



DATA NOTE

The genome sequence of the Eurasian Spoonbill, *Platalea leucorodia* Linnaeus, 1758 (Pelecaniformes: Threskiornithidae)

[version 1; peer review: 3 approved]

Michelle F. O'Brien¹, Rosa Lopez Colom¹,
Natural History Museum Genome Acquisition Lab,
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

¹Wildfowl & Wetlands Trust, Slimbridge, Gloucestershire, England, UK

V1 First published: 15 Sep 2025, 10:511
<https://doi.org/10.12688/wellcomeopenres.24883.1>
Latest published: 15 Sep 2025, 10:511
<https://doi.org/10.12688/wellcomeopenres.24883.1>

Abstract

We present a genome assembly from an individual female *Platalea leucorodia* (Eurasian Spoonbill; Chordata; Aves; Pelecaniformes; Threskiornithidae). The assembly contains two haplotypes with total lengths of 1 345.83 megabases and 1 190.44 megabases. Most of haplotype 1 (95.57%) is scaffolded into 37 chromosomal pseudomolecules, including the W and Z sex chromosomes. Haplotype 2 was assembled to scaffold level. The mitochondrial genome has also been assembled, with a length of 17.15 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

Platalea leucorodia; Eurasian Spoonbill; genome sequence; chromosomal; Pelecaniformes



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

Open Peer Review

Approval Status

	1	2	3
version 1			
15 Sep 2025	view	view	view
1. Subir Shakya , Harvard University, Cambridge, USA			
LMU Munich, Planegg-Martinsreid, Germany			
2. Ricardo Jorge Lopes , Universidade de Lisboa, Lisboa, Portugal			
3. Bengt Hansson , Lund University, Lund, Sweden			

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: **O'Brien MF:** Investigation, Resources, Writing – Original Draft Preparation, Writing – Review & Editing; **Lopez Colom R:** Investigation, Resources, Writing – Original Draft Preparation, 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>].

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

Copyright: © 2025 O'Brien MF *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: O'Brien MF, Lopez Colom R, Natural History Museum Genome Acquisition Lab *et al.* **The genome sequence of the Eurasian Spoonbill, *Platalea leucorodia* Linnaeus, 1758 (Pelecaniformes: Threskiornithidae) [version 1; peer review: 3 approved]** Wellcome Open Research 2025, 10:511 <https://doi.org/10.12688/wellcomeopenres.24883.1>

First published: 15 Sep 2025, 10:511 <https://doi.org/10.12688/wellcomeopenres.24883.1>

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; Aequornithes; Pelecaniformes; Threskiornithidae; *Platalea*; *Platalea leucorodia* Linnaeus, 1758 (NCBI:txid257867)

Background

The Eurasian Spoonbill (*Platalea leucorodia*) is a tall white bird, with a white feather crest on the head and a yellow area at the base of the neck seen in breeding adults. They stand 70–95 cm tall, with a weight of 1130–1960 g and wingspan of 115–135 cm. Males are larger than females in this species and have longer bills and legs (Matheu *et al.*, 2020). These birds reach sexual maturity at 3 to 4 years of age and the oldest ringed bird was reported to be 28 years of age (Matheu *et al.*, 2020).

This species inhabits shallow, large wetlands which include rivers, marshes, larger lakes and flooded areas. It can also be found in estuarine habitats, tidal creeks and lagoons (particularly during the winter months). Nesting takes place on islands in lakes or rivers and nests are made in dense plant matter such as reedbeds, bushes or trees (Matheu *et al.*, 2020). Nests are often colonial and usually 1–2 m apart. There are usually 3 or 4 eggs in a clutch and incubation can be up to 25 days with fledging taking place after 45 to 50 days (Cramp & Simmons, 1977).

Their diet consists of invertebrates, amphibians and fish and occasional algae or plant material. Foraging takes place as the bird wades through water (often in small flocks) and sweeps the bill from side to side. The significance of bill shape has a number of hypotheses including potential reduction in turbulence (Swennen & Yu, 2008) or generating suction (Weihs & Katzir, 2008).

Distribution of the Eurasian spoonbill is relatively complex, with the species breeding across Eurasia and wintering in North Africa, sub-Saharan Africa and Eurasia (Xi *et al.*, 2021). There are currently three subspecies that have been identified: a West African subspecies (*Platalea leucorodia balsaci*), which breeds mainly in Mauritania and islands along the coast (Piersma *et al.*, 2012); a Red Sea subspecies (*P. l. archeri*), which breeds in the Red Sea and along the coast of Somalia; and the Eurasian Spoonbill (*P. l. leucorodia*) which has a wide distribution throughout Eurasia. This third group is the nominate subspecies and consists of four biogeographical populations currently (Wetlands International, 2021). First, a West Europe/West Mediterranean and West Africa population, which breeds along the western coast of Europe and winters in the western Mediterranean and the western coast of Africa (numbers are estimated at 11 300 birds) (Champagnon *et al.*, 2019); second, a Southeast European/Mediterranean and tropical African population, wintering in

the Mediterranean or tropical areas of northern Africa (numbers are estimated at 11 700 birds) (Champagnon *et al.*, 2019); third, a West/Southwest and South Asian population, which breeds across Central and Southwest Asia and winters in Southwest and South Asia (numbers are estimated at 23 000 birds) (Wetlands International, 2021); and finally, an East Asian population, breeding mostly in Northeast Asia, Mongolia and China, and wintering in Southeast China, South Korea and Japan, (numbers are estimated at 20 000 birds) (Xi *et al.*, 2021).

Analysis of the genetic differences between subspecies and the genetic differentiation between geographic populations is particularly important in this species to ensure that subspecies are correctly recorded (Piersma *et al.*, 2012). Hybridisation has also been recorded between Eurasian and Black-faced spoonbills (Matheu *et al.*, 2020).

The main threats to this species include habitat loss and degradation, overfishing and disturbance, poaching and collision with overhead power lines (del Hoyo *et al.*, 1992; Hancock *et al.*, 1992; Triplet *et al.*, 2008).

This species is listed as Least Concern on the IUCN Red List both globally and in Europe (BirdLife International, 2019; BirdLife International, 2021). This is due to its extensive range, stable trend in population numbers and relative abundance. We present a chromosomally complete genome sequence for *Platalea leucorodia*, the Eurasian Spoonbill. The assembly was produced using the Tree of Life pipeline from a specimen collected in Newgrounds, Gloucester, England, United Kingdom (Figure 1), as part of the Darwin Tree of Life project.

Methods

Sample acquisition

The specimen used for genome sequencing was an adult female *Platalea leucorodia* (specimen ID NHMUK014561650, ToLID bPlaLeu2; Figure 1), collected from Newgrounds, Gloucester, England, United Kingdom (latitude 51.73, longitude −2.4) on 2022-02-02. This was a captive specimen found dead. The specimen was collected and identified by Michelle O'Brien



Figure 1. Photograph of the *Platalea leucorodia* (bPlaLeu2) specimen used for genome sequencing.

(Wildfowl & Wetlands Trust). Several small samples of pectoral muscle were taken and stored at -80°C . A sample from the same specimen was used for RNA sequencing. Sample metadata were collected in line with the Darwin Tree of Life project standards described by [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 bPlaLeu2 sample was weighed and [triaged](#) to determine the appropriate extraction protocol. Tissue from the muscle was homogenised by [cryogenic disruption](#) using the Covaris cryoPREP® Automated Dry Pulverizer.

HMW DNA was extracted in the WSI Scientific Operations core 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 [manual 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 15.37 ng/ μL and a yield of 691.65 ng, with a fragment size of 11.8 kb. The 260/280 spectrophotometric ratio was 1.8, and the 260/230 ratio was 2.27.

RNA was extracted from muscle tissue of bPlaLeu2 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 using the Sequel IIe system (Pacific Biosciences, California, USA). The concentration of the library loaded onto the Sequel IIe was in the range 40–135 pM. The SMRT link software, a PacBio web-based end-to-end workflow manager, was used to set-up and monitor the run, and

to perform primary and secondary analysis of the data upon completion.

Hi-C

Sample preparation and crosslinking

The Hi-C sample was prepared from 20–50 mg of frozen muscle tissue of the bPlaLeu2 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 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 6000.

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 6000 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 misassemblies. 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 53 breaks, 235 joins, and removal of 9 haplotypic duplications. 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 *Platalea leucorodia* specimen generated 42.31 Gb (gigabases) from 4.85 million reads, which were used to assemble the genome. GenomeScope2.0 analysis estimated the haploid genome size at 1 231.17 Mb, with a heterozygosity of 0.41% and repeat content of 7.96% (Figure 2). These estimates guided expectations for the assembly. Based

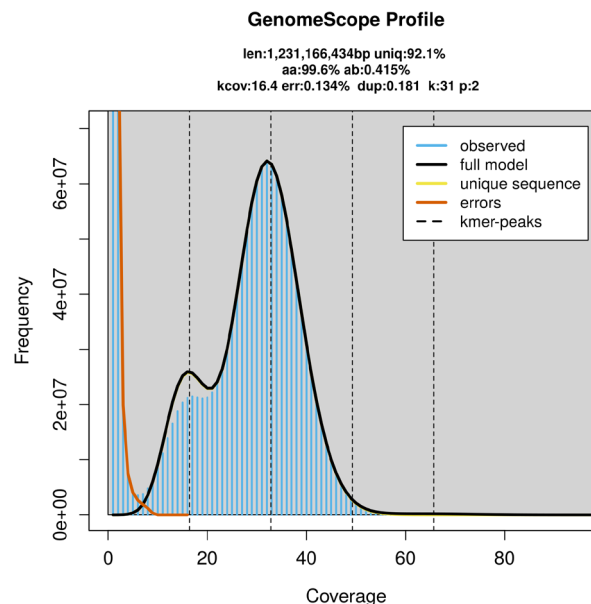


Figure 2. Frequency distribution of k -mers generated using GenomeScope2. The plot shows observed and modelled k -mer spectra, providing estimates of genome size, heterozygosity, and repeat content based on unassembled sequencing reads.

on the estimated genome size, the sequencing data provided approximately 33× coverage. Hi-C sequencing produced 77.64 Gb from 514.17 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.

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 1 345.83 Mb in 668 scaffolds, with 703 gaps, and a scaffold N50 of 94.71 Mb ([Table 2](#)).

Most of the assembly sequence (95.57%) was assigned to 37 chromosomal-level scaffolds, representing 35 autosomes and the W and Z sex chromosomes. These chromosome-level scaffolds, confirmed by Hi-C data, are named according to size ([Figure 3](#); [Table 3](#)). Curation of this assembly resulted in 35 autosomes, plus Z and W. The published karyotype is ± 70 ([Takagi & Sasaki, 1974](#)).

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 62.1, and for haplotype 2, 62.8. When the two haplotypes are combined, the assembly achieves an estimated QV of 62.4. The *k*-mer completeness is 95.62% for haplotype 1, 86.85% for haplotype 2, and 99.61% for the combined haplotypes ([Figure 4](#)).

BUSCO analysis using the aves_odb10 reference set (*n* = 8 338) identified 97.4% of the expected gene set (single = 96.0%, duplicated = 1.4%) 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.

Table 1. Specimen and sequencing data for BioProject PRJEB70851.

Platform	PacBio HiFi	Hi-C	RNA-seq
ToLID	bPlaLeu2	bPlaLeu2	bPlaLeu2
Specimen ID	NHMUK014561650	NHMUK014561650	NHMUK014561650
BioSample (source individual)	SAMEA112468120	SAMEA112468120	SAMEA112468120
BioSample (tissue)	SAMEA112468133	SAMEA112468133	SAMEA112468133
Tissue	muscle	muscle	muscle
Instrument	Sequel IIe	Illumina NovaSeq 6000	Illumina NovaSeq 6000
Run accessions	ERR12353021; ERR12353022	ERR12356293	ERR12356294
Read count total	4.85 million	514.17 million	61.18 million
Base count total	42.31 Gb	77.64 Gb	9.24 Gb

Table 2. Genome assembly statistics.

Assembly name	bPlaLeu2.hap1.1	bPlaLeu2.hap2.1
Assembly accession	GCA_965183815.1	GCA_965183785.1
Assembly level	chromosome	scaffold
Span (Mb)	1 345.83	1 190.44
Number of chromosomes	37	N/A
Number of contigs	1 371	1 012
Contig N50	3.3 Mb	3.6 Mb
Number of scaffolds	668	408
Scaffold N50	94.71 Mb	102.18 Mb
Longest scaffold length (Mb)	169.23	N/A
Sex chromosomes	W and Z	N/A
Organelles	Mitochondrion: 17.15 kb	N/A

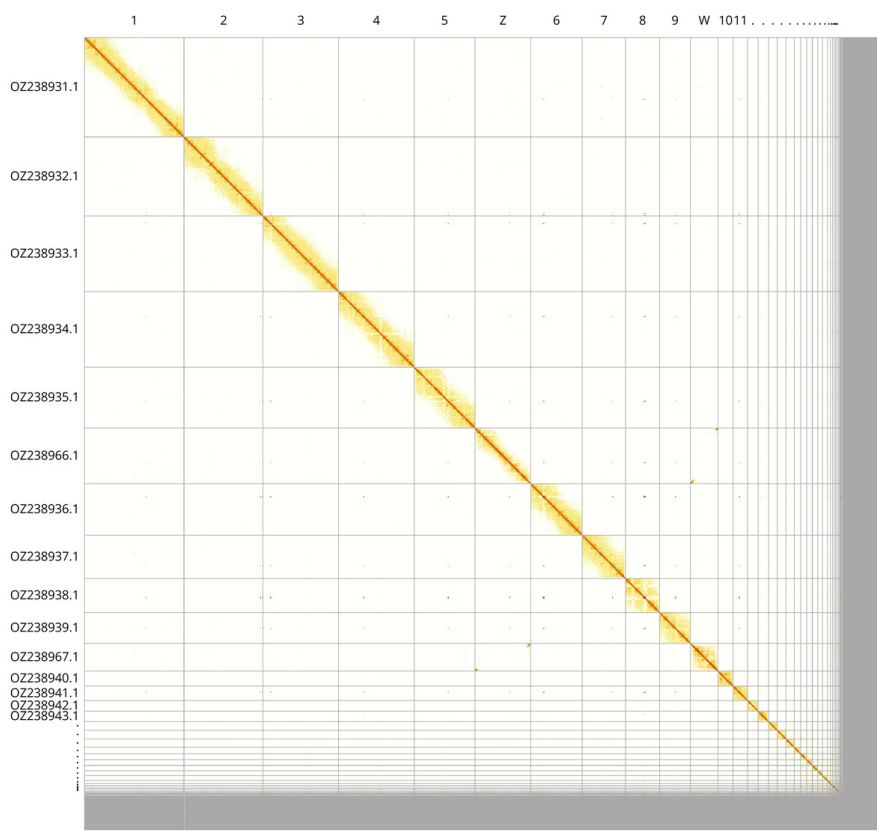


Figure 3. Hi-C contact map of the *Platalea leucorodia* genome assembly. Assembled chromosomes are shown in order of size and labelled along the axes, with a megabase scale shown below. The plot was generated using PretextSnapshot.

Table 3. Chromosomal pseudomolecules in the haplotype 1 genome assembly of *Platalea leucorodia* bPlaLeu2.

INSDC accession	Molecule	Length (Mb)	GC%
OZ238931.1	1	169.66	41
OZ238932.1	2	133.58	41
OZ238933.1	3	128.51	41.50
OZ238934.1	4	128.45	41.50
OZ238935.1	5	103.04	42.50
OZ238936.1	6	87.21	41
OZ238937.1	7	73.81	42.50
OZ238938.1	8	57.75	43.50
OZ238939.1	9	52.14	44
OZ238940.1	10	25.39	45
OZ238941.1	11	24.62	46
OZ238942.1	12	17.91	46.50
OZ238943.1	13	17.68	46.50
OZ238944.1	14	14.84	47.50
OZ238945.1	15	14.57	49
OZ238946.1	16	14.10	47.50
OZ238947.1	17	10.99	48.50
OZ238948.1	18	10.22	51.50

INSDC accession	Molecule	Length (Mb)	GC%
OZ238949.1	19	9.57	53
OZ238950.1	20	9.13	50
OZ238951.1	21	8.08	54.50
OZ238952.1	22	8.07	53
OZ238953.1	23	5.15	58
OZ238954.1	24	3.81	59
OZ238955.1	25	2.64	51
OZ238956.1	26	2.37	62.50
OZ238957.1	27	2.32	54.50
OZ238958.1	28	1.74	61.50
OZ238959.1	29	1.69	62.50
OZ238960.1	30	1.34	53.50
OZ238961.1	31	1.17	54
OZ238962.1	32	0.92	60.50
OZ238963.1	33	0.72	64.50
OZ238964.1	34	0.62	59
OZ238965.1	35	0.52	62.50
OZ238967.1	W	47.12	44
OZ238966.1	Z	94.71	41.50

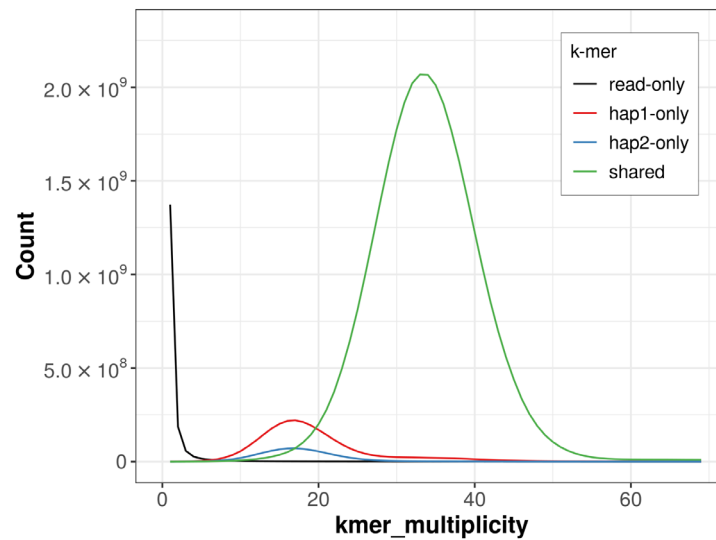


Figure 4. Evaluation of *k*-mer completeness using MerquyFK. This plot illustrates the recovery of *k*-mers from the original read data in the final assemblies. The horizontal axis represents *k*-mer multiplicity, and the vertical axis shows the number of *k*-mers. The black curve represents *k*-mers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

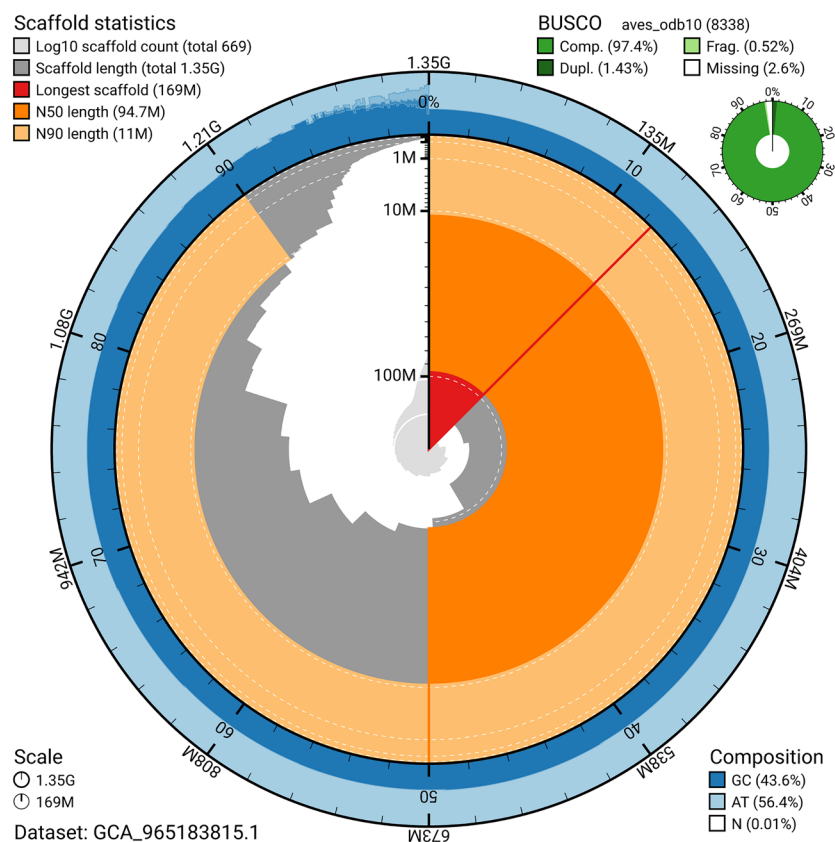


Figure 5. Assembly metrics for bPlaLeu2.hap1.1. The BlobToolKit snail plot provides an overview of assembly metrics and BUSCO gene completeness. The circumference represents the length of the whole genome sequence, and the main plot is divided into 1 000 bins around the circumference. The outermost blue 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](#).

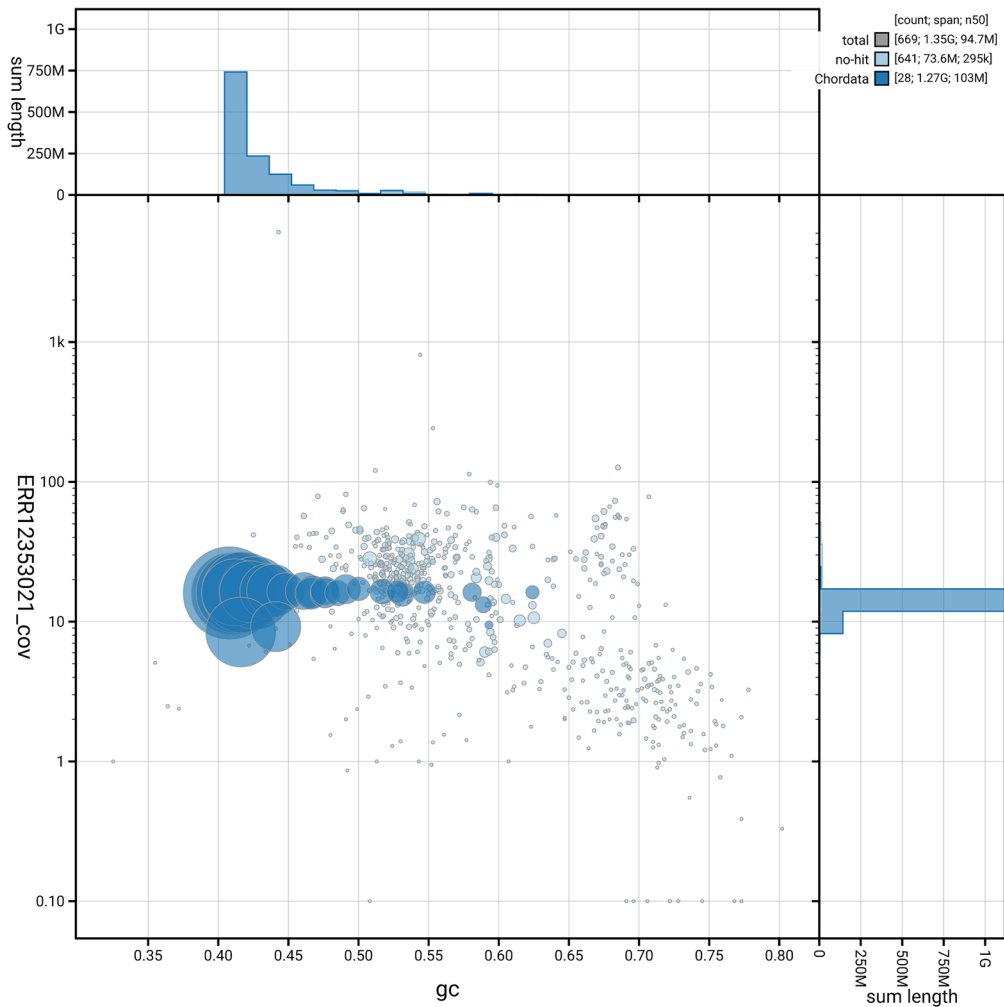


Figure 6. BlobToolKit GC-coverage plot for bPlaLeu2.hap1.1. Blob plot showing sequence coverage (vertical axis) and GC content (horizontal axis). The circles represent scaffolds, with the size proportional to scaffold length and the colour representing phylum membership. The histograms along the axes 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 *Platalea leucorodia* assembly.

Measure	Value	Benchmark
EBP summary (haplotype 1)	6.C.Q62	6.C.Q40
Contig N50 length	3.30 Mb	≥ 1 Mb
Scaffold N50 length	94.71 Mb	= chromosome N50
Consensus quality (QV)	Haplotype 1: 62.1; haplotype 2: 62.8; combined: 62.4	≥ 40
<i>k</i> -mer completeness	Haplotype 1: 95.62%; Haplotype 2: 86.85%; combined: 99.61%	≥ 95%
BUSCO	C:97.4% [S:96.0%; D:1.4%]; F:0.5%; M:2.1%; n:8338	S > 90%; D < 5%
Percentage of assembly assigned to chromosomes	95.57%	≥ 90%

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: *Platalea leucorodia* (Eurasian spoonbill). Accession number [PRJEB70851](#). The genome sequence is released openly for reuse. The *Platalea leucorodia* 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.

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	4.4.5	https://github.com/blobtoolkit/blobtoolkit
BUSCO	5.7.1	https://gitlab.com/ezlab/busco
bwa-mem2	2.2.1	https://github.com/bwa-mem2/bwa-mem2
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	2.28-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

Software	Version	Source
Nextflow	24.10.4	https://github.com/nextflow-io/nextflow
PretextSnapshot	N/A	https://github.com/sanger-tol/PretextSnapshot
PretextView	0.2.5	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	v0.7.1	https://github.com/sanger-tol/blobtoolkit
Seqtk	1.3	https://github.com/lh3/seqtk
Singularity	3.9.0	https://github.com/sylabs/singularity
TreeVal	1.2.0	https://github.com/sanger-tol/treeval
YaHS	1.2a.2	https://github.com/c-zhou/yahs

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](#)

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: **Platalea leucorodia (amended version of 2016 assessment).** 2019. [Publisher Full Text](#)
- BirdLife International: **Platalea leucorodia (Europe assessment).** 2021. [Publisher Full Text](#)
- 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](#)
- Champagnon J, Kralj J, Cano Alonso LS, *et al.*: **Proceedings of the IX workshop of the AEWA Eurasian Spoonbill international expert group.** In: *IUCN-SSC Stork, Ibis and Spoonbill Specialist Group Special Publication 2*, 2019. [Reference Source](#)
- 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](#)
- Cramp S, Simmons KEL: **The birds of the Western Palearctic.** Vol. 1. **Ostrich to ducks. Handbook of the birds of Europe, the Middle East and North Africa.** 1977. [Reference Source](#)
- 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](#)
- del Hoyo J, Elliot A, Sargatal J: **Handbook of the birds of the world, Vol. 1: ostrich to ducks.** Barcelona, Spain: Lynx Edicions, 1992. [Reference Source](#)
- 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](#)
- Hancock JA, Kushlan JA, Kahl MP: **Storks, ibises and spoonbills of the world.** London: Academic Press, 1992. [Reference Source](#)
- 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): g1aa153.
[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](#)

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, 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](#)

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](#)

Matheu E, del Hoyo J, Christie DA, *et al.*: **Eurasian Spoonbill (*Platalea leucorodia*), version 1.0.** In: J. del Hoyo, A. Elliott, J. Sargatal, *et al.* (eds), *Birds of the World*. Ithaca, NY, USA: Cornell Lab of Ornithology, 2020.
[Publisher Full Text](#)

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

Piersma T, van der Velde M, El-Hacen EHM, *et al.*: **Molecular verification of the subspecies status of the Mauritanian Spoonbill *Platalea leucorodia balsaci*.** *Ardea*. 2012; **100**: 131–36.
[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.*: **Merquy: reference-free quality, completeness, and phasing assessment for genome assemblies.** *Genome Biol*. 2020; **21**(1): 245.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free 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](#)

Swennen C, Yu YT: **Bill sweeping in spoonbills *Platalea*: no evidence for an effective suction force at the tip.** *J Avian Biol*. 2008; **39**: 3–6.
[Publisher Full Text](#)

Takagi N, Sasaki M: **A phylogenetic study of bird karyotypes.** *Chromosoma*. 1974; **46**(1): 91–120.
[PubMed Abstract](#) | [Publisher Full Text](#)

Triplet P, Overdijk O, Smart M, *et al.*: **International single species action plan for the conservation of the Eurasian Spoonbill *Platalea leucorodia*.** 2008.
[Reference Source](#)

Uliano-Silva M, Ferreira JGRN, Krashenninnikova 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.** In: *2019 IEEE International Parallel and Distributed Processing Symposium (IPDPS)*. IEEE, 2019; 314–324.
[Publisher Full Text](#)

Weihs D, Katzir G: **Eurasian spoonbills *Platalea leucorodia* are good hydrodynamicists after all: reply to Swennen and Yu.** *J Avian Biol*. 2008; **39**: 7–8.
[Publisher Full Text](#)

Wetlands International: **Waterbird population estimates 5 2021.** 2021.
[Reference Source](#)

Xi J, Deng X, Zhao G, *et al.*: **Migration routes, behaviour and protection status of Eurasian Spoonbills (*Platalea leucorodia*) wintering in China.** *Avian Res*. 2021; **12**: 1–17.
[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 25 October 2025

<https://doi.org/10.21956/wellcomeopenres.27408.r133385>

© 2025 Hansson B. 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.



Bengt Hansson 

Lund University, Lund, Sweden

The manuscript presents a high-quality genome assembly of *Platalea leucorodia* (Eurasian Spoonbill), generated as part of the Darwin Tree of Life project. The authors report two haplotypes with substantial coverage and chromosomal scaffolding, including the W and Z sex chromosomes, as well as a complete mitochondrial genome. The methods appear sound and appropriate for the scope of the project, and the results are clearly presented and well contextualized.

I am in general happy with the content, the methodological approach, and the presentation of the results. The assembly represents a valuable contribution to avian genomics and biodiversity research.

One minor point: the photo in Figure 1 appears to be upside down and should be corrected.

I recommend the manuscript for indexing with no major revisions.

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, population genetics, sex chromosome evolution, avian genomics

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

Reviewer Report 15 October 2025

<https://doi.org/10.21956/wellcomeopenres.27408.r133386>

© 2025 Lopes 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.



Ricardo Jorge Lopes 

CE3C - Centre for Ecology, Evolution and Environmental Changes & CHANGE - Global Change and Sustainability Institute, Departamento de Biologia, Faculdade de Ciências, Universidade de Lisboa, Lisboa, Portugal

This paper describes the procedures and results of a reference genome assembly of a female adult European spoonbill (*Platalea leucorodia*). I find the methodology robust, and well documented and the genome assembly of high quality. It is the second genome for the Genus (There is a genome of *Platalea minor* - black-faced spoonbill), and it would be pertinent to say this and provide some comparative summary, especially if you find significant differences.

I have some minor comments:

1. Sample acquisition

- Concerning the voucher, the bird was found dead, but the cause of death is not mentioned. I think it is important to inform if the death was a random incident, or caused by a pathogen of other infirmity. Was a necropsy made? If yes, was there any confirmation of cause of death?

- Do you have information about the captivity status? Was the bird born in captivity? If so, please inform about the age. If not, please give some information about the origin of the bird.

2. Assembly curation

Did the ASCC pipeline identified any cobiont? If so, please report them. Does the submitted Run accessions files are real raw files or the files after going through this pipeline?

3. Data availability section

Should the mention of *Platalea leucorodia* be in italic? Please check.

4. Figure 6

BlobToolKit GC-coverage plot - Design wise, the legend position, as it is, the break of the border of the plot is not "nice".

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

Yes

Are the protocols appropriate and is the work technically sound?

Partly

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 work mainly with Conservation Genetics, actively promoting the production of high quality reference genomes.

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 06 October 2025

<https://doi.org/10.21956/wellcomeopenres.27408.r133388>

© 2025 Shakya S. 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.



Subir Shakya 

¹ Harvard University, Cambridge, Massachusetts, USA

² Department of Evolutionary Biology, LMU Munich, Planegg-Martinsreid, Bavaria, Germany

Background: If you would like a description of the bird, then perhaps including the spoon shaped bill is a must compared to the other traits described.

Also, some of the background paragraphs are one line long. Perhaps coalesce a few of them into coherent paragraphs.

Most of the methods and results are pretty standard and I do not have issues regarding any of them. One thing I may like to see is a circos plot comparing the scaffolds to the other ibis genome, since in the ibis there is a neo-sex chromosome and thus a comparison would help visualize if the neo-sex chromosome is shared in the entire family or if it is species specific.

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

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.
