



DATA NOTE

The genome sequence of the Feathered Gothic, *Tholera decimalis* (Poda, 1761) [version 1; peer review: 2 approved]

Douglas Boyes¹⁺, Peter W.H. Holland²,
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

¹UK Centre for Ecology & Hydrology, Wallingford, England, UK

²University of Oxford, Oxford, England, UK

+ Deceased author

V1 First published: 05 May 2023, 8:200
<https://doi.org/10.12688/wellcomeopenres.19395.1>
Latest published: 05 May 2023, 8:200
<https://doi.org/10.12688/wellcomeopenres.19395.1>

Abstract

We present a genome assembly from an individual female *Tholera decimalis* (the Feathered Gothic; Arthropoda; Insecta; Lepidoptera; Noctuidae). The genome sequence is 1,334.1 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 15.4 kilobases in length. Gene annotation of this assembly on Ensembl identified 12,771 protein coding genes.

Keywords

Tholera decimalis, Feathered Gothic, genome sequence, chromosomal, Lepidoptera



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

Open Peer Review

Approval Status

	1	2
version 1 05 May 2023	 view	 view

1. **Hidemasa Bono** , Hiroshima University, Higashi-Hiroshima, Japan
Hiroshima University, Higashi-Hiroshima, Japan
Research Organization of Information and Systems, Mishima, Japan

2. **Jacqueline Heckenhauer**, Senckenberg Research Institute and Natural History Museum Frankfurt, Frankfurt, Germany

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; **Holland PWH:** 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, <https://doi.org/10.35802/206194>) and the Darwin Tree of Life Discretionary Award (218328, <https://doi.org/10.35802/218328>). *The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.*

Copyright: © 2023 Boyes D *et al.* This is an open access article distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

How to cite this article: Boyes D, Holland PWH, University of Oxford and Wytham Woods Genome Acquisition Lab *et al.* **The genome sequence of the Feathered Gothic, *Tholera decimalis* (Poda, 1761) [version 1; peer review: 2 approved]** Wellcome Open Research 2023, 8:200 <https://doi.org/10.12688/wellcomeopenres.19395.1>

First published: 05 May 2023, 8:200 <https://doi.org/10.12688/wellcomeopenres.19395.1>

Species taxonomy

Eukaryota; Metazoa; Ecdysozoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Lepidoptera; Glossata; Ditrysia; Noctuoidea; Noctuidae; Hadeninae; *Tholera*; *Tholera decimalis* (Poda, 1761) (NCBI:txid988041).

Background

The use of a network of standardised light-traps operated every night since 1970, coupled with millions of individual records collated by amateur and professional recorders, has permitted analysis of long-term trends in moth abundance. These studies indicate that over 60% of larger moth species have decreased in abundance in Britain in the past half century (Conrad *et al.*, 2006; Fox *et al.*, 2019; Randle *et al.*, 2019). The picture for geographical distribution within Britain is more complex, with roughly equal numbers of species expanding their range or facing range contraction (Randle *et al.*, 2019). The Feathered Gothic *Tholera decimalis* is a moth that has had mixed fortunes in this period and may provide a useful study case. The quantitative data indicate that *T. decimalis* has suffered a large decline in abundance since 1970, while its geographic range showed a large contraction followed by a more recent expansion from 2000 to 2016 (Randle *et al.*, 2019).

Currently, *T. decimalis* is found widely across England and Wales, particularly in southern counties, but it is now absent from most of Scotland and very rare in Ireland and Northern Ireland (National Biodiversity Data Centre, 2023; Randle *et al.*, 2019; Thompson & Nelson, 2003). The species has a patchy distribution across Eurasia with records concentrated in the Netherlands, Austria, Switzerland and Scandinavia; there are sporadic records further east through Russia to Mongolia (GBIF Secretariat, 2023). The moth is associated with rough grassland, downland and open woodland, with adults laying eggs during August and September. The eggs overwinter before the hatched larvae feed on grasses between March and July (Stokoe, 1948). The adult moth has bold white markings along the wing veins, streaked over ornately patterned rich brown forewings. Males have very pronounced bipectinate (feathered) antennae; this feature, not seen in females, is likely an adaptation to accommodate increased numbers of olfactory receptors for pheromone detection. Further research is needed into the chemical biology of this species, particularly as electroantennogram recording showed no response to 30 putative pheromone components that elicit responses in related species (Renou *et al.*, 1991).

A genome sequence for *T. decimalis* will enable studies into the biochemical basis of olfactory reception and molecular adaptations for grass feeding, and may facilitate future research into the biological factors affecting responses to environmental and land use changes.

Genome sequence report

The genome was sequenced from one female *Tholera decimalis* (Figure 1) collected from Wytham Woods, Oxfordshire, UK (latitude 51.77, longitude -1.34). A total of 45-fold coverage in Pacific Biosciences single-molecule HiFi long reads and 29-fold coverage in 10X Genomics read clouds were generated. Primary assembly contigs were scaffolded with chromosome conformation Hi-C data. Manual assembly curation corrected



Figure 1. Photograph of the *Tholera decimalis* (ilThoDeci1) specimen used for genome sequencing.

58 missing or mis-joins and removed 22 haplotypic duplications, reducing the assembly length by 1.26% and the scaffold number by 25%.

The final assembly has a total length of 1,334.1 Mb in 84 sequence scaffolds with a scaffold N50 of 44.3 Mb (Table 1). Most (99.75%) 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 2–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 64.6 with *k*-mer completeness of 100%, and the assembly has a BUSCO v5.3.2 completeness of 99.0% (single = 97.9%, duplicated = 1.1%), 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/988041>.

Genome annotation report

The *T. decimalis* assembly GCA_943138885.1 genome was annotated using the Ensembl rapid annotation pipeline (Table 1; https://rapid.ensembl.org/Tholera_decimalis_GCA_943138885.1/Info/Index). The resulting annotation includes 22,825 transcribed mRNAs from 12,771 protein-coding and 1,968 non-coding genes.

Methods

Sample acquisition and nucleic acid extraction

A male *Tholera decimalis* (ilThoDeci1) was collected from Wytham Woods, Oxfordshire (biological vice-county: Berkshire), UK (latitude 51.77, longitude -1.34) on 8 September 2020. The specimen was taken from woodland habitat by Douglas Boyes (University of Oxford) using a light trap. The

Table 1. Genome data for *Tholera decimalis*, ilThoDeci1.2.

Project accession data		
Assembly identifier	ilThoDeci1.2	
Species	<i>Tholera decimalis</i>	
Specimen	ilThoDeci1	
NCBI taxonomy ID	988041	
BioProject	PRJEB52581	
BioSample ID	SAMEA8603189	
Isolate information	ilThoDeci1, female: head and thorax (genome sequencing and Hi-C scaffolding); abdomen (RNA sequencing)	
Assembly metrics*		Benchmark
Consensus quality (QV)	64.6	≥ 50
k-mer completeness	100%	≥ 95%
BUSCO**	C:99.0%[S:97.9%,D:1.1%], F:0.3%,M:0.7%,n:5,286	C ≥ 95%
Percentage of assembly mapped to chromosomes	99.75%	≥ 95%
Sex chromosomes	Z chromosome	<i>localised homologous pairs</i>
Organelles	Mitochondrial genome assembled	<i>complete single alleles</i>
Raw data accessions		
PacificBiosciences SEQUEL II	ERR9709328, ERR9709329, ERR9709330, ERR9709331, ERR9709332	
10X Genomics Illumina	ERR9682493–ERR9682496	
Hi-C Illumina	ERR9682498	
PolyA RNA-Seq Illumina	ERR9682497	
Genome assembly		
Assembly accession	GCA_943138885.2	
Accession of alternate haplotype	GCA_943138005.2	
Span (Mb)	1,334.1	
Number of contigs	311	
Contig N50 length (Mb)	10.2	
Number of scaffolds	84	
Scaffold N50 length (Mb)	44.3	
Longest scaffold (Mb)	55.9	
Genome annotation		
Number of protein-coding genes	12,771	
Number of non-coding genes	1,968	
Number of gene transcripts	22,825	

* 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/ilThoDeci1.2/dataset/CALPBR02/busco>.

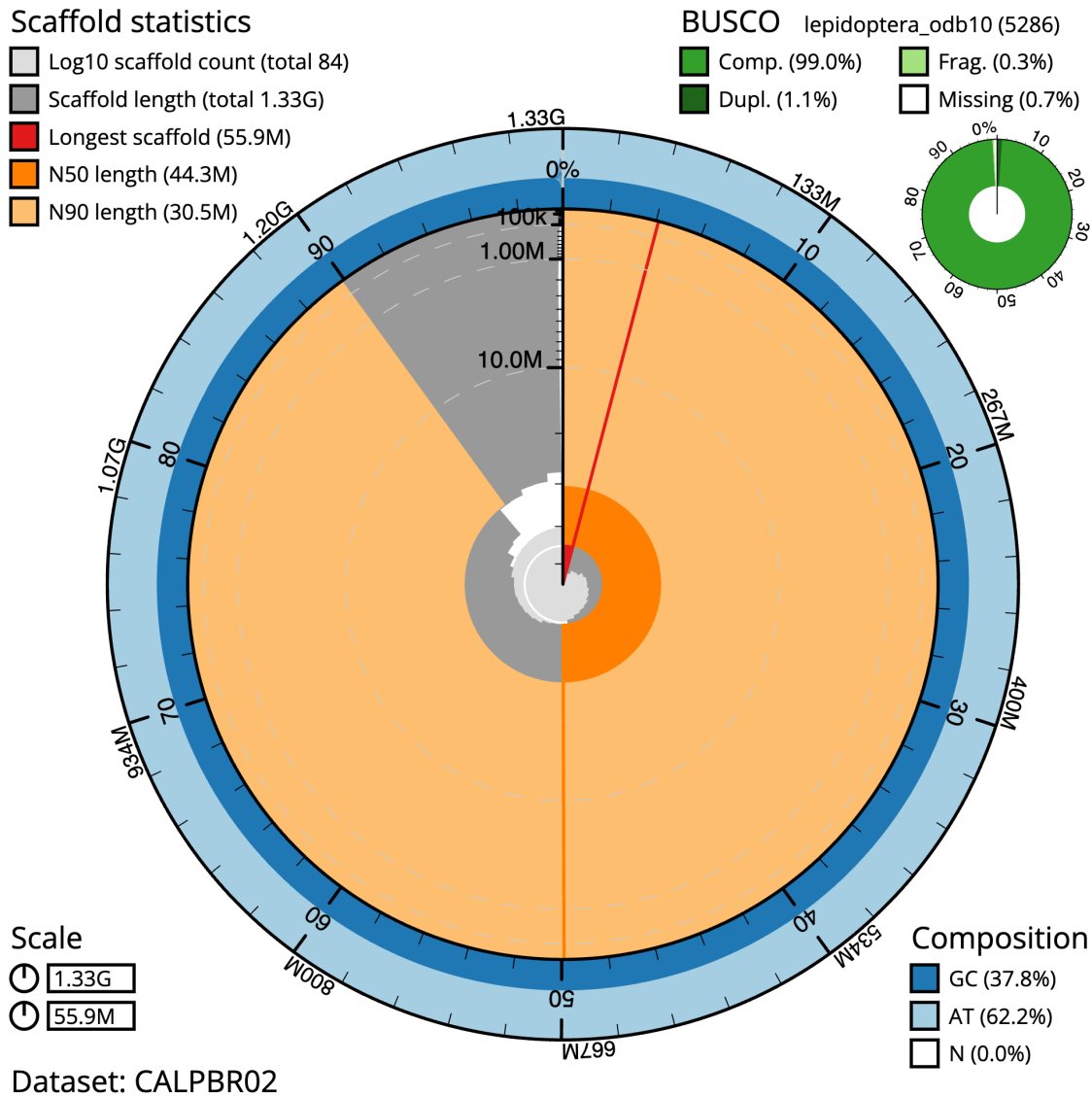


Figure 2. Genome assembly of *Tholera decimilis*, iIThoDeci1.2: metrics. The BlobToolKit Snailplot shows N50 metrics and BUSCO gene completeness. The main plot is divided into 1,000 size-ordered bins around the circumference with each bin representing 0.1% of the 1,334,081,899 bp assembly. The distribution of scaffold lengths is shown in dark grey with the plot radius scaled to the longest scaffold present in the assembly (55,901,591 bp, shown in red). Orange and pale-orange arcs show the N50 and N90 scaffold lengths (44,326,127 and 30,461,373 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/iIThoDeci1.2/dataset/CALPBR02/snail>.

specimen was identified by the collector and snap-frozen on dry ice.

DNA was extracted at the Tree of Life laboratory, Wellcome Sanger Institute (WSI). The iIThoDeci1 sample was weighed and dissected on dry ice with tissue set aside for Hi-C sequencing. Head and thorax tissue was cryogenically disrupted to a fine powder using a Covaris cryoPREP Automated Dry Pulveriser, receiving multiple impacts. 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

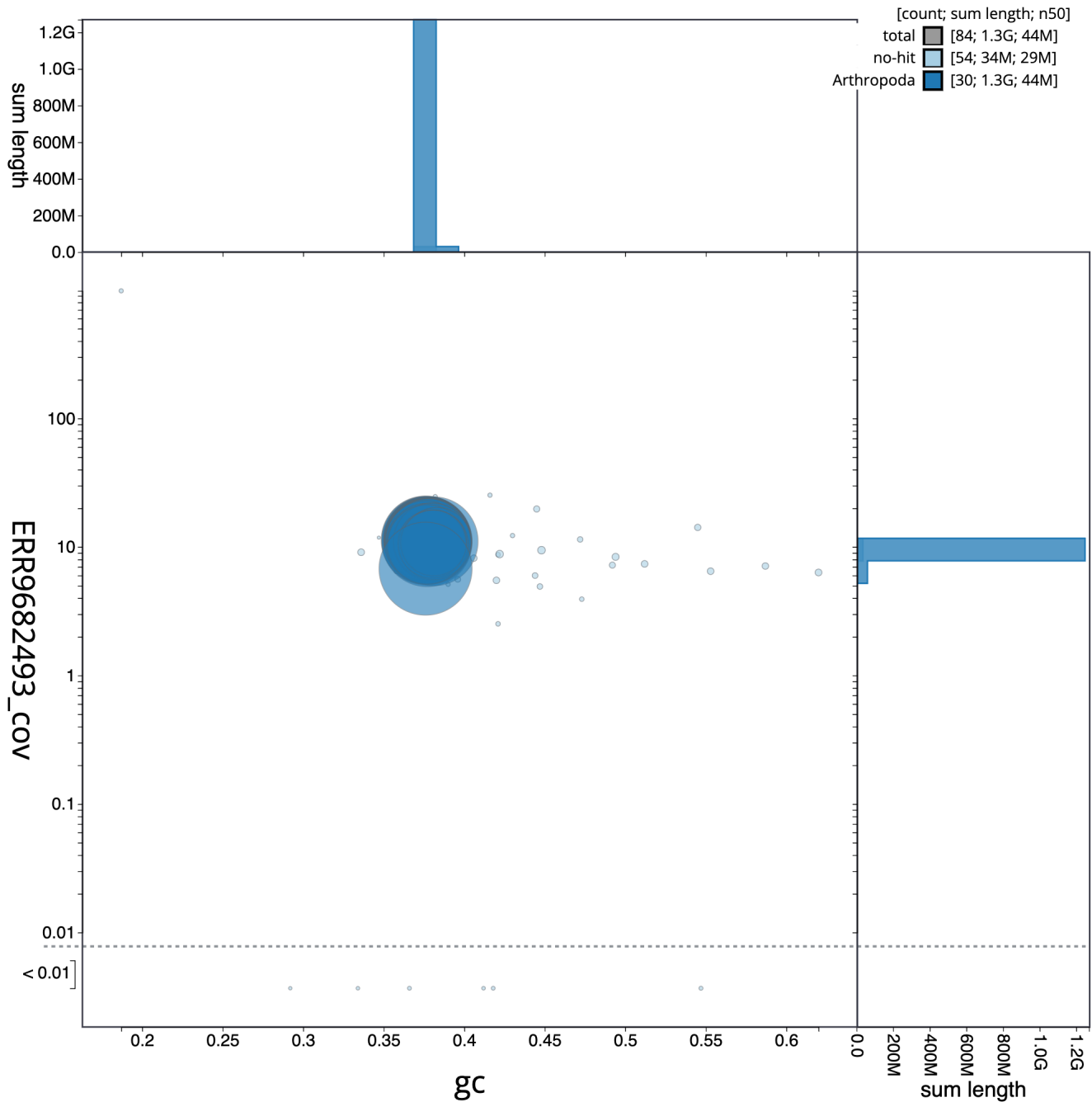


Figure 3. Genome assembly of *Tholera decimalis*, ilThoDeci1.2: GC coverage. BlobToolKit GC-coverage plot. Scaffolds are coloured by phylum. Circles are sized in proportion to scaffold length. Histograms show the distribution of scaffold length sum along each axis. An interactive version of this figure is available at <https://blobtoolkit.genomehubs.org/view/ilThoDeci1.2/dataset/CALPBR02/blob>.

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 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 ilThoDeci1 in the Tree of Life Laboratory at the WSI using TRIzol, according to the manufacturer’s instructions. RNA was then eluted in 50 μ l RNase-free water and its concentration assessed using a Nanodrop spectrophotometer and Qubit Fluorometer using the Qubit RNA Broad-Range (BR) Assay kit. Analysis of

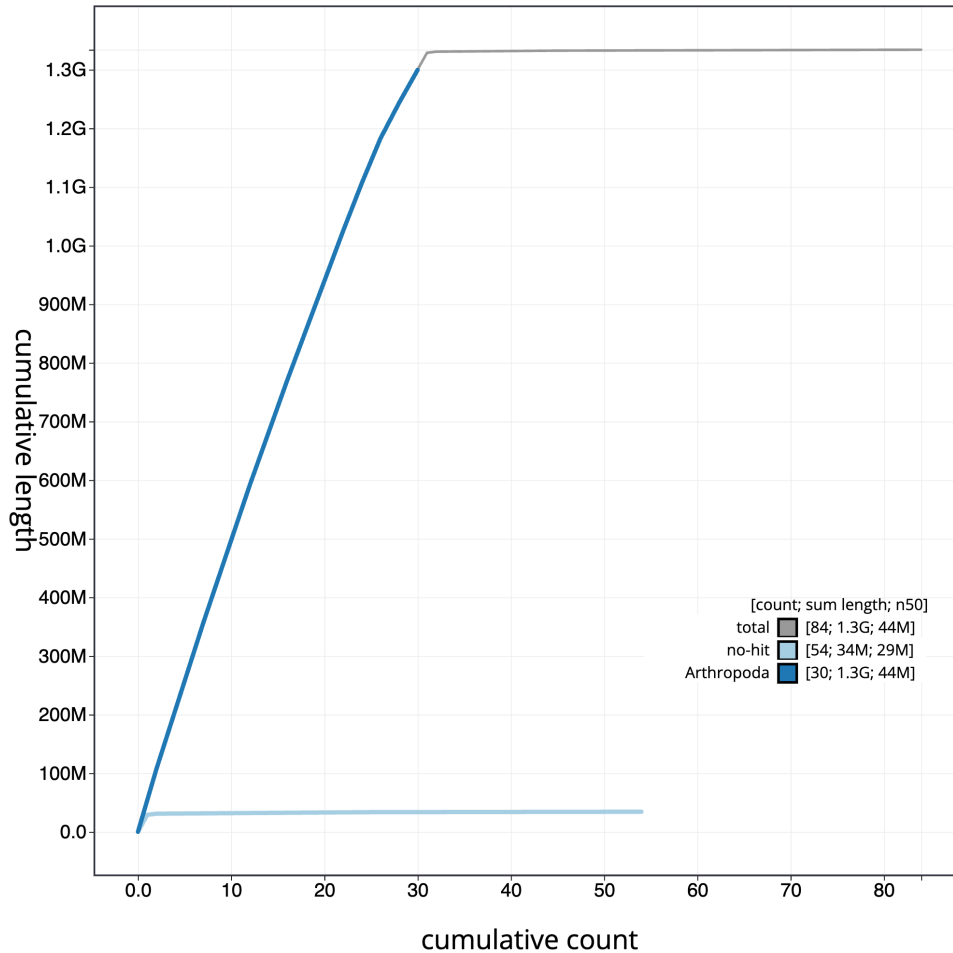


Figure 4. Genome assembly of *Tholera decimalis*, ilThoDeci1.2: cumulative sequence. BlobToolKit cumulative sequence plot. The grey line shows cumulative length for all scaffolds. Coloured lines show cumulative lengths of scaffolds assigned to each phylum using the buscogenes taxrule. An interactive version of this figure is available at <https://blobtoolkit.genomehubs.org/view/ilThoDeci1.2/dataset/CALPBR02/cumulative>.

the integrity of the RNA was done using 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. DNA sequencing was performed by the Scientific Operations core at the WSI on Pacific Biosciences SEQUEL II (HiFi), Illumina HiSeq 4000 (RNA-Seq) and Illumina NovaSeq 6000 (10X) instruments. Hi-C data were also generated from head tissue of ilThoDeci1 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 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.*, 2022), 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. To evaluate the assembly, MerquryFK was used to estimate consensus quality (QV) scores and *k*-mer completeness (Rhie *et al.*, 2020). 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 software tool versions and sources.

Genome annotation

The Ensembl gene annotation system (Aken *et al.*, 2016) was used to generate annotation for the *Tholera decimalis* assembly

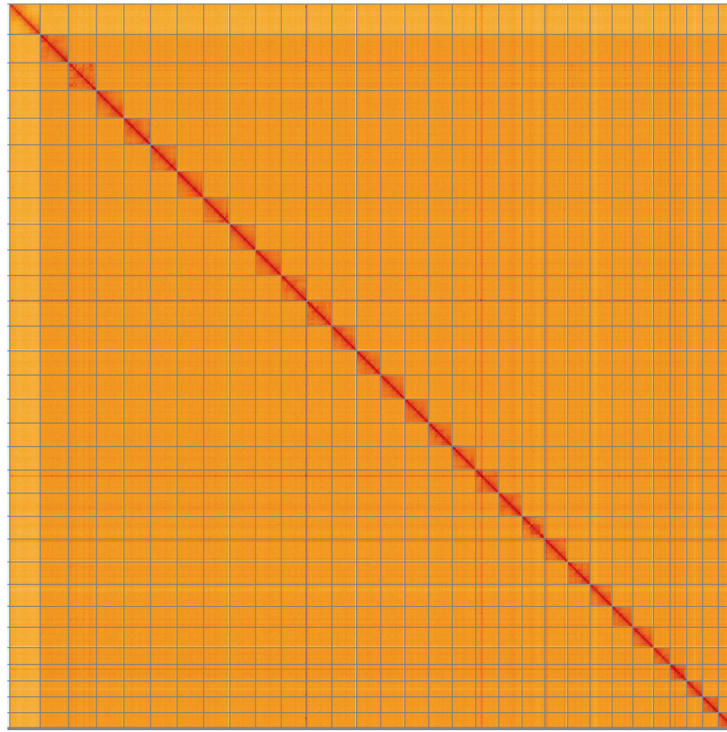


Figure 5. Genome assembly of *Tholera decimalis*, ilThoDeci1.2: Hi-C contact map. Hi-C contact map of the ilThoDeci1.2 assembly, visualised using HiGlass. Chromosomes are shown in order of size from left to right and top to bottom. An interactive version of this figure may be viewed at https://genome-note-higlass.tol.sanger.ac.uk/l/?d=IsQwpVaRQ36c521_VU0W0w.

Table 2. Chromosomal pseudomolecules in the genome assembly of *Tholera decimalis*, ilThoDeci1.

INSDC accession	Chromosome	Size (Mb)	GC%
OW964549.2	1	52.14	37.7
OW964550.2	2	50.54	37.6
OW964551.2	3	50.86	38.1
OW964552.2	4	49.27	37.6
OW964553.2	5	48.6	37.7
OW964554.2	6	48.73	37.7
OW964555.2	7	47.62	37.8
OW964556.2	8	48.04	37.7
OW964557.2	9	47.04	37.7
OW964558.2	10	46.7	37.8
OW964559.2	11	46.23	37.7
OW964560.2	12	44.29	37.7
OW964561.2	13	45.48	37.8
OW964562.2	14	44.01	37.7
OW964563.2	15	44.33	37.8
OW964564.2	16	43.19	37.7

INSDC accession	Chromosome	Size (Mb)	GC%
OW964565.2	17	42.57	37.8
OW964566.2	18	42.87	37.8
OW964567.2	19	41.98	37.7
OW964568.2	20	42.31	37.8
OW964569.2	21	41.18	37.6
OW964570.2	22	38.34	38
OW964571.2	23	42.05	37.7
OW964572.2	24	37.87	38
OW964573.2	25	40.27	37.7
OW964574.2	26	30.36	38.3
OW964575.2	27	30.46	38.1
OW964576.2	28	29.18	38
OW964577.2	29	29.08	37.8
OW964578.2	30	27.34	38.1
OW964548.2	Z	55.9	37.6
OW964579.2	MT	0.02	18.8
-	unplaced	5.23	41.3

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	0.16.1-r375	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
YaHS	yahs-1.1.91eebc2	https://github.com/c-zhou/yahs

(GCA_943138885.1). 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](#)).

Ethics and compliance issues

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](#). 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. All efforts are undertaken to minimise the suffering of animals used for sequencing. 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: *Tholera decimalis* (feathered gothic). Accession number [PRJEB52581](#); <https://identifiers.org/ena.embl/PRJEB52581>. (Wellcome Sanger Institute, 2022)

The genome sequence is released openly for reuse. The *Tholera decimalis* 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

- 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](#)
- Conrad KF, Warren MS, Fox R, *et al.*: **Rapid declines of common, widespread British moths provide evidence of an insect biodiversity crisis.** *Biological Conservation*. 2006; **132**(3): 279–291.
[Publisher Full Text](#)
- Fox R, Parsons MS, Harrower CA: **A review of the status of the macro-moths of Great Britain.** Dorset, UK: Butterfly Conservation, 2019.
[Reference Source](#)
- Garrison E, Marth G: **Haplotype-based variant detection from short-read sequencing.** 2012.
[Publisher Full Text](#)
- GBIF Secretariat: **Tholera decimialis Poda, 1761.** *GBIF Backbone Taxonomy*. 2023; (Accessed: 3 April 2023).
[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](#)
- Howe K, Chow W, Collins J, *et al.*: **Significantly improving the quality of genome assemblies through curation.** *GigaScience*. Oxford University Press, 2021; **10**(1): g1aa153.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Kerpedjiev P, Abdennur N, Lekschas F, *et al.*: **HiGlass: Web-based visual exploration and analysis of genome interaction maps.** *Genome Biol*. 2018; **19**(1): 125.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Manni M, Berkeley MR, Seppely 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](#)
- National Biodiversity Data Centre: **Feathered Gothic (Tholera decimialis).** National Biodiversity Data Centre, Ireland. 2023; (Accessed: 5 January 2023).
[Reference Source](#)
- 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](#)
- Renou M, Lucas P, Dore JC, *et al.*: **A comparative study of sex pheromone reception in the Hadeninae (Lepidoptera: Noctuidae).** *Physiol Entomol*. 1991; **16**(1): 87–97.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Rhie A, Walenz BP, Koren S, *et al.*: **Mercury: Reference-free quality, completeness, and phasing assessment for genome assemblies.** *Genome Biol*. 2020; **21**(1): 245.
[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](#)
- 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](#)
- Stokoe WJ: **The Caterpillars of British Moths. Series 1.** London and New York: Warne and Co, 1948.
[Reference Source](#)
- Thompson RS, Nelson B: **Feathered Gothic.** In: *The Butterflies and Moths of Northern Ireland*. Ulster: Ulster Museum, 2003; (Accessed: 5 January 2023).
[Reference Source](#)
- Uliano-Silva M, Ferreira JGRN, Krashennikova K, *et al.*: **MitoHiFi: a python pipeline for mitochondrial genome assembly from PacBio High Fidelity reads.** *bioRxiv*. [Preprint]. 2022.
[Publisher 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](#)
- Wellcome Sanger Institute: **The genome sequence of the Feathered Gothic, Tholera decimialis (Poda, 1761).** European Nucleotide Archive. [dataset], accession number PRJEB52581, 2022.
- Zhou C, McCarthy SA, Durbin R: **YaHS: yet another Hi-C scaffolding tool.** *Bioinformatics*. Edited by C. Alkan, 2023; **39**(1): btac808.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)

Open Peer Review

Current Peer Review Status:  

Version 1

Reviewer Report 04 September 2023

<https://doi.org/10.21956/wellcomeopenres.21483.r66112>

© 2023 Heckenhauer J. This is an open access peer review report distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.



Jacqueline Heckenhauer

Senckenberg Research Institute and Natural History Museum Frankfurt, Frankfurt, Germany

In their data note, the authors present the genome sequence of the primary assembly and contigs of the alternate haplotype of *Tholera decimalis* using a *de novo* assembly method following the approach used by the Darwin Tree of Life Project. The presented genome is of high quality and released openly for reuse. Therefore this high-quality reference genome is very beneficial to the field.

However, I suggest that the authors include a short summary of repetitive DNA components of the assembly and submit the library of repetitive element (RE) sequences to a public repository of repetitive elements (e.g., Dfam). This way the community would even more benefit from this data. The diversity of available insect genomes has rapidly expanded, but the rate of community contributions to RE databases which are important for RE annotation has not kept pace, preventing high-resolution study of REs in many groups.

Was the genome size of this specimen / species estimated by flow cytometry or similar?

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: evolution, biodiversity genomics

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

Reviewer Report 23 May 2023

<https://doi.org/10.21956/wellcomeopenres.21483.r57063>

© 2023 Bono H. This is an open access peer review report distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.



Hidemasa Bono 

¹ Genome Editing Innovation Center, Hiroshima University, Higashi-Hiroshima, Hiroshima, Japan

² Graduate School of Integrated Sciences for Life, Hiroshima University, Higashi-Hiroshima, Hiroshima, Japan

³ Database Center for Life Science, Research Organization of Information and Systems, Mishima, Shizuoka, Japan

Authors present the genome sequencing of the feathered gothic *Theolera decimalis*. The reads used in this study were sufficiently deep (45-fold coverage in PacBio HiFi) with Hi-C reads. The assembly constructed was very good judged from BUSCO v5.3.2 completeness score (99.0%). The computational programs used in that process were clearly summarized in Table 3 with precise version information. Reviewer has no more requests on the manuscript as a genome sequencing report.

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 informatics for non-model organisms

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.