DATA NOTE

The genome sequence of the Rosy Footman, *Miltochrista miniata* (Forster, 1771) [version 1; peer review: awaiting peer review]

Callum J. Macgregor[1,2], Ilik Saccheri[3], Barry Fox[1,4], Douglas Boyes[5+],
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]Leverhulme Centre for Anthropocene Biodiversity, University of York, York, England, UK
[2]British Trust for Ornithology Cymru, Bangor, Wales, UK
[3]Department of Evolution, Ecology and Behaviour, University of Liverpool, Liverpool, England, UK
[4]Rothamsted Research, Rothamsted Research, England, UK
[5]UK Centre for Ecology & Hydrology, Wallingford, England, UK

[+] Deceased author

v1 **First published:** 28 Dec 2023, **8**:582
https://doi.org/10.12688/wellcomeopenres.20505.1
Latest published: 28 Dec 2023, **8**:582
https://doi.org/10.12688/wellcomeopenres.20505.1

Abstract

We present a genome assembly from an individual female *Miltochrista miniata* (the Rosy Footman; Arthropoda; Insecta; Lepidoptera; Erebidae). The genome sequence is 1057.0 megabases in span. Most of the assembly is scaffolded into 32 chromosomal pseudomolecules, including the W and Z sex chromosomes. The mitochondrial genome has also been assembled and is 15.93 kilobases in length. Gene annotation of this assembly on Ensembl identified 23,879 protein coding genes.

Keywords

Miltochrista miniata, rosy footman, 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.

This article is included in the Tree of Life gateway.

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

Author roles: Macgregor CJ: Investigation, Resources, Writing – Original Draft Preparation, Writing – Review & Editing; **Saccheri I**: Investigation, Resources, Writing – Review & Editing; **Fox B**: Investigation, Resources; **Boyes D**: Investigation, Resources;

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: Macgregor CJ, Saccheri I, Fox B *et al.* **The genome sequence of the Rosy Footman, *Miltochrista miniata* (Forster, 1771) [version 1; peer review: awaiting peer review]** Wellcome Open Research 2023, **8**:582 https://doi.org/10.12688/wellcomeopenres.20505.1

First published: 28 Dec 2023, **8**:582 https://doi.org/10.12688/wellcomeopenres.20505.1

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; Erebidae; Arctiinae; Lithosiini; *Miltochrista*; *Miltochrista miniata* (Forster, 1771) (NCBI:txid987980).

Background

The Rosy Footman, *Miltochrista miniata* (Forster, 1771), is a small moth of the family Erebidae (Figure 1). It is distributed across the temperate Palaearctic in Europe and Asia, but has been split from *Miltochrista rosaria* Butler, 1877, which was formerly considered to be an eastern Asian aberration of *M. miniata*. It is restricted to southern counties within the UK (Randle *et al.*, 2019). The larvae feed on lichens, and therefore the species may be found in a range of habitats, but especially forests. A lichen polyphenolic, physodic acid, is sequestered by larvae and detectable in adults, suggesting that the bright colouration from which the species gets its English name may be an honest aposematic signal (Scott *et al.*, 2014). Abundance and distribution within the UK have both increased, reflecting the recent positive trends for many other lichenivorous species (Conrad *et al.*, 2004), which may be linked to improvements in air quality including reductions in sulphur dioxide pollution (Pescott *et al.*, 2015). It is also significantly advancing its phenology (Macgregor *et al.*, 2019). The chromosome number has been described as 31 (Belyakova & Lukhtanov, 1996).

Here we present a chromosomally complete genome sequence for *Miltochrista miniata*, 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.

Figure 1. Image of *Miltochrista miniata* (not the specimen used for genome sequencing). Photograph by Kurt Kulac (CC-BY-SA-2.5).

Genome sequence report

The genome was sequenced from one female *Miltochrista miniata* collected from Winterslow, England (51.09, 1.64). A total of 34-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 72 missing joins or mis-joins and removed 28 haplotypic duplications, reducing the assembly length by 1.64% and the scaffold number by 5.22%, and reducing the scaffold N50 by 0.98%.

The final assembly has a total length of 1057.0 Mb in 109 sequence scaffolds with a scaffold N50 of 37.5 Mb (Table 1). 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 (99.59%) of the assembly sequence was assigned to 32 chromosomal-level scaffolds, representing 30 autosomes and the W and Z sex chromosomes. 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 64.2 with *k*-mer completeness of 100%, and the assembly has a BUSCO v5.3.2 completeness of 98.8% (single = 97.9%, duplicated = 0.9%), 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/987980.

Genome annotation report

The *Miltochrista miniata* genome assembly (GCA_933228765.1) was annotated using the Ensembl rapid annotation pipeline (Table 1; https://rapid.ensembl.org/Miltochrista_miniata_GCA_933228765.1/Info/Index). The resulting annotation includes 24,074 transcribed mRNAs from 23,879 protein-coding genes.

Methods

Sample acquisition and nucleic acid extraction

Two female *Miltochrista miniata* specimens were collected from Winterslow, England, UK (latitude 51.09, longitude 1.64) on 2016-07-27, using a light trap. The specimens were collected by Barry Fox (University of York) and identified by Ilik Saccheri (University of Liverpool) and stored in a –80 °C freezer. Specimen ID SAN0001385, (ToLID ilMilMini1) was used for DNA sequencing and specimen ID SAN0001386, (ToLID ilMilMini3) was used for Hi-C sequencing. The specimen used for RNA sequencing (specimen ID Ox001696, ToLID ilMilMini5) was collected

Table 1. Genome data for *Miltochrista miniata*, ilMilMini2.1.

Project accession data		
Assembly identifier	ilMilMini2.1	
Species	*Miltochrista miniata*	
Specimen	ilMilMini2	
NCBI taxonomy ID	987980	
BioProject	PRJEB50744	
BioSample ID	SAMEA9252593	
Isolate information	ilMilMini2, female: whole organism (DNA) ilMilMini3, female: whole organism (Hi-C data) ilMilMini5: abdomen (RNA sequencing)	
Assembly metrics*		***Benchmark***
Consensus quality (QV)	64.2	*≥ 50*
k-mer completeness	100%	*≥ 95%*
BUSCO**	C:98.8%[S:97.9%,D:0.9%], F:0.3%,M:0.9%,n:5,286	*C ≥ 95%*
Percentage of assembly mapped to chromosomes	99.59%	*≥ 95%*
Sex chromosomes	W and Z chromosomes	*localised homologous pairs*
Organelles	Mitochondrial genome assembled	*complete single alleles*
Raw data accessions		
PacificBiosciences SEQUEL II	ERR8575383, ERR8575380, ERR8575381, ERR8575382	
10X Genomics Illumina	ERR8571666, ERR8571667, ERR8571669, ERR8571668	
Hi-C Illumina	ERR8571670	
PolyA RNA-Seq Illumina	ERR10890680	
Genome assembly		
Assembly accession	GCA_933228765.1	
Accession of alternate haplotype	GCA_933228775.1	
Span (Mb)	1,057.0	
Number of contigs	184	
Contig N50 length (Mb)	16.5	
Number of scaffolds	109	
Scaffold N50 length (Mb)	37.5	
Longest scaffold (Mb)	47.8	
Genome annotation		
Number of protein-coding genes	23,879	
Number of gene transcripts	24,074	

* 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/Miltochrista%20miniata/dataset/CAKOGA01/busco.

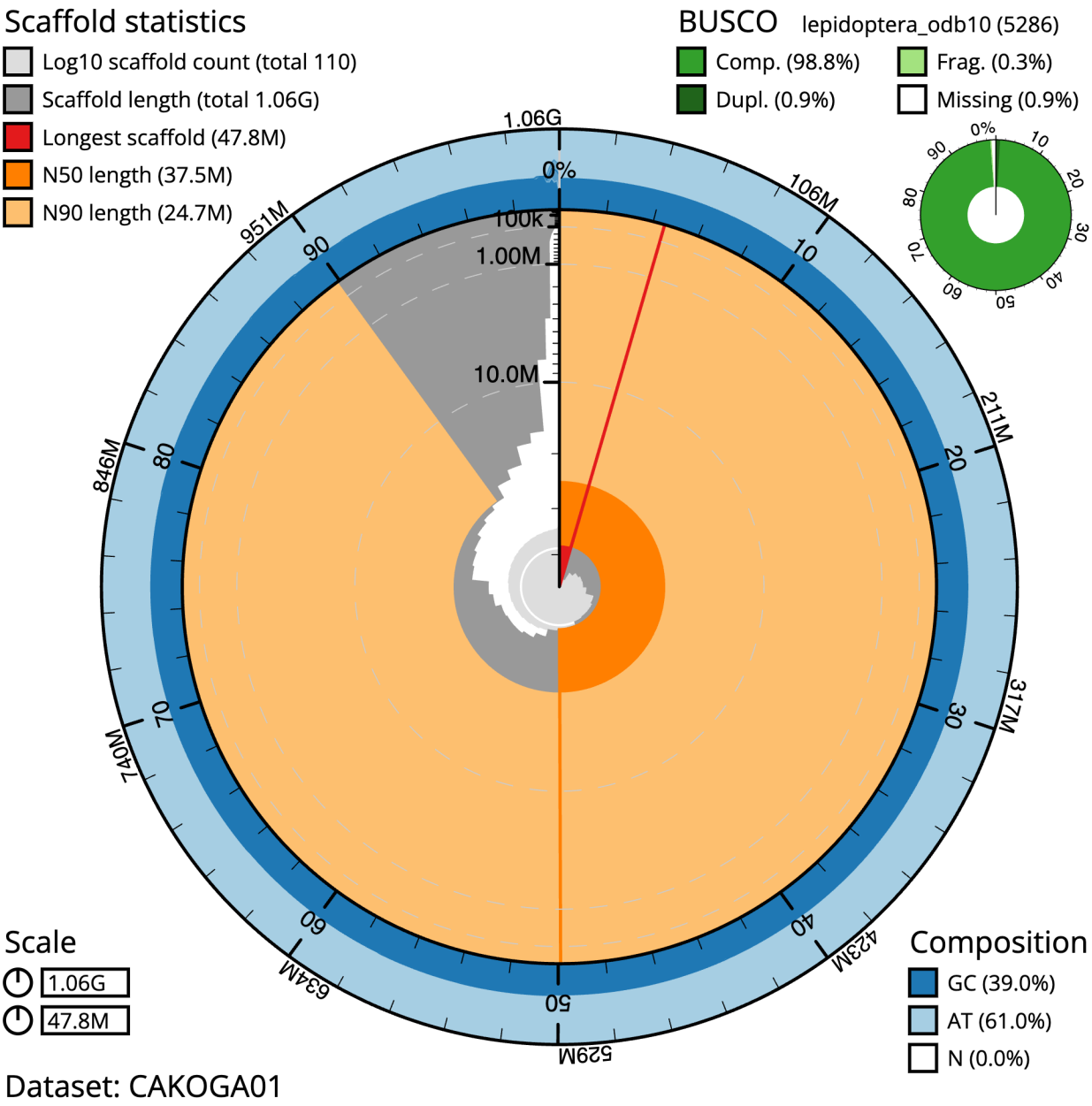

Figure 2. Genome assembly of *Miltochrista miniata*, ilMilMini2.1: metrics. The BlobToolKit Snailplot shows N50 metrics and BUSCO gene completeness. The main plot is divided into 1,000 size-ordered bins around the circumference with each bin representing 0.1% of the 1,057,002,560 bp assembly. The distribution of scaffold lengths is shown in dark grey with the plot radius scaled to the longest scaffold present in the assembly (47,752,028 bp, shown in red). Orange and pale-orange arcs show the N50 and N90 scaffold lengths (37,487,092 and 24,710,470 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/Miltochrista%20miniata/dataset/CAKOGA01/snail.

from Wytham Woods (latitude 51.77, longitude –1.34) on 2021-07-17 by Douglas Boyes.

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; HMW DNA fragmentation; and fragmented DNA clean-up. The ilMilMini1 sample was weighed and dissected on dry ice with tissue set aside for Hi-C sequencing (as per the protocol at https://dx.doi.org/10.17504/protocols.io.x54v9prmqg3e/v1). For sample homogenisation, tissue was homogenised using a Nippi Powermasher

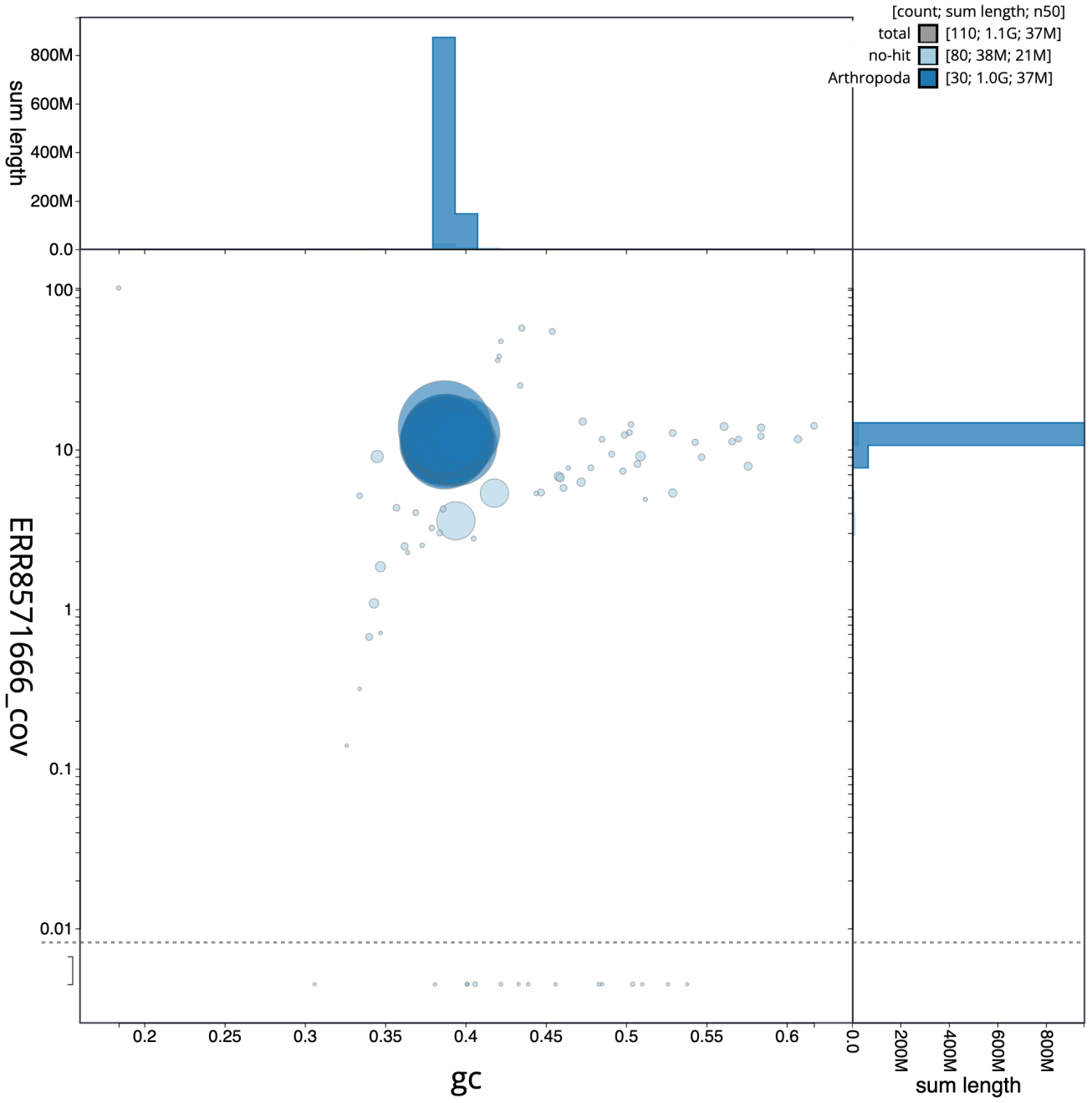

Figure 3. Genome assembly of *Miltochrista miniata*, ilMilMini2.1: BlobToolKit GC-coverage plot. Scaffolds are coloured by phylum. Circles are sized in proportion to scaffold length. Histograms show the distribution of scaffold length sum along each axis. An interactive version of this figure is available at https://blobtoolkit.genomehubs.org/view/Miltochrista%20miniata/dataset/CAKOGA01/blob.

fitted with a BioMasher pestle, following the protocol at https://dx.doi.org/10.17504/protocols.io.5qpvo3r19v4o/v1. HMW DNA was extracted by means of the Automated MagAttract protocol (https://dx.doi.org/10.17504/protocols.io.kxygx3y4dg8j/v1). HMW DNA was sheared into an average fragment size of 12–20 kb in a Megaruptor 3 system with speed setting 30, following the HMW DNA Fragmentation: Diagenode Megaruptor®3 for PacBio HiFi protocol (https://dx.doi.org/10.17504/protocols.io.8epv5x2zjg1b/v1). Sheared DNA was purified by solid-phase reversible immobilisation (SPRI) protocol (protocol at https://dx.doi.org/10.17504/protocols.io.kxygx3y1dg8j/v1). 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

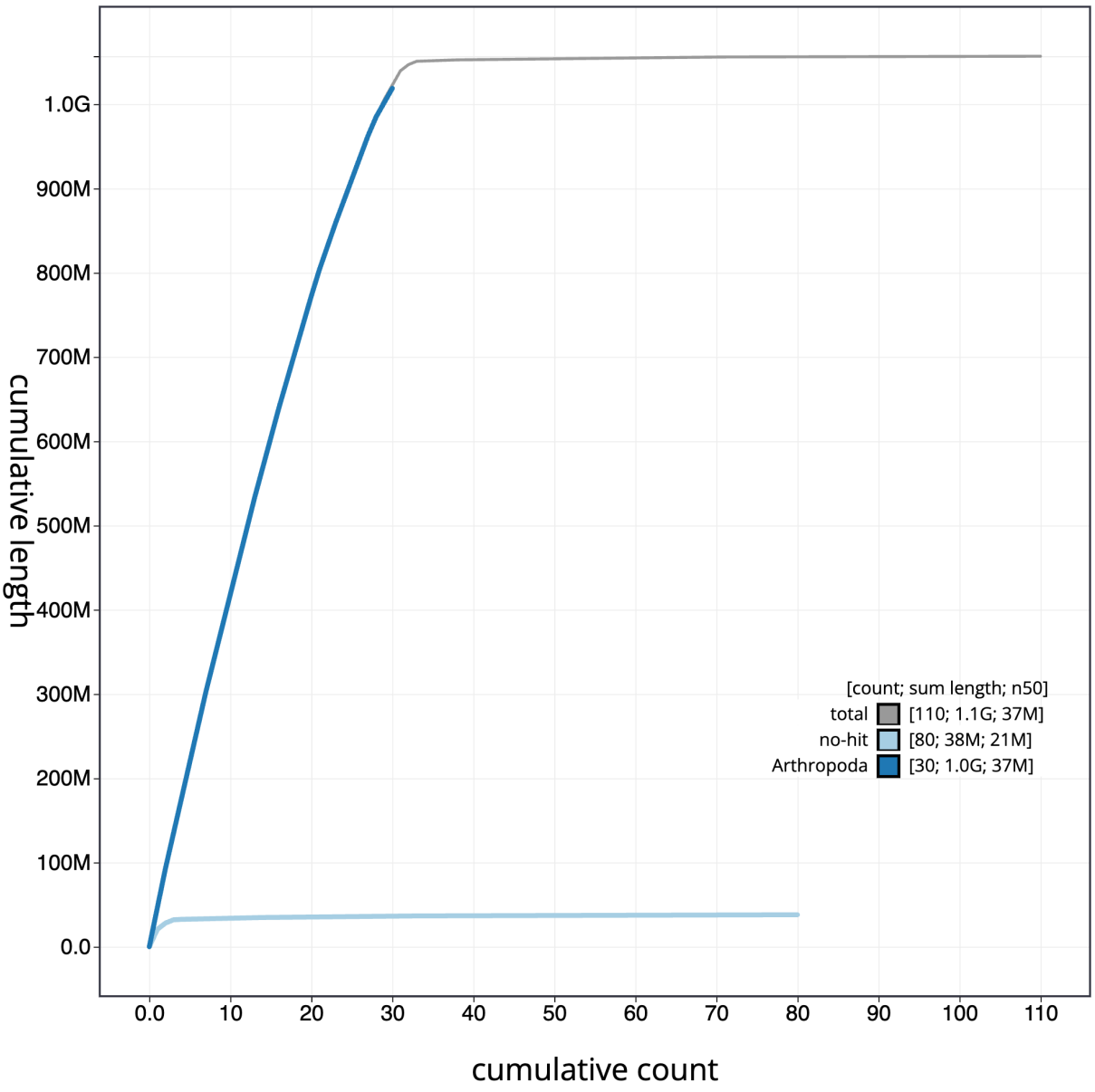

Figure 4. Genome assembly of *Miltochrista miniata*, ilMilMini2.1: BlobToolKit cumulative sequence plot. The grey line shows cumulative length for all scaffolds. Coloured lines show cumulative lengths of scaffolds assigned to each phylum using the buscogenes taxrule. An interactive version of this figure is available at https://blobtoolkit.genomehubs.org/view/Miltochrista%20miniata/dataset/CAKOGA01/cumulative.

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 abdomen tissue of ilMilMini5 in the Tree of Life Laboratory at the WSI using the RNA Extraction: Automated MagMax™ *mir*Vana protocol (https://dx.doi.org/10.17504/protocols.io.6qpvr36n3vmk/v1). The RNA concentration was assessed using a Nanodrop spectrophotometer and Qubit Fluorometer using the Qubit RNA Broad-Range (BR) 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 Tree of Life laboratory are publicly available on protocols.io: https://dx.doi.org/10.17504/protocols.io.8epv5xxy6g1b/v1.

Sequencing

Pacific Biosciences HiFi circular consensus and 10X Genomics read cloud DNA sequencing libraries were constructed

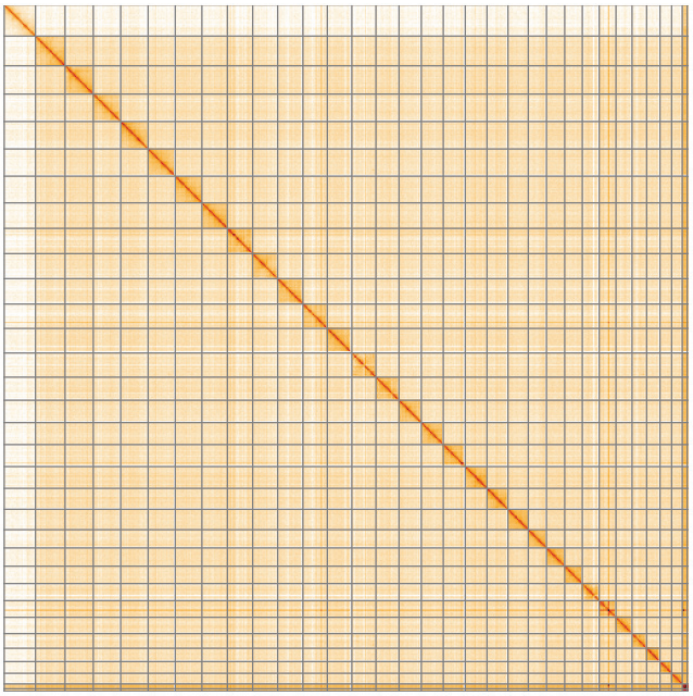

Figure 5. Genome assembly of *Miltochrista miniata*, ilMilMini2.1: Hi-C contact map of the ilMilMini2.1 assembly, visualised using HiGlass. Chromosomes are shown in order of size from left to right and top to bottom. An interactive version of this figure may be viewed at https://genome-note-higlass.tol.sanger.ac.uk/l/?d=KvrRZPsGQx-RHC5ayviYDw.

Table 2. Chromosomal pseudomolecules in the genome assembly of *Miltochrista miniata*, ilMilMini2.

INSDC accession	Chromosome	Length (Mb)	GC%
OW121748.1	1	45.43	39.0
OW121749.1	2	43.25	39.0
OW121750.1	3	42.32	38.5
OW121751.1	4	41.83	39.0
OW121752.1	5	41.71	38.5
OW121753.1	6	40.89	38.5
OW121754.1	7	39.01	38.5
OW121755.1	8	37.49	38.5
OW121756.1	9	38.64	38.5
OW121757.1	10	38.6	38.5
OW121758.1	11	38.44	38.5
OW121759.1	12	37.68	38.5
OW121760.1	13	37.07	39.5
OW121761.1	14	35.35	39.0
OW121762.1	15	34.21	38.5
OW121763.1	16	33.68	38.5
OW121764.1	17	33.63	39.0
OW121765.1	18	33.35	39.0
OW121766.1	19	32.13	38.5
OW121767.1	20	31.3	39.0
OW121768.1	21	27.97	38.5
OW121769.1	22	27.58	39.5
OW121770.1	23	27.01	39.0
OW121771.1	24	26.15	39.5
OW121772.1	25	25.5	40.0
OW121773.1	26	24.71	39.0
OW121774.1	27	22.38	40.0
OW121775.1	28	20.74	39.5
OW121776.1	29	17.7	40.0
OW121777.1	30	16.23	39.5
OW121778.1	W	7.39	39.5
OW121747.1	Z	47.75	38.5
OW121779.1	MT	0.02	18.5

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), Illumina NovaSeq 6000 (RNA-Seq) and Illumina NovaSeq 6000 (10X) instruments. Hi-C data were also generated from whole organism tissue of ilMilMini3 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 YaHS (Zhou *et al.*, 2023). 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 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 *Miltochrista miniata* assembly (GCA_933228765.1) 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.

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

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
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
YaHS	yahs-1.1.91eebc2	https://github.com/c-zhou/yahs

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: *Miltochrista miniata* (rosy footman). Accession number PRJEB50744; https://identifiers.org/ena.embl/PRJEB50744 (Wellcome Sanger Institute, 2022). The genome sequence is released openly for reuse. The *Miltochrista miniata* 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

Belyakova NA, Lukhtanov VA: **Karyotypic features of some families of noctuoid and bombycoid complexes.** *Entomol Rev.* 1996; **75**: 55–74.
Reference Source

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

Conrad KF, Woiwod IP, Parsons M, *et al.*: **Long-term population trends in widespread British moths.** *J Insect Conserv.* 2004; **8**(2–3): 119–136.
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

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

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

Macgregor CJ, Thomas CD, Roy DB, *et al.*: **Climate-induced phenology shifts linked to range expansions in species with multiple reproductive cycles per year.** *Nat Commun.* 2019; **10**(1): 4455.
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

Pescott OL, Simkin JM, August TA, *et al.*: **Air pollution and its effects on lichens, bryophytes, and lichen-feeding Lepidoptera: review and evidence from biological records.** *Biol J Linn Soc.* 2015; **115**(3): 611–635.
Publisher Full Text

Randle Z, Evans-Hill LJ, Parsons MS, *et al.*: **Atlas of Britain & Ireland's Larger Moths.** Newbury: NatureBureau, 2019.
Reference Source

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**

Scott CH, Zaspel JM, Chialvo P, *et al.*: **A preliminary molecular phylogenetic assessment of the lichen moths (Lepidoptera: Erebidae: Arctiinae: Lithosiini) with comments on palatability and chemical sequestration.** *Syst Entomol.* 2014; **39**(2): 286–303.
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**

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

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 Rosy Footman, *Miltochrista miniata* (Forster, 1771).** European Nucleotide Archive. [dataset], accession number PRJEB50744, 2022.

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**