




DATA NOTE

The genome sequence of the Hebrew Character, *Orthosia gothica* (Linnaeus, 1758) [version 1; peer review: 2 approved]

Douglas Boyes¹⁺, Peter W.H. Holland ²,
University of Oxford and Wytham Woods Genome Acquisition Lab,
Darwin Tree of Life Barcoding collective,
Wellcome Sanger Institute Tree of Life Management, Samples and Laboratory
team,
Wellcome Sanger Institute Scientific Operations: Sequencing Operations,
Wellcome Sanger Institute Tree of Life Core Informatics team,
Tree of Life Core Informatics collective, Darwin Tree of Life Consortium

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

²University of Oxford, Oxford, England, UK

+ Deceased author

V1 First published: 19 Feb 2024, 9:90
<https://doi.org/10.12688/wellcomeopenres.20904.1>
Latest published: 19 Feb 2024, 9:90
<https://doi.org/10.12688/wellcomeopenres.20904.1>

Abstract

We present a genome assembly from an individual female *Orthosia gothica* (the Hebrew character; Arthropoda; Insecta; Lepidoptera; Noctuidae). The genome sequence is 1,065.1 megabases in span. Most of the assembly is scaffolded into 37 chromosomal pseudomolecules, including the Z sex chromosome. The mitochondrial genome has also been assembled and is 15.38 kilobases in length. Gene annotation of this assembly on Ensembl identified 13,691 protein coding genes.

Keywords



Orthosia gothica, Hebrew character, genome sequence, chromosomal, Lepidoptera



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

Open Peer Review

Approval Status  

	1	2
version 1 19 Feb 2024	 view	 view

1. **Ricardo Rodriguez-de-la-vega** ,
Université Paris-Saclay, Paris, France
2. **Alessandro Grapputo** , Università degli
Studi di Padova, Padua, Italy

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: © 2024 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 Hebrew Character, *Orthosia gothica* (Linnaeus, 1758) [version 1; peer review: 2 approved]** Wellcome Open Research 2024, 9:90 <https://doi.org/10.12688/wellcomeopenres.20904.1>

First published: 19 Feb 2024, 9:90 <https://doi.org/10.12688/wellcomeopenres.20904.1>

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; Obtectomera; Noctuoidea; Noctuidae; Hadeninae; *Orthosia*; *Orthosia gothica* (Linnaeus, 1758) (NCBI:txid43337).

Background

The Hebrew Character moth *Orthosia gothica* is a spring-flying moth found across Eurasia from Portugal and Ireland to Japan, and from the north of Scandinavia to the south of Italy (GBIF Secretariat, 2023). The moth gets its common name from a distinctive black or brown mark on each forewing in the shape of the letter ן (nun) in the Hebrew alphabet. The species name *gothica* derives from the same marking, likening the shape to a Gothic arch. In Britain and Ireland the flight period spans from March to May, with the species recorded from all regions and almost all habitats, sometimes abundantly (MothsIreland, 2023; NBN Atlas Partnership, 2023). The adults are attracted to light and can also be found feeding on nectar at swallow catkins (South, 1961).

The larvae of *O. gothica* are polyphagous, feeding on leaves of herbaceous plants and deciduous trees, and the species has been recorded as a minor pest in orchards and forestry (Bues *et al.*, 1994). Steps taken toward potential pest control include a synthetic male-attractant chemical blend (Tøth *et al.*, 1992) and bisexual lures based on floral compounds or chemicals from rotting fruit (Szanyi *et al.*, 2022). The species has been used as a model system to study biochemical responses of lepidopteran larvae to insecticides (Egaas, 2000; Egaas *et al.*, 1992).

A genome sequence of the Hebrew Character *Orthosia gothica* was determined as part of the Darwin Tree of Life project. The genome sequence will facilitate research into adaptations to polyphagy and the biochemical basis of insecticide response, and will contribute to the growing set of resources for studying molecular evolution in the Lepidoptera.

Genome sequence report

The genome was sequenced from one female *Orthosia gothica* (Figure 1) collected from Wytham Woods, Oxfordshire, UK (51.77, -1.34). A total of 21-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 172 missing joins or mis-joins and removed 85 haplotypic duplications, reducing the assembly length by 4.93% and the scaffold number by 20.14%.

The final assembly has a total length of 1,065.1 Mb in 337 sequence scaffolds with a scaffold N50 of 32.4 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



Figure 1. Photograph of the *Orthosia gothica* (ilOrtGoth1) specimen used for genome sequencing.

assembly plot in Figure 4 shows curves for subsets of scaffolds assigned to different phyla. Most (97.61%) of the assembly sequence was assigned to 37 chromosomal-level scaffolds, representing 36 autosomes and the Z sex chromosome. The Z chromosome was identified based on coverage; a W chromosome was not identified. 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 62.8 with *k*-mer completeness of 100.0%, and the assembly has a BUSCO v5.3.2 completeness of 98.9% (single = 97.1%, duplicated = 1.8%), 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/43337>.

Genome annotation report

The *Orthosia gothica* genome assembly (GCA_949775005.1) was annotated using the Ensembl rapid annotation pipeline (Table 1; https://rapid.ensembl.org/Orthosia_gothica_GCA_949775005.1/Info/Index). The resulting annotation includes 24,253 transcribed mRNAs from 13,691 protein-coding and 2,098 non-coding genes.

Methods

Sample acquisition and nucleic acid extraction

A female *Orthosia gothica* (specimen ID Ox001095, ToLID ilOrtGoth1) was collected from Wytham Woods, Oxfordshire (biological vice-county Berkshire), UK (latitude 51.77, longitude -1.34) on 2021-03-31 using a light trap. The specimen was collected and identified by Douglas Boyes (University of Oxford) and preserved on dry ice.

Table 1. Genome data for *Orthosia gothica*, ilOrtGoth1.1.

Project accession data		
Assembly identifier	ilOrtGoth1.1	
Species	<i>Orthosia gothica</i>	
Specimen	ilOrtGoth1	
NCBI taxonomy ID	43337	
BioProject	PRJEB54944	
BioSample ID	SAMEA10107018	
Isolate information	ilOrtGoth1, female: thorax (DNA sequencing), head (Hi-C sequencing), abdomen (RNA sequencing)	
Assembly metrics*		Benchmark
Consensus quality (QV)	62.8	≥ 50
<i>k</i> -mer completeness	100.0%	≥ 95%
BUSCO**	C:98.9%[S:97.1%,D:1.8%], F:0.2%,M:1.0%,n:5,286	C ≥ 95%
Percentage of assembly mapped to chromosomes	97.61%	≥ 95%
Sex chromosomes	Z	<i>localised homologous pairs</i>
Organelles	Mitochondrial genome: 15.38 kb	<i>complete single alleles</i>
Raw data accessions		
PacificBiosciences SEQUEL II	ERR10008899	
Hi-C Illumina	ERR9988142	
PolyA RNA-Seq Illumina	ERR10890698	
Genome assembly		
Assembly accession	GCA_949775005.1	
<i>Accession of alternate haplotype</i>	GCA_949774995.1	
Span (Mb)	1,065.1	
Number of contigs	1,142	
Contig N50 length (Mb)	2.3	
Number of scaffolds	337	
Scaffold N50 length (Mb)	32.4	
Longest scaffold (Mb)	64.64	
Genome annotation		
Number of protein-coding genes	13,691	
Number of non-coding genes	2,098	
Number of gene transcripts	24,253	

* 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/ilOrtGoth1_1/dataset/ilOrtGoth1_1/busco.

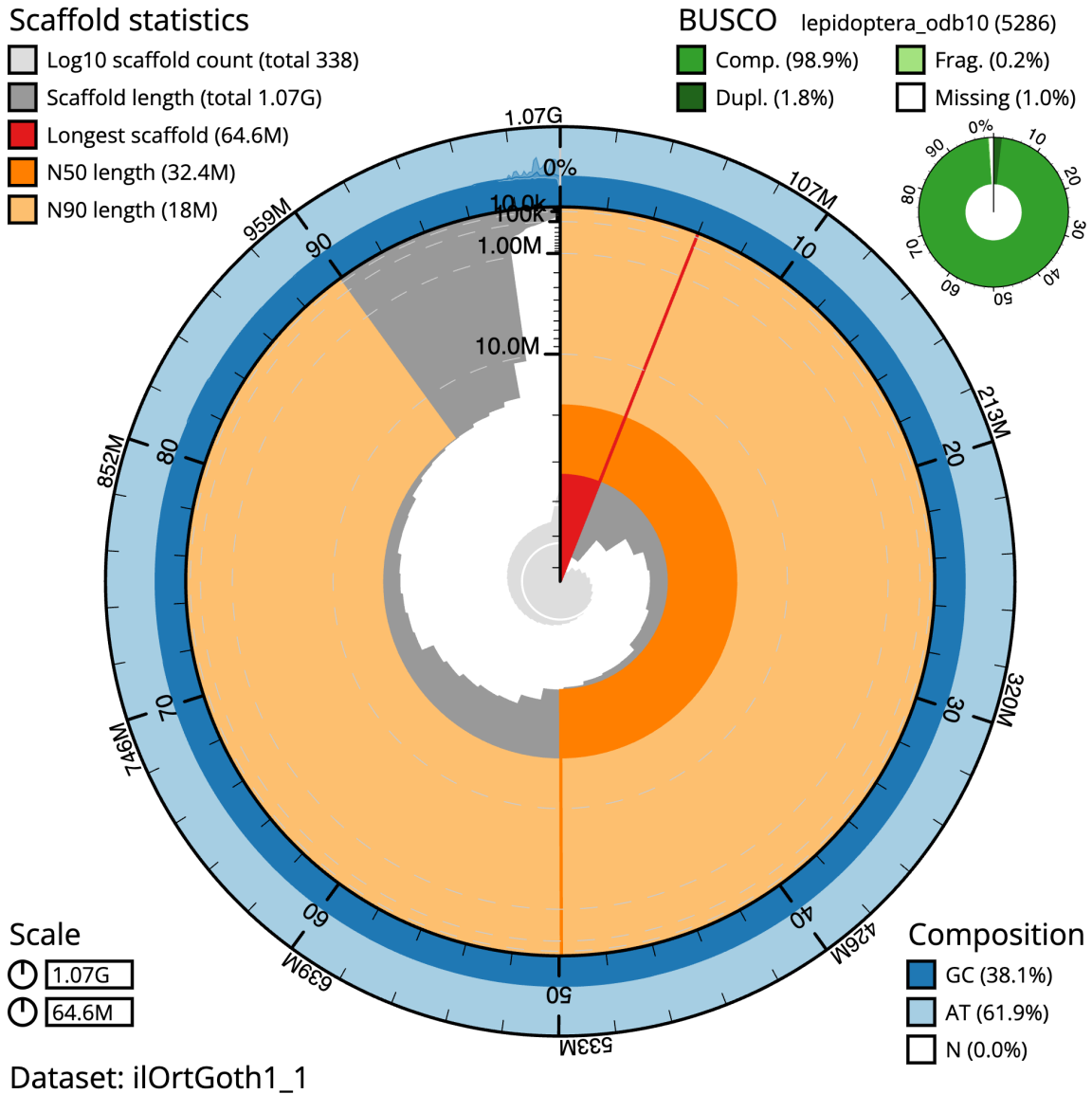


Figure 2. Genome assembly of *Orthosia gothica*, ilOrtGoth1.1: metrics. The BlobToolKit Snailplot shows N50 metrics and BUSCO gene completeness. The main plot is divided into 1,000 size-ordered bins around the circumference with each bin representing 0.1% of the 1,065,129,911 bp assembly. The distribution of scaffold lengths is shown in dark grey with the plot radius scaled to the longest scaffold present in the assembly (64,640,039 bp, shown in red). Orange and pale-orange arcs show the N50 and N90 scaffold lengths (32,428,790 and 17,961,581 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/ilOrtGoth1_1/dataset/ilOrtGoth1_1/snail.

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 ilOrtGoth1 sample was weighed and dissected on dry ice (Jay *et al.*, 2023). For sample homogenisation, thorax tissue

was cryogenically disrupted using the Covaris cryoPREP® Automated Dry Pulverizer (Narváez-Gómez *et al.*, 2023).

HMW DNA was extracted in the WSI Scientific Operations core using the Automated MagAttract v2 protocol (Oatley *et al.*, 2023). The DNA was sheared into an average fragment

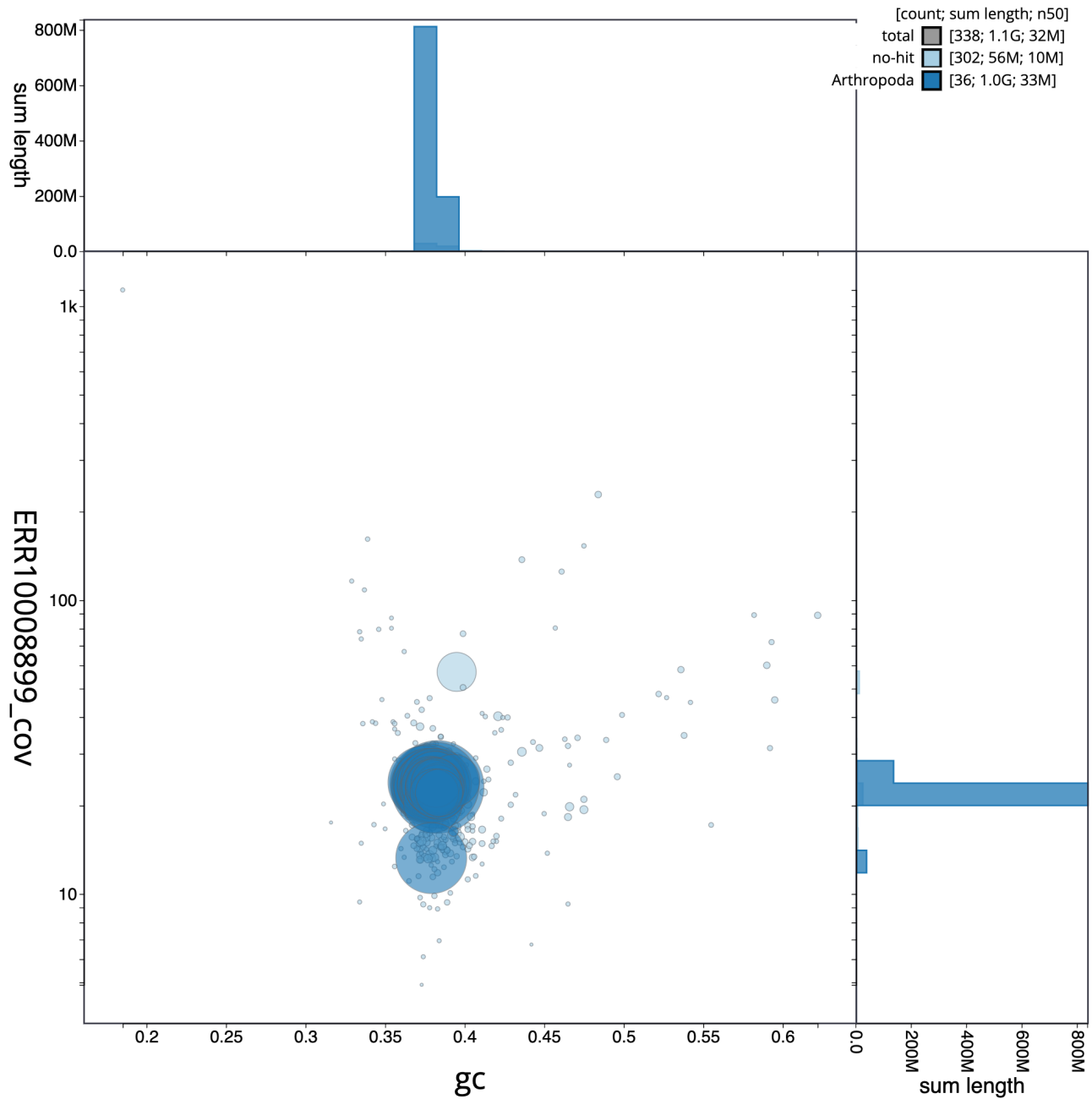


Figure 3. Genome assembly of *Orthosia gothica*, iOrtGoth1.1: BlobToolKit GC-coverage plot. Scaffolds are coloured by phylum. Circles are sized in proportion to scaffold length. Histograms show the distribution of scaffold length sum along each axis. An interactive version of this figure is available at https://blobtoolkit.genomehubs.org/view/iOrtGoth1_1/dataset/iOrtGoth1_1/blob.

size of 12–20 kb in a Megaruptor 3 system with speed setting 31 (Bates *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 abdomen tissue of iOrtGoth1 in the Tree of Life Laboratory at the WSI using the RNA Extraction: Automated MagMax™ *mir*-Vana protocol (do Amaral *et al.*, 2023). The RNA concentration was assessed using a Nanodrop spectrophotometer and a Qubit Fluorometer using the

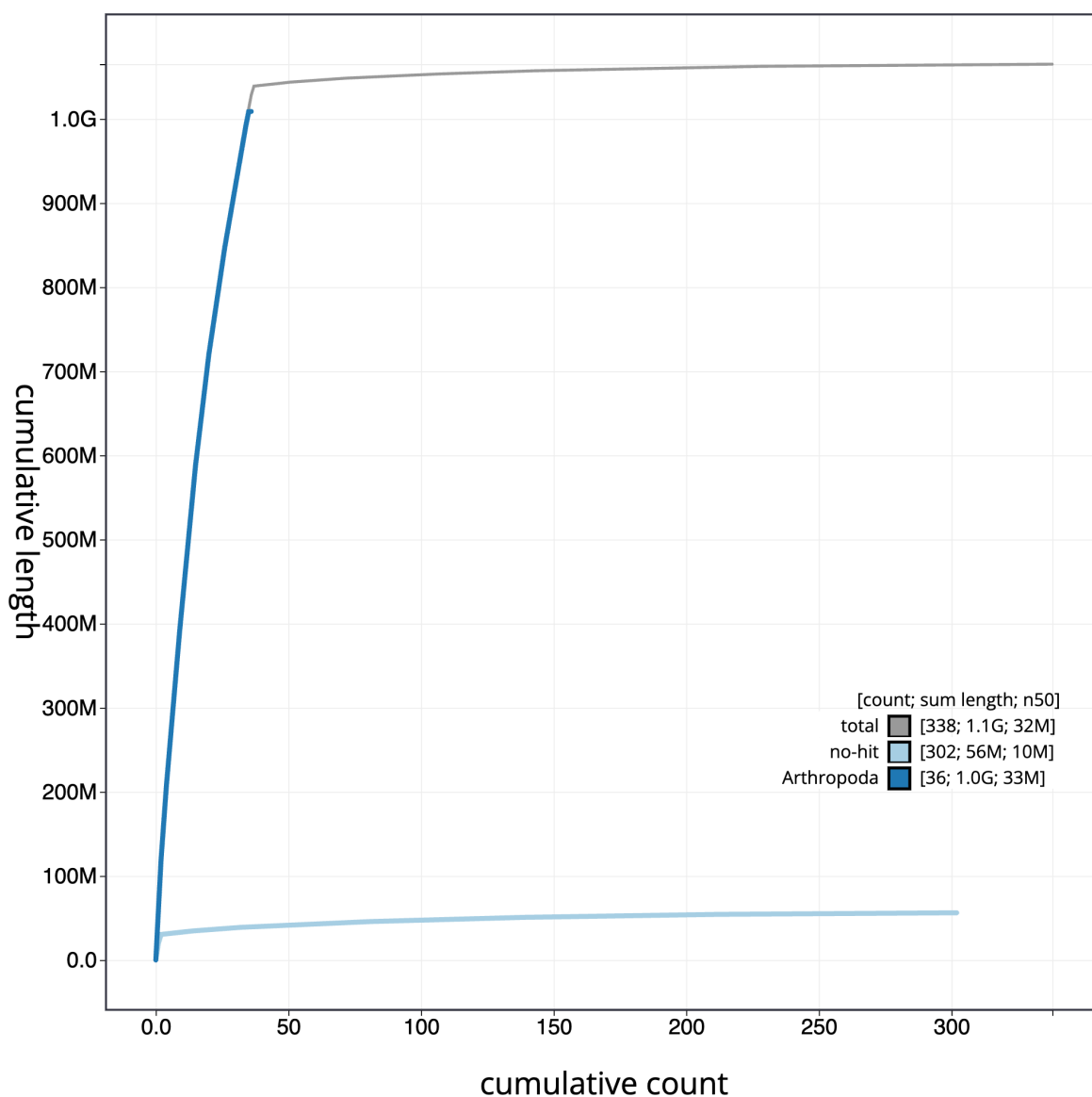


Figure 4. Genome assembly of *Orthosia gothica*, iOrtGoth1.1: BlobToolKit cumulative sequence plot. The grey line shows cumulative length for all scaffolds. Coloured lines show cumulative lengths of scaffolds assigned to each phylum using the buscogenes taxrule. An interactive version of this figure is available at https://blobtoolkit.genomehubs.org/view/iOrtGoth1_1/dataset/iOrtGoth1_1/cumulative.

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.*, 2023).

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 head tissue of iOrtGoth1 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

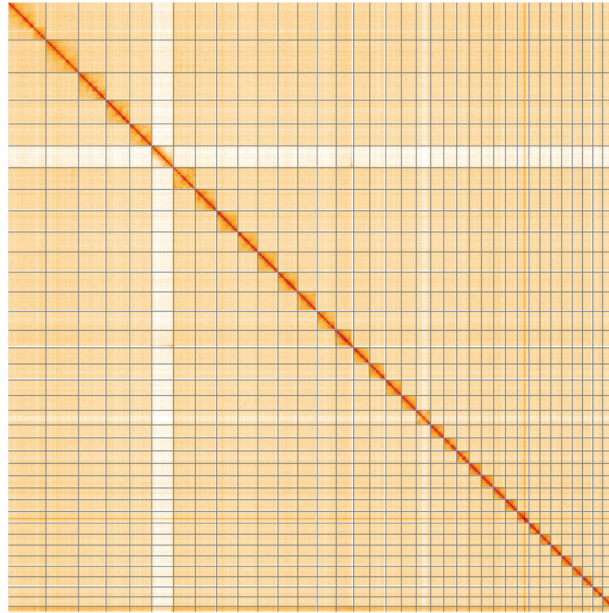


Figure 5. Genome assembly of *Orthosia gothica*, iOrtGoth1.1: Hi-C contact map of the iOrtGoth1.1 assembly, visualised using HiGlass. Chromosomes are shown in order of size from left to right and top to bottom. An interactive version of this figure may be viewed at <https://genome-note-higlass.tol.sanger.ac.uk/l/?d=YA-Wg3v2T3qyxzkeuf2aZA>.

Table 2. Chromosomal pseudomolecules in the genome assembly of *Orthosia gothica*, iOrtGoth1.

INSDC accession	Chromosome	Length (Mb)	GC%
OX459034.1	1	64.64	38.5
OX459035.1	2	55.54	38.0
OX459036.1	3	46.83	38.0
OX459037.1	4	40.27	37.5
OX459038.1	5	37.61	38.0
OX459040.1	6	37.07	38.0
OX459041.1	7	36.2	37.5
OX459042.1	8	36.02	37.5
OX459043.1	9	34.26	37.5
OX459044.1	10	34.19	38.0
OX459045.1	11	33.82	38.0
OX459046.1	12	32.99	38.0
OX459047.1	13	32.43	38.0
OX459048.1	14	29.68	38.0
OX459049.1	15	27.52	38.0
OX459050.1	16	27.19	38.0
OX459051.1	17	26.44	38.0
OX459052.1	18	25.41	38.0

INSDC accession	Chromosome	Length (Mb)	GC%
OX459053.1	19	24.73	38.0
OX459054.1	20	22.14	38.0
OX459055.1	21	21.87	38.5
OX459056.1	22	21.43	38.0
OX459057.1	23	21.02	38.0
OX459058.1	24	20.8	38.0
OX459059.1	25	20.05	38.5
OX459060.1	26	20.01	38.0
OX459061.1	27	19.83	39.5
OX459062.1	28	18.96	38.0
OX459063.1	29	18.49	38.5
OX459064.1	30	18.29	38.0
OX459065.1	31	17.96	38.5
OX459066.1	32	17.86	38.5
OX459067.1	33	17.27	38.0
OX459068.1	34	16.75	38.0
OX459069.1	35	15.89	38.5
OX459070.1	36	10.48	39.5
OX459039.1	Z	37.11	38.0
OX459071.1	MT	0.02	18.5

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 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 bwamem2 (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 gene annotation system (Aken *et al.*, 2016) was used to generate annotation for the *Orthosia gothica* assembly (GCA_949775005.1). 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).

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 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.

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

Data availability

European Nucleotide Archive: *Orthosia gothica* (Hebrew character). Accession number PRJEB54944; <https://identifiers.org/ena.embl/PRJEB54944> (Wellcome Sanger Institute, 2023). The genome sequence is released openly for reuse. The *Orthosia gothica* 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](#)
- Bates A, Clayton-Lucey I, Howard C: **Sanger Tree of Life HMW DNA Fragmentation: Diagenote Megaruptor[®] for LI PacBio.** *protocols.io.* 2023. [Publisher 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](#)
- Bues R, Poitout S, Toubon JF, et al.: **Regulation of the unicyclic pest *Orthosia gothica* L (Lepidoptera, Noctuidae).** *Bulletin de La Société Zoologique de France - Evolution et Zoologie.* 1994; **119**: 201–209.
- 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, Yatsenko H, Jay J, et al.: **Sanger Tree of Life Wet Laboratory Protocol Collection V.1.** *protocols.io.* 2023. [Publisher Full Text](#)
- Di Tommaso P, Chatzou M, Floden EW, et al.: **Nextflow enables reproducible computational workflows.** *Nat Biotechnol.* 2017; **35**(4): 316–319. [PubMed Abstract](#) | [Publisher Full Text](#)
- do Amaral RJV, Bates A, Denton A, et al.: **Sanger Tree of Life RNA Extraction: Automated MagMax[™] mirVana.** *protocols.io.* 2023. [Publisher Full Text](#)
- Egaas E: **Effects of atrazine, endosulfan and butylated hydroxyanisole on glutathione-S-transferases in *Orthosia gothica*.** *Comp Biochem Physiol C Toxicol Pharmacol.* 2000; **127**(2): 117–22. [PubMed Abstract](#) | [Publisher Full Text](#)
- Egaas E, Svendsen NO, Kobro S, et al.: **Glutathione S-transferases in endosulfan-treated red sword grass moth (*Xylena vetusta* Hb.) and Hebrew character moth (*Orthosia gothica* L.) reared on leaves from apple (*Malus domestica* cult.) or willow (*Salix caprea* L.).** *Comp Biochem Physiol C Toxicol Pharmacol.* 1992; **101**(1): 143–150. [Publisher Full Text](#)
- GBIF Secretariat: ***Orthosia gothica* Linnaeus, 1758.** *GBIF Backbone Taxonomy.* 2023; [Accessed 30 December 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): gjaa153. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Jay J, Yatsenko H, Narváez-Gómez JP, et al.: **Sanger Tree of Life Sample Preparation: Triage and Dissection.** *protocols.io.* 2023. [Publisher Full Text](#)
- Kerpedjiev P, Abdennur N, Lekschas F, et al.: **HiGlass: web-based visual exploration and analysis of genome interaction maps.** *Genome Biol.* 2018; **19**(1): 125. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Manni M, Berkeley MR, 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](#)
- MothsIreland: **Hebrew Character (*Orthosia gothica*).** *MothsIreland.* 2023; [Accessed 30 December 2023]. [Reference Source](#)
- Narváez-Gómez JP, Mbye H, Oatley G, et al.: **Sanger Tree of Life Sample Homogenisation: Covaris cryoPREP[®] Automated Dry Pulverizer V.1.** *protocols.io.* 2023. [Publisher Full Text](#)
- NBN Atlas Partnership: ***Orthosia gothica* (Linnaeus, 1758) Hebrew Character.** *NBN Atlas.* 2023; [Accessed 30 December 2023]. [Reference Source](#)
- Oatley G, Denton A, Howard C: **Sanger Tree of Life HMW DNA Extraction: Automated MagAttract v.2.** *protocols.io.* 2023. [Publisher Full Text](#)
- Rao SSP, Huntley MH, Durand NC, et al.: **A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping.** *Cell.* 2014; **159**(7): 1665–1680. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Rhie A, McCarthy SA, Fedrigo O, et al.: **Towards complete and error-free genome assemblies of all vertebrate species.** *Nature.* 2021; **592**(7856): 737–746. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Rhie A, Walenz BP, Koren S, et al.: **Merqury: Reference-free quality, completeness, and phasing assessment for genome assemblies.** *Genome Biol.* 2020; **21**(1): 245. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)

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

[PubMed Abstract](#) | [Publisher Full Text](#)

South R: **Moths of the British Isles.** New edition. London: Frederick Warne and Co, 1961.

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.

[Reference Source](#)

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

[Publisher Full Text](#)

Szanyi S, Varga Z, Nagy A, et al.: **Bisexual lures and their comparison with synthetic sex attractants for trapping *Orthosia* species (Lepidoptera: Noctuidae).** *J Appl Entomol*. 2022; **146**(9): 1109–1115.

[Publisher Full Text](#)

Tóth M, Szöcs G, Bengtsson M: **A Sex Attractant for Male *Orthosia gothica* L.**

(Lepidoptera: Noctuidae). *Zeitschrift Fur Naturforschung - Section C Journal of Biosciences*. 1992; **47**(7–8): 613–617.

[Publisher Full Text](#)

Uliano-Silva M, Ferreira JGRN, Krashennikova 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 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 Hebrew Character. *Orthosia gothica* (Linnaeus, 1758).** European Nucleotide Archive, [dataset], accession number PRJEB54944. 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](#)

Open Peer Review

Current Peer Review Status:  

Version 1

Reviewer Report 17 October 2024

<https://doi.org/10.21956/wellcomeopenres.23130.r99894>

© 2024 Grapputo A. 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.



Alessandro Grapputo 

Dipartimento di Biologia, Universita degli Studi di Padova, Padua, Veneto, Italy

The paper by Boyes and colleagues present the reference genome of the Hebrew Character moth *Orthosia gothica* a spring-flying moth found across Eurasia. The data are well presented and accessible. The assembled genome is of high quality representing the reference genome for the specie. The paper is well written. The methods of DNA/RNA extraction, sequence acquisition, genome assembly pipeline and annotation with the Ensembl rapid annotation pipeline follow the classic protocol of most Data Note papers published by Wellcome Open Research. In summary the note is sound and good to be indexed.

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 ecology

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 10 September 2024

<https://doi.org/10.21956/wellcomeopenres.23130.r91876>

© 2024 **Rodriguez-de-la-vega R.** 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.



Ricardo Rodriguez-de-la-vega 

Écologie Systématique et Évolution, CNRS, Université Paris-Saclay, Paris, France

Comments on Boyes et al., "The genome sequence of the Hebrew character..."

In this data note, the Darwin Tree of Life's inordinate fondness for moths continues with the release of what is approximately the 314th moth genome*.

As is typical for the consortium's outputs, the combination of HiFi and Hi-C sequencing (both from the same specimen in this case) and manual curation has produced a chromosome-level assembly, which can be reused for further analysis. While the genome excels in all metrics presented in the data note, I would encourage readers to explore the extensive online materials for a more nuanced understanding of the genome's quality and specific features. To point out just two: heterozygosity seems unusually high, and the Hi-C fragment separation distribution is closer to the "bad example" when compared to other genome assemblies within the Noctuidae family.

A telling detail, only briefly mentioned in the hyper-standardised wording common in DToL data notes, is that the sequenced individual was a female, alas no W chromosome was identified, even though combined HiFi and Hi-C placed over 97.6% of the assembled contigs onto chromosomes.

Another not discussed point is about the large contig (11 Mb) with twice as much coverage as the chromosome anchored contigs in the GC-coverage, which might be a cobiont. The DToL consortium might need to consider updating their Data Note standards to start discussing these no-hit contigs.

Mind that in the text, the Z chromosome is said to have been identified based on coverage, but Table 1 suggests identification by homologous pairs.

I would like to thank the people in charge of peer reviewers assignments and DToL's commitment to presenting common names for giving me the opportunity to discover the world of butterfly/moth vernacular naming in the UK.

*DToL gateway queried on August 30th, 2024.

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: Evolutionary genomics, bioinformatics

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.
