DATA NOTE

The genome sequence of the Golden Argent moth, *Argyresthia goedartella* (Linnaeus, 1758) [version 1; peer review: awaiting peer review]

Douglas Boyes[1+], Clare Boyes[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]Independent researcher, Welshpool, Wales, UK

[+] Deceased author

v1 **First published:** 15 Apr 2024, **9**:196
https://doi.org/10.12688/wellcomeopenres.21227.1
Latest published: 15 Apr 2024, **9**:196
https://doi.org/10.12688/wellcomeopenres.21227.1

Abstract

We present a genome assembly from an individual female *Argyresthia goedartella* (the Golden Argent; Arthropoda; Insecta; Lepidoptera; Argyresthiidae). The genome sequence is 1,108.8 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 17.96 kilobases in length. Gene annotation of this assembly on Ensembl identified 13,530 protein coding genes.

Keywords

Argyresthia goedartella, Golden Argent moth, genome sequence, chromosomal, Lepidoptera

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.

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

Author roles: Boyes D: Investigation, Resources; **Boyes C**: 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 [206194, https://doi.org/10.35802/206194] and the Darwin Tree of Life Discretionary Award [218328, https://doi.org/10.35802/218328].

How to cite this article: Boyes D, Boyes C, University of Oxford and Wytham Woods Genome Acquisition Lab *et al.* **The genome sequence of the Golden Argent moth, *Argyresthia goedartella* (Linnaeus, 1758) [version 1; peer review: awaiting peer review]** Wellcome Open Research 2024, **9**:196 https://doi.org/10.12688/wellcomeopenres.21227.1

First published: 15 Apr 2024, **9**:196 https://doi.org/10.12688/wellcomeopenres.21227.1

Species taxonomy

Eukaryota; Opisthokonta; Metazoa; Eumetazoa; Bilateria; Protostomia; Ecdysozoa; Panarthropoda; Arthropoda; Mandibulata; Pancrustacea; Hexapoda; Insecta; Dicondylia; Pterygota; Neoptera; Endopterygota; Amphiesmenoptera; Lepidoptera; Glossata; Neolepidoptera; Heteroneura; Ditrysia; Yponomeutoidea; Argyresthiidae; *Argyresthia*; *Argyresthia goedartella* (Linnaeus, 1758) (NCBI:txid572725).

Background

Argyresthia goedartella, the Golden Argent, is a common micro-moth in the family Argyresthiidae, and is widely distributed throughout the British Isles and Europe. It is also found in North America (GBIF Secretariat, 2024). The tiny adult (forewing length 5–6 mm) has a number of forms, which occur throughout its range (Emmet, 1996). The most common form is relatively easy to identify with a golden Y-shaped mark across the middle of the white forewing. The top and bottom of the forewing are also golden (Sterling *et al.*, 2023).

The pinkish caterpillars feed on the terminal buds and catkins of alder and birch. They overwinter in a shoot or catkin until late spring when the larva descends on a silken thread to pupate under the bark. They occasionally pupate in a fungus *Fomitopsis betulinus* (Langmaid *et al.*, 2018). Pupation may be delayed for several weeks, and several larvae can sometimes be found together under bark. In Norway, the larvae and pupae of *A. geodartella* have been found to be an important food source for Lesser Spotted Woodpeckers: this research demonstrated a strong influence on breeding success and adult woodpecker survival in early spring (Selås *et al.*, 2008). The adult moth emerges between June to August and can be found by beating the host plants. It will fly on sunny afternoons and also comes to light (Emmet, 1996).

The genome of *Argyresthia goedartella* was sequenced as part of the Darwin Tree of Life Project, a collaborative effort to sequence all the named eukaryotic species in the Atlantic Archipelago of Britain and Ireland. The genome sequence will be useful for research into colour variation in moths. Here we present a chromosomally complete genome sequence for *Argyresthia goedartella* based on one female specimen from Wytham Woods, Oxfordshire, UK.

Genome sequence report

The genome was sequenced from one female *Argyresthia goedartella* (Figure 1) collected from Marley Fen, Wytham Woods, Oxfordshire, UK (51.77, –1.31). A total of 29-fold coverage in Pacific Biosciences single-molecule HiFi long reads was generated. Primary assembly contigs were scaffolded with chromosome conformation Hi-C data. Manual assembly curation corrected 62 missing joins or mis-joins and removed 13 haplotypic duplications, reducing the assembly length by 0.83% and the scaffold number by 13.21%, also decreasing the scaffold N50 by 0.48%.

The final assembly has a total length of 1,108.8 Mb in 91 sequence scaffolds with a scaffold N50 of 37.0 Mb (Table 1).

Figure 1. Photograph of the *Argyresthia goedartella* (ilArgGoed1) specimen used for genome sequencing.

The snail plot 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 (98.45%) 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). The order and orientation of chromosome 1 is uncertain from 42 Mb to the end. 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 64.0 with *k*-mer completeness of 100.0%, and the assembly has a BUSCO v5.3.2 completeness of 97.5% (single = 96.1%, duplicated = 1.4%), 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/572725.

Genome annotation report

The *Argyresthia goedartella* genome assembly (GCA_949825045.1) was annotated at the European Bioinformatics Institute (EBI) on Ensembl Rapid Release. The resulting annotation includes 25,747 transcribed mRNAs from 13,530 protein-coding and 2,470 non-coding genes (Table 1; https://rapid.ensembl.org/Argyresthia_goedartella_GCA_949825045.1/Info/Index).

Table 1. Genome data for *Argyresthia goedartella*, ilArgGoed1.1.

Project accession data		
Assembly identifier	ilArgGoed1.1	
Species	*Argyresthia goedartella*	
Specimen	ilArgGoed1	
NCBI taxonomy ID	572725	
BioProject	PRJEB59390	
BioSample ID	SAMEA7520176	
Isolate information	ilArgGoed1: whole organism (DNA sequencing) ilArgGoed2: whole organism (Hi-C sequencing) ilArgGoed3: whole organism (RNA sequencing)	
Assembly metrics*		***Benchmark***
Consensus quality (QV)	64.0	*≥ 50*
k-mer completeness	100.0%	*≥ 95%*
BUSCO**	C:97.5%[S:96.1%,D:1.4%],F:0.6%, M:1.9%,n:5,286	*C ≥ 95%*
Percentage of assembly mapped to chromosomes	98.45%	*≥ 95%*
Sex chromosomes	Z	*localised homologous pairs*
Organelles	Mitochondrial genome: 17.96 kb	*complete single alleles*
Raw data accessions		
PacificBiosciences SEQUEL II	ERR10925356, ERR10925357	
Hi-C Illumina	ERR10851534	
PolyA RNA-Seq Illumina	ERR10851535	
Genome assembly		
Assembly accession	GCA_949825045.1	
Accession of alternate haplotype	GCA_949825015.1	
Span (Mb)	1108.8	
Number of contigs	217	
Contig N50 length (Mb)	11.1	
Number of scaffolds	91	
Scaffold N50 length (Mb)	37.0	
Longest scaffold (Mb)	50.05	
Genome annotation		
Number of protein-coding genes	13,530	
Number of non-coding genes	2,470	
Number of gene transcripts	25,747	

* 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/ilArgGoed1_1/dataset/ilArgGoed1_1/busco.

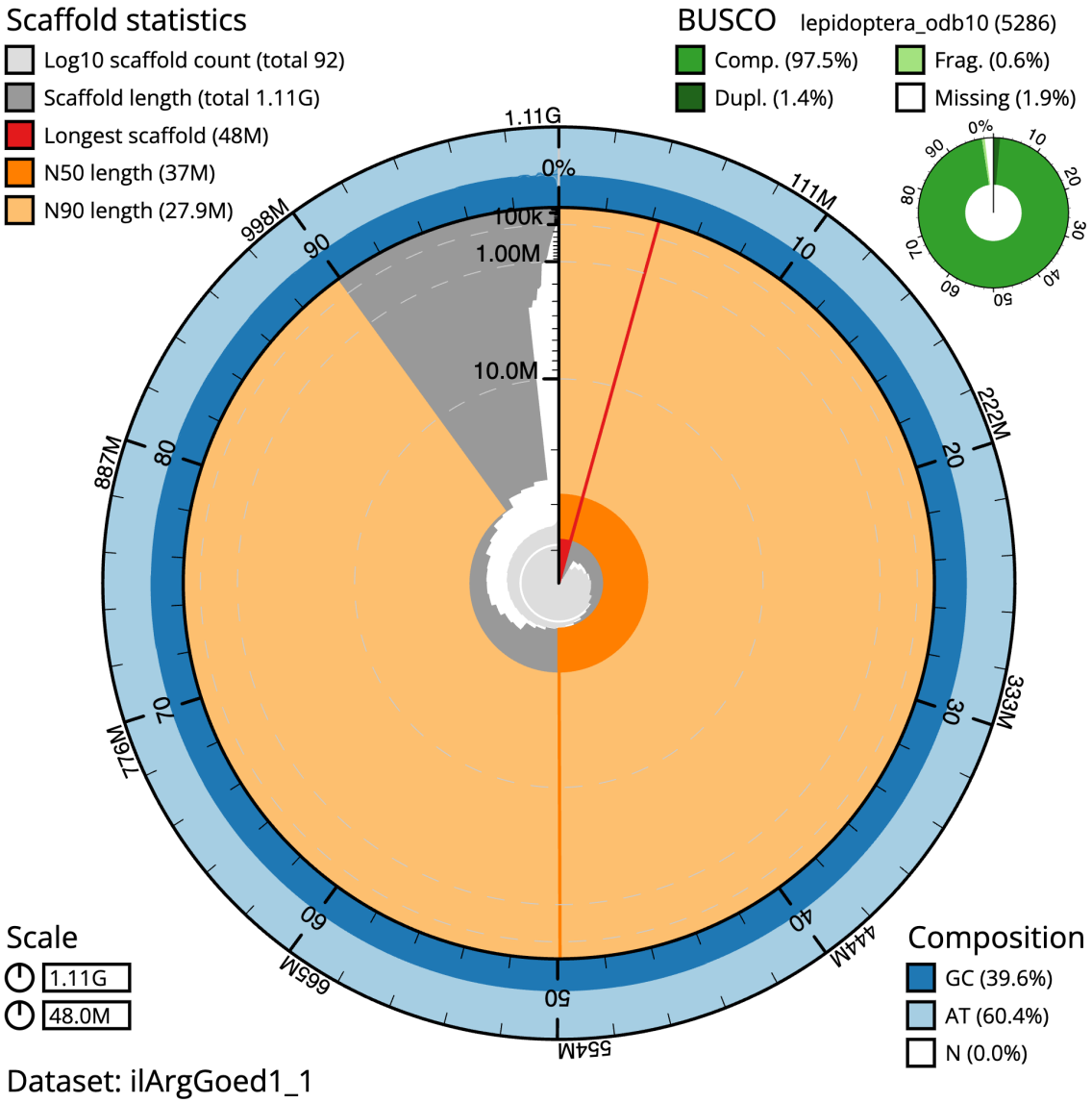

Figure 2. Genome assembly of *Argyresthia goedartella*, ilArgGoed1.1: metrics. The BlobToolKit snail plot shows N50 metrics and BUSCO gene completeness. The main plot is divided into 1,000 size-ordered bins around the circumference with each bin representing 0.1% of the 1,108,819,053 bp assembly. The distribution of scaffold lengths is shown in dark grey with the plot radius scaled to the longest scaffold present in the assembly (48,028,177 bp, shown in red). Orange and pale-orange arcs show the N50 and N90 scaffold lengths (36,986,452 and 27,886,200 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/ilArgGoed1_1/dataset/ilArgGoed1_1/snail.

Methods

Sample acquisition and nucleic acid extraction

The specimen used for DNA sequencing was a female *Argyresthia goedartella* (specimen ID Ox000216, ToLID ilArgGoed1), collected from Marley Fen, Wytham Woods, Oxfordshire (biological vice-county Berkshire), UK (latitude 51.77, longitude –1.31) on 2019-08-24, using a light trap. The specimens used for Hi-C sequencing (specimen ID Ox001867, ToLID ilArgGoed2) and for RNA sequencing (specimen ID Ox001868, ToLID ilArgGoed3) were collected from Wytham Woods, Oxfordshire, UK (latitude 51.77, longitude –1.34) on 2021-08-12 using a light trap. All specimens were collected and identified by Douglas Boyes (University of Oxford), and then preserved on dry ice.

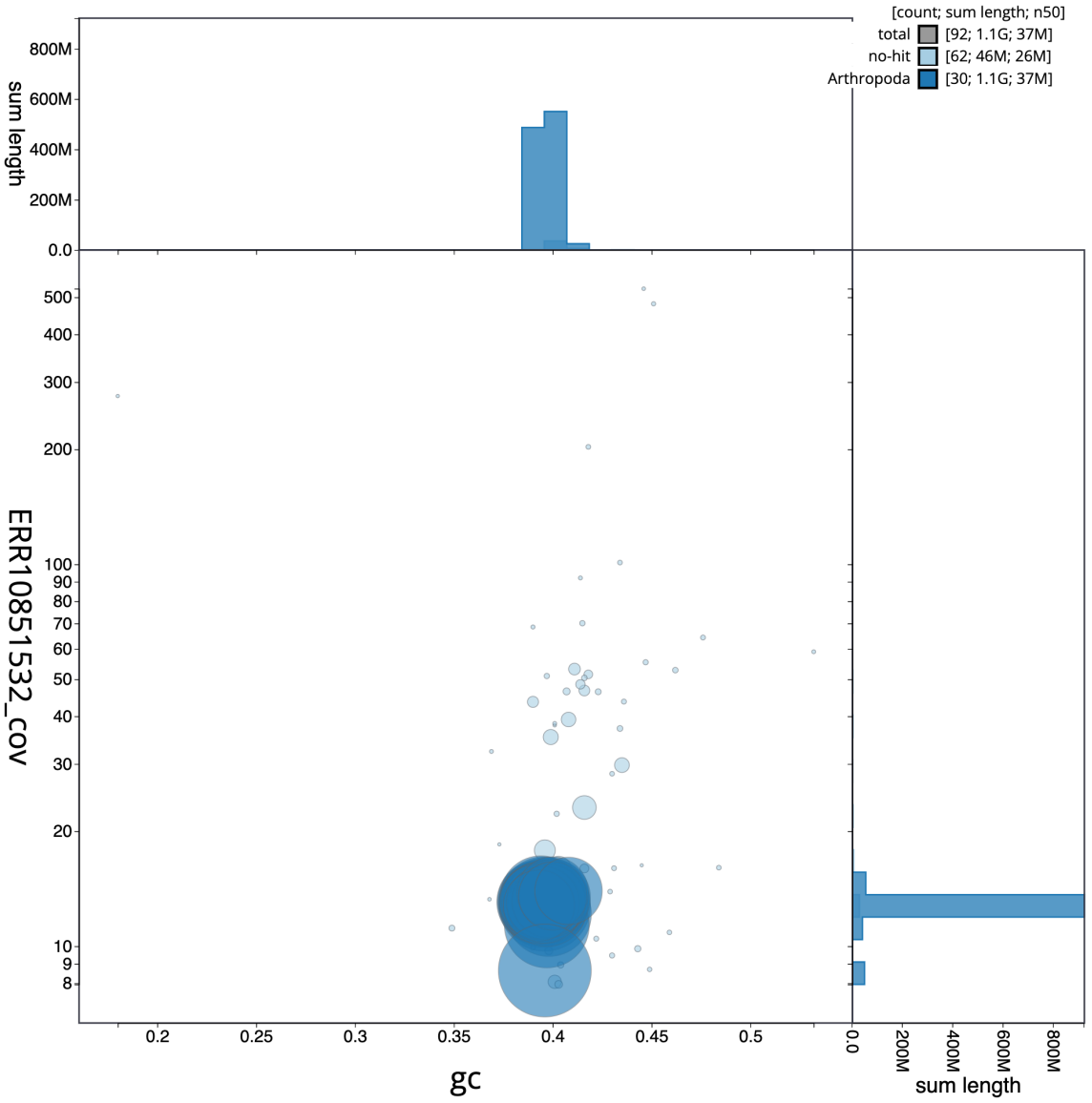

Figure 3. Genome assembly of *Argyresthia goedartella*, ilArgGoed1.1: BlobToolKit GC-coverage plot. Sequences are coloured by phylum. Circles are sized in proportion to sequence length. Histograms show the distribution of sequence length sum along each axis. An interactive version of this figure is available at https://blobtoolkit.genomehubs.org/view/ilArgGoed1_1/dataset/ilArgGoed1_1/blob.

The workflow for high molecular weight (HMW) DNA extraction at the Wellcome Sanger Institute (WSI) includes a sequence of core procedures: sample preparation; sample homogenisation, DNA extraction, fragmentation, and clean-up. In sample preparation, the ilArgGoed1 sample was weighed and dissected on dry ice (Jay *et al.*, 2023). Tissue from the whole organism was homogenised using a PowerMasher II tissue disruptor (Denton *et al.*, 2023a). HMW DNA was extracted using the Automated MagAttract v1 protocol (Sheerin *et al.*, 2023). DNA was 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 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 the whole organism tissue of ilArgGoed3 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

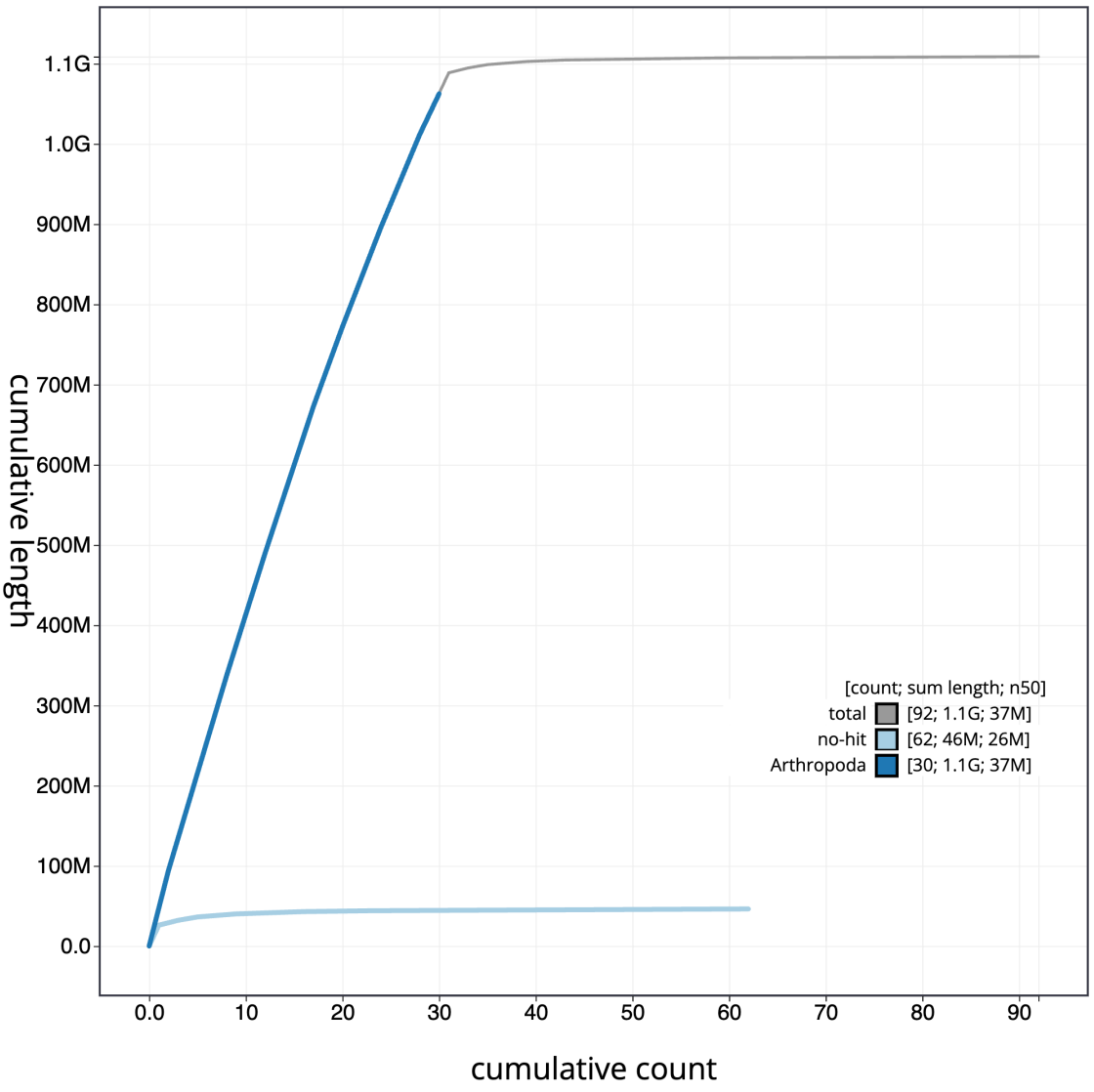

Figure 4. Genome assembly of *Argyresthia goedartella*, ilArgGoed1.1: BlobToolKit cumulative sequence plot. The grey line shows cumulative length for all sequences. Coloured lines show cumulative lengths of sequences assigned to each phylum using the buscogenes taxrule. An interactive version of this figure is available at https://blobtoolkit.genomehubs.org/view/ilArgGoed1_1/dataset/ilArgGoed1_1/cumulative.

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.

Protocols developed by the WSI Tree of Life laboratory are publicly available on protocols.io (Denton *et al.*, 2023b).

Sequencing

Pacific Biosciences HiFi circular consensus DNA sequencing libraries were constructed according to the manufacturers' instructions. Poly(A) RNA-Seq libraries were constructed using the NEB Ultra II RNA Library Prep kit. DNA and RNA sequencing was performed by the Scientific Operations core at the WSI on Pacific Biosciences SEQUEL II (HiFi) and Illumina NovaSeq 6000 (RNA-Seq) instruments. Hi-C data were also generated from the whole organism tissue of ilArgGoed2 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). The assembly was then scaffolded with Hi-C data (Rao *et al.*, 2014) using YaHS (Zhou *et al.*, 2023). The assembly was checked for contamination and corrected using the TreeVal pipeline (Pointon *et al.*, 2023). Manual curation was performed using JBrowse2 (Diesh *et al.*, 2023), HiGlass (Kerpedjiev *et al.*, 2018) and PretextView

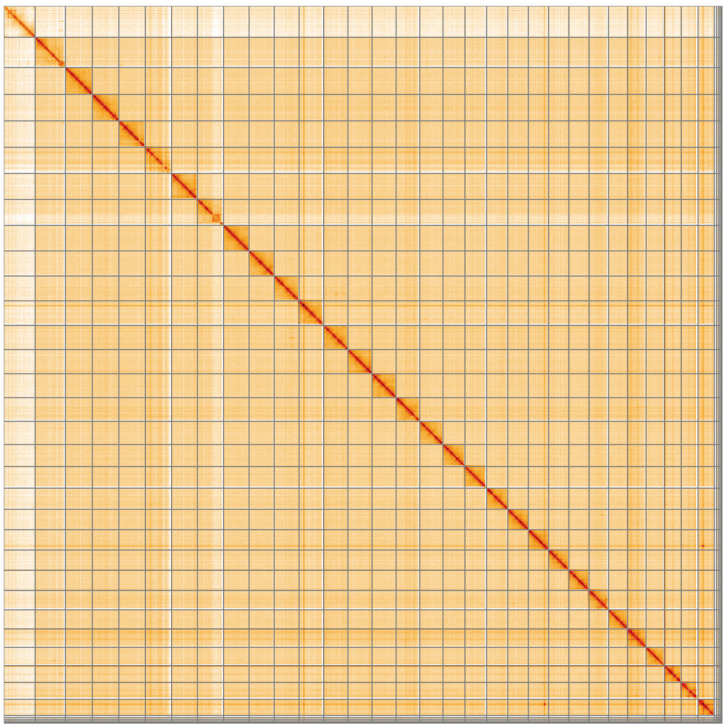

Figure 5. Genome assembly of *Argyresthia goedartella*, ilArgGoed1.1: Hi-C contact map of the ilArgGoed1.1 assembly, visualised using HiGlass. Chromosomes are shown in order of size from left to right and top to bottom. An interactive version of this figure may be viewed at https://genome-note-higlass.tol.sanger.ac.uk/l/?d=fu7E0qweQI6dfWlLcNrzNQ.

Table 2. Chromosomal pseudomolecules in the genome assembly of *Argyresthia goedartella*, ilArgGoed1.

INSDC accession	Chromosome	Length (Mb)	GC%
OX463807.1	1	46.37	39.5
OX463808.1	2	41.23	39.5
OX463809.1	3	40.64	39.5
OX463810.1	4	40.55	39.5
OX463811.1	5	40.14	40.0
OX463812.1	6	39.86	39.5
OX463813.1	7	39.81	39.5
OX463814.1	8	39.16	39.5
OX463815.1	9	38.51	39.5
OX463816.1	10	38.15	39.5
OX463817.1	11	37.58	39.5
OX463818.1	12	37.16	39.5
OX463819.1	13	36.99	39.5
OX463820.1	14	36.83	39.5
OX463821.1	15	36.21	39.5
OX463822.1	16	35.64	39.5
OX463823.1	17	33.65	39.5
OX463824.1	18	33.3	39.5
OX463825.1	19	32.33	39.5
OX463826.1	20	31.41	39.5
OX463827.1	21	31.17	40.0
OX463828.1	22	31.05	39.5
OX463829.1	23	30.89	39.0
OX463830.1	24	29.81	40.0
OX463831.1	25	29.36	39.5
OX463832.1	26	28.23	40.0
OX463833.1	27	27.89	39.5
OX463834.1	28	26.07	40.0
OX463835.1	29	25.65	40.0
OX463836.1	30	24.95	41.0
OX463806.1	Z	48.03	39.5
OX463837.1	MT	0.02	18.0

(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 Ensembl Genebuild annotation system (Aken *et al.*, 2016) was used to generate annotation for the *Argyresthia goedartella* assembly (GCA_949825045.1) in Ensembl Rapid Release at the EBI. Annotation was created primarily through alignment of transcriptomic data to the genome, with gap filling via protein-to-genome alignments of a select set of proteins from UniProt (UniProt Consortium, 2019).

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.

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

Table 3. Software tools: versions and sources.

Software tool	Version	Source
BlobToolKit	4.2.1	https://github.com/blobtoolkit/blobtoolkit
BUSCO	5.3.2	https://gitlab.com/ezlab/busco
Hifiasm	0.16.1-r375	https://github.com/chhylp123/hifiasm
HiGlass	1.11.6	https://github.com/higlass/higlass
Merqury	MerquryFK	https://github.com/thegenemyers/MERQURY.FK
MitoHiFi	2	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
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
TreeVal	-	https://github.com/sanger-tol/treeval
YaHS	1.1a.2	https://github.com/c-zhou/yahs

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: *Argyresthia goedartella* (golden argent). Accession number PRJEB59390; https://identifiers.org/ena.embl/PRJEB59390 (Wellcome Sanger Institute, 2023). The genome sequence is released openly for reuse. The *Argyresthia goedartella* 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

Aken BL, Ayling S, Barrell D, *et al.*: **The Ensembl gene annotation system.** *Database (Oxford).* 2016; **2016**: baw093.
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

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

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

Diesh C, Stevens GJ, Xie P, *et al.*: **JBrowse 2: a modular genome browser with views of synteny and structural variation.** *Genome Biol.* 2023; **24**(1): 74.
PubMed Abstract | Publisher Full Text | Free 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

Emmet AM: **The Moths and Butterflies of Great Britain and Ireland.** Yponomeutidae - Elachistidae. Colchester: Harley Books, 1996.

GBIF Secretariat: ***Argyresthia goedartella* (Linnaeus, 1758).** *GBIF Backbone Taxonomy.* 2024; [Accessed 28 February 2024].
Reference Source

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

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

Langmaid JR, Palmer S, Young MR: **A Field Guide to the Smaller Moths of Great Britain and Ireland.** 3rd ed, British Entomological and Natural History Society, 2018.
Reference Source

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

Pointon DL, Eagles W, Sims Y, *et al.*: **sanger-tol/treeval v1.0.0 – Ancient Atlantis.** 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

Selås V, Steen R, Kobro S, *et al.*: **Direct and indirect weather impacts on spring populations of lesser spotted woodpecker (*Dendrocopos minor*) in Norway.** *Scand J For Res.* 2008; **23**(2): 148–153.
Publisher Full Text

Sheerin E, Sampaio F, Oatley G, *et al.*: **Sanger Tree of Life HMW DNA**

Extraction: Automated MagAttract v.1. *protocols.io.* 2023.
Publisher 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**

Sterling P, Parsons M, Lewington R: **Field Guide to the Micro-moths of Great Britain and Ireland.** London: Bloomsbury Publishing, 2023.

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.
Publisher Full Text

Surana P, Muffato M, Sadasivan Baby C: **sanger-tol/genomenote (v1.0.dev).** *Zenodo.* 2023b.
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**

UniProt Consortium: **UniProt: a worldwide hub of protein knowledge.** *Nucleic Acids Res.* 2019; **47**(D1): D506–D515.
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 Golden Argent moth, *Argyresthia goedartella* (Linnaeus, 1758).** European Nucleotide Archive, [dataset], accession number PRJEB59390, 2023.

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**