ARPHA Preprints

PENSOFT

Data Paper (Biosciences)

Author-formatted document posted on 08/04/2025

Published in a RIO article collection by decision of the collection editors.

DOI: https://doi.org/10.3897/arphapreprints.e155085

ERGA-BGE Reference Genome of the Western Montpellier Snake (*Malpolon monspessulanus*), a Key Species for Evolutionary and Venom Studies

Salvador Carranza, Daniel Fernández-Guiberteau, Laura Blasón, Javier Palma-Guerrero, Rosa Fernández, Rita Monteiro, Astrid Böhne, Laura Aguilera, Marta Gut, Francisco Câmara Ferreira, Fernando Cruz, Jèssica Gómez-Garrido, Tyler Alioto, Leanne Haggerty, Swati Sinha, Fergal Martin, Diego De Panis

ARPHA Preprints *Author-formatted document posted on 08/04/2025.* DOI: https://doi.org/10.3897/arphapreprints.e155085

GENOME REPORT

ERGA-BGE Reference Genome of the Western Montpellier Snake (*Malpolon monspessulanus*), a Key Species for Evolutionary and Venom Studies

Salvador Carranza[1], Daniel Fernández-Guiberteau[2], Laura Blasón[2], Javier Palma Guerrero[1], Rosa Fernández[1], Rita Monteiro[3], Astrid Böhne[3], Laura Aguilera[4,5], Marta Gut[4,5], Francisco Câmara Ferreira[4,5], Fernando Cruz[4,5], Jèssica Gómez-Garrido[4,5], Tyler S. Alioto[4,5], Leanne Haggerty[6], Swati Sinha[6], Fergal Martin[6], Diego De Panis[7,8*]

[1] Institut de Biologia Evolutiva (CSIC-Universitat Pompeu Fabra), Passeig marítim de la Barceloneta 37-49, 08003, Barcelona, Spain
[2] Centre de Recerca i Educació Ambiental de Calafell (CREAC, GRENP-Ajuntament de Calafell), C/ Dinamarca 8, 43882, Segur de Calafell, Tarragona, Spain
[3] Leibniz Institute for the Analysis of Biodiversity Change, Museum Koenig Bonn, Adenauerallee 127, 53113 Bonn, Germany
[4] Centro Nacional de Análisis Genómico (CNAG), C/Baldiri Reixac, 4, 08028 Barcelona, Spain
[5] Universitat de Barcelona (UB), 08028, Barcelona, Spain
[6] European Molecular Biology Laboratory, European Bioinformatics Institute, Wellcome Genome Campus, Hinxton, Cambridge, UK
[7] Leibniz Institute for Zoo and Wildlife Research, Alfred-Kowalke-Str 17, 10315 Berlin, Germany
[8] Berlin Center for Genomics in Biodiversity Research (BeGenDiv), Königin-Luise-Str 6-8, 14195 Berlin, Germany
* To whom correspondence should be addressed: panis@izw-berlin.de

Abstract

Malpolon monspessulanus is a large, rear-fanged snake widely distributed across Mediterranean habitats, where it plays an essential ecological role by controlling populations of small vertebrates. The reference genome of this species offers a crucial resource for uncovering the genomic foundations underlying venom evolution in rear-fanged snakes. A total of 23 contiguous chromosomal pseudomolecules (21 autosomes and two sex chromosomes) were assembled from the genome sequence. This chromosome-level assembly encompasses 1.8 Gb, composed of 140 contigs and 60 scaffolds, with contig and scaffold N50 values of 32.7 Mb and 121.9 Mb, respectively.

Keywords

Malpolon monspessulanus, genome assembly, genome annotation, European Reference Genome Atlas, Biodiversity Genomics Europe, Earth Biogenome Project, Psammophiidae, Western Montpellier snake

ARPHA Preprints *Author-formatted document posted on 08/04/2025.* DOI: https://doi.org/10.3897/arphapreprints.e155085

Introduction

Malpolon monspessulanus, commonly known as the Western Montpellier snake, is a large, slender, and fast-moving member of the Psammophiidae family, capable of exceeding 2 meters in length, making it one of Europe's largest snakes. It has large eyes and a distinctive, prominent ridge formed by the preocular and supraocular scales, resulting in a concave top of the head. Its long, smooth-scaled body is typically olive or brown, with lighter coloration on the underside (Speybroeck et al., 2016). Though harmless to humans, it possesses rear fangs and produces a mild venom, primarily used to subdue small prey such as lizards and rodents (Pleguezuelos & Salvador Milla, 2021).

Widely distributed across the southwestern Mediterranean, *M. monspessulanus* ranges from extreme northwest Italy and southeastern France across the entire Iberian Peninsula to the Mediterranean-climate zones of northwestern Africa, including Morocco and Algeria, extending as far east as Algiers (Carranza et al., 2006). This species thrives in a variety of habitats, including scrublands, rocky areas, and agricultural fields, where it serves as a top predator, maintaining balance of small vertebrates such as lizards, rodents, and birds (Speybroeck et al., 2016). Its presence also supports biodiversity, as it serves as prey for larger predators, contributing to the overall health and functionality of Mediterranean ecosystems (Pleguezuelos & Salvador Milla, 2021).

Listed as Least Concern on the IUCN Red List (Martínez-Solano et al., 2008), the Western Montpellier snake benefits from a wide distribution, tolerance to moderate habitat disturbance, and presumed large populations. Its adaptability to various environments and stable population trends contribute to its current conservation status (Speybroeck et al., 2016).

The Western Montpellier snake also hosts a variety of parasites, including cestodes (*Diplopylidium acanthotetrum*), nematodes (*Kalicephalus viperae*, *Ophidascaris sp.*, *Hexametra quadricornis*), acanthocephalans (*Centrorhynchus aluconis*), and mites (*Ophionyssus natricis*) (Pleguezuelos & Salvador Milla, 2021).

Developing a high-quality reference genome for *M. monspessulanus* is crucial for understanding the genetic basis of its venom system, particularly the evolution of toxin gene families. Identifying these genes will provide insights into how venom components have evolved. This genomic resource will provide a foundation for studying the molecular evolution of venom in rear-fanged snakes, helping to uncover the genetic pathways that have shaped specific toxin families. Additionally, it will offer broader implications for understanding the genetic diversity and evolutionary history of venomous species. More generally, reptiles remain underrepresented among vertebrates in terms of high-quality genome assemblies, and snake phylogenomics is an especially active field where new data can significantly advance our understanding of their evolution (Srodawa et al., 2023).

The generation of this reference resource was coordinated by the European Reference Genome Atlas (ERGA) initiative's Biodiversity Genomics Europe (BGE) project, supporting ERGA's aims of promoting transnational cooperation to promote advances in the application of genomics technologies to protect and restore biodiversity (Mazzoni et al., 2023). This species falls within the regional reach of the Catalan Initiative for the Earth BioGenome Project (CBP), which is linked to ERGA (Corominas et al., 2024).

Materials & Methods

ERGA's sequencing strategy includes Oxford Nanopore Technology (ONT) and/or Pacific

ARPHA Preprints *Author-formatted document posted on 08/04/2025.* DOI: https://doi.org/10.3897/arphapreprints.e155085

Biosciences (PacBio) for long-read sequencing, along with Hi-C sequencing for chromosomal architecture, Illumina Paired-End (PE) for polishing (i.e. recommended for ONT-only assemblies), and RNA sequencing for transcriptomic profiling, to facilitate genome assembly and annotation.

Sample and Sampling Information

On July 16th, 2023, an adult female of *Malpolon monspessulanus* was collected by hand by Salvador Carranza and Daniel Fernández-Guiberteau in Parc de la Sinia, Calafell, Tarragona (Spain). Species identification was based on an identification key (Pleguezuelos & Salvador Milla, 2021). Sampling was authorized under permit number SF/0179/23 issued by Generalitat de Catalunya, Departament d'Acció Climàtica, Alimentació i Agenda Rural, Direcció General de Polítiques Ambientals i Medi Natural. Euthanasia was performed on site by first administering anesthetic agents (alfaxalone [Alfaxan], 10 mg/kg intramuscular), and once the animal was fully anesthetized, the spinal cord was severed at the neck to ensure death. Following euthanasia, the specimen's tissues (kidney, heart, liver, lung, muscle, ovary, pancreas, and blood) were immediately snap-frozen and stored in liquid nitrogen for long-term preservation.

Vouchering information

Physical reference materials for the sequenced individual (Figure 1) have been deposited in Museo Nacional de Ciencias Naturales (CSIC, mncn.csic.es/en), under the accession number MNCN52207.

Frozen reference tissue material of the sampled individual is available at the the biobank of the Museo Nacional de Ciencias Naturales under the voucher IDs MNCN-ADN-151757 and MNCN-ADN-151758.

Data Availability

Malpolon monspessulanus and the related genomic study were assigned to Tree of Life ID (ToLID) rMalMon1 and all sample, sequence, and assembly information are available under the umbrella BioProject PRJEB77784. The sample information is available at the following BioSample accessions: SAMEA114541112, SAMEA114541114, SAMEA114541115, SAMEA114541117, SAMEA114541118, and SAMEA114541123. The genome assembly is accessible from ENA under accession number GCA_964265115.1 and the annotated genome is available through the Ensembl Rapid Release page (https://projects.ensembl.org/erga-bge/). Sequencing data produced as part of this project are available from ENA at the following accessions: ERX12752244, ERX13166519, ERX13166520, ERX13166521, ERX13549168, ERX13549169, ERX13549170, and ERX13549171. Documentation related to the genome assembly and curation can be found in the ERGA Assembly Report (EAR) document available at github.com/ERGA-consortium/EARs/tree/main/Assembly_Reports/Malpolon_monspessulanus/rMalMon1. Further details and data about the project are hosted on the ERGA portal at portal.erga-biodiversity.eu/data_portal/184164.

Genetic Information

The estimated genome size, based on ancestral taxa, is 1.48G, while the estimation based on reads kmer profiling is 1.6 Gbp. This is a diploid genome with a haploid number of 22 chromosomes (2n=44), including WZ sex chromosomes in females. Information for this species was retrieved from Genomes on a Tree (Challis et al., 2023).

DNA/RNA processing

DNA was extracted from the pancreas using the Blood & Cell Culture DNA Midi Kit (Qiagen) following the manufacturer's instructions. DNA quantification was performed using a

ARPHA Preprints *Author-formatted document posted on 08/04/2025.* DOI: https://doi.org/10.3897/arphapreprints.e155085

Qubit dsDNA BR Assay Kit (Thermo Fisher Scientific), and DNA integrity was assessed using a Femto Pulse system (Genomic DNA 165 Kb Kit, Agilent). DNA was stored at 4ºC until use.

RNA was extracted using an RNeasy Mini Kit (Qiagen) according to the manufacturer's instructions. RNA was extracted from four different specimen body parts: lung, liver, ovary and pancreas. RNA quantification was performed using the Qubit RNA BR Kit and RNA integrity was assessed using a Bioanalyzer 2100 system (RNA 6000 Nano Kit, Agilent). RNA was stored at -80ºC until use.

Library Preparation and Sequencing

A long-read whole genome library was prepared using the SQK-LSK114 kit and sequenced on a PromethION P24 A series instrument (Oxford Nanopore Technologies). For short-read whole genome sequencing (WGS), a library was prepared using the KAPA Hyper Prep Kit (Roche). A Hi-C library preparation, using liver tissue, was conducted with the Dovetail Omni-C Kit (Cantata Bio) and further processed with the KAPA Hyper Prep Kit for Illumina sequencing (Roche). The RNA libraries were prepared with the KAPA mRNA Hyper Prep Kit (Roche). All the short-read libraries were sequenced on the Illumina NovaSeq 6000 instrument. In total, 93x Oxford Nanopore, 62x Illumina WGS shotgun, and 51x HiC data were sequenced to generate the assembly.

Genome Assembly Methods

{The genome was assembled using the CNAG CLAWS pipeline (Gomez-Garrido, 2024). Briefly, reads were preprocessed for quality and length using Trim Galore v0.6.7 and Filtlong v0.2.1, and initial contigs were assembled using NextDenovo v2.5.0, followed by polishing of the assembled contigs using HyPo v1.0.3 and scaffolding with YaHS v1.2a. Finally, assembled scaffolds were curated via manual inspection using Pretext v0.2.5 with Sanger's Rapid Assembly Curation Toolkit (gitlab.com/wtsi-grit/rapid-curation) to remove any false joins and incorporate any sequences not automatically scaffolded into their respective locations in the chromosomal pseudomolecules (or super-scaffolds). Finally, the mitochondrial genome was assembled as a single circular contig of {size} bp using the FOAM pipeline v0.5 (github.com/cnag-aat/FOAM) and included in the released assembly (accession number: OZ184096.1). Summary analysis of the released assembly was performed using the ERGA-BGE Genome Report ASM Galaxy workflow (De Panis, 2024b), incorporating tools such as BUSCO v5.5, Merqury v1.3, and others (see reference for the full list of tools).

Genome Annotation Methods

A gene set was generated using the Ensembl Gene Annotation system (Aken et al., 2016), primarily by aligning publicly available short-read RNA-seq data from BioSample: SAMEA114541112, SAMEA114541115, SAMEA114541118, and SAMEA114541123 to the genome. Gaps in the annotation were filled via protein-to-genome alignments of a select set of vertebrate proteins from UniProt (The UniProt Consortium, 2019), which had experimental evidence at the protein or transcript level. At each locus, data were aggregated and consolidated, prioritising models derived from RNA-seq data, resulting in a final set of gene models and associated non-redundant transcript sets. To distinguish true isoforms from fragments, the likelihood of each open reading frame (ORF) was evaluated against known vertebrate proteins. Low-quality transcript models, such as those showing evidence of fragmented ORFs, were removed. In cases where RNA-seq data were fragmented or absent, homology data were prioritised, favouring longer transcripts with strong intron support from short-read data. The resulting gene models were classified into three

ARPHA Preprints *Author-formatted document posted on 08/04/2025.* DOI: https://doi.org/10.3897/arphapreprints.e155085

categories: protein-coding, pseudogene, and long non-coding. Models with hits to known proteins and few structural abnormalities were classified as protein-coding. Models with hits to known proteins but displaying abnormalities, such as the absence of a start codon, non-canonical splicing, unusually small intron structures (<75 bp), or excessive repeat coverage, were reclassified as pseudogenes. Single-exon models with a corresponding multi-exon copy elsewhere in the genome were classified as processed (retrotransposed) pseudogenes. Models that did not fit any of the previously described categories did not overlap protein-coding genes, and were constructed from transcriptomic data were considered potential lncRNAs. Potential lncRNAs were further filtered to remove single-exon loci due to their unreliability. Putative miRNAs were predicted by performing a BLAST search of miRBase (Kozomara et al., 2019) against the genome, followed by RNAfold analysis (Gruber et al., 2008). Other small non-coding loci were identified by scanning the genome with Rfam (Kalvari et al., 2018) and passing the results through Infernal (Nawrocki & Eddy, 2013). Summary analysis of the released annotation was performed using the ERGA-BGE Genome Report ANNOT Galaxy workflow (De Panis, 2024a), incorporating tools such as AGAT v1.2, OMArk v0.3, and others (see reference for the full list of tools).

Results

Genome Assembly

The genome assembly has a total length of 1,752,167,081 bp in 61 scaffolds including the mitogenome (Figures 2 and 3), with a GC content of 42.27%. It features a contig N50 of 32,743,049 bp (L50=13) and a scaffold N50 of 121,923,953 bp (L50=5). There are 80 gaps, totaling 16,000 bp in cumulative size. The single-copy gene content analysis using the tetrapoda database with BUSCO resulted in 96.0% completeness (95.1% single and 0.9% duplicated). 95.47% of reads k-mers were present in the assembly and the assembly has a base accuracy Quality Value (QV) of 47.56 as calculated by Merqury.

Genome Annotation

The genome annotation consists of 16,508 protein-coding genes with an associated 41,078 transcripts, in addition to 51 pseudogenes, and 837 non-coding RNA genes of various types (Table 1). Using the longest isoform per transcript, the single-copy gene content analysis using the Tetrapoda database with BUSCO resulted in 91.1% completeness. Using the OMAmer Episquamata database for OMArk resulted in 94.75% completeness and 98.14% consistency (Table 2).

ARPHA Preprints *Author-formatted document posted on 08/04/2025.* DOI: https://doi.org/10.3897/arphapreprints.e155085

Figure 1. Image of the sequenced individual of *Malpolon monspessulanus*. The image, along with another, is available in ERGA's EBI BioImageArchive dataset (ebi.ac.uk/biostudies/bioimages/studies/S-BIAD1012?query=ERGA) under accession ID SAMEA114541114.

ARPHA Preprints *Author-formatted document posted on 08/04/2025.* DOI: https://doi.org/10.3897/arphapreprints.e155085

Table 1. Statistics from assembled gene models

	No. genes	No. transcripts	Mean* gene length (bp)	No. single-exon genes	Mean* exons per transcript
Protein-coding	16,508	41,078	36,338	652	14.5
Pseudogenes	51	51	13,194	5	13.0
lncRNA	363	431	10,069	86	2.5
snRNA	158	158	127	158	1.0
snoRNA	182	182	110	182	1.0
rRNA	52	52	185	52	1.0
miRNA	57	57	86	57	1.0
scRNA	15	15	194	15	1.0
Other non-coding	10	10	105-296	10	1.0-1.0

*Combined categories show the range of the mean values

Table 2. Annotation completeness and consistency scores calculated by BUSCO run in protein mode (Tetrapoda) and OMArk (Episquamata)

	Complete	Singular	Duplicated	Fragmented	Missing
BUSCO	4,840 (91.1%)	4,780 (90.0%)	60 (1.1%)	124 (2.3%)	346 (6.6%)
OMArk	12,621 (94.75%)	12,366 (92.84%)	255 (1.91%)	-	698 (5.24%)
	Consistent	**Inconsistent**	**Contaminants**	**Unknown**	
OMArk	16,201 (98.14%)	172 (1.04%)	0 (0.00%)	135 (0.82%)	

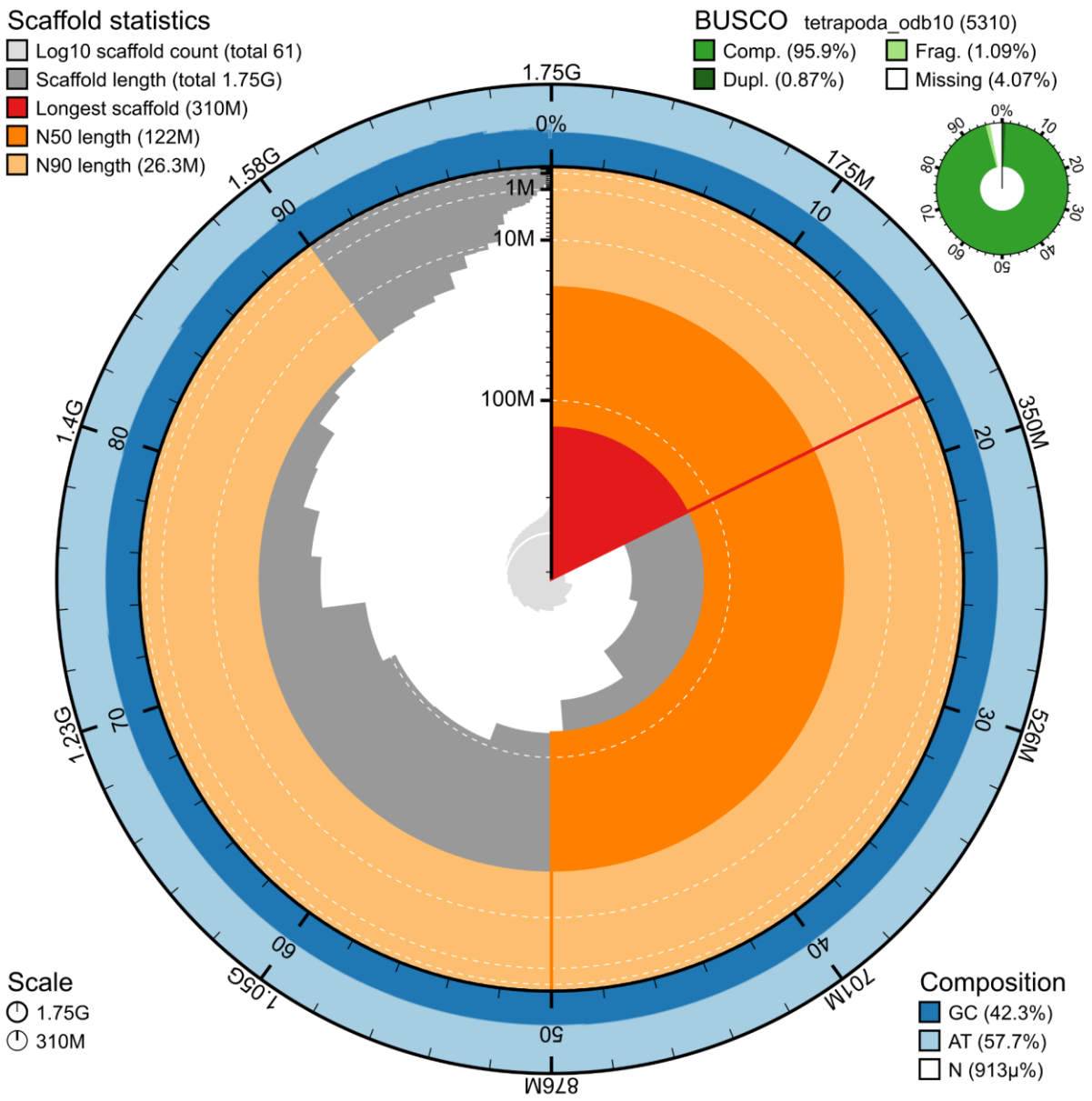

Figure 2. Snail plot summary of assembly statistics. The main plot is divided into 1,000 size-ordered bins around the circumference, with each bin representing 0.1% of the 1,752,167,081 bp assembly including the mitochondrial genome. The distribution of sequence lengths is shown in dark grey, with the plot radius scaled to the longest sequence present in the assembly (310,195,023 bp, shown in red). Orange and pale-orange arcs show the scaffold N50 and N90 sequence lengths (121,923,953 and 26,303,670 bp), respectively. The pale grey spiral shows the cumulative sequence 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 found in the assembled genome from the Tetrapoda database (odb10) is shown on the top right.

ARPHA Preprints *Author-formatted document posted on 08/04/2025.* DOI: https://doi.org/10.3897/arphapreprints.e155085

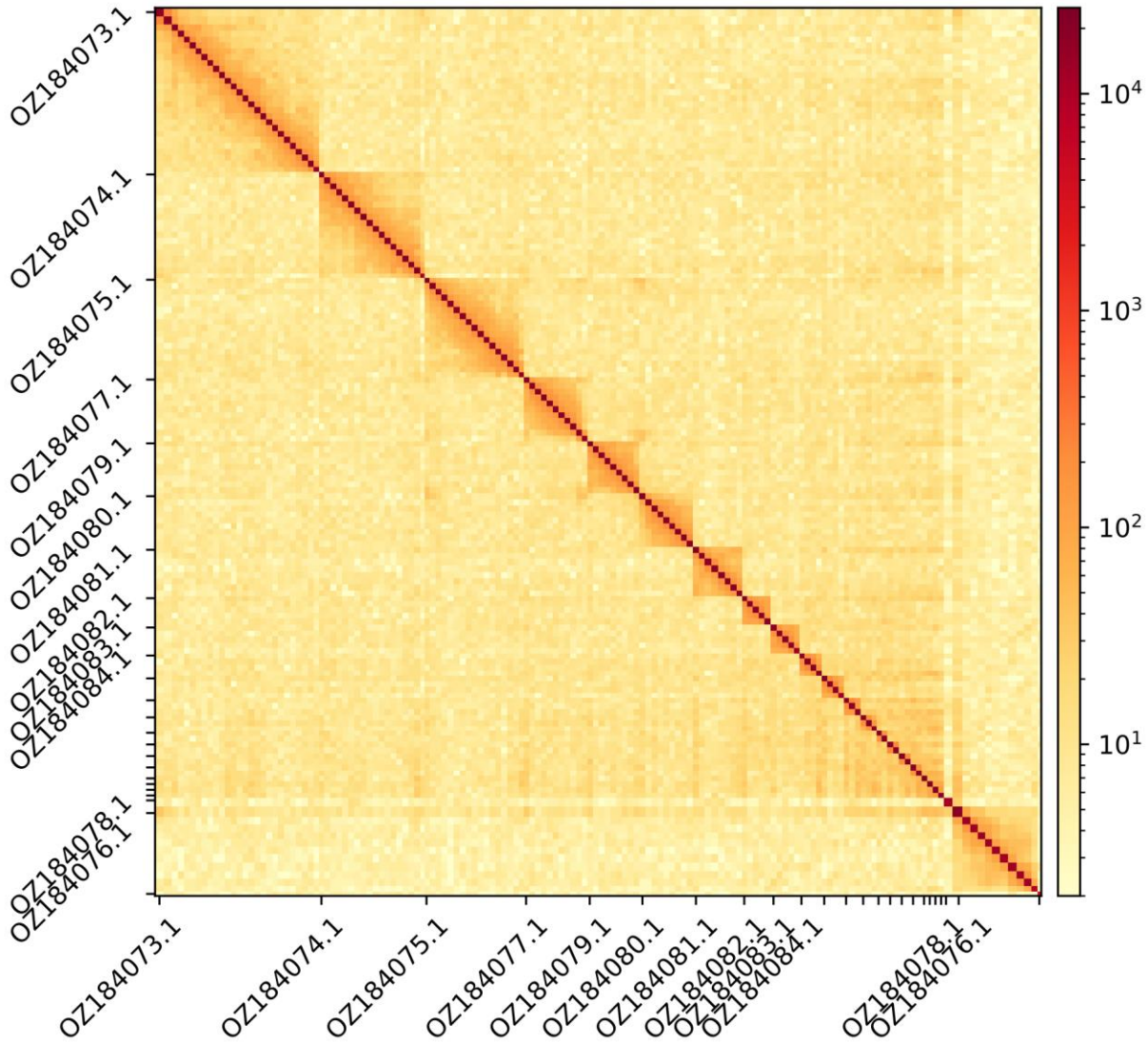

Figure 3. Hi-C contact map showing spatial interactions between regions of the genome. The diagonal corresponds to intra-chromosomal contacts, depicting chromosome boundaries. The frequency of contacts is shown on a logarithmic heatmap scale. Hi-C matrix bins were merged into a 200 kb bin size for plotting. Due to space constraints on the axes, only the GenBank names of the 10th largest autosomes and the WZ sex chromosomes are shown.

ARPHA Preprints *Author-formatted document posted on 08/04/2025.* DOI: https://doi.org/10.3897/arphapreprints.e155085

Acknowledgements

We would like to express our gratitude to Jhulyana López Caro and Eduardo Cáceres for their invaluable assistance in the field and laboratory. We acknowledge the support of the Freiburg Galaxy Team: Saim Momin and Björn Grüning, Bioinformatics, University of Freiburg (Germany), funded by the German Federal Ministry of Education and Research BMBF grant 031 A538A de.NBI-RBC and the Ministry of Science, Research and the Arts Baden-Württemberg (MWK) within the framework of LIBIS/de.NBI Freiburg. We would like to acknowledge the assembly reviewer, Tom Mathers from the Wellcome Sanger Institute.

Conflict of Interest

The authors declare no conflict of interest related to this study. The funding sources had no involvement in the study design, collection, analysis, or interpretation of data; in the writing of the manuscript; or in the decision to submit the article for publication. All authors have participated sufficiently in the work to take public responsibility for the content and agree to the submission of this manuscript.

Funder Information

Biodiversity Genomics Europe (**Grant no.101059492**) is funded by Horizon Europe under the Biodiversity, Circular Economy and Environment call (REA.B.3); co-funded by the Swiss State Secretariat for Education, Research and Innovation (SERI) under contract numbers 22.00173 and 24.00054; and by the UK Research and Innovation (UKRI) under the Department for Business, Energy and Industrial Strategy's Horizon Europe Guarantee Scheme. This research was also made possible by funding from grant PID2021-128901NB-100 MCIN/AEI/10.13039/501100011033 and by ERDF, a way of making Europe.

Author Contributions

SC and DFG collected and identified the species. SC, DFG, and LB sampled and preserved biological material and provided metadata. JPG and RF coordinated the sample collection. JPG, RF, RM, and AB provided support for sampling, metadata, and management. LA and MG extracted DNA, prepared libraries, and performed sequencing. FCF, FC, and JGG performed genome assembly and curation under the supervision of TA. LH, SS, and FM performed genome annotation. DDP conducted the analysis and generated the report. All authors contributed to writing, reviewing, and editing this genome note, and all have read and approved the final version.

Literature Cited

Aken, B. L., Ayling, S., Barrell, D., Clarke, L., Curwen, V., Fairley, S., Fernandez Banet, J., Billis, K., García Girón, C., Hourlier, T., Howe, K., Kähäri, A., Kokocinski, F., Martin, F. J., Murphy, D. N., Nag, R., Ruffier, M., Schuster, M., Tang, Y. A., … Searle, S. M. J. (2016). The Ensembl gene annotation system. *Database*, *2016*, baw093. https://doi.org/10.1093/database/baw093

Carranza, S., Arnold, E. N., & Pleguezuelos, J. M. (2006). Phylogeny, biogeography, and evolution of two

ARPHA Preprints *Author-formatted document posted on 08/04/2025.* DOI: https://doi.org/10.3897/arphapreprints.e155085

Mediterranean snakes, *Malpolon monspessulanus* and *Hemorrhois hippocrepis* (Squamata, Colubridae), using mtDNA sequences. *Molecular Phylogenetics and Evolution*, *40*(2), 532–546. https://doi.org/10.1016/j.ympev.2006.03.028

Challis, R., Kumar, S., Sotero-Caio, C., Brown, M., & Blaxter, M. (2023). Genomes on a Tree (GoaT): A versatile, scalable search engine for genomic and sequencing project metadata across the eukaryotic tree of life. *Wellcome Open Research*, *8*, 24. https://doi.org/10.12688/wellcomeopenres.18658.1

Corominas, M., Marquès-Bonet, T., Arnedo, M. A., Bayés, M., Belmonte, J., Escrivà, H., Fernández, R., Gabaldón, T., Garnatje, T., Germain, J., Niell, M., Palero, F., Pons, J., Puigdomènech, P., Initiative For The Earth BioGenome Project, T. C., Catalan initiative for the Earth BioGenome Project, Arroyo, V., Cuevas-Caballé, C., Obiol, J. F., … Guigó, R. (2024). The Catalan initiative for the Earth BioGenome Project: Contributing local data to global biodiversity genomics. *NAR Genomics and Bioinformatics*, *6*(3), lqae075. https://doi.org/10.1093/nargab/lqae075

De Panis, D. (2024a). *ERGA-BGE Genome Report ANNOT analyses*. WorkflowHub. https://doi.org/10.48546/WORKFLOWHUB.WORKFLOW.1096.1

De Panis, D. (2024b). *ERGA-BGE Genome Report ASM analyses (one-asm WGS Illumina PE + HiC)*. WorkflowHub. https://doi.org/10.48546/WORKFLOWHUB.WORKFLOW.1103.2

Gomez-Garrido, J. (2024). *CLAWS (CNAG's long-read assembly workflow in Snakemake)*. WorkflowHub. https://doi.org/10.48546/WORKFLOWHUB.WORKFLOW.567.2

Gruber, A. R., Lorenz, R., Bernhart, S. H., Neuböck, R., & Hofacker, I. L. (2008). The Vienna RNA Websuite. *Nucleic Acids Research*, *36*(suppl_2), W70–W74. https://doi.org/10.1093/nar/gkn188

Kalvari, I., Nawrocki, E. P., Argasinska, J., Quinones-Olvera, N., Finn, R. D., Bateman, A., & Petrov, A. I. (2018). Non-Coding RNA Analysis Using the Rfam Database. *Current Protocols in Bioinformatics*, *62*(1), e51. https://doi.org/10.1002/cpbi.51

Kozomara, A., Birgaoanu, M., & Griffiths-Jones, S. (2019). miRBase: From microRNA sequences to function. *Nucleic Acids Research*, *47*(D1), D155–D162. https://doi.org/10.1093/nar/gky1141

Martínez-Solano, I., Corti, C., Mellado, V. P., Sá-Sousa, P., Pleguezuelos, J. M., & Cheylan, M. (2008). IUCN Red List of Threatened Species: Malpolon monspessulanus. *IUCN Red List of Threatened Species*. https://www.iucnredlist.org/en

Mazzoni, C., Ciofi, C., & Waterhouse, R. (2023). Biodiversity: An atlas of European reference genomes. *Nature*, *619*, 252–252. https://doi.org/10.1038/d41586-023-02229-w

Nawrocki, E. P., & Eddy, S. R. (2013). Infernal 1.1: 100-fold faster RNA homology searches. *Bioinformatics*, *29*(22), 2933–2935. https://doi.org/10.1093/bioinformatics/btt509

Pleguezuelos, J. M., & Salvador Milla, A. (2021). *Culebra bastarda – Malpolon monspessulanus (Hermann, 1804)*. https://doi.org/10.20350/digitalCSIC/8807

Speybroeck, J., Beukema, W., Bok, B., & Van Der Voort, J. (2016). *Field Guide to the Amphibians and Reptiles of Britain and Europe*.

ARPHA Preprints *Author-formatted document posted on 08/04/2025.* DOI: https://doi.org/10.3897/arphapreprints.e155085

Srodawa, K., Cerda, P. A., Davis Rabosky, A. R., & Crowe-Riddell, J. M. (2023). Evolution of Three-Finger Toxin Genes in Neotropical Colubrine Snakes (Colubridae). *Toxins*, *15*(9), Article 9. https://doi.org/10.3390/toxins15090523

The UniProt Consortium. (2019). UniProt: A worldwide hub of protein knowledge. *Nucleic Acids Research*, *47*(D1), D506–D515. https://doi.org/10.1093/nar/gky1049