DATA NOTE

The genome sequence of Geoffroy's Bat, *Myotis emarginatus* (Geoffroy, 1806) (Chiroptera: Vespertilionidae)

[version 1; peer review: 2 approved, 1 approved with reservations]

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:** 02 May 2026, **11**:258
https://doi.org/10.12688/wellcomeopenres.26492.1
Latest published: 02 May 2026, **11**:258
https://doi.org/10.12688/wellcomeopenres.26492.1

Abstract

We present a genome assembly from an individual female *Myotis emarginatus* (Geoffroy's Bat; Chordata; Mammalia; Chiroptera; Vespertilionidae). The assembly contains two haplotypes with total lengths of 2 110.52 megabases and 2 084.88 megabases. Most of haplotype 1 (96.18%) is scaffolded into 22 chromosomal pseudomolecules, including the X sex chromosome, while 94.98% of haplotype 2 is scaffolded into 22 chromosomal pseudomolecules, including the X sex chromosome. The mitochondrial genome has also been assembled, with a length of 17.1 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

Myotis emarginatus, Geoffroy's Bat, genome sequence, chromosomal, Chiroptera

Open Peer Review

Approval Status ✓ ✓ ?

	1	2	3
version 1 02 May 2026	✓ view	✓ view	? view

1. **Heliana Dundarova**, Bulgarian Academy of Sciences, Sofia, Bulgaria
2. **Keping Sun**, Northeast Normal University, Changchun, China
3. **Michael R Buchalski**, Wildlife Health Laboratory, California Department of Fish and Wildlife, Sacramento, USA
Samantha Capel, California Department of Fish and Wildlife, Sacramento, Sacramento, USA

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.

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**: Resources, 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 received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement BATSPEAK:101001702). ECT is a Wellcome collaborator and the 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 Geoffroy's Bat, *Myotis emarginatus* (Geoffroy, 1806) (Chiroptera: Vespertilionidae) [version 1; peer review: 2 approved, 1 approved with reservations]** Wellcome Open Research 2026, **11**:258 https://doi.org/10.12688/wellcomeopenres.26492.1

First published: 02 May 2026, **11**:258 https://doi.org/10.12688/wellcomeopenres.26492.1

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; *Myotis*; *Myotis emarginatus* (E.Geoffroy, 1806) (NCBI:txid109480).

Background

Geoffroy's bat (*Myotis emarginatus*) is a medium-sized bat (adult weight 7–10 g, forearm length 38–43 mm) characterised by relatively long, deeply notched ears and a unique woolly fur. Its pelage is essentially yellowish-brown with little contrast between dorsal and ventral sides. Geoffroy's bat is the only European representative of the African clade of *Myotis* (Morales *et al.*, 2024), sometimes classified in the subgenus *Chrysopteron* (Csorba *et al.*, 2014).

Myotis emarginatus preys on a variety of arthropods, but primarily feeds on spiders and flies, which it gleans directly from hard surfaces and leaves. Its preference for blood-feeding stable flies (e.g. *Stomoxys calcitrans*) makes it a valuable ally to farmers. Favoured hunting grounds include all types of forests and pasturelands.

In early summer, females gather in nursery colonies located in underground cavities or human-made structures. These roosts are often shared with other species, such as rhinolophid bats, forming mixed colonies. In autumn, Geoffroy's bats mate at swarming sites, which serve as hotspots for genetic exchange (Foley *et al.*, 2024).

This thermophilous species is distributed in continental Europe and the Mediterranean Basin and extends its range into the Middle East and Afghanistan. Although it is globally considered as Least Concern (Lc) under IUCN criteria, the species is declining in many northern European countries.

We present a chromosome-level genome sequence for *Myotis emarginatus*, generated using the Tree of Life pipeline from a specimen collected from Anglefort, France (Figure 1). This assembly was generated as part of the Darwin Tree of Life Project, which aims to generate high-quality reference genomes for all named eukaryotic species in Britain and Ireland to support research, conservation, and the sustainable use of biodiversity (Darwin Tree of Life Project Consortium, 2022).

Methods

Sample acquisition

The genome is based on an immature female *Myotis emarginatus* (specimen ID SAN00003684, ToLID mMyoEma1; Figure 1) which was found injured by a cat at Anglefort, France on 2021-08-16. It had a broken wing and developed sever malformation which necessitated euthanasia after one month of captivity. After dissection, the carcass was deposited in the collections of the Natural History Museum of Geneva under voucher number MHNG-MAMO-3012.090. Organs were immediately frozen on dry ice and stored at −80 °C. The specimen was identified by Manuel Ruedi based on morphological and mitochondrial DNA characters [GenBank OQ706656 and OQ885399; Ruedi *et al.* (2023)]. 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 mMyoEma1 sample was weighed and triaged to determine the appropriate extraction protocol. Heart tissue was homogenised by powermashing using a PowerMasher II tissue disruptor. HMW DNA was extracted using version 2 of the Manual MagAttract 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 28.8 ng/μL and a yield of 3 744.00 ng.

RNA was extracted from spleen tissue of mMyoEma1 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.

Figure 1. Photograph of the *Myotis emarginatus* (mMyoEma1) specimen used for genome sequencing.

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.

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 mMyoEma1 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

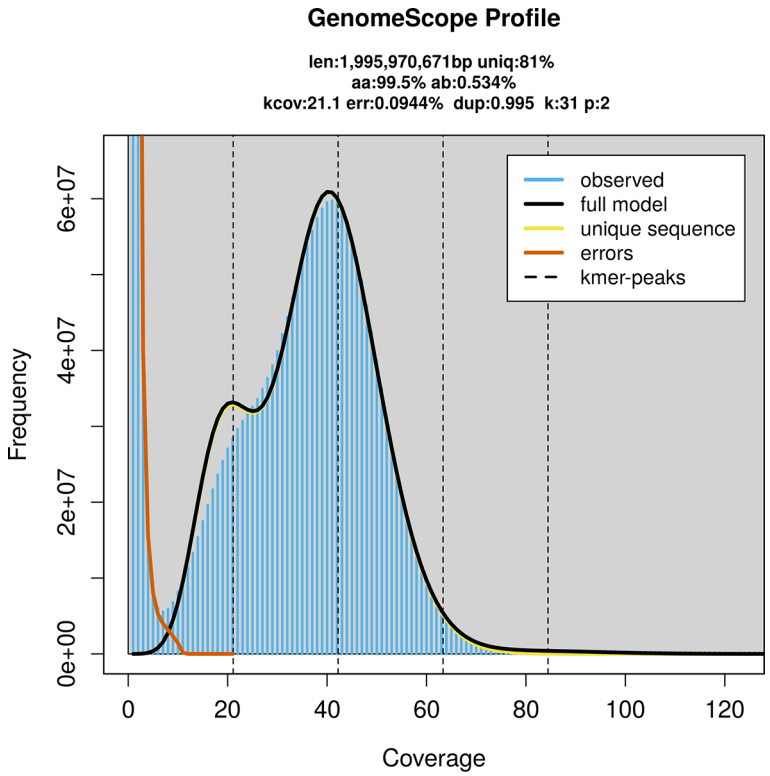

Figure 2. Frequency distribution of *k*-mers generated using GenomeScope2. The plot shows observed and modelled *k*-mer spectra, providing estimates of genome size, heterozygosity, and repeat content based on unassembled sequencing reads.

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 18 breaks and 82 joins. This reduced the scaffold count by 10.8% and increased the scaffold N50 by 1.5%. 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 MerquryFK 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.

Table 1. Specimen and sequencing data for *Myotis emarginatus* (BioProject PRJEB81650).

Platform	PacBio HiFi	Hi-C	RNA-seq
ToLID	mMyoEma1	mMyoEma1	mMyoEma1
Specimen ID	SAN00003684	SAN00003684	SAN00003684
BioSample (source individual)	SAMEA115534660	SAMEA115534660	SAMEA115534660
BioSample (tissue)	SAMEA115534662	SAMEA115534662	SAMEA115534665
Tissue	heart	heart	spleen
Instrument	Revio	Illumina NovaSeq X	Illumina NovaSeq X
Run accessions	ERR13900463	ERR13907247	ERR15140912
Read count total	6.74 million	1 058.16 million	108.83 million
Base count total	86.98 Gb	159.78 Gb	16.43 Gb

Table 2. Genome assembly statistics for *Myotis emarginatus*.

Genome assembly	Haplotype 1	Haplotype 2
Assembly name	mMyoEma1.hap1.2	mMyoEma1.hap2.2
Assembly accession	GCA_965115925.2	GCA_965115945.2
Assembly level	chromosome	chromosome
Span (Mb)	2 110.52	2 084.88
Number of chromosomes	22	22
Number of contigs	531	682
Contig N50	48.68 Mb	35.85 Mb
Number of scaffolds	305	418
Scaffold N50	99.61 Mb	96.7 Mb
Longest scaffold length (Mb)	228.73	229.98
Sex chromosomes	X	X
Organelles	Mitochondrion: 17.1 kb	-

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 *Myotis emarginatus* specimen generated 86.98 Gb (gigabases) from 6.74 million reads, which were used to assemble the genome. GenomeScope2.0 analysis estimated the haploid genome size at 1 995.97 Mb, with a heterozygosity of 0.53% and repeat content of 19.05% (Figure 2). These estimates guided expectations for the assembly. Based on the estimated genome size, the sequencing data provided approximately 42× coverage. Hi-C sequencing produced 159.78 Gb from 1 058.16 million reads, which were used to scaffold the assembly. RNA sequencing data were

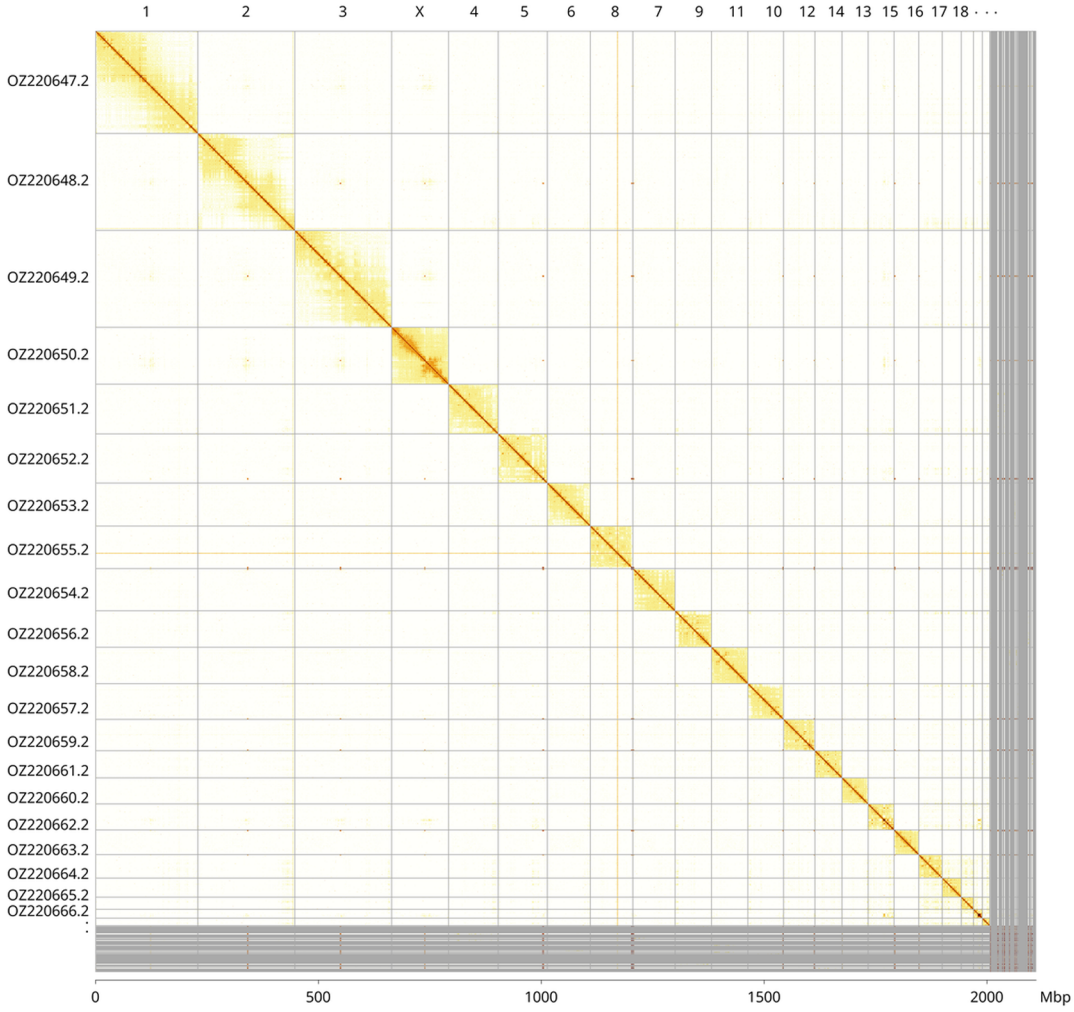

Figure 3. Hi-C contact map of the *Myotis emarginatus* genome assembly. Assembled chromosomes are shown in order of size and labelled along the axes, with a megabase scale shown below. The plot was generated using PretextSnapshot.

Table 3. Chromosomal pseudomolecules in both haplotypes of the genome assembly of *Myotis emarginatus*, mMyoEma1.

Haplotype 1				Haplotype 2			
INSDC accession	**Name**	**Length (Mb)**	**GC%**	**INSDC accession**	**Name**	**Length (Mb)**	**GC%**
OZ220647.2	1	228.73	41.50	OZ251378.1	1	229.98	41.50
OZ220648.2	2	217.92	42	OZ251379.1	2	213.61	42
OZ220649.2	3	217.38	41	OZ251380.1	3	212.67	41
OZ220651.2	4	116.24	42	OZ251381.1	4	115.11	42
OZ220652.2	5	109.27	44	OZ251382.1	5	109.92	43.50
OZ220653.2	6	99.61	41	OZ251383.1	6	96.70	41.50
OZ220654.2	7	95.97	43	OZ251385.1	7	94.06	43.50
OZ220655.2	8	98.24	41	OZ251384.1	8	96.53	41
OZ220656.2	9	86.54	43.50	OZ251386.1	9	78.93	43
OZ220657.2	10	82.15	42	OZ251388.1	10	80.90	41.50
OZ220658.2	11	82.92	43	OZ251387.1	11	80.58	42
OZ220659.2	12	68.35	43.50	OZ251389.1	12	67.51	43.50
OZ220660.2	13	60.78	44.50	OZ251391.1	13	58.75	44
OZ220661.2	14	61.24	43.50	OZ251390.1	14	58.20	44
OZ220662.2	15	59.17	47	OZ251392.1	15	55.74	46.50
OZ220663.2	16	53.98	42.50	OZ251393.1	16	52.68	42
OZ220664.2	17	52.94	47	OZ251394.1	17	48.68	47
OZ220665.2	18	42.90	46	OZ251395.1	18	43.61	46
OZ220666.2	19	28.21	49	OZ251396.1	19	27.60	49.50
OZ220667.2	20	19.86	48	OZ251397.1	20	17.15	48
OZ220668.2	21	19.79	45.50	OZ251398.1	21	16.86	47
OZ220650.2	X	127.72	41	OZ251399.1	X	124.37	40.50

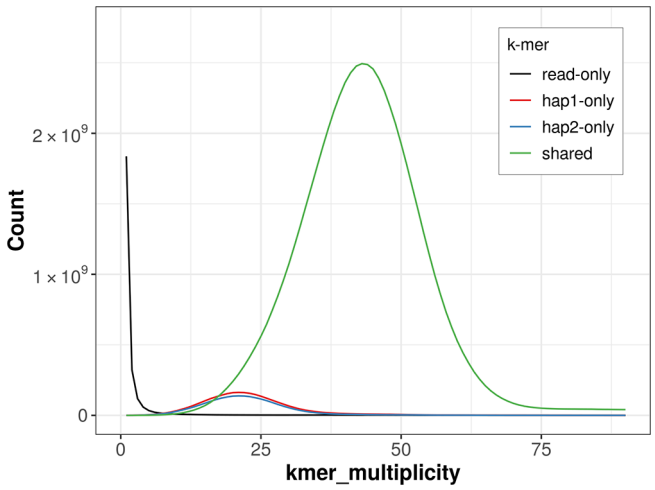

Figure 4. Evaluation of *k*-mer completeness using MerquryFK. This plot illustrates the recovery of *k*-mers from the original read data in the final assemblies. The horizontal axis represents *k*-mer multiplicity, and the vertical axis shows the number of *k*-mers. The black curve represents *k*-mers 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.

also generated and are available in public sequence repositories. Table 1 summarises the specimen and sequencing details.

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 2 110.52 Mb in 305 scaffolds, with 226 gaps, and a scaffold N50 of 99.61 Mb (Table 2).

Most of the haplotype 1 assembly sequence (96.18%) 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). Chromosome X was identified through synteny analysis with *Myotis daubentonii* (GCA_963259705.1). The exact order and orientation of the contigs in the following regions are

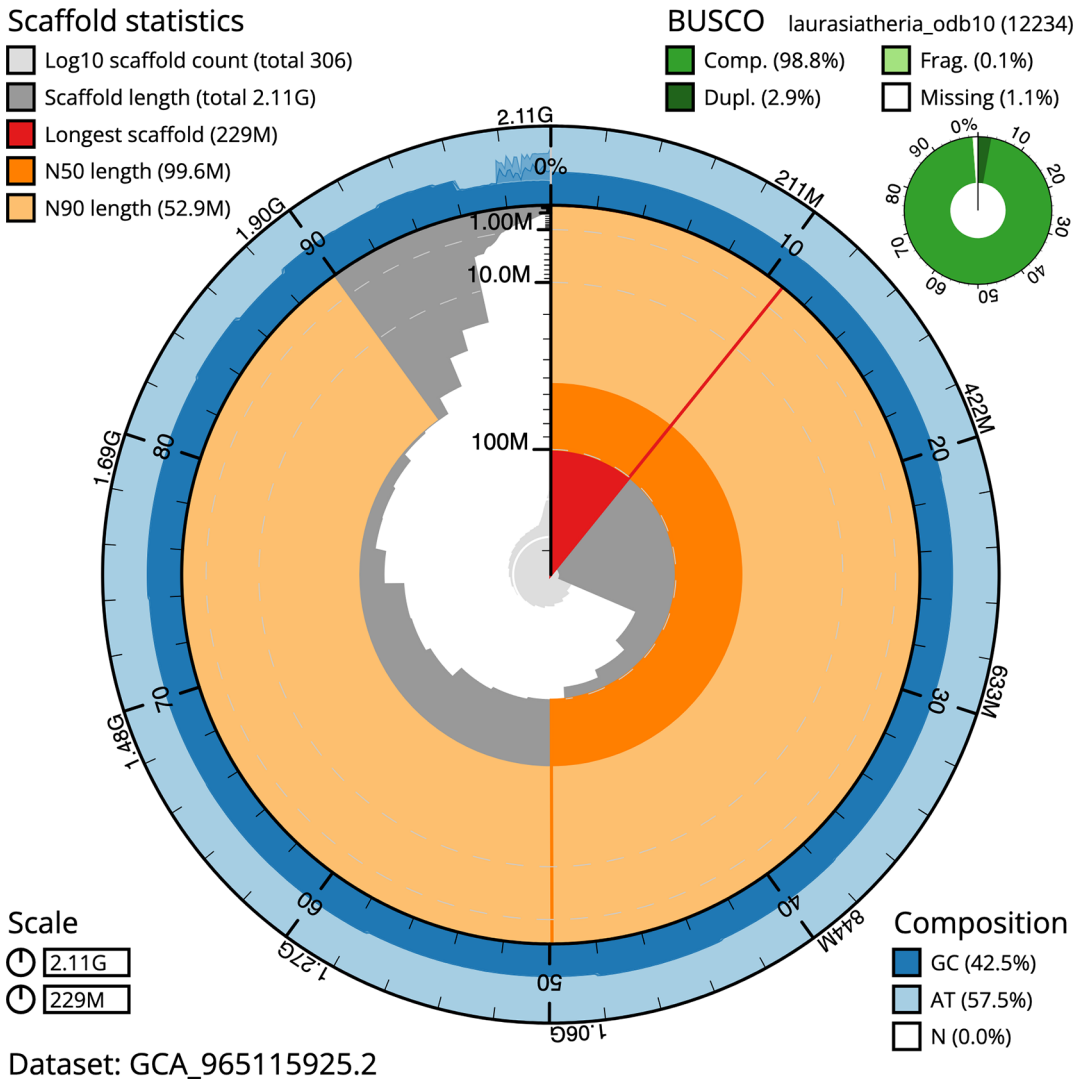

Figure 5. Assembly metrics for mMyoEma1.hap1.2. The BlobToolKit snail plot provides an overview of assembly metrics and BUSCO gene completeness. The circumference represents the length of the whole genome sequence, and the main plot is divided into 1 000 bins around the circumference. The outermost blue 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.

unknown: chromosome 1 (99 058–104 530 kbp), chromosome 6 (97 067–208.5 bp), and chromosome 1 (91.5–102 953 kbp).

The mitochondrial genome was also assembled (length 17.1 kb, OZ220669.2). This sequence is included as a contig in the multifasta file of the genome submission and as a standalone record.

Assembly quality metrics

For haplotype 1, the estimated QV is 59.2, and for haplotype 2, 59.2. When the two haplotypes are combined, the assembly achieves an estimated QV of 59.2. The *k*-mer completeness is 93.39% for haplotype 1, 92.20% for haplotype 2, and 99.50% for the combined haplotypes (Figure 4).

BUSCO analysis using the laurasiatheria_odb10 reference set (n = 12 234) identified 98.8% of the expected gene set (single = 95.9%, duplicated = 2.9%) in haplotype 1. For haplotype 2, BUSCO v.6.0.0 analysis identified 96.5% of the expected gene set (single = 93.5%, duplicated = 3.1%). The snail plot in Figure 5 summarises the scaffold length

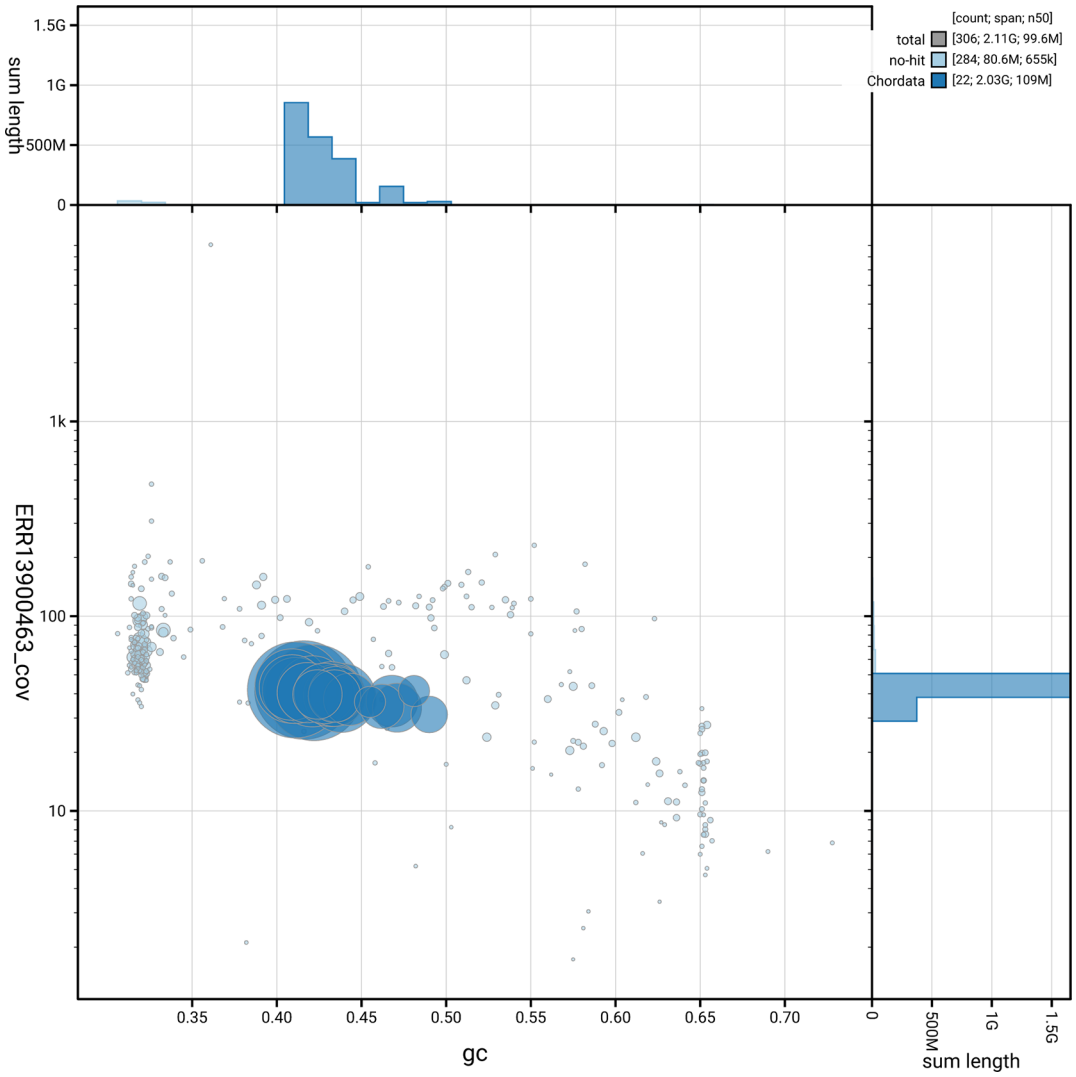

Figure 6. BlobToolKit blob plot for mMyoEma1.hap1.2. The plot shows base coverage (vertical axis) and GC content (horizontal axis). The circles represent scaffolds, with the size proportional to scaffold length and the colour representing phylum membership. The histograms along the axes display the total 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 *Myotis emarginatus* assembly.**

Measure	Value	Benchmark
EBP summary (haplotype 1)	7.C.Q59	6.C.Q40
Contig N50 length	48.68 Mb	≥ 1 Mb
Scaffold N50 length	99.61 Mb	= chromosome N50
Consensus quality (QV)	Haplotype 1: 59.2; haplotype 2: 59.2; combined: 59.2	≥ 40
k-mer completeness	Haplotype 1: 93.39%; Haplotype 2: 92.20%; combined: 99.50%	≥ 95%
BUSCO	C:98.8% [S:95.9%, D:2.9%], F:0.1%, M:1.1%, n:12 234	S > 90%; D < 5%
Percentage of assembly assigned to chromosomes	96.18%	≥ 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 used for *Myotis emarginatus*.**

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	1.2a.2	https://github.com/c-zhou/yahs

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.

Table 4 lists the assembly metric benchmarks adapted from Rhie *et al.* (2021) and the Earth BioGenome Project Report on Assembly Standards September 2024. The EBP metric, calculated for the haplotype 1, is **7.C.Q59**, meeting the recommended reference standard.

Data availability

European Nucleotide Archive: Myotis emarginatus (Geoffroy's bat). Accession number PRJEB81650. The genome sequence is released openly for reuse. The *Myotis emarginatus* genome sequencing initiative is part of the Darwin Tree of Life Project (PRJEB40665), the Sanger Institute Tree of Life Programme (PRJEB43745), 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.

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.

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

Csorba G, Chou C-H, Ruedi M, *et al.*: **The reds and the yellows: A review of Asian *Chrysopteron* jentink, 1910 (Chiroptera: Vespertilionidae: *Myotis*).** *J. Mammal.* 2014; **95**(4): 663–678.
Publisher 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

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

Foley NM, Harris AJ, Bredemeyer KR, *et al.*: **Karyotypic stasis and swarming influenced the evolution of viral tolerance in a species-rich bat radiation.** *Cell Genomics.* 2024; **4**(2): 100482.
PubMed Abstract | Publisher Full Text | Free 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): 10.5555/2600239.2600241.

Morales AE, Burbrink FT, Segall M, *et al.*: **Distinct genes with similar functions underlie convergent evolution in *Myotis* bat ecomorphs.** *Mol. Biol. Evol.* 2024; **41**(9): msae165.
PubMed Abstract | Publisher Full Text | Free Full Text

Ranallo-Benavidez TR, Jaron KS, Schatz MC: **GenomeScope 2.0 and Smudgeplot for reference-free profiling of polyploid genomes.** ***Nature.*** *Communications.* 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

Ruedi M, Manzinalli J, Dietrich A, *et al.*: **Shortcomings of DNA barcodes: A perspective from the mammal fauna of Switzerland.** *Hystrix, the Italian Journal of Mammalogy.* 2023; **34**(1): 54–61.
Publisher Full Text

Schoch CL, Ciufo S, Domrachev M, *et al.*: **NCBI taxonomy: A comprehensive update on curation, resources and tools.** *Database.* 2020; **2020**, **2020**: baaa062.
PubMed Abstract | Publisher Full Text | Free 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 05 June 2026

https://doi.org/10.21956/wellcomeopenres.29173.r154838

?

Michael R Buchalski
Wildlife Health Laboratory, California Department of Fish and Wildlife, Sacramento, CA, USA
Samantha Capel
California Department of Fish and Wildlife, Sacramento, Sacramento, California, USA

This article presents a reference genome assembly for *Myotis emarginatus*, detailing the sequencing and assembly methodology and resultant assembly quality. All key methodology and results are present, but there are a few details missing:

1. Tissue type from which RNA was extracted is not specified in-text within the "RNA library preparation and sequencing" Methods subsection

2. Scaffold and contig L50 scores should be included in Table 2

3. Why wasn't the RNA seq data used to annotate the genome? You may want to specify as readers will be curious.

4. You claim to have a chromosome-level assembly but make no reference to known Myotis karyotypes and how they relate to your final assembly structure

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: Population genomics, conservation genomics, bat genomics

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

Reviewer Report 01 June 2026

https://doi.org/10.21956/wellcomeopenres.29173.r154837

Keping Sun
Northeast Normal University, Changchun, Jilin, China

In the Introduction section, please add relevant literature to support the description of the diet of this species.

In the sample acquisition section, please indicate which organs were collected.

For the species identification, the authors state: “The specimen was identified by Manuel Ruedi based on morphological and mitochondrial DNA characters [GenBank OQ706656 and OQ885399; Ruedi et al. (2023)].” Please specify which mitochondrial gene(s) were used, or whether the entire mitochondrial genome was sequenced.

For the DNA extraction, please indicate which organs were used for DNA extraction. For the RNA extraction, please clarify why RNA was extracted from spleen tissue.

References with missing page numbers:

Danecek P, Bonfield JK, Liddle J, *et al.*: Twelve years of SAMtools and BCFtools. *GigaScience.* 2021; **10** (2).
Howe K, Chow W, Collins J, *et al.*: Significantly improving the quality of genome assemblies through curation. *GigaScience.* 2021; **10**(1)
Merkel D: Docker: Lightweight Linux containers for consistent development and deployment. *Linux J.* 2014; **2014**(239): 10.5555/2600239.2600241.
Rhie A, Walenz BP, Koren S, *et al.*: Merqury: Reference-free quality, completeness, and phasing assessment for genome assemblies. *Genome Biol.* 2020; **21**(1).
Zhou C, McCarthy SA, Durbin R: YaHS: Yet another Hi-C scaffolding tool. *Bioinformatics.* 2023; **39**(1).

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.

Reviewer Report 27 May 2026

https://doi.org/10.21956/wellcomeopenres.29173.r154832

Heliana Dundarova
Institute of Biodiversity and Ecosystem Research, Bulgarian Academy of Sciences, Sofia, Bulgaria

This data note presents a chromosome-level genome assembly of *Myotis emarginatus* generated within the framework of the Darwin Tree of Life Project. Overall, this is a very solid and well-prepared genome note that provides an important genomic resource for future studies on bat evolution, comparative genomics, ecology, and conservation.

It is clearly written, the methodology is appropriate and described in sufficient detail, and the assembly statistics demonstrate a high-quality genome. The use of PacBio HiFi sequencing combined with Hi-C scaffolding and manual curation resulted in an assembly with excellent continuity and completeness.

The manuscript follows current standards for genome resource publications and will be useful to a broad community of researchers working on bats and vertebrate genomics.

The background section is concise but informative, especially regarding the ecology and phylogenetic position of *M. emarginatus*. The discussion of the species as the only European representative of the African clade of *Myotis* adds evolutionary relevance to the study.

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: Ecology, evolution, and conservation of Eurasian bats, with particular focus on the role of bats as hosts and vectors of RNA viruses.

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.