OXFORD

Resource Article: Genomes Explored

Chromosome-level genome assembly of Lilford's wall lizard, *Podarcis lilfordi* (Günther, 1874) from the Balearic Islands (Spain)

Jessica Gomez-Garrido[1], Fernando Cruz[1], Tyler S. Alioto[1], Nathalie Feiner[2], Tobias Uller[2], Marta Gut[1,3], Ignacio Sanchez Escudero[1], Giacomo Tavecchia[4,*], Andreu Rotger[4], Katherin Eliana Otalora Acevedo[5,6], and Laura Baldo[5,7,*]

[1]CNAG-CRG, Centre for Genomic Regulation (CRG), Barcelona Institute of Science and Technology (BIST), 08028 Barcelona, Spain
[2]Department of Biology, Lund University, Lund, Sweden
[3]Universitat Pompeu Fabra (UPF), Barcelona, Spain
[4]Animal Demography and Ecology Unit, IMEDEA, CSIC-UIB, Esporles, Spain
[5]Department of Evolutionary Biology, Ecology and Environmental Sciences, University of Barcelona, Barcelona, Spain
[6]Fundación Motiva Inteligencia Colectiva, Biodiversity Branch, Tunja, Boyacá, Colombia
[7]Institute for Research on Biodiversity (IRBio), University of Barcelona, Barcelona, Spain

*To whom correspondence should be addressed. Tel. +34 9340 21515. Email: baldo.laura@ub.edu; g.tavecchia@imedea.uib-csic.es

Abstract

The Mediterranean lizard *Podarcis lilfordi* is an emblematic species of the Balearic Islands. The extensive phenotypic diversity among extant isolated populations makes the species a great insular model system for eco-evolutionary studies, as well as a challenging target for conservation management plans. Here we report the first high-quality chromosome-level assembly and annotation of the *P. lilfordi* genome, along with its mitogenome, based on a mixed sequencing strategy (10X Genomics linked reads, Oxford Nanopore Technologies long reads and Hi-C scaffolding) coupled with extensive transcriptomic data (Illumina and PacBio). The genome assembly (1.5 Gb) is highly contiguous (N50 = 90 Mb) and complete, with 99% of the sequence assigned to candidate chromosomal sequences and >97% gene completeness. We annotated a total of 25,663 protein-coding genes translating into 38,615 proteins. Comparison to the genome of the related species *Podarcis muralis* revealed substantial similarity in genome size, annotation metrics, repeat content, and a strong collinearity, despite their evolutionary distance (~18–20 MYA). This genome expands the repertoire of available reptilian genomes and will facilitate the exploration of the molecular and evolutionary processes underlying the extraordinary phenotypic diversity of this insular species, while providing a critical resource for conservation genomics.

Key words: Hi-C, Oxford Nanopore Technologies (ONT), RNAseq, mitochondrial genome, reptile

1. Introduction

Podarcis lilfordi (Günther, 1874), also known as Lilford's wall lizard, is an endemic lizard of the Balearic Islands (Spain), currently confined to the Cabrera archipelago and several islets surrounding Menorca and Mallorca.[1] Due to a patchy distribution (~43 isolated populations) and threats from habitat loss, the species is currently listed as endangered by the IUCN (https://www.iucnredlist.org/species/17795/7481971). An extensive morphological diversity has resulted in the description of a large number (currently 23) of subspecies or morphotypes, but current understanding considers each population/islet as a distinct evolutionary unit.[1] The species exhibits several insular characteristics that likely reflect the lack of natural predators, including high population densities,[2] high survival and low fecundity[2–5] compared to their continental relatives. The insularity and large phenotypic diversity of *P. lilfordi* have made it a popular model for

studies of ecological adaptation and demographic dynamics of terrestrial vertebrates, as well as for testing predictions of how evolution proceeds on islands. During the past decades, substantial efforts have been directed toward understanding the species' morphological diversity,[1,3,6] trophic ecology,[7] life history traits and demographic resilience,[3,8] genetics[5,9] and associated gut microbes.[10–12] Despite this extensive knowledge and the endemic character of *P. lilfordi*, genomic information on the species has been limited.[9]

Here we present the complete genome sequence of *P. lilfordi* (both nuclear and mitochondrial) from a single female specimen (the heterogametic sex, ZW) collected from Aire Island (Menorca), where the species was first described (Gunther, 1874). We used a sequencing strategy combining long and short reads, coupled with RNAseq data from multiple tissues and specimens, to achieve a highly robust assembly and annotation. The *P. lilfordi* genome, along with

that of its continental relative *Podarcis muralis*,[13] will extend current genomic resources within this highly diverse lizard genus[14,15] and provide a genetic framework to begin understanding the processes behind the remarkable diversification of the Lilford's wall lizard. In addition, the genome will provide a critical tool for the development of a conservation plan based on population-level genomics (accounting for levels of genetic diversity, drift, and accumulation of deleterious mutations) for an accurate assessment of the species adaptive potential and resilience in face of current challenges by global change and increasing human pressure in the Balearic Islands.

2. Materials and methods

2.1. Sample collection and processing

Five adult specimens of *P. lilfordi (*subspecies *lilfordi)* were collected in April 2021 on Aire Island, to the southeast of Menorca Island (Spain). The island is one of the largest and most densely populated (surface area 34 ha, 4,098.60 lizards/ha).[4] Specimens were caught using pitfall traps containing fruit juices placed along paths and vegetation edges. All specimens were sexed according to visual examination of femoral pores and morphology,[16] weighed and body size measured as snout to vent length (see Supplementary Table S1 for metadata). Individuals were sacrificed by cervical dislocation, immediately stored on dry ice, and kept at –80°C until processing. Under a sterile hood, each frozen specimen was rapidly dissected to extract all major organs, including heart, liver, lungs, kidneys, brain, testicles/ovaries, intestine and muscle tissue from the tail. All tissues were immediately stored at –80°C.

A single female adult specimen (the heterogametic sex, ZW) was chosen for genome sequencing. Tissues were sent to the Centre Nacional d'Anàlisi Genòmica (CNAG-CRG) for DNA extraction and sequencing. Samples from the remaining specimens were shipped to Lund University (Sweden) for RNA extraction. RNA sequencing was performed at the SNP&SEQ Technology Platform (for short reads), and at the Uppsala Genome Center (for long reads) (SciLifeLabs, Sweden).

Specimen sampling was granted by the Species Protection Service (Department of Agriculture, Environment and Territory, Government of the Balearic Islands), under permit CAP03/2021 (to LB). The specimen remains were deposited at the Museum of Natural History of Barcelona (Spain), under voucher name MZB 2022-5701.

2.2. Genomic DNA extractions

High-molecular-weight (HMW) gDNA was extracted from frozen liver using the Nanobind tissue kit (Circulomics) following the manufacturer's protocol. Briefly, two cryopreserved liver aliquots of 38 mg and 24 mg were homogenized under cryogenic conditions on dry ice, using a mortar and pestle. The pulverized tissue was collected into 1.5 ml tubes with lysis buffer (Circulomics). Nanobind disk (Circulomics) was used on fresh supernatant for the gDNA binding. The HMW gDNA eluate was quantified by Qubit DNA BR Assay kit (Thermo Fisher Scientific) and the DNA purity was evaluated using Nanodrop 2000 (Thermo Fisher Scientific) UV/Vis measurements. The gDNA integrity was evaluated with pulsed-field gel electrophoresis SeaKem® GOLD Agarose 1% (Lonza), using the Pippin Pulse (Sage Science). The gDNA samples were stored at 4°C.

2.3. Genome sequencing

The complete genome sequence was achieved employing a mixed sequencing strategy, combining the use of 10X linked short reads (Illumina NovaSeq 6000, 2 × 150 bp) for base accuracy, long reads (Oxford Nanopore Technologies, ONT) for high contiguity and repeat resolution, and Hi-C (Illumina NovaSeq 6000, 2 × 150 bp) for chromosome-level scaffolding (Fig. 1).

The linked reads library was prepared using the Chromium Controller instrument (10X Genomics) and Genome Reagent Kits v2 (10X Genomics) following the manufacturer's protocol. Briefly, 10 ng of HMW gDNA was portioned in GEM reactions including a unique barcode (Gemcode) after loading onto a chromium controller chip. The droplets were then recovered, isothermally incubated, fractured and the intermediate DNA library was then purified and size-selected using Silane and Solid Phase reverse immobilisation (SPRI) beads. Illumina-compatible paired-end sequencing libraries were prepared following 10X Genomics recommendations and validated on an Agilent 2100 BioAnalyzer with the DNA 7500 assay (Agilent). The library was sequenced on NovaSeq 6000 (Illumina, 2 × 151 bp) following the manufacturer's protocol for dual indexing. Image analysis, base calling and quality scoring of the run were processed using the manufacturer's software Real Time Analysis (RTA v3.4.4) and followed by the generation of FASTQ files.

The ONT libraries were prepared using the 1D Sequencing kit SQK-LSK110. Briefly, 1.0 μg of the HMW gDNA was DNA-repaired and DNA-end-repaired using NEBNext FFPE DNA Repair Mix and the NEBNext UltraII End Repair/dA-Tailing Module (New England Biolabs, NEB), followed by ligation of sequencing adaptors. The library was purified with 0.4 X AMPure XP Beads and eluted in Elution Buffer. Four sequencing runs were performed on GridION Mk1 (ONT) using an R9.4.1 flow cell (FLO-MIN106D, ONT) on the MinKNOW platform version 4.2.5 for real-time monitoring. Data were collected for 110 h and base called with Guppy version 4.3.4 in high accuracy mode using the dna_r9.4.1_450bps_modbases_dam-dcm-cpg_hac.cfg model.

The Hi-C libraries were prepared using the Omni-C kit (Dovetail Genomics), following the manufacturer's protocol. Briefly, frozen muscle tissue (heart) was pulverized using a mortar and pestle immersed in a liquid nitrogen bath. Chromatin was fixed in place with formaldehyde (Sigma Aldrich), digested with DNase I and DNA extracted. DNA ends were repaired, and a biotinylated bridge adapter was ligated followed by proximity ligation of adapter-containing ends. After reverse crosslinking, the DNA was purified and followed by the preparation of Illumina-compatible paired-end sequencing libraries (omitting the fragmentation step). Biotinylated chimeric molecules were isolated using streptavidin beads before PCR enrichment of the library. The library was sequenced on NovaSeq 6000 (Illumina, 2 × 151 bp) following the manufacturer's protocol for dual indexing.

2.4. Transcriptome sequencing

RNA was extracted from five different tissues (heart, kidney, liver, lungs and tail) of five specimens using the RNeasy Mini Kits (Qiagen) and including an on-column DNA digestion step. RNA concentration was measured using the Qubit RNA High Sensitivity assay and RNA quality was assessed using the Eukaryote Total RNA Nano assay on the Agilent
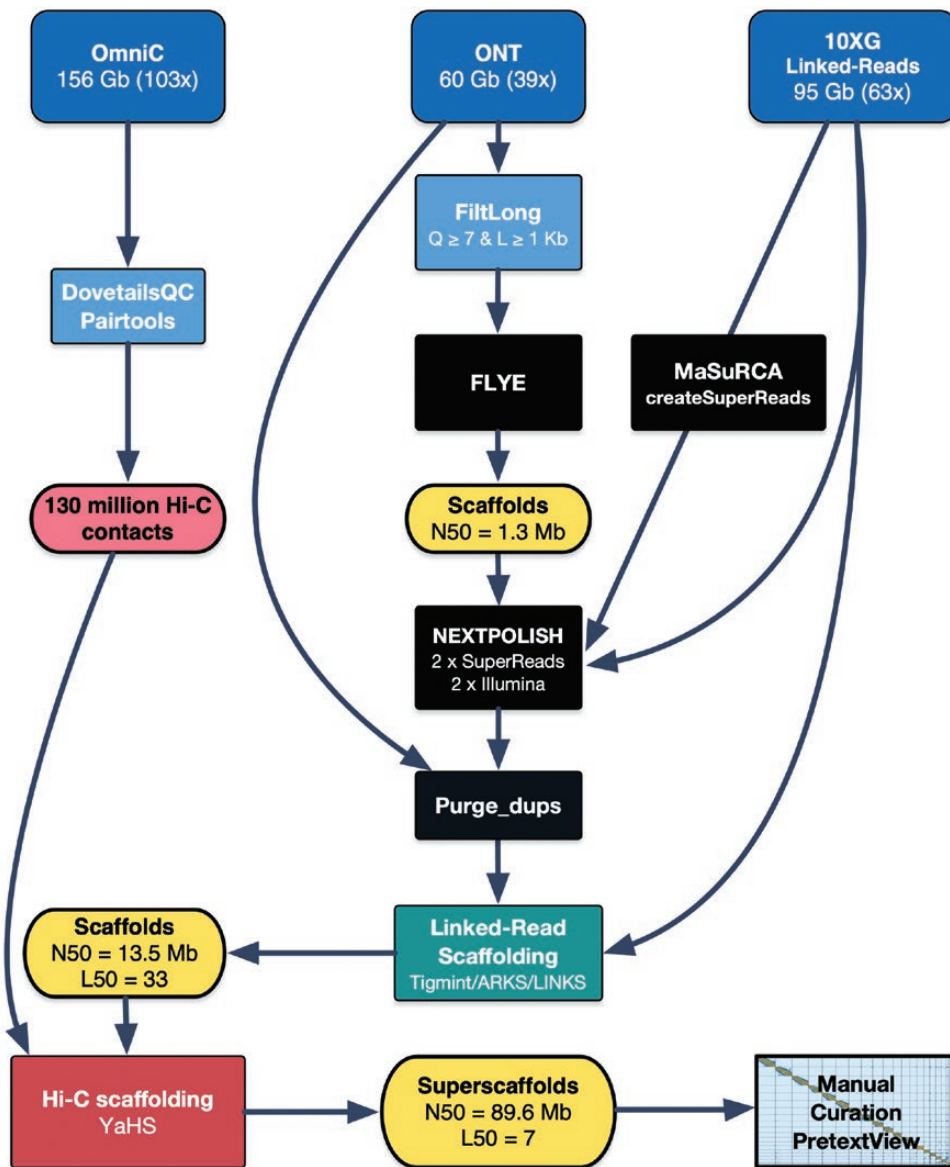
Figure 1. Assembly workflow. Summary of the main steps followed to obtain the curated chromosome-level assembly. See the Snakemake pipeline for detailed information (https://github.com/cnag-aat/assembly_pipeline).

2100 Bioanalyzer system. For short-read RNA sequencing, we pooled samples of two to four individuals per tissue in equimolar amounts (Supplementary Table S2). These five tissue-specific samples were subjected to Illumina's TruSeq Stranded mRNA library preparation kit and sequenced on NovaSeq 6000 (Illumina, 2 × 150 bp). For long-read RNA sequencing, we pooled five individual tissue samples in equimolar amounts (Supplementary Table S2) and subjected this single, mixed sample to PacBio Iso-Seq sequencing.

2.5. Nuclear genome assembly

The nuclear genome was assembled following the workflow summarized in Fig. 1. To increase reproducibility, the main steps have been included into a snakemake[17] pipeline (https://github.com/cnag-aat/assembly_pipeline). A few modifications were added for this project which are detailed below, along with a summary of major steps. All intermediate and final assemblies produced during the process were evaluated as

indicated in the pipeline description, using BUSCO[18,19] v 4.0.6 with vertebrata_odb10, fasta-stats.py (https://github.com/galaxyproject/tools-iuc), Nseries.pl and Merqury[20] v1.1. Note that our final assembly statistics was further evaluated with updated versions of Merqury v1.3 and BUSCO version 5.0.4 with vertebrata_odb10.

Before assembly, the 10X Illumina reads were preprocessed using LongRanger Basic v2.2.2 and the nanopore reads were filtered using FiltLong v0.2.0 with options '--min_length 1000 --min_mean_q 80'. The resulting set of filtered nanopore reads accounted for 50.8 Gb (about 34x coverage) with a mean Phred-scale read quality of 9.8 and a read length N50 of 31.4 Kb (Supplementary Table S3).

First, we assembled the filtered nanopore reads using Flye v2.8.3 with options '--nano-raw -i 2' (the latest performs two final rounds of polishing in the output assembly) obtaining a base assembly that comprised a total of 1.54 Gb, with an N50 of 1.36 Mb and a consensus quality (QV) of 27.19 (Fig. 1 and Supplementary Table S4).

Second, to maximize the QV of the Flye assembly, we tested several polishing tools and strategies (Supplementary Table S4). The best results were obtained using a combined strategy that exclusively used 10X Illumina reads to improve the QV. We used MaSuRCA[21,22] version 4.0.4 using *createSuperReadsForDirectory.perl* to obtain high-quality SuperReads with a read N50 of 539 bp. The SuperReads were then aligned to the genome with Minimap2[23] v2.17 and the Illumina reads were aligned with BWA-MEM[24] v0.7.15. Finally, we ran NextPolish[25] v1.3.1 with these SuperReads, treating them as long reads, and the debarcoded 10X linked reads as paired-end Illumina for polishing the Flye assembly. In total, four polishing iterations were performed with NextPolish: two with the SuperReads and another two with the Illumina reads.

Third, to obtain a haploid reference (removing alternate haplotigs and other artificial duplications), we purged the assembly using purge_dups[26] v1.2.5 with cutoffs -l 5 -m 24 -u 78. This step removed 4,289 scaffolds accounting for 94,919,135 bp.

Fourth, the purged assembly was corrected and scaffolded with 10X Linked-Reads using the Faircloth's Lab pipeline (http://protocols.faircloth-lab.org/en/latest/protocols-computer/assembly/assembly-scaffolding-with-arks-and-links.html#). This pipeline includes Tigmint[27] v1.1.2, ARKS[28] v1.0.3 and LINKS[29] v1.8.5. Tigmint was used to identify and correct mis-assemblies using the linked reads in the purged assembly. The corrected assembly was then scaffolded using ARKS and LINKS, resulting in 2,783 scaffolds accounting for 1,460,390,273 bp with an N50 = 13.51 Mb (Supplementary Table S4).

Finally, the Omni-C reads were then mapped to the assembly using BWA-MEM and pre-processed using the Dovetail pipeline (https://omni-c.readthedocs.io/en/latest/fastq_to_bam.html). The filtering of the alignments was done with the default minimum mapping quality of 40. After the removal of PCR duplicates (46.34%, see Supplementary Table S5), 130,208,672 read pairs remained and were used as input to YaHS[30] v1.1 scaffolder with default parameters. We performed two rounds of assembly error correction and made 15 breaks, followed by ten rounds of scaffolding from higher to lower resolution (10 Mb down to 10 Kb), which produced an assembly with a span of 1,460,440,873 bp with an N50 of 89.54 Mb. The high contiguity achieved is also reflected in the scaffold L90 of 17, a value close to the number of chromosomes (n = 19) (Fig. 1 and Supplementary Table S4).

2.6. Manual curation

To guide manual curation of the assembly, we computed the mean Illumina coverage for all scaffolds (using BWA-MEM v0.7.15, SAMtools[31] v1.9 and BlobTools[32] v1.1) and the whole-genome alignments (WGAs) against the genome assemblies of two species belonging to family *Lacertidae*: a ZZ male of *P. muralis* (PodMur1.0) and a ZW female of *Lacerta agilis* (rLacAgi1.pri). The WGAs were produced with nucmer4[33] and visualized with Dot (https://github.com/MariaNattestad/dot). This approach allowed us to 'scaffotype' (i.e. assign to a chromosome) the largest 20 superscaffolds (Supplementary Tables S6 and S7). Finally, the location of gaps (fasta-stats.py) and telomeres (https://github.com/tolkit/telomeric-identifier), together with the Illumina coverage, were added to the contact map using PretextGraph (https://github.com/wtsi-hpag/PretextGraph). Manual curation was then performed using

PretextView (https://github.com/wtsi-hpag/PretextView). Given the high quality and contiguity of the YaHS assembly, the curation involved only 17 edits.

The Blobtoolkit[34] pipeline was then run on the curated assembly (Supplementary Fig. S1), using the NCBI nt database (updated on September 2022) and several BUSCO odb10 databases (sauropsida, vertebrata, metazoa, eukaryota, chlorophyta, fungi and bacteria). Removal of six contaminated scaffolds (based on GC cutoff of 0.3–0.65) (see Supplementary Table S8) resulted in the final assembly rPodLil1.2.

Finally, by comparing Illumina and ONT coverage estimates, we calculated the ratio of coverage for each sex chromosome with respect to the autosomal mean coverage (Supplementary Table S9). Whole genome alignments of the *P. lilfordi* (rPodLil1.2) against *P. muralis* (PodMur1.0) were performed with Minimap2 using the '-x asm5' option and visualized with the *pafr* R package.

2.7. Nuclear genome annotation

Gene annotation is very sensitive to both the tools and evidence-based data used. To produce an accurate gene annotation here we (i) sequenced and aligned transcriptome data to the *P. lilfordi* assembly, (ii) inferred protein content from comparative analyses against closely related annotated genomes and (iii) performed *ab initio* gene predictions (see flowchart in Supplementary Fig. S2).

As a first step, we masked all the repeat regions as they can include Open-Reading Frames (ORFs) and share certain domain homology with coding genes. For this purpose, we searched for repeats with RepeatMasker v4-1-2 (http://www.repeatmasker.org) using the custom RepBase repeat library available for *Podarcis,* along with a specific repeat library generated with RepeatModeler[35] v1.0.11 for our assembly. After excluding those repeats that were part of repetitive protein families (performing a BLAST search against Uniprot; last accessed March 2022), a final repeat annotation was produced using RepeatMasker. As this repeat annotation was produced mainly to aid in the genome annotation, low-complexity repeats were not annotated. The *P. muralis* genome was similarly processed for repeat annotation and comparative analysis.

Second, to obtain a comprehensive catalogue of the gene isoforms, we generated transcriptomic data from several tissues and multiple individuals using both regular Illumina RNA-seq and long-read PacBio Isoseq technologies (Supplementary Table S2). While regular Illumina provides an ample coverage of the full transcriptome potential, long-reads can cover an entire isoform with a single read, limiting the production of chimeric variants during transcript assembling. Short and long RNA reads were mapped to the genome assembly using STAR[36] v-2.7.2a and Minimap2 v2.14 (with '-x splice:hq -uf' options), respectively. To infer gene models from the mappings, StringTie[37] v2.1.4 was run on each BAM file, setting the option '--conservative' only for the long-read libraries in order to disable the process of combining reads for transcript assembling, therefore retaining the full-length transcript sequences. Finally, all the models produced by Stringtie were combined using TACO[38] v0.6.3 and the resulting 'gtf' file was input into the Program to Assemble Spliced Alignments (PASA[39] v2.4.1) to produce PASA assemblies for annotation. High-quality junctions used during the annotation process were obtained with Portcullis[40]

v1.2.0 after STAR and Minimap2 mapping. Additionally, the *TransDecoder* program (embedded in the PASA package) was run on the PASA assemblies to detect coding regions in the transcripts.

Third, we used proteomes from closely related species as a guide for predicting gene content. For this purpose, we downloaded the complete proteomes of *P. muralis, Pogona vitticeps* and *Pantherophis guttatus* from Uniprot in April 2022 and aligned them to the *P. lilfordi* repeat-masked assembly using Spaln[41] v2.4.03.

Fourth, we performed *ab initio* gene predictions on the final assembly using three different programs: GeneID[42] v1.4, Augustus[43] v3.3.4 and Genemark-ES[44] v2.3e, with and without incorporating junction evidence from the RNAseq data. The gene predictors were run with trained parameters on humans, except for Genemark, which runs in a self-trained mode.

Final genome annotation was achieved by combining all the data into consensus CDS models using EvidenceModeler-1.1.1[39] (EVM). Additionally, untranslated regions (UTRs) and alternative splicing forms were annotated via two rounds of PASA annotation updates (Supplementary Fig. S2).

Functional annotation was performed on the annotated proteins with Blast2GO.[45] First, a DIAMOND Blastp[46] search was made against the NCBI nr database (last accessed May 2022). Furthermore, InterProScan[47] was run to detect protein domains on the annotated proteins. All these data were combined by Blast2GO, which produced the final functional annotation.

General statistics on the genome and individual chromosomes were computed with in-house PERL scripts. The sex chromosome W presented 14 transposon-derived, short, only *ab-initio* single-copy genes. Due to the high repetitiveness of the chromosome, these genes were considered as artefacts and removed from the final annotation.

The annotation of non-coding RNAs (ncRNAs) was obtained as follows. First, the program CMsearch[48] v1.1 from the Infernal[49] package was run against the RFAM database of RNA families v12.0. Additionally, transfer RNA genes were identified by tRNAscan-SE[50] v2.08. Long non-coding RNAs (lncRNAs) were identified as those expressed transcripts (assembled by PASA) longer than 200 bp that were not included in the protein-coding annotation and not covered by a small ncRNA in more than 80% of their length. The resulting transcripts were clustered into genes (i.e. same gene assignment) using shared splice sites or significant sequence overlap.

2.8. Mitogenome assembly and annotation

To obtain the mitochondrial sequences, all ONT reads, previously filtered for whole-genome assembly with FiltLong v.0.2.0 to be at least 1 Kb long and have a mean quality of 7, were mapped with Minimap2 against the *P. muralis* complete mitochondrial genome (NC_011607.1; 17,311 bp) with options: '-t $THREADS -ax map-ont $DATABASE $READS'. We retained all reads with mapping quality = 12 (relatively unique) and at least 800 exact matches to the mitochondrial genome reference; these included 8,093 reads and a total of 51,692,448 bp (estimated mitochondrial coverage 2,986x).

All filtered ONT reads were assembled with Flye[51] v2.9 using the options: 'flye --meta --scaffold -t 12 -i 2 -g 25k --nano-raw'. The '--meta' option is the most appropriate for uneven coverage samples and two polishing iterations were

run with the ONT reads on the final assembly with '-i 2'. Finally, the output assembly was screened for circular contigs, resulting in eight linear contigs and a single circular contig 17,112 bp long (contig_1).

As the reference mitogenome (*P. muralis*) is relatively distant (18–20 MYA), the identification of Illumina reads mapping outside the most conserved regions of the organelle is not straightforward. To overcome this issue, we mapped all the Illumina data (previously de-barcoded PE 2 × 150 bp reads 10x linked-reads) to our complete Flye long-read assembly with gem-mapper, with ≤2% mismatches. Finally, a total of 326,105 read pairs were collected for further polishing (estimated coverage of the mitochondrial genome is 5,255.81x).

To further improve the sequence accuracy of the assembled mitochondrial genome, we performed two additional rounds of polishing on contig_1 with the selected Illumina reads using NextPolish v1.1.0 with Illumina PE 2 × 150 bp with options: '-paired -max_depth 5000'. The polished assembly was evaluated with Merqury v1.1 using '$k = 21$' on the mitochondrial Illumina reads, *dnadiff* from MUMmer[33] package 4.0.0beta2 and fasta-stats.py. The resulting circular chromosome was rotated and oriented according to the *P. muralis* reference, after detecting the appropriate Origin coordinates with the *dnadiff*.

The annotation of the mitogenome was performed using the MITOS[52] Web Server (http://mitos.bioinf.uni-leipzig.de/). Manual curation was performed by checking the predicted sequences and comparing the annotation to that of the *P. muralis* mitogenome. As a result, one partial tRNA-Asp was removed from the annotation due to its absence in the *P. murallis* mitogenome annotation and to the high e-value reported by MITOS (e-value = 0.04071).

3. Results and discussion

3.1. Genome assembly

Genome sequencing yielded a total of 95 Gb of Illumina data (2 × 150 bp), 156.2 Gb of Omni-C data (2 × 150 bp) and 60 Gb of ONT data (Fig. 1). Assembly of the ONT data, followed by polishing, scaffolding, and manual curation resulted in a highly contiguous and complete assembly (rPodLil1.2) of 2,148 scaffolds accounting for 1.46 Gb, in line with the *C-value* and assembly span of *P. muralis* (1.51 Gb for PodMur1.0)[13] (Fig. 2). It has a contig N50 of 1.48 Mb, scaffold N50 of 89.64 Mb (≥10 Mb) and QV of 40 (Supplementary Table S4), meeting the minimum quality requirement of 6.C.Q40 (megabase contig N50 and chromosomal-scale scaffold N50, with less than 1/10,000 error rate) established by the Earth Biogenome Project (EBP) for eukaryotic species with sufficient DNA and tissue.[53] Although the obtained contig N50 is shorter than most values found for other published chromosome-level *Lacertidae* assemblies (Supplementary Table S10), the high contig quality obtained and the use of Hi-C data allowed us to confidently assign most of the sequences (98.70%) to candidate chromosomes. In fact, the resulting assembly is consistent with the karyotype (2*n* = 38),[54] including 18 autosomes and two sex chromosomes (Fig. 2). Moreover, the assembly is highly complete, with 96.7% single copy complete genes, 87% k-mer completeness, and a low false duplication rate of 0.68% (Fig. 2).

Assignment to chromosomes was done via whole-genome alignments to the available genome assemblies of *P. muralis* (a
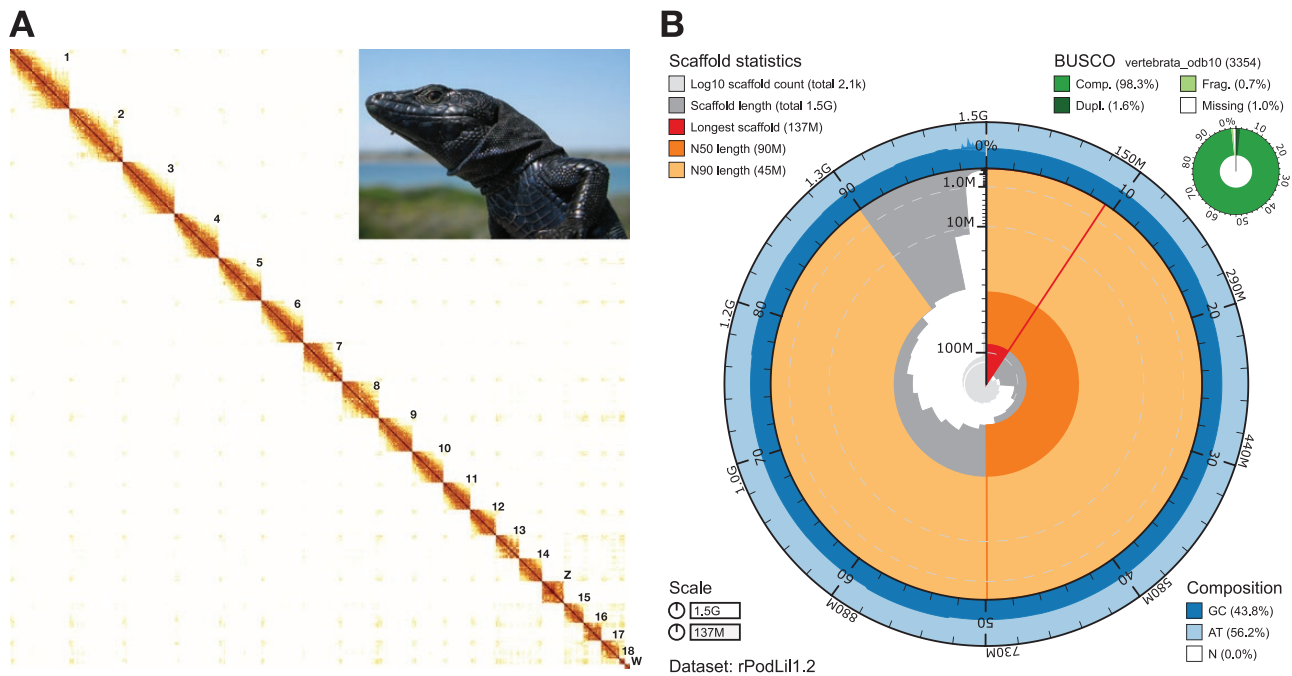
Figure 2. Visual summary of the rPodLil1.2 assembly. (A) Hi-C contact map of the genome assembly visualized in PretextView. The map shows 20 superscaffolds, ordered from longest to shortest, corresponding to the 18 autosomes and the two sex chromosomes (Z and W); the latter can be better visualized in Supplementary Fig. S5. Additionally, there are a total of 44 short unlocalized scaffolds within the main chromosomes, and 2,084 short unplaced scaffolds (not visible at the resolution shown). (B) Snailplot summarising assembly metrics, including scaffold statistics, BUSCO completeness, total size and base composition. The main plot is divided into 1,000 size-ordered bins around the circumference with each bin representing 0.1% of the 1.460.085.851 bp assembly. The distribution of record lengths is shown in dark grey with the plot radius scaled to the longest record present in the assembly (137 Mb, shown in red). Orange and pale-orange arcs show the N50 and N90 record lengths (90 Mb and 45 Mb), respectively. The pale grey spiral shows the cumulative record 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 vertebrata_odb10 set is shown in the top right.

male) and *L. agilis* (rLacAgi1.pri, a female). The alignments revealed a high level of collinearity to both assemblies for the autosomes as well as the Z chromosome (shown for *P. muralis* in Supplementary Fig. S4). A total of 21 inversions greater than 1 Mbp in length were observed with respect to the *P. muralis* assembly, only two of which were larger than 5 Mbp. There was no evidence of translocations or large deletions. Sex chromosome assignment was additionally supported by manual curation, showing the expected drop in sequencing coverage (Supplementary Table S9) and the typical pattern of Hi-C contacts (Supplementary Fig. S5A). Sex chromosomes in lacertids (ZZ and ZW) are known to differ in gene copy numbers, with males (ZZ) showing twice as many genes as females (ZW) due to W degeneration.[53] Moreover, the W is mostly heterochromatic and shows a typical high content in repetitive sequences,[55,56] reducing the number of informative Hi-C read pairs for scaffolding in this chromosome. As expected, the Z chromosome of *P. lilfordi* was assembled into one superscaffold of 50.7 Mb in length that aligns well to both the Z chromosome of *P. muralis* and to that of *L. agilis* (Supplementary Fig. S5B), while the W chromosome was partially assembled into a 12.3 Mb superscaffold that aligns best to the superscaffold W of *L. agilis*, albeit with low identity (Supplementary Fig. S5C).

3.2. Genome annotation

While many genome assemblies have been recently released, their capability to answer certain biologically relevant questions is frequently limited by the lack of genome annotation.

Here, we provide a robust annotation for *P. lilfordi* by combining different approaches including newly generated transcriptome data, comparative proteomics, and *ab initio* gene prediction.

Overall, we annotated a total of 25,663 protein-coding genes that produce 43,578 transcripts (1.7 transcripts per gene) encoding 38,615 unique protein products (Table 1). We were able to assign functional labels to 72% (29,273) of the annotated proteins. The annotated transcripts contain 11 exons on average, with 91% of them being multi-exonic. In addition, we annotated 47,052 non-coding transcripts, including 12,785 lncRNAs and 34,267 sncRNAs (Table 1).

Running BUSCO on both the annotated protein-coding transcripts and protein sets showed a gene completeness of 97.9% and 97.1%, respectively, using the vertebrata_odb10 database. These results are in line with the genome BUSCO completeness of 98.3% and demonstrate the high accuracy of the genome annotation pipeline. Minor differences in the results obtained using the three sources of evidence are likely due to the algorithm that BUSCO uses, and the threshold established to consider a gene absent, fragmented or complete.

Comparison of *P. lilfordi* and *P. muralis* genome annotation statistics reveals only few differences (Table 1). Gene content is comparable, although we annotated ~3,000 extra protein-coding genes for *P. lilfordi*. Aligning these genes back to the PodMur1.0 genome assembly indicates that only 600 are unique to *P. lilfordi* (i.e. do not align to the *P. muralis* assembly). Most of the *P. lilfordi* proteins (96.5%) have a complete ORF (versus 83.7% in *P. muralis*) supporting our

high-quality assembly and annotation (Table 1). Moreover, the median UTR length is four times longer in *P. lilfordi* than in *P. muralis*, which is likely due to the usage of long-read transcript sequencing technologies (PacBio) for our annotation. Although most of the observed differences can be ascribed to the use of different evidence sources and annotation pipelines, additional genome resequencing data will be critical for validation.

3.3. Repetitive elements

Around 39% of the assembled genome was repetitive: 6% was annotated as repeats using the 'podarcis' Repbase library and an additional 33% was classified as repeats using the repeat library obtained after running Repeat Modeler (https://www.repeatmasker.org/RepeatModeler/) against the assembly (Table 2). These repeats were represented by transposable elements, including short, interspersed elements (SINEs), long interspersed nuclear elements (LINEs), long terminal repeat retrotransposons (LTRs) and DNA transposons. The repeat content landscape of *P. lilfordi* was highly comparable to that of *P. muralis* (38–39%), in line with their similar genome size[57]: most of the repeats were classified as LINEs (~12%), followed by DNA transposons (~7%) and SINEs (4–5%) (Table 2). Percentage of LTR was slightly higher in *P. lilfordi* than *P. muralis* (2.37% against 1.36%). At present, we cannot confidently assess whether these minor differences between species are real or a methodological bias (methods can particularly affect repeat-resolution).

3.4. Mitogenome

The resulting mitogenome assembly is a single circular contig 17,251 bp long with a QV of 44.12. Its GC content is 38.73%, very similar to the 38.55% observed in the *P. muralis* mitogenome. As expected, the average identity between both mitogenomes is 88.74%. The origin of replication was identified and used to linearize the circular chromosome by comparison to the *P. muralis* mitogenome. After annotation, it was confirmed that this was the origin of the tRNA-Phe (MT-TF) that is used as a standard origin in vertebrate mitogenomes.[58] The final annotation of the mitochondrial chromosome contained 13 protein-coding genes, 2 rRNAs and 22 tRNAs, in line with expectations for vertebrates.

4. Conclusions

Here we announce and publicly release the *Podarcis lilfordi* reference nuclear and mitochondrial genome assembly and annotation. This is the first chromosome-level reference for an endemic reptile species released within the framework of the Catalan Initiative for the Earth Biogenome Project (CBP). The *P. lilfordi* genome is the second complete genome within the

Table 1. Final genome annotation statistics for the *P. lilfordi* assembly (rPodLil1.2) and comparison to *P. muralis* (PodMur1.0)

	rPodLil1.2	PodMur1.0
Genome size (bp)	1,460,085,851	1,511,002,858
Number of protein-coding genes	25,663	22,062
Median gene length (bp)	13,456	11,808
Number of transcripts	43,578	37,240
Number of proteins	38,615	36,445
Proteins with complete ORFs	37,249	30,503
Functionally annotated proteins	29,273	25,768
Number of coding exons	232,494	219,498
Median UTR length (bp)	2,138	505
Median intron length (bp)	1,297	1,245
Exons/transcript	10.90	12.12
Transcripts/gene	1.70	1.69
Multi-exonic transcripts (proportion)	0.91	0.92
Gene density (gene/Mb)	17.58	14.60

Table 2. Repeat content of the *P. lilfordi* assembly and comparison with *P. muralis*

	rPodLil1.2 (bp)	PodMur1.0 (bp)
SINEs	58,508,447 (4.0%)	71,879,703 (4.7%)
LINEs	186,782,953 (12.8%)	182,775,149 (12.1%)
LTR elements	34,584,027 (2.4%)	20,566,821 (1.4%)
DNA transposons	105,432,906 (7.2%)	110,903,635 (7.3%)
Unclassified	180,164,025 (12.3%)	181,606,330 (12.0%)
Total interspersed repeats	565,472,358 (38.7%)	567,731,638 (37.6%)

Percentages are estimated with respect to the total genome size

highly diverse genus *Podarcis*, along with that of *P. muralis*. Despite their evolutionary divergence (18–20 MYA[14]), the two species exhibit substantial conservation in genome organisation and overall annotation. Given its high quality, contiguity and annotation, the *P. lilfordi* genome sequence will represent a valuable resource for evolutionary and conservation genomics studies. The resource will facilitate comparative genomics of Lacertidae and reptiles in general, and aid in the understanding of the genetic bases of vertebrate insular adaptation (i.e. the island syndrome) and demographic resilience. Additionally, the genome will represent a critical reference to explore the genetic diversity of this endemic species, its adaptive potential and evidence for local adaptation, along with its ability to respond to current threats by human pressure in the Balearic Islands. Finally, we expect that future genome analyses will have a critical impact on conservation management and policy decisions on this endangered species.

Author contributions

L.B., G.T., N.F. and T.U. conceived and advised the project. L.B. and K.E.O.A. performed sample collection and preparation. M.G., I.S.E. and N.F. performed data production. J.G.G., F.C. and T.S.A. performed data processing and genome assembly. J.G.G. and T.S.A. performed data processing and genome annotation. All authors wrote and approved the final manuscript.

Data availability

All the sequencing data generated in this study are available at the European Nucleotide Archive (ENA) (https://www.ebi.ac.uk/ena/browser/home) under the accession number PRJEB50294. The assembly and annotation can also be found at ENA under the accession number GCA_947686815.1.

In addition, a genome browser, a BLAST Sequence Server and annotation data can be found at https://denovo.cnag.cat/podarcis. A Snakemake pipeline for genome assembly is available at https://github.com/cnag-aat/assembly_pipeline.

Supplementary Data

Supplementary data are available at *DNARES* online.

Table S1. Sampled specimen information
Table S2. RNA samples and quality
Table S3. ONT read statistics
Table S4. Assembly statistics, including intermediate steps
Table S5. Omni-C mapping statistics
Table S6. Scaffotyping based on Whole Genome Alignments (Yash vs PodMur1.0)
Table S7. Scaffotyping based on Whole Genome Alignments (Yash vs rLacAgi1)
Table S8. Contaminated scaffolds
Table S9. Ratio of sex chromosome coverage with respect to the autosomes mean coverage (based on ONT and Illumina reads)
Table S10. Major assembly statistics of publicly available Lacertidae genomes
Figure S1. Hexagon-binned blob plot of base coverage of filtered ONT reads against GC proportion for scaffolds in assembly rPodLil1.1. Scaffolds are coloured by phylum (according to Blast hits against the nt database; last accessed September 2022) and binned at a resolution of 30 divisions on each axis. Coloured hexagons within each bin are sized in proportion to the sum of individual scaffold lengths on a square-root scale, ranging from 1,018 to 693,200,870. Histograms show the distribution of scaffold length sum along each axis.
Figure S2. Annotation flowchart based on the non-decontaminated assembly (rPodLil1.1).
Figure S3. K-mer comparison between the Illumina reads and the rPodLil1.2 assembly. Stacked histogram of k-mer distributions obtained by comparing the assembly with Merqury v1.1 using $k = 21$ on the 10X Illumina reads. Artificial duplications corresponding to duplicate k-mers are shown in blue above the main peak (~40x). They only account for 0.68% of the k-mers.
Figure S4. Whole genome alignment of *P. lilfordi* to *P. muralis*. Chromosomal sequences, named according to corresponding chromosomal sequences in *P. muralis*, are ordered from largest to smallest in *P. lilfordi* and oriented with respect to *P. muralis*, which inverts the order of chromosomes 16 and 17. Alignments longer than 100 kb were selected and visualized with the *pafr* R library.

Figure S5. Sex chromosome assignment. (A) Hi-C contact map showing scaffolds corresponding to the sexual chromosomes. Illumina coverage is plotted in pink. (B) Alignment of the scaffold assigned to the Z chromosome in *L. agilis* against the corresponding scaffold in *P. lilfordi*. (C) Alignment of the scaffold assigned to the W chromosome in *L. agilis* against the corresponding scaffold in *P. lilfordi*.

References

1. Pérez-Cembranos, A., Pérez-Mellado, V., Alemany, I., et al. 2020, Morphological and genetic diversity of the Balearic lizard, *Podarcis lilfordi* (Günther, 1874): is it relevant to its conservation?, *Divers. Distrib.*, **26**, 1122–41.

2. Castilla, A.M. and Bauwens, D. 2000, Reproductive characteristics of the Island lacertid lizard *Podarcis lilfordi, J. Herpetol.*, **34**, 390–6.

3. Rotger, A., Igual, J. M., and Tavecchia, G. 2020, Contrasting size-dependent life history strategies of an insular lizard. *Curr. Zool.*, **66**, 625–33.

4. Pérez-Mellado, V., Hernández-Estévez, J.A., García-Díez, T., et al. 2008, Population density in *Podarcis lilfordi* (Squamata, Lacertidae), a lizard species endemic to small islets in the Balearic Islands (Spain), *Amphib-Reptilia*, **29**, 49–60.

5. Terrasa, B., Pérez-Mellado, V., Brown, R.P., Picornell, A., Castro, J.A., and Ramon, M.M. 2009, Foundations for conservation of intraspecific genetic diversity revealed by analysis of phylogeographical structure in the endangered endemic lizard *Podarcis lilfordi*, *Divers. Distrib.*, **15**, 207–21.

6. Rotger, A., Tenan, S., Igual, J.-M., Bonner, S., and Tavecchia, G. 2023, Life span, growth, senescence and island syndrome: accounting for imperfect detection and continuous growth, *J. Anim. Ecol.*, **92**, 183–94.

7. Alemany, I., Pérez-Cembranos, A., Pérez-Mellado, V., et al. 2022, DNA metabarcoding the diet of *Podarcis* lizards endemic to the Balearic Islands Fuller, R., (ed.), *Curr. Zool.*, zoac073. doi:10.1093/cz/zoac073

8. Rotger, A., Igual, J.M., Genovart, M., et al. 2021, Contrasting adult body-size in sister populations of the balearic lizard, *Podarcis lilfordi* (Günther 1874) suggests anthropogenic selective pressures, *Herpetol. Monogr.*, **35**, 53–64.

9. Bassitta, M., Brown, R.P., Pérez-Cembranos, A., et al. 2021, Genomic signatures of drift and selection driven by predation and human pressure in an insular lizard, *Sci. Rep.*, **11**.

10. Alemany, I., Pérez-Cembranos, A., Pérez-Mellado, V., et al. 2022, Faecal microbiota divergence in allopatric populations of *Podarcis lilfordi* and *P. pityusensis*, two lizard species endemic to the Balearic Islands, *Microb. Ecol.* doi:10.1007/s00248-022-02019-3

11. Baldo, L., Riera, J.L., Mitsi, K., and Pretus, J.L. 2018, Processes shaping gut microbiota diversity in allopatric populations of the endemic lizard *Podarcis lilfordi* from Menorcan islets (Balearic Islands), *FEMS Microbiol. Ecol.*, **94**, 1–14.

12. Baldo, L., Tavecchia, G., Rotger Vallespir, A., Igual, J., and Riera, J. 2022, Insular holobionts: the role of phylogeography and seasonal fluctuations in shaping the gut microbiotas of the Balearic Wall lizard *Podarcis lilfordi, Anim. Ecol.*

13. Andrade, P., Pinho, C., Pérez i de Lanuza, G., et al. 2019, Regulatory changes in pterin and carotenoid genes underlie balanced color polymorphisms in the wall lizard, *Proc. Natl. Acad. Sci. U.S.A.*, **116**, 5633–42.

14. Yang, W., Feiner, N., Pinho, C., et al. 2021, Extensive introgression and mosaic genomes of Mediterranean endemic lizards, *Nat. Commun.*, **12**, 2762.

15. Yang, W., Feiner, N., Salvi, D., et al. 2022, Population genomics of wall lizards reflects the dynamic history of the Mediterranean basin, *Mol. Biol. Evol.*, **39**, msab311.

16. Salvador, A. 2009, Lagartija balear – Podarcis lilfordi. En: Enciclopedia Virtual de los Vertebrados Españoles. Salvador, A., Marco, A. (Eds.). Museo Nacional de Ciencias Naturales, Madrid.

17. Mölder, F., Jablonski, K.P., Letcher, B., et al. 2021, Sustainable data analysis with Snakemake, *F1000Res*, **10**, 33.

18. Manni, M., Berkeley, M.R., Seppey, M., Simão, F.A., and Zdobnov, E.M. 2021, BUSCO update: novel and streamlined workflows along with broader and deeper phylogenetic coverage for scoring of eukaryotic, prokaryotic, and viral genomes, *Mol. Biol. Evol.*, **38**, 4647–54.

19. Simão, F.A., Waterhouse, R.M., Ioannidis, P., Kriventseva, E.V., and Zdobnov, E.M. 2015, BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs, *Bioinformatics*, **31**, 3210–2.

20. Rhie, A., Walenz, B.P., Koren, S., and Phillippy, A.M. 2020, Merqury: reference-free quality, completeness, and phasing assessment for genome assemblies, *Genome Biol.*, **21**, 245.

21. Zimin, A.V., Puiu, D., Luo, M.-C., et al. 2017, Hybrid assembly of the large and highly repetitive genome of *Aegilops tauschii*, a progenitor of bread wheat, with the MaSuRCA mega-reads algorithm, *Genome Res.*, **27**, 787–92.

22. Zimin, A.V., Marçais, G., Puiu, D., Roberts, M., Salzberg, S.L., and Yorke, J.A. 2013, The MaSuRCA genome assembler, *Bioinformatics*, **29**, 2669–77.

23. Li, H. 2018, Minimap2: pairwise alignment for nucleotide sequences, *Bioinformatics*, **34**, 3094–100.

24. Li, H. 2013, May 26, Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *arXiv*.

25. Hu, J., Fan, J., Sun, Z., and Liu, S. 2020, NextPolish: a fast and efficient genome polishing tool for long-read assembly, *Bioinformatics*, **36**, 2253–5.

26. Guan, D., McCarthy, S.A., Wood, J., Howe, K., Wang, Y., and Durbin, R. 2020, Identifying and removing haplotypic duplication in primary genome assemblies, *Bioinformatics*, **36**, 2896–8.

27. Jackman, S.D., Coombe, L., Chu, J., et al. 2018, Tigmint: correcting assembly errors using linked reads from large molecules, *BMC Bioinf.*, **19**, 393.

28. Coombe, L., Zhang, J., Vandervalk, B.P., et al. 2018, ARKS: chromosome-scale scaffolding of human genome drafts with linked read kmers, *BMC Bioinf.*, **19**, 234.

29. Warren, R.L., Vandervalk, B.P., Jones, S.J., and Birol, I. 2015, LINKS: scaffolding genome assemblies with kilobase-long nanopore reads, *bioRxiv*.

30. Zhou, C., McCarthy, S. A., and Durbin, R. 2022, YaHS: yet another Hi-C scaffolding tool. *bioRxiv*, 2022.06.09.495093.

31. Li, H., Handsaker, B., Wysoker, A., et al.; 1000 Genome Project Data Processing Subgroup. 2009, The sequence alignment/map format and SAMtools, *Bioinformatics*, **25**, 2078–9.

32. Laetsch, D.R. and Blaxter, M.L. 2017, BlobTools: interrogation of genome assemblies [version 1; peer review: 2 approved with reservations], *F1000Research*, **6**, 1287.

33. Marçais, G., Delcher, A.L., Phillippy, A.M., Coston, R., Salzberg, S.L., and Zimin, A. 2018, MUMmer4: a fast and versatile genome alignment system, *PLoS Comput. Biol.*, **14**, e1005944.

34. Challis, R., Richards, E., Rajan, J., Cochrane, G., and Blaxter, M. 2020, BlobToolKit – interactive quality assessment of genome assemblies, *G3 (Bethesda)*, **10**, 1361–74.

35. Flynn, J.M., Hubley, R., Goubert, C., et al. 2020, RepeatModeler2 for automated genomic discovery of transposable element families. Proc Natl Acad Sci U S A, 117, 9451–7.

36. Dobin, A., Davis, C.A., Schlesinger, F., et al. 2013, STAR: ultrafast universal RNA-seq aligner, *Bioinformatics*, **29**, 15–21.

37. Pertea, M., Pertea, G.M., Antonescu, C.M., Chang, T.-C., Mendell, J.T., and Salzberg, S.L. 2015, StringTie enables improved reconstruction of a transcriptome from RNA-seq reads, *Nat. Biotechnol.*, **33**, 290–5.

38. Niknafs, Y.S., Pandian, B., Iyer, H.K., Chinnaiyan, A.M., and Iyer, M.K. 2017, TACO produces robust multisample transcriptome assemblies from RNA-seq, *Nat. Methods*, **14**, 68–70.

39. Haas, B.J., Salzberg, S.L., Zhu, W., et al. 2008, Automated eukaryotic gene structure annotation using EVidenceModeler and the program to assemble spliced alignments, *Genome Biol.*, **9**, R7.

40. Mapleson, D., Venturini, L., Kaithakottil, G., and Swarbreck, D. 2018, Efficient and accurate detection of splice junctions from RNA-seq with Portcullis, *GigaScience*, **7**.

41. Iwata, H. and Gotoh, O. 2012, Benchmarking spliced alignment programs including Spaln2, an extended version of Spaln that incorporates additional species-specific features, *Nucleic Acids Res.*, **40**, e161.

42. Alioto, T., Blanco, E., Parra, G., and Guigó, R. 2018, Using geneid to Identify Genes, *Curr Protoc Bioinformatics*, **64**, e56.

43. Stanke, M., Schöffmann, O., Morgenstern, B., and Waack, S. 2006, Gene prediction in eukaryotes with a generalized hidden Markov model that uses hints from external sources, *BMC Bioinf.*, **7**, 62.

44. Lomsadze, A., Burns, P.D., and Borodovsky, M. 2014, Integration of mapped RNA-Seq reads into automatic training of eukaryotic gene finding algorithm. *Nucleic Acids Res*, **42**, e119.

45. Conesa, A., Götz, S., García-Gómez, J.M., Terol, J., Talón, M., and Robles, M. 2005, Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research, *Bioinformatics*, **21**, 3674–6.

46. Buchfink, B., Reuter, K., and Drost, H.-G. 2021, Sensitive protein alignments at tree-of-life scale using DIAMOND, *Nat. Methods*, **18**, 366–8.

47. Jones, P., Binns, D., Chang, H.-Y., et al. 2014, InterProScan 5: genome-scale protein function classification, *Bioinformatics*, **30**, 1236–40.

48. Cui, X., Lu, Z., Wang, S., Jing-Yan Wang, J., and Gao, X. 2016, CMsearch: simultaneous exploration of protein sequence space and structure space improves not only protein homology detection but also protein structure prediction, *Bioinformatics*, **32**, i332–40.

49. Nawrocki, E.P. and Eddy, S.R. 2013, Infernal 1.1: 100-fold faster RNA homology searches, *Bioinformatics*, **29**, 2933–5.

50. Chan, P.P. and Lowe, T.M. 2019, tRNAscan-SE: searching for tRNA genes in genomic sequences, *Methods Mol. Biol.*, **1962**, 1–14.

51. Kolmogorov, M., Yuan, J., Lin, Y., and Pevzner, P.A. 2019, Assembly of long, error-prone reads using repeat graphs, *Nat. Biotechnol.*, **37**, 540–6.

52. Bernt, M., Donath, A., Jühling, F., et al. 2013, MITOS: improved de novo metazoan mitochondrial genome annotation, *Mol. Phylogenet. Evol.*, **69**, 313–9.

53. Rovatsos, M., Vukić, J., Mrugała, A., Suwala, G., Lymberakis, P., and Kratochvíl, L. 2019, Little evidence for switches to environmental sex determination and turnover of sex chromosomes in lacertid lizards, *Sci. Rep.*, **9**, 7832.

54. Vujošević, M. and Blagojević, J. 1999, The distribution of constitutive heterochromatin and nucleolus organizers in lizards of the family Lacertidae (Sauria), *Genetika*, **31**, 269–76.

55. Suwala, G., Altmanová, M., Mazzoleni, S., et al. 2020, Evolutionary variability of W-linked repetitive content in lacertid lizards, *Genes*, **11**, 531.

56. Olmo, E., Odierna, G., and Capriglione, T. 1987, Evolution of sex-chromosomes in lacertid lizards, *Chromosoma*, **96**, 33–8.

57. Pasquesi, G.I.M., Adams, R.H., Card, D.C., et al. 2018, Squamate reptiles challenge paradigms of genomic repeat element evolution set by birds and mammals, *Nat. Commun.*, **9**, 2774.

58. Formenti, G., Rhie, A., Balacco, J., et al.; Vertebrate Genomes Project Consortium. 2021, Complete vertebrate mitogenomes reveal widespread repeats and gene duplications, *Genome Biol.*, **22**, 120.