A high-quality reference genome for the critically endangered Aeolian wall lizard, *Podarcis raffonei*

Maëva Gabrielli[1*], Andrea Benazzo[1*], Roberto Biello[1], Lorena Ancona[2], Silvia Fuselli[1], Alessio Iannucci[3], Jennifer Balacco[4], Jacqueline Mountcastle[4], Alan Tracey[5], Gentile Francesco Ficetola[6,7], Daniele Salvi[8], Marco Sollitto[9], Olivier Fedrigo[4], Giulio Formenti[3,4], Erich D. Jarvis[4,10], Marco Gerdol[9], Claudio Ciofi[3], Emiliano Trucchi[2], Giorgio Bertorelle[1]

1 - Department of Life Sciences and Biotechnology, University of Ferrara, Ferrara, Italy

2 - Department of Life and Environmental Sciences, Marche Polytechnic University, Ancona, Italy

3 - Department of Biology, University of Florence, Florence, Italy

4 - Vertebrate Genome Laboratory, The Rockefeller University, New York, NY, USA

5 – Tree of Life, Wellcome Sanger Institute, Cambridge, United Kingdom

6 - Department of Environmental Sciences and Policy, University of Milan, Milan, Italy

7 - Laboratoire d'Ecologie Alpine (LECA), CNRS, Université Grenoble Alpes and Université Savoie Mont Blanc, Grenoble, France

8 - Department of Health, Life & Environmental Sciences - University of L'Aquila, L'Aquila, Italy

9 - Department of Life Sciences, University of Trieste, Trieste, Italy

10 - Howard Hughes Medical Institute, Chevy Chase, MD, USA

* These authors contributed equally

Abstract

The Aeolian wall lizard, *Podarcis raffonei*, is an endangered species endemic to the Aeolian archipelago, Italy, where it is present only in three tiny islets and a narrow promontory of a larger island. Because of the extremely limited area of occupancy, severe population fragmentation and observed decline, it has been classified as Critically Endangered by the International Union for the Conservation of Nature (IUCN). Using Pacific Biosciences (PacBio) High Fidelity (HiFi) long read sequencing, Bionano optical mapping and Arima chromatin conformation capture sequencing (Hi-C), we produced a high-quality, chromosome-scale reference genome for the Aeolian wall lizard, including Z and W sexual chromosomes. The final assembly spans 1.51 Gb across 28 scaffolds with a contig N50 of 61.4 Mb, a scaffold N50 of 93.6 Mb, and a BUSCO completeness score of 97.3%. This genome constitutes a valuable resource for the species to guide potential conservation efforts and more generally for the squamate reptiles that are underrepresented in terms of available high-quality genomic resources.

Keywords

Conservation genetics, Endemixit, Lacertids, De novo assembly, PacBio HiFi, Hi-C

Introduction

The Aeolian wall lizard *Podarcis raffonei* (Figure 1A) is one of the most endangered vertebrate in Europe (Gippoliti et al. 2017). It is endemic to four islands of the Aeolian archipelago, located North-East of Sicily, with an extremely restricted distribution range including three islets less than 0.01 km² (La Canna, Scoglio Faraglione, Strombolicchio) and a larger island (Vulcano, 21.2 km²) where it currently occupies nonetheless a very limited area (Bonardi et al. 2022). The total area of occupancy has been estimated to be as small as 5,000 m² (Ficetola et al. 2021) and the total population size is estimated at about 2,000 individuals (Capula et al. 2002; Cascio et al. 2014; Ficetola et al. 2018; 2021; Gippoliti et al. 2017). As a result, the Aeolian wall lizard has been listed as Critically Endangered in the Red List of Endangered Species of the IUCN (2009).

The main threats to its survival include interactions with the invasive Italian wall lizard *Podarcis siculus*, combined with habitat degradation (Capula et al. 2002). This is particularly visible on the island Vulcano, where the intense habitat change that occurred in the last 50 years may have favoured the spread of *Podarcis siculus*, leading to a sharp decline in the *Podarcis raffonei* population (Capula et al. 2002).

The production of highly contiguous genomes has greatly accelerated in the last decade, refining our understanding of the genomic basis of organismal traits, the chromosome evolution, and allowing the detection of natural selection through genomic scans (Geneva et al. 2022). Furthermore, reference genomes can be key for conservation genomics as they may permit, in combination with whole-genome resequencing data, to assess genetic diversity, investigate inbreeding depression or characterise deleterious mutations (Formenti, Theissinger, et al. 2022). High-quality reference genomes are unevenly distributed across the tree of life, and some clades, such as the squamate reptiles, are underrepresented (Pinto et al. 2022; Card, Jennings, and Edwards 2023).

Here, we present a high-quality chromosome-scale reference genome for the Aeolian wall lizard, produced as part of the Endemixit project (www.endemixit.com). Our final genome assembly spans 1.51 Gb across 28 scaffolds, with a scaffold N50 of 93.6 Mb and a BUSCO completeness score of 97.3%. This high-quality reference genome is a valuable resource to assess the genetic diversity in the four extant populations of the Aeolian wall lizard and better develop the conservation strategy for this species.

Methods

Biological Materials

An adult female was collected on the 31st of July, 2020 by D. Salvi on the stack of La Canna (38°34'56.13"N – 14°31'16.61"E ; See Supplementary Figure 1), in the Aeolian archipelago, in a small terrace at 50 m a.s.l. on the eastern slope of the stack, reached by climbing with the technical assistance of the mountain guide Lorenzo Inzigneri. A piece of tail was cut and immediately frozen in liquid nitrogen until the final storage at -80°C.

Nucleic acid extraction, library preparation, and sequencing

All the following steps were carried out at the Vertebrate Genomes Project (VGP, https://vertebrategenomesproject.org/) lab. High Molecular Weight (HMW) DNA was extracted from muscle with the Circulomics HMW DNA extraction standard TissueRuptor protocol with the Nanobind Tissue Big DNA Kit (PN NB-900-701-01). DNA absorbance was checked as quality and purity control with Nanodrop and average fragment length was verified with a Pulsed Field Gel Electrophoresis (PFGE).

Genomic data from three different sequencing technologies were used for the assembly: PacBio HiFi reads, Bionano optical maps, and Hi-C reads from Arima Genomics.

PacBio HiFi libraries were prepared using the Pacific Biosciences Express Template Prep Kit 2.0. The library was then size selected (>10 kb) using the Circulomics Short Read Eliminator. The PacBio library was sequenced on two PacBio 8M v3 SMRT Cells on a PacBio Sequel II and one PacBio 8M SMRT Cell on a PacBio Sequel IIe using the sequencing kit 2.0 and a 30-h movie.

An aliquot of the HMW DNA was labelled for Bionano Genomics optical mapping using the Bionano Prep Direct Label and Stain (DLS) Protocol and run on one Saphyr instrument chip flowcell.

Hi-C libraries were generated by Arima Genomics (https://arimagenomics.com/) using muscle in-vivo cross-linking with the Arima-HiC kit with 2-enzymes proximity ligation. Proximally ligated DNA was subjected to shearing, size selection (~200-600 bp) with SPRI beads, and enrichment with streptavidin beads for the biotin-labelled DNA. KAPA Hyper Prep kit was employed to generate libraries compatible with Illumina technologies. Libraries were amplified through PCR, purified with SPRI beads and

sequenced on an Illumina HiSeq X (~60X coverage) after a quality check with Bioanalyzer and qPCR.

Nuclear genome assembly

The genome of the Aeolian wall lizard was assembled following the VGP assembly pipeline v2.0 (Rhie et al. 2021), as outlined in Table 1. Briefly, PacBio HiFi long reads were processed using hifiasm (Cheng et al. 2021; 2022) producing a set of primary contigs representing the initial haploid assembly and separating alternative haplotypic variants. Primary contigs were then processed with purge_dups (Guan et al. 2020) to identify residual haplotype duplication in the assembly. Such duplicated sequences were moved to the alternate assembly that was then exposed to a second round of purge_dups to obtain the final set of non-redundant haplotypic variants. Primary contigs were anchored to scaffolds using Bionano optical maps, adjusting the gap size according to the observed optical distance with the bionano_solve pipeline v3.6.1 (Chan et al. 2018). A second round of scaffolding was performed using Hi-C data. Paired-end reads were aligned to the primary assembly using the Arima genomics' pipeline (https://github.com/ArimaGenomics/mapping_pipeline) and the obtained contact data was used to guide the scaffolding procedure using salsa2 (Ghurye et al. 2017; 2019). Hi-C contact maps were generated and visually inspected using PretextSuite (https://github.com/wtsi-hpag/PretextView; https://github.com/wtsi-hpag/PretextMap; https://github.com/wtsi-hpag/PretextSnapshot) before and after the last scaffolding step. The resulting primary and alternate assemblies were screened for residual contaminations (Howe et al. 2021) and manual curation was performed on the primary assembly using the gEVAL browser release 73 (Howe et al. 2021), PretextView and HiGlass (Kerpedjiev et al. 2018) to anchor scaffolds to chromosomes and check their coherence.

Genome size estimation and quality assessment

We estimated the genome size from the PacBio HiFi reads using a k-mer based approach. The distribution of k-mers of length 21 was generated using meryl v1.3 (Miller et al. 2008) and Genomescope2.0 (Ranallo-Benavidez, Jaron, and Schatz 2020) was subsequently used to infer the genome length, genome-wide heterozygosity and error rate.

We assessed the quality of our genome assembly using two independent methods. First, we used the BUSCO quality control tool to check for genome completeness using a set of conserved single-copy orthologous genes. We ran BUSCO v5.3.2 (Manni et al. 2021) in the

genome mode with default parameters on the tetrapod dataset (tetrapoda_odb10) that contains 5,310 orthologous genes. Second, we used Mercury v1.3 (Rhie et al. 2020) to estimate the base level accuracy (QV) and the assembly completeness comparing the k-mers in the assembly and those observed in the HiFi reads. All assembly metrics were computed using gfastats v1.2.3 (Formenti, Abueg, et al. 2022).

Identification of repetitive elements and gene annotation

To identify repetitive elements, we first generated a de novo repeat library using the Extensive de-novo TE Annotator (EDTA) v1.9.9 (Ou et al. 2019) and DeepTE (Yan, Bombarely, and Li 2020) to refine classifications within this library. We then used the final library to mask the genome with RepeatMasker v4.1.2 (Smit, Hubley, and Green, n.d.). We used the same pipeline to identify repeats in the genome of *Podarcis muralis* (assembly PodMur_1.0; Andrade et al. 2019).

For gene prediction, we first downloaded RNA-seq reads available on NCBI from various tissues of closely related species (four species of the genus *Podarcis*; See Supplementary Table 1). Quality control and trimming for adapters and low-quality bases (quality score <20) of the raw reads were performed using fastqc v0.11.8 (Andrews 2010) and TrimGalore v0.5.0 (https://github.com/FelixKrueger/TrimGalore), respectively. High-quality reads were then mapped to the soft-masked assembly with hisat2 v2.1.0 (Kim, Langmead, and Salzberg 2015), and sorted with samtools v1.10 (Li et al. 2009). All the BAM files were filtered to remove invalid splice junctions with Portcullis v1.1.2 (Mapleson et al. 2018). Filtered RNA-seq alignments were passed to Braker v2.1.6 (Hoff et al. 2016; 2019), together with amino acid sequences of the whole exome of 22 closely related species from the order Squamata belonging to 11 families including three Lacertidae (*Podarcis muralis*, *Lacerta agilis* and *Zootoca vivipara*; See Supplementary Table 2). The Braker gene prediction pipeline was run with the options "--softmasking --prg=gth --gth2traingenes". The resulting gene set was further filtered by evidence, keeping only gene predictions supported by RNA-seq or protein evidence using a BRAKER2 script (selectSupportedSubsets.py). The completeness of the final gene set was checked with BUSCO v5.3.2 (Manni et al. 2021) using the longest transcript of each gene as the representative transcript.

Mitochondrial genome sequencing and assembly

To characterize the entire sequence of the mitochondrial DNA via Sanger sequencing, we designed four different, and partially overlapping, amplicons of expected length between 4

and 7.3 Kb. Primers were designed based on mitochondrial DNA sequences of congeneric species (*Podarcis siculus* NC_011609.1, *Podarcis muralis* NC_011607 and NC_011609). Amplifications were carried out starting from 50 ng of extracted DNA, in a 50 µl reaction with 0.2 µM primers and 1.25 u of PrimeSTAR GXL DNA Polymerase. Amplification primers and additional internal primers were used for Sanger sequencing reactions (See Supplementary Table 3). Fragments were visually inspected and manually assembled to reconstruct the mitochondrial sequence.

Comparative analyses with *Podarcis muralis*

We performed a synteny comparison with the *Podarcis muralis* assembly (PodMur_1.0; Andrade et al. 2019), the only chromosome-scale assembly presently available for the *Podarcis* genus. Phylogenetic reconstructions based on whole-genome data suggest that the two species diverged ~18 Mya during Miocene (Yang et al. 2021). We used minimap2 (Li 2018) to map the genome assembly of *Podarcis raffonei* to the genome reference of *Podarcis muralis* allowing a maximum sequence divergence of 5% (parameter -x asm20). We then filtered the alignment by mapping quality (>60) and length of the mapped fragments (>1 Mb) and plotted the alignment between the 18 autosomes and Z sexual chromosome (the W chromosome being absent from the *Podarcis muralis* assembly) using Circos v0.69-8 (Krzywinski et al. 2009). Synteny between the two species was finally used to annotate the scaffolds of the *Podarcis raffonei* assembly as chromosomes.

Results

The final genome size (1.51 Gb) is in agreement with the size estimated from the k-mer analysis with Genomescope2.0 (Figure 1B) and very close to the genome size of *Podarcis muralis* (1.51 Gb, Andrade et al., 2019). The k-mer spectrum shows a bimodal distribution with two major peaks, at ~20 and ~40-fold coverage, corresponding to heterozygous and homozygous states, respectively. Based on PacBio HiFi reads, we estimated a 0.159% sequencing error rate and a 0.177% nucleotide heterozygosity rate (Figure 1B). The mitochondrial genome size is 17,038 bp, in agreement with the mitochondrial genome size of other species of *Podarcis* (17,311 bp for *P. muralis* and 17,297 bp for *P. siculus*; Podnar et al., 2009). The primary assembly contains 28 scaffolds for a total length of 1.51 Gb, with a contig N50 of 61.4 Mb, a scaffold N50 of 93.6 Mb, a longest contig size of 104.8 Mb, and a longest

scaffold size of 139.1 Mb (Table 2; Figure 1C). The alternate assembly contains 4,811 scaffolds spanning 182 Mb, having a N50 of 38.4 Kb.

This assembly is highly contiguous, as shown in the Hi-C contact map (Figure 1D), with the 20 first scaffolds being of chromosome length and corresponding to the 18 autosomes and the two sexual chromosomes Z and W (See Supplementary Table 4). The sequencing depth of the HiFi reads along chromosomes is approximately uniform and does not reveal discrepancies in the assembly (See Supplementary Figure 2). The completeness of the assembly is very high, with a BUSCO completeness score of 97.3% ([Single-copy:96.0%,Duplicated:1.3%],Fragmented:0.6%,Missing:2.1%) using the tetrapod gene set and a k-mer completeness of 99.5%. Per base quality (QV) as estimated by Merqury is 62, corresponding to less than one incorrect nucleotide per megabase.

In total, 22,463 protein-coding genes were predicted. The BUSCO completeness of the gene annotation using the same tetrapod gene set was 92.1% ([Single-copy:91.1%,Duplicated:1.0%],Fragmented:3.9%,Missing:4.0%). The identification of repetitive elements resulted in a 48.2% repeat content, falling within the range of repeat contents for other squamate species (24.4%-73.0%; Pasquesi et al. 2018). In Lacertidae and Teiidae, the repeat content was estimated to be 45.1% and 44.5%for *Podarcis muralis* and *Salvator merianae* (Roscito et al. 2018), respectively (See Supplementary Tables 5 and 6). The major class of repetitive elements was constituted by LTR elements and DNA transposons (See Supplementary Table 5).

The alignment of the genomes of *P. muralis and P. raffonei* revealed a very high congruency in the chromosomal organisation (Figure 2). The only chromosomal segment that did not map to the homologous chromosome from the other species was a 1.5 Mb segment of the chromosome 2 of *P. raffonei* that mapped to the chromosome 18 of *P. muralis*. We analysed the depth of coverage profile and the reads mapping in the edges of this segment of chromosome 2 in *P. raffonei* and did not find any discrepancies in the assembly (See Supplementary Figure 3). The two species have a similar number of genes (24,656 protein-coding genes were predicted in *P. muralis*; Andrade et al. 2019).

Discussion

We present here the first chromosome-scale genome assembly for the Aeolian wall lizard (scaffold N50 of 93.6 Mb). Several metrics indicate that our genome assembly possesses a

very high quality being chromosome-scale, accurate and complete. It constitutes a useful resource for squamates, a group composed of ~11,000 species for which only 29 high-quality genome assemblies are currently available (Card, Jennings, and Edwards 2023). In comparison to the other squamates, the *P. raffonei* assembly has a high scaffold N50 and the highest BUSCO completeness score (See Supplementary Table 6).

The alignment between the genomes of *P. raffonei* and *P. muralis* showed a very high synteny, suggesting that both assemblies are structurally accurate and that the two species share a very similar chromosomal organisation. Only one segment of the chromosome 2 of *P. raffonei* mapped to the chromosome 18 of *P. muralis*. This finding could be a biological chromosomal rearrangement between these two species (that belong to distinct clades of the genus *Podarcis*; Salvi et al. 2021; Yang et al. 2021) or a disjunction in the genome assembly of *P. muralis*.

The genome assembly of the Aeolian wall lizard, one of the most endangered vertebrate species in Europe, is a useful resource to better plan conservation efforts. Previous studies have highlighted that the Aeolian wall lizard exhibits low levels of genetic diversity and that the populations inhabiting different islands show a very reduced gene flow, constituting additional threats to this species (Capula 2004). Accordingly, our genome assembly suggests a very low heterozygosity (0.177% as estimated by Genomescope), the lowest value documented among seven species belonging to distinct squamate families (See Supplementary Table 6). The genome resequencing of several individuals from different islands is in progress to comprehensively characterise the genetic diversity of this species and evaluate its extinction risk.

Fundings

This work was supported by the University of Ferrara (Italy) and funded by the MIUR PRIN 2017 grant 201794ZXTL to G.B.

Acknowledgements

We thank the mountain guide Lorenzo Inzigneri for its help in the sampling in La Canna. Sampling authorization was given under Prot. 45382, Rif. 35644/2019.

Data availability

Raw sequencing data, primary genome assembly and mitochondrial DNA sequence are available under NCBI BioProject PRJNA916649 and PRJNA839511. Gene annotations and RepeatMasker output are available on https://zenodo.org/record/7473296.

References

Andrade, Pedro, Catarina Pinho, Guillem Pérez i de Lanuza, Sandra Afonso, Jindřich Brejcha, Carl-Johan Rubin, Ola Wallerman, et al. 2019. 'Regulatory Changes in Pterin and Carotenoid Genes Underlie Balanced Color Polymorphisms in the Wall Lizard'. *Proceedings of the National Academy of Sciences* 116 (12): 5633–42. https://doi.org/10.1073/pnas.1820320116.

Andrews, S. 2010. 'FastQC: A Quality Control Tool for High Throughput Sequence Data. Available Online at: Http://Www.Bioinformatics.Babraham.Ac.Uk/Projects/Fastqc/'. 2010.

Bonardi, Anna, Gentile Francesco Ficetola, Edoardo Razzetti, Claudia Canedoli, Mattia Falaschi, Elia Lo Parrino, Noemi Rota, Emilio Padoa-Schioppa, and Roberto Sindaco. 2022. 'ReptIslands: Mediterranean Islands and the Distribution of Their Reptile Fauna'. *Global Ecology and Biogeography* 31 (5): 840–47. https://doi.org/10.1111/geb.13490.

Capula, Massimo. 2004. 'Low Genetic Variation in a Critically Endangered Mediterranean Lizard: Conservation Concerns for *Podarcis Raffonei* (Reptilia, Lacertidae)'. *Italian Journal of Zoology* 71 (sup1): 161–66. https://doi.org/10.1080/11250003.2004.9525553.

Capula, Massimo, Luca Luiselli, Marco A. Bologna, and Arianna Ceccarelli. 2002. 'The Decline of the Aeolian Wall Lizard, *Podarcis Raffonei*: Causes and Conservation Proposals'. *Oryx* 36 (1): 66–72. https://doi.org/10.1017/S0030605302000108.

Card, Daren C., W. Bryan Jennings, and Scott V. Edwards. 2023. 'Genome Evolution and the Future of Phylogenomics of Non-Avian Reptiles'. *Animals* 13 (3): 471. https://doi.org/10.3390/ani13030471.

Cascio, Pietro Lo, Flavia Grita, Livia Guarino, and Claudia Speciale. 2014. 'A Little Is Better than None: New Insights into the Natural History of the Aeolian Wall Lizard *Podarcis Raffonei* from La Canna Stack (Squamata Sauria)', 12.

Chan, Saki, Ernest Lam, Michael Saghbini, Sven Bocklandt, Alex Hastie, Han Cao, Erik Holmlin, and Mark Borodkin. 2018. 'Structural Variation Detection and Analysis Using Bionano Optical Mapping'. In *Copy Number Variants: Methods and Protocols*, edited by Derek M. Bickhart, 193–203. Methods in Molecular Biology. New York, NY: Springer. https://doi.org/10.1007/978-1-4939-8666-8_16.

Cheng, Haoyu, Gregory T. Concepcion, Xiaowen Feng, Haowen Zhang, and Heng Li. 2021. 'Haplotype-Resolved de Novo Assembly Using Phased Assembly Graphs with Hifiasm'. *Nature Methods* 18 (2): 170–75. https://doi.org/10.1038/s41592-020-01056-5.

Cheng, Haoyu, Erich D. Jarvis, Olivier Fedrigo, Klaus-Peter Koepfli, Lara Urban, Neil J. Gemmell, and Heng Li. 2022. 'Haplotype-Resolved Assembly of Diploid Genomes without Parental Data'. *Nature Biotechnology* 40 (9): 1332–35. https://doi.org/10.1038/s41587-022-01261-x.

Ficetola, Gentile Francesco, Benedetta Barzaghi, Andrea Melotto, Martina Muraro, Enrico Lunghi, Claudia Canedoli, Elia Lo Parrino, et al. 2018. 'N -Mixture Models Reliably Estimate the Abundance of Small Vertebrates'. *Scientific Reports* 8 (1): 10357. https://doi.org/10.1038/s41598-018-28432-8.

Ficetola, Gentile Francesco, Iolanda Silva-Rocha, Miguel A Carretero, Leonardo Vignoli, Roberto Sacchi, Andrea Melotto, Stefano Scali, and Daniele Salvi. 2021. 'Status of the Largest Extant Population of the Critically Endangered Aeolian Lizard *Podarcis Raffonei* (Capo Grosso, Vulcano Island)'. *PLOS ONE*, 15.

Formenti, Giulio, Linelle Abueg, Angelo Brajuka, Nadolina Brajuka, Cristóbal Gallardo-Alba, Alice Giani, Olivier Fedrigo, and Erich D Jarvis. 2022. 'Gfastats: Conversion, Evaluation and Manipulation of Genome Sequences Using Assembly Graphs'. *Bioinformatics* 38 (17): 4214–16. https://doi.org/10.1093/bioinformatics/btac460.

Formenti, Giulio, Kathrin Theissinger, Carlos Fernandes, Iliana Bista, Aureliano Bombarely, Christoph Bleidorn, Claudio Ciofi, et al. 2022. 'The Era of Reference Genomes in Conservation Genomics'. *Trends in Ecology & Evolution* 37 (3): 197–202. https://doi.org/10.1016/j.tree.2021.11.008.

Geneva, Anthony J., Sungdae Park, Dan G. Bock, Pietro L. H. de Mello, Fatih Sarigol, Marc Tollis, Colin M. Donihue, et al. 2022. 'Chromosome-Scale Genome Assembly of the Brown Anole (*Anolis Sagrei*), an Emerging Model Species'. *Communications Biology* 5 (1): 1–13. https://doi.org/10.1038/s42003-022-04074-5.

Ghurye, Jay, Mihai Pop, Sergey Koren, Derek Bickhart, and Chen-Shan Chin. 2017. 'Scaffolding of Long Read Assemblies Using Long Range Contact Information'. *BMC Genomics* 18 (1): 527. https://doi.org/10.1186/s12864-017-3879-z.

Ghurye, Jay, Arang Rhie, Brian P. Walenz, Anthony Schmitt, Siddarth Selvaraj, Mihai Pop, Adam M. Phillippy, and Sergey Koren. 2019. 'Integrating Hi-C Links with Assembly Graphs for Chromosome-Scale Assembly'. *PLOS Computational Biology* 15 (8): e1007273. https://doi.org/10.1371/journal.pcbi.1007273.

Gippoliti, Spartaco, Massimo Capula, Gentile F. Ficetola, Daniele Salvi, and Franco Andreone. 2017. 'Threatened by Legislative Conservationism? The Case of the Critically Endangered Aeolian Lizard'. *Frontiers in Ecology and Evolution* 5. https://doi.org/10.3389/fevo.2017.00130.

Guan, Dengfeng, Shane A. McCarthy, Jonathan Wood, Kerstin Howe, Yadong Wang, and Richard Durbin. 2020. 'Identifying and Removing Haplotypic Duplication in Primary Genome Assemblies'. *Bioinformatics (Oxford, England)* 36 (9): 2896–98. https://doi.org/10.1093/bioinformatics/btaa025.

Hoff, Katharina J., Simone Lange, Alexandre Lomsadze, Mark Borodovsky, and Mario Stanke. 2016. 'BRAKER1: Unsupervised RNA-Seq-Based Genome Annotation with GeneMark-ET and AUGUSTUS'. *Bioinformatics (Oxford, England)* 32 (5): 767–69. https://doi.org/10.1093/bioinformatics/btv661.

Hoff, Katharina J., Alexandre Lomsadze, Mark Borodovsky, and Mario Stanke. 2019. 'Whole-Genome Annotation with BRAKER'. *Methods in Molecular Biology (Clifton, N.J.)* 1962: 65–95. https://doi.org/10.1007/978-1-4939-9173-0_5.

Howe, Kerstin, William Chow, Joanna Collins, Sarah Pelan, Damon-Lee Pointon, Ying Sims, James Torrance, Alan Tracey, and Jonathan Wood. 2021. 'Significantly Improving the Quality of Genome Assemblies through Curation'. *GigaScience* 10 (1): giaa153. https://doi.org/10.1093/gigascience/giaa153.

IUCN. 2009. '*Podarcis Raffonei*. The IUCN Red List of Threatened Species'. https://doi.org/e.T61552A12514822.

Kerpedjiev, Peter, Nezar Abdennur, Fritz Lekschas, Chuck McCallum, Kasper Dinkla, Hendrik Strobelt, Jacob M. Luber, et al. 2018. 'HiGlass: Web-Based Visual Exploration and Analysis of Genome Interaction Maps'. *Genome Biology* 19 (1): 125. https://doi.org/10.1186/s13059-018-1486-1.

Kim, Daehwan, Ben Langmead, and Steven L. Salzberg. 2015. 'HISAT: A Fast Spliced Aligner with Low Memory Requirements'. *Nature Methods* 12 (4): 357–60. https://doi.org/10.1038/nmeth.3317.

Krzywinski, Martin I., Jacqueline E. Schein, Inanc Birol, Joseph Connors, Randy Gascoyne, Doug Horsman, Steven J. Jones, and Marco A. Marra. 2009. 'Circos: An Information Aesthetic for Comparative Genomics'. *Genome Research*, June. https://doi.org/10.1101/gr.092759.109.

Li, Heng. 2018. 'Minimap2: Pairwise Alignment for Nucleotide Sequences'. Edited by Inanc Birol. *Bioinformatics* 34 (18): 3094–3100. https://doi.org/10.1093/bioinformatics/bty191.

Li, Heng, Bob Handsaker, Alec Wysoker, Tim Fennell, Jue Ruan, Nils Homer, Gabor Marth, Goncalo Abecasis, Richard Durbin, and 1000 Genome Project Data Processing Subgroup. 2009. 'The Sequence Alignment/Map Format and SAMtools'. *Bioinformatics (Oxford, England)* 25 (16): 2078–79. https://doi.org/10.1093/bioinformatics/btp352.

Manni, Mosè, Matthew R Berkeley, Mathieu Seppey, Felipe A Simão, and Evgeny M Zdobnov. 2021. 'BUSCO Update: Novel and Streamlined Workflows along with Broader and Deeper Phylogenetic Coverage for Scoring of Eukaryotic, Prokaryotic, and Viral Genomes'. *Molecular Biology and Evolution* 38 (10): 4647–54. https://doi.org/10.1093/molbev/msab199.

Mapleson, Daniel, Luca Venturini, Gemy Kaithakottil, and David Swarbreck. 2018. 'Efficient and Accurate Detection of Splice Junctions from RNA-Seq with Portcullis'. *GigaScience* 7 (12). https://doi.org/10.1093/gigascience/giy131.

Miller, Jason R., Arthur L. Delcher, Sergey Koren, Eli Venter, Brian P. Walenz, Anushka Brownley, Justin Johnson, Kelvin Li, Clark Mobarry, and Granger Sutton. 2008. 'Aggressive Assembly of Pyrosequencing Reads with Mates'. *Bioinformatics (Oxford, England)* 24 (24): 2818–24. https://doi.org/10.1093/bioinformatics/btn548.

Ou, Shujun, Weija Su, Yi Liao, Kapeel Chougule, Jireh R. A. Agda, Adam J. Hellinga, Carlos Santiago Blanco Lugo, et al. 2019. 'Benchmarking Transposable Element Annotation Methods for Creation of a Streamlined, Comprehensive Pipeline'. *Genome Biology* 20 (1): 275. https://doi.org/10.1186/s13059-019-1905-y.

Pasquesi, Giulia I. M., Richard H. Adams, Daren C. Card, Drew R. Schield, Andrew B. Corbin, Blair W. Perry, Jacobo Reyes-Velasco, et al. 2018. 'Squamate Reptiles Challenge Paradigms of Genomic Repeat Element Evolution Set by Birds and Mammals'. *Nature Communications* 9 (1): 2774. https://doi.org/10.1038/s41467-018-05279-1.

Pinto, Brendan J, Shannon E Keating, Stuart V Nielsen, Daniel P Scantlebury, Juan D Daza, and Tony Gamble. 2022. 'Chromosome-Level Genome Assembly Reveals Dynamic Sex Chromosomes in

Neotropical Leaf-Litter Geckos (Sphaerodactylidae: *Sphaerodactylus*)'. *Journal of Heredity* 113 (3): 272–87. https://doi.org/10.1093/jhered/esac016.

Podnar, M., W. Pinsker, and W. Mayer. 2009. 'Complete Mitochondrial Genomes of Three Lizard Species and the Systematic Position of the Lacertidae (Squamata)*'. *Journal of Zoological Systematics and Evolutionary Research* 47 (1): 35–41. https://doi.org/10.1111/j.1439-0469.2008.00515.x.

Ranallo-Benavidez, T. Rhyker, Kamil S. Jaron, and Michael C. Schatz. 2020. 'GenomeScope 2.0 and Smudgeplot for Reference-Free Profiling of Polyploid Genomes'. *Nature Communications* 11 (1): 1432. https://doi.org/10.1038/s41467-020-14998-3.

Rhie, Arang, Shane A. McCarthy, Olivier Fedrigo, Joana Damas, Giulio Formenti, Sergey Koren, Marcela Uliano-Silva, et al. 2021. 'Towards Complete and Error-Free Genome Assemblies of All Vertebrate Species'. *Nature* 592 (7856): 737–46. https://doi.org/10.1038/s41586-021-03451-0.

Rhie, Arang, Brian P. Walenz, Sergey Koren, and Adam M. Phillippy. 2020. 'Merqury: Reference-Free Quality, Completeness, and Phasing Assessment for Genome Assemblies'. *Genome Biology* 21 (1): 245. https://doi.org/10.1186/s13059-020-02134-9.

Roscito, Juliana G, Katrin Sameith, Martin Pippel, Kees-Jan Francoijs, Sylke Winkler, Andreas Dahl, Georg Papoutsoglou, Gene Myers, and Michael Hiller. 2018. 'The Genome of the Tegu Lizard *Salvator Merianae*: Combining Illumina, PacBio, and Optical Mapping Data to Generate a Highly Contiguous Assembly'. *GigaScience* 7 (12): giy141. https://doi.org/10.1093/gigascience/giy141.

Salvi, Daniele, Catarina Pinho, Joana Mendes, and D. James Harris. 2021. 'Fossil-Calibrated Time Tree of Podarcis Wall Lizards Provides Limited Support for Biogeographic Calibration Models'. *Molecular Phylogenetics and Evolution* 161 (August): 107169. https://doi.org/10.1016/j.ympev.2021.107169.

Smit, A.F.A., R. Hubley, and P. Green. n.d. 'RepeatMasker'. http://repeatmasker.org.

Yan, Haidong, Aureliano Bombarely, and Song Li. 2020. 'DeepTE: A Computational Method for de Novo Classification of Transposons with Convolutional Neural Network'. Edited by Alfonso Valencia. *Bioinformatics* 36 (15): 4269–75. https://doi.org/10.1093/bioinformatics/btaa519.

Yang, Weizhao, Nathalie Feiner, Catarina Pinho, Geoffrey M. While, Antigoni Kaliontzopoulou, D. James Harris, Daniele Salvi, and Tobias Uller. 2021. 'Extensive Introgression and Mosaic Genomes of Mediterranean Endemic Lizards'. *Nature Communications* 12 (1): 2762. https://doi.org/10.1038/s41467-021-22949-9.

Tables and Figures:

Table 1: Pipeline and software used for the genome assembly.

Assembly	Software	Version
K-mer counting	meryl	1.3
Estimation of genome size and heterozygosity	GenomeScope2	2.0
De novo assembly (contigging)	HiFiasm	0.16.1-r375
Remove low-coverage, duplicated contigs	purge_dups	1.2.5
Scaffolding		
Bionano Scaffolding	bionano_solve	3.6.1
Hi-C mapping for SALSA	Arima Genomics mapping pipeline	Commit 2e74ea4
Hi-C Scaffolding	salsa2	2.3
Hi-C Contact map generation		
Short-read alignment	bwa	0.7.17
SAM/BAM processing	samtools	1.10
Pairs processing	bedtools	2.30
Contact map visualization	PretextView	0.2.2
	PretextMap	0.1.8
	PretextSnapshot	0.0.4
Genome assembly refinement		
Manual curation and contamination screening	gEVAL	release 73
Genome quality assessment		
Basic assembly metrics	gfastats	1.2.3
Assembly completeness	BUSCO	5.3.2
	Merqury	1.3
Repeat element identification		
Repeat identification	EDTA	1.9.9
	DeepTE	Commit babd65e
Repeat annotation	RepeatMasker	4.1.2
Gene annotation		
RNA-seq read quality control	fastqc	0.11.8
	TrimGalore	0.5.0
Mapping RNA-seq reads-genome	hisat2	2.1.0

Filtering splice junctions	Portcullis	1.1.2
Gene prediction	Braker	2.1.6
Comparison to *P. muralis*		
Genome–genome alignment	minimap2	2.22
Synteny visualisation	Circos	0.69-8

Table 2: Genome assembly statistics.

Measure	rPodRaf1
Total length	1.513 Gb
Number of scaffolds	28
Scaffold L50/N50	7 scaffolds ; 93.6 Mb
Longest scaffold	139.1 Mb
Number of contigs	53
Contig L50/N50	10 contigs ; 61.4 Mb
Longest contig	104.8 Mb
BUSCO completeness:	97.3%
Single copy	5,095
Duplicated	67
Fragmented	34
Missing	114
Total	5,310

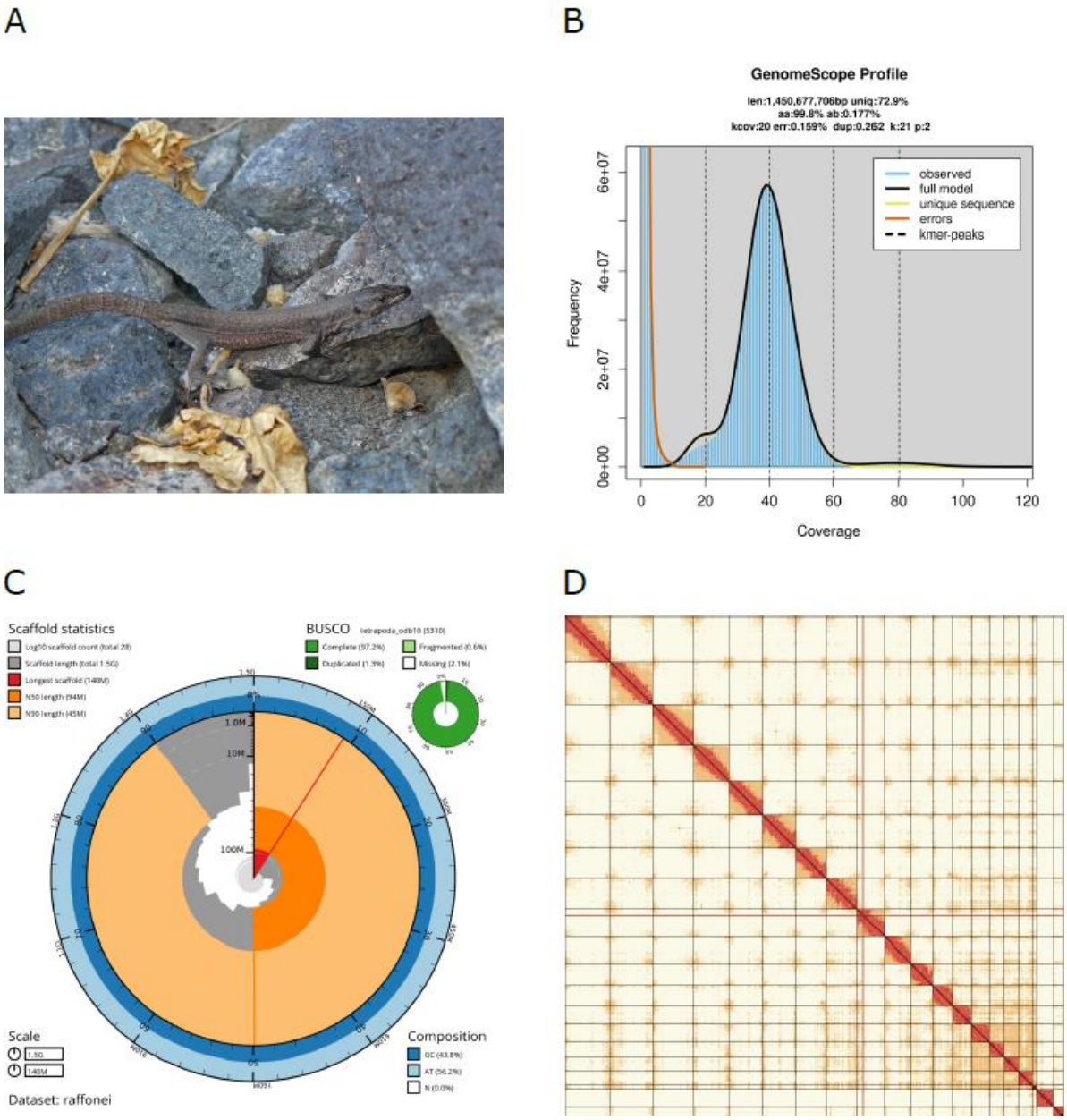

Figure 1: (A) Photography of an individual of *Podarcis raffonei*, on La Canna stack (Photo credit: Daniele Salvi), and visual overview of genome assembly metrics: (B) K-mer spectra output and corresponding genome size and heterozygosity estimated with GenomeScope2.0. (C) BlobToolKit Snail plot showing a graphical representation of the quality metrics presented in Table 2 for the *Podarcis raffonei* primary assembly (rPodRaf1.pri). (D) Hi-C contact map for the 20 scaffolds of the primary genome assembly generated with PretextSnapshot.

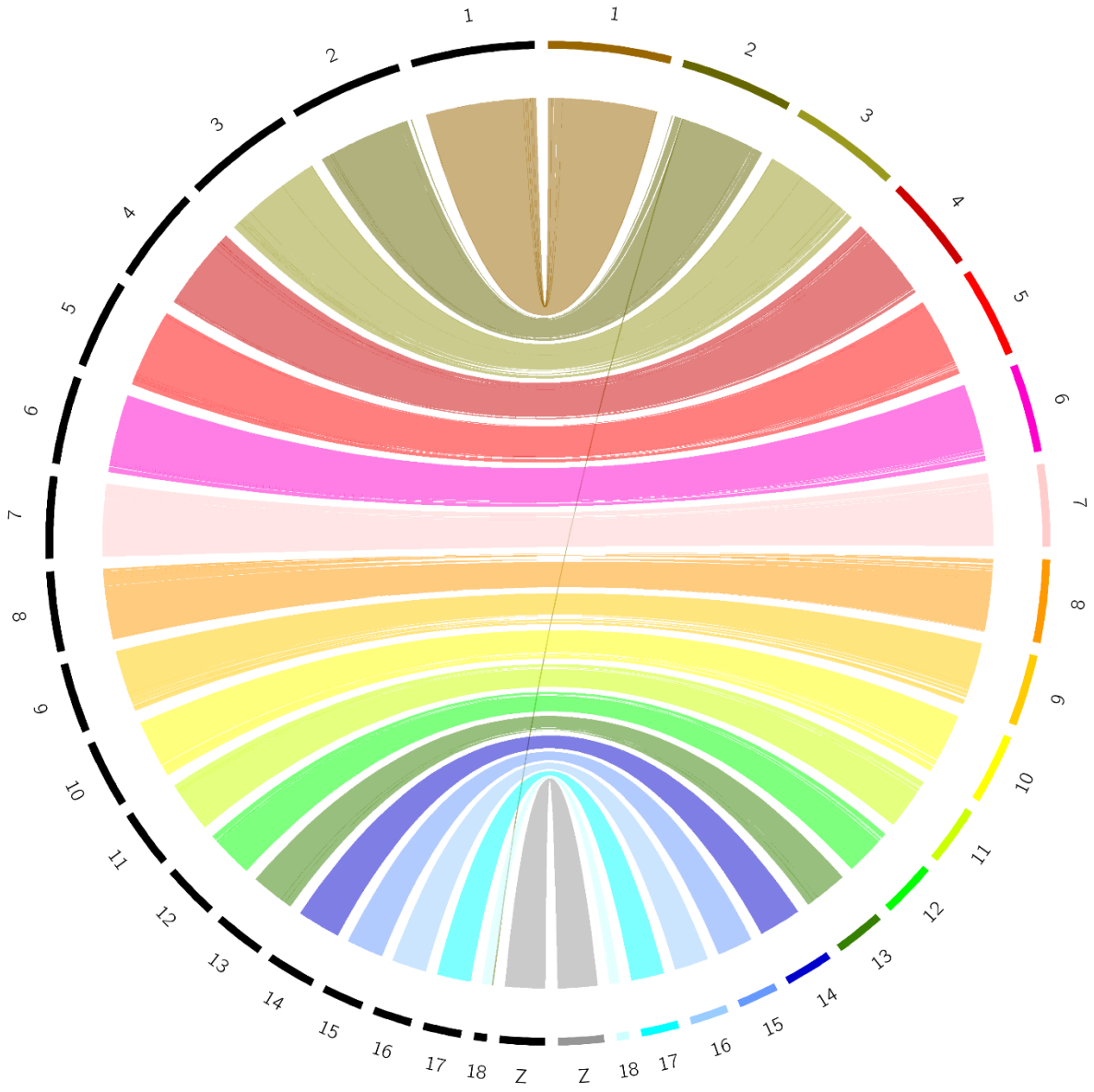

Figure 2: Comparison of the chromosomal structure between the 18 autosomal chromosomes and Z chromosome between *P. raffonei* (right) and *P. muralis* (left). The different colours correspond to the different chromosomes of *P. raffonei*. The chromosomes were aligned using minimap2 and the resulting alignment between fragments longer than 1 Mb is represented with a ribbon plot using Circos.