



DATA NOTE

The genome sequence of the Little Owl, *Athene noctua* (Scopoli, 1769) (Strigiformes: Strigidae)

[version 1; peer review: 2 approved]

Elizabeth Blows¹, Irene Fabiola Roman Maldonado ²,
Natural History Museum Genome Acquisition Lab,
Darwin Tree of Life Barcoding Collective,
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

¹The Raptor Foundation, Woodhurst, Cambridgeshire, England, UK
²Wellcome Sanger Institute, Hinxton, Cambridgeshire, England, UK

V1 First published: 08 Sep 2025, 10:492
<https://doi.org/10.12688/wellcomeopenres.24873.1>
Latest published: 08 Sep 2025, 10:492
<https://doi.org/10.12688/wellcomeopenres.24873.1>

Abstract
We present a genome assembly from an individual male *Athene noctua* (Little Owl; Chordata; Aves; Strigiformes; Strigidae). The assembly contains two haplotypes with total lengths of 1 355.86 megabases and 1 302.38 megabases. Most of haplotype 1 (91.82%) is scaffolded into 41 chromosomal pseudomolecules, including the Z sex chromosome. Haplotype 2 was assembled to scaffold level. The mitochondrial genome has also been assembled, with a length of 18.17 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
Athene noctua; Little Owl; genome sequence; chromosomal; Strigiformes



This article is included in the [Tree of Life](#) gateway.

Open Peer Review

Approval Status  

	1	2
version 1		
08 Sep 2025	view	view

1. **Marija Tanaskovic**, University of Belgrade, Beograd, Serbia
2. **Rauri C.K. Bowie** , University of California Berkeley, Berkeley, USA

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

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

Author roles: Blows E: Investigation, Resources; Roman Maldonado IF: Writing – Original Draft Preparation;

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>].

The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Copyright: © 2025 Blows E *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: Blows E, Roman Maldonado IF, Natural History Museum Genome Acquisition Lab *et al.* **The genome sequence of the Little Owl, *Athene noctua* (Scopoli, 1769) (Strigiformes: Strigidae) [version 1; peer review: 2 approved]** Wellcome Open Research 2025, 10:492 <https://doi.org/10.12688/wellcomeopenres.24873.1>

First published: 08 Sep 2025, 10:492 <https://doi.org/10.12688/wellcomeopenres.24873.1>

Species taxonomy

Eukaryota; Opisthokonta; Metazoa; Eumetazoa; Bilateria; Deuterostomia; Chordata; Craniata; Vertebrata; Gnathostomata; Teleostomi; Euteleostomi; Sarcopterygii; Dipnotetrapodomorpha; Tetrapoda; Amniota; Sauropsida; Sauria; Archelosauria; Archosauria; Dinosauria; Saurischia; Theropoda; Coelurosauria; Aves; Neognathae; Neoaves; Telluraves; Strigiformes; Strigidae; *Athene*; *Athene noctua* (Scopoli, 1769) (NCBI:txid126797)

Background

Athene noctua (also known as the Little Owl) is a small, relatively long-legged nocturnal owl with a short tail and yellow eyes. It is found across Europe, Asia and North Africa (GBIF Secretariat, 2023). Introduced to the United Kingdom in the 19th century, it has since spread across much of England and Wales.

This species of owl is strongly associated with open agricultural landscapes. In northern and central Europe, population sizes are declining rapidly, likely due to environmental modification, habitat loss, degradation, and pollution (Van Nieuwenhuysen *et al.*, 2003).

A. noctua is a carnivorous species. Its diet includes insects, earthworms and small vertebrates (amphibians, reptiles, birds and mammals). The mammals they prey on include mice, rats, voles, shrews, moles and rabbits (Šálek *et al.*, 2016). The species is monogamous, breeding pairs remain together for at least one breeding season (Génot *et al.*, 1997). During the courtship period, the male owl defends his territory and calls to attract a partner. Once a potential mate has been found, the pair bond begins. After successful mating, the female remains in the nest and increasingly ‘begs’ her partner for food. She lays a clutch of three to five eggs (occasionally only two, or as many as eight). After the young have fledged, the bond between the pair weakens and they become more aggressive towards each other. Often, the pair separates at this point, although not always (Johnson *et al.*, 2009; Van Nieuwenhuysen *et al.*, 2008).

Due to the severe decline in the population of little owls, their protection should be treated as a high conservation priority (Šálek & Lövy, 2012). Analysis of the mortality of adult little owls shows two distinct peaks: a winter peak associated with severe winters involving long-standing snow cover, and a summer peak connected with the exhaustion of birds after the breeding season (Šálek & Lövy, 2012). *Athene noctua* was most recently assessed for the IUCN Red List of Threatened Species in 2018, and is listed as “Least Concern” due to its large population size, stable population trend and large range (BirdLife International, 2018).

In addition to the complete genome submitted here, the Bird 10 000 Genomes Project (B10K) generated a scaffold-level genome of the little owl, thereby enriching our genetic knowledge of this species (data obtained via NCBI datasets, O’Leary *et al.*, 2024). The additional sequencing work presented here allows us to refine the existing genome assembly, expand our understanding of biodiversity, and improve genomic technologies.

The access to a complete genome of *Athene noctua* will provide a broader set of tools to understand its adaptations to the multiple regions it inhabits. This will help us find ways to protect crops and learn more about the ecosystems we all depend on. In addition, genomic information is important to understand its evolutionary history, adaptations to a nocturnal lifestyle, and for informing conservation efforts.

We present a chromosome-level genome sequence for *Athene noctua*, generated using the Tree of Life pipeline from a specimen collected in Broad Balk, Ely, Stretham, United Kingdom (Figure 1).

Methods

Sample acquisition

The specimen used for genome sequencing was an adult male *Athene noctua* (specimen ID NHMUK015060570, ToLID bAthNoc1; Figure 1), collected from Broad Balk, Ely, Stretham, United Kingdom (latitude 52.35, longitude 0.224) on 2021-07-09. The specimen was found dead, and formally identified by Liz Blows (Raptor Foundation). The same specimen was used for RNA sequencing. For the Darwin Tree of Life sampling and metadata approach, refer to Lawniczak *et al.* (2022).

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 bAthNoc1 sample was weighed and triaged to determine the appropriate extraction protocol. Heart tissue was homogenised by cryogenic disruption using the Covaris cryoPREP® Automated Dry Pulverizer. HMW DNA was extracted using the Automated MagAttract v2 protocol. DNA was sheared into an



Figure 1. Photograph of the *Athene noctua* by Arturo Nikolai (not the specimen used for genome sequencing).

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 23.6 ng/μL and a yield of 3 068.00 ng. The 260/280 spectrophotometric ratio was 1.84, and the 260/230 ratio was 2.64.

RNA was also extracted from heart tissue of bAthNoc1 in the Tree of Life Laboratory at the WSI using the [RNA Extraction: Automated MagMax™ mirVana](#) 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.

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 25M 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 of the bAthNoc1 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 to 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 and 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 to generate 150-bp 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](#); [Cheng et al., 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 mis-assemblies. The scaffolded assemblies were evaluated using Gfastats ([Formenti et al., 2022](#)), BUSCO ([Manni et al., 2021](#)) and MERQURY.FK ([Rhie et al., 2020](#)).

The mitochondrial genome was assembled using MitoHiFi (Uliano-Silva *et al.*, 2023), which runs MitoFinder (Allio *et al.*, 2020) and uses these annotations to select the final mitochondrial contig and to ensure the general quality of the sequence.

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. MicroFinder (Mathers *et al.*, 2025) was used to order scaffolds prior to 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 19 breaks and 162 joins. The curation process is documented at <https://gitlab.com/wtsi-grit/rapid-curation>. PretextViewSnapshot 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 *Athene noctua* specimen generated 72.21 Gb (gigabases) from 6.79 million reads, which were used to assemble the genome. GenomeScope2.0 analysis estimated the haploid genome size at 1 287.60 Mb, with a heterozygosity of 0.20% and repeat content of 12.32% (Figure 2). These estimates guided expectations for the assembly. Based on the estimated genome size, the sequencing data provided approximately 55× coverage. Hi-C sequencing produced 167.90 Gb from 1 111.92 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.

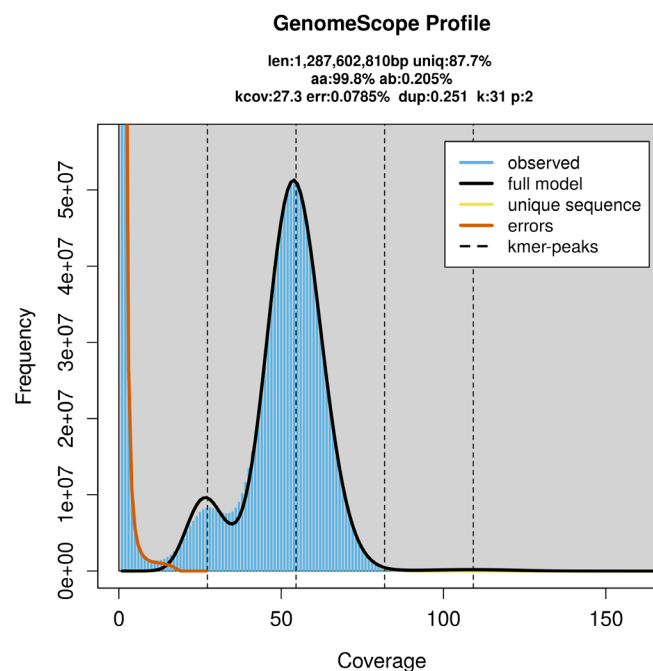


Figure 2. Frequency distribution of *k*-mers generated using GenomeScope2. The plot shows observed and modelled *k*-mer spectra, providing estimates of genome size, heterozygosity, and repeat content based on unassembled sequencing reads.

Table 1. Specimen and sequencing data for BioProject PRJEB74980.

Platform	PacBio HiFi	Hi-C	RNA-seq
ToLID	bAthNoc1	bAthNoc1	bAthNoc1
Specimen ID	NHМУK015060570	NHМУK015060570	NHМУK015060570
BioSample (source individual)	SAMEA114594456	SAMEA114594456	SAMEA114594456
BioSample (tissue)	SAMEA114594473	SAMEA114594473	SAMEA114594473
Tissue	heart	heart	heart
Instrument	Revio	Illumina NovaSeq X	Illumina NovaSeq X
Run accessions	ERR12921322	ERR12945476	ERR13493945
Read count total	6.79 million	1 111.92 million	86.97 million
Base count total	72.21 Gb	167.90 Gb	13.13 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 1 355.86 Mb in 716 scaffolds, with 572 gaps, and a scaffold N50 of 86.33 Mb (Table 2).

Most of the assembly sequence (91.82%) was assigned to 41 chromosomal-level scaffolds, representing 40 autosomes and the Z sex chromosome. These chromosome-level scaffolds, confirmed by Hi-C data, are named according to size (Figure 3; Table 3). The Z chromosome was identified based on alignment with the genome assembly GCA_024206055.2 (*Gallus gallus* GGSWU).

The mitochondrial genome was also assembled. This sequence is included as a contig in the multifasta file of the genome submission and as a standalone record.

For haplotype 1, the estimated QV is 62.2, and for haplotype 2, 62.7. When the two haplotypes are combined, the assembly achieves an estimated QV of 62.4. The *k*-mer completeness is 95.74% for haplotype 1, 95.58% for haplotype 2, and 99.80% for the combined haplotypes (Figure 4).

BUSCO analysis using the aves_odb10 reference set ($n = 8\,338$) identified 97.3% of the expected gene set (single = 96.9%, duplicated = 0.4%) for haplotype 1. 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.

Table 4 lists the assembly metric benchmarks adapted from Rhie *et al.* (2021) the Earth BioGenome Project Report on Assembly Standards September 2024. The EBP metric, calculated for the haplotype 1, is **6.C.Q62**, meeting the recommended reference standard.

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

Table 2. Genome assembly statistics.

Assembly name	bAthNoc1.hap1.1	bAthNoc1.hap2.1
Assembly accession	GCA_965140245.1	GCA_965140185.1
Assembly level	chromosome	scaffold
Span (Mb)	1 355.86	1 302.38
Number of chromosomes	41	N/A
Number of contigs	1 288	1 212
Contig N50	4.05 Mb	3.94 Mb
Number of scaffolds	716	660
Scaffold N50	86.33 Mb	86.09 Mb
Longest scaffold length (Mb)	265.81	N/A
Sex chromosomes	Z	N/A
Organelles	Mitochondrion: 18.17 kb	N/A

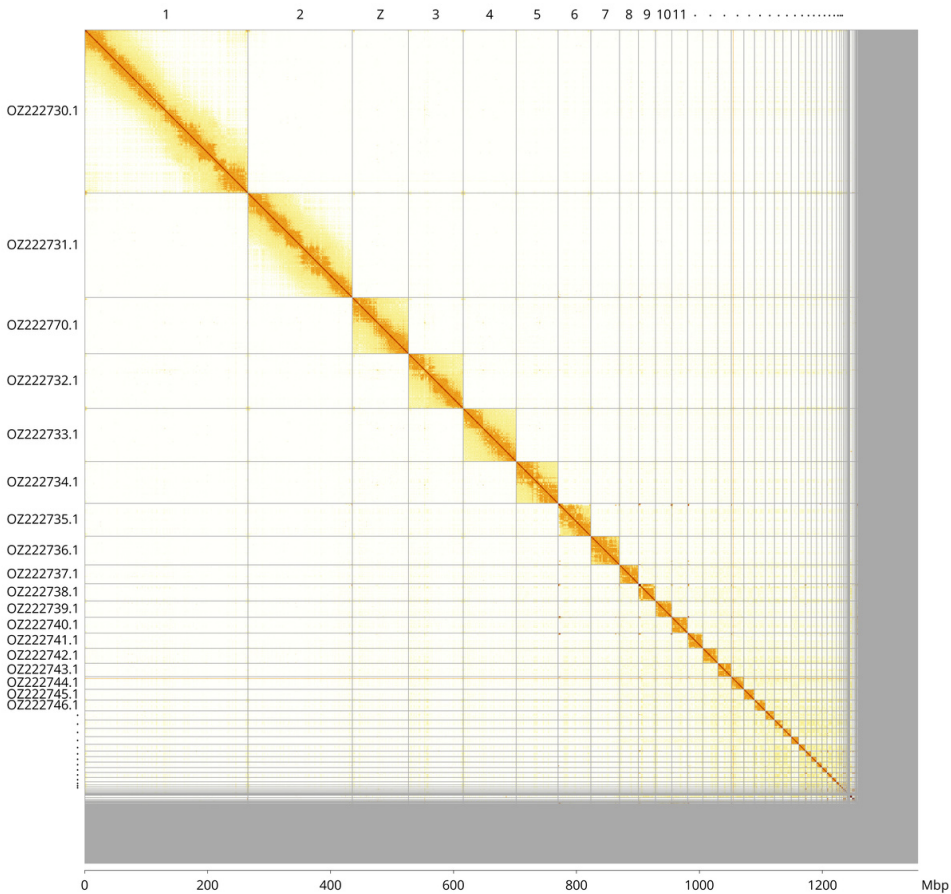


Figure 3. Hi-C contact map of the *Athene noctua* genome assembly. Assembled chromosomes are shown in order of size and labelled along the axes, with a megabase scale shown below. The plot was generated using PretextSnapshot.

Table 3. Chromosomal pseudomolecules in the haplotype 1 genome assembly of *Athene noctua* bAthNoc1.

INSDC accession	Molecule	Length (Mb)	GC%
OZ222730.1	1	265.81	40.50
OZ222731.1	2	169.93	40
OZ222732.1	3	88.85	41
OZ222733.1	4	86.33	40
OZ222734.1	5	68.16	42
OZ222735.1	6	53.35	42.50
OZ222736.1	7	46.44	41.50
OZ222737.1	8	30.66	42.50
OZ222738.1	9	28.06	43.50
OZ222739.1	10	26.49	44.50
OZ222740.1	11	25.61	43.50
OZ222741.1	12	24.72	46
OZ222742.1	13	24.51	43
OZ222743.1	14	21.79	42.50
OZ222744.1	15	20.54	45.50
OZ222745.1	16	17.50	45.50
OZ222746.1	17	17.37	45.50
OZ222747.1	18	14.91	47
OZ222748.1	19	13.75	47

INSDC accession	Molecule	Length (Mb)	GC%
OZ222749.1	20	13.42	47.50
OZ222750.1	21	12.58	42.50
OZ222751.1	22	10.79	47.50
OZ222752.1	23	9.29	51
OZ222753.1	24	8.84	51
OZ222754.1	25	8.52	52.50
OZ222755.1	26	8.37	49.50
OZ222756.1	27	8.18	53
OZ222757.1	28	7.08	49.50
OZ222758.1	29	4.64	56.50
OZ222759.1	30	3.39	59
OZ222760.1	31	3.10	57
OZ222761.1	32	2.15	56
OZ222762.1	33	2.08	60
OZ222763.1	34	1.68	59.50
OZ222764.1	35	1.35	55.50
OZ222765.1	36	1.17	56.50
OZ222766.1	37	0.75	59.50
OZ222767.1	38	0.73	51
OZ222768.1	39	0.57	61
OZ222769.1	40	0.26	61.50
OZ222770.1	Z	91.26	41

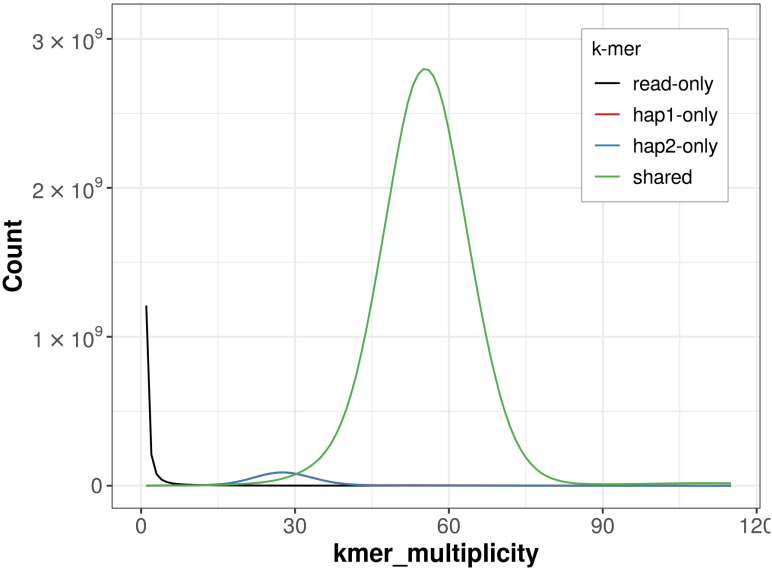


Figure 4. Evaluation of *k*-mer completeness using MerquryFK. This plot illustrates the recovery of *k*-mers from the original read data in the final assemblies. The horizontal axis represents *k*-mer multiplicity, and the vertical axis shows the number of *k*-mers. The black curve represents *k*-mers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

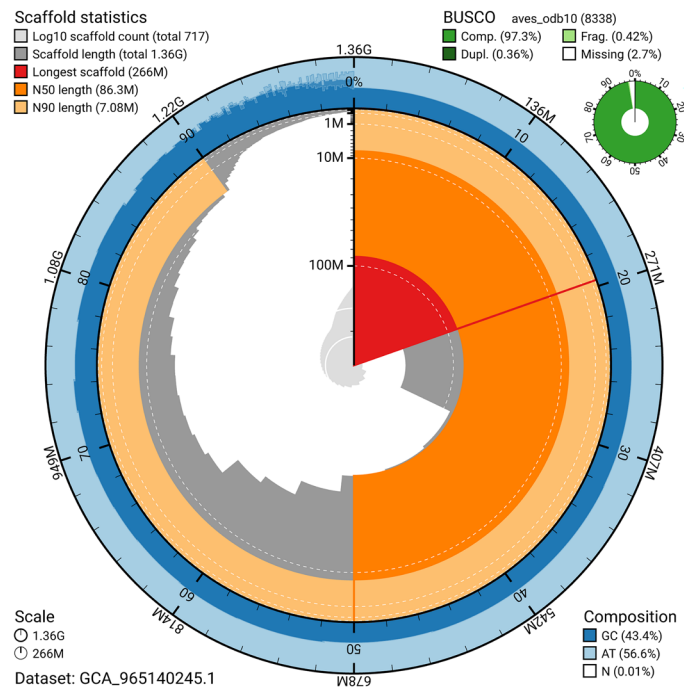


Figure 5. Assembly metrics for bAthNoc1.hap1.1. The BlobToolKit snail plot provides an overview of assembly metrics and BUSCO gene completeness. The circumference represents the length of the whole genome sequence, and the main plot is divided into 1 000 bins around the circumference. The outermost blue 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](#).

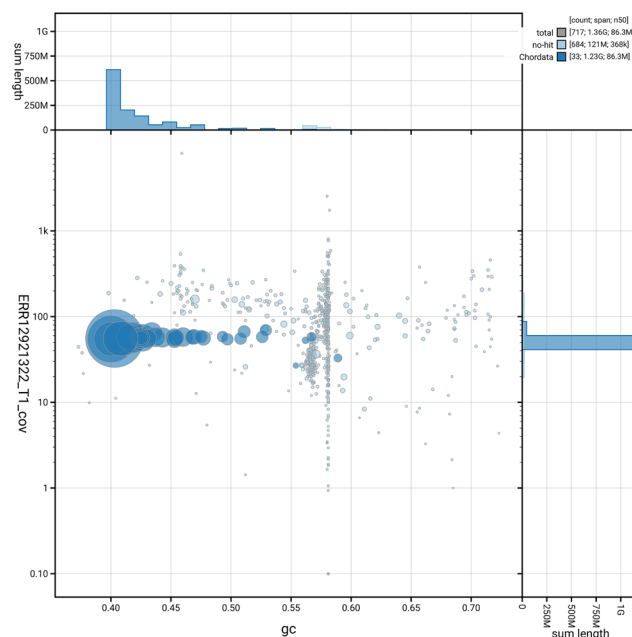


Figure 6. BlobToolKit GC-coverage plot for bAthNoc1.hap1.1. Blob plot showing sequence coverage (vertical axis) and GC content (horizontal axis). The circles represent scaffolds, with the size proportional to scaffold length and the colour representing phylum membership. The histograms along the axes 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 *Athene noctua* assembly.

Measure	Value	Benchmark
EBP summary (haplotype 1)	6.C.Q62	6.C.Q40
Contig N50 length	4.05 Mb	≥ 1 Mb
Scaffold N50 length	86.33 Mb	= chromosome N50
Consensus quality (QV)	Haplotype 1: 62.2; haplotype 2: 62.7; combined: 62.4	≥ 40
<i>k</i> -mer completeness	Haplotype 1: 95.74%; Haplotype 2: 95.58%; combined: 99.80%	≥ 95%
BUSCO	C:97.3% [S:96.9%; D:0.4%]; F:0.4%; M:2.3%; n:8 338	S > 90%; D < 5%
Percentage of assembly assigned to chromosomes	91.82%	≥ 90%

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: *Athene noctua*. Accession number [PRJEB74980](#). The genome sequence is released openly for reuse. The *Athene noctua* genome sequencing initiative is part of the Darwin Tree of Life Project (PRJEB40665), the Sanger Institute Tree of Life Programme (PRJEB43745) and Vertebrate Genomes Project (PRJNA489243). 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 [Table 1](#) and [Table 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.

Author information

Contributors are listed at the following links:

- Members of the [Natural History Museum Genome Acquisition Lab](#)
- 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](#)

Table 5. Software versions and sources.

Software	Version	Source
BEDTools	2.30.0	https://github.com/arq5x/bedtools2
BLAST	2.14.0	ftp://ftp.ncbi.nlm.nih.gov/blast/executables/blast/
BlobToolKit	4.3.9	https://github.com/blobtoolkit/blobtoolkit
BUSCO	5.5.0	https://gitlab.com/ezlab/busco
bwa-mem2	2.2.1	https://github.com/bwa-mem2/bwa-mem2

Software	Version	Source
Cooler	0.8.11	https://github.com/open2c/cooler
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
GoaT CLI	0.2.5	https://github.com/genomehubs/goat-cli
Hifiasm	0.19.8-r603	https://github.com/chhylp123/hifiasm
HiGlass	1.13.4	https://github.com/higlass/higlass
MercuryFK	1.1.2	https://github.com/thegenemyers/MERQURY.FK
Minimap2	2.24-r1122	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	23.10.0	https://github.com/nextflow-io/nextflow
PretextViewSnapshot	N/A	https://github.com/sanger-tol/PretextViewSnapshot
PretextView	0.2.5	https://github.com/sanger-tol/PretextView
samtools	1.19.2	https://github.com/samtools/samtools
sanger-tol/ascc	0.1.0	https://github.com/sanger-tol/ascc
sanger-tol/blobtoolkit	0.6.0	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

References

- Allio R, Schomaker-Bastos A, Romiguier J, *et al.*: **MitoFinder: efficient automated large-scale extraction of mitogenomic data in target enrichment phylogenomics.** *Mol Ecol Resour.* 2020; **20**(4): 892–905.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- 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 MJ, 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](#)
- BirdLife International: **Species factsheet: Little owl *Athene noctua*.** 2018.
[Reference Source](#)
- Buchfink B, Reuter K, Drost HG: **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 (Bethesda).* 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): 1332–35.
[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): giab008.
[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](#)

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](#)

GBIF Secretariat: ***Athene noctua* (Scopoli, 1769).** In: *GBIF Backbone Taxonomy*, 2023.

[Reference Source](#)

Génot JC, Julliard M, Van Nieuwenhuysse D: **Little owl, *Athene noctua*.** In: W. J. M. Hagemeijer and M. J. Blair (eds), *The EBCC atlas of European breeding birds: Their Distribution and Abundance*. London: T & AD Poyser, 1997; 408–9.

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): gaa153.

[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)

Johnson DH, Van Nieuwenhuysse D, Génot JC: **Survey protocol for the Little owl *Athene noctua*.** *Ardea.* 2009; **97**(4): 403–12.

[Publisher 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](#)

Lawniczak MKN, Davey RP, Rajan J, *et al.*: **Specimen and sample metadata standards for biodiversity genomics: a proposal from the Darwin Tree of Life project [version 1; peer review: 2 approved with reservations].** *Wellcome Open Res.* 2022; **7**: 187.

[Publisher 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, Seppely 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](#)

Mathers TC, Paulini M, Sotero-Caio CG, *et al.*: **MicroFinder: conserved gene-set mapping and assembly ordering for manual curation of bird microchromosomes.** *bioRxiv.* 2025.

[Publisher Full Text](#)

Merkel D: **Docker: lightweight Linux containers for consistent development and deployment.** *Linux J.* 2014; **2014**(239): 2.

[Reference Source](#)

O'Leary NA, Cox E, Holmes JB, *et al.*: **Exploring and retrieving sequence and metadata for species across the Tree of Life with NCBI datasets.** *Sci Data.* 2024; **11**(1): 732.

[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.** *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.*: **Mercury: reference-free quality, completeness, and phasing assessment for genome assemblies.** *Genome Biol.* 2020; **21**(1): 245.

[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)

Šálek M, Chrenková M, Dobrý M, *et al.*: **Scale-dependent habitat associations of a rapidly declining farmland predator, the Little owl *Athene noctua*, in contrasting agricultural landscapes.** *Agric Ecosyst Environ.* 2016; **224**: 56–66.

[Publisher Full Text](#)

Šálek M, Lövy M: **Spatial ecology and habitat selection of Little owl *Athene noctua* during the breeding season in Central European farmland.** *Bird Conserv Int.* 2012; **22**(3): 328–38.

[Publisher Full Text](#)

Schoch CL, Ciufo S, Domrachev M, *et al.*: **NCBI Taxonomy: a comprehensive update on curation, resources and tools.** *Database (Oxford).* 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](#)

Van Nieuwenhuysse D, Génot JC, Johnson DH: **The Little owl: conservation, ecology and behavior of *Athene noctua*.** Cambridge, New York: Cambridge University Press, 2008.

[Reference Source](#)

Van Nieuwenhuysse D, van Harxen R, Johnson DH: **The Little owl: population dynamics, behavior and management of *Athene noctua*.** Cambridge University Press, 2023.

[Publisher Full Text](#)

Vasimuddin M, Misra S, Li H, *et al.*: **Efficient architecture-aware acceleration of BWA-MEM for multicore systems.** In: *2019 IEEE International Parallel and Distributed Processing Symposium (IPDPS)*. IEEE, 2019; 314–324.

[Publisher Full Text](#)

Zhou C, McCarthy SA, Durbin R: **YaHS: Yet another Hi-C Scaffolding tool.** *Bioinformatics.* 2023; **39**(1): btac808.

[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)

Open Peer Review

Current Peer Review Status:  

Version 1

Reviewer Report 13 November 2025

<https://doi.org/10.21956/wellcomeopenres.27397.r132129>

© 2025 Bowie 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.



Rauri C.K. Bowie 

University of California Berkeley, Berkeley, USA

This data note presents a high-quality reference genome for the Little Owl. The methods are clearly detailed and meet the standards expected for a genome report. I have only a few minor comments for the authors to consider.

1. **Sample acquisition:** The authors should explicitly state whether a voucher (physical) specimen is available. I assume it is, given that a Natural History Museum number – NHMUK015060570 – is reported, but this could be made clearer. Are additional tissue samples available should someone want to perform future analyses on this individual?
2. **Assembly quality assessment.** What database did the BlobToolKit query for the BUSCO lineages? The GoaT database?
3. **Data availability:** *Athene noctua* should be in italics.

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: Evolutionary Biology, Genomics, Birds

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 30 September 2025

<https://doi.org/10.21956/wellcomeopenres.27397.r132126>

© 2025 Tanaskovic M. 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.



Marija Tanaskovic

University of Belgrade, Beograd, Serbia

The manuscript describes the sequencing, assembly, and curation of a high-quality reference genome for *Athene noctua*. The authors apply state-of-the-art approaches, including PacBio HiFi sequencing, Hi-C scaffolding, and RNA-seq for annotation support. The assembly reaches Earth BioGenome Project (EBP) reference standards (6.C.Q62) with excellent continuity and completeness (scaffold N50 of 86.3 Mb, 97.3% BUSCO completeness, QV > 62). The work is clearly presented and will be of value to the community as an important resource for avian genomics and conservation biology.

The study is technically sound, methodologically detailed, and reproducible. However, there are several areas where additional clarification and standardization of reporting would improve the manuscript's clarity and alignment with current best practices.

Major Comments

1. **Sequencing depth reporting** PacBio coverage is reported (55×), but Hi-C and RNA-seq depth are not explicitly linked to genome size or downstream use. For Hi-C, please provide approximate coverage relative to genome size. For RNA-seq, indicate total yield (e.g., millions of reads per library) to support downstream annotation.
2. **Haplotype handling** The assembly strategy generates two haplotypes, with haplotype 1 curated to chromosome level and haplotype 2 scaffold-level. Please state explicitly which haplotype is considered the reference assembly for submission and downstream use, and how the second haplotype will be provided.
3. **Contamination assessment** Methods describe the use of ASCC and BlobToolKit for contamination screening, but results are not reported. Please summarize contamination screening outcomes in the results (e.g., confirmation of minimal non-target sequences).
4. **Assembly curation** Manual curation involved 19 breaks and 162 joins. Please provide a brief explanation of the types of misassemblies detected (e.g., chimeric joins, collapsed repeats) and the criteria for corrections, as this adds confidence in the final assembly.

5. Minor Comments

- **QC metrics** QC reporting is inconsistent across DNA, RNA, and library preparations. For RNA, integrity values (RIN/RQN) should be reported, as they are provided for DNA purity and fragment size. For libraries, fragment size distributions and molarity should be reported uniformly.
- **Redundancy in phrasing** The phrase "following the manufacturer's instructions" is repeated frequently. This could be streamlined to improve readability (e.g., state once per subsection).
- **Reference consistency** References to Rhie et al. (2021) and EBP standards should be

formatted consistently. Ensure all software tools and pipelines (e.g., Qubit software version, TapeStation kits) are cited with versions

Recommendation Minor revisions The genome assembly is of very high quality and meets the standards for publication as a reference genome. The manuscript requires only modest revisions for clarity, consistency, and completeness in reporting QC metrics, sequencing depth, contamination screening results, and haplotype designation

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 genetics

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.
