Check for updates

DATA NOTE

The genome sequence of the Ash-bark Knot-horn, *Euzophera pinguis* (Haworth, 1811) [version 1; peer review: awaiting peer review]

Douglas Boyes[1+], 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 programme,
Wellcome Sanger Institute Scientific Operations: DNA Pipelines collective,
Tree of Life Core Informatics collective, Darwin Tree of Life Consortium

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

[+] Deceased author

v1 **First published:** 26 May 2023, **8**:228
https://doi.org/10.12688/wellcomeopenres.19393.1
Latest published: 26 May 2023, **8**:228
https://doi.org/10.12688/wellcomeopenres.19393.1

Abstract

We present a genome assembly from an individual male *Euzophera pinguis* (the Ash-bark Knot-horn; Arthropoda; Insecta; Lepidoptera; Pyralidae). The genome sequence is 464.7 megabases in span. Most of the assembly is scaffolded into 30 chromosomal pseudomolecules, including the Z sex chromosome. The mitochondrial genome has also been assembled and is 15.2 kilobases in length.

Keywords

Euzophera pinguis, Ash-bark Knot-horn, genome sequence, chromosomal, Lepidoptera

This article is included in the Tree of Life gateway.

Open Peer Review

Approval Status *AWAITING PEER REVIEW*

Any reports and responses or comments on the article can be found at the end of the article.

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

Author roles: Boyes D: Investigation, Resources; **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 Ash-bark Knot-horn, *Euzophera pinguis* (Haworth, 1811) [version 1; peer review: awaiting peer review]** Wellcome Open Research 2023, **8**:228 https://doi.org/10.12688/wellcomeopenres.19393.1

First published: 26 May 2023, **8**:228 https://doi.org/10.12688/wellcomeopenres.19393.1

Species taxonomy

Eukaryota; Metazoa; Ecdysozoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Lepidoptera; Glossata; Ditrysia; Pyraloidea; Pyralidae; Phycitinae; *Euzophera*; *Euzophera pinguis* (Haworth, 1811) (NCBI:txid1366396).

Background

Euzophera pinguis is a small moth in the family Pyralidae found widely across Europe, with scattered records from Russia and north Africa (GBIF Secretariat, 2022). In Britain, the moth has been recorded from south and central England and Wales, primarily from woodlands where stands of ash are present, although it is absent from many regions and is never present in large numbers (NBN Atlas Partnership, 2021). The first record for Ireland was in 2021 and there are few records from Scotland (Cubitt, 2021; Merne & O'Donnell, 2021).

The moth is known by several common names including the Ash-bark Knot-horn, Tabby Knot-horn, Cherry Bark Moth and Olive Pyralid Moth. As these names suggest, the moth can be associated with a range of woody host plants, varying with location. In Britain, the adult moth is on the wing in July and August and lays eggs on ash trees. The larvae burrow into the trunk where they feed on the inner bark layers, ejecting frass from the entrance hole. Larvae take one or two years to develop, pupating in the burrow, and can cause death of the ash tree (Beirne, 1952). In Spain and Tunisia, the moth has two generations per year and is a pest of olive trees; the larvae burrow under the bark causing considerable damage, particularly to young olive trees (Durán *et al.*, 1998; Jardak & Ksantini, 1996). In 2017, the species was noted as a new invasive pest of olive trees in Lebanon (Moussa *et al.*, 2017). In Bulgaria, *E. pinguis* has been reported as a pest of almond trees, also through the larvae burrowing under the bark of the trunk and larger branches (Ivanov, 1974). A synthetic pheromone has proved successful at disrupting mating under field test conditions in Spain (Ortiz *et al.*, 2004).

A complete genome sequence of *E. pinguis* will aid research into the biochemical adaptations permitting larval feeding on bark and will facilitate efforts to establish specific control measures.

Genome sequence report

The genome was sequenced from one male *Euzophera pinguis* (Figure 1) collected from Wytham Woods, Oxfordshire, UK (latitude 51.77, longitude –1.34) A total of 32-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 nine missing joins or mis-joins and removed five haplotypic duplications, reducing the scaffold number by 8.06%, and increasing the scaffold N50 by 3.11%.

The final assembly has a total length of 464.7 Mb in 57 sequence scaffolds with a scaffold N50 of 17.1 Mb (Table 1). Most (99.72%) of the assembly sequence was assigned to 30

Figure 1. Photograph of the *Euzophera pinguis* (ilEuzPing2) specimen used for genome sequencing.

chromosomal-level scaffolds, representing 29 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 61.6 with *k*-mer completeness of 100%, and the assembly has a BUSCO v5.3.2 completeness of 98.9% (single = 98.5%, duplicated = 0.3%), 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/1366396.

Methods

Sample acquisition and nucleic acid extraction

A male *Euzophera pinguis* (individual ilEuzPing2, specimen Ox000586) was collected from Wytham Woods, Oxfordshire (biological vice-county: Berkshire), UK (latitude 51.77, longitude –1.34) by Douglas Boyes (University of Oxford) on 5 July 2020. The specimen was taken from woodland habitat using a light trap. The specimen was also identified by Douglas Boyes and was snap-frozen on dry ice.

The ilEuzPing2 sample was prepared by the Tree of Life laboratory, Wellcome Sanger Institute (WSI). The sample was weighed and dissected on dry ice with tissue set aside for Hi-C sequencing. Whole organism tissue was disrupted using a Nippi Powermasher fitted with a BioMasher pestle. DNA was extracted at the Wellcome Sanger Institute (WSI) Scientific Operations core using the Qiagen MagAttract HMW DNA kit, according to the manufacturer's instructions.

Table 1. Genome data for *Euzophera pinguis*, ilEuzPing2.1.

Project accession data		
Assembly identifier	ilEuzPing2.1	
Species	*Euzophera pinguis*	
Specimen	ilEuzPing2	
NCBI taxonomy ID	1366396	
BioProject	PRJEB56485	
BioSample ID	SAMEA7701450	
Isolate information	ilEuzPing2, male (genome sequencing and Hi-C scaffolding)	
Assembly metrics*		***Benchmark***
Consensus quality (QV)	61.6	*≥ 50*
k-mer completeness	100%	*≥ 95%*
BUSCO**	C:98.9%[S:98.5%,D:0.3%], F:0.3%,M:0.8%,n:5,286	*C ≥ 95%*
Percentage of assembly mapped to chromosomes	99.72%	*≥ 95%*
Sex chromosomes	Z chromosome	*localised homologous pairs*
Organelles	Mitochondrial genome assembled	*complete single alleles*
Raw data accessions		
PacificBiosciences SEQUEL II	ERR10357394	
Hi-C Illumina	ERR10323137	
Genome assembly		
Assembly accession	GCA_947363495.1	
Accession of alternate haplotype	GCA_947365835.1	
Span (Mb)	464.7	
Number of contigs	129	
Contig N50 length (Mb)	7.7	
Number of scaffolds	57	
Scaffold N50 length (Mb)	17.1	
Longest scaffold (Mb)	28.3	

* 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/ilEuzPing2.1/dataset/CANASO01/busco.

Sequencing

Pacific Biosciences HiFi circular consensus 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) instrument. Hi-C data were also generated from tissue of ilEuzPing2 that had been set aside, 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

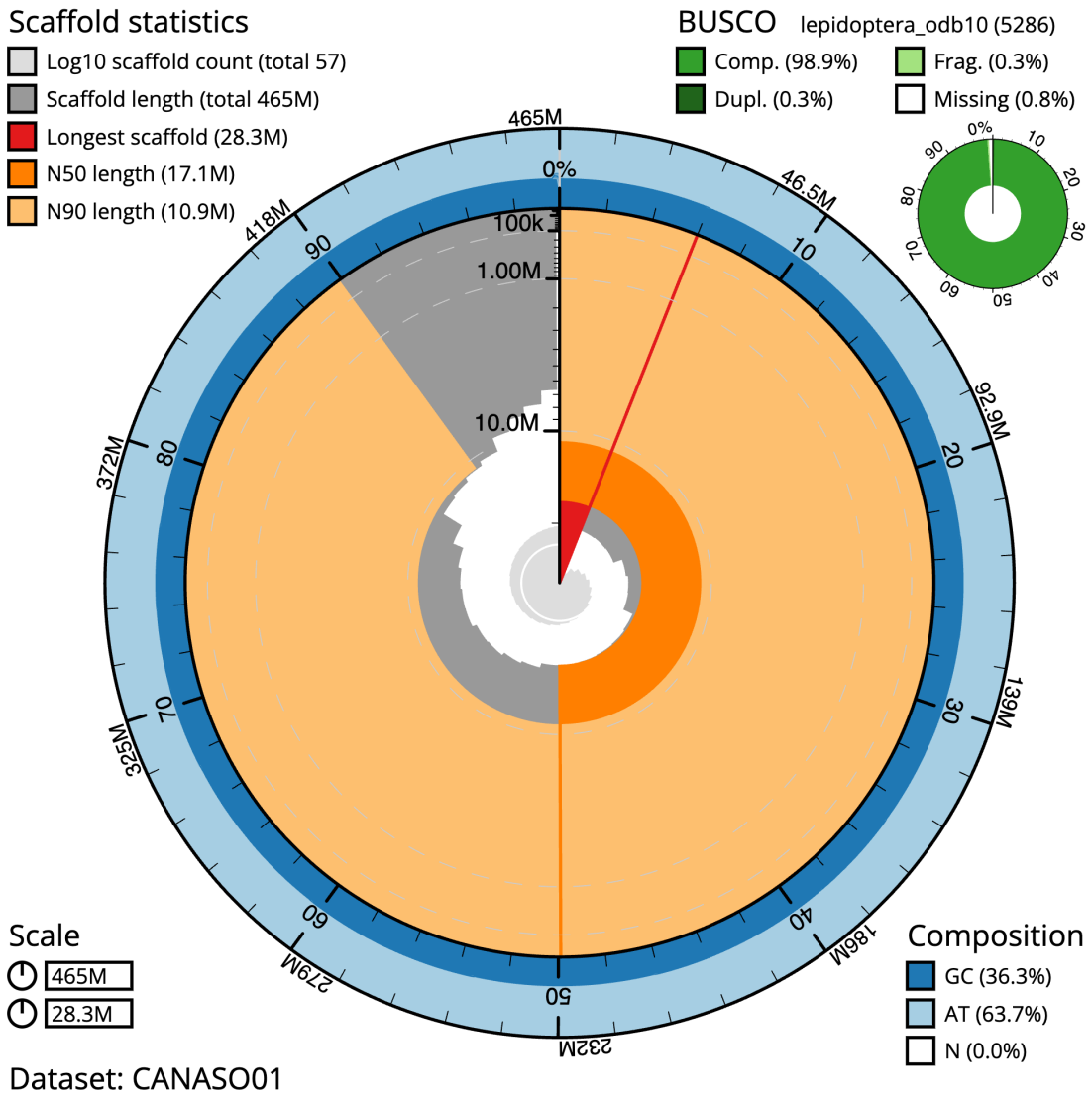

Figure 2. Genome assembly of *Euzophera pinguis*, ilEuzPing2.1: metrics. The BlobToolKit Snailplot shows N50 metrics and BUSCO gene completeness. The main plot is divided into 1,000 size-ordered bins around the circumference with each bin representing 0.1% of the 464,729,775 bp assembly. The distribution of scaffold lengths is shown in dark grey with the plot radius scaled to the longest scaffold present in the assembly (28,280,186 bp, shown in red). Orange and pale-orange arcs show the N50 and N90 scaffold lengths (17,073,911 and 10,923,421 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/ilEuzPing2.1/dataset/CANASO01/snail.

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 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;

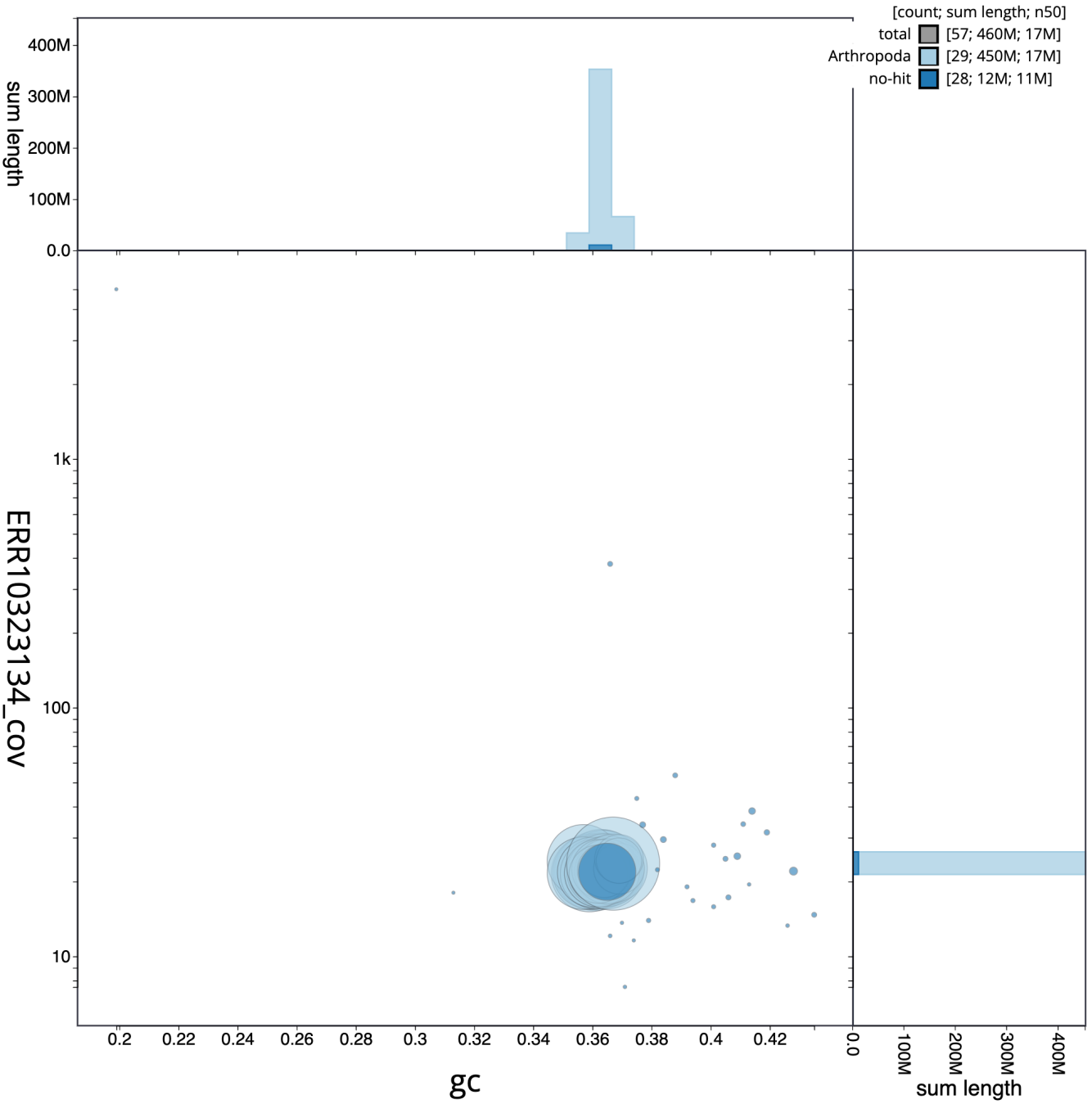

Figure 3. Genome assembly of *Euzophera pinguis*, ilEuzPing2.1: BlobToolKit GC-coverage plot. Scaffolds are coloured by phylum. Circles are sized in proportion to scaffold length. Histograms show the distribution of scaffold length sum along each axis. An interactive version of this figure is available at https://blobtoolkit.genomehubs.org/view/ilEuzPing2.1/dataset/CANASO01/blob.

Simão *et al.*, 2015) were calculated. Table 3 contains a list of software tool versions and sources.

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,

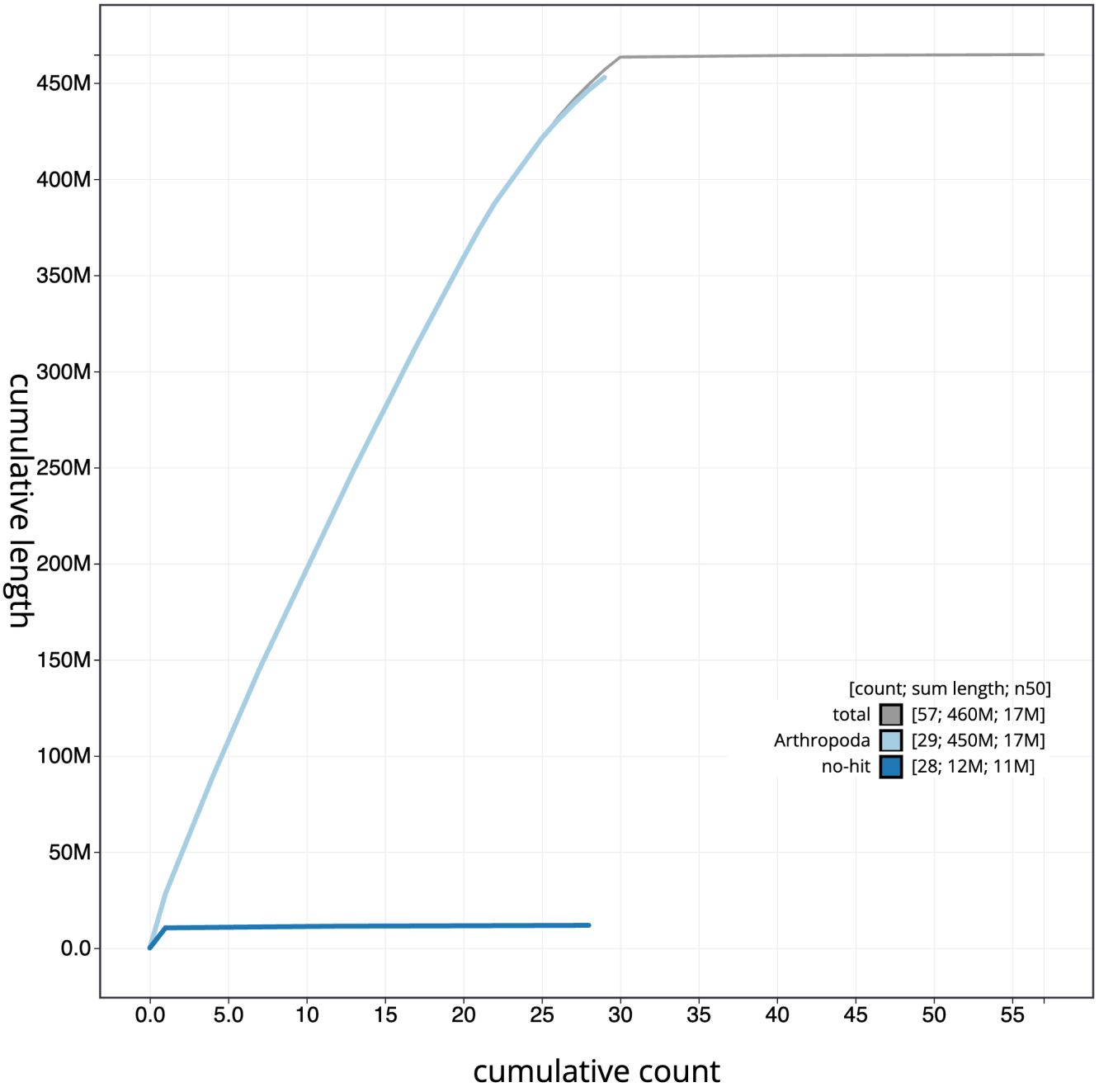

Figure 4. Genome assembly of *Euzophera pinguis*, ilEuzPing2.1: BlobToolKit cumulative sequence plot. The grey line shows cumulative length for all scaffolds. Coloured lines show cumulative lengths of scaffolds assigned to each phylum using the buscogenes taxrule. An interactive version of this figure is available at https://blobtoolkit.genomehubs.org/view/ilEuzPing2.1/dataset/CANASO01/cumulative.

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.

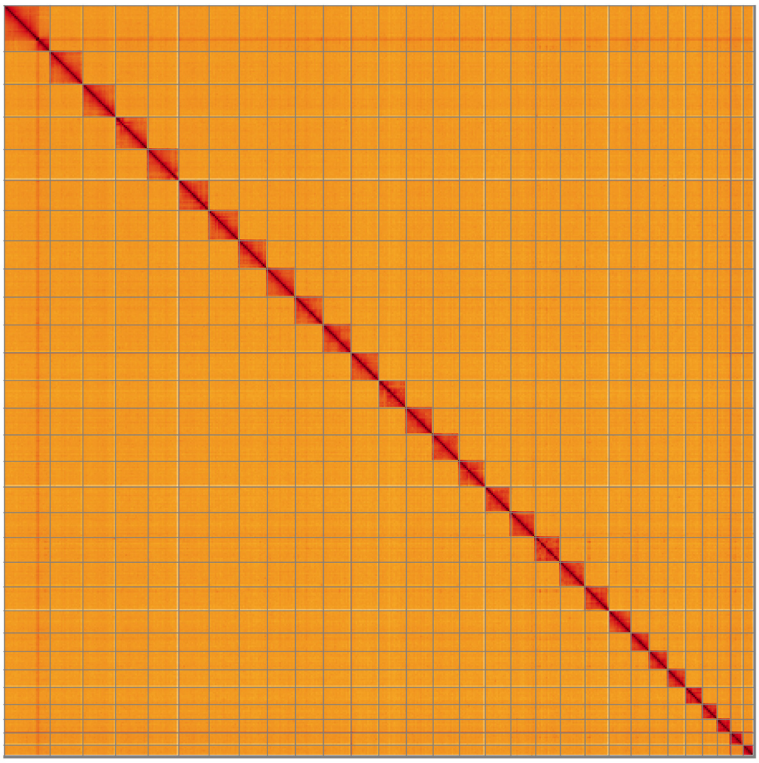

Figure 5. Genome assembly of *Euzophera pinguis*, ilEuzPing2.1: Hi-C contact map of the ilEuzPing2.1 assembly, visualised using HiGlass. Chromosomes are shown in order of size from left to right and top to bottom. An interactive version of this figure may be viewed at https://genome-note-higlass.tol.sanger.ac.uk/l/?d=YOoKUrcxQ-6Yi_TGrGF0XQ.

Table 2. Chromosomal pseudomolecules in the genome assembly of *Euzophera pinguis*, ilEuzPing2.

INSDC accession	Chromosome	Size (Mb)	GC%
OX376155.1	1	20.27	36.1
OX376156.1	2	20.2	36.3
OX376157.1	3	20.12	36.3
OX376158.1	4	18.99	35.9
OX376159.1	5	18.72	36.4
OX376160.1	6	18.66	36
OX376161.1	7	17.45	35.9
OX376162.1	8	17.31	36.5
OX376163.1	9	17.24	36.3
OX376164.1	10	17.17	36.3
OX376165.1	11	17.12	35.7
OX376166.1	12	17.07	35.7
OX376167.1	13	16.56	36.4
OX376168.1	14	16.3	36.5
OX376169.1	15	15.93	36.1
OX376170.1	16	15.84	36
OX376171.1	17	15.35	36.5
OX376172.1	18	15.32	36.3
OX376173.1	19	15.16	36.7
OX376174.1	20	14.8	36.2
OX376175.1	21	13.65	36.5
OX376176.1	22	11.5	36.5
OX376177.1	23	11.3	36.3
OX376178.1	24	10.92	36.2
OX376179.1	25	10.51	36.5
OX376180.1	26	9.25	36.4
OX376181.1	27	8.28	36.9
OX376182.1	28	7.59	36.9
OX376183.1	29	6.57	36.9
OX376154.1	Z	28.28	36.7
OX376184.1	MT	0.02	20.2

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

Data availability

European Nucleotide Archive: *Euzophera pinguis* (ash-bark knot-horn). Accession number PRJEB56485; https://identifiers.org/ena.embl/PRJEB56485. (Wellcome Sanger Institute, 2022)

The genome sequence is released openly for reuse. The *Euzophera pinguis* 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. The genome will be annotated using available RNA-Seq data and presented through the Ensembl pipeline at the European Bioinformatics Institute. 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

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

Beirne BP: **British Pyralid and Plume Moths.** London: Warne and Co, 1952; 208.
Reference Source

Bernt M, Donath A, Jühling F, *et al.*: **MITOS: Improved *de novo* metazoan mitochondrial genome annotation.** *Mol Phylogenet Evol.* 2013; **69**(2): 313–9.
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

Cubitt M: **Pyralidae, Scottish Micro Moth Distribution Maps.** 2021; (Accessed: 10 December 2022).
Reference Source

Durán JM, Alvarado M, Serrano A, *et al.*: **Contribucion al conocimiento de *Euzophera pinguis* (Haworth, [1881])(Lep: *Pyralidae*), plaga del olivo.** *Bol San Veg Plagas.* 1998; **24**: 267–278.
Reference Source

GBIF Secretariat: ***Euzophera pinguis* Haworth, 1811.** *GBIF Backbone Taxonomy.* 2022; (Accessed: 21 January 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): giaa153.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

Ivanov I: **Euzophera pinguis Hw. as a pest of almond.** *Rastitelna Zashchita.* 1974; **33**: 38–39.

Jardak T, Ksantini M: **La estructuración de la protección fitosanitaria del olivo en Túnez: Elementos básicos y necesidades económicas y ecológicas.** *Olivae: revista oficial del Consejo Oleícola Internacional.* 1996; **61**: 24–33.
Reference Source

Kerpedjiev P, Abdennur N, Lekschas F, *et al.*: **HiGlass: Web-based visual exploration and analysis of genome interaction maps.** *Genome Biol.* 2018; **19**(1): 125.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

Manni M, Berkeley MR, Seppey M, *et al.*: **BUSCO Update: Novel and Streamlined Workflows along with Broader and Deeper Phylogenetic Coverage for Scoring of Eukaryotic, Prokaryotic, and Viral Genomes.** *Mol Biol Evol.* 2021; **38**(10): 4647–4654.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

Merne C, O'Donnell G: **62.048 Euzophera Pinguis. 1st Irish Records!** Moths Ireland, 2021; (Accessed: 5 April 2023).
Reference Source

Moussa Z, Choueiri E, Youssef A, *et al.*: **First record of Olive Pyralid Moth, *Euzophera pinguis* (Haworth, 1811) on olive trees in Lebanon (Lepidoptera, Pyralidae).** *Bulletin de la Société entomologique de France.* 2017; **122**: 57–60.
Reference Source

NBN Atlas Partnership: ***Euzophera pinguis* (Haworth, 1811) Ash-bark Knot-horn.** NBN Atlas, 2021; (Accessed: 21 January 2023).
Reference Source

Ortiz A, Quesada A, Sanchez A: **Potential for Use of Synthetic Sex Pheromone for Mating Disruption of the Olive Pyralid Moth, *Euzophera pinguis*.** *J Chem Ecol.* 2004; **30**(5): 991–1000.
PubMed Abstract | **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**

Uliano-Silva M, Ferreira JGRN, Krasheninnikova K, *et al.*: **MitoHiFi: a python pipeline for mitochondrial genome assembly from PacBio High Fidelity reads.** *bioRxiv.* [Preprint], 2022.
Publisher Full Text

Wellcome Sanger Institute: **The genome sequence of the Ash-bark Knot-horn, *Euzophera pinguis* (Haworth, 1811).** European Nucleotide Archive. [dataset], accession number PRJEB56485, 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**