DATA NOTE

# The genome sequence of the Ruddy Flat-body, *Agonopterix subpropinquella* (Stainton, 1849) [version 1; peer review: 2 approved]

Douglas Boyes[1+], James Hammond[2],
University of Oxford and Wytham Woods Genome Acquisition Lab,
Darwin Tree of Life Barcoding collective,
Wellcome Sanger Institute Tree of Life programme,
Wellcome Sanger Institute Scientific Operations: DNA Pipelines collective,
Tree of Life Core Informatics collective, Darwin Tree of Life Consortium

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

[+] Deceased author

**v1** **First published:** 23 Oct 2023, **8**:487
https://doi.org/10.12688/wellcomeopenres.20170.1
**Latest published:** 23 Oct 2023, **8**:487
https://doi.org/10.12688/wellcomeopenres.20170.1

## Abstract

We present a genome assembly from an individual male *Agonopterix subpropinquella* (the Ruddy Flat-body; Arthropoda; Insecta; Lepidoptera; Depressariidae). The genome sequence is 667.9 megabases in span. Most of the assembly is scaffolded into 28 chromosomal pseudomolecules, including the Z sex chromosome. The mitochondrial genome has also been assembled and is 16.5 kilobases in length. Gene annotation of this assembly on Ensembl identified 18,796 protein coding genes.

## Keywords

Agonopterix subpropinquella, Ruddy Flat-body, genome sequence, chromosomal, Lepidoptera

This article is included in the Tree of Life gateway.

**Open Peer Review**

**Approval Status** ✓ ✓

| | 1 | 2 |
|---|---|---|
| **version 1** 23 Oct 2023 | ✓ view | ✓ view |

1. **Simo Njabulo Maduna**, Norwegian Institute of Bioeconomy Research, Svanvik, Norway

2. **Naciye Sena Çağatay**, University of Liverpool, Liverpool, UK

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; **Hammond J**: 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 (206194) and the Darwin Tree of Life Discretionary Award (218328).
*The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.*



**How to cite this article:** Boyes D, Hammond J, University of Oxford and Wytham Woods Genome Acquisition Lab *et al.* **The genome sequence of the Ruddy Flat-body, *Agonopterix subpropinquella* (Stainton, 1849) [version 1; peer review: 2 approved]** Wellcome Open Research 2023, **8**:487 https://doi.org/10.12688/wellcomeopenres.20170.1

**First published:** 23 Oct 2023, **8**:487 https://doi.org/10.12688/wellcomeopenres.20170.1

## Species taxonomy

Eukaryota; Metazoa; Eumetazoa; Bilateria; Protostomia; Ecdysozoa; Panarthropoda; Arthropoda; Mandibulata; Pancrustacea; Hexapoda; Insecta; Dicondylia; Pterygota; Neoptera; Endopterygota; Amphiesmenoptera; Lepidoptera; Glossata; Neolepidoptera; Heteroneura; Ditrysia; Gelechioidea; Depressariidae; Depressariinae; *Agonopterix*; *Agonopterix subpropinquella* (Stainton, 1849) (NCBI:txid1857958).

## Background

*Agonopterix subpropinquella*, the Ruddy Flat-body, is a small moth of the Depressariidae family. The species has a pan-European, mainly coastal, distribution (GBIF Secretariat, 2023), and is found widely across Britain and Ireland (Harper *et al.*, 2002). Like many species in its genus, the imago is drab-coloured and indistinctly patterned, indeed the species' scientific name likely being a reference to its similarity to other members of the genus (Emmet, 1991). However, the species also shows a distinctive form f. *rhodochrella,* with exaggerated blackish colouring on the head, thorax, and forewing (Harper *et al.*, 2002).

In common with many members of its genus, the species overwinters as an adult (Harper *et al.*, 2002). Adults are on the wing between August and May, hibernating over the winter, and may be disturbed from their hibernation by beating thatch or dense vegetation over the winter (Sterling & Parsons, 2018). Eggs are lain in May on knapweeds (*Centaurea* spp.) or thistles (*Cirsium* spp.) (Harper *et al.*, 2002). The larva is green and initially mines the foodplant before feeding in a silken spinning (Harper *et al.*, 2002). Larvae feed between June and July, pupating from July to August in earth or amongst detritus (Harper *et al.*, 2002).

The genome of the ruddy flat-body, *Agonopterix subpropinquella*, was sequenced as part of the Darwin Tree of Life Project, a collaborative effort to sequence all named eukaryotic species in the Atlantic Archipelago of Britain and Ireland. Here we present a chromosomally complete genome sequence for *Agonopterix subpropinquella*, based on one male specimen from Wytham Woods, Oxfordshire.

## Genome sequence report

The genome was sequenced from one male *Agonopterix subpropinquella* (Figure 1) collected from Wytham Woods, Oxfordshire, UK (51.77, –1.34). A total of 28-fold coverage in Pacific Biosciences single-molecule HiFi long reads and 63-fold coverage in 10X Genomics read clouds were generated. Primary assembly contigs were scaffolded with chromosome conformation Hi-C data. Manual assembly curation corrected 56 missing joins or mis-joins and removed 12 haplotypic duplications, reducing the assembly length by 0.43% and the scaffold number by 50.79%.

The final assembly has a total length of 667.9 Mb in 31 sequence scaffolds with a scaffold N50 of 25.1 Mb (Table 1). Most (99.98%) of the assembly sequence was assigned to 28 chromosomal-level scaffolds, representing 27 autosomes and the Z sex chromosome. A summary of the assembly statistics is shown in Figure 2, 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. 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.

**Figure 1. Photograph of the *Agonopterix subpropinquella* (ilAgoSubp1) specimen used for genome sequencing.**

The estimated Quality Value (QV) of the final assembly is 56.5 with *k*-mer completeness of 99.99%, and the assembly has a BUSCO v5.3.2 completeness of 98.7% (single = 98.0%, duplicated = 0.7%), using the lepidoptera_odb10 reference set ($n$ = 5,286).

Metadata for specimens, spectral estimates, sequencing runs, contaminants and pre-curation assembly statistics can be found at https://links.tol.sanger.ac.uk/species/1857958.

## Genome annotation report

The *Agonopterix subpropinquella* genome assembly (GCA_922987775.1) was annotated using the Ensembl rapid annotation pipeline (Table 1; https://rapid.ensembl.org/Agonopterix_subpropinquella_GCA_922987775.1/Info/Index). The resulting annotation includes 18,967 transcribed mRNAs from 18,796 protein-coding genes.

## Methods

### Sample acquisition and nucleic acid extraction

A male *Agonopterix subpropinquella* (specimen ID Ox000822, ToLID ilAgoSubp1) was collected from Wytham Woods, Oxfordshire (biological vice-county Berkshire), UK (latitude 51.77, longitude –1.34) on 2020-08-01 using a light trap. The specimen was collected and identified by Douglas Boyes (University of Oxford) and preserved on dry ice.

DNA was extracted at the Tree of Life laboratory, Wellcome Sanger Institute (WSI). The ilAgoSubp1 sample was weighed and dissected on dry ice with tissue set aside for Hi-C

**Table 1. Genome data for *Agonopterix subpropinquella*, ilAgoSubp1.1.**

| **Project accession data** | | |
|---|---|---|
| Assembly identifier | ilAgoSubp1.1 | |
| Assembly release date | 2021-12-19 | |
| Species | *Agonopterix subpropinquella* | |
| Specimen | ilAgoSubp1 | |
| NCBI taxonomy ID | 1857958 | |
| BioProject | PRJEB47465 | |
| BioSample ID | SAMEA7746629 | |
| Isolate information | ilAgoSubp1, male: whole organism (DNA sequencing and Hi-C data) | |
| **Assembly metrics*** | | ***Benchmark*** |
| Consensus quality (QV) | 56.5 | *≥ 50* |
| *k*-mer completeness | 99.99% | *≥ 95%* |
| BUSCO** | C:98.7%[S:98.0%,D:0.7%], F:0.2%,M:1.1%,n:5,286 | *C ≥ 95%* |
| Percentage of assembly mapped to chromosomes | 99.98% | *≥ 95%* |
| Sex chromosomes | Z chromosome | *localised homologous pairs* |
| Organelles | Mitochondrial genome assembled | *complete single alleles* |
| **Raw data accessions** | | |
| PacificBiosciences SEQUEL II | ERR6808069, ERR6939284 | |
| 10X Genomics Illumina | ERR6747928, ERR6747929, ERR6747930, ERR6747931 | |
| Hi-C Illumina | ERR6747927 | |
| **Genome assembly** | | |
| Assembly accession | GCA_922987775.1 | |
| *Accession of alternate haplotype* | GCA_922987765.1 | |
| Span (Mb) | 667.9 | |
| Number of contigs | 95 | |
| Contig N50 length (Mb) | 11.6 | |
| Number of scaffolds | 31 | |
| Scaffold N50 length (Mb) | 25.1 | |
| Longest scaffold (Mb) | 69.3 | |
| **Genome annotation** | | |
| Number of protein-coding genes | 18,796 | |
| Number of gene transcripts | 18,967 | |

* 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 v5.3.2. 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/Agonopterix%20subpropinquella/dataset/CAKLPO01.1/busco.

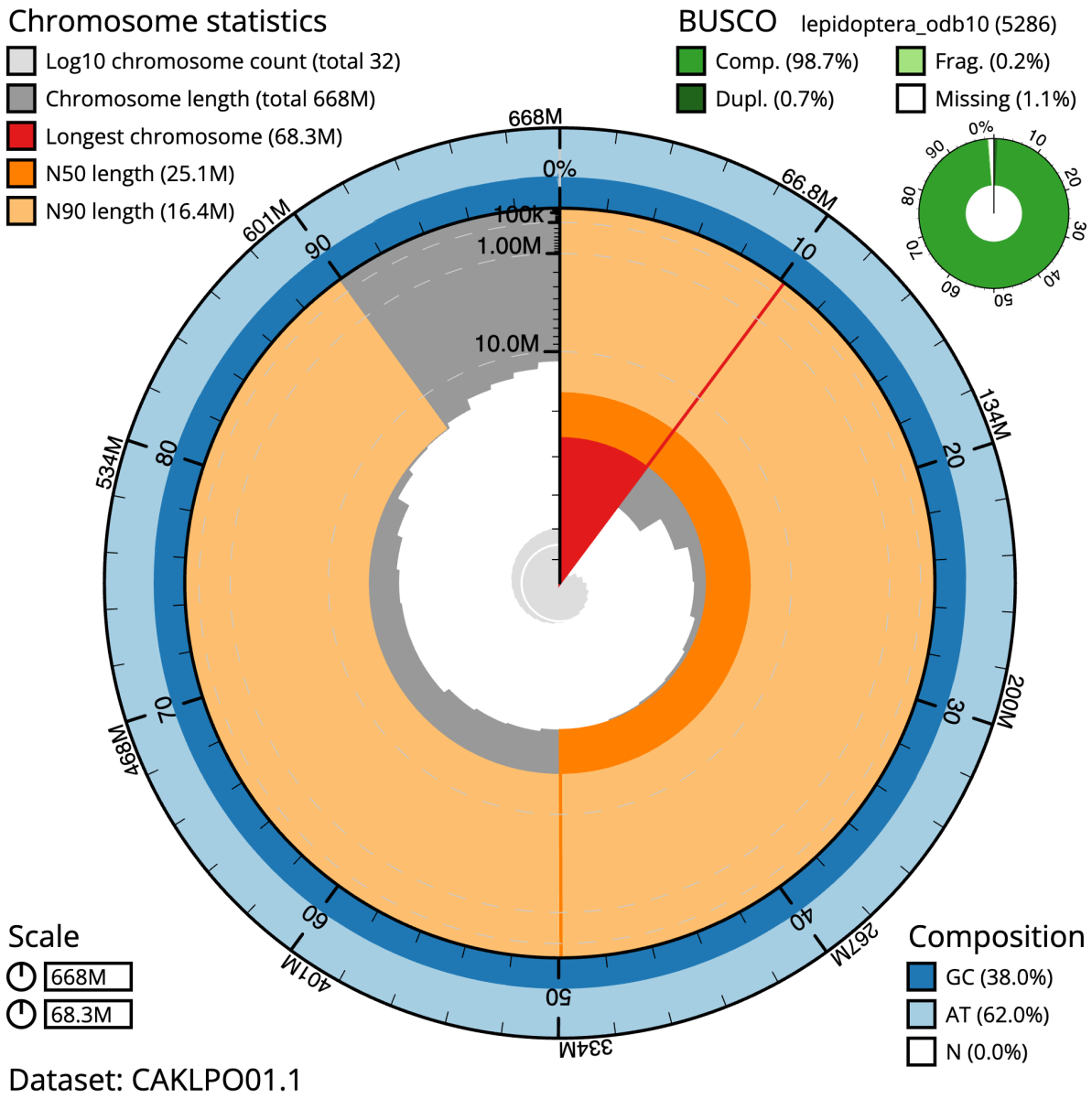


**Figure 2. Genome assembly of *Agonopterix subpropinquella*, ilAgoSubp1.1: metrics.** The BlobToolKit Snailplot shows N50 metrics and BUSCO gene completeness. The main plot is divided into 1,000 size-ordered bins around the circumference with each bin representing 0.1% of the 667,905,724 bp assembly. The distribution of scaffold lengths is shown in dark grey with the plot radius scaled to the longest scaffold present in the assembly (68,253,362 bp, shown in red). Orange and pale-orange arcs show the N50 and N90 scaffold lengths (25,090,608 and 16,437,072 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/Agonopterix%20subpropinquella/dataset/CAKLPO01.1/snail.

sequencing. Tissue from the whole organism was disrupted using a Nippi Powermasher fitted with a BioMasher pestle. High molecular weight (HMW) DNA was extracted using the Qiagen MagAttract HMW DNA extraction kit. Low molecular weight DNA was removed from a 20 ng aliquot of extracted DNA using the 0.8X AMpure XP purification kit prior to 10X Chromium sequencing; a minimum of 50 ng DNA was submitted for 10X sequencing. HMW DNA was sheared into an average fragment size of 12–20 kb in a Megaruptor 3 system with speed setting 30. Sheared DNA was purified by solid-phase reversible immobilisation using AMPure PB beads with a 1.8X ratio of beads to sample to remove the shorter fragments and concentrate the DNA sample. The concentration of the sheared and purified DNA was assessed using a

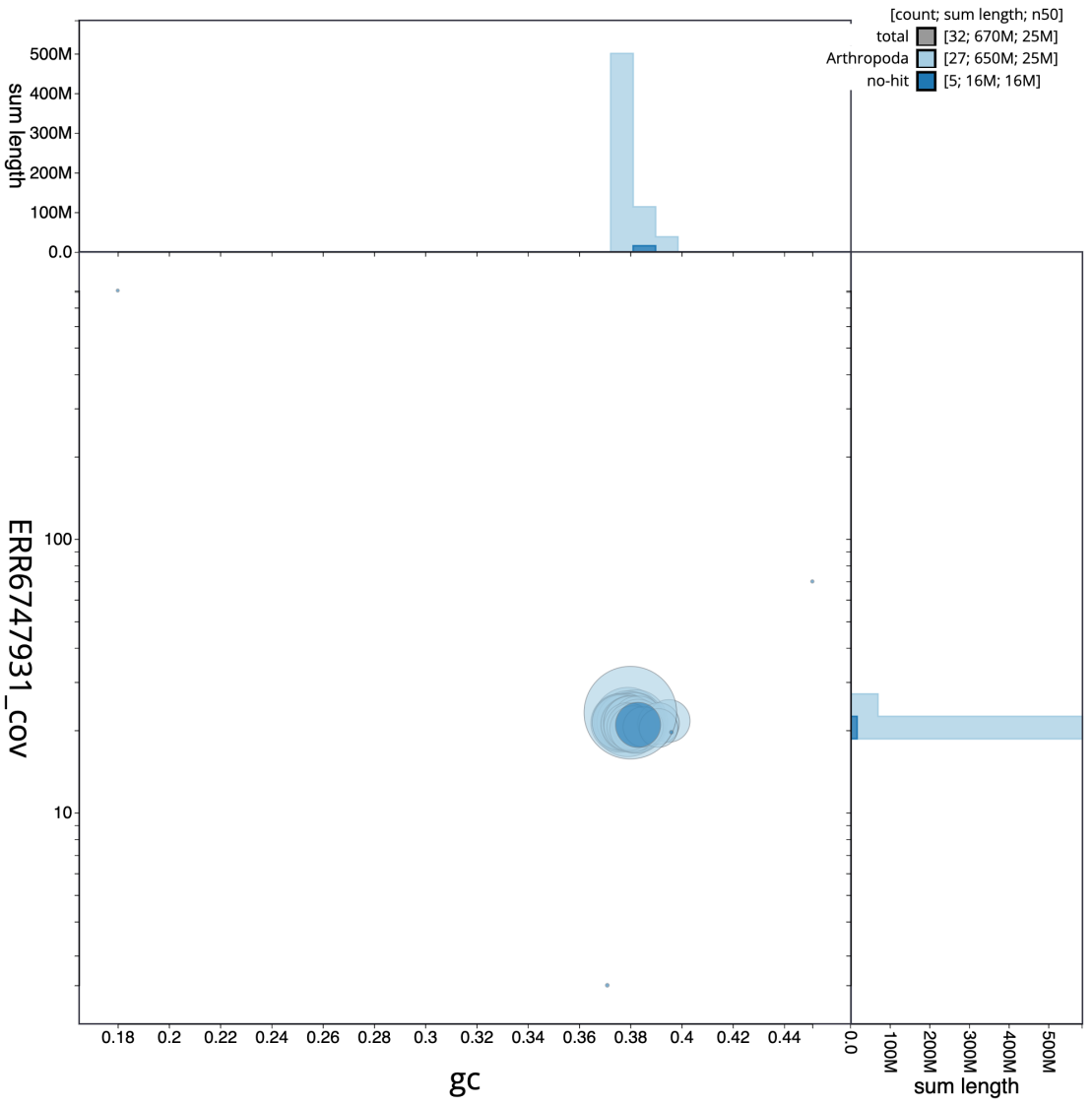


**Figure 3. Genome assembly of *Agonopterix subpropinquella*, ilAgoSubp1.1: BlobToolKit GC-coverage plot.** Scaffolds are coloured by phylum. Circles are sized in proportion to scaffold length. Histograms show the distribution of scaffold length sum along each axis. An interactive version of this figure is available at https://blobtoolkit.genomehubs.org/view/Agonopterix%20subpropinquella/dataset/CAKLPO01.1/blob.

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.

## Sequencing

Pacific Biosciences HiFi circular consensus and 10X Genomics read cloud DNA sequencing libraries were constructed according to the manufacturers' instructions. DNA sequencing was performed by the Scientific Operations core at the WSI on Pacific Biosciences SEQUEL II (HiFi) and Illumina NovaSeq 6000 (10X) instruments. Hi-C data were also generated from remaining tissue of ilAgoSubp1 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 as described previously (Howe *et al.*, 2021). Manual curation was performed using HiGlass (Kerpedjiev *et al.*, 2018) and Pretext (Harry, 2022). The mitochondrial genome was assembled using MitoHiFi (Uliano-Silva *et al.*, 2023), which

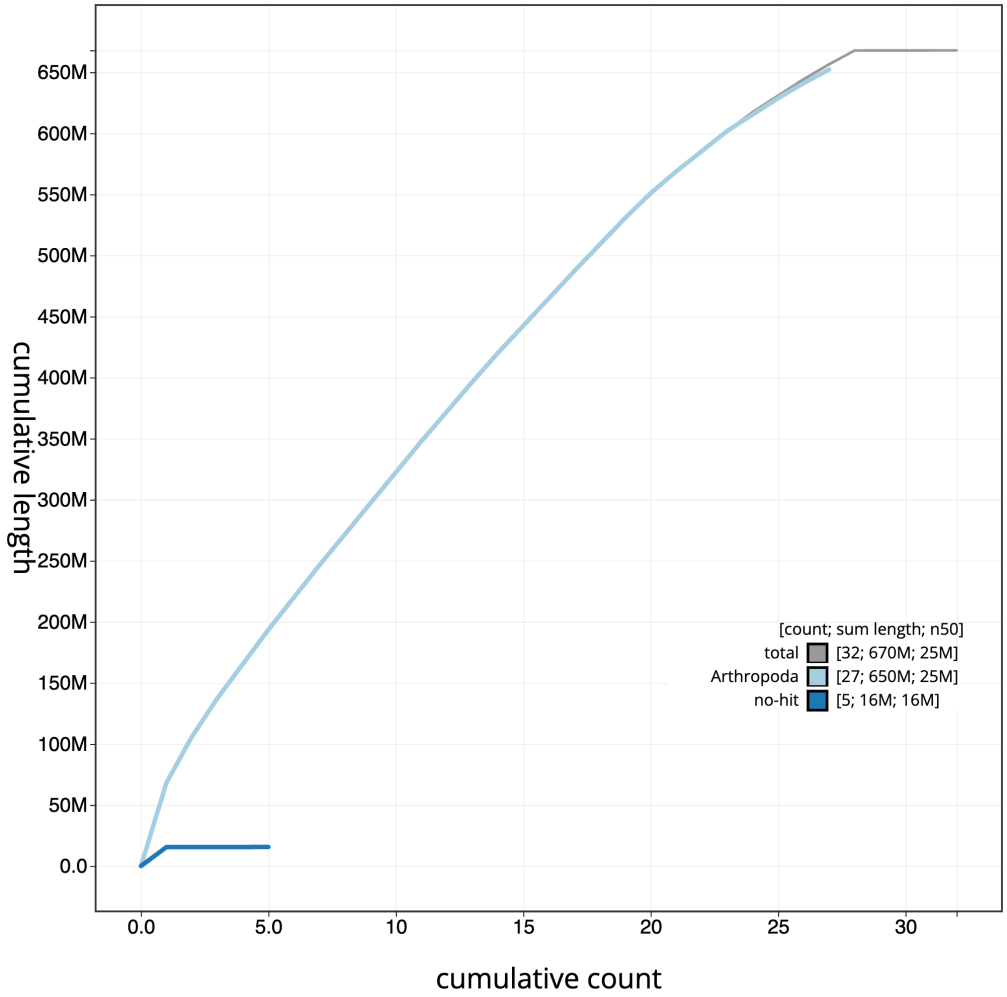


**Figure 4. Genome assembly of *Agonopterix subpropinquella*, ilAgoSubp1.1: BlobToolKit cumulative sequence plot.** The grey line shows cumulative length for all scaffolds. Coloured lines show cumulative lengths of scaffolds assigned to each phylum using the buscogenes taxrule. An interactive version of this figure is available at https://blobtoolkit.genomehubs.org/view/Agonopterix%20subpropinquella/dataset/CAKLPO01.1/cumulative.

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 *Agonopterix subpropinquella* assembly (GCA_922987775.1) 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.

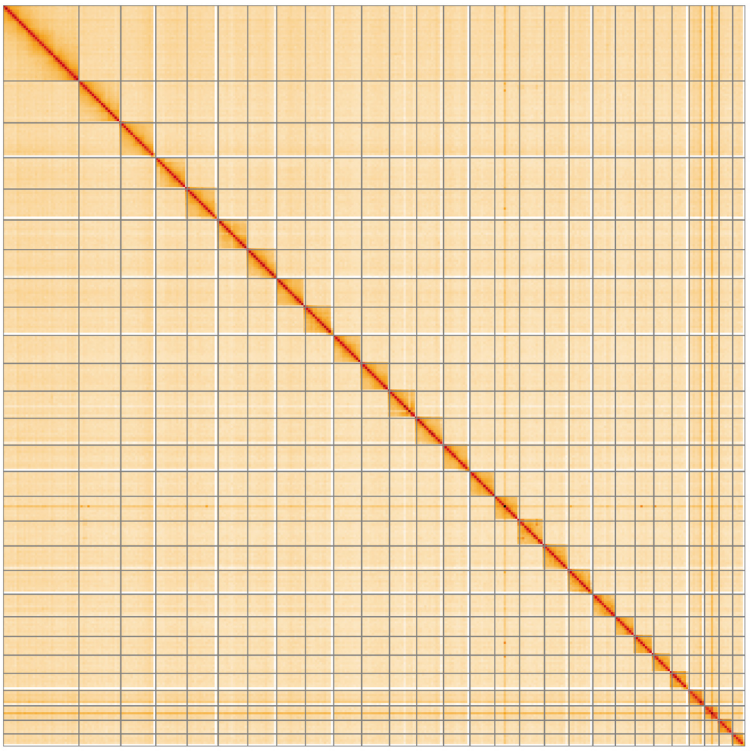

**Figure 5. Genome assembly of *Agonopterix subpropinquella*, ilAgoSubp1.1: Hi-C contact map of the ilAgoSubp1.1 assembly, visualised using HiGlass.** Chromosomes are shown in order of size from left to right and top to bottom. An interactive version of this figure may be viewed at https://genome-note-higlass.tol.sanger.ac.uk/l/?d=RQh-QoTJRnastq9ymSznYA.

**Table 2. Chromosomal pseudomolecules in the genome assembly of *Agonopterix subpropinquella*, ilAgoSubp1.**

| INSDC accession | Chromosome | Length (Mb) | GC% |
|---|---|---|---|
| OV277694.1 | 1 | 37.69 | 38.0 |
| OV277695.1 | 2 | 31.47 | 38.0 |
| OV277696.1 | 3 | 28.22 | 38.0 |
| OV277697.1 | 4 | 27.9 | 38.0 |
| OV277698.1 | 5 | 26.69 | 37.5 |
| OV277699.1 | 6 | 26.05 | 38.0 |
| OV277700.1 | 7 | 25.78 | 38.0 |
| OV277701.1 | 8 | 25.49 | 37.5 |
| OV277702.1 | 9 | 25.12 | 38.0 |
| OV277703.1 | 10 | 25.09 | 38.0 |
| OV277704.1 | 11 | 24.47 | 37.5 |
| OV277705.1 | 12 | 24.19 | 37.5 |
| OV277706.1 | 13 | 23.65 | 38.0 |
| OV277707.1 | 14 | 22.38 | 38.0 |
| OV277708.1 | 15 | 22.37 | 38.0 |
| OV277709.1 | 16 | 22.35 | 38.0 |
| OV277710.1 | 17 | 22.07 | 38.0 |
| OV277711.1 | 18 | 21.46 | 38.5 |
| OV277712.1 | 19 | 20.32 | 38.5 |
| OV277713.1 | 20 | 17.76 | 38.0 |
| OV277714.1 | 21 | 16.71 | 38.0 |
| OV277715.1 | 22 | 16.44 | 38.0 |
| OV277716.1 | 23 | 15.51 | 38.5 |
| OV277717.1 | 24 | 13.79 | 39.5 |
| OV277718.1 | 25 | 13.08 | 39.0 |
| OV277719.1 | 26 | 12.23 | 38.5 |
| OV277720.1 | 27 | 11.25 | 39.0 |
| OV277693.1 | Z | 68.25 | 38.0 |
| OV277721.1 | MT | 0.02 | 18.0 |

**Table 3. Software tools: versions and sources.**

| Software tool | Version | Source |
|---|---|---|
| BlobToolKit | 4.0.7 | 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 |
| Hifiasm | 0.15.3-r339 | 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 | 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 |
| 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: *Agonopterix subpropinquella* (ruddy flat-body). Accession number PRJEB47465; https://identifiers.org/ena.embl/PRJEB47465. (Wellcome Sanger Institute, 2021) The genome sequence is released openly for reuse. The *Agonopterix subpropinquella* 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.4789928.

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 programme are listed here: https://doi.org/10.5281/zenodo.4783585.

Members of Wellcome Sanger Institute Scientific Operations: DNA Pipelines collective are listed here: https://doi.org/10.5281/zenodo.4790455.

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

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

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

Emmet AM: **The Scientific Names of the British Lepidoptera - their History and Meaning.** Colchester: Harley Books, 1991.
Reference Source

Garrison E, Marth G: **Haplotype-based variant detection from short-read sequencing.** 2012.
Publisher Full Text

GBIF Secretariat: ***Agonopterix subpropinquella* Stainton, 1849.** *GBIF Backbone Taxonomy. Checklist dataset.* 2023; [Accessed 29 September 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

Harper MW, Langmaid JR, Emmet AM: **Oecophoridae.** In: Emmet, A. M. and Langmaid, J. R. (eds.) *The Moths and Butterflies of Great Britain and Ireland. Vol. 4 Pt. 1): Volume 4 Part 1 (Oecophoridae to Scythrididae, excluding Gelechiidae).* Colchester, UK: Harley Books, 2002.
Reference Source

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

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

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

Sterling P, Parsons M: **Field Guide to the Micro-moths of Great Britain and Ireland.** London: Bloomsbury, 2018.

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].
Reference Source

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 Md, 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 Ruddy Flat-body, *Agonopterix subpropinquella* (Stainton, 1849).** European Nucleotide Archive. [dataset], accession number PRJEB47465, 2021.

# Open Peer Review

## Current Peer Review Status: ✔ ✔

**Version 1**

Reviewer Report 06 May 2024

https://doi.org/10.21956/wellcomeopenres.22333.r75821



✔ **Naciye Sena Çağatay**
University of Liverpool, Liverpool, England, UK

The complete genome of the Ruddy Flat-body *Agonopterix subpropinquella* was assembled with Boyes and Hammond using PacBio HiFi sequencing, 10X Genomics and Hi-C Illumina sequencing techniques. The analysis were explained very well in each method. The genome assembly is of high quality and the figures shown are informative. I have a minor revision below.
Background
Can you check these words.
Like many species in its genus, the imago is ”image”
Eggs are lain in May on knapweeds “laid”
Q1:
I was just wondering, there is no information about morphological identification. How do you decide this species? Why are you using male sample?

**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:*** Molecular entomology

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

Reviewer Report 27 March 2024

https://doi.org/10.21956/wellcomeopenres.22333.r75813



**Simo Njabulo Maduna**
Norwegian Institute of Bioeconomy Research, Svanvik, Norway

The near-complete genome of the Ruddy Flat-body *Agonopterix subpropinquella* was assembled by Boyes and Hammond through the utilization of BacBio HiFi sequencing, as well as 10X Genomics and Hi-C Illumina sequencing techniques. The authors present the initial and superior reference genome for the Ruddy Flat-body. Their manuscript is skilfully composed and includes an elaborate bioinformatics pipeline and succinct findings, accompanied by genome assembly metric benchmarks to facilitate comprehension. Below I present minor revisions.

ABSTRACT

Include a sentence on the sequencing strategy employed to generate the reported chromosomal-level genome assembly for the target species.

BACKGROUND

Kindly incorporate a sentence regarding the genome sizes of moths, accompanied by a few examples, to provide the reader with an understanding of the average genome size of moths.

Change “...Like many species in its genus...” to “...Like many of its congeners...”
Question: What was the rationale behind selecting a homogametic male for sequencing rather than a heterogametic female? The reviewer is merely inquisitive given that the research gap is now on the assembly of the W sex chromosome.

**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:*** Genome Assembly; Ecological Genomics; Conservation Genomics; Invasive Genomics; Population Genomics; Phylogenomics; Molecular Phylogenetics

**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.**