DATA NOTE

REVISED ERGA-CBP chromosome-level genome assembly of the blind scorpion *Belisarius xambeui* Simon, 1879 (Belisariidae, Scorpiones), a singular scorpion in Europe

[version 2; peer review: 3 approved, 1 approved with reservations]
Previously titled: "ERGA-BGE chromosome-level genome assembly of the blind scorpion *Belisarius xambeui* Simon, 1879 (Belisariidae, Scorpiones), the most singular scorpion in Europe"

Sara Guirao-Rico[1,2], Vadim A. Pisarenco[1,2], Paula Escuer[3], Pau Balart-García[4], Marc Domènech[1,2], Adrià Bellvert[2,5], Silvia Adrián-Serrano[2,6], Alejandro Sánchez-Gracia[1,2], Miquel A. Arnedo[2,6], Julio Rozas[1,2]

[1]Universitat de Barcelona Departament de Genetica Microbiologia i Estadistica, Barcelona, Catalonia, 08028, Spain
[2]Universitat de Barcelona Institut de Recerca de la Biodiversitat, Barcelona, Catalonia, 08028, Spain
[3]Universite de Neuchatel Institut de Biologie, Neuchâtel, Canton of Neuchâtel, Switzerland
[4]Fish Evolution Research Group, Department of Zoology, Faculty of Science, Charles University, Prague, Czech Republic
[5]Universitat Greifswald Zoologisches Institut und Museum, Greifswald, Mecklenburg-Vorpommern, Germany
[6]Departament de Biologia Evolutiva Ecologia i Ciencies Ambientals, Universitat de Barcelona, Barcelona, Spain

v2 **First published:** 17 Feb 2026, **6**:53
https://doi.org/10.12688/openreseurope.22872.1
Latest published: 21 Apr 2026, **6**:53
https://doi.org/10.12688/openreseurope.22872.2

Abstract

We present a chromosome-level reference genome for the blind scorpion *Belisarius xambeui* Simon, 1879 (Belisariidae, Scorpiones). The genome size estimated by flow cytometry (4.32 Gb) closely matches the final assembly size (3.98 Gb). The final assembly comprises 19,045 scaffolds, including 56 chromosome-level scaffolds (pseudochromosomes) that account for 90.08% of the total assembly. Nucleotide diversity across the genome was low, with an average π of 0.0018.

Keywords

Long-read sequencing, reference genome, scorpion, European Reference Genome Atlas, Biodiversity Genomics Europe, Earth Biogenome

Open Peer Review

Approval Status ✓ ✓ ? ✓

	1	2	3	4
version 2 (revision) 21 Apr 2026				✓ view
version 1 17 Feb 2026	✓ view	✓ view	? view	✓ view

1. **Cheng-Min Shi**, Hebei Agricultural University, Baoding, China
2. **Carlos E. Santibáñez-López**, Western Connecticut State University, Danbury, USA
3. **Carlos Fernando Prada Quiroga**, University of Tolima, Ibague, Colombia
4. **Javier Blasco-Aróstegui**, Universidade de Lisboa, Lisbon, Portugal

This article is included in the Horizon Europe gateway.

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

Corresponding authors: Miquel A. Arnedo (marnedo@ub.edu), Julio Rozas (jrozas@ub.edu)

Author roles: Guirao-Rico S: Data Curation, Formal Analysis, Investigation, Supervision, Visualization, Writing – Original Draft Preparation, Writing – Review & Editing; **Pisarenco VA**: Data Curation, Formal Analysis, Investigation, Writing – Original Draft Preparation, Writing – Review & Editing; **Escuer P**: Data Curation, Writing – Review & Editing; **Balart-García P**: Resources, Writing – Review & Editing; **Domènech M**: Resources, Writing – Review & Editing; **Bellvert A**: Resources, Writing – Review & Editing; **Adrián-Serrano S**: Formal Analysis, Resources, Writing – Review & Editing; **Sánchez-Gracia A**: Funding Acquisition, Supervision, Writing – Original Draft Preparation, Writing – Review & Editing; **Arnedo MA**: Funding Acquisition, Project Administration, Resources, Supervision, Visualization, Writing – Review & Editing; **Rozas J**: Funding Acquisition, Supervision, Visualization, Writing – Original Draft Preparation, Writing – Review & Editing

Competing interests: No competing interests were disclosed.

Grant 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 study was supported by the Institut d'Estudis Catalans (IEC) under the Catalan Initiative for the Earth Biogenome Project (CBP; primera convocàtoria). The CPB is part of the global Earth BioGenome Project (EBP) and it is affiliated with the European Reference Genome Atlas (ERGA). This work was also supported by Ministerio de Ciencia e Innovación (MCIN/AEI/10.13039/501100011033) grants (PID2019-103947GB, PID2019-105794GB, PID2022-137758NB-I00, PID2022-138477NB, and an FPI fellowship to V.A.P. PRE2020-095592) and FPU fellowship to SAS (FPU18/04086). Additional funding was also provided by Comissió Interdepartamental de Recerca i Innovació Tecnològica (2021SGR00279, 2021SGR0689).

How to cite this article: Guirao-Rico S, Pisarenco VA, Escuer P *et al.* **ERGA-CBP chromosome-level genome assembly of the blind scorpion *Belisarius xambeui* Simon, 1879 (Belisariidae, Scorpiones), a singular scorpion in Europe [version 2; peer review: 3 approved, 1 approved with reservations]** Open Research Europe 2026, **6**:53 https://doi.org/10.12688/openreseurope.22872.2

First published: 17 Feb 2026, **6**:53 https://doi.org/10.12688/openreseurope.22872.1

REVISED **Amendments from Version 1**

1) The title has been revised. In the new version, the term "BGE" has been removed because the report was not funded by BGE. Additionally, the second part of the title has been modified following a reviewer's suggestion.

2) We have addressed the majority of the reviewers' comments, resolved their concerns, and incorporated additional relevant references.

3) Figures 1 and 2 have been improved and updated in accordance with the reviewers' recommendations.

4) All accession numbers have now been included.

Any further responses from the reviewers can be found at the end of the article

Background

Scorpions represent an iconic lineage of arthropods, long recognized for their unique body plan, ancient fossil record, and venom potency. The blind scorpion *Belisarius xambeui* Simon 1879 is a small species (~3 cm) that constitutes a narrow Catalan endemism (Simon, 1879; Lourenço, 2015). It is distributed in the eastern Pyrenees and Pre-Pyrenees and northernmost part of the Catalan Pre-Costal mountain range (Figure 1). It is usually found under leaf litter and buried stones in humid and shady habitats, as well as in caves. Its morphology is highly modified, characterized by the absence of eyes, and the extreme reduction in the number of pectinal teeth (only 4–6), which serve as the chemosensory organ (Figure 2). This species was the first troglomorphic scorpion ever described, and it is considered one of the most cryptic in Europe, with very small and highly threatened populations (Vignoli *et al.*, 2008). Moreover, its low fertility (5–24 eggs per clutch, Auber, 1959) may further compromise its long-term survival. Indeed, the species is listed as threatened with extinction in the Catalan catalog of threatened wild fauna (https://portaldogc.gencat.cat/utilsEADOP/PDF/8758/1927723.pdf).

The taxonomic position of *B. xambeui* remains uncertain. It was initially allied with the family Troglotayosicidae, which comprises a single genus of humicolous and cave-dwelling species native to Ecuador and Colombia, in South America. More recently, however, the family Belisariidae originally described by Lourenço (1998) has been resurrected to

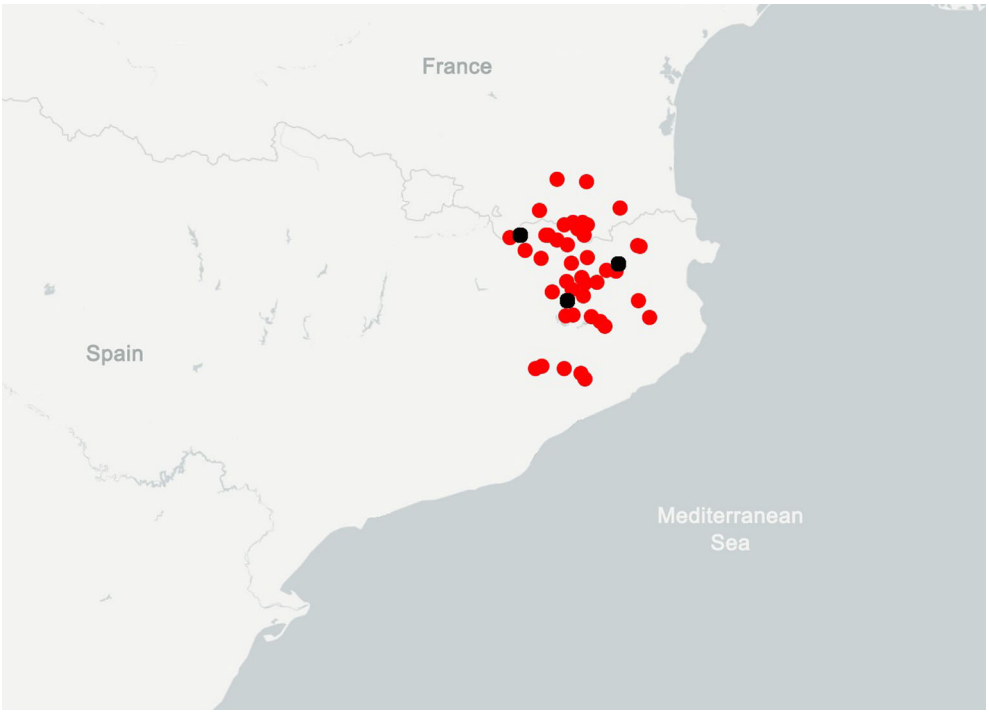

Figure 1. Distribution of *B. xambeui*. Modified from GBIF.org, it includes *B. xambeui* sampling sites (black points) and records from museums, zoological collections and citizen science platforms (red points).

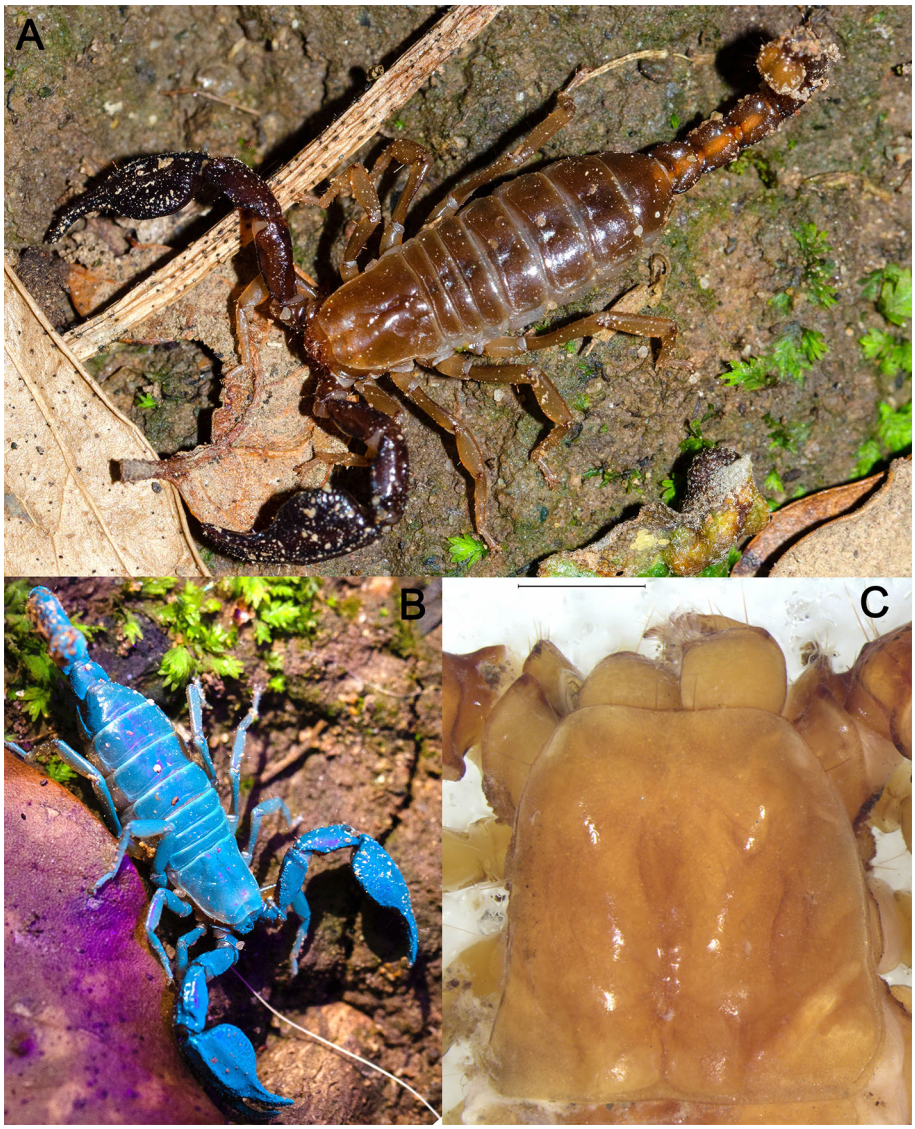

Figure 2. A) An adult specimen of *B. xambeui* in its habitat, in Osona (Catalonia). B) A specimen of *B. xambeui* glowing blue under black light (UV-A). Compounds in the scorpion's cuticle absorb ultraviolet light and re-emit it as visible fluorescence. C) The eyeless carapace of a young specimen under the stereomicroscope; scale bar: 1 mm.

accommodate the blind scorpion along with a recently described cave-dwelling species from southern Iberian Peninsula (*B. ibericus* Lourenço, 2015), as well as the newly established genus and species *Sardoscorpius troglophilus* Tropea & Onnis, 2020, found in caves in Sardinia (Tropea & Onnis, 2020).

This new genome, distantly related to the species with previously available genomic data, will provide new insights into the phylogenetic position of *B. xambeui* within the scorpions tree of life. At present, genomic data are available for only seven scorpion species, namely the hadrurid *Hadrurus arizonensis* Ewing, 1928, the buthids *Olivierus martensii* (Karsch, 1879) (=*Mesobuthus martensii*), *Olivierus przewalskii* (=*Mesobuthus przewalskii*) (Birula, 1897), *Androctonus mauritanicus* (Pocock, 1902), *Centruroides vittatus* (Say, 1821), *Centruroides sculpturatus* Ewing, 1928, and the euscorpiid *Euscorpius feti* Tropea, 2013 (Cao *et al.*, 2013; Schwager *et al.*, 2017; El Ghoubali *et al.*, 2022; Bryant *et al.*, 2024; Yamashita *et al.*, 2024; Su *et al.*, 2025; National Center for Biotechnology Information (NCBI) genome dataset, https://www.ncbi.nlm.nih.gov/datasets/genome/). Scorpions comprise two major clades: Buthida, which includes buthids and other families with generally stronger venoms, and Iurida, which includes belisariids, hadrurids

and euscorpiids, among others, commonly with less potent venoms. Recent estimates suggest that these two main lineages diverged approximately 291–320 million years ago (Santibáñez-López *et al.*, 2022). The new genome of *B. xambeui* will also yield valuable information on the genomic basis of fundamental biological traits, such as changes in the venom composition and sensory perception (i.e., vision and chemosensation) (Garb *et al.*, 2018). More broadly, the inclusion of this genome will substantially enhance the currently limited genomic resources available for non-insect arthropods. Despite the remarkable diversity and ecological relevance of chelicerates (about 100,000 species), less than 100 species have been sequenced at the chromosome level. Within the Scorpiones order (about 2,500 species), the *Belisarius* genome represents the third high-quality reference genome. It thus constitutes an important resource for studying chelicerate genome structure and evolution, offering a valuable genome for the broader scientific community.

Material and methods

Sample collection, processing and genome size estimation

Three specimens of *B. xambeui* (Figure 2) were collected in three different locations in Catalonia (Spain) and in three different years. One individual, sampled in 2019 in the Cova de la Mosquera (Beuda, Girona), was used for the PacBio genome sequencing, and whole-genome Illumina resequencing. The second individual, sampled in 2020 in Cantonigròs, was used for flow cytometry. The third one, sampled in 2021 in Queralbs, was used to generate the Hi-C libraries. Because specimens of *B. xambeui* are rare, and in order to secure the highest quality of DNA, we used specimens in the best condition of preservation, which motivated the decision to use specimens from different localities. Sampling was conducted with authorization from the Generalitat de Catalunya's Departament de Territori i Sostenibilitat, and the help of the Federació Catalana d'Espeleologia. All individuals except the one used for flow cytometry were preserved either in ethanol 100% at -20ºC or flash frozen at -80ºC (Table S1). We estimated the genome size from a single adult individual (Table S1) using flow cytometry as in Sánchez-Herrero *et al.* (2019). *B. xambeui* specimens are easily identifiable by the absence of eyes and the reduced number of teeth on the pectines, which clearly distinguish them from the other two scorpion species inhabiting the region, *Buthus occitanus* (Pierre-Joseph Amoreux, 1789) and *Tetratrichobothrius flavicaudis* De Geer, 1778. Furthermore, the species identity of all individuals was confirmed in the laboratory using the mitochondrial cytochrome c oxidase subunit I (COI) as DNA barcode prior to DNA sequencing. The sex of the sequenced individuals is unknown, because a careful examination under the microscope was not possible since the specimens needed to remain frozen in order to ensure correct preservation for genome sequencing. Vouchers are deposited at the Centre de Recursos de Biodiversitat Animal of the Universitat de Barcelona (codes MZB2026-0200 and MZB2026-0201).

DNA extraction

We extracted a high molecular weight genomic DNA (HMW gDNA) from a single flash-frozen individual using the Blood & Cell Culture DNA Mini Kit (Quiagen) for both long and short read sequencing (Table S2). We measured the DNA extraction quality using a NanoDrop spectrophotometer, yielding absorbance ratios of 260/280 = 1.74 and 260/230 = 2.26. We built Arima Hi-C libraries at the Genomics Unit of the Institute of Evolutionary Biology (Spain; https://www.ibe.upf-csic.es/genomics-unit) following the manufacturer's protocol (Arima Genomics) using a single, flash-frozen whole specimen (Table S2).

Genome sequencing

HMW gDNA was sequenced using PacBio CLR technology, which resulted in average coverage of 84x complemented by short-read sequencing on the Illumina NovaSeq 6000 platform (150PE; ~27x of coverage), both carried out at Novogene. For the scaffolding process, the Arima Hi-C libraries were sequenced using Illumina NovaSeq 6000 platform (150PE; 66x of coverage) at the Centre Nacional d'Anàlisi Genòmica (CNAG) sequencing unit (Spain) (Table S2).

Genome assembly and manual curation

De novo genome assembly was performed in two steps. First, PacBio reads were converted to FASTA format using the commands `bam2fq` from samtools v.1.9 and `seq -A` from seqtk v.1.3-r117-dirty. The resulting FASTA reads were assembled with wtdbg2 v.2.5 (Ruan & Li, 2020). To ensure optimal assembly quality, we generated six independent assemblies, differing in three key parameters of wtdbg2: minimum read length (`-L`), minimum alignment length (`-l`), and enabling or not the realignment mode (`-R`) (Table S3a). PacBio reads were aligned back to each assembly with minimap v.2.22 (Li, 2018) for polishing. Assembly statistics, including the number of contigs, mean and the median of the contig length, and N50, were computed using the script `seq_n50.pl` (provided by wtdbg2 software) (Table S3b). Assembly completeness was determined using BUSCO-v2.5.4 (Seppey *et al.*, 2019), with the odb10 databases (Arachnida; 2,934 genes; Arthropoda; 1,013 genes and Eukaryota; 255 genes) (Kriventseva *et al.*, 2019).

Potential haplotype duplications (i. e., haplotigs and contig overlaps) were evaluated with purge_dups v1.2.6 (Guan *et al.*, 2020), with default parameters and the cutoffs rechecked `-m 48 -u 288`. minimap v.2.22 (Li, 2018) was also used within

the purge_dups pipeline. To assess and remove putative organismal contaminations, we used the BlobTools2 from BlobTools kit pipeline (Kumar *et al.*, 2013; Laetsch & Blaxter, 2017), using the NCBI nucleotide and UniProt databases (updated in July 2021 and June 2021, respectively) (https://www.ncbi.nlm.nih.gov/nucleotide/, The UniProt Consortium, 2021), along with the aforementioned BUSCO odb10 databases.

To further improve the quality of the assembly, we addressed potential mitochondrial contamination. Putative mitochondrial contigs were identified to enable their removal or masking from the nuclear genome assembly. Specifically, all assembled contigs were queried using BLAST against two reference datasets: i) the *B. xambeui* mitochondrial assembly, and ii) a set of 11 complete mitochondrial genomes from 10 different scorpion species (Table S4). Only BLAST hits with an E-value $\leq 1\times10^{-50}$ were considered significant and thus subject to masked or removal. Hits shorter than 1,000 bp were disregarded. For contigs exhibiting significant mitochondrial hits, a conditional masking strategy was applied: i) if the non-mitochondrial portion of the contig exceeded 3,000 bp, the identified mitochondrial segment was excised and replaced with 15-base 'N' placeholder sequence to preserve the putative nuclear portion of the contig; ii) if the remaining non-mitochondrial portion of the contig was 3,000 bp or less, the entire contig was discarded from the assembly. This criterion ensured the retention of high-confidence nuclear contigs while removing sequences likely to be mitochondrial or too short for reliable assembly.

We obtained a chromosome-level assembly by scaffolding the draft genome using the Arima Hi-C data. First, raw Hi-C reads were preprocessed and quality-trimmed using fastp with default parameters (Chen *et al.*, 2018). Second, trimmed paired Hi-C reads were mapped against the draft assembly using bwa mem-v0.7.17-r1198-dirty (Li & Durbin, 2009), designating the 5' coordinate as the primary alignment and skipping rescuing mates and read pairing (`-5SP`). At this step, the minimum mapping quality was set to 0 (`-T0`). Third, high-quality ligation events (Hi-C pairs) were identified from the resulting SAM file using the pairtools parse command from Pairtools-v1.1.2 (Open2C *et al.*, 2024). In this step, reads with a minimal mapping score below 30 were considered as multi-mapping (`-min-mapq 30`). In addition, we filtered out complex walks (multiple ligation events), retaining only the 5'-most unique alignment on each side (`-walks-policy 5unique`), and ignoring gaps of up 30 bp between alignments (`-max-inter-align-gap 30`). Fourth, the resulting pairsam file was sorted with ("pairtools sort"), PCR duplicates were removed ("pairtools dedup"), and the final BAM files were generated ("pairtools split"). Finally, the wtdbg2 draft genome assembly and the processed Hi-C BAM file were used as input for YaHS-v1.2.2 (Zhou *et al.*, 2023), run with default parameters, to generate the chromosome-scale scaffolds. The resulting assembly was visually inspected and manually curated by generating the Hi-C contact map using JuicerTools-v1.2.2 (Durand *et al.*, 2016b) and examined them in JuiceBox-v1.11.08 (Durand *et al.*, 2016a). This inspection enabled the identification of potential misjoints and misplaced scaffolds. Manual curation resolved two misjoins and corrected a few incorrectly placed scaffolds.

Repetitive elements analysis

We used RepeatModeler-v2.0.6 (Flynn *et al.*, 2020) for *de novo* identification of repetitive elements (REs), using the LTR discovery pipeline (`-LTRStruct`). Sequences classified as "unknown" were subsequently reclassified with DeepTE (Yan *et al.*, 2020) using the metazoan model (`-sp M`). The resulting RE libraries were merged into a single comprehensive database, which was then used as input for masking with RepeatMasker-v4.1.7 (http://repeatmasker.org), configured with Dfam 3.8 (Storer *et al.*, 2021), an executed using the sensitive search mode (`-s`). Transposable elements (TEs) identified by RepeatMasker were further classified using custom Python scripts based on their class, subclass, order and superfamily, following the hierarchical classification of Wicker *et al.* (2007) and Wells & Feschotte (2020).

Nucleotide diversity

We estimated the per-site nucleotide diversity (π) with ANGSD v0.940-dirty (Korneliussen *et al.*, 2014), using Illumina resequencing data from a single individual, sequenced at an average coverage depth of ~27X. Firstly, we assessed the quality of raw reads using FASTQC v0.11.9 (Andrews (2010), http://www.bioinformatics.babraham.ac.uk/projects/fastqc/). Reads were trimmed with Trimmomatic v0.39 (Bolger *et al.*, 2014), to filter out adapters, reads shorter than 50bp (`MINLEN:50`), and low-quality bases using a sliding window of 4bp with an average Phred score of 15 (`SLIDINGWINDOW:4:15`). We also removed leading and trailing bases with low quality scores (<3) (`LEADING:3, TRAILING:3`), as well as all missing bases. Secondly, filtered reads were mapped against the reference genome of *B. xambeui* using bwa mem v0.7.16 (Li & Durbin, 2009). Next, the generated SAM file was converted to BAM format, sorted and indexed using samtools v1.21 (Danecek*et al.*, 2021). Thirdly, we estimated the sample allele frequency (SAF) likelihoods with ANGSD using the BAM file as input. We applied specific quality filters to filter out positions with very low or very high coverage (`-setMinDepth 10, -setMaxDepth 90`). In addition, we only retained positions with high base calling quality, properly paired and high mapping quality reads (`-minQ 20 -only_proper_pairs -minMapQ 20`). We set the mapQ parameter for excessive mismatches (`-C 50`) and the qscores around indels (`-baq 1`). We discarded reads that did not map uniquely (`-uniqueOnly 1`), as well as those not primary, failure and duplicate

reads (`-remove_bads 1`). We set the *B. xambeui* reference genome as both the reference (`-ref`) and the ancestral (`-anc`) sequence, and estimated the genotype likelihoods per-site using the GATK algorithm (`-GL 2`) assuming Hardy-Weinberg Equilibrium (`-doSaf 1`). In addition, we used the `-doCounts 1` option to count the bases per site after applying the abovementioned depth and base quality filters. The resulting SAF file was used as input for the realSFS program to estimate the folded site frequency spectrum (SFS) (`-fold 1`), and to calculate the thetas per site (realSFS saf2theta). Finally, we retrieved the nucleotide diversity (π; Nei, 1987) for each chromosome using the thetaStat program (thetaStat do_stat) within ANGSD (see also (Kousathanas *et al.*, 2017)).

Mitochondrial assembly and annotation

To assemble the mitogenome of *B. xambeui*, we first performed a BLAST using a mtDNA sequence of the scorpion *Centruroides vittatus* (GenBank ID: MF975702.1) to retrieve all publicly available complete scorpion mitochondrial genomes, resulting in a dataset of 11 complete mtDNA sequences (Table S4). These genomes were subsequently used as queries in a BLAST search against PacBio long-read dataset from *B. xambeui* to identify mitochondrial reads, applying an E-value threshold of < 1e-4. All reads with significant matches against these scorpion mitogenomes were subsequently assembled *de novo* using wtdbg2 v.2.5 (Ruan & Li, 2020), and the resulting assembly was polished using PacBio long

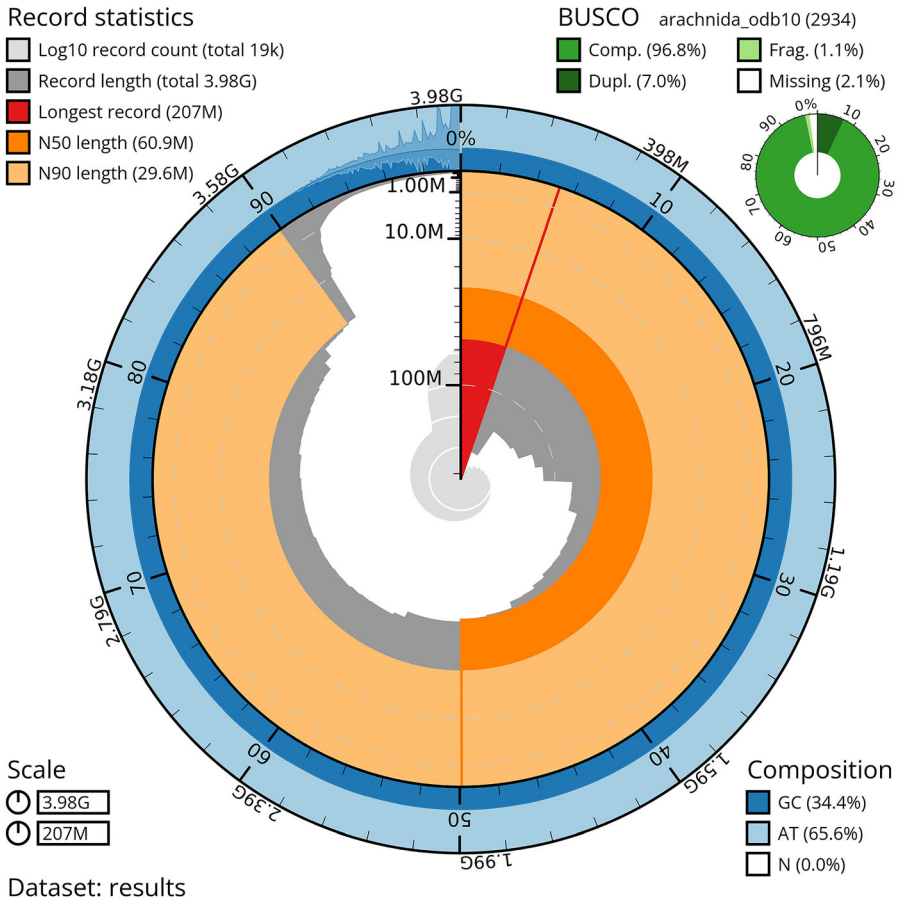

Figure 3. Snail plot summary of genome assembly of *B. xambeui*. The BlobToolKit snail plot shows N50 metrics and BUSCO gene completeness. The main plot is divided into 1,000 bins around the circumference with each bin representing 0.1% of the 3,979,917,391 bp assembly. The distribution of sequence lengths is shown in dark grey, with the plot radius scaled to the longest sequence in the assembly (207,115,667 bp, shown in red). Orange and pale-orange arcs indicate the N50 and N90 sequence lengths (60,866,488 and 29,580,904 bp), respectively. The pale grey spiral shows the cumulative sequence count on a log scale, with white scale lines indicating successive orders of magnitude. The blue and pale-blue outer bands show the distribution of GC, AT and N percentages across the same bins as the inner plot. A summary of complete, fragmented, duplicated and missing BUSCO genes from the arachnida_odb10 dataset is shown in the top right.

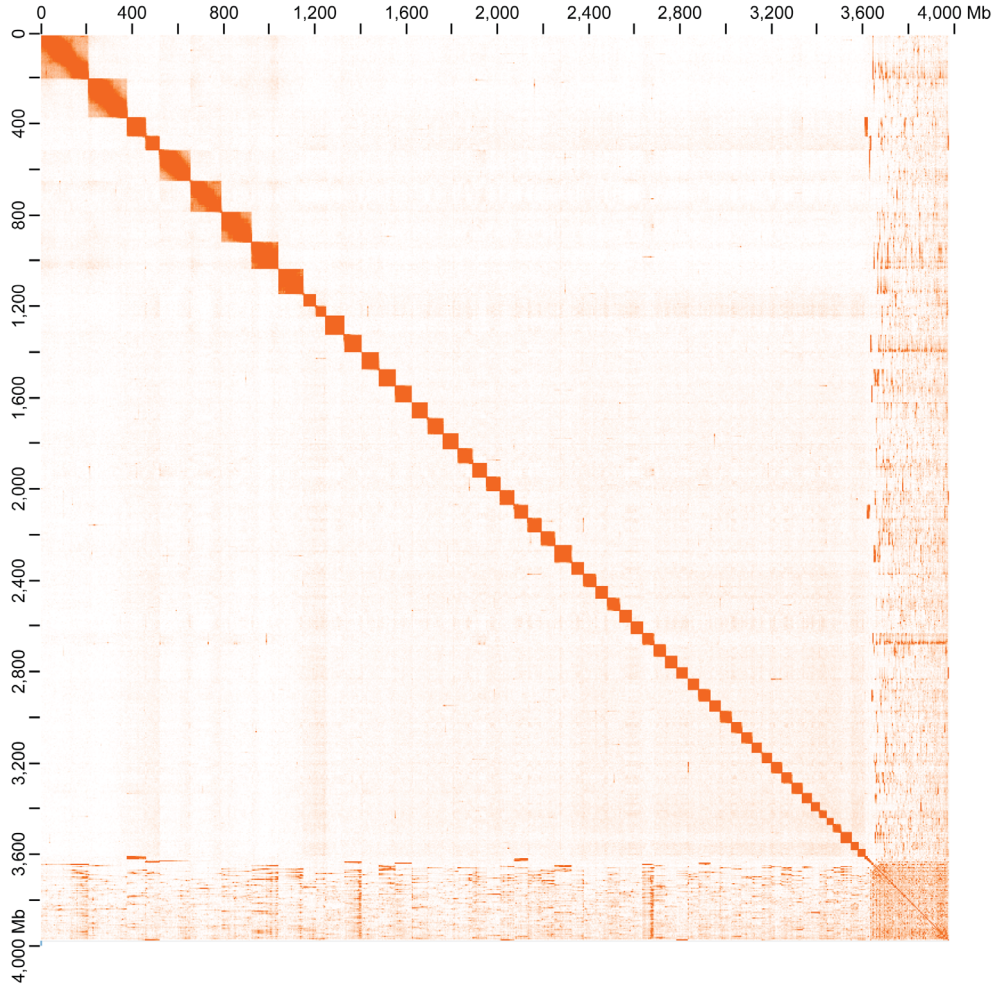

Figure 4. Hi-C contact map of the chromosome-level genome assembly of *B. xambeui* visualized using JuiceBox.

Table 1. Genome assembly and annotation statistics.

Assembly level	Chromosome
Assembly size (Mb)	3,979.92
GC Content (%)	34.37%
Number of scaffolds	19,045
Longest scaffold (Mb)	207.1
Scaffold N50 (Mb) / L50	60.9 / 21
Scaffold N90 (Mb) / L90	29.6 / 55
Chromosomes[a] & Sequence in chromosomes (%)	56 / 90.08%
BUSCO completeness[b]	
Genome	C:96.8% [S:89.8%,D:7.0%], F:1.1%,M:2.1%,n:2,934
Repetitive regions (%)	68.4%
Nucleotide diversity	0.0018

[a]Haploid number of chromosomes.
[b]BUSCO scores based on the arachnida_odb10 BUSCO set using version 5.4.2. C = complete [S = single copy, D = duplicated], F = fragmented, M = missing, n = number of orthologues in comparison.

reads. Mitochondrial genome annotation was performed using MITOS2 (Bernt *et al.*, 2013, available at http://mitos2.bioinf.uni-leipzig.de/index.py), with default parameters, to infer protein-coding genes (PCGs) and ribosomal RNAs (rRNAs). Transfer RNAs (tRNAs) were annotated using both MITOS2 and ARAGORN (Laslett & Canback, 2004), allowing inference of their primary sequences and secondary structures. Gene boundaries were subsequently manually curated based on alignments between the mitochondrial genome of species *B. xambeui* and eight scorpion mitogenomes (Table S5) retrieved from public databases.

Results

Nuclear genome sequence report

Flow cytometry analysis estimated the *B. xambeui* genome size at 4.32 Gb, approximately four times larger than those reported for other scorpion species (0.88 Gb–1.32 Gb) (Gregory (2026), http://www.genomesize.com), and nearly two times than those estimated in *Hadrurus arizonensis* based on the total assembly size (Bryant *et al.*, 2024), and similar to that reported for *Euscorpius feti* (4.4 Gb) (National Center for Biotechnology Information (NCBI) genome dataset, https://www.ncbi.nlm.nih.gov/datasets/genome/). The optimal assembly (Table S3b), selected based on its superior contiguity and completeness, comprised 24,679 contigs, 373 of which were identified as potential contaminants (assigned to Proteobacteria and other non-target taxa, based on BLAST and Diamond hits) and were excluded. We also identified 96 contigs with significant mitochondrial BLAST hits (E-value < 1e-50), 7 were completely removed, 21 were masked and 89 were disregarded since they did not pass the filter of the BLAST E-value or the total length that match mtDNA was below 1,000 bp. The resulting *B. xambeui* genome assembly was ~3.97 Gb, including 24,299 contigs, with a contig N50 ~1.07 Mb. This assembly had a high completeness, with 91.4%, 89.8% and 87.9% of complete BUSCO genes identified across the Arachnida, Arthropoda and Eukaryota datasets, respectively. After the assembly and polishing step, we used BlobTools2 from BlobTools kit (Kumar *et al.*, 2013; Laetsch & Blaxter, 2017) to assess the assembly quality, identify and remove potential contaminant scaffolds.

The scaffolded assembly comprised 19,045 scaffolds with a N50 of ~207 Mb, including approximately 56 chromosome-level scaffolds (pseudochromosomes) that contain 90.08% of the assembled sequences (Figure 3, Figure 4, Table 1 and Table S6). Remarkably, the complete BUSCO scores increased to 96.8%, 96.5% and 97.3% across the Arachnida, Arthropoda and Eukaryota datasets, respectively.

Mitogenome sequence report

The mitochondrial genome assembly of *B. xambeui* yielded an almost complete mitogenome of 14,529 bp, consistent with sizes reported for related scorpions. Assembly accuracy was validated by BLAST comparison against the available COI sequence of *B. xambeui* (GenBank ID: JN018156), showing near-identical sequence similarity. We found that the assembly was likely incomplete at the *nad5* gene, with a missing fragment of ~1,018 bp, that the *atp6* gene includes a deletion that disrupted the reading frame. Automatic annotation recovered all protein coding genes, both rRNAs, and all tRNAs except *trnK*, which was manually annotated by comparison with homologous sequences. Manual curation confirmed canonical secondary structures for all tRNAs and corrected minor issues in *trnF*, *trnL1*, *trnM* and *trnS1*. Among the 22 tRNAs, five (*trnC*, *trnD*, *trnF*, *trnG*, *trnH*) lacked the T arm, *trnS1* lacked the D arm, whereas the remaining had full structures. *trnF* showed two possible secondary structures, and *trnR* exhibited an unusual "bubble" caused by an asymmetric mismatch in the acceptor stem. Notably, we identified a >600 bp insertion between *trnW* and *trnC*, which is absent from other scorpion mitogenomes, with similarity to the mitochondrial control region. We validated this issue by PCR, using primers designed at the coding regions of *nad2* and *cox1*, confirming that this insertion is not an assembly artifact, but likely represents a duplicated fragment of the control region that has been inserted between *nad2* and *cox1* in *B. xambeui.*

Nucleotide diversity

We found that the levels of nucleotide diversity are low, with an average of $\pi = 0.0018$, relatively homogeneous across chromosomes (ranging from a maximum of $\pi = 0.0028$ and a minimum of $\pi = 0.0011$. The values are nearly identical when excluding repetitive regions (Table S7). Although these values indicate a somewhat reduced level of genetic variation, they are likely not low enough to raise immediate concern regarding a severe loss of polymorphism or compromised adaptive potential. However, since these estimates are derived from a single individual they should be interpreted with caution, as they may not fully capture population-level variation. Moreover, although nucleotide diversity could be an important criterion in conservation assessments, it should not be considered as the sole factor guiding scientific-based conservation recommendations. Although some studies have reported a relationship between nucleotide diversity and evolutionary responses, there is no simple general relationship between genetic diversity and the risk of species extinction (Ørsted *et al.*, 2019; Teixeira & Huber, 2021; Jeon *et al.*, 2024).

Data availability

Genome sequencing data have been submitted to the European Nucleotide Archive (ENA) (https://www.ebi.ac.uk/ena/browser/home) under the umbrella study PRJEB54934, including PRJEB54932 (raw sequencing data) and PRJEB54933 (chromosome-level genome assembly). The genome assembly is also accessible through NCBI (GCA_982267015.1).

Genomic data includes PacBio CLR long-reads (ERR10089999, ERR10090000, ERR10090001), Illumina short-reads (ERR10088334, ERR10088335, ERR10088336, ERR10088337) and Arima Hi-C data (ERR16820444-16820459). Information of the DNA sample of the individual PB0001 has been deposited under the accession number ERS12764351. Additional information of the genome assembly is available at https://github.com/molevol-ub/Belisarius_xambeui.

References

Auber M: **Observations sur le biotope et la biologie du scorpion aveugle: *Belisarius xambeui* E. Simon.** *Vie et Milieu.* 1959; **10**(2): 160–167.

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

Bolger AM, Lohse M, Usadel B: **Trimmomatic: a flexible trimmer for Illumina sequence data.** *Bioinformatics.* 2014; **30**(15): 2114–2120.
PubMed Abstract | Publisher Full Text | Free Full Text

Bryant MJ, Coello AM, Glendening AM, *et al.*: **Unveiling the genetic blueprint of a desert scorpion: a chromosome-level genome of *hadrurus arizonensis* provides the first reference for parvorder iurida.** *Genome Biol Evol.* 2024; **16**(5): evae097.
PubMed Abstract | Publisher Full Text | Free Full Text

Cao Z, Yu Y, Wu Y, *et al.*: **The genome of Mesobuthus martensii reveals a unique adaptation model of arthropods.** *Nat. Commun.* 2013; **4**: 2602.
PubMed Abstract | Publisher Full Text | Free Full Text

Chen S, Zhou Y, Chen Y, *et al.*: **fastp: an ultra-fast all-in-one FASTQ preprocessor.** *Bioinformatics.* 2018; **34**(17): i884–i890.
PubMed Abstract | Publisher Full Text | Free Full Text

Danecek P, Bonfield JK, Liddle J, *et al.*: **Twelve years of SAMtools and BCftools.** *GigaScience.* 2021; **10**: giab008.
PubMed Abstract | Publisher Full Text | Free Full Text

Durand NC, Robinson JT, Shamim MS, *et al.*: **Juicebox provides a visualization system for Hi-C contact maps with unlimited zoom.** *Cell Syst.* 2016a; **3**(1): 99–101.
PubMed Abstract | Publisher Full Text | Free Full Text

Durand NC, Shamim MS, Machol I, *et al.*: **Juicer provides a one-click system for analyzing loop-resolution Hi-C experiments.** *Cell Syst.* 2016b; **3**(1): 95–98.
PubMed Abstract | Publisher Full Text | Free Full Text

El Ghoubali D, Pirro S, Sehli S, *et al.*: **The complete genome sequence of Androctonus mauritanicus, the Moroccan black thick-tailed scorpion.** *Biodivers. Genomes.* 2022.
Publisher Full Text

Flynn JM, Hubley R, Goubert C, *et al.*: **RepeatModeler2 for automated genomic discovery of Transposable Element families.** *Proc Natl Acad Sci U S A.* 2020; **117**(17): 9451–9457.
PubMed Abstract | Publisher Full Text | Free Full Text

Garb JE, Sharma PP, Ayoub NA: **Recent progress and prospects for advancing arachnid genomics.** *Curr Opin Insect Sci.* 2018; **25**: 51–57.
PubMed Abstract | Publisher Full Text | Free Full Text

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

Jeon JY, Black AN, Heenkenda EJ, *et al.*: **Genomic diversity as a key conservation criterion: proof-of-concept from mammalian whole-genome resequencing data.** *Evol Appl.* 2024; **17**(9): e70000.
PubMed Abstract | Publisher Full Text | Free Full Text

Korneliussen TS, Albrechtsen A, Nielsen R: **ANGSD: Analysis of Next Generation Sequencing Data.** *BMC Bioinformatics.* 2014; **15**(1): 356.
PubMed Abstract | Publisher Full Text | Free Full Text

Kousathanas A, Leuenberger C, Link V, *et al.*: **Inferring heterozygosity from ancient and low coverage genomes.** *Genetics.* 2017; **205**(1): 317–332.
PubMed Abstract | Publisher Full Text | Free Full Text

Kriventseva EV, Kuznetsov D, Tegenfeldt F, *et al.*: **OrthoDB v10: sampling the diversity of animal, plant, fungal, protist, bacterial and viral genomes for evolutionary and functional annotations of orthologs.** *Nucleic Acids Res.* 2019; **47**(D1): D807–D811.
PubMed Abstract | Publisher Full Text | Free Full Text

Kumar S, Jones M, Koutsovoulos G, *et al.*: **Blobology: exploring raw genome data for contaminants, symbionts and parasites using taxon-annotated gc-coverage plots.** *Front Genet.* 2013; **4**: 237.
PubMed Abstract | Publisher Full Text | Free Full Text

Laetsch DR, Blaxter ML: **Blobtools: interrogation of genome assemblies [version 1; peer review: 2 approved with reservations].** *F1000Research.* 2017; **6**: 1287.
Publisher Full Text

Laslett D, Canback B: **ARAGORN, a program to detect tRNA genes and tmRNA genes in nucleotide sequences.** *Nucleic Acids Res.* 2004; **32**(1): 11–16.
PubMed Abstract | Publisher Full Text | Free Full Text

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

Li H, Durbin R: **Fast and accurate short read alignment with burrows–wheeler transform.** *Bioinformatics.* 2009; **25**(14): 1754–1760.
PubMed Abstract | Publisher Full Text | Free Full Text

Lourenço WR: **Panbiogeography, disrupted distributions and the concept of relictual family in scorpions.** *Biogeographica (Paris).* 1998; **74**(3): 133–144.

Lourenço WR: **The genus Belisarius Simon, 1879 (Scorpiones: Troglotayosicidae), with the description of a new vicariant species from the south of Spain.** *Comptes Rendus Biologies.* 2015; **338**: 362–367.
PubMed Abstract | Publisher Full Text

Nei M: **Molecular evolutionary genetics.** New York, NY: Columbia University Press, 1987.
Reference Source

Open2C, Abdennur N, Fudenberg G, *et al.*: **Pairtools: from sequencing data to chromosome contacts.** *PLoS Comput Biol.* 2024; **20**(5): e1012164.
PubMed Abstract | Publisher Full Text | Free Full Text

Ørsted M, Hoffmann AA, Sverrisdóttir E, *et al.*: **Genomic variation predicts adaptive evolutionary responses better than population bottleneck history.** *PLoS Genet.* 2019; **15**(6): e1008205.
PubMed Abstract | Publisher Full Text | Free Full Text

Ruan J, Li H: **Fast and accurate long-read assembly with wtdbg2.** *Nat Methods.* 2020; **17**(2): 155–158.
PubMed Abstract | Publisher Full Text | Free Full Text

Sánchez-Herrero JF, Frías-López C, Escuer P, *et al.*: **The draft genome sequence of the spider *Dysdera silvatica* (Araneae, Dysderidae): a valuable resource for functional and evolutionary genomic studies in chelicerates.** *GigaScience.* 2019; **8**(8): giz099.
PubMed Abstract | Publisher Full Text | Free Full Text

Schwager EE, Sharma PP, Clarke T, *et al.*: **The house spider genome reveals an ancient whole-genome duplication during arachnid evolution.** *BMC Biol.* 2017; **15**: 62.
PubMed Abstract | Publisher Full Text | Free Full Text

Seppey M, Manni M, Zdobnov EM: **BUSCO: assessing genome assembly and annotation completeness**. In: Kollmar M, editor. *Gene Prediction: Methods and Protocols.* New York, NY: Springer, 2019; 227–245.
Publisher Full Text

Santibáñez-López CE, Aharon S, Ballesteros JA, *et al.*: **Phylogenomics of Scorpions Reveal Contemporaneous Diversification of Scorpion Mammalian Predators and Mammal-Active Sodium Channel Toxins.** *Syst. Biol.* 2022; **71**: 1281–1289.
PubMed Abstract | Publisher Full Text

Simon E: **3e Ordre – Scorpiones.** In: *Les Arachnides de France. VII. Contenant les Ordres des Chernetes, Scorpiones et Opiliones.* Roret, Paris, 79–115. 1879.

Storer J, Hubley R, Rosen J, *et al.*: **The Dfam community resource of transposable element families, sequence models, and genome annotations.** *Mob DNA.* 2021; **12**(1): 2.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

Su Y-K, Xiu M-H, Yang H-Y, *et al.*: **A chromosome-level genome assembly for the desert scorpion Mesobuthus przewalskii from Asian drylands.** *J. Hered.* 2025; **116**: 532–539.
PubMed Abstract | **Publisher Full Text**

Teixeira JC, Huber CD: **The inflated significance of neutral genetic diversity in conservation genetics.** *Proc Natl Acad Sci U S A.* 2021; **118**(10): e2015096118.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

Tropea G, Onnis C: **A remarkable discovery of a new scorpion genus and species from Sardinia (Scorpiones: Chactoidea: Belisariidae).** *Arachnida – Rivista Aracnologica Italiana.* 2020; **26**: 3–25.

UniProt Consortium: **UniProt: the Universal Protein Knowledgebase in 2021.** *Nucleic Acids Res.* 2021; **49**(D1): D480–D489.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

Vignoli V, Klann AE, Michalik P: **Spermatozoa and sperm packages of the European troglophylous scorpion *Belisarius xambeui* Simon, 1879 (Troglotayosicidae, Scorpiones).** *Tissue Cell.* 2008; **40**(6): 411–416.
PubMed Abstract | **Publisher Full Text**

Wells JN, Feschotte C: **A field guide to eukaryotic Transposable Elements.** *Annu Rev Genet.* 2020; **54**: 539–561.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

Wicker T, Sabot F, Hua-Van A, *et al.*: **A unified classification system for eukaryotic transposable elements.** *Nat Rev Genet.* 2007; **8**(12): 973–982.
PubMed Abstract | **Publisher Full Text**

Yamashita T, Rhoads DD, Pummill J: **A robust genome assembly with transcriptomic data from the striped bark scorpion, Centruroides vittatus.** *G3 Genes|Genomes|Genetics.* 2024; **14**: jkae120.

Yan H, Bombarely A, Li S: **DeepTE: a computational method for *de novo* classification of transposons with Convolutional Neural Network.** *Bioinformatics.* 2020; **36**(15): 4269–4275.
PubMed Abstract | **Publisher Full Text**

Zhou C, McCarthy SA, Durbin R: **YaHS: Yet another Hi-C Scaffolding tool.** *Bioinformatics.* 2023; **39**(1): btac808.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

Open Peer Review

Current Peer Review Status: ✔ ✔ ? ✔

Version 2

Reviewer Report 23 April 2026

https://doi.org/10.21956/openreseurope.25418.r72791

✔ **Javier Blasco-Aróstegui**
Universidade de Lisboa, Lisbon, Lisbon, Portugal

The authors have appropriately addressed all the suggested edits and the manuscript is more than ready for indexing, comprising a fantastic contribution to scorpiology.

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Ecology - Evolution - Phylogenetics - Systematics

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.

Version 1

Reviewer Report 10 March 2026

https://doi.org/10.21956/openreseurope.24731.r69973

✔ **Javier Blasco-Aróstegui**
Universidade de Lisboa, Lisbon, Lisbon, Portugal

The article provides a reference chromosome-level genome for the troglomorphic scorpion *Belisarius xambeui*, endemic to France and Spain, and belonging to an incertae sedis lineage within the scorpion Tree of Life. This work has the potential to shed light on the evolution of this group

and help clarify its phylogenetic position within the Scorpion Tree of Life. I believe this contribution is timely and appropriate, and I therefore recommend the article for indexing in Open Research Europe.

However, several minor edits and comments should be addressed prior to acceptance. These are important to clarify certain aspects of the manuscript and to ensure that relevant information is presented accurately.

Title: It is arguably *"the most singular scorpion in Europe"*, considering that there are supposedly two species in the genus *Belisarius*, and the closely related species *Sardoscorpius troglophilus* might be even more "singular" due to its rarity. I would recommend avoiding this second fragment of the title.

"Bauplan"
Consider replacing this term with "body plan," which is more widely used and readily understood in the literature.

"Stiff bristles"
I am not certain that this represents a diagnostic character for this taxon. Please provide an appropriate reference if this feature is indeed considered diagnostic. Moreover, European euscorpiids also possess bristles on the chelicerae. Other characters may be more informative for defining this lineage of scorpions, such as the trichobothrial pattern, which is typically regarded as more diagnostic.

"it is considered the most cryptic in Europe"
This appears to be a rather strong statement that is supported only by an older reference (Vignoli et al., 2008). Some populations of its congener *B. ibericus* appear to be even more relictual or cryptic, as they have not been rediscovered since their original description to the best of my knowledge. A similar case could be made for *S. troglophilus*, which has never been recaptured and is known only from the original material. Other European examples include the troglobitic *Euscorpius studentium* from Montenegro, which is known from only two adult specimens and a few juveniles. The authors may wish to soften this statement or provide additional support.

"Moreover, its low fertility (5–24 eggs per clutch) may further compromise its long-term survival"
Please provide a reference supporting this statement.

"It was initially allied with the family Troglotayosicidae, which comprises a single genus of cave-dwelling species native to Ecuador and Colombia"
This statement requires clarification. The genus *Troglotayosicus* largely comprises humicolous species inhabiting leaf litter, and only one species (*Troglotayosicus vachoni*) is truly cave-dwelling.

"the new family Belisariidae has been created to include the blind scorpion along with a recently described cave-dwelling species from southern Iberian Peninsula (*B. ibericus* Lourenço, 2015)"
Please note that the family Belisariidae was originally described by Lourenço (1998). It was only recently resurrected to accommodate the species mentioned here. The wording should be adjusted to reflect this taxonomic history.

"At present, genomic data are available for only six scorpion species"
In a brief search, I found that a chromosome-level genome assembly is also available for *Olivierus przewalskii (= Mesobuthus przewalskii*; Su et al., 2025). The authors may wish to include this reference and, ideally, cite the other available scorpion genomes as well.

"Sex of the individuals collected is unknown"
It should be possible to determine the sex of the individuals used for the genome assembly by examining the number and morphology of the pectinal teeth, as well as the genital operculum. The literature on *Belisarius* should provide information on the typical counts for males and females. See for example:
Refer to reference 1&2

Individuals from different localities
According to the information available, *Belisarius xambeui* populations appear to represent a clear example of a genetic bottleneck. It may therefore be important to justify this in the methodology as to why individuals from different localities were used rather than specimens from the same population, as otherwise this could have potentially influenced the genomic analyses.

Figure 1
A more detailed map would be appropriate for the article. The current map lacks the sampling localities used in the study; they should be in a different color or shape. It may also be useful to include an elevation layer, which would help illustrate the altitudinal gradient across the distribution of the species. Additionally, if occurrence data were obtained from GBIF, the authors should ensure that records corresponding to museum coordinates rather than actual collection localities were filtered out.

References
The authors may also wish to cite several relevant works on the taxa in question, including:(Refer to reference 3, 4, & 5)

References
1. Borelli, A.: Descrizione del maschio del Belisarius xambeui E. Sim. (Arachnida, Scorpiones) de la Catalogna settentrionale. *Publ. Junta Cien. Nat. Barcelona 4: 1–6*. 1924.
2. Vachon, M.: Remarques sur le scorpion aveugle du Roussillon: Belisarius xambeui E. Simon. *Bull. Mus. Natl. Hist. Nat. 16: 298–305*. 1945.
3. Simon, E.: 3e Ordre – Scorpiones. In: Les Arachnides de France. VII. Contenant les Ordres des Chernetes, Scorpiones et Opiliones. *Roret, Paris, pp. 79–115*. 1879.
4. Lourenço, W. R.: The genus Belisarius Simon, 1879 (Scorpiones: Troglotayosicidae), with the description of a new vicariant species from southern Spain. *Comptes Rendus Biologies, 338: 362–367* . 2015.
5. Tropea, G., Onnis, C.: A remarkable discovery of a new scorpion genus and species from Sardinia (Scorpiones: Chactoidea: Belisariidae). *Arachnida – Rivista Aracnologica Italiana, 26: 3–25*. 2020.

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

Yes

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

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

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

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Ecology - Evolution - Phylogenetics - Systematics

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.

Author Response 20 Mar 2026

Julio Rozas

The article provides a reference chromosome-level genome for the troglomorphic scorpion Belisarius xambeui, endemic to France and Spain, and belonging to an incertae sedis lineage within the scorpion Tree of Life. This work has the potential to shed light on the evolution of this group and help clarify its phylogenetic position within the Scorpion Tree of Life. I believe this contribution is timely and appropriate, and I therefore recommend the article for indexing in Open Research Europe.

However, several minor edits and comments should be addressed prior to acceptance. These are important to clarify certain aspects of the manuscript and to ensure that relevant information is presented accurately.

Title: It is arguably “the most singular scorpion in Europe”, considering that there are supposedly two species in the genus Belisarius, and the closely related species Sardoscorpius troglophilus might be even more “singular” due to its rarity. I would recommend avoiding this second fragment of the title.
We agree, we decided to modify this second fragment of the title to: … a singular scorpion in Europe.
“Bauplan”
Consider replacing this term with “body plan,” which is more widely used and readily understood in the literature.
We replaced the term as suggested.

“Stiff bristles”
I am not certain that this represents a diagnostic character for this taxon. Please provide an appropriate reference if this feature is indeed considered diagnostic. Moreover, European euscorpiids also possess bristles on the chelicerae. Other characters may be more

informative for defining this lineage of scorpions, such as the trichobothrial pattern, which is typically regarded as more diagnostic.
We agree. Thus, we have removed this part in the new versión of the ms.

“it is considered the most cryptic in Europe”
This appears to be a rather strong statement that is supported only by an older reference (Vignoli et al., 2008). Some populations of its congener B. ibericus appear to be even more relictual or cryptic, as they have not been rediscovered since their original description to the best of my knowledge. A similar case could be made for S. troglophilus, which has never been recaptured and is known only from the original material. Other European examples include the troglobitic Euscorpius studentium from Montenegro, which is known from only two adult specimens and a few juveniles. The authors may wish to soften this statement or provide additional support.
We agree, we have modified this sentence in the new versión of the ms.

“Moreover, its low fertility (5–24 eggs per clutch) may further compromise its long-term survival”
Please provide a reference supporting this statement.
We added Auber 1959 as a reference to support this statement.

“It was initially allied with the family Troglotayosicidae, which comprises a single genus of cave-dwelling species native to Ecuador and Colombia”
This statement requires clarification. The genus Troglotayosicus largely comprises humicolous species inhabiting leaf litter, and only one species (Troglotayosicus vachoni) is truly cave-dwelling.
We thank the reviewer for the clarification. We added in the text that this genus also comprises humicolous species.

“the new family Belisariidae has been created to include the blind scorpion along with a recently described cave-dwelling species from southern Iberian Peninsula (B. ibericus Lourenço, 2015)”
Please note that the family Belisariidae was originally described by Lourenço (1998). It was only recently resurrected to accommodate the species mentioned here. The wording should be adjusted to reflect this taxonomic history.
We agree with the reviewer and corrected the sentence in the text as suggested.

“At present, genomic data are available for only six scorpion species”
In a brief search, I found that a chromosome-level genome assembly is also available for Olivierus przewalskii (= Mesobuthus przewalskii; Su et al., 2025). The authors may wish to include this reference and, ideally, cite the other available scorpion genomes as well.
Added
“Sex of the individuals collected is unknown”
It should be possible to determine the sex of the individuals used for the genome assembly by examining the number and morphology of the pectinal teeth, as well as the genital operculum. The literature on Belisarius should provide information on the typical counts for males and females. See for example:
Refer to reference 1&2

We thank the reviewer for the information and references. Unfortunately, a careful and detailed examination was not possible since the specimens needed to be flash frozen in order to ensure an optimal preservation for high molecular weight DNA extraction.

Individuals from different localities
According to the information available, Belisarius xambeui populations appear to represent a clear example of a genetic bottleneck. It may therefore be important to justify this in the methodology as to why individuals from different localities were used rather than specimens from the same population, as otherwise this could have potentially influenced the genomic analyses.
We agree with the reviewer. However, because specimens of *Belisarius xambeui* are rare, and in order to secure the highest quality of DNA, we used specimens in the best condition of preservation, which motivated the decision to use specimens from different localities.

Figure 1
A more detailed map would be appropriate for the article. The current map lacks the sampling localities used in the study; they should be in a different color or shape. It may also be useful to include an elevation layer, which would help illustrate the altitudinal gradient across the distribution of the species. Additionally, if occurrence data were obtained from GBIF, the authors should ensure that records corresponding to museum coordinates rather than actual collection localities were filtered out.
We have included a new versión of this figure in the new versión of the ms.

References
The authors may also wish to cite several relevant works on the taxa in question, including:(Refer to reference 3, 4, & 5)
Added

Competing Interests: No competing interests were disclosed.

Reviewer Report 04 March 2026

https://doi.org/10.21956/openreseurope.24731.r69968

?

Carlos Fernando Prada Quiroga
University of Tolima, Ibague, Colombia

The manuscript entitled *"ERGA-BGE-CBP chromosome-level genome assembly of the blind scorpion* Belisarius xambeui *Simon, 1879 (Belisariidae, Scorpiones), the most singular scorpion in Europe"*

presents the sequencing and chromosome-level assembly of the genome of this species. The study is generally well structured and adequately supported.

However, several methodological aspects require clarification and additional detail before the manuscript can be considered suitable for indexing.

1. Sample Collection, Processing, and Genome Size Estimation:

a) The section would benefit from a more detailed description of the taxonomic identification process. Please specify how the individuals were identified (e.g., diagnostic morphological characters, taxonomic keys, expert validation). Additionally, indicate whether voucher specimens were deposited in a recognized biological collection, including accession numbers if available. This information is essential to ensure traceability and reproducibility.

b) The authors state that “the sex of these individuals is unknown.” Although sexual dimorphism in this taxon is minimal, it is important to clarify that sex determination may still be feasible through detailed examination of the genital operculum. While there are no obvious external morphological differences between males and females, sex identification can be achieved through careful analysis of genital structures (see, for example, Lourenço, 2015). The authors should clarify whether such examination was attempted and, if not, justify its omission.

2. Sampling Localities for PacBio and Hi-C

The manuscript indicates that one individual from Cova de la Mosquera (Beuda, Girona) was used for PacBio sequencing, while a third individual from Cantonigròs (Santa Maria de Corcó) was used for Hi-C scaffolding. The rationale for selecting individuals from different localities should be explicitly justified.

Using specimens from distinct populations may introduce structural variation that could potentially affect genome scaffolding and assembly accuracy. Please clarify whether population-level genomic variation was considered and how potential structural differences were controlled for or evaluated.

3. Figure Quality

Figure 2C appears to be of moderate to low resolution. It is recommended that the authors replace this image with a higher-quality photograph in which diagnostic morphological features of the organism are more clearly visible.

4. DNA Extraction and Quantification

In the DNA extraction section, please provide detailed information on DNA quantification procedures, including:

The method used for quantification (e.g., fluorometric, spectrophotometric).

The equipment and kits employed.

Quality assessment metrics (e.g., A260/280, A260/230 ratios, fragment size evaluation).Including this information will improve methodological transparency and reproducibility.

5. Genome Annotation

More importantly, the authors should provide a clear justification for not performing gene annotation on the assembled genome. Even in draft assemblies, structural and functional annotation is generally considered a fundamental component of genome characterization. If annotation was not conducted due to assembly limitations (e.g., fragmentation, incomplete scaffolding) or other constraints, this should be explicitly stated and discussed. Otherwise, the absence of gene annotation significantly limits the biological interpretation and utility of the genomic resource

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

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

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

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

Competing Interests: No competing interests were disclosed.

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however I have significant reservations, as outlined above.

Author Response 18 Mar 2026

Julio Rozas

1a). Sample Collection, Processing, and Genome Size Estimation:

a) The section would benefit from a more detailed description of the taxonomic identification process. Please specify how the individuals were identified (e.g., diagnostic morphological characters, taxonomic keys, expert validation). Additionally, indicate whether voucher specimens were deposited in a recognized biological collection, including accession numbers if available. This information is essential to ensure traceability and reproducibility.

We agree. We have added a sentence deling with that in the new versión of the ms. 1b) The authors state that “the sex of these individuals is unknown.” Although sexual dimorphism in this taxon is minimal, it is important to clarify that sex determination may still be feasible through detailed examination of the genital operculum. While there are no obvious external

morphological differences between males and females, sex identification can be achieved through careful analysis of genital structures (see, for example, Lourenço, 2015). The authors should clarify whether such examination was attempted and, if not, justify its omission.
We have added a sentence deling with that in the new versión of the ms.
2. Sampling Localities for PacBio and Hi-C

The manuscript indicates that one individual from Cova de la Mosquera (Beuda, Girona) was used for PacBio sequencing, while a third individual from Cantonigròs (Santa Maria de Corcó) was used for Hi-C scaffolding. The rationale for selecting individuals from different localities should be explicitly justified.

Using specimens from distinct populations may introduce structural variation that could potentially affect genome scaffolding and assembly accuracy. Please clarify whether population-level genomic variation was considered and how potential structural differences were controlled for or evaluated.
We have justified the reasons of our samplins strategy in the new versión of the ms.
3. Figure Quality

Figure 2C appears to be of moderate to low resolution. It is recommended that the authors replace this image with a higher-quality photograph in which diagnostic morphological features of the organism are more clearly visible.
We have included a new versión of this figure in the new versión of the ms.
4. DNA Extraction and Quantification

In the DNA extraction section, please provide detailed information on DNA quantification procedures, including:

The method used for quantification (e.g., fluorometric, spectrophotometric).

The equipment and kits employed.

Quality assessment metrics (e.g., A260/280, A260/230 ratios, fragment size evaluation).Including this information will improve methodological transparency and reproducibility.
We have included this information in the new versión of the ms. We measured the DNA extraction quality uisng a NanoDrop spectrophotomer, yielding absorbance ratios of 260/280 = 1.74 and 260/230 = 2.26
5. Genome Annotation

More importantly, the authors should provide a clear justification for not performing gene annotation on the assembled genome. Even in draft assemblies, structural and functional annotation is generally considered a fundamental component of genome characterization. If annotation was not conducted due to assembly limitations (e.g., fragmentation, incomplete scaffolding) or other constraints, this should be explicitly stated and discussed. Otherwise, the absence of gene annotation significantly limits the biological interpretation and utility of the genomic resource We completly agree. However, we have not yet finished

this bioinformatic task. Once completed we will include the information in the Belisarius repository on GitHub (indicated in the ms).

Competing Interests: No competing interests were disclosed.

Reviewer Report 03 March 2026

https://doi.org/10.21956/openreseurope.24731.r69972

Carlos E. Santibáñez-López
Western Connecticut State University, Danbury, Connecticut, USA

This manuscript represents a significant contribution to the field of arachnid genomics. The authors have provided a robust framework for understanding scorpion genome architecture, and the methodology is both sound and transparent. The commitment to data sharing is commendable. I recommend this manuscript for indexing pending the following minor revisions to ensure the literature and divergence estimates are fully up to date.

The authors discuss the unique life history of *Belisarius xambeui*; however, the claim regarding the **low fertility** of this species requires a formal citation. Please include the appropriate primary literature to support this observation.

The manuscript notes that there are currently six published scorpion genomes. However, the authors should incorporate the recently published genome of ***Olivierus przewalskii*** (= *Mesobuthus przewalskii*; Su et al., 2025).

The divergence estimates provided for the split between **Buthida** and **Iurida** should be updated to reflect recent large-scale phylogenomic analyses. I recommend using the estimate of **291–320 Ma (Carboniferous–Permian)** as recovered by the comprehensive analysis in Santibáñez-López et al. (2022).

References

1. Phylogenomics of scorpions reveal contemporaneous diversification of scorpion mammalian predators and mammal-active sodium channel toxins. Systematic Biology, 71(6), 1281-1289.
2. A chromosome-level genome assembly for the desert scorpion Mesobuthus przewalskii from Asian drylands. Journal of Heredity,.

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

Yes

Are the protocols appropriate and is the work technically sound?

Yes

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

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

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Arachnid Evolution - Genomics - Venomics

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.

Author Response 20 Mar 2026

Julio Rozas

This manuscript represents a significant contribution to the field of arachnid genomics. The authors have provided a robust framework for understanding scorpion genome architecture, and the methodology is both sound and transparent. The commitment to data sharing is commendable. I recommend this manuscript for indexing pending the following minor revisions to ensure the literature and divergence estimates are fully up to date.

1. The authors discuss the unique life history of Belisarius xambeui; however, the claim regarding the low fertility of this species requires a formal citation. Please include the appropriate primary literature to support this observation.
We have added a sentence dealing with that in the new version of the ms.

2. The manuscript notes that there are currently six published scorpion genomes. However, the authors should incorporate the recently published genome of Olivierus przewalskii (= Mesobuthus przewalskii; Su et al., 2025).
We have modified this sentence in the new version of the ms.

3. The divergence estimates provided for the split between Buthida and Iurida should be updated to reflect recent large-scale phylogenomic analyses. I recommend using the estimate of 291–320 Ma (Carboniferous–Permian) as recovered by the comprehensive analysis in Santibáñez-López et al. (2022).
We agree. We have modified this sentence in the new version of the ms.

Competing Interests: No competing interests were disclosed.

Reviewer Report 27 February 2026

https://doi.org/10.21956/openreseurope.24731.r69975

Cheng-Min Shi
Hebei Agricultural University, Baoding, China

This data note reports a genome assembly for Belisarius *xambeui,* a scorpion species that is endemic to Catalan. *Belisarius xambeui* belongs to the species-poor family Belisariidae, which represents a unique lineage of scorpions as signified by troglomorphism. The assembly is an important contribution to the genome resources not only to scorpions but also to troglomorphic arthropods. The genome reported is the largest one of the sequenced genomes of scorpions to date. In general, the manuscript is well written. I only have the following remarks.
1. The last sentence of the Abstract should be better expressed as ''The average nucleotide diversity π across the genome is 0.0018 (range 0.0011~0.0028)'. Without a reference species or population, it is not prudent to say the diversity is low or not.
2. Background--- At present, genomic data are available for at least 7 scorpion species. See Su et al. 2024. Journal of Heredity 116(4): 532-539 for a chromosome-level genome assembly for *Mesobuthus przewalskii* (Buthidae).
3. Data availability ---accessions for mitochondrial genome and HiC data should be provided.

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

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

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

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

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Scorpion molecular ecology and evolutionary genomics

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.

Author Response 18 Mar 2026

Julio Rozas

1. The last sentence of the Abstract should be better expressed as ''The average nucleotide diversity π across the genome is 0.0018 (range 0.0011~0.0028)'. Without a reference species or population, it is not prudent to say the diversity is low or not. We agree. We have added a

sentence deling with that in the new versión of the ms.
2. Background--- At present, genomic data are available for at least 7 scorpion species. See Su et al. 2024. Journal of Heredity 116(4): 532-539 for a chromosome-level genome assembly for Mesobuthus przewalskii (Buthidae). Added
3. Data availability ---accessions for mitochondrial genome and HiC data should be provided. Absolutely. The accession numbers have already been included.

Competing Interests: No competing interests were disclosed.

Author Response 19 Mar 2026

Julio Rozas

1. The last sentence of the Abstract should be better expressed as ''The average nucleotide diversity π across the genome is 0.0018 (range 0.0011~0.0028)'. Without a reference species or population, it is not prudent to say the diversity is low or not. We agree. We have added a sentence deling with that in the new versión of the ms.
2. Background--- At present, genomic data are available for at least 7 scorpion species. See Su et al. 2024. Journal of Heredity 116(4): 532-539 for a chromosome-level genome assembly for Mesobuthus przewalskii (Buthidae). Added
3. Data availability ---accessions for mitochondrial genome and HiC data should be provided. Absolutely. The accession numbers have already been included.

Competing Interests: No competing interests were disclosed.