided expectations for the assembly. Based on the estimated genome size, the sequencing data provided approximately 15× coverage. Hi-C sequencing produced 221.39 Gb from 1 466.18 million reads, which were used to scaffold the assembly. Table 1 summarises the specimen and sequencing details.

Assembly statistics

The genome was assembled into two haplotypes using Hi-C phasing. Haplotype 1 was curated to chromosome level, while haplotype 2 was assembled to scaffold level. The final assembly has a total length of 1 462.31 Mb in 257 scaffolds, with 1 024 gaps, and a scaffold N50 of 89.16 Mb (Table 2).

Most of the assembly sequence (99.26%) 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 3; Table 3). The Z and W chromosomes were identified based on its single-copy status within a diploid assembly and synteny analysis with *Podarcis siculus* (GCA_964188175.1) (Feiner *et al.*, 2025). The exact order and orientation of the contigs are uncertain in the following regions: chromosome 6 (85 700–89 500 Kbp), chromosome 7 (75 200–1 200 kbp), chromosome 12 (8 200–11 700 kbp), chromosome 14 (9 500 –12 400 kbp).

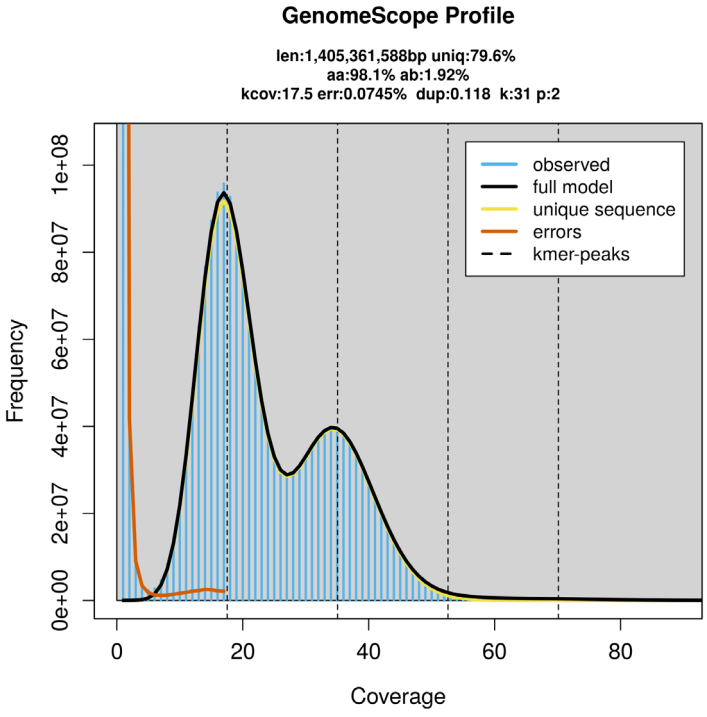

Figure 2. Frequency distribution of *k*-mers generated using GenomeScope2. The plot shows observed and modelled *k*-mer spectra, providing estimates of genome size, heterozygosity, and repeat content based on unassembled sequencing reads.

Table 1. **Specimen and sequencing data for BioProject PRJEB76520.**

Platform	PacBio HiFi	Hi-C
ToLID	rPodTil1	rPodTil1
Specimen ID	SAN25001764	SAN25001764
BioSample (source individual)	SAMEA114217799	SAMEA114217799
BioSample (tissue)	SAMEA115336781	SAMEA114217803
Tissue	terminal body	terminal body
Instrument	Revio	Illumina NovaSeq X
Run accessions	ERR13265034; ERR13265035	ERR13301067
Read count total	8.30 million	1 466.18 million
Base count total	50.67 Gb	221.39 Gb

Table 2. **Genome assembly statistics.**

Metric	Haplotype 1	Haplotype 2
Assembly name	rPodTil1.hap1.1	rPodTil1.hap2.2
Assembly accession	GCA_965153285.1	GCA_965153305.2
Assembly level	chromosome	chromosome
Span (Mb)	1 462.31	1 394.94
Number of chromosomes	20	18
Number of contigs	1 281	1 310
Contig N50	2.56 Mb	2.54 Mb
Number of scaffolds	257	307
Scaffold N50	89.16 Mb	90.89 Mb
Longest scaffold length (Mb)	135.08	135.23
Sex chromosomes	W and Z	N/A
Organelles	Mitochondrion: 17.19 kb	N/A

Chromosome 18 is notably small, measuring only 12,718 kbp in the assembled genome.

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.

For haplotype 1, the estimated QV is 64.6, and for haplotype 2, 64.6. When the two haplotypes are combined, the assembly achieves an estimated QV of 64.6. The *k*-mer completeness is 70.43% for haplotype 1, 67.43% for haplotype 2, and 99.77% for the combined haplotypes (Figure 4).

BUSCO analysis using the sauropsida_odb10 reference set (n = 7 480) identified 94.9% of the expected gene set (single = 93.1%, duplicated = 1.8%) for haplotype 1. The snail plot in Figure 5 summarises the scaffold length distribution and other assembly statistics for haplotype 1. The blob plot in Figure 6 shows the distribution of scaffolds by GC proportion and coverage for haplotype 1.

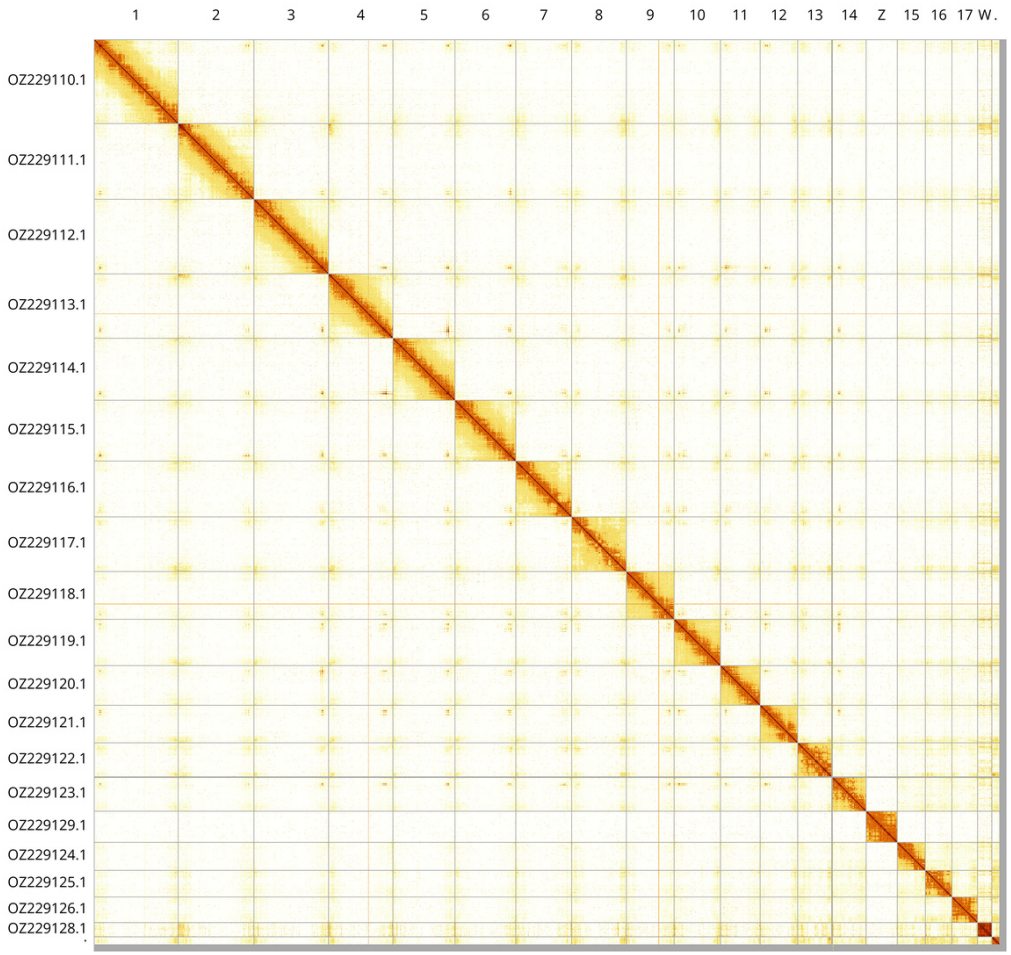

Figure 3. Hi-C contact map of the *Podarcis tiliguerta* genome assembly. Assembled chromosomes are shown in order of size and labelled along the axes. The plot was generated using PretextSnapshot.

Table 3. Chromosomal pseudomolecules in both haplotypes of the genome assembly of *Calluna vulgaris*, ddCalVulg4.

Haplotype 1				Haplotype 2			
INSDC accession	Name	Length (Mb)	GC%	INSDC accession	Name	Length (Mb)	GC%
OZ229110.1	1	135.08	43.50	OZ229093.1	1	135.23	43.50
OZ229111.1	2	121.46	44	OZ229094.1	2	124.15	44
OZ229112.1	3	119.35	43	OZ229095.1	3	120.05	43
OZ229113.1	4	103.49	43	OZ229096.1	4	103.17	43
OZ229114.1	5	99.17	43.50	OZ229097.1	5	98.92	43.50
OZ229115.1	6	97.79	43.50	OZ229098.1	6	98.44	43.50

Haplotype 1				Haplotype 2			
INSDC accession	Name	Length (Mb)	GC%	INSDC accession	Name	Length (Mb)	GC%
OZ229116.1	7	89.16	43	OZ229099.1	7	91.04	43
OZ229117.1	8	87.59	45.50	OZ229100.1	8	87.33	46
OZ229118.1	9	76.80	43.50	OZ229101.1	9	77.10	43.50
OZ229119.1	10	73.99	43.50	OZ229102.1	10	74.60	43.50
OZ229120.1	11	63.69	43.50	OZ229103.1	11	63.77	43.50
OZ229121.1	12	60.19	44	OZ229104.1	12	61.03	43.50
OZ229122.1	13	55.57	45.50	OZ229105.1	13	55.59	45.50
OZ229123.1	14	54.02	45.50	OZ229106.1	14	52.88	45.50
OZ229124.1	15	44.89	46	OZ229107.1	15	44.86	46
OZ229125.1	16	42.81	47	OZ229108.1	16	42.25	47
OZ229126.1	17	41.15	45	OZ229109.1	17	40.65	44.50
OZ229127.1	18	12.75	49	OZ271920.1	18	12.74	49.50
OZ229128.1	W	22.58	45.50				
OZ229129.1	Z	50.01	44.50				

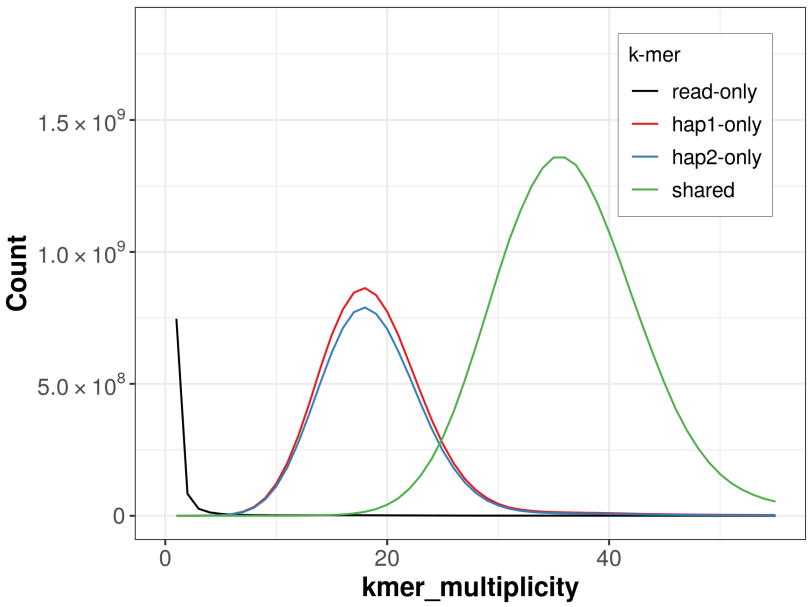

Figure 4. Evaluation of *k*-mer completeness using MerquryFK. This plot illustrates the recovery of *k*-mers from the original read data in the final assemblies. The horizontal axis represents *k*-mer multiplicity, and the vertical axis shows the number of *k*-mers. The black curve represents *k*-mers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

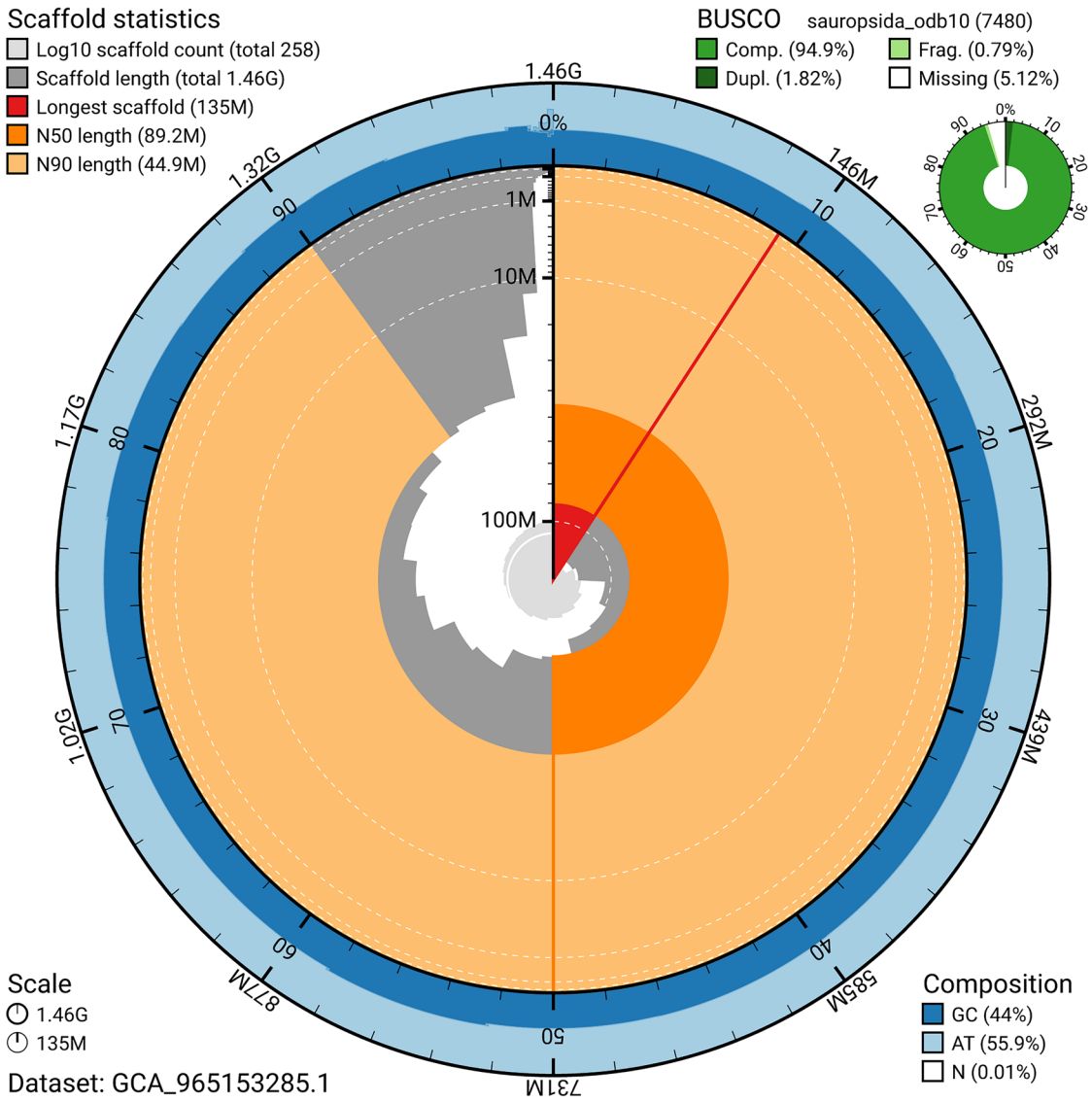

Figure 5. Assembly metrics for rPodTil1.hap1.1. The BlobToolKit snail plot provides an overview of assembly metrics and BUSCO gene completeness. The circumference represents the length of the whole genome sequence, and the main plot is divided into 1 000 bins around the circumference. The outermost blue tracks display the distribution of GC, AT, and N percentages across the bins. Scaffolds are arranged clockwise from longest to shortest and are depicted in dark grey. The longest scaffold is indicated by the red arc, and the deeper orange and pale orange arcs represent the N50 and N90 lengths. A light grey spiral at the centre shows the cumulative scaffold count on a logarithmic scale. A summary of complete, fragmented, duplicated, and missing BUSCO genes in the set is presented at the top right. An interactive version of this figure can be accessed on the BlobToolKit viewer.

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

Wellcome Sanger Institute – Legal and Governance

The materials that have contributed to this genome note have been supplied by a Tree of Life collaborator. The Wellcome Sanger Institute employs a process whereby due diligence is carried out proportionate to the nature of the materials

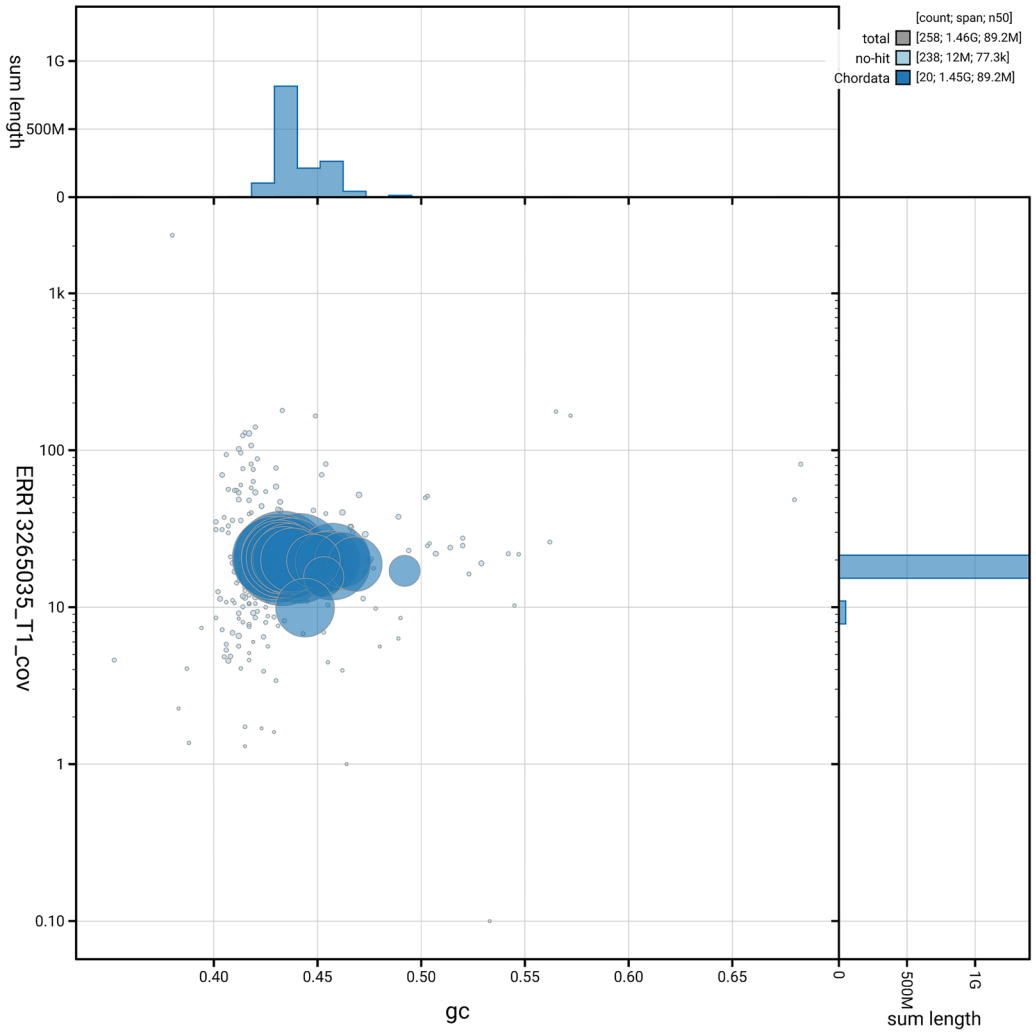

Figure 6. BlobToolKit GC-coverage plot for rPodTil1.hap1.1. Blob plot showing sequence coverage (vertical axis) and GC content (horizontal axis). The circles represent scaffolds, with the size proportional to scaffold length and the colour representing phylum membership. The histograms along the axes display the total length of sequences distributed across different levels of coverage and GC content. An interactive version of this figure is available on the BlobToolKit viewer.

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

Measure	Value	Benchmark
EBP summary (haplotype 1)	6.C.Q64	6.C.Q40
Contig N50 length	2.56 Mb	≥ 1 Mb
Scaffold N50 length	89.16 Mb	= chromosome N50
Consensus quality (QV)	Haplotype 1: 64.6; haplotype 2: 64.6; combined: 64.6	≥ 40
k-mer completeness	Haplotype 1: 70.43%; Haplotype 2: 67.43%; combined: 99.77%	≥ 95%
BUSCO	C:94.9% [S:93.1%; D:1.8%]; F:0.8%; M:4.3%; n:7 480	S > 90%; D < 5%
Percentage of assembly assigned to chromosomes	99.26%	≥ 90%

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 tiliguerta (Tyrrhenian wall lizard). Accession number PRJEB76520. The genome sequence is released openly for reuse. The *Podarcis tiliguerta* genome sequencing initiative is part of the Sanger Institute Tree of Life Programme (PRJEB43745) and the Vertebrate Genomes Project (PRJNA489243). All raw sequence data and the assembly have been deposited in INSDC databases. The genome will be annotated using available RNA-Seq data and presented through the Ensembl pipeline at the European Bioinformatics Institute. Raw data and assembly accession identifiers are reported in Table 1 and Table 2.

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

Author information

Contributors are listed at the following links:

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

Table 5. Software versions and sources.

Software	Version	Source
BEDTools	2.30.0	https://github.com/arq5x/bedtools2
BLAST	2.14.0	ftp://ftp.ncbi.nlm.nih.gov/blast/executables/blast+/
BlobToolKit	4.3.9	https://github.com/blobtoolkit/blobtoolkit
BUSCO	5.5.0	https://gitlab.com/ezlab/busco
bwa-mem2	2.2.1	https://github.com/bwa-mem2/bwa-mem2
Cooler	0.8.11	https://github.com/open2c/cooler
DIAMOND	2.1.8	https://github.com/bbuchfink/diamond
fasta_windows	0.2.4	https://github.com/tolkit/fasta_windows
FastK	1.1	https://github.com/thegenemyers/FASTK
GenomeScope2.0	2.0.1	https://github.com/tbenavi1/genomescope2.0
Gfastats	1.3.6	https://github.com/vgl-hub/gfastats
GoaT CLI	0.2.5	https://github.com/genomehubs/goat-cli
Hifiasm	0.19.8-r603	https://github.com/chhylp123/hifiasm
HiGlass	1.13.4	https://github.com/higlass/higlass
MerquryFK	1.1.2	https://github.com/thegenemyers/MERQURY.FK
Minimap2	2.24-r1122	https://github.com/lh3/minimap2
MitoHiFi	3	https://github.com/marcelauliano/MitoHiFi
MultiQC	1.14; 1.17 and 1.18	https://github.com/MultiQC/MultiQC

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

References

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

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

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

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

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

Bowles P: **Podarcis tiliguerta.** The IUCN Red List of Threatened Species 2024: e.T157293A137862171, 2024.
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

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

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

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

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

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

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

Feiner N, Uller T, Salvi D, *et al.*: **The genome sequence of the italian wall lizard, *podarcis siculus* (rafinesque-schmaltz, 1810) [version 1; peer review: 1 approved with reservations].** *Wellcome Open Res.* 2025; **10**: 255.
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

Gomez-Garrido J, Cruz F, Alioto TS, *et al.*: **Chromosome-level genome assembly of lilford's wall lizard, *Podarcis lilfordi* (Günther, 1874) from the Balearic Islands (Spain).** *DNA Res.* 2023; **30**(3): dsad008.
PubMed Abstract | Publisher Full Text | Free Full Text

Harris D, Pinho C, Carretero M, *et al.*: **Determination of genetic diversity within the insular lizard *Podarcis tiliguerta* using mtDNA sequence data, with a reassessment of the phylogeny of *Podarcis*.** *Amphibia-Reptilia.* 2005; **26**: 401–7.
Publisher Full Text

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

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

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

Meier J, Feiner N, Uller T, *et al.*: **The genome sequence of the ibiza wall lizard, *podarcis pityusensis* (boscá, 1883) [version 1; peer review: 2 approved, 1 approved with reservations].** *Wellcome Open Research.* 2025; **10**: 235.
PubMed Abstract | Publisher Full Text | Free Full Text

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

Mertens R: **Unterlagen zu einer "Herpetologia Tyrrhenia" V. Die Amphibien und Reptilien Korsikas.** *Senckenbergiana Biologica.* 1957; **38**(3–4).

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**

Rodríguez V, Buades JM, Brown RP, *et al.*: **Evolutionary history of *Podarcis tiliguerta* on Corsica and Sardinia.** *BMC Evol Biol.* 2017; **17**(1): 27.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

Salvi D, Pinho C, Harris DJ: **Digging up the roots of an insular hotspot of genetic diversity: decoupled mito-nuclear histories in the evolution of the Corsican-Sardinian endemic lizard *Podarcis tiliguerta.*** *BMC Evol Biol.* 2017; **17**(1): 63.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

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

Senczuk G, Castiglia R, Colangelo P, *et al.*: **The role of island physiography in maintaining genetic diversity in the endemic Tyrrhenian wall lizard (*Podarcis tiliguerta*).** *J Zool.* 2019; **309**(2): 140–51.
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 13 October 2025

https://doi.org/10.21956/wellcomeopenres.27377.r133864

?

Raul Araya Donoso
Yale University, New Haven, USA

This is a review of a Data Note submitted to Welcome Open Research by Feiner et al. The authors report the sequencing and assembly of a phased reference genome for the Tyrrhenian Wall Lizard, Podarcis tiliguerta.

Overall, the note is well written, the methods are adequate for their aims, and the generation of reference genomes for this group of lizards is a big contribution to understanding their evolutionary history.

I have only minor comments to improve the general delivery of the results.

Background
I want to acknowledge the authors for a well-detailed description of the species and its evolutionary context. This really helps the reader to understand the relevance of the generation of this dataset, which is challenging for “Data Notes”.

Methods
In the Genome Assembly section. The assembly evaluation methods can be removed to avoid redundancy. Since the section that comes after (“Assembly quality assessment”) describes these methods in detail.

The method used for the identification and naming of chromosomes methods (size and synteny to Podarcis siculus) should be moved to this section.

Genome Sequence Report
Are you currently able to identify which chromosomes are missing from haplotype 2 and present in haplotype 1? If the authors can easily identify it, it would be useful information.

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

Yes

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

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

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

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Evolutionary genomics

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however I have significant reservations, as outlined above.

Reviewer Report 09 October 2025

https://doi.org/10.21956/wellcomeopenres.27377.r133869

Andreas Kautt
Washington University in St. Louis, St. Louis, USA

The authors generated a high-quality haplotype-phased genome assembly of the Tyrrhenian Wall Lizard endemic to a few islands in the Mediterranean Sea. The assembly was generated using state-of-the-art methods, is characterized by excellent quality metrics, and will provide an important resource for evolutionary genomic and phylogeographic studies. Overall, this is very nice work and I only have two related comments concerning the small size of chr18 and two clarification requests for the authors to consider.

Main comments:
- Do the authors have any insights into why chr18 is so small? Have they examined synteny with other available Podarcis assemblies? Did the majority of chr18 end up in haplotype2? It might be worthwhile to investigate this further, as something appears to be missing in haplotype1.

- Related to the previous point, a BUSCO score of ~95%, while good overall, seems somewhat low for an assembly of this quality and raises the question of whether it is genes on (the full) chr18 that are predominantly the ones that are missing. One way to check this would be to assess whether the BUSCO score of the "primary" assembly (i.e., non-haplotype-resolved hifiasm output) is considerably better than hap1. Another approach would be to check if the missing genes are

syntenic in other closely related species, which would support the hypothesis that a chromosome segment is missing from the assembly rather than genes scattered throughout.

Minor comments:
- The reference to the Darwin Tree of Life Project seems a bit odd given that this species is endemic to Corsica and Sicily, while the Darwin Tree of Life Project "aims to generate high-quality reference genomes for all named eukaryotic species in Britain and Ireland." I leave it to the authors to determine if it is appropriate to acknowledge the Darwin Tree of Life Project, but maybe they can briefly explain how this assembly fits its scope.

- I assume hifiasm was run with default settings. Was the expected haploid genome size specified? Please specify.

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

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

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

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

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Evolutionary genomics

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 07 October 2025

https://doi.org/10.21956/wellcomeopenres.27377.r133867

Anthony Snead
New York University, NY, NY, USA

This work by Feiner and colleagues critical provide a critical genomic resource for understanding macroevolutionary dynamics across wall lizards, while opening the door to population genomic studies using the Tyrrhenian Wall Lizard (*Podarcis tiliguerta*). They present a chromosome-level haplotype-phased reference grade genome assembly for *Podarcis tiliguerta* with 99.26% of the

assembly sequence assigned to 20 chromosomal-level scaffolds representing the 18 autosomes and two sex chromosomes in haplotype 1. The data and assembly quality checks are robust and this promising to be a critical resource for future evolutionary genomic studies. There are only a couple of clarifying comments on specific lines that I have for the authors.

1. The authors state they collected standard morphometric measurements of the specimen. Standard morphometric measurements could differ between labs. I assumed weight and snout to vent length; however, researchers often use calipers to measure limb lengths head width or other measurements. I think clarifying that point and pointing to a table or supplementary file with that data could aid researchers who may want to use that data in the future.
2. The authors mention that the orientation of some contigs are uncertain. I would assume this is due to the manual curation processes, but I think it would help readers to briefly mention why the exact order and orientation of these areas are uncertain.

Overall, I find this to be a robust data resource, and I am excited to see this genomic resource released.

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: Population genomics, comparative genomics, 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.