Check for updates

DATA NOTE

The genome sequence of the Skyros Wall Lizard, *Podarcis gaigeae* (Werner, 1930)

[version 1; peer review: awaiting peer review]

Nathalie Feiner[1,2], Tobias Uller[2], Javier Abalos[2,3], Kostas Sagonas[4], Panayiotis Pafilis[5,6], Joana Meier[7],
Wellcome Sanger Institute Tree of Life Management, Samples and Laboratory team,
Wellcome Sanger Institute Scientific Operations: Sequencing Operations,
Wellcome Sanger Institute Tree of Life Core Informatics team,
Tree of Life Core Informatics collective

[1]Max Planck Institute for Evolutionary Biology, Plön, Germany
[2]Department of Biology, Lund University, Lund, Sweden
[3]Ethology Lab, Cavanilles Institute of Biodiversity and Evolutionary Biology, University of Valencia, Valencia, Spain
[4]School of Biology, Aristotle University of Thessaloniki, Thessaloniki, Greece
[5]Department of Biology, National and Kapodistrian University of Athens, Athens, Greece
[6]Museum of Zoology, National and Kapodistrian University of Athens, Athens, Greece
[7]Tree of Life, Wellcome Sanger Institute, Hinxton, England, UK

v1 **First published:** 07 May 2025, **10**:234
https://doi.org/10.12688/wellcomeopenres.24139.1
Latest published: 07 May 2025, **10**:234
https://doi.org/10.12688/wellcomeopenres.24139.1

Abstract

We present a genome assembly from a female specimen of *Podarcis gaigeae* (Skyros Wall Lizard; Chordata; Lepidosauria; Squamata; Lacertidae). The assembly contains two haplotypes with total lengths of 1,514.62 megabases and 1,422.31 megabases. Most of haplotype 1 (99.33%) is scaffolded into 20 chromosomal pseudomolecules, including the W and Z sex chromosomes. Most of haplotype 2 (99.58%) is scaffolded into 18 chromosomal pseudomolecules. The mitochondrial genome has also been assembled, with a length of 20.31 kilobases.

Keywords
Podarcis gaigeae, Skyros Wall Lizard, genome sequence, chromosomal, Squamata

Open Peer Review

Approval Status *AWAITING PEER REVIEW*

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

This article is included in the Tree of Life gateway.

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; **Sagonas K**: Resources, Writing – Review & Editing; **Pafilis P**: Resources, Writing – Review & Editing; **Meier J**: Investigation, Resources, Writing – Review & Editing;

Competing interests: No competing interests were disclosed.

Grant information: This work was supported by Wellcome through core funding to the Wellcome Sanger Institute (220540). N.F. was funded by Starting Grants from the European Research Council (948126) and the Swedish Research Council (2020-03650). J.A. was funded by a Margarita Salas contract from the Spanish Ministry of Science (MS21-053). K.S. was supported by the Hellenic Foundation for Research and Innovation (H.F.R.I.) under the “2nd Call for H.F.R.I. Research Projects to support PostDoctoral Researchers” (Project Number: 00167).
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 Skyros Wall Lizard, *Podarcis gaigeae* (Werner, 1930) [version 1; peer review: awaiting peer review]** Wellcome Open Research 2025, **10**:234 https://doi.org/10.12688/wellcomeopenres.24139.1

First published: 07 May 2025, **10**:234 https://doi.org/10.12688/wellcomeopenres.24139.1

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 gaigeae* (Werner, 1930) (NCBI:txid90521)

Background

The Skyros wall lizard *Podarcis gaigeae* (Werner, 1930), named after the American herpetologist Helen Thompson Gaige, is a diurnal lacertid lizard endemic to the Skyros archipelago in the Aegean Sea, Greece. It is broadly distributed on the main island of Skyros, all its surrounding islets, and on Piperi, an island ca. 50 km north of Skyros. The Piperi population is classified as subspecies *P. g. weigandi*. The Skyros wall lizard is listed in Annex III of the Berne Convention. It is classified as a Vulnerable Species (VU) by the IUCN (Bowles, 2024) due to the limited distribution and the possibility of introduction of predators in the distribution areas, and it is referred to as Least Concern (LC) by the Red Book of Endangered Animals in Greece (Legakis & Maragos, 2009)

This species is one of 28 currently described species of the genus *Podarcis*. *Podarcis gaigeae* is the sister species to *Podarcis milensis* (Poulakakis *et al.*, 2005; Yang *et al.*, 2021), which is found on the southwestern-most island in the Cyclades island group. Thus, *P. gaigeae* is not most closely related to the broadly distributed species *P. erhardii*, which is in closer geographic proximity to *P. gaigeae* relative to *P. milensis*. This suggests that the evolutionary history that gave rise to the *Podarcis* lineages inhabiting the islands of the Aegean Sea today involves ancient range expansion and contraction of different lineages, in addition to the isolation caused by rising sea levels.

The species is found in a variety of habitats, including Mediterranean shrubland, grassland and sandy or rocky shore areas. Individuals can be active all year, but reproduction is generally restricted to spring and early summer. *P. gaigeae* is oviparous and females lay between one and five eggs per clutch (Pafilis *et al.*, 2011), and likely multiple clutches per year (Valakos *et al.*, 2008). Meiri *et al.* (2020) provide an average clutch size of 2.47.

Skyros wall lizards show a wide variety of colours, colour patterns and body sizes. For example, males on the islet Exo Diavates reach body sizes of up to 90 mm snout-to-vent length (svl; mean: 85 mm), while main island males reach a maximum of 73 mm svl (mean: 61 mm). Also, females are on average 14 mm larger on Exo Diavates (mean: 70 mm) compared to the main island (mean: 56 mm), illustrating the large variation in body size (Lymberakis *et al.*, 2018; Pafilis *et al.*, 2011). The dorsum is mostly showing a reticulated pattern in brown, green and sometimes yellow hues, often with a green-to-brown gradient from head to tail. A complete lack of patterning ('concolor') is observed in several populations and can be locally abundant. 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. Ventral colours are variable, and yellow, orange and white morphs, and also mixtures of two of these colours, have been described (Runemark *et al.*, 2010).

P. gaigeae is the second species in the Balkan group of *Podarcis* (currently consisting of 10 species) after *P. cretensis* (Poulakakis *et al.*, 2024) for which a reference genome is constructed. This reference genome will be a valuable resource, both for studies within the species *P. gaigeae* targeting conservation and population genomics and for comparative studies above the species level.

Genome sequence report

Sequencing data

The genome of a specimen of *Podarcis gaigeae* (Figure 1) was sequenced using Pacific Biosciences single-molecule HiFi long reads, generating 64.62 Gb (gigabases) from 7.97 million reads. GenomeScope analysis of the PacBio HiFi data estimated the haploid genome size at 1,432.67 Mb, with a heterozygosity of 0.39% and repeat content of 19.80%. 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 44.0x coverage of the genome. Chromosome conformation Hi-C sequencing produced 192.16 Gb from 1,272.60 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

Figure 1. **A**) Photographs of a male and **B**) a female *P. gaigeae* from Molos islet, Skyros, Greece. **C**) Photograph of a male (left) and a female (right) from Exo Diavates. Photographs by Javier Abalos (**A** and **B**) and Panayiotis Pafilis (**C**).

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

Project information			
Study title	Podarcis gaigeae (Skyros wall lizard)		
Umbrella BioProject	PRJEB73729		
Species	*Podarcis gaigeae*		
BioSpecimen	SAMEA114217796		
NCBI taxonomy ID	90521		
Specimen information			
Technology	**ToLID**	**BioSample accession**	**Organism part**
PacBio long read sequencing	rPodGai1	SAMEA114217800	terminal body
Hi-C sequencing	rPodGai1	SAMEA114217800	terminal body
Sequencing information			
Platform	**Run accession**	**Read count**	**Base count (Gb)**
Hi-C Illumina NovaSeq X	ERR12743669	1.27e+09	192.16
PacBio Revio	ERR12736923	7.97e+06	64.62

assembly was improved by manual curation, which corrected 135 misjoins or missing joins and removed 1 haplotypic duplications. These interventions decreased the scaffold count by 20.68%. The final assembly has a total length of 1,514.62 Mb in 210 scaffolds, with 883 gaps, and a scaffold N50 of 91.93 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 (99.33%) was assigned to 20 chromosomal-level scaffolds, representing 19 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 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 64.9, and for haplotype 2, 64.6. When the two haplotypes are combined, the assembly achieves an estimated QV of 64.7. The *k*-mer recovery for haplotype 1 is 91.72%, and for haplotype 2 87.81%, while the combined haplotypes have a *k*-mer recovery of 99.72%. BUSCO v.5.5.0 analysis using the saurospida_odb10 reference set (n = 7,480) identified 95.2% of the expected gene set (single = 93.3%, duplicated = 1.9%) 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.Q64**.

Methods

Sample acquisition

The specimen, an adult female *Podarcis gaigeae* lizard (specimen ID SAN25001761, ToLID rPodGai1) was collected on 2022-05-18 from a grassland habitat at sea level close to the chapel Agios Fokas on the North-West coast of the main island of Skyros (latitude: 38.875748; longitude: 24.477255). We selected a female since this is the heterogametic sex in *Podarcis*. 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

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

Genome assembly	**Haplotype 1**	**Haplotype 2**
Assembly name	rPodGai1.hap1.2	rPodGai1.hap2.1
Assembly accession	GCA_964106915.2	GCA_964106785.1
Assembly level	chromosome	chromosome
Span (Mb)	1,514.62	1,422.31
Number of contigs	1,093	910
Number of scaffolds	210	127
Longest scaffold (Mb)	138.85	137.19
Assembly metrics (benchmark)	**Haplotype 1**	**Haplotype 2**
Contig N50 length (≥ 1 Mb)	2.86 Mb	3.1 Mb
Scaffold N50 length (= chromosome N50)	91.93 Mb	91.28 Mb
Consensus quality (QV) (≥ 40)	64.9	64.6
k-mer completeness	91.72%	87.81%
Completed *k*-mer completeness (≥ 95%)	99.72%	
BUSCO* (S > 90%; D < 5%)	C:95.2%[S:93.3%,D:1.9%], F:0.8%,M:4.0%,n:7,480	C:92.2%[S:90.6%,D:1.6%], F:1.0%,M:6.9%,n:7,480
Percentage of assembly mapped to chromosomes (≥ 90%)	99.33%	99.58%
Sex chromosomes (localised homologous pairs)	W and Z	-
Organelles (one complete allele)	Mitochondrial genome: 20.31 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.

specimen was released again at the site of capture. Field work was conducted under the permit ID ΥΠΕΝ/ΔΔΔ/16809/663.

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 rPodGai1 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

Hi-C data were generated from the terminal body of the rPodGai1 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 a final formaldehyde concentration of 2%. The tissue was then homogenised using the Diagnocine Power Masher-II. The crosslinked DNA

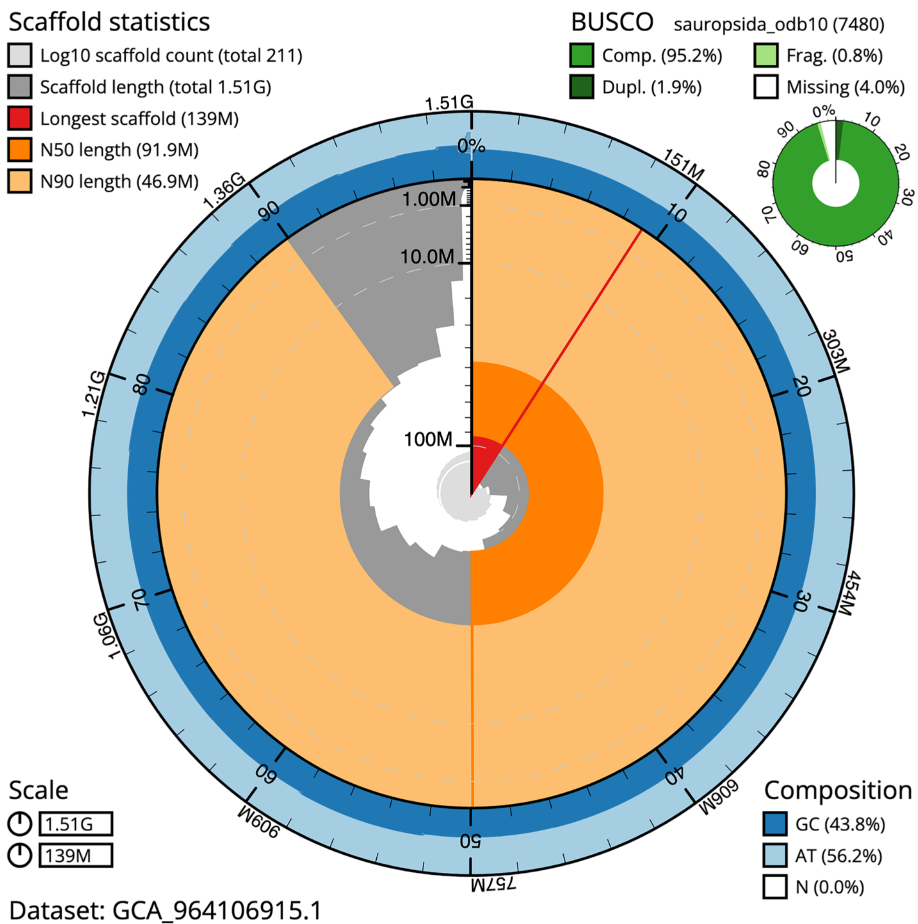

Figure 2. Genome assembly of *Podarcis gaigeae*, rPodGai1.hap1.2: metrics. The BlobToolKit snail plot provides an overview of assembly metrics and BUSCO gene completeness. The circumference represents the length of the whole genome sequence, and the main plot is divided into 1,000 bins around the circumference. 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.

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, 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.

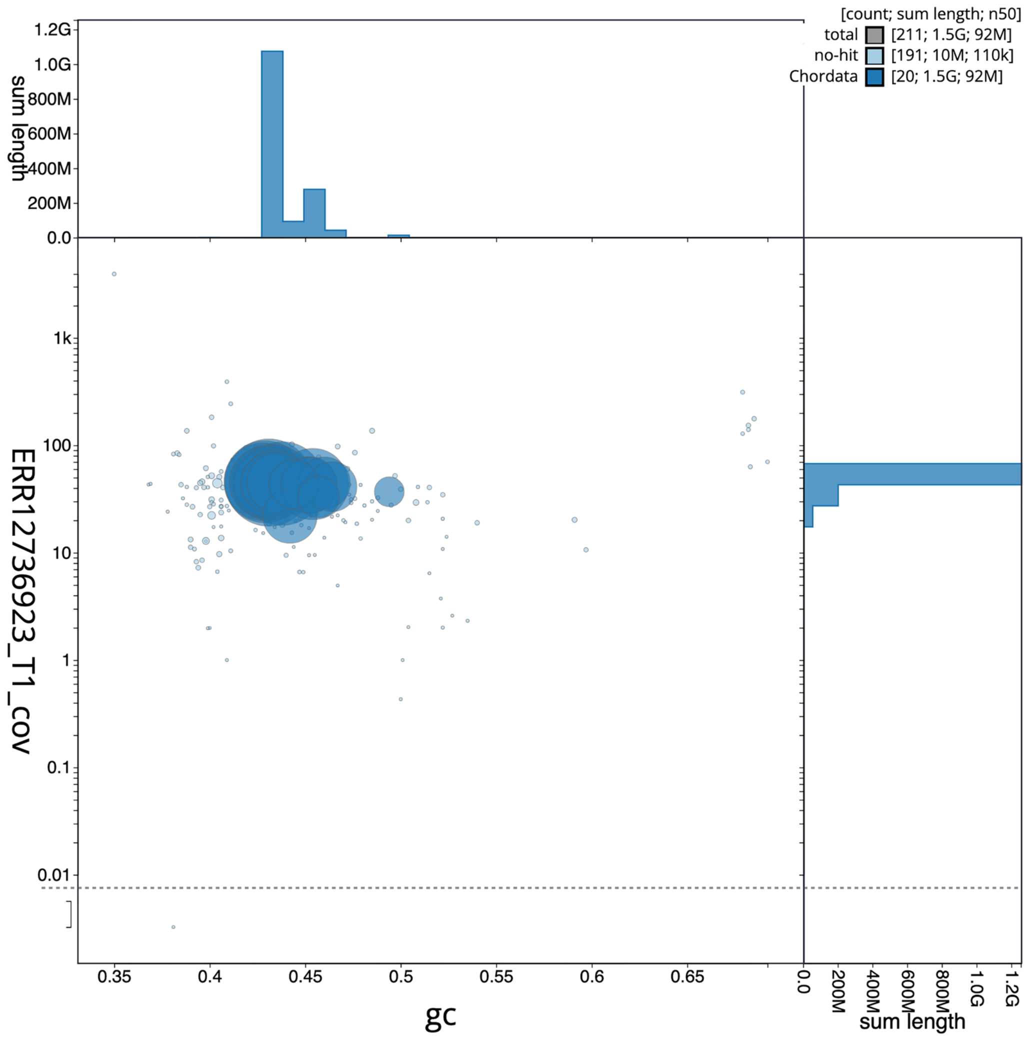

Figure 3. Genome assembly of *Podarcis gaigeae*, rPodGai1.hap1.2: BlobToolKit GC-coverage plot. Blob plot showing sequence coverage (vertical axis) and GC content (horizontal axis). The circles represent scaffolds, with the size proportional to scaffold length and the colour representing phylum membership. The histograms along the axes display the total length of sequences distributed across different levels of coverage and GC content.

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, the biotinylated DNA constructs were fragmented using a Covaris E220 sonicator and size-selected to 400–600 bp using SPRISelect beads. DNA was then enriched using Arima-HiC v2 Enrichment beads. The NEBNext Ultra II DNA Library Prep Kit (New England Biolabs) was used for end repair, A-tailing, and adapter ligation, following a modified protocol in which library preparation is carried out while the DNA remains bound to

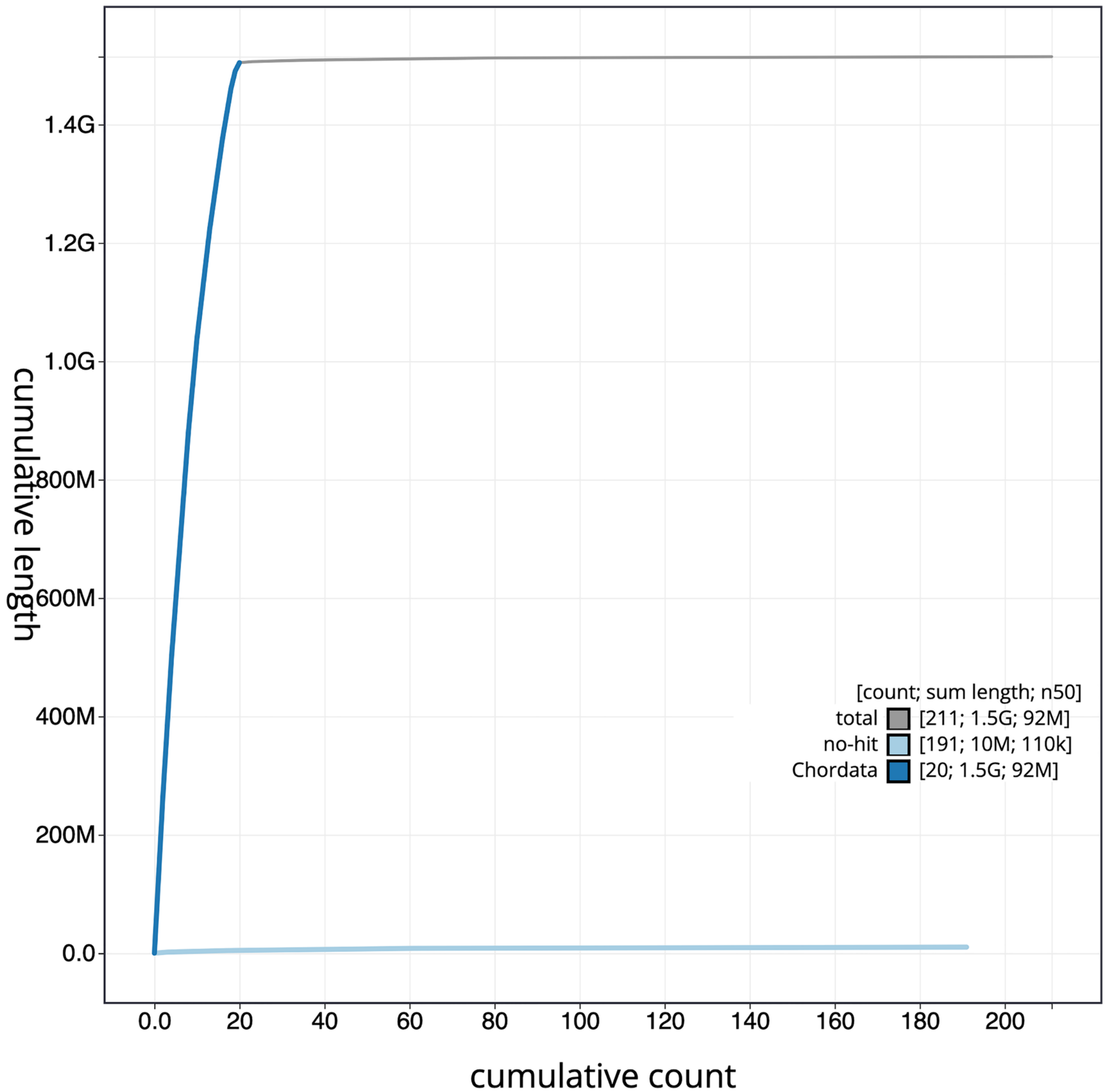

Figure 4. Genome assembly of *Podarcis gaigeae,* rPodGai1.hap1.2: BlobToolKit cumulative sequence plot. The grey line shows cumulative length for all scaffolds. Coloured lines show cumulative lengths of scaffolds assigned to each phylum using the buscogenes taxrule.

the enrichment beads. PCR amplification was performed using KAPA HiFi HotStart mix and custom dual-indexed adapters (Integrated DNA Technologies) in a 96-well plate format. Depending on sample concentration and biotinylation percentage determined at the crosslinking stage, samples were amplified for 10–16 PCR cycles. Post-PCR clean-up was carried out using SPRISelect beads. The libraries were quantified using the Accuclear Ultra High Sensitivity dsDNA Standards Assay kit (Biotium) and normalised to 10 ng/μL before sequencing. Hi-C sequencing was performed on the Illumina NovaSeq X instrument using 150 bp paired-end reads.

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),

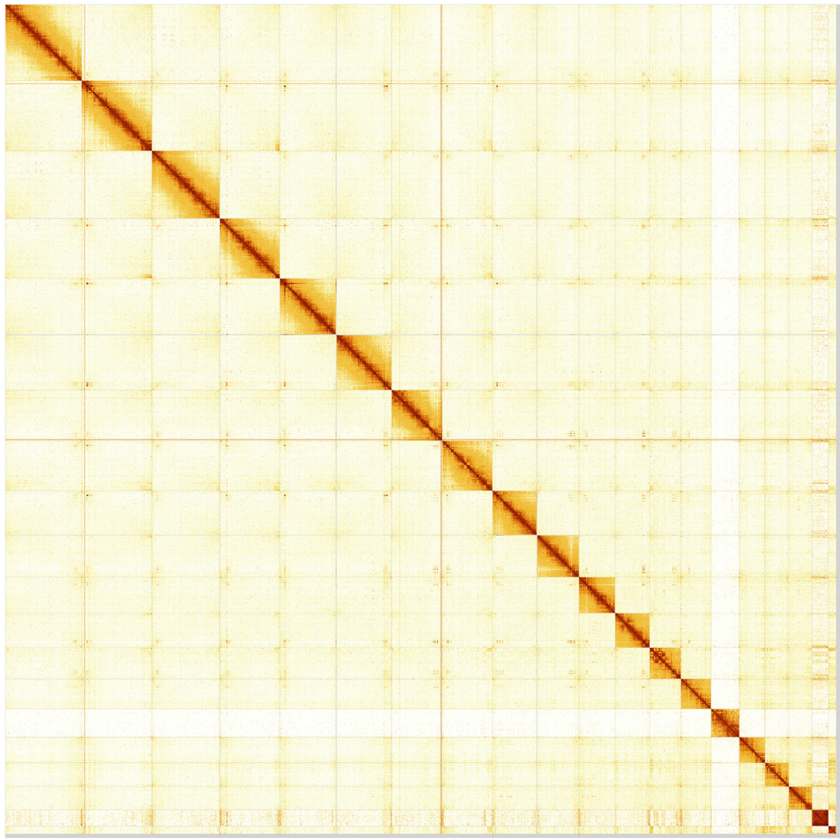

Figure 5. Genome assembly of *Podarcis gaigeae*, rPodGai1.hap1.2: Hi-C contact map of the rPodGai1.hap1.2 assembly, produced in PretextView. Chromosomes are shown in order of size from left to right and top to bottom.

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 datain 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

Table 3. Chromosomal pseudomolecules in the genome assembly of *Podarcis gaigeae*, rPodGai1.

Haplotype 1				Haplotype 2			
INSDC accession	Name	Length (Mb)	GC%	INSDC accession	Name	Length (Mb)	GC%
OZ066914.1	1	138.85	43	OZ066879.1	1	137.19	43
OZ066915.1	2	127.28	44	OZ066880.1	2	124.56	44
OZ066916.1	3	122.63	43	OZ066881.1	3	123.29	43
OZ066917.1	4	108.66	43	OZ066882.1	4	106.82	43
OZ066918.1	5	101.56	43	OZ066883.1	5	101.64	43
OZ066919.1	6	99.79	43.5	OZ066884.1	6	100.51	43.5
OZ066920.1	7	91.93	43	OZ066885.1	7	91.28	43
OZ066921.1	8	90.96	45.5	OZ066886.1	8	90.61	45.5
OZ066922.1	9	80.5	43	OZ066887.1	9	79.57	43
OZ066923.1	10	75.82	43.5	OZ066888.1	10	76.2	43.5
OZ066924.1	11	65.26	43	OZ066889.1	11	64.96	43
OZ066925.1	12	62.69	43.5	OZ066890.1	12	63.81	43.5
OZ066926.1	13	56.93	45	OZ066891.1	13	56.5	45
OZ066927.1	14	54.28	45	OZ066892.1	14	53.92	45
OZ066929.1	15	46.85	46	OZ066893.1	15	46.13	46
OZ066930.1	16	43.03	46.5	OZ066894.1	16	43.2	46.5
OZ066931.1	17	42.22	44.5	OZ066895.1	17	42.14	44.5
OZ066933.1	18	14.21	49.5	OZ066896.1	18	13.98	49.5
OZ066932.1	W	29.52	45.5				
OZ066928.1	Z	51.47	44				
OZ066935.1	MT	0.02	35.5				

(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.

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	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

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 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 gaigeae* (Skyros wall lizard). Accession number PRJEB73729; https://identifiers.org/ena.embl/PRJEB73729. The genome sequence is released openly for reuse. The *Podarcis gaigeae* genome assembly is provided by the Wellcome Sanger Institute Tree of Life Programme (https://www.sanger.ac.uk/programme/tree-of-life/) and is part of 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.

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

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

Bowles P: **Podarcis gaigeae.** *IUCN Red List of Threatened Species.* 2024; **2024**: e.T61547A137855287.
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

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

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

Legakis A, Maragos P: **The Red Book of endangered animals of Greece.** Athens, Greece: Hellenic Zoological Society, 2009.
Reference Source

Li H: **Minimap2: pairwise alignment for nucleotide sequences.** *Bioinformatics.* 2018; **34**(18): 3094–3100.
PubMed Abstract | Publisher Full Text | Free Full Text

Lymberakis P, Pafilis P, Poulakakis N, *et al.*: **The Amphibians and Reptiles of the Aegean Sea.** In: *Biogeography and Biodiversity of the Aegean*. 2018; 169–189.
Reference Source

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

Pafilis P, Foufopoulos J, Sagonas K, *et al.*: **Reproductive biology of insular

reptiles: marine subsidies modulate expression of the 'Island syndrome'. *Copeia.* 2011; **2011**(4): 545–552.
Publisher Full Text

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

Poulakakis N, Lymberakis P, Valakos E, *et al.*: **Phylogenetic relationships and biogeography of *Podarcis* species from the Balkan Peninsula, by bayesian and maximum likelihood analyses of mitochondrial DNA sequences.** *Mol Phylogenet Evol.* 2005; **37**(3): 845–857.
PubMed Abstract | **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**

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**

Runemark A, Hansson B, Pafilis P, *et al.*: **Island biology and morphological divergence of the Skyros wall lizard *Podarcis gaigeae*: a combined role for local selection and genetic drift on color morph frequency divergence?** *BMC Evol Biol.* 2010; **10**(1): 269.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

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

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

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

Uliano-Silva M, Ferreira JGRN, Krasheninnikova K, *et al.*: **MitoHiFi: a python pipeline for mitochondrial genome assembly from PacBio high fidelity reads.** *BMC Bioinformatics.* 2023; **24**(1): 288.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

Valakos ED, Pafilis P, Lymberakis P, *et al.*: **The Amphibians and Reptiles of Greece.** Chimaira, 2008.
Reference Source

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**