DATA NOTE

The genome sequence of Nathusius' pipistrelle, *Pipistrellus nathusii* (Keyserling & Blasius, 1839) (Chiroptera: Vespertilionidae)

[version 1; peer review: 2 approved]

Manuel Ruedi[1], Meike Mai[2], Emma C. Teeling[3,4], Sonja C. Vernes[2],
Bat1K Consortium,
Wellcome Sanger Institute Tree of Life Management, Samples and Laboratory team,
Wellcome Sanger Institute Scientific Operations: Sequencing Operations,
Wellcome Sanger Institute Tree of Life Core Informatics team,
Tree of Life Core Informatics collective, Darwin Tree of Life Consortium

[1]Muséum d'histoire naturelle de Genève, Geneva, Switzerland
[2]School of Biology, University of St Andrews, St Andrews, Scotland, UK
[3]School of Biology and Environmental Science, University College Dublin, Dublin, Ireland
[4]Tree of Life Programme, Wellcome Sanger Institute, Hinxton, England, UK

v1 **First published:** 22 Apr 2026, **11**:234
https://doi.org/10.12688/wellcomeopenres.26312.1
Latest published: 22 Apr 2026, **11**:234
https://doi.org/10.12688/wellcomeopenres.26312.1

Abstract

We present a genome assembly from an individual female *Pipistrellus nathusii* (Nathusius' pipistrelle; Chordata; Mammalia; Chiroptera; Vespertilionidae). The assembly contains two haplotypes with total lengths of 1 961.71 megabases and 1 889.00 megabases. Most of haplotype 1 (92.66%) is scaffolded into 22 chromosomal pseudomolecules, including the X sex chromosome. Haplotype 2 was assembled to scaffold level. The mitochondrial genome has also been assembled, with a length of 17.27 kilobases. This assembly was generated as part of the Darwin Tree of Life project, which produces reference genomes for eukaryotic species found in Britain and Ireland.

Keywords

Pipistrellus nathusii, Nathusius' pipistrelle, genome sequence, chromosomal, Chiroptera

Open Peer Review

Approval Status ✓ ✓

	1	2
version 1 22 Apr 2026	✓ view	✓ view

1. **Luciano Chaves Franco Filho**, Instituto Evandro Chagas, Ananindeua, Brazil

2. **Xiuguang Mao**, East China Normal University, Shanghai, China

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

This article is included in the Tree of Life gateway.

This article is included in the Bat1K genomes collection.

Corresponding author: Darwin Tree of Life Consortium (mark.blaxter@sanger.ac.uk)

Author roles: Ruedi M: Investigation, Resources, Writing – Original Draft Preparation, Writing – Review & Editing; **Mai M**: Writing – Review & Editing; **Teeling EC**: Resources, Writing – Review & Editing; **Vernes SC**: Resources, Writing – Review & Editing;

Competing interests: No competing interests were disclosed.

Grant information: This work was supported by Wellcome through core funding to the Wellcome Sanger Institute (220540) and the Darwin Tree of Life Discretionary Award [218328, https://doi.org/10.35802/218328]. SCV was supported by a UKRI Future Leaders Fellowship (MR/T021985/1) and an ERC Consolidator Grant (101001702; BATSPEAK). ECT is supported by an Irish Research Council Laureate Award (IRCLA/2017/58) and Science Foundation Ireland Future Frontiers (19/FFP/6790).
The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

How to cite this article: Ruedi M, Mai M, Teeling EC *et al.* **The genome sequence of Nathusius' pipistrelle, *Pipistrellus nathusii* (Keyserling & Blasius, 1839) (Chiroptera: Vespertilionidae) [version 1; peer review: 2 approved]** Wellcome Open Research 2026, **11**:234 https://doi.org/10.12688/wellcomeopenres.26312.1

First published: 22 Apr 2026, **11**:234 https://doi.org/10.12688/wellcomeopenres.26312.1

Species taxonomy

Eukaryota; Opisthokonta; Metazoa; Eumetazoa; Bilateria; Deuterostomia; Chordata; Craniata; Vertebrata; Gnathostomata; Teleostomi; Euteleostomi; Sarcopterygii; Dipnotetrapodomorpha; Tetrapoda; Amniota; Mammalia; Theria; Eutheria; Boreoeutheria; Laurasiatheria; Chiroptera; Yangochiroptera; Vespertilionidae; *Pipistrellus*; *Pipistrellus nathusii* (Keyserling & Blasius, 1839) (NCBI:txid59473).

Background

Nathusius' pipistrelle, *Pipistrellus nathusii*, (Figure 1) is a small, uniform brown bat (adult weight 6–10 g) distinguished from other pipistrelles by longer, narrower wings (fifth digit >43 mm), hairy uropatagium and unique dentition (Russ, 2023). It feeds on insects in a variety of forested habitats, including in urbanised areas. This common bat is found throughout most of Europe and western Asia and is classified as Least Concern (LC) under IUCN criteria. It is one of few temperate bats making long-distance migrations between north-eastern Europe where females build nursery colonies, and central or southern Europe where they overwinter. Unlike in other European bats, females often give birth to twins.

Phylogenetic reconstructions based on mitochondrial genes have classified Nathusius' pipistrelle as a close relative to noctules (genus *Nyctalus*) (Roehrs *et al.*, 2010). However, more recent data based on few nuclear genes rather place it basal to other European pipistrelles, from which it diverged some 11 MYA (Dool & Puechmaille, 2024).

Another chromosome-level assembly for this species is also available (GCA_963693515.1, GCA_963693525.1; submitted by the Leibniz Institute for Zoo and Wildlife Research). This sampling is part of the Bat1K Project (Teeling *et al.*, 2018) and the Darwin Tree of Life Project (DToL) (Darwin Tree of Life Project Consortium, 2022). The Bat1K is a collaborative effort to sequence all extant bat species, and DToL aims to sequence all named eukaryotic species in the Atlantic Archipelago of Britain and Ireland.

Methods

Sample acquisition

The genome is based on an adult female *Pipistrellus nathusii* (specimen ID SAN00003686, ToLID mPipNat1), collected from Geneva, Switzerland, on 2024-02-16. The bat was found on the ground with the right wing broken, and had to be euthanised. Upon dissection, organs were placed immediately in Eppendorf tubes placed in dry ice and transferred to a − 80 °C freezer within 1 h. The specimen was identified by Manuel Ruedi and vouchered at the Natural History Museum of Geneva (specimen number MHNG-MAMO-3018.095). The same specimen was used for RNA sequencing.

Nucleic acid extraction

Protocols for high molecular weight (HMW) DNA extraction developed at the Wellcome Sanger Institute (WSI) Tree of Life Core Laboratory are available on protocols.io (Howard *et al.*, 2025). The mPipNat1 sample was weighed and triaged to determine the appropriate extraction protocol. Tissue from the heart was homogenised by powermashing using a PowerMasher II tissue disruptor. HMW DNA was extracted using the Automated MagAttract v2 protocol. DNA was sheared into an average fragment size of 12–20 kb following the Megaruptor®3 for LI PacBio protocol. Sheared DNA was purified by automated SPRI (solid-phase reversible immobilisation). The concentration of the sheared and purified DNA was assessed using a Nanodrop spectrophotometer and Qubit Fluorometer using the Qubit dsDNA High Sensitivity Assay kit. Fragment size distribution was evaluated by running the sample on the FemtoPulse system. For this sample, the final post-shearing DNA had a Qubit concentration of 19.6 ng/μL and a yield of 2 548.00 ng.

RNA was extracted from heart tissue of mPipNat1 in the Tree of Life Laboratory at the WSI using the RNA Extraction: Automated MagMax™ *mir*Vana protocol. The RNA concentration was assessed using a Nanodrop spectrophotometer and a Qubit Fluorometer using the Qubit RNA Broad-Range Assay kit. Analysis of the integrity of the RNA was done using the Agilent RNA 6000 Pico Kit and Eukaryotic Total RNA assay.

PacBio HiFi library preparation and sequencing

Library preparation and sequencing were performed at the WSI Scientific Operations core. Libraries were prepared using the SMRTbell Prep Kit 3.0 (Pacific Biosciences, California, USA), following the manufacturer's instructions. The kit includes reagents for end repair/A-tailing, adapter ligation, post-ligation SMRTbell bead clean-up, and nuclease treatment. Size selection and clean-up were performed using diluted AMPure PB beads (Pacific Biosciences). DNA concentration was quantified using a Qubit Fluorometer v4.0 (ThermoFisher Scientific) and the Qubit 1X dsDNA HS assay kit. Final library fragment size was assessed with the Agilent Femto Pulse Automated Pulsed Field CE Instrument (Agilent Technologies) using the gDNA 55 kb BAC analysis kit.

Figure 1. Photograph of *Pipistrellus nathusii* (not the individual used for genome sequencing). Photograph by Manuel Ruedi.

The sample was sequenced on a Revio instrument (Pacific Biosciences). The prepared library was normalised to 2 nM, and 15 µL was used for making complexes. Primers were annealed and polymerases bound to generate circularised complexes, following the manufacturer's instructions. Complexes were purified using 1.2X SMRTbell beads, then diluted to the Revio loading concentration (200–300 pM) and spiked with a Revio sequencing internal control. The sample was sequenced on a Revio 25 M SMRT cell. The SMRT Link software (Pacific Biosciences), a web-based workflow manager, was used to configure and monitor the run and to carry out primary and secondary data analysis.

Hi-C

Sample preparation and crosslinking

The Hi-C sample was prepared from 20–50 mg of frozen heart tissue from the mPipNat1 sample using the Arima-HiC v2 kit (Arima Genomics). Following the manufacturer's instructions, tissue was fixed and DNA crosslinked using TC buffer to a final formaldehyde concentration of 2%. The tissue was homogenised using the Diagnocine Power Masher-II. Crosslinked DNA was digested with a restriction enzyme master mix, biotinylated, and ligated. Clean-up was performed with SPRISelect beads before library preparation. DNA concentration was measured with the Qubit Fluorometer (Thermo Fisher Scientific) and Qubit HS Assay Kit. The biotinylation percentage was estimated using the Arima-HiC v2 QC beads.

Hi-C library preparation and sequencing

Biotinylated DNA constructs were fragmented using a Covaris E220 sonicator and size selected to 400–600 bp using SPRISelect beads. DNA was enriched with Arima-HiC v2 kit Enrichment beads. End repair, A-tailing, and adapter ligation were carried out with the NEBNext Ultra II DNA Library Prep Kit (New England Biolabs), following a modified protocol where library preparation occurs while DNA remains bound to the Enrichment beads. Library amplification was performed using KAPA HiFi HotStart mix and a custom Unique Dual Index (UDI) barcode set (Integrated DNA Technologies). Depending on sample concentration and biotinylation percentage determined at the crosslinking stage, libraries were amplified with 10–16 PCR cycles. Post-PCR clean-up was performed with SPRISelect beads. Libraries were quantified using the AccuClear Ultra High Sensitivity dsDNA Standards Assay Kit (Biotium) and a FLUOstar Omega plate reader (BMG Labtech).

Prior to sequencing, libraries were normalised to 10 ng/µL. Normalised libraries were quantified again to create equimolar and/or weighted 2.8 nM pools. Pool concentrations were checked using the Agilent 4200 TapeStation

(Agilent) with High Sensitivity D500 reagents before sequencing. Sequencing was performed using paired-end 150 bp reads on the Illumina NovaSeq X.

RNA library preparation and sequencing

Libraries were prepared using the NEBNext® Ultra™ II Directional RNA Library Prep Kit for Illumina (New England Biolabs), following the manufacturer's instructions. Poly(A) mRNA in the total RNA solution was isolated using oligo (dT) beads, converted to cDNA, and uniquely indexed; 14 PCR cycles were performed. Libraries were size-selected to produce fragments between 100–300 bp. Libraries were quantified, normalised, pooled to a final concentration of 2.8 nM, and diluted to 150 pM for loading. Sequencing was carried out on the Illumina NovaSeq X, generating paired-end reads.

Genome assembly

Prior to assembly of the PacBio HiFi reads, a database of *k*-mer counts ($k = 31$) was generated from the filtered reads using FastK. GenomeScope2 (Ranallo-Benavidez *et al.*, 2020) was used to analyse the *k*-mer frequency distributions, providing estimates of genome size, heterozygosity, and repeat content.

The HiFi reads were assembled using Hifiasm in Hi-C phasing mode (Cheng *et al.*, 2021, 2022), producing two haplotypes. Hi-C reads (Rao *et al.*, 2014) were mapped to the primary contigs using bwa-mem2 (Vasimuddin *et al.*, 2019). Contigs were further scaffolded with Hi-C data in YaHS (Zhou *et al.*, 2023), using the --break option for handling potential misassemblies. The scaffolded assemblies were evaluated using Gfastats (Formenti *et al.*, 2022), BUSCO (Manni *et al.*, 2021) and MerquryFK (Rhie *et al.*, 2020).

The mitochondrial genome was assembled using MitoHiFi (Uliano-Silva *et al.*, 2023).

Assembly curation

The assembly was decontaminated using the Assembly Screen for Cobionts and Contaminants (ASCC) pipeline. TreeVal was used to generate the flat files and maps for use in curation. Manual curation was conducted primarily in PretextView and HiGlass (Kerpedjiev *et al.*, 2018). Scaffolds were visually inspected and corrected as described by Howe *et al.* (2021). Manual corrections included 11 breaks and 24 joins. This reduced the scaffold count by 3.9% and increased the scaffold N50 by 3.2%. The curation process is described at https://gitlab.com/wtsi-grit/rapid-curation. PretextSnapshot was used to generate a Hi-C contact map of the final assembly.

Assembly quality assessment

The Merqury.FK tool (Rhie *et al.*, 2020) was run in a Singularity container (Kurtzer *et al.*, 2017) to evaluate *k*-mer completeness and assembly quality for both haplotypes using the *k*-mer databases ($k = 31$) computed prior to genome assembly. The analysis outputs included assembly QV scores and completeness statistics.

The genome was analysed using the BlobToolKit pipeline, a Nextflow implementation of the earlier Snakemake version (Challis *et al.*, 2020). The pipeline aligns PacBio reads using minimap2 (Li, 2018) and SAMtools (Danecek *et al.*, 2021) to generate coverage tracks. It runs BUSCO (Manni *et al.*, 2021) using lineages identified from the NCBI Taxonomy (Schoch *et al.*, 2020). For the three domain-level lineages, BUSCO genes are aligned to the UniProt Reference Proteomes database (Bateman *et al.*, 2023) using DIAMOND blastp (Buchfink *et al.*, 2021). The genome is divided into chunks based on the density of BUSCO genes from the closest taxonomic lineage, and each chunk is aligned to the UniProt Reference Proteomes database with DIAMOND blastx. Sequences without hits are chunked using seqtk and aligned to the NT database with blastn (Altschul *et al.*, 1990). The BlobToolKit suite consolidates all outputs into a blobdir for visualisation. The BlobToolKit pipeline was developed using nf-core tooling (Ewels *et al.*, 2020) and MultiQC (Ewels *et al.*, 2016), with containerisation through Docker (Merkel, 2014) and Singularity (Kurtzer *et al.*, 2017).

Genome sequence report

Sequence data

PacBio sequencing of the *Pipistrellus nathusii* specimen generated 93.23 Gb (gigabases) from 7.23 million reads, which were used to assemble the genome. GenomeScope2.0 analysis estimated the haploid genome size at 1 767.67 Mb, with a heterozygosity of 0.72% and repeat content of 16.40% (Figure 2). These estimates guided expectations for the assembly. Based on the estimated genome size, the sequencing data provided approximately 51× coverage. Hi-C sequencing produced 240.67 Gb from 1 593.83 million reads, which were used to scaffold the assembly. RNA sequencing data were also generated and are available in public sequence repositories. Table 1 summarises the specimen and sequencing details.

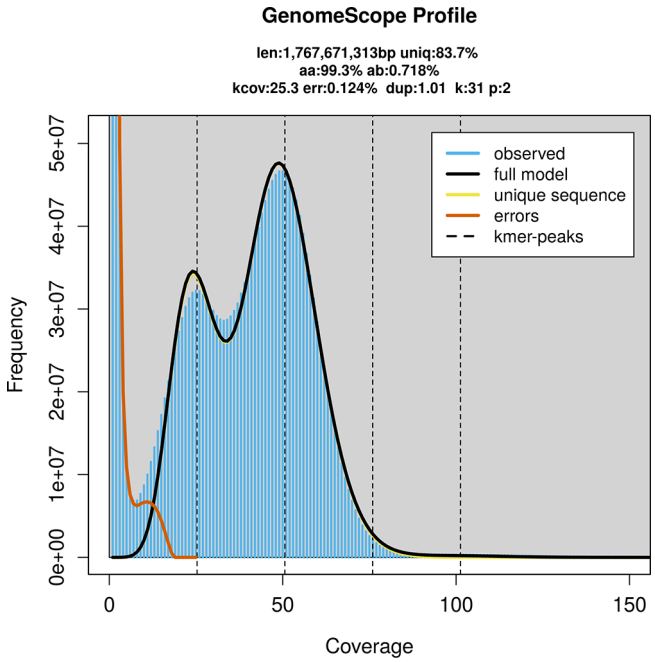

Figure 2. Frequency distribution of *k*-mers generated using GenomeScope2. The plot shows observed and modelled *k*-mer spectra, providing estimates of genome size, heterozygosity, and repeat content based on unassembled sequencing reads.

Table 1. Specimen and sequencing data for BioProject PRJEB81652.

Platform	PacBio HiFi	Hi-C	RNA-seq
ToLID	mPipNat1	mPipNat1	mPipNat1
Specimen ID	SAN00003686	SAN00003686	SAN00003686
BioSample (source individual)	SAMEA115534661	SAMEA115534661	SAMEA115534661
BioSample (tissue)	SAMEA115534667	SAMEA115534667	SAMEA115534666
Tissue	heart	heart	heart
Instrument	Revio	Illumina NovaSeq X	Illumina NovaSeq X
Run accessions	ERR13900464	ERR13907248	ERR15140913
Read count total	7.23 million	1 593.83 million	96.15 million
Base count total	93.23 Gb	240.67 Gb	14.52 Gb

Assembly statistics

The genome was assembled into two haplotypes using Hi-C phasing. Haplotype 1 was curated to chromosome level, while haplotype 2 was assembled to scaffold level. The final assembly has a total length of 1 961.71 Mb in 396 scaffolds, with 307 gaps, and a scaffold N50 of 89.97 Mb (Table 2).

Most of the haplotype 1 assembly sequence (92.66%) was assigned to 22 chromosomal-level scaffolds, representing 21 autosomes and the X sex chromosome. These chromosome-level scaffolds, confirmed by Hi-C data, are named according to synteny (Figure 3; Table 3). The X chromosome was identified based on synteny with the genome of *Myotis nattereri* (GCA_964212035.1).

The mitochondrial genome was also assembled (length 17.27 kb, OZ209765.2). This sequence is included as a contig in the multifasta file of the genome submission and as a standalone record.

Table 2. **Genome assembly statistics.**

Assembly name	mPipNat1.hap1.2	mPipNat1.hap2.2
Assembly accession	GCA_964656225.2	GCA_964656075.2
Assembly level	chromosome	scaffold
Span (Mb)	1 961.71	1 889.00
Number of chromosomes	22	scaffold-level
Number of contigs	703	759
Contig N50	25.16 Mb	25.75 Mb
Number of scaffolds	396	453
Scaffold N50	89.97 Mb	87.7 Mb
Longest scaffold length (Mb)	211.51	-
Sex chromosomes	X	-
Organelles	Mitochondrion: 17.27 kb	-

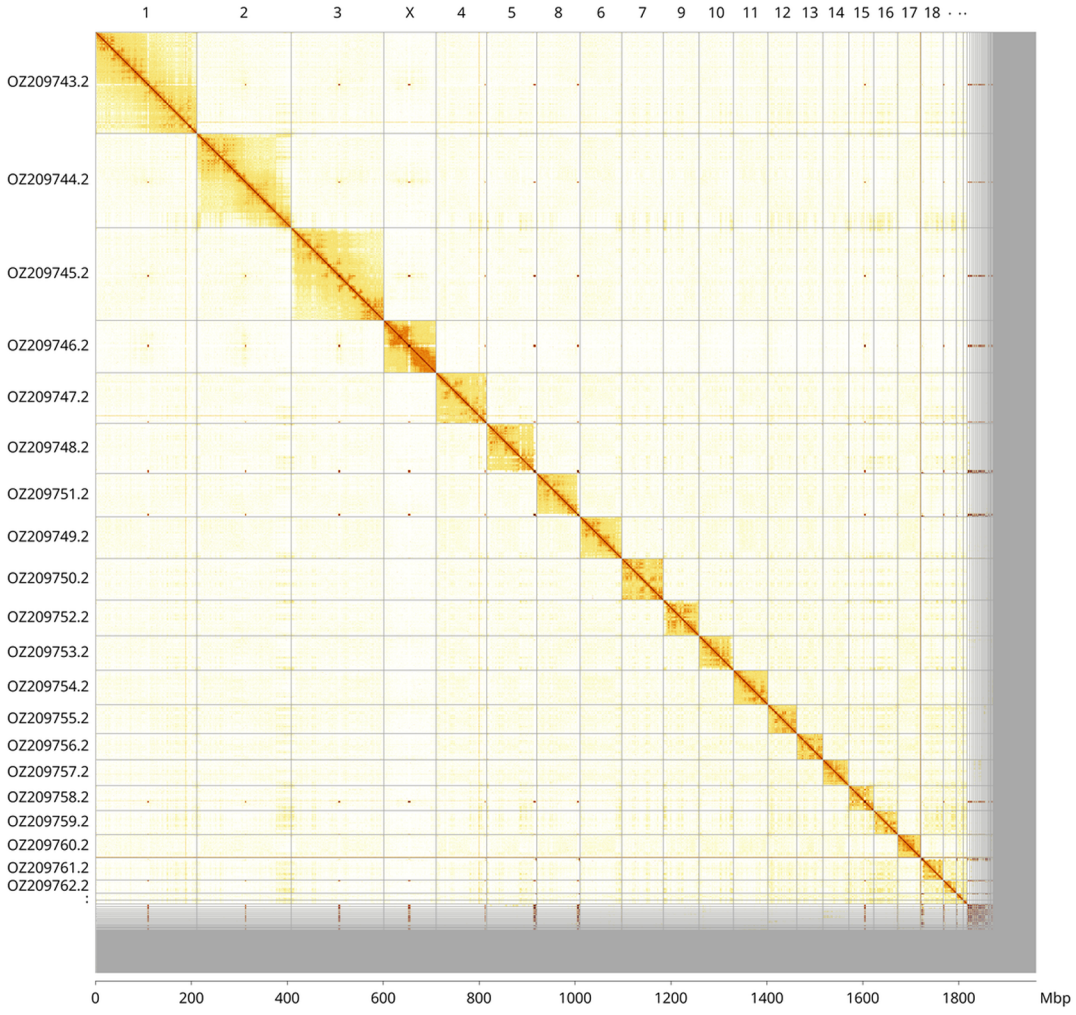

Figure 3. **Hi-C contact map of the *Pipistrellus nathusii* genome assembly.** Assembled chromosomes are shown in order of size and labelled along the axes, with a megabase scale shown below. The plot was generated using PretextSnapshot.

Table 3. Chromosomal pseudomolecules in the haplotype 1 genome assembly of *Pipistrellus nathusii* mPipNat1.

INSDC accession	Molecule	Length (Mb)	GC%
OZ209743.2	1	211.51	41
OZ209744.2	2	196.81	41.50
OZ209745.2	3	193.32	40.50
OZ209747.2	4	105.69	42
OZ209748.2	5	104.51	43.50
OZ209749.2	6	87.16	41
OZ209750.2	7	86.80	43
OZ209751.2	8	89.97	41
OZ209752.2	9	74.36	43.50
OZ209753.2	10	71.96	43
OZ209754.2	11	71.89	41.50
OZ209755.2	12	60.09	43
OZ209756.2	13	54.45	44
OZ209757.2	14	54.14	44
OZ209758.2	15	51.71	47
OZ209759.2	16	49.88	47.50
OZ209760.2	17	48.34	41.50
OZ209761.2	18	46.67	47
OZ209762.2	19	27.41	48.50
OZ209763.2	20	14.37	47
OZ209764.2	21	7.98	51
OZ209746.2	X	108.69	39.50

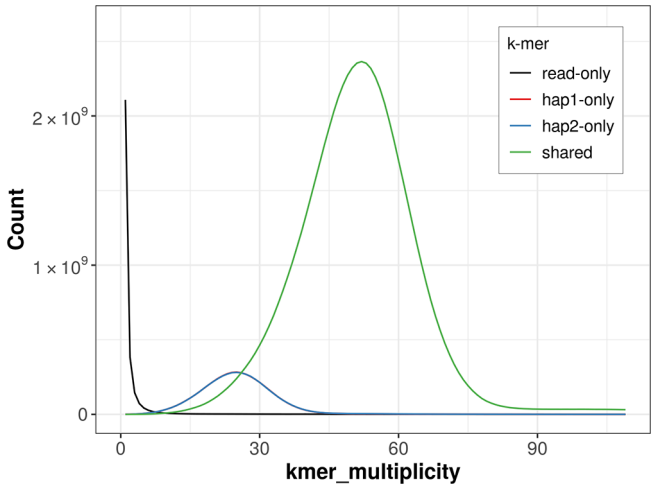

Figure 4. Evaluation of *k*-mer completeness using MerquryFK. This plot illustrates the recovery of *k*-mers from the original read data in the final assemblies. The horizontal axis represents *k*-mer multiplicity, and the vertical axis shows the number of *k*-mers. The black curve represents *k*-mers that appear in the reads but are not assembled. The green curve corresponds to *k*-mers shared by both haplotypes, and the red and blue curves show *k*-mers found only in one of the haplotypes.

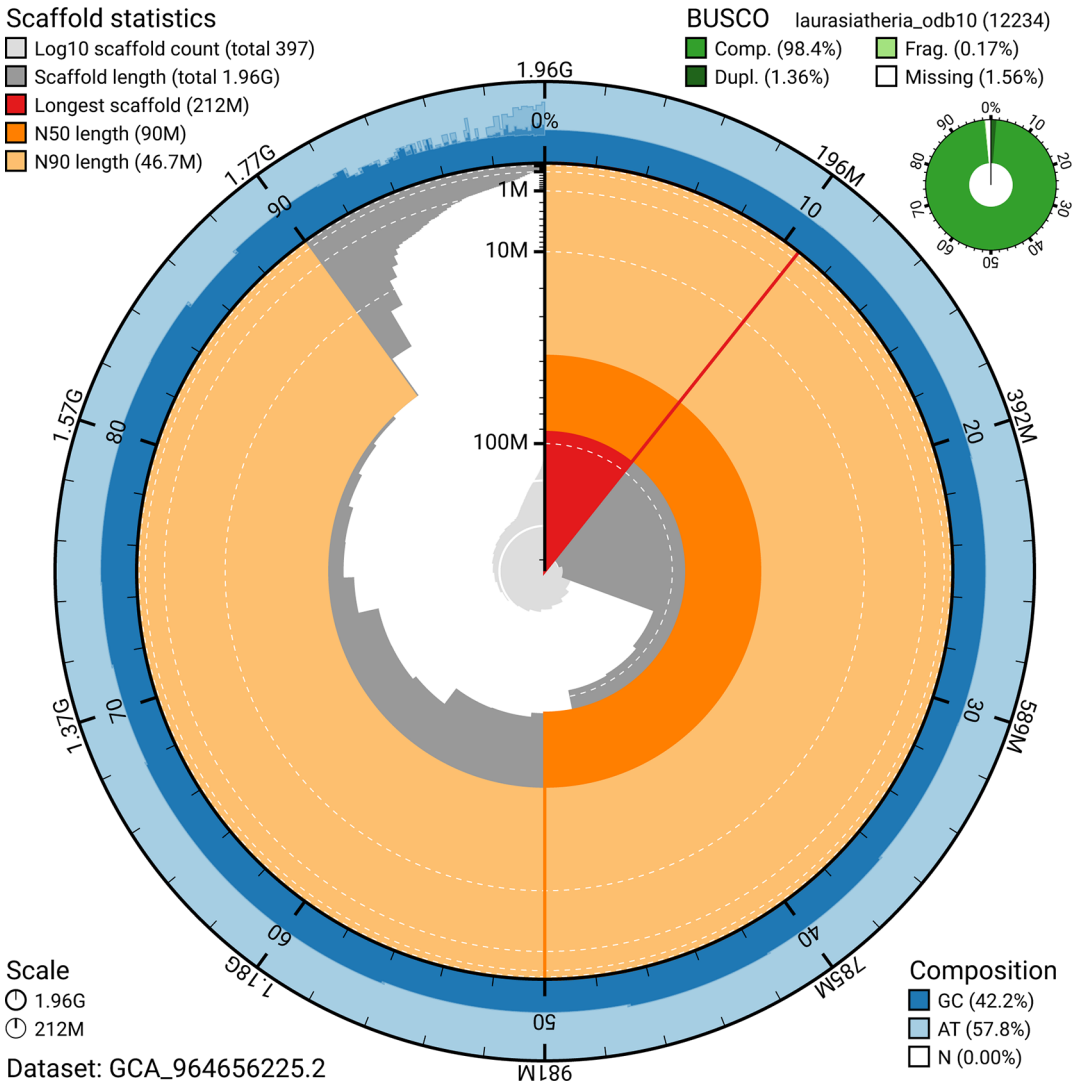

Figure 5. Assembly metrics for mPipNat1.hap1.2. The BlobToolKit snail plot provides an overview of assembly metrics and BUSCO gene completeness. The circumference represents the length of the whole genome sequence, and the main plot is divided into 1 000 bins around the circumference. The outermost blue tracks display the distribution of GC, AT, and N percentages across the bins. Scaffolds are arranged clockwise from longest to shortest and are depicted in dark grey. The longest scaffold is indicated by the red arc, and the deeper orange and pale orange arcs represent the N50 and N90 lengths. A light grey spiral at the centre shows the cumulative scaffold count on a logarithmic scale. A summary of complete, fragmented, duplicated, and missing BUSCO genes in the set is presented at the top right. An interactive version of this figure can be accessed on the BlobToolKit viewer.

Assembly quality metrics

For haplotype 1, the estimated QV is 61.2, and for haplotype 2, 61.4. When the two haplotypes are combined, the assembly achieves an estimated QV of 61.3. The *k*-mer completeness is 87.67% for haplotype 1, 87.68% for haplotype 2, and 99.57% for the combined haplotypes (Figure 4).

BUSCO analysis using the laurasiatheria_odb10 reference set (n = 12 234) identified 98.4% of the expected gene set (single = 97.1%, duplicated = 1.4%) in haplotype 1. For haplotype 2, BUSCO v.6.0.0 analysis identified 98.4% of the expected gene set (single = 97.1%, duplicated = 1.3%). The snail plot in Figure 5 summarises the scaffold length distribution and other assembly statistics for haplotype 1. The blob plot in Figure 6 shows the distribution of scaffolds by GC proportion and coverage for haplotype 1.

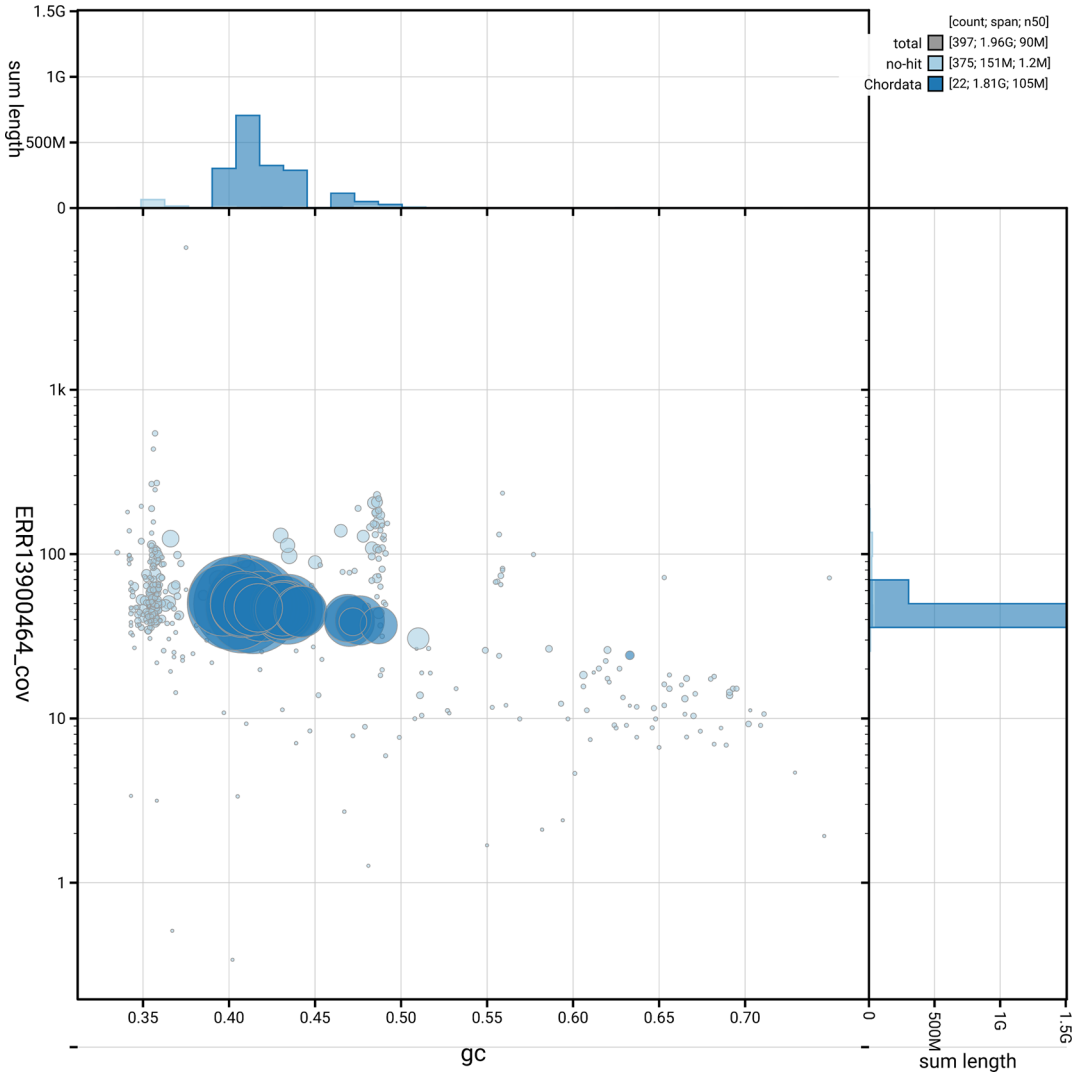

Figure 6. BlobToolKit GC-coverage plot for mPipNat1.hap1.2. Blob plot showing sequence coverage (vertical axis) and GC content (horizontal axis). The circles represent scaffolds, with the size proportional to scaffold length and the colour representing phylum membership. The histograms along the axes display the total length of sequences distributed across different levels of coverage and GC content. An interactive version of this figure is available on the BlobToolKit viewer.

Table 4. Earth biogenome project summary metrics for the *Pipistrellus nathusii* assembly.

Measure	Value	Benchmark
EBP summary (haplotype 1)	7.C.Q61	6.C.Q40
Contig N50 length	25.16 Mb	≥ 1 Mb
Scaffold N50 length	89.97 Mb	= chromosome N50
Consensus quality (QV)	Haplotype 1: 61.2; haplotype 2: 61.4; combined: 61.3	≥ 40
k-mer completeness	Haplotype 1: 87.67%; Haplotype 2: 87.68%; combined: 99.57%	≥ 95%
BUSCO	C:100.0% [S:91.1%, D:8.9%], F:0.0%, M:0.0%, n:954	S > 90%; D < 5%
Percentage of assembly assigned to chromosomes	92.66%	≥ 90%

Notes: The EBP summary uses log10(Contig N50); chromosome-level (C) or log10(Scaffold N50); Q (Merqury QV). BUSCO: C = complete; S = single-copy; D = duplicated; F = fragmented; M = missing; n = orthologues

Table 5. Software versions and sources.

Software	Version	Source
BLAST	2.14.0	ftp://ftp.ncbi.nlm.nih.gov/blast/executables/blast+/
BlobToolKit	4.4.6	https://github.com/blobtoolkit/blobtoolkit
BUSCO	6.0.0	https://gitlab.com/ezlab/busco
bwa-mem2	2.2.1	https://github.com/bwa-mem2/bwa-mem2
DIAMOND	2.1.8	https://github.com/bbuchfink/diamond
fasta_windows	0.2.4	https://github.com/tolkit/fasta_windows
FastK	1.1	https://github.com/thegenemyers/FASTK
GenomeScope2.0	2.0.1	https://github.com/tbenavi1/genomescope2.0
Gfastats	1.3.6	https://github.com/vgl-hub/gfastats
Hifiasm	0.19.8-r603	https://github.com/chhylp123/hifiasm
HiGlass	1.13.4	https://github.com/higlass/higlass
MerquryFK	1.1.2	https://github.com/thegenemyers/MERQURY.FK
Minimap2	2.28-r1209	https://github.com/lh3/minimap2
MitoHiFi	3	https://github.com/marcelauliano/MitoHiFi
MultiQC	1.14; 1.17 and 1.18	https://github.com/MultiQC/MultiQC
Nextflow	24.10.4	https://github.com/nextflow-io/nextflow
PretextSnapshot	0.0.5	https://github.com/sanger-tol/PretextSnapshot
PretextView	1.0.3	https://github.com/sanger-tol/PretextView
samtools	1.21	https://github.com/samtools/samtools
sanger-tol/ascc	0.1.0	https://github.com/sanger-tol/ascc
sanger-tol/blobtoolkit	v0.9.0	https://github.com/sanger-tol/blobtoolkit
sanger-tol/curationpretext	1.4.2	https://github.com/sanger-tol/curationpretext
Seqtk	1.3	https://github.com/lh3/seqtk
Singularity	3.9.0	https://github.com/sylabs/singularity
TreeVal	1.4.0	https://github.com/sanger-tol/treeval
YaHS	1.2.2	https://github.com/c-zhou/yahs

Table 4 lists the assembly metric benchmarks adapted from Rhie *et al.* (2021) and the Earth BioGenome Project Report on Assembly Standards January 2026. The EBP metric, calculated for the haplotype 1, is **7.C.Q61**, meeting the recommended reference standard.

Author information

Contributors are listed at the following links:

- Members of the Wellcome Sanger Institute Tree of Life Management, Samples and Laboratory team
- Members of Wellcome Sanger Institute Scientific Operations – Sequencing Operations
- Members of the Wellcome Sanger Institute Tree of Life Core Informatics team
- Members of the Tree of Life Core Informatics collective
- Members of the Darwin Tree of Life Consortium

Wellcome sanger institute – legal and governance

The materials that have contributed to this genome note have been supplied by a Darwin Tree of Life Partner. The submission of materials by a Darwin Tree of Life Partner is subject to the **'Darwin Tree of Life Project Sampling Code of Practice'**, which can be found in full on the Darwin Tree of Life website. By agreeing with and signing up to the Sampling Code of Practice, the Darwin Tree of Life Partner agrees they will meet the legal and ethical requirements and standards set out within this document in respect of all samples acquired for, and supplied to, the Darwin Tree of Life Project. Further, the Wellcome Sanger Institute employs a process whereby due diligence is carried out proportionate to the nature of the materials themselves, and the circumstances under which they have been/are to be collected and provided for use. The purpose of this is to address and mitigate any potential legal and/or ethical implications of receipt and use of the materials as part of the research project, and to ensure that in doing so we align with best practice wherever possible. The overarching areas of consideration are:

- Ethical review of provenance and sourcing of the material
- Legality of collection, transfer and use (national and international)

Each transfer of samples is further undertaken according to a Research Collaboration Agreement or Material Transfer Agreement entered into by the Darwin Tree of Life Partner, Genome Research Limited (operating as the Wellcome Sanger Institute), and in some circumstances, other Darwin Tree of Life collaborators.

Data availability

European Nucleotide Archive: Pipistrellus nathusii (Nathusius's pipistrelle). Accession number PRJEB81652. The genome sequence is released openly for reuse. The *Pipistrellus nathusii* genome sequencing initiative is part of the Darwin Tree of Life Project (PRJEB40665), the Sanger Institute Tree of Life Programme (PRJEB43745), the Vertebrate Genomes Project (PRJNA489243) and the Bat1K Project (PRJNA489245). All raw sequence data and the assembly have been deposited in INSDC databases. The genome will be annotated using available RNA-Seq data and presented through the Ensembl pipeline at the European Bioinformatics Institute. Raw data and assembly accession identifiers are reported in Tables 1 and 2.

Production code used in genome assembly at the WSI Tree of Life is available at https://github.com/sanger-tol. Table 5 lists software versions used in this study.

Acknowledgements

We thank the coordination centre for bat conservation in Geneva (CCO-GE), which provided permissions for collection procedures.

References

Altschul SF, Gish W, Miller W, *et al.*: **Basic Local Alignment Search Tool.** *J. Mol. Biol.* 1990; **215**(3): 403–410.
PubMed Abstract | Publisher Full Text

Bateman A, Martin M-J, Orchard S, *et al.*: **UniProt: The Universal Protein Knowledgebase in 2023.** *Nucleic Acids Res.* 2023; **51**(D1): D523–D531.
PubMed Abstract | Publisher Full Text | Free Full Text

Buchfink B, Reuter K, Drost H-G: **Sensitive protein alignments at tree-of-life scale using DIAMOND.** *Nat. Methods.* 2021; **18**(4): 366–368.
PubMed Abstract | Publisher Full Text | Free Full Text

Challis R, Richards E, Rajan J, *et al.*: **BlobToolKit – interactive quality assessment of genome assemblies.** *G3: Genes, Genomes, Genetics.* 2020; **10**(4): 1361–1374.
PubMed Abstract | 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 | Free Full Text

Cheng H, Jarvis ED, Fedrigo O, *et al.*: **Haplotype-resolved assembly of diploid genomes without parental data.** *Nat. Biotechnol.* 2022; **40**(9): 1332–1335.
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**(2).
PubMed Abstract | Publisher Full Text | Free Full Text

Darwin Tree of Life Project Consortium: **Sequence locally, think globally: The Darwin Tree of Life Project.** *Proc. Natl. Acad. Sci. USA.* 2022; **119**(4): e2115642118.
PubMed Abstract | Publisher Full Text | Free Full Text

Dool SE, Puechmaille SJ: **Challenges in the Vespertilionidae phylogeny: Resolving *Pipistrellus nathusii* placement and affirming generic status for Asian pipistrelles.** *J. Mammal.* 2024; **106**(2): 457–467.
Publisher Full Text

Ewels P, Magnusson M, Lundin S, *et al.*: **MultiQC: Summarize analysis results for multiple tools and samples in a single report.** *Bioinformatics.* 2016; **32**(19): 3047–3048.
PubMed Abstract | Publisher Full Text | Free Full Text

Ewels PA, Peltzer A, Fillinger S, *et al.*: **The nf-core framework for community-curated bioinformatics pipelines.** *Nat. Biotechnol.* 2020; **38**(3): 276–278.
PubMed Abstract | Publisher Full Text

Formenti G, Abueg L, Brajuka A, *et al.*: **Gfastats: Conversion, evaluation and manipulation of genome sequences using assembly graphs.**

Bioinformatics. 2022; **38**(17): 4214–4216.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

Howard C, Denton A, Jackson B, *et al.*: **On the path to reference genomes for all biodiversity: Lessons learned and laboratory protocols created in the Sanger Tree of Life core laboratory over the first 2000 species.** *bioRxiv.* 2025.
Publisher Full Text

Howe K, Chow W, Collins J, *et al.*: **Significantly improving the quality of genome assemblies through curation.** *GigaScience.* 2021; **10**(1).
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

Kerpedjiev P, Abdennur N, Lekschas F, *et al.*: **HiGlass: Web-based visual exploration and analysis of genome interaction maps.** *Genome Biol.* 2018; **19**(1): 125.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

Kurtzer GM, Sochat V, Bauer MW: **Singularity: Scientific containers for mobility of compute.** *PLOS ONE.* 2017; **12**(5): e0177459.
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**

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** | **Free Full Text**

Merkel D: **Docker: Lightweight Linux containers for consistent development and deployment.** *Linux J.* 2014; **2014**(239).
Publisher Full Text

Ranallo-Benavidez TR, Jaron KS, Schatz MC: **GenomeScope 2.0 and Smudgeplot for reference-free profiling of polyploid genomes.** *Nat. Commun.* 2020; **11**(1): 1432.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

Rao SSP, Huntley MH, Durand NC, *et al.*: **A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping.** *Cell.* 2014; **159**(7): 1665–1680.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

Rhie A, McCarthy SA, Fedrigo O, *et al.*: **Towards complete and error-free genome assemblies of all vertebrate species.** *Nature.* 2021; **592**(7856): 737–746.
PubMed Abstract | **Publisher Full Text** | **Free 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).
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

Roehrs ZP, Lack JB, Van Den Bussche RA: **Tribal phylogenetic relationships within Vespertilioninae (Chiroptera: Vespertilionidae) based on mitochondrial and nuclear sequence data.** *J. Mammal.* 2010; **91**(5): 1073–1092.
Publisher Full Text

Russ J: **Nathusius's pipistrelle *Pipistrellus nathusii* (Keyserling and Blasius, 1839).** Russo D, editor. *Handbook of the Mammals of Europe.* Switzerland: Springer Nature; 2023; 529–54.
Publisher Full Text

Schoch CL, Ciufo S, Domrachev M, *et al.*: **NCBI taxonomy: A comprehensive update on curation, resources and tools.** *Database.* 2020; **2020**: baaa062.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

Teeling EC, Vernes SC, Dávalos LM, *et al.*: **Bat biology, genomes, and the Bat1K project: To generate chromosome-level genomes for all living bat species.** *Annu. Rev. Anim. Biosci.* 2018; **6**: 23–46.
PubMed Abstract | **Publisher Full Text**

Uliano-Silva M, Ferreira JGRN, Krasheninnikova K, *et al.*: **MitoHiFi: A Python pipeline for mitochondrial genome assembly from PacBio high fidelity reads.** *BMC Bioinformatics.* 2023; **24**(1): 288.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

Vasimuddin M, Misra S, Li H, *et al.*: **Efficient architecture-aware acceleration of BWA-MEM for multicore systems.** *2019 IEEE International Parallel and Distributed Processing Symposium (IPDPS).* IEEE; 2019; 314–24.
Publisher Full Text

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

Open Peer Review

Current Peer Review Status: ✔ ✔

Version 1

Reviewer Report 15 June 2026

https://doi.org/10.21956/wellcomeopenres.28977.r157928

✔ **Xiuguang Mao**
East China Normal University, Shanghai, China

This manuscript provides a high-quality chromosome-level genome assembly for *Pipistrellus nathusii*. Together with the previous high-quality assembly of this species, these genome resources will greatly contribute to bat research. For the current version of the manuscript, I have only three minor comments:
1) "The mitochondrial genome was assembled using MitoHiFi." How much PacBio HiFi data (all or a subset — please be specific) was used for the mitochondrial genome assembly?
2) "The final assembly has a total length of 1,961.71 Mb," which is larger than the haploid genome size estimated by GenomeScope 2.0 analysis (1,767.67 Mb). What is the reason for this inconsistency?
3) In Table 3, if the chromosomes are numbered based on their length, why is Chr8 (89.97 Mb) longer than Chr7 (86.80 Mb)?

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: Comparative genomics, genome assembly, evolutionary biology, speciation, bats

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.

Reviewer Report 12 May 2026

https://doi.org/10.21956/wellcomeopenres.28977.r154089

Luciano Chaves Franco Filho
Instituto Evandro Chagas, Ananindeua, Brazil

The manuscript presents a chromosome-level genome assembly of *Pipistrellus nathusii* (Nathusius' pipistrelle), generated using PacBio HiFi, Hi-C, and RNA-seq technologies within the Bat1K and Darwin Tree of Life initiatives. The manuscript is well organized and presents relevant genomic data for *Pipistrellus nathusii*. The methodologies and assembly metrics are adequately described, and the study provides a useful genomic resource for the scientific community. I consider the manuscript suitable for indexing following standard editorial evaluation.

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: I have expertise in genomics, metagenomics, molecular biology, and molecular diagnostics applied to infectious diseases and public health. My work focuses on genomic surveillance, pathogen detection, sequencing technologies, and the application of molecular approaches to investigate microbial diversity, antimicrobial resistance, and epidemiological monitoring.

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.