DATA NOTE

The genome sequence of the Cretan wall lizard, *Podarcis cretensis* (Wettstein, 1952)

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

Nikos Poulakakis[1,2], Petros Lymberakis[2], Manos Stratakis[1,2], Iliana Bista[3-5], Wellcome Sanger Institute Tree of Life Management, Samples and Laboratory team, Wellcome Sanger Institute Scientific Operations: Sequencing Operations, Wellcome Sanger Institute Tree of Life Core Informatics team, Tree of Life Core Informatics collective

[1]Biology Department, University of Crete, Heraklion, Greece
[2]Natural History Museum of Crete, University of Crete, Heraklion, Greece
[3]LOEWE Centre for Translational Biodiversity Genomics, Frankfurt, Germany
[4]Senckenberg Research Institute, Frankfurt, Germany
[5]Naturalis Biodiversity Center, Leiden, South Holland, The Netherlands

v1 **First published:** 20 Mar 2024, **9**:161
https://doi.org/10.12688/wellcomeopenres.21176.1
Latest published: 20 Mar 2024, **9**:161
https://doi.org/10.12688/wellcomeopenres.21176.1

Abstract

We present a genome assembly from an individual male *Podarcis cretensis* (the Cretan wall lizard; Chordata; Lepidosauria; Squamata; Lacertidae). The genome sequence is 1,507.6 megabases in span. Most of the assembly is scaffolded into 19 chromosomal pseudomolecules, including the Z sex chromosome. The mitochondrial genome has also been assembled and is 17.09 kilobases in length.

Keywords

Podarcis cretensis, Cretan wall lizard, genome sequence, chromosomal, Squamata

This article is included in the Tree of Life gateway.

Open Peer Review

Approval Status ? ?

	1	2
version 1 20 Mar 2024	? view	? view

1. **Terry Bertozzi**, South Australian Museum, Adelaide, Australia
The University of Adelaide School of Biological Sciences, Adelaide, Australia

2. **Hayden R. Davis**, University of Washington, Seattle, USA

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

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

Author roles: Poulakakis N: Investigation, Resources, Writing – Original Draft Preparation, Writing – Review & Editing; **Lymberakis P**: Investigation, Resources, Writing – Original Draft Preparation; **Stratakis M**: Investigation, Resources; **Bista I**: 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 [206194, https://doi.org/10.35802/206194]. Biodiversity Genomics Europe. HORIZON-CL6-2021-BIODIV-01-01 - European participation in global biodiversity genomics endeavours aimed at identifying all biodiversity on Earth. HORIZON-RIA. High mountains, shallow waters: the effects of climate change on organisms living in 'extreme' and sensitive ecosystems. Funding Agency: Green Fund, Ministry of Environment & Energy (3333/2022). IB was supported by the Centre for Translational Biodiversity Genomics (LOEWE-TBG) through the program LOEWE–Landes-Offensive zur Entwicklung Wissenschaftlich-ökonomischer Exzellenz of Hesse's Ministry of Higher Education, Research, and the Arts.
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: 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 Research 2024, **9**:161 https://doi.org/10.12688/wellcomeopenres.21176.1

First published: 20 Mar 2024, **9**:161 https://doi.org/10.12688/wellcomeopenres.21176.1

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 cretensis* (Wettstein, 1952) (NCBI:txid1330250).

Background

Podarcis cretensis (Figure 1), the Cretan wall lizard, is a species of lacertid lizard endemic to the island of Crete (including its satellite islands) in Greece. Crete is an island that hosts many endemic species due to its long-lasting geographical isolation. At the same time, it experiences continuous human activity during at least the past 5,000 years, with increasing effects on local habitats and biodiversity. It is also lying at a geographic region that is expected to experience severe effects from on-going climate change, leading to increased drought.

Until 2007, the wall lizard populations of Crete were described as subspecies of *Podarcis erhardii*. But analyses on genetic and morphological data revealed that they constitute a distinct species, closely related to *Podarcis peloponnesiacus* from Peloponnese and *Podarcis levendis* (from the islands of Pori and Lagouvardos) (Lymberakis *et al.*, 2008).

The target species is Endangered [B1ab(iii) ver 3.1; but it needs a reconsideration] and is protected under Appendix II of Bern and Annex IV of the EU Habitats Directive (as part of *P. erhardii*). Its range includes sites that are in the NATURA 2000 network. Its populations are believed to have declined during the past few decades, especially due to the fragmentation and/or destruction of its habitat (Lymberakis, 2009).

It is present only on the western part of Crete and satellite islets at the East of Crete from sea level to 2,000m asl. The population size has not been quantified but is very patchily distributed.

Figure 1. Photographs of *Podarcis cretensis* by Nikos Poulakakis (not the specimen used for genome sequencing).

The snout-vent length (SVL) of adults is up to 7.1 cm, with the tail almost twice as long (Lymberakis *et al.*, 2008). Males are larger than females. Its usual basic ground colour is brown or greyish, but may also be olive-brown, olive-green or brown-black with many dark markings and spots (Pafilis & Maragou, 2020; Valakos *et al.*, 2008). Two narrow light dorsolateral stripes with darker edges are often present and are more clearly defined in females. The belly is unspotted, white, light grey, bluish or yellow and during the breeding season may range from orange to brick red in males. These may also carry large blue spots on the hind legs and flanks. The juveniles look much like females. It is a diurnal species that may be active throughout the year. In summer, it seeks shelter during the hot midday period and reappears in the afternoon. The mating season is from February to May. The species produces at least two clutches per year with 1 to 5 eggs per clutch. It occupies shrubland, rocky areas, and dry riverbeds with low, dense and xeric vegetation (Mediterranean-type shrubby vegetation). It feeds mainly on small invertebrates, especially beetles, land snails and spiders. It sometimes feeds on smaller lizards, while nectar-feeding populations have also been observed.

Evolutionary adaptation to thermal environments is probably the most important factor that influences the fate and geographic range of a species in response to climate change. While impacts of climate change are expected in all taxa, their nature and strength are likely to vary among species. Failure to adjust or adapt culminates in demographic collapse and extinction, which is more likely to happen to species with narrow climatic and habitat requirements, limited dispersal abilities, and ectothermal physiology (Massot *et al.*, 2008). Lizards are considered "the new amphibians", because of their elevated extinction risk, as 20% of lizard species or 39% of local populations are predicted to become extinct worldwide by 2080 (Sinervo *et al.*, 2010). Thus, it is crucially important to investigate how climate change will impact ectothermic animals, especially reptiles and particularly lizards. The genome of *P. cretensis* will be used in a current project, funded by Green Fund (Greece), which aims to assess the biological responses of organisms to climate change, monitoring the changes in biodiversity dynamics and identifying selective pressures on areas that can potentially act as refugia of tomorrow's biota.

Genome sequence report

The genome was sequenced from one male *Podarcis cretensis* collected from Anopoli, Chania, Crete, Greece (35.33, 24.05). A total of 27-fold coverage in Pacific Biosciences single-molecule HiFi long reads was generated. Primary assembly contigs were scaffolded with chromosome conformation Hi-C data. Manual assembly curation corrected 9 missing joins or mis-joins and removed one haplotypic duplication, reducing the scaffold number by 5.56%.

The final assembly has a total length of 1507.6 Mb in 50 sequence scaffolds with a scaffold N50 of 94.6 Mb (Table 1). The snail plot in Figure 2 provides a summary of the assembly statistics, while the distribution of assembly scaffolds on GC proportion and coverage is shown in Figure 3. The cumulative assembly plot in Figure 4 shows curves for subsets of scaffolds assigned to different phyla. Most (99.79%) of the assembly sequence was assigned to 19 chromosomal-level scaffolds, representing 18 autosomes and the Z sex chromosome. Chromosome-scale scaffolds confirmed by the Hi-C data are named in order of size (Figure 5; Table 2). Chromosome Z was assigned based on synteny to the genome

Table 1. Genome data for *Podarcis cretensis*, rPodCre2.1.

Project accession data		
Assembly identifier	rPodCre2.1	
Species	*Podarcis cretensis*	
Specimen	rPodCre2	
NCBI taxonomy ID	1330250	
BioProject	PRJEB61854	
BioSample ID	SAMEA12110019	
Isolate information	rPodCre2, male: muscle (DNA, Hi-C and RNA sequencing)	
Assembly metrics*		***Benchmark***
Consensus quality (QV)	65.1	*≥ 50*
k-mer completeness	100.0%	*≥ 95%*
BUSCO**	C:95.4%[S:93.7%,D:1.7%], F:0.9%,M:3.8%,n:7,480	*C ≥ 95%*
Percentage of assembly mapped to chromosomes	99.79%	*≥ 95%*
Sex chromosomes	Z chromosome	*localised homologous pairs*
Organelles	Mitochondrial genome: 17.09 kb	*complete single alleles*
Raw data accessions		
PacificBiosciences SEQUEL II	ERR11413981, ERR11413982	
Hi-C Illumina	ERR11439636, ERR11439637, ERR11439635	
PolyA RNA-Seq Illumina	ERR11439638	
Genome assembly		
Assembly accession	GCA_951804945.1	
Accession of alternate haplotype	GCA_951804955.1	
Span (Mb)	1,507.6	
Number of contigs	98	
Contig N50 length (Mb)	65.7	
Number of scaffolds	50	
Scaffold N50 length (Mb)	94.6	
Longest scaffold (Mb)	141.6	

* Assembly metric benchmarks are adapted from column VGP-2020 of "Table 1: Proposed standards and metrics for defining genome assembly quality" from Rhie *et al.* (2021).

** BUSCO scores based on the sauropsida_odb10 BUSCO set using v5.3.2. C = complete [S = single copy, D = duplicated], F = fragmented, M = missing, n = number of orthologues in comparison. A full set of BUSCO scores is available at https://blobtoolkit.genomehubs.org/view/Podarcis%20cretensis/dataset/rPodCre2_1/busco.

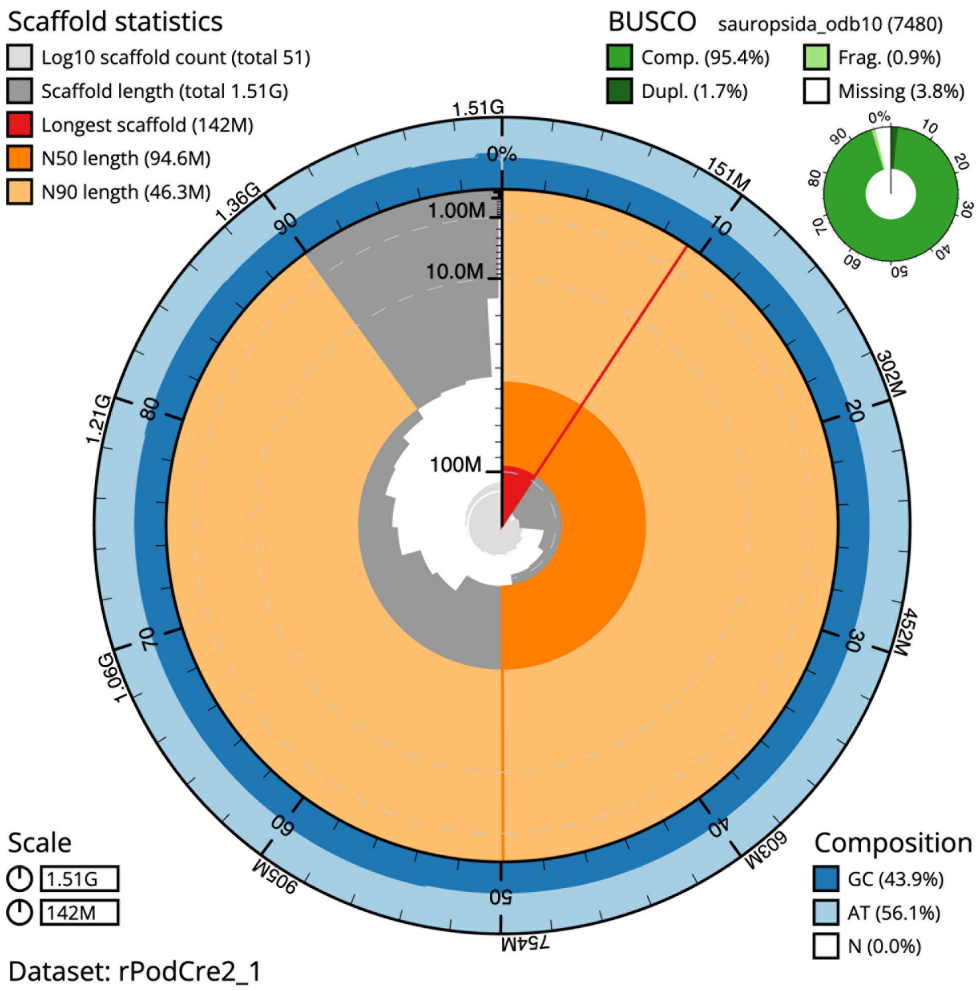

Figure 2. Genome assembly of *Podarcis cretensis*, rPodCre2.1: metrics. The BlobToolKit snail plot shows N50 metrics and BUSCO gene completeness. The main plot is divided into 1,000 size-ordered bins around the circumference with each bin representing 0.1% of the 1,507,627,055 bp assembly. The distribution of scaffold lengths is shown in dark grey with the plot radius scaled to the longest scaffold present in the assembly (141,635,579 bp, shown in red). Orange and pale-orange arcs show the N50 and N90 scaffold lengths (94,621,998 and 46,264,000 bp), respectively. The pale grey spiral shows the cumulative scaffold count on a log scale with white scale lines showing successive orders of magnitude. The blue and pale-blue area around the outside of the plot shows the distribution of GC, AT and N percentages in the same bins as the inner plot. A summary of complete, fragmented, duplicated and missing BUSCO genes in the sauropsida_odb10 set is shown in the top right. An interactive version of this figure is available at https://blobtoolkit.genomehubs.org/view/Podarcis%20cretensis/dataset/rPodCre2_1/snail.

assembly of *Podarcis muralis nigriventris* (GCA_014706415.1). While not fully phased, the assembly deposited is of one haplotype. Contigs corresponding to the second haplotype have also been deposited. The mitochondrial genome was also assembled and can be found as a contig within the multifasta file of the genome submission.

The estimated Quality Value (QV) of the final assembly is 65.1, with *k*-mer completeness of 100.0%, and the assembly has a BUSCO v5.3.2 completeness of 95.4% (single = 93.7%, duplicated = 1.7%), using the sauropsida_odb10 reference set (n = 7,480).

Metadata for specimens, barcode results, spectra estimates, sequencing runs, contaminants and pre-curation assembly statistics are given at https://tolqc.cog.sanger.ac.uk/erga/reptiles/Podarcis_cretensis/.

Methods

Sample acquisition

The specimen, an adult male *Podarcis cretensis* Lacertid lizard (Biospecimen ID SAMEA12110018, ToLID rPodCre2) was collected on the 23 July 2020 from a high-altitude shrub habitat (ca 2000m) over the village of Anopoli, Chania Greece. It was transported to NHMC facilities and euthanised

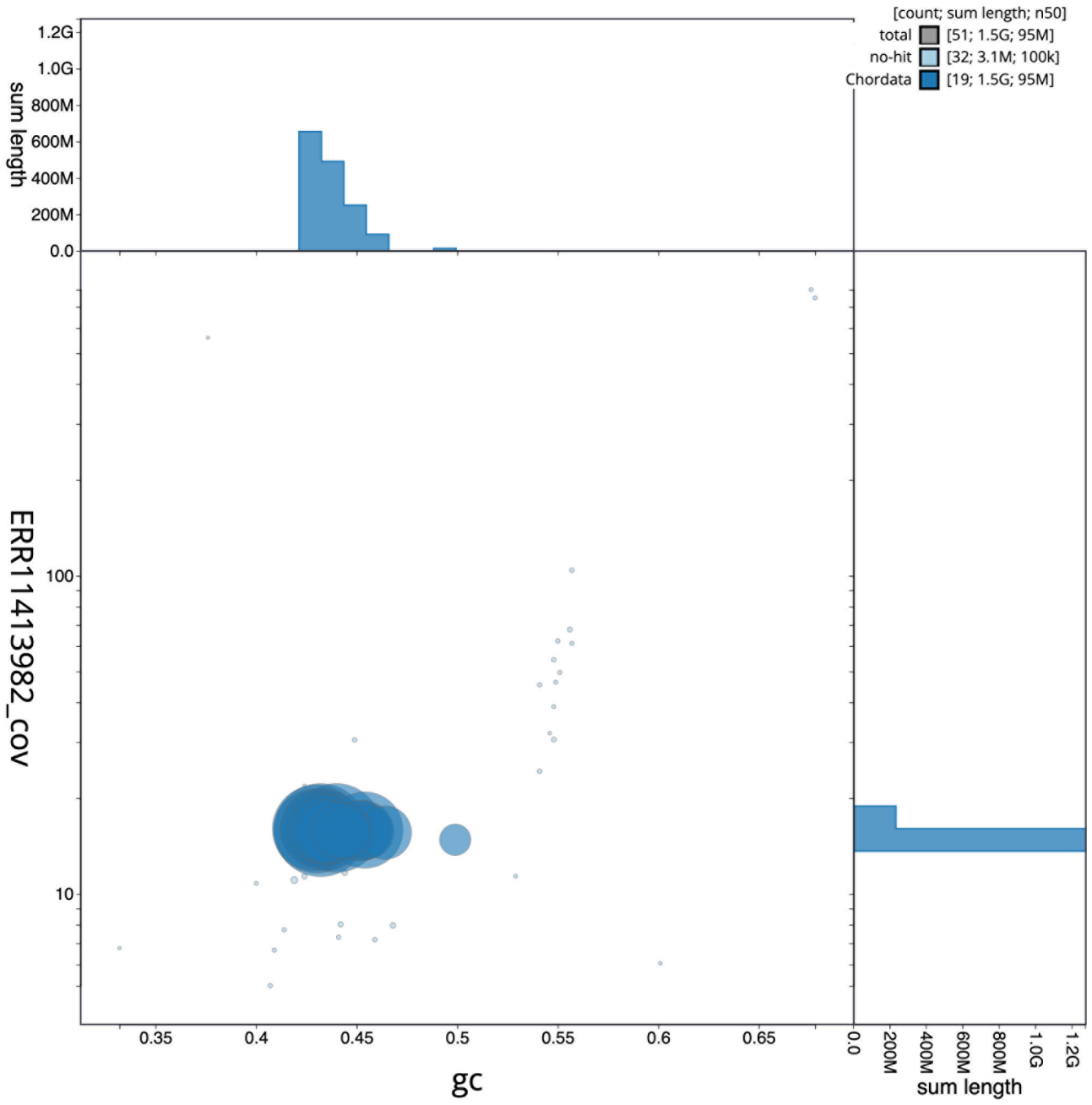

Figure 3. Genome assembly of *Podarcis cretensis*, rPodCre2.1: BlobToolKit GC-coverage plot. Scaffolds are coloured by phylum. Circles are sized in proportion to scaffold length. Histograms show the distribution of scaffold length sum along each axis. An interactive version of this figure is available at https://blobtoolkit.genomehubs.org/view/Podarcis%20cretensis/dataset/rPodCre2_1/blob.

by concussion, according to the European Commission directive (European Commission, Directorate-General for Environment, Close, B., Baumans, V., Banister, K., Euthanasia of experimental animals, Publications Office, 1997), and the brain was removed immediately after to ensure the animal was deceased. Brain, liver and heart tissue were immediately placed in –80 °C and the remaining part in 96% ethanol. The specimen collection number is NHMC80.3.51.2950.

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 core procedures: sample preparation; sample homogenisation, DNA extraction, fragmentation, and clean-up. In sample preparation, the rPodCre2 sample was weighed and dissected on dry ice (Jay *et al.*, 2023). For sample homogenisation, muscle tissue

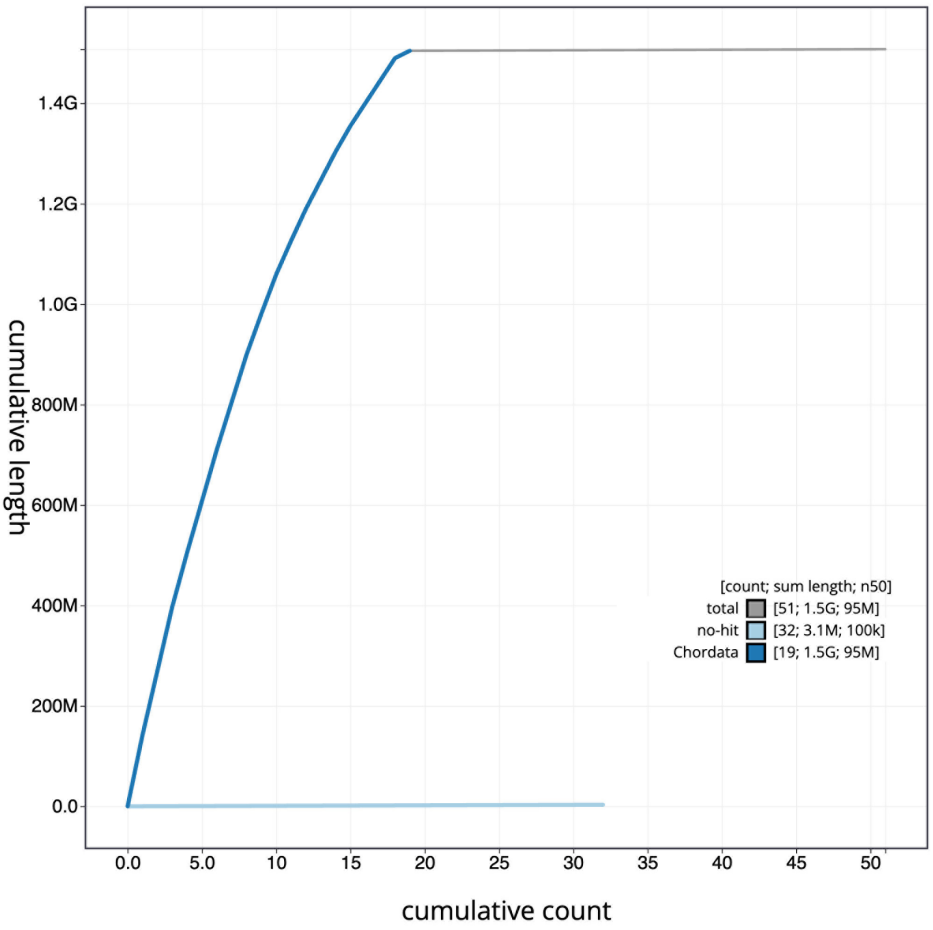

Figure 4. Genome assembly of *Podarcis cretensis*, rPodCre2.1: BlobToolKit cumulative sequence plot. The grey line shows cumulative length for all scaffolds. Coloured lines show cumulative lengths of scaffolds assigned to each phylum using the buscogenes taxrule. An interactive version of this figure is available at https://blobtoolkit.genomehubs.org/view/Podarcis%20cretensis/dataset/rPodCre2_1/cumulative.

was cryogenically disrupted using the Covaris cryoPREP® Automated Dry Pulverizer (Narváez-Gómez *et al.*, 2023).

HMW DNA was extracted using the Automated MagAttract v2 protocol (Oatley *et al.*, 2023a). HMW DNA was sheared into an average fragment size of 12–20 kb in a Megaruptor 3 system with speed setting 31 (Bates *et al.*, 2023). Sheared DNA was purified by solid-phase reversible immobilisation (Oatley *et al.*, 2023b): in brief, the method employs a 1.8X ratio of AMPure PB beads to sample to eliminate shorter fragments and concentrate the DNA. The concentration of the sheared and purified DNA was assessed using a Nanodrop spectrophotometer and Qubit Fluorometer and Qubit dsDNA High Sensitivity Assay kit. Fragment size distribution was evaluated by running the sample on the FemtoPulse system.

RNA was extracted from muscle tissue of rPodCre2 in the Tree of Life Laboratory at the WSI using the RNA Extraction: Automated MagMax™ *mir*Vana protocol (do Amaral *et al.*, 2023). The RNA concentration was assessed using a Nanodrop spectrophotometer and a Qubit Fluorometer using the Qubit RNA Broad-Range Assay kit. Analysis of the integrity of the RNA was done using the Agilent RNA 6000 Pico Kit and Eukaryotic Total RNA assay.

Protocols developed by the Wellcome Sanger Institute (WSI) Tree of Life core laboratory have been deposited on protocols.io (Denton *et al.*, 2023).

Sequencing

Pacific Biosciences HiFi circular consensus DNA sequencing libraries were constructed according to the manufacturers' instructions. Poly(A) RNA-Seq libraries were constructed using the NEB Ultra II RNA Library Prep kit. DNA and RNA sequencing was performed by the Scientific Operations core at the WSI on Pacific Biosciences SEQUEL II (HiFi) and Illumina NovaSeq 6000 (RNA-Seq) instruments. Hi-C data were also generated from muscle tissue of rPodCre2 using the Arima2 kit and sequenced on the Illumina NovaSeq 6000, Illumina NovaSeq 6000, Illumina NovaSeq 6000 instrument.

Genome assembly, curation and evaluation

Assembly was carried out with Hifiasm (Cheng *et al.*, 2021) and haplotypic duplication was identified and removed with

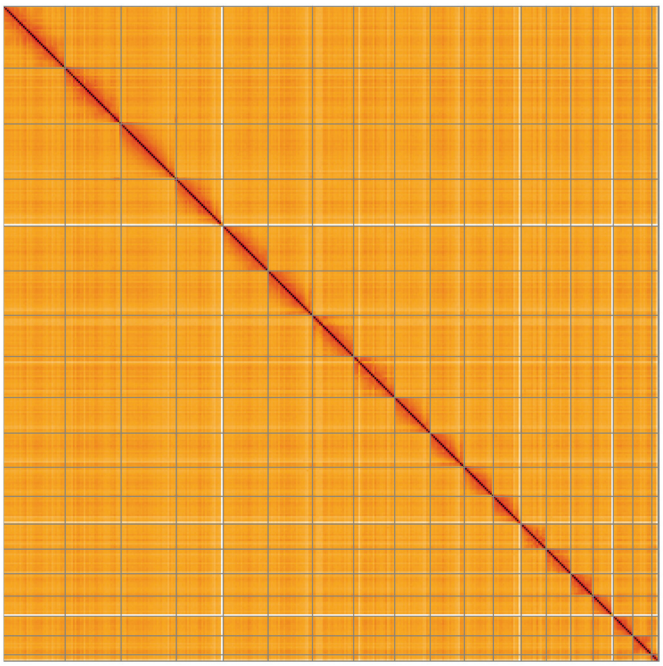

Figure 5. Genome assembly of *Podarcis cretensis*, rPodCre2.1: Hi-C contact map of the rPodCre2.1 assembly, visualised using HiGlass. Chromosomes are shown in order of size from left to right and top to bottom. An interactive version of this figure may be viewed at https://genome-note-higlass.tol.sanger.ac.uk/l/?d=Z_EqgAS4Re-IRqSnIxDZFw.

Table 2. Chromosomal pseudomolecules in the genome assembly of *Podarcis cretensis*, rPodCre2.

INSDC accession	Chromosome	Length (Mb)	GC%
OX638149.1	1	141.64	43.0
OX638150.1	2	128.45	44.0
OX638151.1	3	126.79	43.0
OX638152.1	4	107.92	43.0
OX638153.1	5	103.04	43.0
OX638154.1	6	102.1	43.5
OX638155.1	7	94.62	43.0
OX638156.1	8	94.37	45.5
OX638157.1	9	81.67	43.0
OX638158.1	10	78.44	43.5
OX638159.1	11	66.83	43.5
OX638160.1	12	63.68	43.5
OX638161.1	13	57.67	45.5
OX638162.1	14	56.36	45.0
OX638164.1	15	46.26	45.5
OX638165.1	16	45.08	46.5
OX638166.1	17	43.57	44.5
OX638167.1	18	14.62	50.0
OX638163.1	Z	51.45	44.0
OX638168.1	MT	0.02	38.5

purge_dups (Guan *et al.*, 2020). The assembly was then scaffolded with Hi-C data (Rao *et al.*, 2014) using YaHS (Zhou *et al.*, 2023). The assembly was checked for contamination and corrected as described previously (Howe *et al.*, 2021). Manual curation was performed using HiGlass (Kerpedjiev *et al.*, 2018) and PretextView (Harry, 2022). The mitochondrial genome was assembled using MitoHiFi (Uliano-Silva *et al.*, 2023), which runs MitoFinder (Allio *et al.*, 2020) or MITOS (Bernt *et al.*, 2013) and uses these annotations to select the final mitochondrial contig and to ensure the general quality of the sequence.

A Hi-C map for the final assembly was produced using bwa-mem2 (Vasimuddin *et al.*, 2019) in the Cooler file format (Abdennur & Mirny, 2020). To assess the assembly metrics, the *k*-mer completeness and QV consensus quality values were calculated in Merqury (Rhie *et al.*, 2020). This work was done using Nextflow (Di Tommaso *et al.*, 2017) DSL2 pipelines "sanger-tol/readmapping" (Surana *et al.*, 2023a) and "sanger-tol/genomenote" (Surana *et al.*, 2023b). The genome was analysed within the BlobToolKit environment (Challis *et al.*, 2020) and BUSCO scores (Manni *et al.*, 2021; Simão *et al.*, 2015) were calculated.

Table 3 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 this is

Table 3. Software tools: versions and sources.

Software tool	Version	Source
BlobToolKit	4.2.1	https://github.com/blobtoolkit/blobtoolkit
BUSCO	5.3.2	https://gitlab.com/ezlab/busco
Hifiasm	0.16.1-r375	https://github.com/chhylp123/hifiasm
HiGlass	1.11.6	https://github.com/higlass/higlass
Merqury	MerquryFK	https://github.com/thegenemyers/MERQURY.FK
MitoHiFi	3	https://github.com/marcelauliano/MitoHiFi
PretextView	0.2	https://github.com/wtsi-hpag/PretextView
purge_dups	1.2.5	https://github.com/dfguan/purge_dups
sanger-tol/genomenote	v1.0	https://github.com/sanger-tol/genomenote
sanger-tol/readmapping	1.1.0	https://github.com/sanger-tol/readmapping/tree/1.1.0
YaHS	1.2a.2	https://github.com/c-zhou/yahs

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 cretensis* (Cretan wall lizard). Accession number PRJEB61854; https://identifiers.org/ena.embl/PRJEB61854 (Wellcome Sanger Institute, 2023). The genome sequence is released openly for reuse. The *Podarcis cretensis* genome sequencing initiative is part of the European Reference Genome Atlas Pilot Project (https://www.erga-biodiversity.eu/pilot-project). All raw sequence data and the assembly have been deposited in INSDC databases. Raw data and assembly accession identifiers are reported in Table 1.

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.10066175.

Members of Wellcome Sanger Institute Scientific Operations: Sequencing Operations are listed here: https://doi.org/10.5281/zenodo.10043364.

Members of the Wellcome Sanger Institute Tree of Life Core Informatics team are listed here: https://doi.org/10.5281/zenodo.10066637.

Members of the Tree of Life Core Informatics collective are listed here: https://doi.org/10.5281/zenodo.5013541.

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

Bates A, Clayton-Lucey I, Howard C: **Sanger Tree of Life HMW DNA Fragmentation: Diagenode Megaruptor®3 for LI PacBio.** *protocols.io.* 2023.
Publisher Full Text

Bernt M, Donath A, Jühling F, *et al.*: **MITOS: Improved *de novo* metazoan mitochondrial genome annotation.** *Mol Phylogenet Evol.* 2013; **69**(2): 313–319.
PubMed Abstract | Publisher 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

Denton A, Yatsenko H, Jay J, *et al.*: **Sanger Tree of Life Wet Laboratory Protocol Collection V.1.** *protocols.io.* 2023.
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

do Amaral RJV, Bates A, Denton A, *et al.*: **Sanger Tree of Life RNA Extraction: Automated MagMax™ mirVana.** *protocols.io.* 2023.
Publisher Full Text

Guan D, McCarthy SA, Wood J, *et al.*: **Identifying and removing haplotypic duplication in primary genome assemblies.** *Bioinformatics.* 2020; **36**(9): 2896–2898.
PubMed Abstract | Publisher Full Text | Free Full Text

Harry E: **PretextView (Paired REad TEXTure Viewer): A desktop application for viewing pretext contact maps.** 2022; [Accessed 19 October 2022].
Reference Source

Howe K, Chow W, Collins J, *et al.*: **Significantly improving the quality of genome assemblies through curation.** *GigaScience.* Oxford University Press, 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

Lymberakis P, Poulakakis N, Kaliontzopoulou A, *et al.*: **Two new species of *Podarcis* (Squamata; Lacertidae) from Greece.** *Syst Biodivers.* 2008; **6**(3): 307–318.
Publisher Full Text

Lymberakis P: **Podarcis cretensis**. The IUCN Red List of Threatened Species, 2009; [Accessed: 23 January 2024].
Publisher Full Text

Manni M, Berkeley MR, Seppey M, *et al.*: **BUSCO update: Novel and streamlined workflows along with broader and deeper phylogenetic coverage for scoring of eukaryotic, prokaryotic, and viral genomes.** *Mol Biol Evol.* 2021; **38**(10): 4647–4654.
PubMed Abstract | Publisher Full Text | Free Full Text

Massot M, Clobert J, Ferrière R: **Climate warming, dispersal inhibition and extinction risk.** *Glob Change Biol.* 2008; **14**(3): 461–469.
Publisher Full Text

Narváez-Gómez JP, Mbye H, Oatley G, *et al.*: **Sanger Tree of Life Sample Homogenisation: Covaris cryoPREP® Automated Dry Pulverizer V.1.** *protocols.io.* 2023.
Publisher Full Text

Oatley G, Denton A, Howard C: **Sanger Tree of Life HMW DNA Extraction: Automated MagAttract v.2.** *protocols.io.* 2023a.
Publisher Full Text

Oatley G, Sampaio F, Howard C: **Sanger Tree of Life Fragmented DNA clean up: Automated SPRI.** *protocols.io.* 2023b.
Publisher Full Text

Pafilis P, Maragou P: **Atlas of Amphibians and Reptiles of Greece [Greek]**. Nicosia, Cyprus: Broken - Hill Publishers Ltd, 2020.

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

Simão FA, Waterhouse RM, Ioannidis P, *et al.*: **BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs.** *Bioinformatics.* 2015; **31**(19): 3210–3212.
PubMed Abstract | Publisher Full Text

Sinervo B, Méndez-de-la-Cruz F, Miles DB, *et al.*: **Erosion of lizard diversity by climate change and altered thermal niches.** *Science*. 2010; **328**(5980): 894–899.
PubMed Abstract | Publisher Full Text

Surana P, Muffato M, Qi G: **sanger-tol/readmapping: sanger-tol/readmapping v1.1.0 - Hebridean Black (1.1.0).** *Zenodo.* 2023a.
Publisher Full Text

Surana P, Muffato M, Sadasivan Baby C: **sanger-tol/genomenote (v1.0.dev).** *Zenodo.* 2023b.
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 E, Pafilis P, Sotiropoulos K, *et al.*: **Reptiles and amphibians of Greece.** Frankfurt am Main: 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

Wellcome Sanger Institute: **The genome sequence of the Cretan wall lizard, *Podarcis cretensis* (Wettstein, 1952).** European Nucleotide Archive. [dataset], accession number PRJEB61854, 2023.

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 07 December 2024

https://doi.org/10.21956/wellcomeopenres.23421.r108664

?

Hayden R. Davis
University of Washington, Seattle, Washington, USA

The authors present a very important and fairly high quality reference genome for an important species, *Podarcis cretensis*. The authors use a combination of long read, Hi-C, and RNAseq data. The methods applied are appropriate for the study and they used pipelines that are publicly available, making reproduction of their results realistic. In addition to assembling one full haplotype, the authors also published the contigs for the second haplotype.

Other than a couple minor comments that I feel should be addressed, this manuscript is in good condition.

1. Including an explanation for why the authors selected a specimen from a high-altitude shrub habitat would be useful.

2. I'm not seeing a collection permit for this specimen. I'm not sure if the 'materials contributed by a ToL collaborator' also refers to the specimen, but it seems that it was collected for the purposes of this study, so there should be a collection permit.

3. Was RepeatModeler and RepeatMasker used prior to annotation?

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?
Partly

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

Yes

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: bioinformatics; population genetics; molecular ecology

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 22 August 2024

https://doi.org/10.21956/wellcomeopenres.23421.r93810

?

Terry Bertozzi
[1] Evolutionary Biology Unit, South Australian Museum, Adelaide, South Australia, Australia
[2] Molecular and Biomedical Science, The University of Adelaide School of Biological Sciences, Adelaide, South Australia, Australia

The authors report on the sequencing and genome assembly of *Podarcis cretensis*, an endemic lizard from Crete. The genome size is within the range expected for squamates and appears to be of excellent quality and completeness based on the metrics provided. The methods used are appropriate for the scope of work and the datasets and metadata are well laid out. However, there are some minor issues that should be addressed to improve the clarity and repeatability of the work. I have addressed these below titled by their relevant sections.

Genome sequence report
It should be specified that genome size given is haploid. That is the norm but with the onset of T2T, it won't be long before diploid/polyploid genomes are described.

The Y axis of Figure 3 could be labelled more clearly with the word coverage.

Background
I am unclear what "B1ab(iii) ver 3.1" refers to? Further the manuscript states that the conservation status of the species needs reconsidering but there is no context. A brief a statement as to why would be beneficial.

I assume "Appendix II of Bern" should be "Appendix II of the Bern Convention"

Expand "NATURA 2000 network" to "NATURA 2000 protected areas network" which provides more context for the rationale.

The fourth paragraph should be incorporated into the third, given the topics are related.

Sequencing
Add “respectively” to the end of the third sentence to make it crystal clear which sequencing technology was being used for which purpose.

Final sentence - Illumina Novaseq 6000 is repeated three times. It would be better in my opinion to indicate that either, three sequencing libraries were made and all were sequenced using that platform or the same library was sequenced three times using that platform.

It is unclear how the HiC and RNAseq was sequenced on the Illumina platform. I assume they are paired end reads but I could not find that information nor the number of sequencing cycles used in each run?

The Runid links in ToI data access pages appear to have link errors and run accessions are pending – please check.

Genome assembly, curation and evaluation
There are no bioinformatics details provided for the assembly or scaffolding steps that I could find. While the software and version numbers are provided, there is no way to repeat the analyses as no software parameters have been provided. If the defaults are used, this should be explicitly stated. Further, the HiC data needs to be mapped to the assembly before it can be scaffolded with YaHS -how was this done?

Given that RNA was sequenced, I find it strange that there has been no attempt to at least perform a cursory gene prediction/annotation or use it in any way. Similarly, there is no discussion of the mitochondrial genome other than reporting its size. It would be preferable to have some indication of its quality, for example were the expected genes/tRNAs present.

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

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

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

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

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Bioinformatics, 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, however I have

significant reservations, as outlined above.