DATA NOTE

The genome sequence of the Holly Tortrix, *Rhopobota naevana* (Hubner, 1817) (Lepidoptera: Tortricidae)

[version 1; peer review: 2 approved]

Douglas Boyes[1+], Finley Hutchinson[2], Liam M. Crowley[3], Cian D. Williams[4], Clare 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]UK Centre for Ecology & Hydrology, Wallingford, England, UK
[2]University of Exeter, Penryn, England, UK
[3]University of Oxford, Oxford, England, UK
[4]University of Cambridge, Cambridge, England, UK
[5]Independent researcher, Welshpool, Wales, UK

[+] Deceased author

v1 **First published:** 27 Jun 2026, **11**:400
https://doi.org/10.12688/wellcomeopenres.26882.1
Latest published: 27 Jun 2026, **11**:400
https://doi.org/10.12688/wellcomeopenres.26882.1

Abstract

We present a genome assembly from an individual male *Rhopobota naevana* (Holly Tortrix; Arthropoda; Insecta; Lepidoptera; Tortricidae). The genome sequence has a total length of 581.80 megabases. Most of the assembly (99.48%) is scaffolded into 28 chromosomal pseudomolecules, including the Z sex chromosome. The mitochondrial genome has also been assembled, with a length of 16.5 kilobases. This assembly was generated as part of the Darwin Tree of Life project, which produces genomes for eukaryotic species found in Britain and Ireland.

Keywords

Rhopobota naevana, Holly Tortrix, genome sequence, chromosomal, Lepidoptera

Open Peer Review

Approval Status ✓ ✓

	1	2
version 1 27 Jun 2026	✓ view	✓ view

1. **Sivasankaran Kuppusamy**, Loyola College, Chennai, India

2. **Nicola J Nadeau**, The University of Sheffield, Sheffield, UK

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: Boyes D: Investigation, Resources; **Hutchinson F**: Investigation, Resources; **Crowley LM**: Investigation, Resources; **Williams CD**: Writing – Original Draft Preparation; **Boyes C**: Writing – Original Draft Preparation;

Competing interests: No competing interests were disclosed.

Grant information: This work was supported by Wellcome through core funding to the Wellcome Sanger Institute (220540) 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.

How to cite this article: Boyes D, Hutchinson F, Crowley LM *et al.* **The genome sequence of the Holly Tortrix, *Rhopobota naevana* (Hubner, 1817) (Lepidoptera: Tortricidae) [version 1; peer review: 2 approved]** Wellcome Open Research 2026, **11**:400 https://doi.org/10.12688/wellcomeopenres.26882.1

First published: 27 Jun 2026, **11**:400 https://doi.org/10.12688/wellcomeopenres.26882.1

Species taxonomy

Eukaryota; Opisthokonta; Metazoa; Eumetazoa; Bilateria; Protostomia; Ecdysozoa; Panarthropoda; Arthropoda; Mandibulata; Pancrustacea; Altocrustacea; Allotriocarida; Hexapoda; Insecta; Dicondylia; Pterygota; Neoptera; Eumetabola; Endopterygota; Aparaglossata; Panorpida; Amphiesmenoptera; Lepidoptera; Glossata; Neolepidoptera; Heteroneura; Ditrysia; Apoditrysia; Tortricoidea; Tortricidae; Olethreutinae; Eucosmini; *Rhopobota*; *Rhopobota naevana* (Hubner, 1817) (NCBI:txid572892).

Background

The Holly Tortrix, *Rhopobota naevana* (Hübner, 1817), is a small moth in the family Tortricidae. It is common in the British Isles and has a wide distribution across Europe, Asia and North America, with additional records from Madagascar (GBIF Secretariat, 2025). Adults are variable in colour, but can be recognised by the hooked apex of the forewing and, in females, by a distinctive crease in the folded wing (Bradley *et al.*, 1979; Sterling *et al.*, 2023).

The larvae feed on a range of woody plants, including holly, bilberry, heather and cranberry (Hancock *et al.*, 2015; Huie, 1917). They feed within young leaves or buds, often protected by silk spinnings (Huie, 1917). The species is also known as the Blackheaded Fireworm and is an economically important pest of cranberry, where larval feeding can damage developing leaves, buds and fruit (Deland *et al.*, 2014; Fitzpatrick, 2006; Labarre *et al.*, 2024).

The genome of *Rhopobota naevana* was sequenced as part of the Darwin Tree of Life Project, a collaborative effort to sequence the named eukaryotic species of Britain and Ireland. Here we present a chromosomally complete genome sequence for *R. naevana*, based on one specimen from Wytham Woods, Oxfordshire, UK (Figure 1).

Methods

Sample acquisition and DNA barcoding

The specimen used for genome sequencing was an adult male *Rhopobota naevana* (specimen ID Ox000655, ToLID ilRhoNaev1; Figure 1), collected from Wytham Woods, Oxfordshire, UK (latitude 51.772, longitude −1.338) on 2020-07-20. The specimen was collected and identified by Douglas Boyes. A second specimen was used for Hi-C sequencing (specimen ID Ox003072, ToLID ilRhoNaev2). It was collected from Wytham Woods, Oxfordshire, United Kingdom (latitude 51.772, longitude −1.338) on 2022-07-22. The specimen was collected by Finley Hutchinson and Liam Crowley and identified by Finley Hutchinson.

Figure 1. Photograph of the *Rhopobota naevana* (ilRhoNaev1) specimen used for genome sequencing.

The initial identification was verified by an additional DNA barcoding process according to the framework developed by Twyford *et al.* (2024). A small sample was dissected from the specimen and stored in ethanol, while the remaining parts were shipped on dry ice to the Wellcome Sanger Institute (WSI) (see the protocol). The tissue was lysed, the COI marker region was amplified by PCR, and amplicons were sequenced and compared to the BOLD database, confirming the species identification (Crowley *et al.*, 2023). Following whole genome sequence generation, the relevant DNA barcode region was also used alongside the initial barcoding data for sample tracking at the WSI (Twyford *et al.*, 2024). The standard operating procedures for Darwin Tree of Life barcoding are available on protocols.io.

Nucleic acid extraction

Detailed protocols for nucleic acid extraction developed at the Wellcome Sanger Institute (WSI) Tree of Life Core Laboratory are available on protocols.io (Howard *et al.*, 2025). The ilRhoNaev1 specimen was weighed and triaged to determine the appropriate extraction protocol. Tissue from the whole organism was homogenised by powermashing using a PowerMasher II tissue disruptor. High molecular weight (HMW) DNA was extracted using the Automated MagAttract v2 protocol. DNA was sheared into an average fragment size of 12–20 kb following the Megaruptor®3 for LI PacBio protocol. Sheared DNA was purified by automated SPRI (solid-phase reversible immobilisation). The concentration of the sheared and purified DNA was assessed using a Nanodrop spectrophotometer and Qubit Fluorometer using the Qubit dsDNA High Sensitivity Assay kit. Fragment size distribution was evaluated by running the sample on the FemtoPulse system.

PacBio HiFi library preparation and sequencing

Library preparation and sequencing were performed at the WSI Scientific Operations core. Libraries were prepared using the SMRTbell Prep Kit 3.0 (Pacific Biosciences, California, USA), following the manufacturer's instructions. The kit includes reagents for end repair/A-tailing, adapter ligation, post-ligation SMRTbell bead clean-up, and nuclease treatment. Size selection and clean-up were performed using diluted AMPure PB beads (Pacific Biosciences). DNA concentration was quantified using a Qubit Fluorometer v4.0 (ThermoFisher Scientific) and the Qubit 1X dsDNA HS assay kit. Final library fragment size was assessed with the Agilent Femto Pulse Automated Pulsed Field CE Instrument (Agilent Technologies) using the gDNA 55 kb BAC analysis kit.

The sample was sequenced using the Sequel IIe system (Pacific Biosciences, California, USA). The concentration of the library loaded onto the Sequel IIe was in the range 40–135 pM. SMRT Link software, a PacBio web-based workflow manager, was used to set up and monitor the run, and to perform primary and secondary analysis of the data upon completion.

Hi-C

Sample preparation and crosslinking

The Hi-C sample was prepared from 20–50 mg of frozen tissue from the ilRhoNaev2 sample using the Arima-HiC v2 kit (Arima Genomics). Following the manufacturer's instructions, tissue was fixed and DNA crosslinked using TC buffer to a final formaldehyde concentration of 2%. The tissue was homogenised using the Diagnocine Power Masher-II. Crosslinked DNA was digested with a restriction enzyme master mix, biotinylated, and ligated. Clean-up was performed with SPRISelect beads before library preparation. DNA concentration was measured with the Qubit Fluorometer (Thermo Fisher Scientific) and Qubit HS Assay Kit. The biotinylation percentage was estimated using the Arima-HiC v2 QC beads.

Hi-C library preparation and sequencing

Biotinylated DNA constructs were fragmented using a Covaris E220 sonicator and size selected to 400–600 bp using SPRISelect beads. DNA was enriched with Arima-HiC v2 kit Enrichment beads. End repair, A-tailing, and adapter ligation were carried out with the NEBNext Ultra II DNA Library Prep Kit (New England Biolabs), following a modified protocol where library preparation occurs while DNA remains bound to the Enrichment beads. Library amplification was performed using KAPA HiFi HotStart mix and a custom Unique Dual Index (UDI) barcode set (Integrated DNA Technologies). Depending on sample concentration and biotinylation percentage determined at the crosslinking stage, libraries were amplified with 10–16 PCR cycles. Post-PCR clean-up was performed with SPRISelect beads. Libraries were quantified using the AccuClear Ultra High Sensitivity dsDNA Standards Assay Kit (Biotium) and a FLUOstar Omega plate reader (BMG Labtech).

Prior to sequencing, libraries were normalised to 10 ng/μL. Normalised libraries were quantified again to create equimolar and/or weighted 2.8 nM pools. Pool concentrations were checked using the Agilent 4200 TapeStation (Agilent) with High Sensitivity D500 reagents before sequencing. Sequencing was performed using paired-end 150 bp reads on the Illumina NovaSeq 6000.

Genome assembly

Prior to assembly of the PacBio HiFi reads, a database of *k*-mer counts ($k = 31$) was generated from the filtered reads using FastK. GenomeScope2 (Ranallo-Benavidez *et al.*, 2020) was used to analyse the *k*-mer frequency distributions, providing estimates of genome size, heterozygosity, and repeat content.

The HiFi reads were assembled using Hifiasm (Cheng *et al.*, 2021) with the --primary option. The Hi-C reads (Rao *et al.*, 2014) were mapped to the primary contigs using bwa-mem2 (Vasimuddin *et al.*, 2019), and the contigs were scaffolded in YaHS (Zhou *et al.*, 2023) with the --break option for handling potential misassemblies. The scaffolded assemblies were evaluated using Gfastats (Formenti *et al.*, 2022), BUSCO (Manni *et al.*, 2021) and MerquryFK (Rhie *et al.*, 2020).

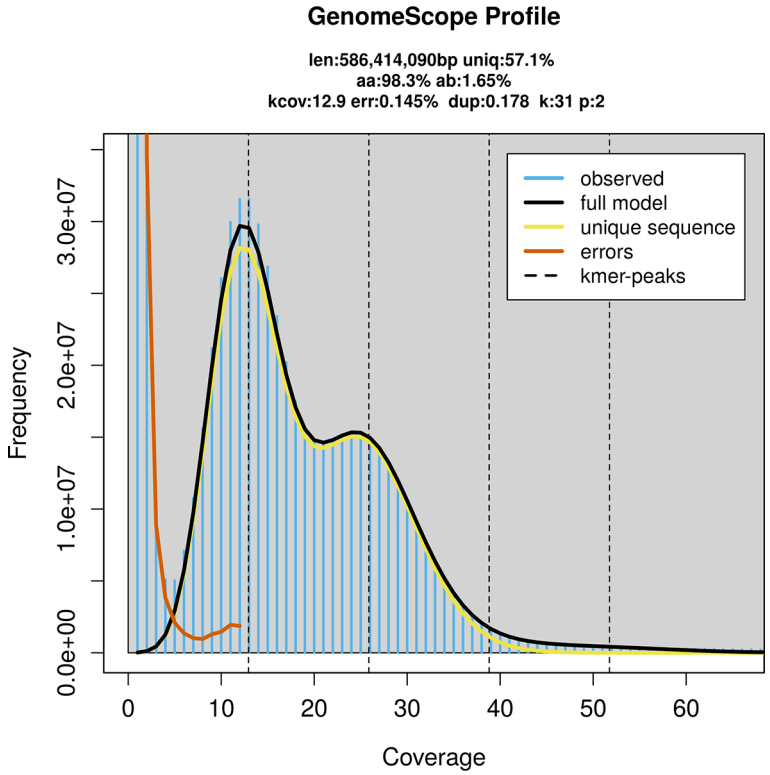

Figure 2. Frequency distribution of *k*-mers generated using GenomeScope2. The plot shows observed and modelled *k*-mer spectra, providing estimates of genome size, heterozygosity, and repeat content based on unassembled sequencing reads.

Table 1. Specimen and sequencing data for *Rhopobota naevana* (BioProject PRJEB55964).

Platform	PacBio HiFi	Hi-C
ToLID	ilRhoNaev1	ilRhoNaev2
Specimen ID	Ox000655	Ox003072
BioSample (source individual)	SAMEA7701517	SAMEA113425680
BioSample (tissue)	SAMEA7701699	SAMEA113425861
Tissue	whole organism	whole organism
Instrument	Sequel	Illumina NovaSeq 6000
Run accessions	ERR10224907	ERR13925921
Read count total	1.19 million reads	385.33 million read pairs
Base count total	15.91 Gb	116.37 Gb

Table 2. **Genome assembly data for *Rhopobota naevana*.**

Genome assembly	Primary assembly
Assembly name	ilRhoNaev1.1
Assembly accession	GCA_965641425.1
Alternate haplotype accession	GCA_965641315.1
Assembly level	chromosome
Span (Mb)	581.80
Number of chromosomes	28
Number of contigs	248
Contig N50	7.16 Mb
Number of scaffolds	67
Scaffold N50	21.26 Mb
Sex chromosomes	Z
Organelles	Mitochondrial genome: 16.50 kb

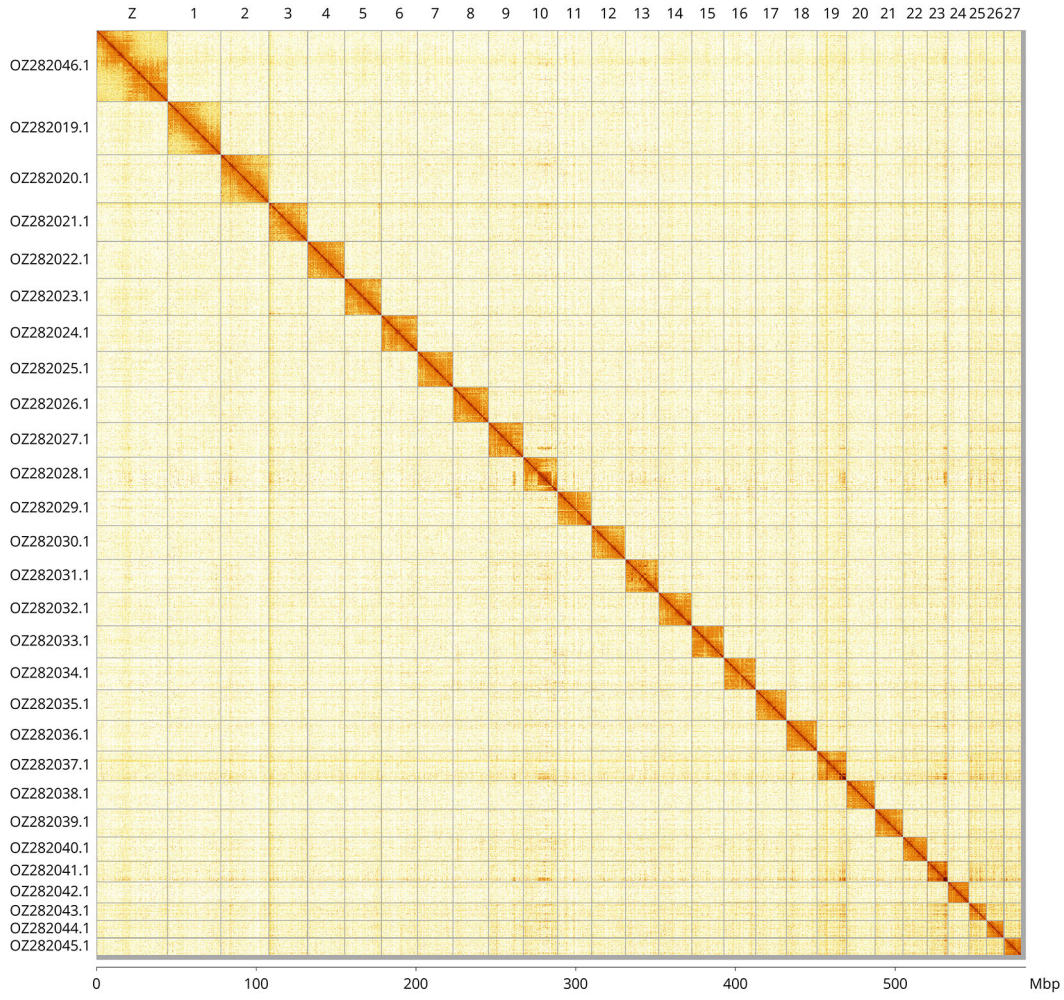

Figure 3. **Hi-C contact map of the *Rhopobota naevana* genome assembly.** Assembled chromosomes are shown in order of size and labelled along the axes, with a megabase scale shown below. The plot was generated using PretextSnapshot.

The mitochondrial genome was assembled using MitoHiFi (Uliano-Silva *et al.*, 2023). The MitoHiFi reference was the mitochondrial genome of *Epinotia rubiginosana* (NC_067880.1).

Assembly curation

The assembly was decontaminated using the Assembly Screen for Cobionts and Contaminants (ASCC) pipeline. TreeVal was used to generate the flat files and maps for use in curation. Manual curation was conducted primarily in PretextView and HiGlass (Kerpedjiev *et al.*, 2018). Scaffolds were visually inspected and corrected as described by Howe *et al.* (2021). Manual corrections included 101 breaks and 100 joins. This reduced the scaffold count by 11.7%, reduced the scaffold N50 by 0.8%, and reduced the total assembly length by 2.4%. The curation process is described at https://gitlab.com/wtsi-grit/rapid-curation. PretextSnapshot was used to generate a Hi-C contact map of the final assembly.

Assembly quality assessment

The MerquryFK tool (Rhie *et al.*, 2020) was run in a Singularity container (Kurtzer *et al.*, 2017) to evaluate *k*-mer completeness and assembly quality for the primary and alternate haplotypes using the *k*-mer database ($k = 31$) computed prior to genome assembly. The analysis outputs included assembly QV scores and completeness statistics.

Table 3. Chromosomal pseudomolecules in the primary genome assembly of *Rhopobota naevana* ilRhoNaev1.

INSDC accession	Molecule	Length (Mb)	GC%
OZ282019.1	1	33.16	38
OZ282020.1	2	30.09	38
OZ282021.1	3	24.23	38.50
OZ282022.1	4	23.14	38.50
OZ282023.1	5	23.01	39
OZ282024.1	6	22.59	38.50
OZ282025.1	7	22.30	38.50
OZ282026.1	8	22.14	38
OZ282027.1	9	21.81	38.50
OZ282028.1	10	21.36	38.50
OZ282029.1	11	21.26	38.50
OZ282030.1	12	21.22	38.50
OZ282031.1	13	20.74	39
OZ282032.1	14	20.73	38.50
OZ282033.1	15	20.11	38.50
OZ282034.1	16	19.90	38.50
OZ282035.1	17	19.19	38.50
OZ282036.1	18	19.18	39
OZ282037.1	19	18.42	39.50
OZ282038.1	20	17.81	39.50
OZ282039.1	21	17.62	39
OZ282040.1	22	15.02	38.50
OZ282041.1	23	13.08	40.50
OZ282042.1	24	13.07	39
OZ282043.1	25	11.13	40
OZ282044.1	26	10.99	39.50
OZ282045.1	27	10.78	39
OZ282046.1	Z	44.67	38.50

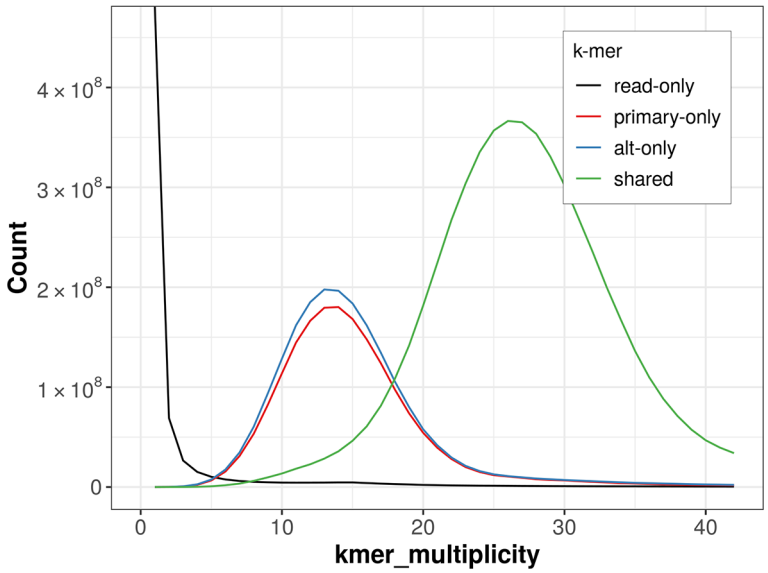

Figure 4. Evaluation of *k*-mer completeness using MerquryFK. This plot illustrates the recovery of *k*-mers from the original read data in the final assemblies. The horizontal axis represents *k*-mer multiplicity, and the vertical axis shows the number of *k*-mers. The black curve represents *k*-mers that appear in the reads but are not assembled. The green curve corresponds to *k*-mers shared by both haplotypes, and the red and blue curves show *k*-mers found only in one of the haplotypes.

The genome was analysed using the BlobToolKit pipeline, a Nextflow implementation of the earlier Snakemake version (Challis *et al.*, 2020). The pipeline aligns PacBio reads using minimap2 (Li, 2018) and SAMtools (Danecek *et al.*, 2021) to generate coverage tracks. It runs BUSCO (Manni *et al.*, 2021) using lineages identified from the NCBI Taxonomy (Schoch *et al.*, 2020). For the three domain-level lineages, BUSCO genes are aligned to the UniProt Reference Proteomes database (Bateman *et al.*, 2023) using DIAMOND blastp (Buchfink *et al.*, 2021). The genome is divided into chunks based on the density of BUSCO genes from the closest taxonomic lineage, and each chunk is aligned to the UniProt Reference Proteomes database with DIAMOND blastx. Sequences without hits are chunked using seqtk and aligned to the NT database with blastn (Altschul *et al.*, 1990). The BlobToolKit suite consolidates all outputs into a blobdir for visualisation. The BlobToolKit pipeline was developed using nf-core tooling (Ewels *et al.*, 2020) and MultiQC (Ewels *et al.*, 2016), with containerisation through Docker (Merkel, 2014) and Singularity (Kurtzer *et al.*, 2017).

Genome sequence report

Sequence data

PacBio sequencing of the *Rhopobota naevana* specimen generated 15.91 Gb (gigabases) from 1.19 million reads, which were used to assemble the genome. GenomeScope2.0 analysis estimated the haploid genome size at 588.45 Mb, with a heterozygosity of 1.65% and repeat content of 42.86% (Figure 2). These estimates guided expectations for the assembly. Based on the estimated genome size, the sequencing data provided approximately 26× coverage. Hi-C sequencing produced 116.37 Gb from 385.33 million reads, which were used to scaffold the assembly. Table 1 summarises the specimen and sequencing details.

Assembly statistics

The primary haplotype was assembled, and contigs corresponding to an alternate haplotype were also deposited in INSDC databases. The final assembly has a total length of 581.80 Mb in 67 scaffolds, with 181 gaps, and a scaffold N50 of 21.26 Mb (Table 2).

Most of the assembly sequence (99.48%) was assigned to 28 chromosomal-level scaffolds, representing 27 autosomes and the Z sex chromosome. These chromosome-level scaffolds, confirmed by Hi-C data, are named according to size (Figure 3; Table 3). The Z chromosomes were identified based on BUSCO gene painting with ancestral Merian elements (Wright *et al.*, 2024). During curation we noted that scaffolds in the following regions have uncertain order and orientation: Chromosome 13 from ~12.25–12.78 Mb and Chromosome 26 from ~10.35–10.70 Mb.

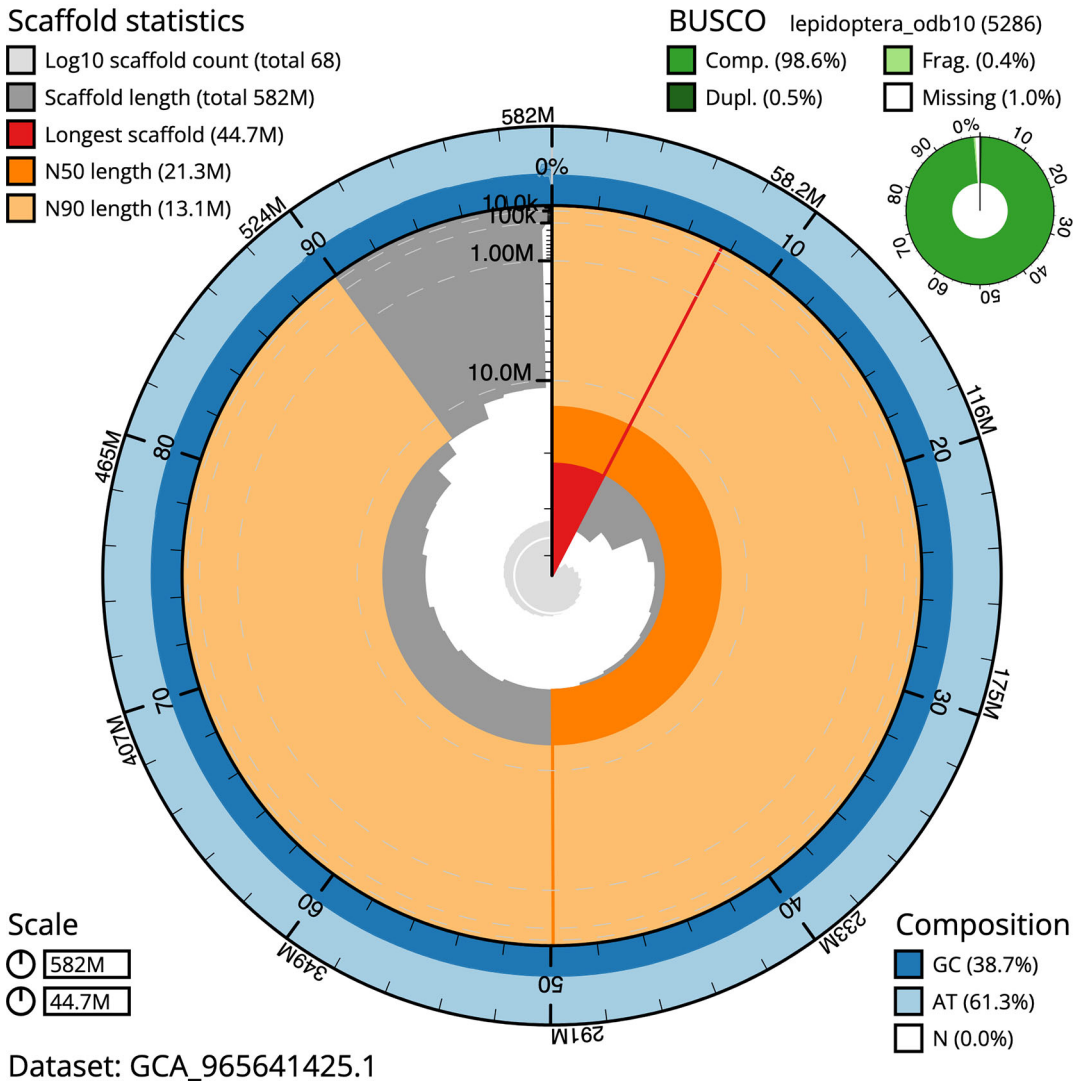

Figure 5. Assembly metrics for ilRhoNaev1.1. The BlobToolKit snail plot provides an overview of assembly metrics and BUSCO gene completeness. The circumference represents the length of the whole genome sequence, and the main plot is divided into 1 000 bins around the circumference. The outermost blue tracks display the distribution of GC, AT, and N percentages across the bins. Scaffolds are arranged clockwise from longest to shortest and are depicted in dark grey. The longest scaffold is indicated by the red arc, and the deeper orange and pale orange arcs represent the N50 and N90 lengths. A light grey spiral at the centre shows the cumulative scaffold count on a logarithmic scale. A summary of complete, fragmented, duplicated, and missing BUSCO genes in the lepidoptera_odb10 set is presented at the top right. An interactive version of this figure can be accessed on the BlobToolKit viewer.

The mitochondrial genome was also assembled (length 16.5 kb, OZ282047.1). This sequence is included as a contig in the multifasta file of the genome submission and as a standalone record.

Assembly quality metrics

The combined primary and alternate assemblies achieve an estimated QV of 65.4. The *k*-mer completeness is 70.31% for the primary assembly, 73.04% for the alternate haplotype, and 98.61% for the combined assemblies (Figure 4).

BUSCO v.5.8.3 analysis using the lepidoptera_odb10 reference set (n = 5 286) identified 98.6% of the expected gene set (single = 98.2%, duplicated = 0.5%). The snail plot in Figure 5 summarises the scaffold length distribution and other assembly statistics for the primary assembly. The blob plot in Figure 6 shows the distribution of scaffolds by GC proportion and coverage.

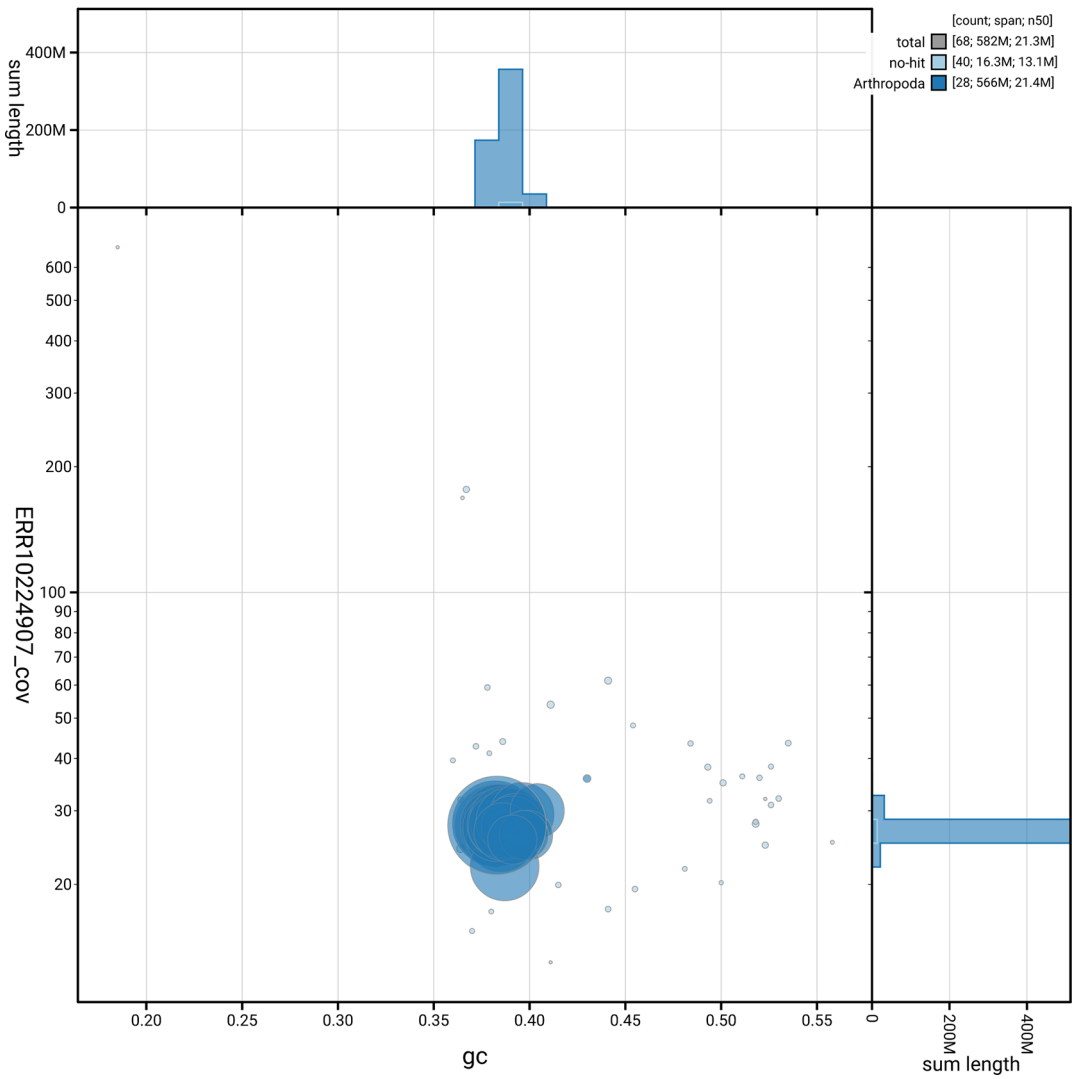

Figure 6. BlobToolKit blob plot for ilRhoNaev1.1. The plot shows base coverage (vertical axis) and GC content (horizontal axis). The circles represent scaffolds, with the size proportional to scaffold length and the colour representing phylum membership. The histograms along the axes display the total length of sequences distributed across different levels of coverage and GC content. An interactive version of this figure is available on the BlobToolKit viewer.

Table 4. Earth Biogenome Project summary metrics for the *Rhopobota naevana* assembly.

Measure	Value	Benchmark
EBP summary (primary)	6.C.Q65	6.C.Q40
Contig N50 length	7.16 Mb	≥ 1 Mb
Scaffold N50 length	21.26 Mb	= chromosome N50
Consensus quality (QV)	Primary: 65.7; alternate: 65.2; combined: 65.4	≥ 40
k-mer completeness	Primary: 70.31%; alternate: 73.04%; combined: 98.61%	≥ 95%
BUSCO	C:98.6% [S:98.2%, D:0.5%], F:0.4%, M:1.0%, n:5 286	S > 90%; D < 5%
Percentage of assembly assigned to chromosomes	99.48%	≥ 90%

Notes: The EBP summary uses log10(Contig N50); chromosome-level (C) or log10(Scaffold N50); Q (Merqury QV). BUSCO: C = complete; S = single-copy; D = duplicated; F = fragmented; M = missing; n = orthologues.

Table 4 lists the assembly metric benchmarks adapted from Rhie *et al.* (2021) and the Earth BioGenome Project Report on Assembly Standards. The EBP metric, calculated for the primary assembly, is **6.C.Q65**, meeting the recommended 6.C.Q40 reference standard.

Author information

Contributors are listed at the following links:

- Members of the University of Oxford and Wytham Woods Genome Acquisition Lab
- Members of the Darwin Tree of Life Barcoding collective
- Members of the Wellcome Sanger Institute Tree of Life Management, Samples and Laboratory team
- Members of Wellcome Sanger Institute Scientific Operations – Sequencing Operations
- Members of the Wellcome Sanger Institute Tree of Life Core Informatics team
- Members of the Tree of Life Core Informatics collective
- Members of the Darwin Tree of Life Consortium

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. 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.

Data availability

European Nucleotide Archive: Rhopobota naevana (holly tortrix). Accession number PRJEB55964; https://identifiers.org/ena.embl/PRJEB55964. The genome sequence is released openly for reuse. The *Rhopobota naevana* genome sequencing initiative is part of the Darwin Tree of Life Project (PRJEB40665), the Sanger Institute Tree of Life Programme (PRJEB43745) and Project Psyche (PRJEB71705). 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 Tables 1 and 2.

Production code used in genome assembly at the WSI Tree of Life is available at https://github.com/sanger-tol. Table 5 lists software versions used in this study.

Table 5. Software versions and sources used for *Rhopobota naevana*.

Software	Version	Source
BLAST	2.14.0	ftp://ftp.ncbi.nlm.nih.gov/blast/executables/blast+/
BlobToolKit	4.4.6	https://github.com/blobtoolkit/blobtoolkit
BUSCO	5.8.3	https://gitlab.com/ezlab/busco
bwa-mem2	2.2.1	https://github.com/bwa-mem2/bwa-mem2
DIAMOND	2.1.8	https://github.com/bbuchfink/diamond
fasta_windows	0.2.4	https://github.com/tolkit/fasta_windows
FastK	1.1	https://github.com/thegenemyers/FASTK
GenomeScope2.0	2.0.1	https://github.com/tbenavi1/genomescope2.0
Gfastats	1.3.6	https://github.com/vgl-hub/gfastats
Hifiasm	0.19.8-r603	https://github.com/chhylp123/hifiasm
HiGlass	1.13.4	https://github.com/higlass/higlass
MerquryFK	1.1.0-c1	https://github.com/thegenemyers/MERQURY.FK
Minimap2	2.24-r1122; 2.28-r1209	https://github.com/lh3/minimap2
MitoHiFi	3.2.2	https://github.com/marcelauliano/MitoHiFi
MultiQC	1.14; 1.17 and 1.18	https://github.com/MultiQC/MultiQC
Nextflow	23.10.0; 24.10.4	https://github.com/nextflow-io/nextflow
PretextSnapshot	0.0.4	https://github.com/sanger-tol/PretextSnapshot
PretextView	1.0.3	https://github.com/sanger-tol/PretextView
samtools	1.17; 1.18; 1.2; 1.21; 1.6	https://github.com/samtools/samtools
sanger-tol/ascc	0.1.0	https://github.com/sanger-tol/ascc
sanger-tol/blobtoolkit	v0.8.0	https://github.com/sanger-tol/blobtoolkit
sanger-tol/curationpretext	1.4.2	https://github.com/sanger-tol/curationpretext
Seqtk	1.4-r122	https://github.com/lh3/seqtk
Singularity	3.9.0	https://github.com/sylabs/singularity
TreeVal	1.1.1	https://github.com/sanger-tol/treeval
YaHS	1.2a.2	https://github.com/c-zhou/yahs

References

Altschul SF, Gish W, Miller W, *et al.*: **Basic Local Alignment Search Tool.** *J. Mol. Biol.* 1990; **215**(3): 403–410.
PubMed Abstract | Publisher Full Text

Bateman A, Martin M-J, Orchard S, *et al.*: **UniProt: The Universal Protein Knowledgebase in 2023.** *Nucleic Acids Res.* 2023; **51**(D1): D523–D531.
PubMed Abstract | Publisher Full Text | Free Full Text

Bradley JD, Tremewan WG, Smith A: *British Tortricoid Moths.* 1979; **2**.

Buchfink B, Reuter K, Drost H-G: **Sensitive protein alignments at tree-of-life scale using DIAMOND.** *Nat. Methods.* 2021; **18**(4): 366–368.
PubMed Abstract | Publisher Full Text | Free Full Text

Challis R, Richards E, Rajan J, *et al.*: **BlobToolKit – interactive quality assessment of genome assemblies.** *G3: Genes, Genomes, Genetics.* 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

Crowley L, Allen H, Barnes I, *et al.*: **A sampling strategy for genome sequencing the British terrestrial Arthropod fauna.** *Wellcome Open Research.* 2023; **8**: 123.
PubMed Abstract | Publisher Full Text | Free Full Text

Danecek P, Bonfield JK, Liddle J, *et al.*: **Twelve years of SAMtools and BCFtools.** *GigaScience.* 2021; **10**(2): giab008.
PubMed Abstract | Publisher Full Text | Free Full Text

Deland J-P, Vanoosthuyse F, Cormier D: **Évaluation de l'efficacité des insecticides biologiques azadirachtine et *B. thuringiensis* var. Kurstaki pour lutter contre la tordeuse des canneberges dans la production de canneberges.** 2014.
Reference Source

Ewels P, Magnusson M, Lundin S, *et al.*: **MultiQC: Summarize analysis results for multiple tools and samples in a single report.** *Bioinformatics.* 2016; **32**(19): 3047–3048.
PubMed Abstract | Publisher Full Text | Free Full Text

Ewels PA, Peltzer A, Fillinger S, *et al.*: **The nf-core framework for community-curated bioinformatics pipelines.** *Nat. Biotechnol.* 2020; **38**(3): 276–278.
PubMed Abstract | PubMed Abstract | Publisher Full Text

Fitzpatrick SM: **Delayed mating reduces fecundity of blackheaded fireworm, *Rhopobota naevana*, on cranberry.** *Entomol. Exp. Appl.* 2006; **120**: 245–250.
Publisher Full Text

Formenti G, Abueg L, Brajuka A, *et al.*: **Gfastats: Conversion, evaluation and manipulation of genome sequences using assembly graphs.** *Bioinformatics.* 2022; **38**(17): 4214–4216.
PubMed Abstract | Publisher Full Text | Free Full Text

GBIF Secretariat: ***Rhopobota naevana* (Hübner, 1817).** 2025.
Reference Source

Hancock EF, Bland KP, Razowski J: *The Moths and Butterflies of Great Britain and Ireland: Tortricidae, Olethreutinae.* Leiden: Brill; 2015.

Howard C, Denton A, Jackson B, *et al.*: **On the path to reference genomes for all biodiversity: Lessons learned and laboratory protocols created in the Sanger Tree of Life core laboratory over the first 2000 species.** *bioRxiv.* 2025.
Publisher Full Text

Howe K, Chow W, Collins J, *et al.*: **Significantly improving the quality of genome assemblies through curation.** *GigaScience.* 2021; **10**(1): giaa153.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

Huie LH: ***Eudemis naevana*, Hb., The Holly Tortrix Moth.** *Proceedings of the Royal Society of Edinburgh.* 1917; **20**(3): 164–178.
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**

Kurtzer GM, Sochat V, Bauer MW: **Singularity: Scientific containers for mobility of compute.** *PLOS ONE.* 2017; **12**(5): e0177459.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

Labarre D, Bernardo-Santos J, Pouët C, *et al.*: ***Rhopobota naevana* (Hübner), Blackheaded Fireworm/tordeuse des canneberges (Lepidoptera: Tortricidae).** *Biological Control Programmes in Canada, 2013–2023.* CABI; 2024; 351–58.
Publisher Full Text

Li H: **Minimap2: Pairwise alignment for nucleotide sequences.** *Bioinformatics.* 2018; **34**(18): 3094–3100.
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**

Merkel D: **Docker: Lightweight Linux containers for consistent development and deployment.** *Linux J.* 2014; **2014**(239): 2.
Publisher Full Text

Ranallo-Benavidez TR, Jaron KS, Schatz MC: **GenomeScope 2.0 and Smudgeplot for reference-free profiling of polyploid genomes.** *Nat. Commun.* 2020; **11**(1): 1432.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

Rao SSP, Huntley MH, Durand NC, *et al.*: **A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping.** *Cell.* 2014; **159**(7): 1665–1680.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

Rhie A, McCarthy SA, Fedrigo O, *et al.*: **Towards complete and error-free genome assemblies of all vertebrate species.** *Nature.* 2021; **592**(7856): 737–746.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

Rhie A, Walenz BP, Koren S, *et al.*: **Merqury: Reference-free quality, completeness, and phasing assessment for genome assemblies.** *Genome Biol.* 2020; **21**(1): 245.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

Schoch CL, Ciufo S, Domrachev M, *et al.*: **NCBI taxonomy: A comprehensive update on curation, resources and tools.** *Database.* 2020; **2020**: baaa062.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

Sterling P, Parsons M, Lewington R: *Field Guide to the Micro-Moths of Great Britain and Ireland.* Dorset: British Wildlife Publishing; 2023.

Twyford AD, Beasley J, Barnes I, *et al.*: **A DNA barcoding framework for taxonomic verification in the Darwin Tree of Life Project.** *Wellcome Open Research.* 2024; **9**: 339.
PubMed Abstract | **Publisher Full Text** | **Free 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.** *2019 IEEE International Parallel and Distributed Processing Symposium (IPDPS).* IEEE; 2019; 314–24.
Publisher Full Text

Wright CJ, Stevens L, Mackintosh A, *et al.*: **Comparative genomics reveals the dynamics of chromosome evolution in Lepidoptera.** *Nature Ecology & Evolution.* 2024; **8**(4): 777–790.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

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 18 August 2026

https://doi.org/10.21956/wellcomeopenres.29601.r162916

✔ **Nicola J Nadeau**
The University of Sheffield, Sheffield, UK

This note reports the reference genome generated for the moth *Rhopobota naevana.* Moths are ecologically important as prey for many larger species, pollinators as adults and pest herbivores as larvae. The note gives useful background information on the ecology of this species.

The data note gives a detailed description of the methods, and a clear report of the data and assembly generated. The work is thorough and conducted to the recognized standards in the field. The assemblies and raw data are mostly available in an accessible form on ENA. However, data for the alternate haplotype (component project PRJEB91880) seems to be missing. The project also seems to include an assembly of the Wolbachia endosymbiont. While a full description of this assembly is presumably going to be published separately, it might be worth including a mention of this in the current note.

I have a couple of other minor suggestions:

"The MitoHiFi reference was the mitochondrial genome of *Epinotia rubiginosana"* - It might be useful to elaborate on the taxonomic relationship of *Epinotia rubiginosana* to *Rhopobota naevana* or say something about how similar their Mito genomes are.

" Z chromosomes were identified based on BUSCO gene painting with ancestral Merian elements" - It would be useful to explain what Merian elements are for those who are not familiar. I also wonder if this chromosome should really be annotated as the Z without any other supporting information (e.g. coverage in females). It would be more conservative to simply give it a number at this stage.

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 genetics/genomics, lepidoptera

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 28 July 2026

https://doi.org/10.21956/wellcomeopenres.29601.r160588

Sivasankaran Kuppusamy
Loyola College, Chennai, tamilnadu, India

The authors have successfully generated a chromosome-level genome assembly for *Rhopobota naevana* (Hubner, 1817), with a total genome size of 581. 80 megabases. They have employed appropriate methodologies for High Molecular Weight DNA extraction, DNA quantification and SMRT library preparation and sequencing using Illumina and PacBio platforms. The genome assembly and annotations were carried out using well established bioinformatic software and pipelines. Overall, the genome analysis is technically sound, and the methodologies adopted are appropriate for producing a reliable chromosome-level genome assembly.
Comments on the manuscript
Authors have generated 132.28 Gb of high-quality sequencing data through Illumina and PacBio pipelines. However, authors have not reported the protein-coding genes non-coding genes and gene transcripts. The details can be included in the genome assembly table.
The Longest scaffold (Mb), GC content (%) and BUCSCO completeness (%) are the important assembly metrics and can be included in the genome assembly table
Overall, the research article was well prepared and the manuscript meets the necessary scientific standard and is suitable for indexing

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: Phylogenetic analysis of superfamily Noctuoidea moths using mitogenome analysis

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.