Check for updates

DATA NOTE

REVISED ERGA-BGE reference genome of the forest-restricted Cretan subspecies of Hanak's bat (*Pipistrellus hanaki creticus*), an IUCN Vulnerable species

[version 3; peer review: 2 approved]

Previously titled ; "ERGA-BGE reference genome of Hanak's bat (*Pipistrellus hanaki*), an IUCN Vulnerable species restricted to forest-like biotopes

Panagiotis Georgakakis[1], Danae Karakasi[1,2], Petros Lymberakis[1], Manolis Papadimitrakis[1], Manos Stratakis[1,2], Eleftherios Bitzilekis[1], Nikolaos Poulakakis[1-3], Astrid Böhne[4], Rita Monteiro[4], Rosa Fernández[5], Nuria Escudero[5], Genoscope Sequencing Team, Alice Moussy[6], Corinne Cruaud[6], Karine Labadie[6], Lola Demirdjian[7], Sophie Mangenot[7], Caroline Belser[7], Patrick Wincker[7], Pedro H. Oliveira[7], Jean-Marc Aury[7], Leanne Haggerty[8], Swati Sinha[8], Fergal Martin[8], Chiara Bortoluzzi[9]

[1]Natural History Museum of Crete, School of Sciences and Engineering, Knossos Avenue, University of Crete, Heraklion, GR-71409, Greece
[2]Department of Biology, School of Sciences and Engineering, Vassilika Vouton, University of Crete, Heraklion, GR-70013, Greece
[3]Foundation for Research and Technology – Hellas (FORTH), Institute of Molecular Biology and Biotechnology (IMBB), Heraklion, GR-70013, Greece
[4]Leibniz Institute for the Analysis of Biodiversity Change, Museum Koenig Bonn, Adenauerallee 127, Bonn, 53113, Germany
[5]Metazoa Phylogenomics Lab, Passeig marítim de la Barceloneta 37-49., Institute for Evolutionary Biology (CSIC-UPF), Barcelona, 08003, Spain
[6]Genoscope, Institut François Jacob, CEA, CNRS, Univ Evry, Université Paris-Saclay, Evry, 91057, France
[7]Génomique Métabolique, Genoscope, Institut François Jacob, CEA, CNRS, Univ Evry, Université Paris-Saclay, Evry, 91057, France
[8]European Molecular Biology Laboratory, European Bioinformatics Institute, Wellcome Genome Campus, Hinxton, Cambridge, CB10 1SD, UK
[9]SIB Swiss Institute of Bioinformatics, Amphipôle, Quartier UNIL-Sorge, Lausanne, 1015, Switzerland

v3 **First published:** 25 Sep 2025, **5**:298
https://doi.org/10.12688/openreseurope.20937.1
Second version: 11 Nov 2025, **5**:298
https://doi.org/10.12688/openreseurope.20937.2
Latest published: 11 Apr 2026, **5**:298
https://doi.org/10.12688/openreseurope.20937.3

Abstract

Hanak's bat (*Pipistrellus hanaki* Hulva and Benda 2004) is one of the most range restricted mammals in Europe, since it occurs only in Cyrenaica, Libya, and Crete (Greece). It is currently classified as 'Vulnerable' on the IUCN Red List, with its foraging habitat threatened

Open Peer Review

Approval Status ✓ ✓

	1	2
version 3 (revision) 11 Apr 2026		
version 2		✓

by a number of human activities. The reference genome of the Cretan subspecies (*Pipistrellus hanaki creticus*) will provide a crucial resource for uncovering the species phylogenetic history and will help assess the degree of genetic isolation among the species populations. A total of 23 contiguous chromosomal pseudomolecules (sex chromosomes included) were assembled from the genome sequence. This chromosome-level assembly encompasses 1.9 Gb, composed of 447 contigs and 141 scaffolds, with contig and scaffold N50 values of 48.7 Mb and 89.1 Mb, respectively.

Keywords
Pipistrellus hanaki, genome assembly, European Reference Genome Atlas, Biodiversity Genomics Europe, Earth Biogenome Project, Vespertilionidae family, Hanak's bat, Νανονυχτερίδα του Hanak

This article is included in the Horizon Europe gateway.

This article is included in the Genome Reports from the Biodiversity Genomics Europe Project collection.

	1	2
(revision) 11 Nov 2025		view
version 1 25 Sep 2025	✔ view	

1. **Richard Orton**, University of Glasgow, Glasgow, UK
2. **Jiang Feng**, Northeast Normal University, Changchun, China

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

Corresponding author: Chiara Bortoluzzi (chiara.bortoluzzi@sib.swiss)

Author roles: Georgakakis P: Investigation, Resources, Writing – Original Draft Preparation, Writing – Review & Editing; **Karakasi D**: Investigation, Methodology, Project Administration, Resources, Supervision, Writing – Review & Editing; **Lymberakis P**: Investigation, Resources, Writing – Review & Editing; **Papadimitrakis M**: Investigation, Resources, Writing – Review & Editing; **Stratakis M**: Investigation, Resources, Writing – Review & Editing; **Bitzilekis E**: Investigation, Resources, Writing – Review & Editing; **Poulakakis N**: Investigation, Resources, Writing – Review & Editing; **Böhne A**: Methodology, Project Administration, Supervision, Writing – Review & Editing; **Monteiro R**: Methodology, Project Administration, Supervision, Writing – Review & Editing; **Fernández R**: Methodology, Project Administration, Supervision, Writing – Review & Editing; **Escudero N**: Methodology, Project Administration, Supervision, Writing – Review & Editing; **Moussy A**: Investigation, Supervision, Writing – Review & Editing; **Cruaud C**: Investigation, Supervision, Writing – Review & Editing; **Labadie K**: Investigation, Supervision, Writing – Original Draft Preparation; **Demirdjian L**: Data Curation, Formal Analysis, Writing – Review & Editing; **Mangenot S**: Data Curation, Formal Analysis, Writing – Review & Editing; **Belser C**: Data Curation, Formal Analysis, Writing – Review & Editing; **Wincker P**: Investigation, Supervision, Writing – Review & Editing; **Oliveira PH**: Investigation, Supervision, Writing – Review & Editing; **Aury JM**: Data Curation, Formal Analysis, Supervision, Writing – Review & Editing; **Haggerty L**: Data Curation, Formal Analysis, Writing – Review & Editing; **Sinha S**: Data Curation, Formal Analysis, Writing – Review & Editing; **Martin F**: Data Curation, Formal Analysis, Writing – Review & Editing; **Bortoluzzi C**: Visualization, Writing – Original Draft Preparation

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 work was supported by the Genoscope, the Commissariat à l'Énergie Atomique et aux Énergies Alternatives (CEA), France Génomique (ANR-10-INBS-09-08), and the exploratory research programme 'ATLASea: Atlas of marine genomes and its targeted project SEQ-Sea (ANR-22-EXAT-0003-SEQ-Sea).
The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Copyright: © 2026 Georgakakis P *et al*. This is an open access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

How to cite this article: Georgakakis P, Karakasi D, Lymberakis P *et al.* **ERGA-BGE reference genome of the forest-restricted Cretan subspecies of Hanak's bat (*Pipistrellus hanaki creticus*), an IUCN Vulnerable species [version 3; peer review: 2 approved]** Open Research Europe 2026, **5**:298 https://doi.org/10.12688/openreseurope.20937.3

First published: 25 Sep 2025, **5**:298 https://doi.org/10.12688/openreseurope.20937.1

REVISED **Amendments from Version 2**

In this revised version of the Genome Report of Hanak's bat, we addressed all reviewers' comments as follow. We refocused the entire Genome Report on the subspecies *Pipistrellus hanaki creticus* in line both reviewers' comments. As a result, we decided to change the title of the Genome Report to reflect this change. We also added a new figure on the species itself, as recommended by both reviewers. Both major and minor suggestions were also addressed.

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

Introduction

Hanak's bat (*Pipistrellus hanaki* Hulva and Benda 2004) or Νανονυχτερίδα του Hanak in Greek, is one of the six west-Palaearctic members of the genus *Pipistrellus* Kaup, 1829 s.str (Figure 1). Phylogenetically, it is the sister species to *Pipistrellus pygmaeus* (Leach, 1825). Hanak's bat is a small bat, with a pale brown to rusty brown pelage coloration, moderately paler and distinctly rustier than in the other two congeneric species, *Pipistrellus pipistrellus* and *Pipistrellus pygmaeus*. The face, wing membranes, ears, and tragi are dark brown, whereas the ear bases and the area around the eyes are slightly paler. In Europe, only a small-sized form of Hanak's bat occurs in Crete, *P. hanaki creticus* Benda, 2009, while the nominotypical form, *P. hanaki hanaki* Hulva and Benda, 2004 lives just in northern Cyrenaica, Libya. These two forms differ in morphometric and genetic traits (Benda et al., 2004, 2008, 2014); while *P. hanaki hanaki* is the largest member of the common pipistrelle (*Pipistrellus pipistrellus*) group, *P. hanaki creticus* has a significant smaller body and skull size and its endemic Cretan population was even suggested to represent a separate species, since its phylogenetic position was not fully resolved, but it was shown to be a sister species to both *P. hanaki* s.str and *P. pygmaeus* (Benda *et al.*, 2014). This phylogenetic reconstruction, however, has not yet been accepted.

The Hanak's bat was first discovered in the north-eastern part of Cyrenaica, Libya (Benda *et al.*, 2004) and was subsequently found solely in the center of the forested northern part of the Cyrenaican plateau (Jebel Al Akhdar Mts). Its presence on Crete was first documented in 2007 (Hulva *et al.*, 2007) from a specimen collected in the western part of the island which was analysed morphologically and genetically. Since then, it has been found in several localities, mostly in forested areas, tree cultivations, and vegetated wetlands of western and central Crete (Benda *et al.*, 2008; Georgiakakis *et al.*, 2023). Although rather widespread on Crete, *Pipistrellus hanaki* was not located in several sampling efforts that were recently undertaken in many islands of south-east Greece, viz. Rhodes and Karpathos (Georgiakakis *et al.*, 2023).

Figure 1. A specimen of the subspecies *Pipistrellus hanaki creticus*. The individual in the picture is a different specimen from the one sampled and sequenced for this study. Photo credit: Panagiotis GEORGIAKAKIS, Natural History Museum of Crete, Crete.

Pipistrellus hanaki is currently classified as 'Vulnerable' on the IUCN Red List (Georgiakakis *et al.*, 2020), reflecting its high risk of extinction in the wild because of, among others, significant habitat loss due to residential and commercial development, forest fires and expansion of cultivations (Georgiakakis *et al.*, 2018). Additionally, it is listed under Annex IV of the Habitats Directive (92/43/EU) and Appendix II of the Bern Convention highlighting its protection and conservation status as a species of European interest and necessitating the proper management of its habitat.

Pipistrellus hanaki creticus utilizes a great variety of roosts (trees, rock fissures, buildings), but it depends on mature trees – native *Quercus* but also cultivated *Olea*, *Ceratonia*, *Ficus* and *Prunus* – for foraging (Georgiakakis *et al.*, 2018). *Pipistrellus h. hanaki* seems to prefer similar habitats in the Mediterranean forests and shrublands of Cyrenaica (Benda et al., 2014). As an insect predator, it plays an important role in the forest ecosystem function and resilience.

Developing a high-quality reference genome for *Pipistrellus hanaki* is essential to advance our understanding of its unique genetic makeup. This genomic resource will also support conservation efforts by providing valuable insights into its phylogenetic relationships within the *P. pipistrellus* group and the genetic differences between *P. hanaki hanaki* and *P. hanaki creticus.* Given that the phylogenetic relationships of the two subspecies are not resolved yet (Benda et al., 2014), sampling and sequencing of the nominotypical subspecies *P. h. hanaki* is needed, to enable inter-subspecies comparisons.

The generation of this reference resource was coordinated by the European Reference Genome Atlas (ERGA) initiative's Biodiversity Genomics Europe (BGE) project, supporting ERGA's aims of promoting transnational cooperation to promote advances in the application of genomics technologies to protect and restore biodiversity (Mazzoni *et al.*, 2023).

Materials & methods

ERGA's sequencing strategy includes Oxford Nanopore Technology (ONT) and/or Pacific Biosciences (PacBio) for long-read sequencing, along with Hi-C sequencing for chromosomal architecture, Illumina Paired-End (PE) for polishing (i.e. recommended for ONT-only assemblies), and RNA sequencing for transcriptome profiling, to facilitate genome assembly and annotation.

Sample and sampling information

On 21 October 2023, one adult male of *Pipistrellus hanaki creticus* was sampled by Panagiotis Georgakakis from the Natural History Museum of Crete (NHMC) of the University of Crete. The species was identified using the identification key of Dietz & Kiefer (2016). The specimen was collected with a mist net in Psiloritis mt., Rouvas forest, Irakleio, Crete, Greece. Sampling was performed under Presidential Decree 67/1981 issued by the Greek Government. The specimen was euthanized by increasing concentration of CO_2, after which it was immediately flash-frozen in liquid nitrogen, and preserved at −80 °C until DNA extraction.

Vouchering information

Physical reference material for the here sequenced specimen has been deposited in the Vertebrates Collections of the NHMC https://www.nhmc.uoc.gr/en/departments/vertebrates under accession ID NHMC.80.5.121.52.

Frozen reference tissue material of muscle and liver is available from the same individual at the Genomics and Genetic Resources Division of the NHMC https://www.nhmc.uoc.gr/en/departments/genomics under accession ID NHMC.80.5.121.52.

Genetic information

The estimated genome size, estimated by Genomes on a Tree (GoaT) (Challis *et al.*, 2023) by ancestral state reconstruction, is 2.12 Gb. This is a diploid genome with a haploid number of 22 chromosomes (2n = 44). All information for this species was retrieved from GoaT.

DNA/RNA processing

DNA was extracted from muscle (45 mg) using a Genomic-tip 100/G Kit (QIAGEN, MD, USA) following manufacturer instructions. DNA fragment size selection was performed using Short Read Eliminator (PacBio, CA, USA). Quantification was performed using a Qubit dsDNA HS Assay kit (Thermo Fisher Scientific) and integrity was assessed in a FemtoPulse system (Agilent). DNA was stored at 4 °C until usage.

RNA was extracted from 10 mg of muscle using the RNeasy Plus Universal kit (Qiagen) following manufacturer instructions. Residual genomic DNA was removed with 6 U of TURBO DNase (2 U/μL) (Thermo Fisher Scientific).

Quantification was performed using a Qubit RNA HS Assay kit and integrity was assessed in a Bioanalyzer system (Agilent). RNA was stored at −80 °C.

Library preparation and sequencing

Long-read DNA libraries were prepared with the SMRTbell prep kit 3.0 following manufacturers' instructions and sequenced on a Revio system (PacBio). Hi-C libraries were generated from muscle (20 mg) of the same individual using the Arima High Coverage HiC kit (following the Animal Tissues low input protocol v01) and sequenced on a NovaSeq 6000 instrument (Illumina) with 2x150 bp read length. Poly(A) RNA-Seq libraries were constructed using the Illumina Stranded mRNA Prep, Ligation Prep kit (Illumina) and sequenced on an Illumina NovaSeq X Plus instrument (Illumina) with 2x150 bp read length. In total, 36x PacBio and 19x HiC data were sequenced to generate the assembly.

Genome assembly methods

The genome of *Pipistrellus hanaki creticus* was assembled using the Genoscope GALOP pipeline (https://workflowhub.eu/workflows/1200). Briefly, raw PacBio HiFi reads were assembled using Hifiasm v0.19.5-r593 (Cheng *et al.*, 2021). Remaining allelic duplications were removed using purge_dups v1.2.5 (Guan *et al.*, 2020) with default parameters and the proposed cutoffs. The purged assembly was scaffolded using YaHS v1.2 (Zhou *et al.*, 2023) and assembled scaffolds were then curated through manual inspection using PretextView v0.2.5 to remove false joins and incorporate sequences not automatically scaffolded into their respective locations within the chromosomal pseudomolecules. Summary analysis of the released assembly was performed using the ERGA-BGE Genome Report ASM Galaxy workflow (10.48546/workflowhub.workflow.1104.1).

Genome annotation methods

A gene set was generated using the Ensembl Gene Annotation system (Aken *et al.*, 2016), primarily by aligning short-read RNA-seq data from BioSample SAMEA115120717 to a previous version of the reference genome (GCA_964339955.1). Gaps in the annotation were filled via protein-to-genome alignments of a select set of vertebrate proteins from UniProt (UniProt Consortium, 2019), which had experimental evidence at the protein or transcript level. At each locus, data were aggregated and consolidated, prioritising models derived from RNA-seq data, resulting in a final set of gene models and associated non-redundant transcript sets. To distinguish true isoforms from fragments, the likelihood of each open reading frame (ORF) was evaluated against known vertebrate proteins. Low-quality transcript models, such as those showing evidence of fragmented ORFs, were removed. In cases where RNA-seq data were fragmented or absent, homology data were prioritised, favouring longer transcripts with strong intron support from short-read data. The resulting gene models were classified into three categories: protein-coding, pseudogene, and long non-coding. Models with hits to known proteins and few structural abnormalities were classified as protein-coding. Models with hits to known proteins but displaying abnormalities, such as the absence of a start codon, non-canonical splicing, unusually small intron structures (<75 bp), or excessive repeat coverage, were reclassified as pseudogenes. Single-exon models with a corresponding multi-exon copy elsewhere in the genome were classified as processed (retrotransposed) pseudogenes. Models that did not fit any of the previously described categories did not overlap protein-coding genes and were constructed from transcriptomic data were considered potential lncRNAs. Potential lncRNAs were further filtered to remove single-exon loci due to their unreliability. Putative miRNAs were predicted by performing a BLAST search of miRBase (Kozomara *et al.*, 2019) against the genome, followed by RNAfold analysis (Gruber *et al.*, 2008). Other small non-coding loci were identified by scanning the genome with Rfam (Kalvari *et al.*, 2018) and passing the results through Infernal (Nawrocki & Eddy, 2013).

Results

Genome assembly

The genome assembly has a total length of 1,892,181,625 bp in 141 scaffolds (Figure 2 and Figure 3), with a GC content of 42.5%. It features a contig N50 of 48,700,701 bp (L50 = 16) and a scaffold N50 of 89,115,682 bp (L50 = 7). There are 306 gaps, totalling 34.1 kb in cumulative size. The single-copy gene content analysis using the Mammalia database with BUSCO (Manni *et al.*, 2021) resulted in 93.3% completeness (92.0% single and 1.3% duplicated). 96.3% of reads k-mers were present in the assembly and the assembly has a base accuracy Quality Value (QV) of 61.4 as calculated by Merqury (Rhie *et al.*, 2020).

Genome annotation

The genome annotation consists of 18,099 protein-coding genes with associated 30,313 transcripts, in addition to 5,633 non-coding genes (Table 1). Using the longest isoform per transcript, the single-copy gene content analysis using the Mammalia odb10 database with BUSCO resulted in 95.3% completeness. Using the OMAmer Metazoa-v2.0.0.h5 database for OMArk (Nevers *et al.*, 2025) resulted in 95.7% completeness and 98.5% consistency (Table 2).

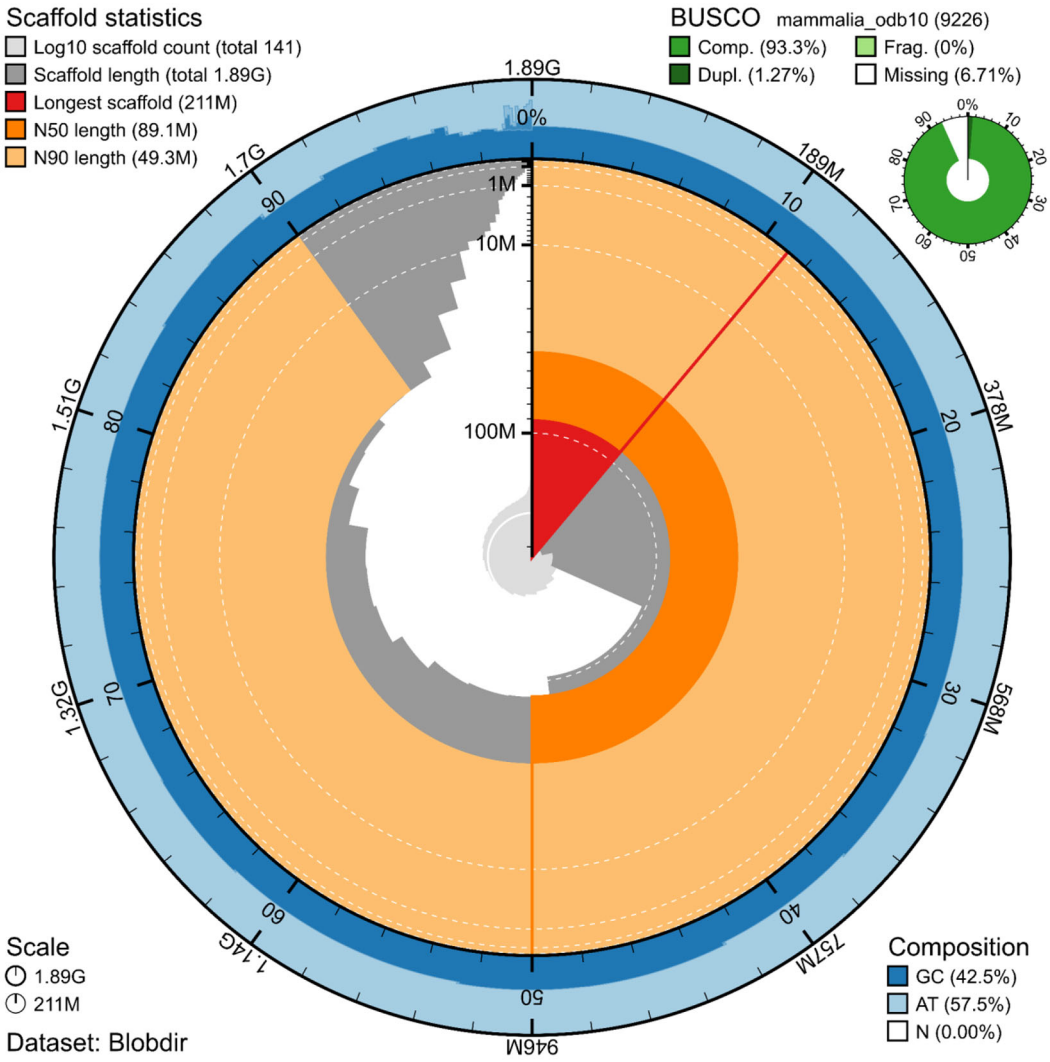

Figure 2. Snail plot summary of assembly statistics. The main plot is divided into 1,000 size-ordered bins around the circumference, with each bin representing 0.1% of the 1,892,181,625 bp assembly. The distribution of sequence lengths is shown in dark grey, with the plot radius scaled to the longest sequence present in the assembly (211 Mb, shown in red). Orange and pale-orange arcs show the scaffold N50 and N90 sequence lengths (89,115,682 and 49,289,172 bp), respectively. The pale grey spiral shows the cumulative sequence count on a log-scale, with white scale lines showing successive orders of magnitude. The blue and pale-blue area around the outside of the plot shows the distribution of GC, AT, and N percentages in the same bins as the inner plot. A summary of complete, fragmented, duplicated, and missing BUSCO genes found in the assembled genome from the Mammalia database (odb10) is shown on the top right.

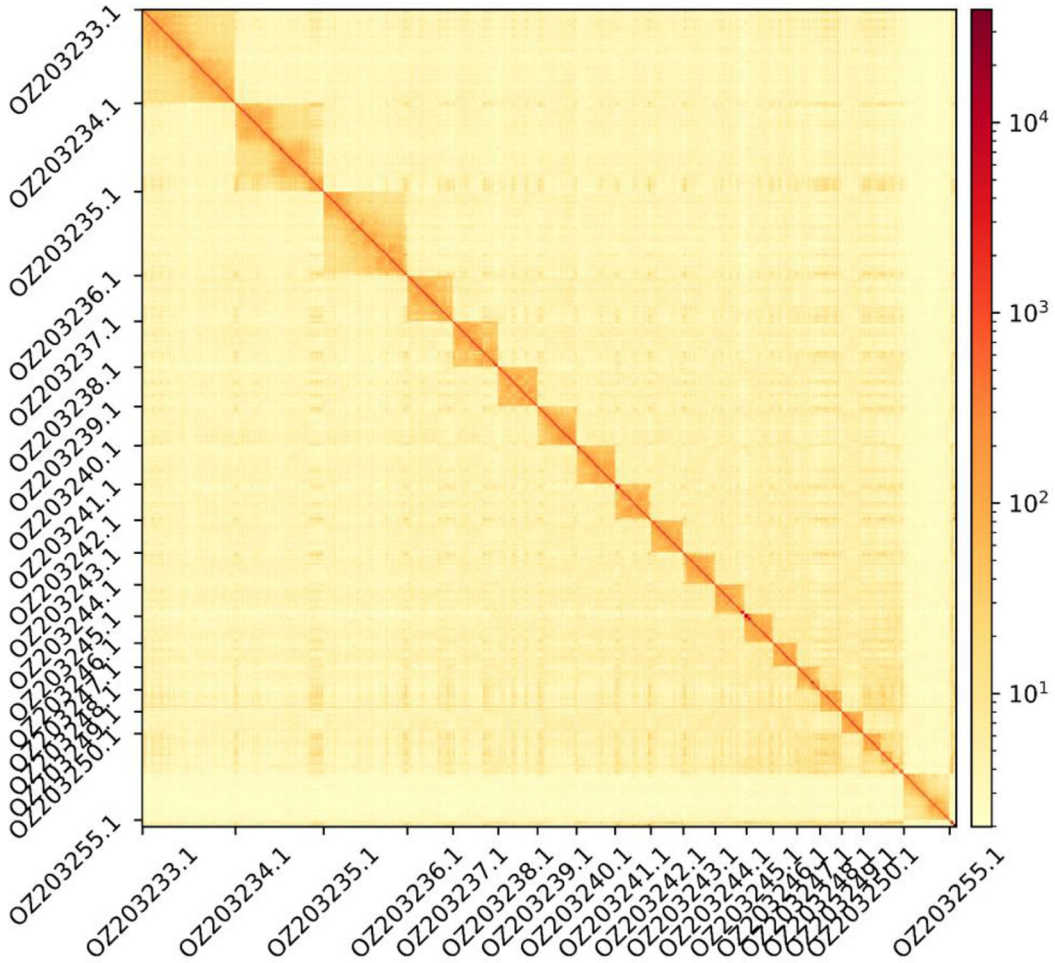

Figure 3. Hi-C contact map showing spatial interactions between regions of the genome. The diagonal corresponds to intra-chromosomal contacts, depicting chromosome boundaries. The frequency of contacts is shown on a logarithmic heatmap scale. Hi-C matrix bins were merged into a 50 kb bin size for plotting. Due to space constraints on the axes, only the GenBank names of the 18th largest autosomes and the Y chromosome (GenBank: OZ203255.1) are shown.

Table 1. Statistics from assembled gene models.

	No. genes	No. transcripts	Mean gene length (bp)	No. single-exon genes	Mean exons per transcript
mRNA	18,099	30,313	41,209	1,346	11.4
pseudogene	386	386	14,593	33	13.3
snoRNA	972	972	123	972	1.0
lncRNA	3,043	3,204	3,256	2,630	1.4
miRNA	135	135	82	135	1.0
snRNA	1,263	1,263	119	1,263	1.0
rRNA	110	110	225	110	1.0
scRNA	77	77	148	77	1.0
Other ncRNA	33	33	96–281	33	1.0

Table 2. Annotation completeness and consistency scores calculated by BUSCO run in protein mode (mammalia_odb10) and OMArk (Metazoa-v2.0.0.h5).

	Complete	Single copy	Duplicated	Fragmented	Missing
BUSCO	8,790 (95.3%)	8,706 (94.4%)	84 (0.9%)	74 (0.8%)	362 (3.9%)
OMArk	13,499 (95.7%)	13,207 (93.7%)	292 (2.0%)	-	596 (4.2%)
	Consistent	**Inconsistent**	**Contaminants**	**Unknown**	
OMArk	17,942 (98.5%)	197 (1.1%)	0.0 (0.0%)	81 (0.4%)	

Data availability

Pipistrellus hanaki creticus and the related genomic study were assigned to Tree of Life ID (ToLID) 'mPipHan1' and all sample, sequence, and assembly information are available under the umbrella BioProject PRJEB77247. The sample information is available at the following BioSample accessions: SAMEA115799862, SAMEA115799867, and SAMEA115120717. The genome assembly is accessible from ENA under accession number GCA_964339955.4 and the annotated genome is available at the Ensembl website (https://projects.ensembl.org/erga-bge/). Sequencing data produced as part of this project are available from ENA at the following accessions: ERX12737184, ERX12737185, ERX14169058, ERX14169059, and ERX12733454. Documentation related to the genome assembly and curation can be found in the ERGA Assembly Report (EAR) document available at https://github.com/ERGA-consortium/EARs/tree/main/Assembly_Reports/Pipistrellus_hanaki/mPipHan1. Further details and data about the project are hosted on the ERGA portal at https://portal.erga-biodiversity.eu/data_portal/412090.

Author contributions

DK coordinated the project; PG collected the species; PG identified the species; DK, MP, MS, EB, PL, and NP sampled and preserved biological material and provided metadata; AsB, RM, RF, and NE provided support in sampling, shipping of biological material, metadata collection, and management; GST extracted DNA, prepared libraries, and performed sequencing under the supervision of AM, CC, KL, PHO, and PW; SM, CB, LD, and JMA performed genome assembly and curation under the supervision of JMA; LH, SS, and FM performed genome annotation; CB generated the analysis and report. All authors contributed to the writing, review, and editing of this genome note and read and approved the final version. This work is part of the species assigned to Genoscope, which was instrumental in the wet lab, sequencing, and assembly processes, and represents a key contribution to BGE's outputs.

Author information

Members of the Genoscope Sequencing Team are listed here: https://zenodo.org/records/14611490.

Acknowledgements

We would like to express our gratitude to M. Kantzaridou and I. Ekklisiarchos for assistance in fieldwork. We acknowledge the support of the Freiburg Galaxy Team: Saim Momin and Björn Grüning, Bioinformatics, University of Freiburg (Germany), funded by the German Federal Ministry of Education and Research BMBF grant 031 A538A de. NBI-RBC and the Ministry of Science, Research and the Arts Baden-Württemberg (MWK) within the framework of LIBIS/de. NBI Freiburg. We would like to acknowledge the assembly reviewer, Tyler S. Alioto from the Centro Nacional de Análisis Genómico (CNAG).

References

Aken BL, Ayling S, Barrell D, *et al.*: **The Ensembl gene annotation system.** *Database (Oxford).* 2016; **2016**: baw093.
PubMed Abstract | Publisher Full Text | Free Full Text

Benda P, Georgiakakis P, Dietz C, *et al.*: **Bats (Mammalia: Chiroptera) of the Eastern Mediterranean and Middle East. Part 7. The bat fauna of Crete, Greece.** *Acta Soc Zool Bohem.* 2008; **72**: 105–190.
Reference Source .

Benda P, Hulva P, Gaisler J: **Systematic status of African populations of *Pipistrellus pipistrellus* complex (Chiroptera: Vespertilionidae), with a description of a new species from Cyrenaica, Libya.** *Acta Chiropterol.* 2004; **6**(2): 193–217.
Publisher Full Text

Benda P, Spitzenberger F, Hanák V, *et al.*: **Bats (Mammalia: Chiroptera) of the Eastern Mediterranean and Middle East. Part 11. On the bat fauna of Libya II.** *Acta Soc Zool Bohem.* 2014; **78**: 1–162.
Reference Source .

Challis R, Kumar S, Sotero-Caio C, *et al.*: **Genomes on a Tree (GoaT): a versatile, scalable search engine for genomic and sequencing project metadata across the eukaryotic Tree of Life [version 1; peer review: 2 approved].** *Wellcome Open Res.* 2023; **8**: 24.
Publisher Full Text | Free Full Text

Cheng H, Concepcion GT, Feng X, *et al.*: **Haplotype-resolved *de novo* assembly using phased assembly graphs with hifiasm.** *Nat. Methods.* 2021; **18**(2): 170–175.
PubMed Abstract | **Publisher Full Text**

Dietz C, Kiefer A: *Bats of Britain and Europe.*: Bloomsbury publishing; 2016.
Reference Source

Georgiakakis P, Cooper-Bohannon R, Aulagnier S, *et al.*: ***Pipistrellus hanaki.*** *IUCN Red List Threatened Species.* 2020; **2020**: e.T136209A22011859.
Publisher Full Text

Georgiakakis P, Kafkaletou Diez A, Salvarina I, *et al.*: **The bats of Greece: an updated review of their distribution, ecology and conservation.** *Animals (Basel).* 2023; **13**(15): 2529.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

Georgiakakis P, Poursanidis D, Kantzaridou M, *et al.*: **The importance of forest conservation for the survival of the range-restricted *Pipistrellus hanaki,* an endemic bat from Crete and Cyrenaica.** *Mamm. Biol.* 2018; **93**: 109–117.
Publisher Full Text

Gruber AR, Lorenz R, Bernhart SH, *et al.*: **The Vienna RNA websuite.** *Nucleic Acids Res.* 2008; **36**: W70–W74.
Publisher Full Text

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

Hulva P, Benda P, Hanak V, *et al.*: **New mitochondrial lineages within the *Pipistrellus pipistrellus* complex from Mediterranean Europe.** *Folia Zool.* 2007; **56**(4): 378.
Reference Source .

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**

Kalvari I, Nawrocki EP, Argasinska J, *et al.*: **Non-coding RNA analysis using the Rfam database.** *Curr. Protoc. Bioinformatics.* 2018; **62**(1): e51.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

Kozomara A, Birgaoanu M, Griffiths-Jones S: **miRBase: from microRNA sequences to function.** *Nucleic Acids Res.* 2019; **47**(D1): D155–D162.
PubMed Abstract | **Publisher Full Text**

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

Mazzoni CJ, Ciofi C, Waterhouse RM: **Biodiversity: an atlas of European reference genomes.** *Nature.* 2023; **619**: 252.
Publisher Full Text

Nawrocki EP, Eddy SR: **Infernal 1.1: 100-fold faster RNA homology searches.** *Bioinformatics.* 2013; **29**(22): 2933–2935.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

Nevers Y, Warwick Vesztrocy A, Rossier V, *et al.*: **Quality assessment of gene repertoire annotations with OMArk.** *Nat. Biotechnol.* 2025; **43**(1): 124–133.
PubMed Abstract | **Publisher Full Text**

Rhie A, Walenz BP, Koren S, *et al.*: **Merqury: reference-free quality, completeness, and phasing assessment for genome assemblies.** *Genome Biol.* 2020; **21**(1): 245.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

UniProt Consortium: **UniProt: a worldwide hub of protein knowledge.** *Nucleic Acids Res.* 2019; **47**(D1): D506–D515.
PubMed Abstract | **Publisher Full Text**

Open Peer Review

Current Peer Review Status: ✔ ✔

Version 2

Reviewer Report 19 November 2025

https://doi.org/10.21956/openreseurope.23670.r63933

© **2025 Feng J.** This is an open access peer review report distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

✔ **Jiang Feng**
[1] Northeast Normal University, Changchun, China
[2] Northeast Normal University, Changchun, China

This manuscript presents a high-quality chromosome-level reference genome assembly and annotation for *Pipistrellus hanaki*—a IUCN Vulnerable species with a narrow distribution. The study follows standardized genomic workflows, including PacBio HiFi sequencing and Hi-C scaffolding, and provides comprehensive assembly statistics and annotation details, which is valuable for enrich genomic resource of this species and supporting its conservation. There are two issues regarding species taxonomic clarity and methodological rational need to be addressed to enhance the manuscript's scientific rigor and practical utility.

1.A critical inconsistency exists between the manuscript's title and the actual sequenced subspecies. The title refers to Hanak's bat (*Pipistrellus hanaki*), but the Introduction explicitly states that the sequenced individual is from Crete, belonging to the subspecies *P. hanaki creticus* (not the nominotypical *P. hanaki hanaki* from Libya). Moreover, the manuscript fails to address whether the genome of *P. hanaki creticus* can fully represent the entire species, especially given previous debates about the Cretan population potentially being a distinct species (Benda et al., 2014). This ambiguity undermines the generalizability of the "reference genome" and confuses readers about its taxonomic scope. Therefore, the authors are suggested to revise the title to explicitly reflect the subspecies, e.g., "ERGA-BGE reference genome of the Cretan subspecies of Hanak's bat (*Pipistrellus hanaki creticus*), an IUCN Vulnerable species restricted to forest-like biotopes". And add a dedicated subsection in the Discussion to address: 1) genetic differences between *P. hanaki creticus* and *P. hanaki hanaki*; 2) limitations of using a single subspecies to represent the entire species, and recommendations for future genome sequencing of the Libyan subspecies to enable inter-subspecies comparisons.

2. The Methods section states that gene annotation relied on "publicly available short-read RNA-seq data from BioSample SAMEA115120717" aligned to a previous version of the genome (GCA_964339955.1), rather than RNA-seq data generated from the newly assembled genome in this study. No explanation is provided for this choice, why public data was prioritized. This raises concerns about annotation accuracy, as public data may not perfectly match the new assembly or

the specific Cretan subspecies. If own RNA-seq data was generated but unused, explain technical issues (low quality? insufficient depth?) and validate the compatibility of the public data with the new assembly (e.g., sequence identity, alignment rate). If no new RNA-seq data was generated, acknowledge this limitation and discuss potential biases (such as public data from a different individual/population affecting gene model accuracy).

3. Other minor issues

The phrase "collected with a mist nest" may contain a spelling error. "Mist net" is the standard term for the tool used in bat sampling (to capture flying individuals).
In Introduction, multiple citations are overly repetitive. For example, "(Benda et al., 2004; Benda et al., 2008; Benda et al., 2014)") should be "(Benda et al., 2004, 2008, 2014)".

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: Animal molecular ecology

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

Author Response 18 Mar 2026

Chiara Bortoluzzi

We thank the reviewer for their feedback. We address all of their comments as below:

1. As pointed out by the reviewer, the manuscript now focuses on the Cretan subspecies of Hanak's bat (*Pipistrellus hanaki creticus*). The manuscript has thus been changed accordingly throughout.
2. As already addressed for reviewer 2, we rephrased the RNA-Seq data statement to prevent further misunderstanding. Briefly, short-read RNA-Seq data produced for the assembled species were used for the genome annotation. These RNA-Seq data are publicly available.
3. We changed the sentence on the collection method.
4. We corrected the citations in the introduction.

Competing Interests: No competing interests were disclosed.

Version 1

Reviewer Report 03 November 2025

https://doi.org/10.21956/openreseurope.22651.r61360

© 2025 Orton R. This is an open access peer review report distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Richard Orton
[1] University of Glasgow, Glasgow, UK
[2] University of Glasgow, Glasgow, UK
[3] University of Glasgow, Glasgow, UK

Overall – I found this to be a very well written report. I have only a couple of minor suggestions:

Introduction – suggestion - a picture of Hanak's bat might be useful here to go with the description

"In Europe, only a small-sized form of Hanak's bat occurs in Crete, *P. hanaki creticus* Benda, 2009, while the nominotypical form, *P. hanaki hanaki* Hulva and Benda, 2004 lives just in northern Cyrenaica, Libya. These two forms differ in morphometric and genetic traits (Benda *et al.*, 2004; Benda *et al.*, 2008; Benda *et al.*, 2014);" - rephrase – presumably the Hulva and Benda inclusions mid sentence are incorrectly formatted refs?

Overall – is there an expectation that the Crete and Libya populations of Hanak's bat are the same species and therefore the same genome? The title of the manuscript is Hanak's bat (*Pipistrellus hanaki*) – but according to the sentence highlighted in the previous point you are sequencing *P. hanaki criticus.* Maybe a sentence somewhere in discussion to address?
"A gene set was generated using the Ensembl Gene Annotation system (Aken *et al.*, 2016), primarily by aligning publicly available short-read RNA-seq data from BioSample SAMEA115120717 to a previous version of the reference genome (GCA_964339955.1)." Clarification - unsure what is happening here – you seem to be using public data with an existing reference genome – why are you not using your own one that was created?

What quality control has been done on the scaffolds/contigs? i.e. often viral transcripts (both true retroviral transcripts which are in then genome, but also erroneous viruses that infected the sequenced host) can be found in provisional genome sequences – have these been checked for (just a suggestion).

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: Bioinformatics, Metagenomics, Sequencing

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 07 Nov 2025

Chiara Bortoluzzi

As suggested by the reviewer, we now present in the Introduction a picture of an Hanak's bat individual. We did not change the formatting of the references when referring to *P. hanaki creticus* and *P. hanaki hanaki,* as we follow here standard international nomenclature rules. We rephrased the section on the gene annotation to avoid any further misunderstanding about the RNA-seq data used in the gene annotation step. We also would like to refer the reviewer to the EAR report (we refer to this in the main text of this Genome Report), as this report provides a comprehensive overview of all contamination analyses performed during the assembly step.

Competing Interests: No competing interests were disclosed.