DATA NOTE

The genome sequence of the Centre-barred Sallow, *Atethmia centrago* (Haworth, 1809) [version 1; peer review: awaiting peer review]

Douglas Boyes[1+], Peter W. H. Holland[2],
University of Oxford and Wytham Woods 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

[1]UK Centre for Ecology & Hydrology, Wallingford, England, UK
[2]University of Oxford, Oxford, England, UK

[+] Deceased author

Open Peer Review

Approval Status *AWAITING PEER REVIEW*

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

Abstract

We present a genome assembly from an individual male *Atethmia centrago* (the Centre-barred Sallow; Arthropoda; Insecta; Lepidoptera; Noctuidae). The genome sequence is 926.6 megabases in span. Most of the assembly is scaffolded into 31 chromosomal pseudomolecules, including the Z sex chromosome. The mitochondrial genome has also been assembled and is 15.57 kilobases in length. Gene annotation of this assembly on Ensembl identified 21,345 protein coding genes.

Keywords

Atethmia centrago, Centre-barred Sallow, genome sequence, chromosomal, Lepidoptera

This article is included in the Tree of Life gateway.

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

Author roles: Boyes D: Investigation, Resources; **Holland PWH**: Writing – Original Draft Preparation, Writing – Review & Editing;

How to cite this article: Boyes D, Holland PWH, University of Oxford and Wytham Woods Genome Acquisition Lab *et al.* **The genome sequence of the Centre-barred Sallow,** *Atethmia centrago* **(Haworth, 1809) [version 1; peer review: awaiting peer review]** Wellcome Open Research 2024, **9**:53 https://doi.org/10.12688/wellcomeopenres.20651.1

First published: 19 Feb 2024, **9**:53 https://doi.org/10.12688/wellcomeopenres.20651.1

Species taxonomy

Eukaryota; Metazoa; Eumetazoa; Bilateria; Protostomia; Ecdysozoa; Panarthropoda; Arthropoda; Mandibulata; Pancrustacea; Hexapoda; Insecta; Dicondylia; Pterygota; Neoptera; Endopterygota; Amphiesmenoptera; Lepidoptera; Glossata; Neolepidoptera; Heteroneura; Ditrysia; Obtectomera; Noctuoidea; Noctuidae; Xyleninae; *Atethmia*; *Atethmia centrago* (Haworth, 1809) (NCBI:txid988071).

Background

The common name 'sallow' has been given to moths from several genera in the family Noctuidae with similar yellowish orange wing colouration; most species of sallow moth fly in late summer or autumn. *Atethmia centrago*, the Centre-barred Sallow, is well known example found in woodlands and gardens across much of Britain and mainland Europe, with additional scattered records from Ukraine, Turkey and Russia (GBIF Secretariat, 2023; NBN Atlas Partnership, 2023). *A. centrago* is distinguished from superficially similar species by a distinctive brown band limited by a proximal straight line spanning the orange-coloured forewings; straight lines are unusual in lepidopteran wing patterns, perhaps because many pattern elements are thought to develop under the influence of diffusing factors acting in round eyespots or oval 'symmetry systems', often bounded by wing veins (see Martin & Reed, 2014; Monteiro *et al.*, 2006). Colour polymorphism has been described in *A. centrago*, particularly in relation to the central cross band on the forewings (Heath & Emmet, 1983).

Atethmia centrago has a single generation per year in Britain. In autumn, the adults lay eggs on ash trees where the newly hatched larvae burrow into the ash buds in winter. Later in development the larvae switch to feeding on flowers and leaves, hiding in bark crevices or descending the tree in the daytime and emerging at night to feed (Heath & Emmet, 1983). As the larvae are thought to be obligate feeders on ash, *A. centrago* may be particularly vulnerable to the effects of ash dieback disease, a fungal pathogen that has spread across Northern Europe (Brooke *et al.*, 2014).

A complete genome sequence for *Atethmia centrago* will be useful for research into constraints limiting larval feeding strategies and facilitate studies into population responses driven by declines in foodplant abundance.

Genome sequence report

The genome was sequenced from one male *Atethmia centrago* (Figure 1) collected from Wytham, Oxfordshire, UK (51.77, –1.31). A total of 29-fold coverage in Pacific Biosciences single-molecule HiFi long reads and 42-fold coverage in 10X Genomics read clouds was generated. Primary assembly contigs were scaffolded with chromosome conformation Hi-C data. Manual assembly curation corrected 30 missing joins or mis-joins, reducing the scaffold number by 29.03%, and increasing the scaffold N50 by 0.60%.

The final assembly has a total length of 926.6 Mb in 44 sequence scaffolds with a scaffold N50 of 31.8 Mb (Table 1).

Figure 1. Photograph of the *Atethmia centrago* (ilAteCent1) specimen used for genome sequencing.

The snailplot in Figure 2 provides a summary of the assembly statistics, while the distribution of assembly scaffolds on GC proportion and coverage is shown in Figure 3. The cumulative assembly plot in Figure 4 shows curves for subsets of scaffolds assigned to different phyla. Most (99.85%) of the assembly sequence was assigned to 31 chromosomal-level scaffolds, representing 30 autosomes and the Z sex chromosome. Chromosome-scale scaffolds confirmed by the Hi-C data are named in order of size (Figure 5; Table 2). While not fully phased, the assembly deposited is of one haplotype. Contigs corresponding to the second haplotype have also been deposited. The mitochondrial genome was also assembled and can be found as a contig within the multifasta file of the genome submission.

The estimated Quality Value (QV) of the final assembly is 57.7 with *k*-mer completeness of 99.99%, and the assembly has a BUSCO v5.3.2 completeness of 99.1% (single = 98.6%, duplicated = 0.5%), using the lepidoptera_odb10 reference set ($n = 5,286$).

Metadata for specimens, barcode results, spectra estimates, sequencing runs, contaminants and pre-curation assembly statistics are given at https://links.tol.sanger.ac.uk/species/988071.

Genome annotation report

The *Atethmia centrago* genome assembly (GCA_905333075.3) was annotated using the Ensembl rapid annotation pipeline (Table 1; https://rapid.ensembl.org/Atethmia_centrago_GCA_905333075.3/Info/Index). The resulting annotation includes 21,555 transcribed mRNAs from 21,345 protein-coding genes.

Methods

Sample acquisition and nucleic acid extraction

A male *Atethmia centrago* (specimen ID Ox000211, ToLID ilAteCent1) was collected from Marley Fen, Wytham, Oxfordshire (biological vice-county Berkshire), UK (latitude 51.77, longitude –1.31) on 2019-08-24 using a light trap. This specimen was used for DNA and Hi-C sequencing. The

Table 1. Genome data for *Atethmia centrago*, ilAteCent1.3.

Project accession data		
Assembly identifier	ilAteCent1.3	
Species	*Atethmia centrago*	
Specimen	ilAteCent1	
NCBI taxonomy ID	988071	
BioProject	PRJEB43531	
BioSample ID	SAMEA7520177	
Isolate information	ilAteCent1, male: head and thorax (DNA sequencing), abdomen (Hi-C sequencing) ilAteCent2: abdomen (RNA sequencing)	
Assembly metrics*		***Benchmark***
Consensus quality (QV)	57.7	*≥ 50*
k-mer completeness	99.99%	*≥ 95%*
BUSCO**	C:99.1%[S:98.6%,D:0.5%], F:0.3%,M:0.6%,n:5,286	*C ≥ 95%*
Percentage of assembly mapped to chromosomes	99.85%	*≥ 95%*
Sex chromosomes	Z	*localised homologous pairs*
Organelles	Mitochondrial genome: 15.57 kb	*complete single alleles*
Raw data accessions		
PacificBiosciences SEQUEL II	ERR6548405, ERR6548404	
10X Genomics Illumina	ERR6054501, ERR6054498, ERR6054499, ERR6054500	
Hi-C Illumina	ERR6054502	
PolyA RNA-Seq Illumina	ERR9434969	
Genome assembly		
Assembly accession	GCA_905333075.3	
Accession of alternate haplotype	GCA_905332945.1	
Span (Mb)	926.6	
Number of contigs	80	
Contig N50 length (Mb)	26.6	
Number of scaffolds	44	
Scaffold N50 length (Mb)	31.8	
Longest scaffold (Mb)	40.62	
Genome annotation		
Number of protein-coding genes	21,345	
Number of gene transcripts	21,555	

* Assembly metric benchmarks are adapted from column VGP-2020 of "Table 1: Proposed standards and metrics for defining genome assembly quality" from (Rhie *et al.*, 2021).

** BUSCO scores based on the lepidoptera_odb10 BUSCO set using version 5.3.2. C = complete [S = single copy, D = duplicated], F = fragmented, M = missing, n = number of orthologues in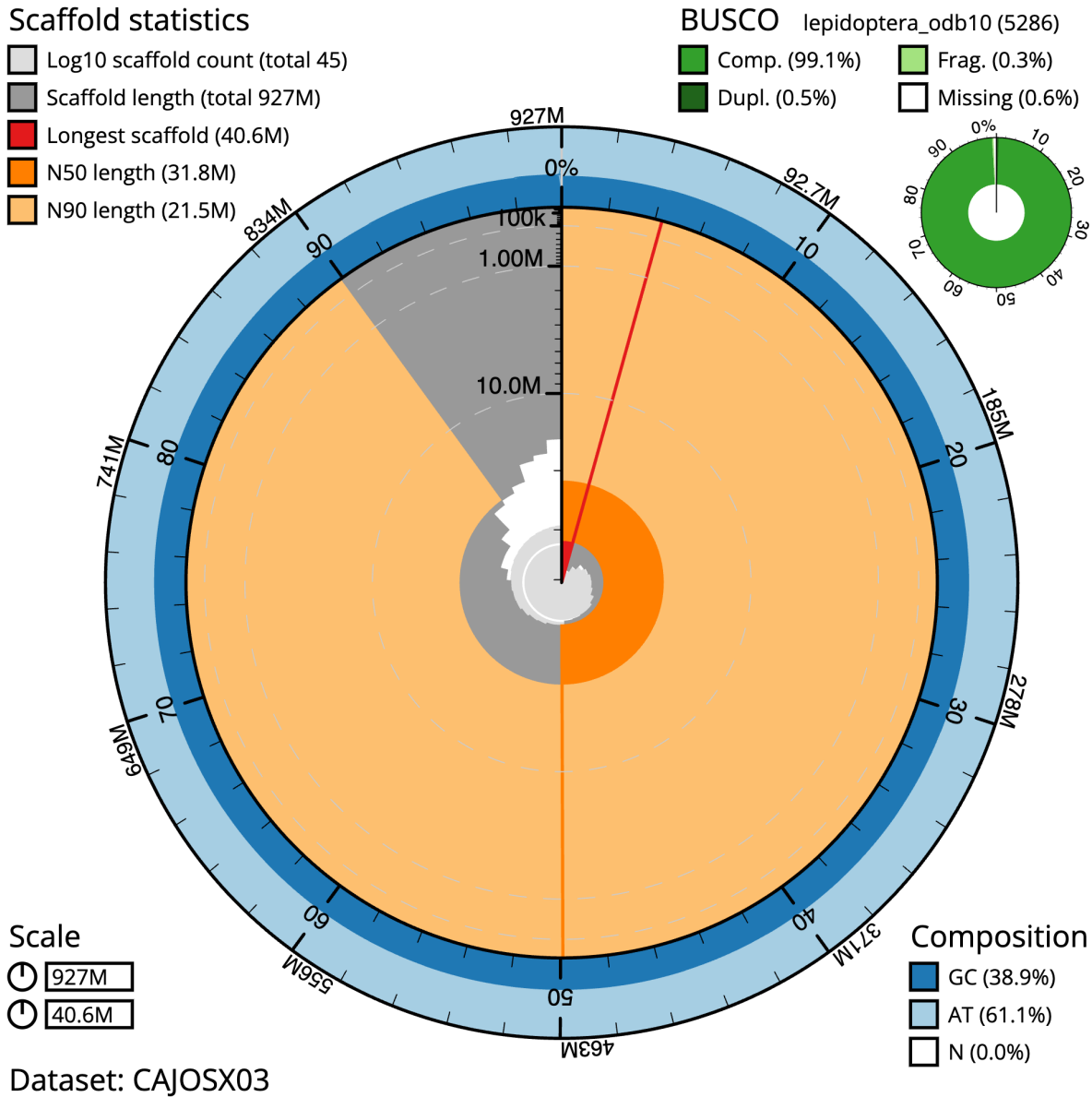 comparison. A full set of BUSCO scores is available at https://blobtoolkit.genomehubs.org/view/CAJOSX03/dataset/CAJOSX03/busco.

Figure 2. Genome assembly of *Atethmia centrago*, ilAteCent1.3: metrics. The BlobToolKit Snailplot shows N50 metrics and BUSCO gene completeness. The main plot is divided into 1,000 size-ordered bins around the circumference with each bin representing 0.1% of the 926,655,388 bp assembly. The distribution of scaffold lengths is shown in dark grey with the plot radius scaled to the longest scaffold present in the assembly (40,619,326 bp, shown in red). Orange and pale-orange arcs show the N50 and N90 scaffold lengths (31,773,998 and 21,540,269 bp), respectively. The pale grey spiral shows the cumulative scaffold count on a log scale with white scale lines showing successive orders of magnitude. The blue and pale-blue area around the outside of the plot shows the distribution of GC, AT and N percentages in the same bins as the inner plot. A summary of complete, fragmented, duplicated and missing BUSCO genes in the lepidoptera_odb10 set is shown in the top right. An interactive version of this figure is available at https://blobtoolkit.genomehubs.org/view/CAJOSX03/dataset/CAJOSX03/snail.

specimen used for RNA sequencing (specimen ID Ox000937, ToLID ilAteCent2) was collected from the same area on 2020-08-31 by Douglas Boyes. The specimens were collected and identified by Douglas Boyes (University of Oxford) and preserved on dry ice.

Protocols developed by the Wellcome Sanger Institute (WSI) Tree of Life core laboratory have been deposited on protocols. io (Denton *et al.*, 2023b). The workflow for high molecular weight (HMW) DNA extraction at the WSI includes a sequence of core procedures: sample preparation; sample homogenisation,
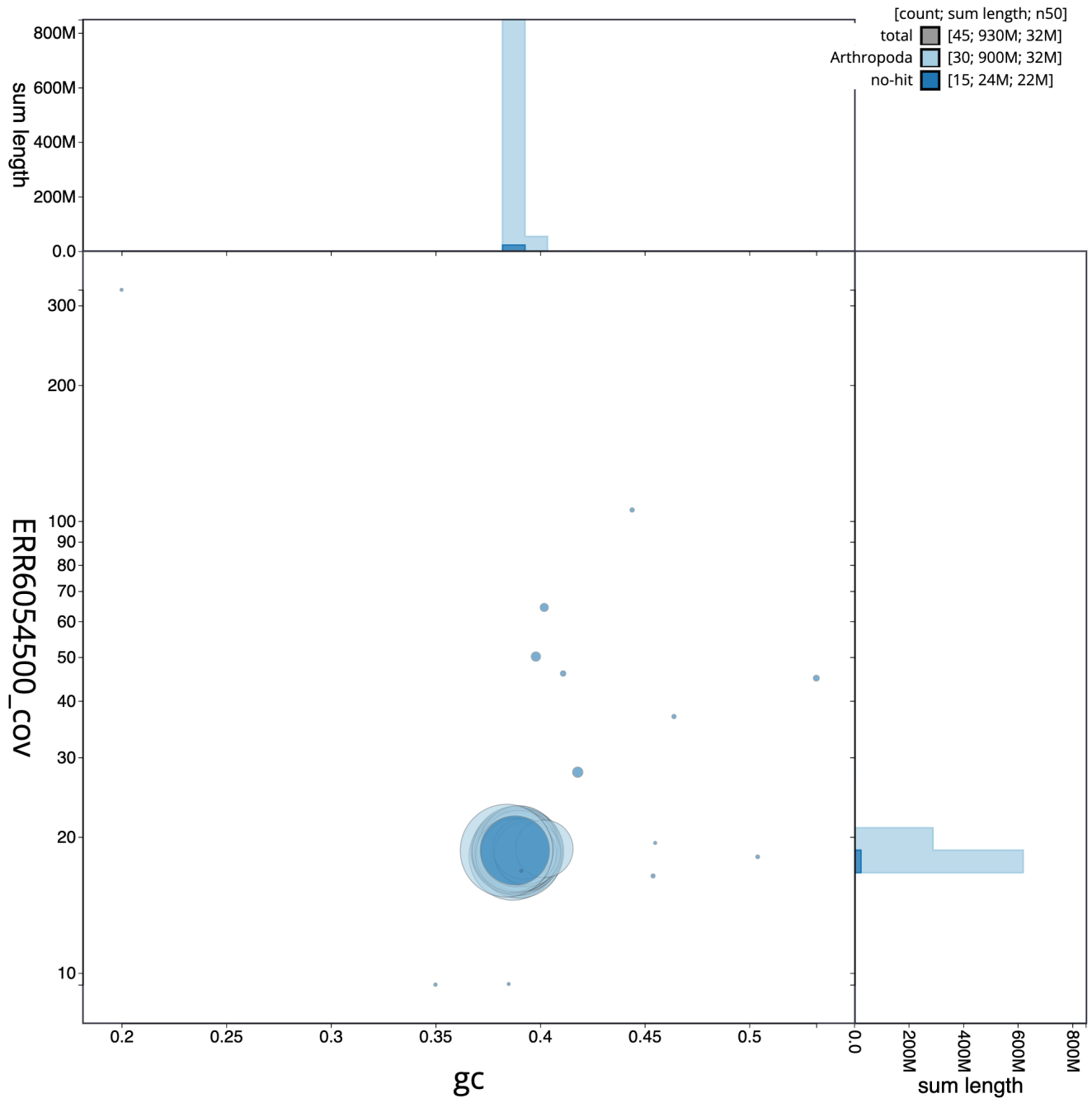
Figure 3. Genome assembly of *Atethmia centrago*, ilAteCent1.3: BlobToolKit GC-coverage plot. Scaffolds are coloured by phylum. Circles are sized in proportion to scaffold length. Histograms show the distribution of scaffold length sum along each axis. An interactive version of this figure is available at https://blobtoolkit.genomehubs.org/view/CAJOSX03/dataset/CAJOSX03/blob.

DNA extraction, fragmentation, and clean-up. In sample preparation, the ilAteCent1 sample was weighed and dissected on dry ice (Jay *et al.*, 2023). Tissue from the head and thorax was homogenised using a PowerMasher II tissue disruptor (Denton *et al.*, 2023a). HMW DNA was extracted using the Automated MagAttract v1 protocol (Oatley *et al.*, 2023), and was then sheared into an average fragment size

of 12–20 kb in a Megaruptor 3 system with speed setting 30 (Todorovic *et al.*, 2023). Sheared DNA was purified by solid-phase reversible immobilisation (Strickland *et al.*, 2023): in brief, the method employs a 1.8X ratio of AMPure PB beads to sample to eliminate shorter fragments and concentrate the DNA. The concentration of the sheared and purified DNA was assessed using a Nanodrop spectrophotometer and Qubit
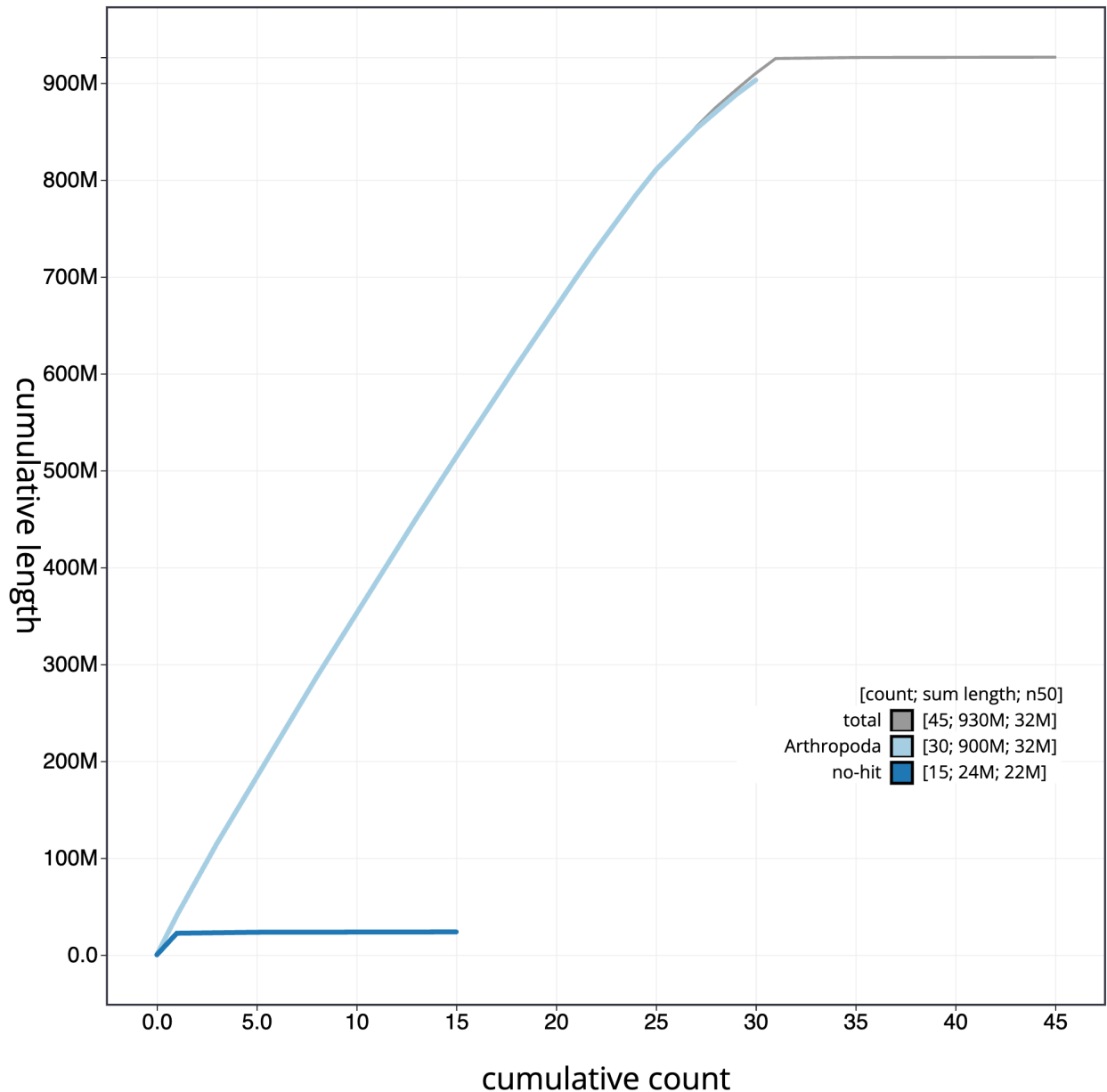
[count; sum length; n50]
total [45; 930M; 32M]
Arthropoda [30; 900M; 32M]
no-hit [15; 24M; 22M]

Figure 4. Genome assembly of *Atethmia centrago*, ilAteCent1.3: BlobToolKit cumulative sequence plot. The grey line shows cumulative length for all scaffolds. Coloured lines show cumulative lengths of scaffolds assigned to each phylum using the buscogenes taxrule. An interactive version of this figure is available at https://blobtoolkit.genomehubs.org/view/CAJOSX03/dataset/CAJOSX03/cumulative.

Fluorometer and Qubit dsDNA High Sensitivity Assay kit. Fragment size distribution was evaluated by running the sample on the FemtoPulse system.

RNA was extracted from abdomen tissue of ilAteCent2 in the Tree of Life Laboratory at the WSI using the RNA Extraction: Automated MagMax™ *mir*Vana protocol (do Amaral *et al.*, 2023). 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.

Sequencing

Pacific Biosciences HiFi circular consensus and 10X Genomics read cloud DNA sequencing libraries were constructed according to the manufacturers' instructions. Poly(A) RNA-Seq libraries were constructed using the NEB Ultra II RNA Library
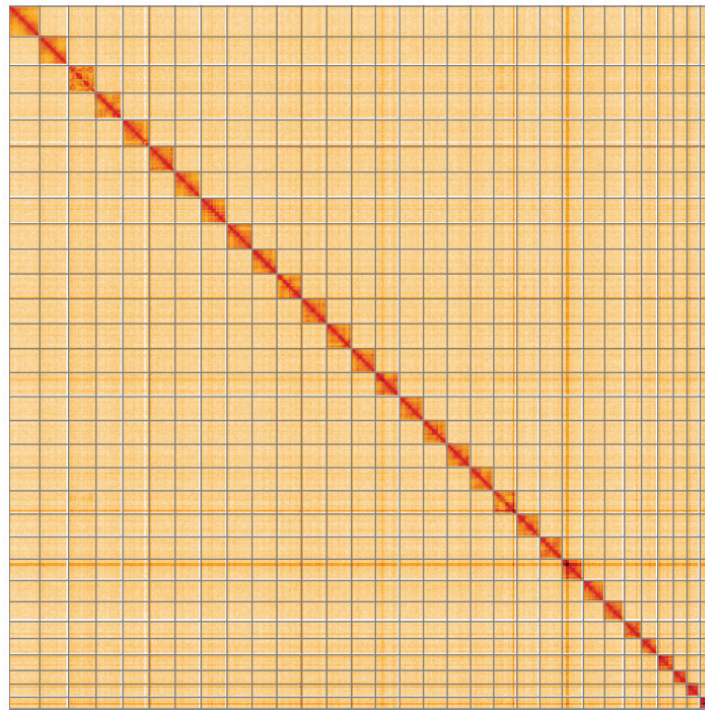
Figure 5. Genome assembly of *Atethmia centrago*, ilAteCent1.3: Hi-C contact map of the ilAteCent1.3 assembly, visualised using HiGlass. Chromosomes are shown in order of size from left to right and top to bottom. An interactive version of this figure may be viewed at https://genome-note-higlass.tol.sanger.ac.uk/l/?d=d_Pn_Zo_STmfSuX844wY8g.

Table 2. Chromosomal pseudomolecules in the genome assembly of *Atethmia centrago*, ilAteCent1.

INSDC accession	Chromosome	Length (Mb)	GC%
HG995367.1	1	37.65	39.0
HG995368.1	2	36.37	38.5
HG995369.1	3	35.02	39.0
HG995370.1	4	34.67	39.0
HG995371.1	5	34.29	39.0
HG995372.1	6	34.2	39.0
HG995373.1	7	33.9	39.0
HG995374.1	8	33.21	38.5
HG995375.1	9	32.77	39.0
HG995376.1	10	32.76	39.0
HG995377.1	11	32.66	39.0
HG995378.1	12	32.56	39.0
HG995379.1	13	31.77	39.0
HG995380.1	14	31.73	39.0
HG995381.1	15	31.41	38.5

INSDC accession	Chromosome	Length (Mb)	GC%
HG995382.1	16	31.24	38.5
HG995383.1	17	30.85	38.5
HG995384.1	18	30.52	39.0
HG995385.1	19	30.43	39.0
HG995386.1	20	30.25	39.0
HG995387.1	21	29.4	39.0
HG995388.1	22	28.02	39.0
HG995389.1	23	27.99	39.0
HG995390.1	24	25.99	38.5
HG995391.1	25	22.31	39.0
HG995392.1	26	21.54	39.0
HG995393.1	27	20.57	39.5
HG995394.1	28	17.95	39.5
HG995395.1	29	17.24	39.0
HG995396.1	30	15.39	40.0
HG995366.1	Z	40.62	38.5
HG998572.2	MT	0.02	20.0

Prep kit. DNA and RNA sequencing was performed by the Scientific Operations core at the WSI on Pacific Biosciences SEQUEL II (HiFi), Illumina HiSeq 4000 (RNA-Seq) and HiSeq X Ten (10X) instruments. Hi-C data were also generated from abdomen tissue of ilAteCent1 using the Arima2 kit and sequenced on the Illumina NovaSeq 6000 instrument.

Genome assembly, curation and evaluation

Assembly was carried out with Hifiasm (Cheng *et al.*, 2021) and haplotypic duplication was identified and removed with purge_dups (Guan *et al.*, 2020). One round of polishing was performed by aligning 10X Genomics read data to the assembly with Long Ranger ALIGN, calling variants with FreeBayes (Garrison & Marth, 2012). The assembly was then scaffolded with Hi-C data (Rao *et al.*, 2014) using SALSA2 (Ghurye *et al.*, 2019). The assembly was checked for contamination and corrected using the gEVAL system (Chow *et al.*, 2016) as described previously (Howe *et al.*, 2021). Manual curation was performed using gEVAL, HiGlass (Kerpedjiev *et al.*, 2018) and Pretext (Harry, 2022). The mitochondrial genome was assembled using MitoHiFi (Uliano-Silva *et al.*, 2023), which runs MitoFinder (Allio *et al.*, 2020) or MITOS (Bernt *et al.*, 2013) and uses these annotations to select the final mitochondrial contig and to ensure the general quality of the sequence.

A Hi-C map for the final assembly was produced using bwa-mem2 (Vasimuddin *et al.*, 2019) in the Cooler file format (Abdennur & Mirny, 2020). To assess the assembly metrics, the *k*-mer completeness and QV consensus quality values were calculated in Merqury (Rhie *et al.*, 2020). This work was done using Nextflow (Di Tommaso *et al.*, 2017) DSL2 pipelines "sanger-tol/readmapping" (Surana *et al.*, 2023a) and "sanger-tol/genomenote" (Surana *et al.*, 2023b). The genome was analysed within the BlobToolKit environment (Challis *et al.*, 2020) and BUSCO scores (Manni *et al.*, 2021; Simão *et al.*, 2015) were calculated.

Table 3 contains a list of relevant software tool versions and sources.

Genome annotation

The BRAKER2 pipeline (Brůna *et al.*, 2021) was used in the default protein mode to generate annotation for the *Atethmia centrago* assembly (GCA_905333075.3) in Ensembl Rapid Release.

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 here. 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.

Table 3. Software tools: versions and sources.

Software tool	Version	Source
BlobToolKit	4.1.7	https://github.com/blobtoolkit/blobtoolkit
BUSCO	5.3.2	https://gitlab.com/ezlab/busco
FreeBayes	1.3.1-17-gaa2ace8	https://github.com/freebayes/freebayes
gEVAL	N/A	https://geval.org.uk/
Hifiasm	0.12	https://github.com/chhylp123/hifiasm
HiGlass	1.11.6	https://github.com/higlass/higlass
Long Ranger ALIGN	2.2.2	https://support.10xgenomics.com/genome-exome/software/pipelines/latest/advanced/other-pipelines
Merqury	MerquryFK	https://github.com/thegenemyers/MERQURY.FK
MitoHiFi	1	https://github.com/marcelauliano/MitoHiFi
PretextView	0.2	https://github.com/wtsi-hpag/PretextView
purge_dups	1.2.3	https://github.com/dfguan/purge_dups
SALSA	2.2	https://github.com/salsa-rs/salsa
sanger-tol/genomenote	v1.0	https://github.com/sanger-tol/genomenote
sanger-tol/readmapping	1.1.0	https://github.com/sanger-tol/readmapping/tree/1.1.0

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: *Atethmia centrago* (centre-barred sallow). Accession number PRJEB43531; https://identifiers.org/ena.embl/PRJEB43531 (Wellcome Sanger Institute, 2022). The genome sequence is released openly for reuse. The *Atethmia centrago* genome sequencing initiative is part of the Darwin Tree of Life (DToL) project. All raw sequence data and the assembly have been deposited in INSDC databases. Raw data and assembly accession identifiers are reported in Table 1.

Author information
Members of the University of Oxford and Wytham Woods Genome Acquisition Lab are listed here: https://doi.org/10.5281/zenodo.7125292.

Members of the Darwin Tree of Life Barcoding collective are listed here: https://doi.org/10.5281/zenodo.4893703.

Members of the Wellcome Sanger Institute Tree of Life Management, Samples and Laboratory team are listed here: https://doi.org/10.5281/zenodo.10066175.

Members of Wellcome Sanger Institute Scientific Operations: Sequencing Operations are listed here: https://doi.org/10.5281/zenodo.10043364.

Members of the Wellcome Sanger Institute Tree of Life Core Informatics team are listed here: https://doi.org/10.5281/zenodo.10066637.

Members of the Tree of Life Core Informatics collective are listed here: https://doi.org/10.5281/zenodo.5013541.

Members of the Darwin Tree of Life Consortium are listed here: https://doi.org/10.5281/zenodo.4783558.

References

Abdennur N, Mirny LA: **Cooler: Scalable storage for Hi-C data and other genomically labeled arrays.** *Bioinformatics.* 2020; **36**(1): 311–316.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

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**

Bernt M, Donath A, Jühling F, *et al.*: **MITOS: Improved *de novo* metazoan mitochondrial genome annotation.** *Mol Phylogenet Evol.* 2013; **69**(2): 313–319.
PubMed Abstract | **Publisher Full Text**

Brooke A, Mitchell RJ, Harmer R: **Ash dieback and biodiversity loss: can management make broadleaved woodlands more resilient?** *Quarterly Journal of Forestry.* 2014; **108**(4): 241–248.
Reference Source

Brůna T, Hoff KJ, Lomsadze A, *et al.*: **BRAKER2: Automatic eukaryotic genome annotation with GeneMark-EP+ and AUGUSTUS supported by a protein database.** *NAR Genom Bioinform.* 2021; **3**(1): lqaa108.
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**

Chow W, Brugger K, Caccamo M, *et al.*: **gEVAL — a web-based browser for evaluating genome assemblies.** *Bioinformatics.* 2016; **32**(16): 2508–2510.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

Denton A, Oatley G, Cornwell C, *et al.*: **Sanger Tree of Life Sample Homogenisation: PowerMash.** *Protocols.io.* 2023a.
Publisher Full Text

Denton A, Yatsenko H, Jay J, *et al.*: **Sanger Tree of Life Wet Laboratory Protocol Collection V.1.** *Protocols.io.* 2023b.
Publisher Full Text

Di Tommaso P, Chatzou M, Floden EW, *et al.*: **Nextflow enables reproducible computational workflows.** *Nat Biotechnol.* 2017; **35**(4): 316–319.
PubMed Abstract | **Publisher Full Text**

do Amaral RJV, Bates A, Denton A, *et al.*: **Sanger Tree of Life RNA Extraction: Automated MagMax™ mirVana.** *protocols.io.* 2023.
Publisher Full Text

Garrison E, Marth G: **Haplotype-based variant detection from short-read sequencing**. 2012; [Accessed 26 July 2023].
Publisher Full Text

GBIF Secretariat: *Atethmia centrago* (Haworth, 1809). GBIF Backbone Taxonomy. 2023; [Accessed 20 November 2023].
Reference Source

Ghurye J, Rhie A, Walenz BP, *et al.*: **Integrating Hi-C links with assembly graphs for chromosome-scale assembly.** *PLoS Comput Biol.* 2019; **15**(8): e1007273.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

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

Harry E: **PretextView (Paired REad TEXTure Viewer): A desktop application for viewing pretext contact maps.** 2022; [Accessed 19 October 2022].
Reference Source

Heath J, Emmet MA: **The Moths and Butterflies of Great Britain and Ireland, Vol. 9 Part 1 (Sphingidae to Noctuidae - Noctuinae and Hadeninae)**. Colchester: Harley Books, 1983.
Reference Source

Howe K, Chow W, Collins J, *et al.*: **Significantly improving the quality of genome assemblies through curation.** *GigaScience.* Oxford University Press, 2021; **10**(1): giaa153.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

Jay J, Yatsenko H, Narváez-Gómez JP, *et al.*: **Sanger Tree of Life Sample Preparation: Triage and Dissection.** *Protocols.io.* 2023.
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**

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**

Martin A, Reed RD: ***Wnt* signaling underlies evolution and development of the butterfly wing pattern symmetry systems.** *Dev Biol.* 2014; **395**(2): 367–78.
PubMed Abstract | **Publisher Full Text**

Monteiro A, Glaser G, Stockslager S, *et al.*: **Comparative insights into questions of lepidopteran wing pattern homology.** *BMC Dev Biol.* 2006; **6**: 52.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

NBN Atlas Partnership: ***Atethmia centrago* (Haworth, 1809) Centre-barred Sallow.** NBN Atlas, 2023; [Accessed 20 November 2023].
Reference Source

Oatley G, Denton A, Howard C: **Sanger Tree of Life HMW DNA Extraction: Automated MagAttract v.2.** *Protocols.io.* 2023.
Publisher 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): 245.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

Simão FA, Waterhouse RM, Ioannidis P, *et al.*: **BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs.** *Bioinformatics.* 2015; **31**(19): 3210–3212.
PubMed Abstract | **Publisher Full Text**

Strickland M, Cornwell C, Howard C: **Sanger Tree of Life Fragmented DNA clean up: Manual SPRI.** *protocols.io.* 2023.
Publisher Full Text

Surana P, Muffato M, Qi G: **sanger-tol/readmapping: sanger-tol/readmapping v1.1.0 - Hebridean Black (1.1.0).** *Zenodo.* 2023a; [Accessed 21 July 2023].
Publisher Full Text

Surana P, Muffato M, Sadasivan Baby C: **sanger-tol/genomenote (v1.0.dev).** *Zenodo.* 2023b; [Accessed 21 July 2023].
Publisher Full Text

Todorovic M, Sampaio F, Howard C: **Sanger Tree of Life HMW DNA Fragmentation: Diagenode Megaruptor®3 for PacBio HiFi.** *protocols.io.* 2023.
Publisher Full Text

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

Vasimuddin M, Misra S, Li H, *et al.*: **Efficient Architecture-Aware Acceleration of BWA-MEM for Multicore Systems.** In: *2019 IEEE International Parallel and Distributed Processing Symposium (IPDPS).* IEEE, 2019; 314–324.
Publisher Full Text

Wellcome Sanger Institute: **The genome sequence of the Centre-barred Sallow, *Atethmia centrago* (Haworth, 1809).** European Nucleotide Archive. [dataset], accession number PRJEB43531, 2022.