DATA NOTE

The genome sequence of the Beautiful Knot-horn moth, *Rhodophaea formosa* Haworth, 1811

[version 1; peer review: 1 approved]

Denise C. Wawman[1],
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]Edward Grey Institute, Department of Biology, University of Oxford, Oxford, England, UK

v1 **First published:** 28 Jun 2024, **9**:344
https://doi.org/10.12688/wellcomeopenres.21492.1
Latest published: 28 Jun 2024, **9**:344
https://doi.org/10.12688/wellcomeopenres.21492.1

Abstract

We present a genome assembly from a male *Rhodophaea formosa* (Beautiful Knot-horn; Arthropoda; Insecta; Lepidoptera; Pyralidae). The genome sequence is 616.3 megabases in span. Most of the assembly is scaffolded into 31 chromosomal pseudomolecules, including the Z sex chromosome. The mitochondrial genome has also been assembled and is 15.44 kilobases in length. Gene annotation of this assembly on Ensembl identified 18,577 protein coding genes.

Keywords

Rhodophaea formosa, Beautiful Knot-horn moth, genome sequence, chromosomal, Lepidoptera

This article is included in the Tree of Life gateway.

Open Peer Review

Approval Status ✓

	1
version 1 28 Jun 2024	✓ view

1. **Savarimuthu Ignacimuthu**, St. Xavier's College, Interdisciplinary Research Centre in Biology, Xavier Research Foundation, Manonmaniam Sundaranar University, Thirunelveli, India
Sivasankaran Kuppusamy, Loyola College, Chennai, India

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: Wawman DC: Investigation, Resources, 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.

How to cite this article: Wawman DC, University of Oxford and Wytham Woods Genome Acquisition Lab, Darwin Tree of Life Barcoding collective *et al.* **The genome sequence of the Beautiful Knot-horn moth, *Rhodophaea formosa* Haworth, 1811 [version 1; peer review: 1 approved]** Wellcome Open Research 2024, **9**:344 https://doi.org/10.12688/wellcomeopenres.21492.1

First published: 28 Jun 2024, **9**:344 https://doi.org/10.12688/wellcomeopenres.21492.1

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; Pyraloidea; Pyralidae; Phycitinae; *Rhodophaea*; *Rhodophaea formosa* Haworth, 1811 (NCBI:txid1870710).

Background

Rhodophaea formosa, the Beautiful Knot-horn, is a micro-moth in the family Pyralidae, its forewing measuring 9–11mm. It has a brownish or pinkish red forewing, with a wide blackish crossband and two whitish cross-lines, one about half way along the wing – just behind the black crossband – and another faint line three-quarters of the way along the wing (Sterling *et al.*, 2023).

Rhodophaea formosa is on the wing around hedgerows from late-May to September where the larva feed on elm *Ulmus procera* from July to October (Sterling *et al.*, 2023). The larvae are the same dark green colour as elm, with fine dotted white lines running from the head to the tip of the abdomen (Emmet, 1981). *R. formosa* is expanding its range northwards and westwards in the United Kingdom: it has recently been recorded from Northumberland and Yorkshire and was first found in Wales in 2006 (Sterling *et al.*, 2023).

For such a small moth, this species has accumulated a wide range of long names over the years. It has previously had the following binominals: *Nephopteryx formosa* (Haworth, 1811), *Oncocera formosa* (Haworth), *Pempelia formosa* (Haworth, 1811), *Phycis formosa* (Haworth, 1811), *Rhodophaea dibaphiella* Hübner, *Rhodophaea dubiella* Duponchel, 1836, *Rhodophaea perfluella* Zincken, 1818 and *Salebria formosa* (Haworth, 1811), with *R. formosa* and *S. formosa* having been used most frequently (GBIF Secretariat, 2024). Many of its names in European languages make reference to *R. formosa* being found on elm trees. Google Translate gives a literal translation of its Danish name, Elmehalvmøl, as "Elm Half Moth"; the Swedish, Almmolnmott, is "Elm Cloud Moth" and the German Feldulmen-Schmalzünsler is "Field Elm Borer", but the Dutch, Veelkleurige lichtmot, translates into "Multicoloured Light Moth".

We present a chromosomally complete genome sequence for *Rhodophaea formosa*, based on one specimen collected using a mercury vapour light trap in a rural garden in the hamlet of Bratton, near Minehead, in Somerset, as part of the Darwin Tree of Life Project.

Genome sequence report

The genome was sequenced from a male *Rhodophaea formosa* (Figure 1) collected from Bratton, Somerset, UK (51.20, –3.51). A total of 40-fold coverage in Pacific Biosciences single-molecule HiFi long reads was generated. Primary assembly contigs were scaffolded with chromosome conformation

Figure 1. Photograph of the *Rhodophaea formosa* (ilRhoForm1) specimen used for genome sequencing.

Hi-C data. Manual assembly curation corrected 14 missing joins or mis-joins and removed 2 haplotypic duplications, reducing the scaffold number by 11.90%, and increasing the scaffold N50 by 0.36%.

The final assembly has a total length of 616.3 Mb in 36 sequence scaffolds with a scaffold N50 of 22.3 Mb (Table 1). The snail plot in Figure 2 provides a summary of the assembly statistics, while the distribution of assembly scaffolds on GC proportion and coverage is shown in Figure 3. The cumulative assembly plot in Figure 4 shows curves for subsets of scaffolds assigned to different phyla. Most (99.96%) of the assembly sequence was assigned to 31 chromosomal-level scaffolds, representing 30 autosomes and the Z sex chromosome. Chromosome-scale scaffolds confirmed by the Hi-C data are named in order of size (Figure 5; Table 2). The Z chromosome was identified based on synteny with *Elegia similella* (GCA_947532085.1). 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 67.0 with *k*-mer completeness of 100.0%, and the assembly has a BUSCO v5.3.2 completeness of 99.0% (single = 98.5%, duplicated = 0.4%), 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/1870710.

Genome annotation report

The *Rhodophaea formosa* genome assembly (GCA_963082605.1) was annotated at the European Bioinformatics Institute (EBI) on Ensembl Rapid Release. The resulting annotation includes 18,819 transcribed mRNAs from 18,577 protein-coding genes (Table 1; https://rapid.ensembl.org/Rhodophaea_formosa_GCA_963082605.1/Info/Index).

Table 1. Genome data for *Rhodophaea formosa*, ilRhoForm1.1.

Project accession data		
Assembly identifier	ilRhoForm1.1	
Species	*Rhodophaea formosa*	
Specimen	ilRhoForm1	
NCBI taxonomy ID	1870710	
BioProject	PRJEB63434	
BioSample ID	SAMEA112226469	
Isolate information	ilRhoForm1, male: whole organism (DNA and HiC sequencing)	
Assembly metrics*		***Benchmark***
Consensus quality (QV)	67.0	*≥ 50*
k-mer completeness	100.0%	*≥ 95%*
BUSCO**	C:99.0%[S:98.5%,D:0.4%], F:0.3%,M:0.8%,n:5,286	*C ≥ 95%*
Percentage of assembly mapped to chromosomes	99.96%	*≥ 95%*
Sex chromosomes	ZZ	*localised homologous pairs*
Organelles	Mitochondrial genome: 15.44 kb	*complete single alleles*
Raw data accessions		
PacificBiosciences SEQUEL II	ERR11593801	
Hi-C Illumina	ERR11606317	
Genome assembly		
Assembly accession	GCA_963082605.1	
Accession of alternate haplotype	GCA_963082635.1	
Span (Mb)	616.3	
Number of contigs	95	
Contig N50 length (Mb)	12.4	
Number of scaffolds	36	
Scaffold N50 length (Mb)	22.3	
Longest scaffold (Mb)	29.66	
Genome annotation		
Number of protein-coding genes	18,577	
Number of gene transcripts	18,819	

* 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/CAUJBE01.1/dataset/CAUJBE01.1/busco.

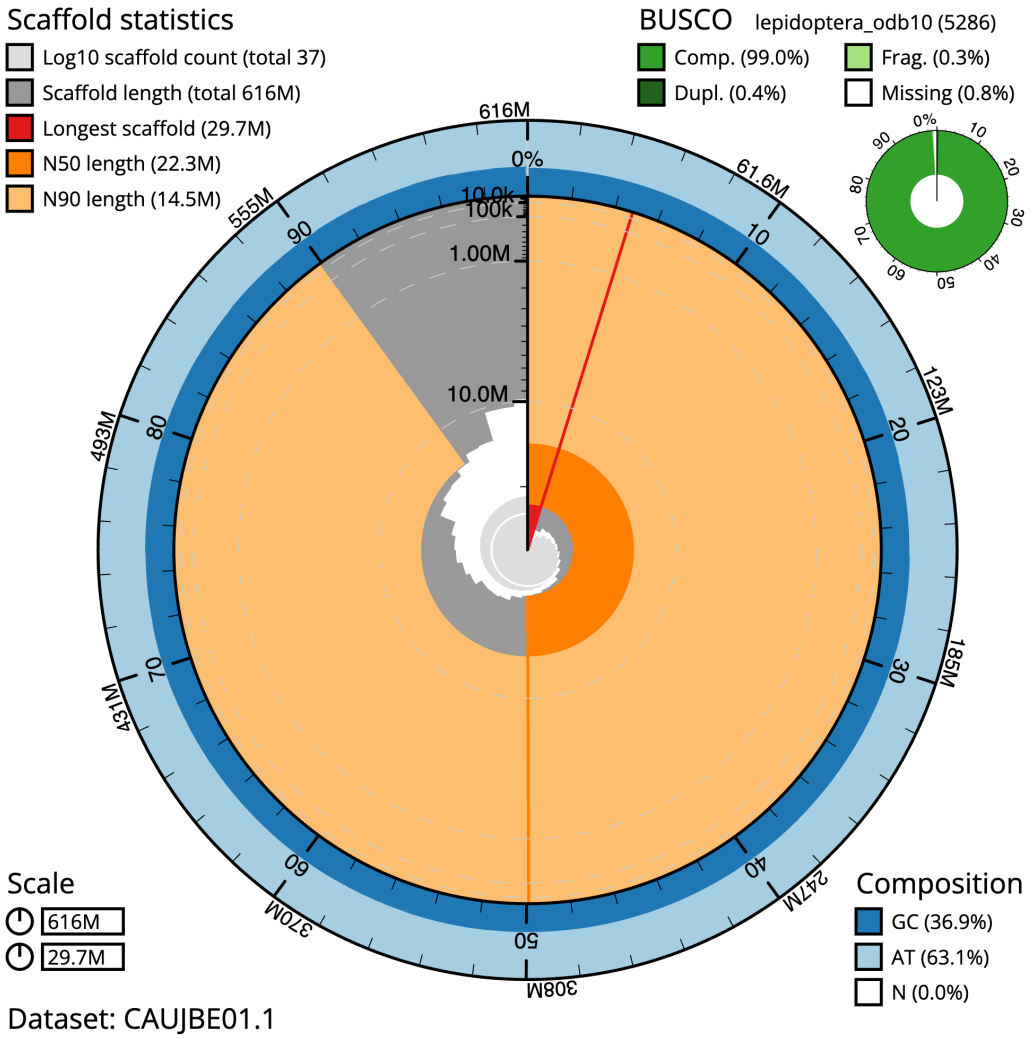

Figure 2. Genome assembly of *Rhodophaea formosa*, ilRhoForm1.1: metrics. The BlobToolKit snail plot shows N50 metrics and BUSCO gene completeness. The main plot is divided into 1,000 size-ordered bins around the circumference with each bin representing 0.1% of the 616,286,378 bp assembly. The distribution of scaffold lengths is shown in dark grey with the plot radius scaled to the longest scaffold present in the assembly (29,656,602 bp, shown in red). Orange and pale-orange arcs show the N50 and N90 scaffold lengths (22,288,753 and 14,491,228 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/CAUJBE01.1/dataset/CAUJBE01.1/snail.

Methods

Sample acquisition and nucleic acid extraction

A male *Rhodophaea formosa* (specimen ID Ox002243, ToLID ilRhoForm1) was collected from Bratton, Somerset, UK (latitude 51.20, longitude –3.51) on 2022-06-20 using a light trap. The specimen was collected and identified by Denise Wawman (University of Oxford) and preserved on dry ice.

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. The sample was prepared for extraction at the WSI Tree of Life Core Laboratory: the ilRhoForm1 sample was weighed and dissected on dry ice (Jay *et al.*, 2023). Tissue of the whole organism was homogenised using a PowerMasher II tissue disruptor (Denton *et al.*, 2023a).

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 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,

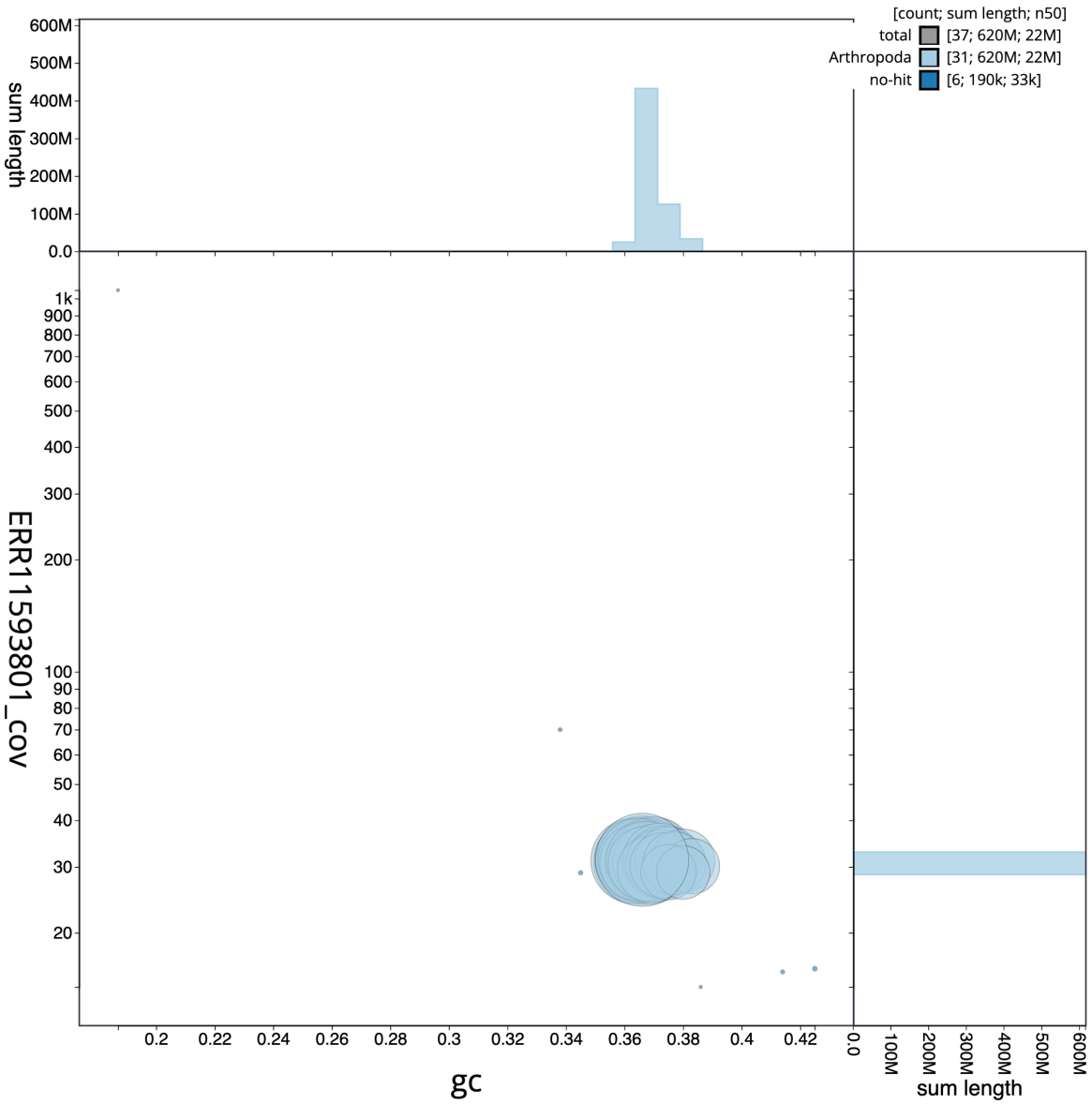

Figure 3. Genome assembly of *Rhodophaea formosa*, ilRhoForm1.1: BlobToolKit GC-coverage plot. Sequences are coloured by phylum. Circles are sized in proportion to sequence length. Histograms show the distribution of sequence length sum along each axis. An interactive version of this figure is available at https://blobtoolkit.genomehubs.org/view/CAUJBE01.1/dataset/CAUJBE01.1/blob.

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.

Protocols developed by the WSI Tree of Life laboratory are publicly available on protocols.io (Denton *et al.*, 2023b).

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 a Pacific Biosciences SEQUEL II instrument. Hi-C data were also generated from remaining tissue of ilRhoForm1 using the Arima2 kit and sequenced on the Illumina NovaSeq 6000 instrument.

Genome assembly, curation and evaluation

Assembly was carried out with Hifiasm (Cheng *et al.*, 2021) and haplotypic duplication was identified and removed with purge_dups (Guan *et al.*, 2020). 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)

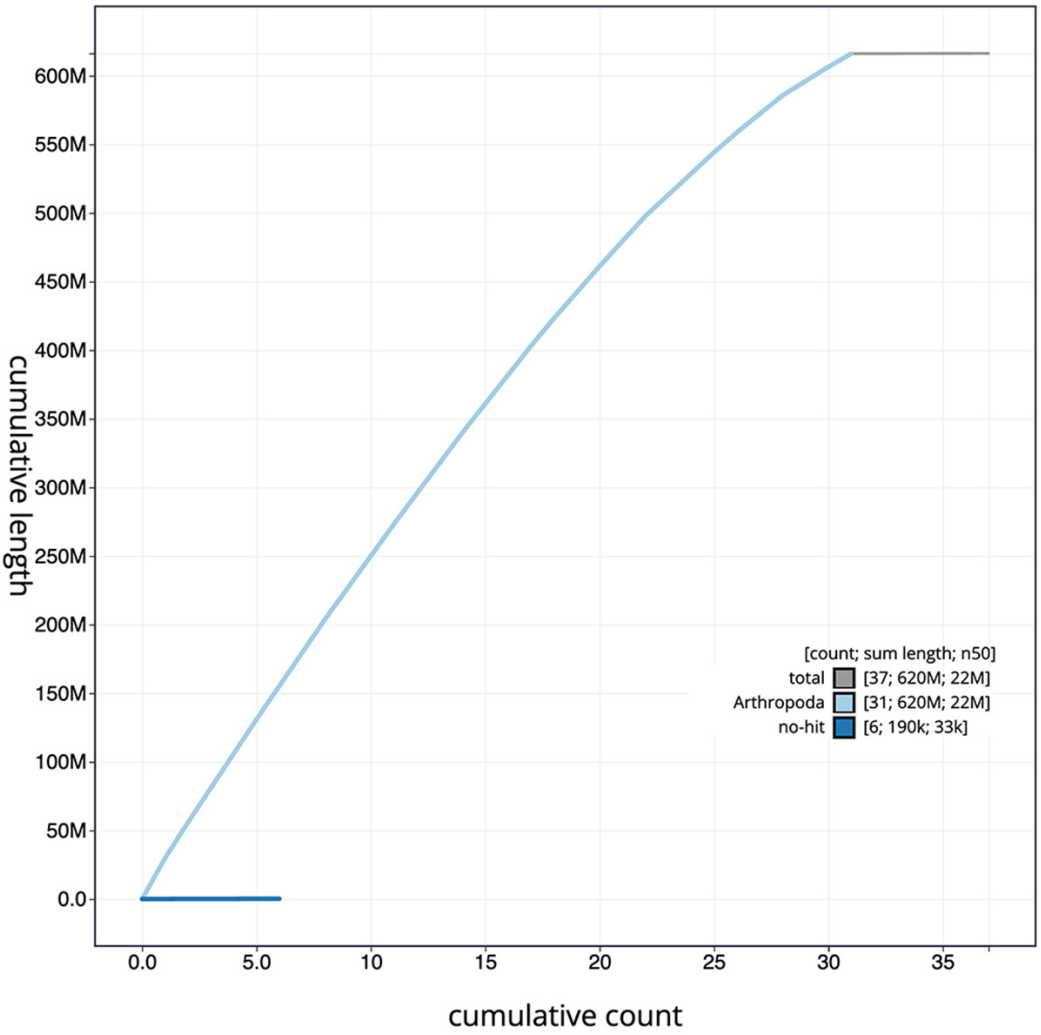

Figure 4. Genome assembly of *Rhodophaea formosa*, ilRhoForm1.1: BlobToolKit cumulative sequence plot. The grey line shows cumulative length for all sequences. Coloured lines show cumulative lengths of sequences assigned to each phylum using the buscogenes taxrule. An interactive version of this figure is available at https://blobtoolkit.genomehubs.org/view/CAUJBE01.1/dataset/CAUJBE01.1/cumulative.

and PretextView (Harry, 2022). The mitochondrial genome was assembled using MitoHiFi (Uliano-Silva *et al.*, 2023), which runs MitoFinder (Allio *et al.*, 2020) or MITOS (Bernt *et al.*, 2013) and uses these annotations to select the final mitochondrial contig and to ensure the general quality of the sequence.

A Hi-C map for the final assembly was produced using bwa-mem2 (Vasimuddin *et al.*, 2019) in the Cooler file format (Abdennur & Mirny, 2020). To assess the assembly metrics, the *k*-mer completeness and QV consensus quality values were calculated in Merqury (Rhie *et al.*, 2020). This work was done using Nextflow (Di Tommaso *et al.*, 2017) DSL2 pipelines "sanger-tol/readmapping" (Surana *et al.*, 2023a) and "sanger-tol/genomenote" (Surana *et al.*, 2023b). The genome was analysed within the BlobToolKit environment (Challis *et al.*, 2020) and BUSCO scores (Manni *et al.*, 2021; Simão *et al.*, 2015) were calculated.

Table 3 contains a list of relevant software tool versions and sources.

Genome annotation

The BRAKER2 pipeline (Brůna *et al.*, 2021) was used in the default protein mode to generate annotation for the *Rhodophaea formosa* assembly (GCA_963082605.1) in Ensembl Rapid Release at the EBI.

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

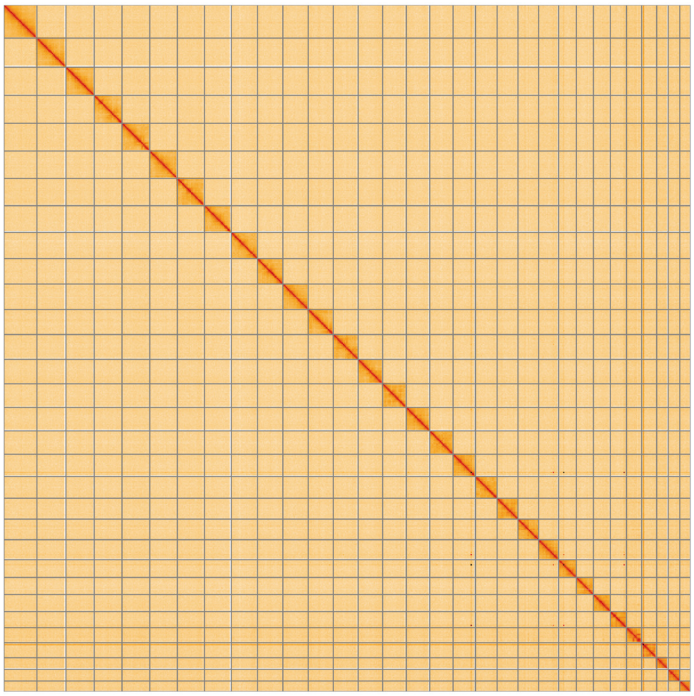

Figure 5. Genome assembly of *Rhodophaea formosa*, ilRhoForm1.1: Hi-C contact map of the ilRhoForm1.1 assembly, visualised using HiGlass. Chromosomes are shown in order of size from left to right and top to bottom. An interactive version of this figure may be viewed at https://genome-note-higlass.tol.sanger.ac.uk/l/?d=ZBU9dlRZRc2pp9Y2fEq_lw.

Table 2. Chromosomal pseudomolecules in the genome assembly of *Rhodophaea formosa*, ilRhoForm1.

INSDC accession	Chromosome	Length (Mb)	GC%
OY720204.1	1	26.04	36.5
OY720205.1	2	25.3	37.0
OY720206.1	3	25.06	36.5
OY720207.1	4	24.87	37.0
OY720208.1	5	24.7	36.5
OY720209.1	6	24.37	37.0
OY720210.1	7	23.98	36.5
OY720211.1	8	23.46	36.5
OY720212.1	9	22.87	36.5
OY720213.1	10	22.78	37.0
OY720214.1	11	22.5	37.0
OY720215.1	12	22.29	36.5
OY720216.1	13	21.97	37.0
OY720217.1	14	21.23	36.5
OY720218.1	15	21.01	37.0
OY720219.1	16	20.98	36.5
OY720220.1	17	20.01	37.0
OY720221.1	18	19.44	37.0
OY720222.1	19	18.72	36.5
OY720223.1	20	18.51	37.0
OY720224.1	21	17.96	37.5
OY720225.1	22	15.9	37.0
OY720226.1	23	15.35	37.0
OY720227.1	24	15.22	37.5
OY720228.1	25	14.49	37.5
OY720229.1	26	13.59	38.0
OY720230.1	27	13.45	37.5
OY720231.1	28	10.37	38.5
OY720232.1	29	10.36	37.5
OY720233.1	30	9.66	38.0
OY720203.1	Z	29.66	36.5
OY720234.1	MT	0.02	19.0

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	0.19.5-r587	https://github.com/chhylp123/hifiasm
HiGlass	1.11.6	https://github.com/higlass/higlass
Merqury	MerquryFK	https://github.com/thegenemyers/MERQURY.FK
MitoHiFi	3	https://github.com/marcelauliano/MitoHiFi
PretextView	0.2	https://github.com/wtsi-hpag/PretextView
purge_dups	1.2.5	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	1.2a.2	https://github.com/c-zhou/yahs

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: *Rhodophaea formosa*. Accession number PRJEB63434; https://identifiers.org/ena.embl/PRJEB63434 (Wellcome Sanger Institute, 2023). The genome sequence is released openly for reuse. The *Rhodophaea formosa* genome sequencing initiative is part of the Darwin Tree of Life (DToL) project. All raw sequence data and the assembly have been deposited in INSDC databases. Raw data and assembly accession identifiers are reported in Table 1.

Author information

Members of the University of Oxford and Wytham Woods Genome Acquisition Lab are listed here: https://doi.org/10.5281/zenodo.7125292.

Members of the Darwin Tree of Life Barcoding collective are listed here: https://doi.org/10.5281/zenodo.4893703.

Members of the Wellcome Sanger Institute Tree of Life Management, Samples and Laboratory team are listed here: https://doi.org/10.5281/zenodo.10066175.

Members of Wellcome Sanger Institute Scientific Operations: Sequencing Operations are listed here: https://doi.org/10.5281/zenodo.10043364.

Members of the Wellcome Sanger Institute Tree of Life Core Informatics team are listed here: https://doi.org/10.5281/zenodo.10066637.

Members of the Tree of Life Core Informatics collective are listed here: https://doi.org/10.5281/zenodo.5013541.

Members of the Darwin Tree of Life Consortium are listed here: https://doi.org/10.5281/zenodo.4783558.

References

Abdennur N, Mirny LA: **Cooler: scalable storage for Hi-C data and other genomically labeled arrays.** *Bioinformatics.* 2020; **36**(1): 311–316.
PubMed Abstract | Publisher Full Text | Free Full Text

Allio R, Schomaker-Bastos A, Romiguier J, *et al.*: **MitoFinder: efficient automated large-scale extraction of mitogenomic data in target enrichment phylogenomics.** *Mol Ecol Resour.* 2020; **20**(4): 892–905.
PubMed Abstract | Publisher Full Text | Free Full Text

Bates A, Clayton-Lucey I, Howard C: **Sanger Tree of Life HMW DNA fragmentation: diagenode Megaruptor®3 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

Brůna T, Hoff KJ, Lomsadze A, *et al.*: **BRAKER2: automatic eukaryotic genome annotation with GeneMark-EP+ and AUGUSTUS supported by a protein database.** *NAR Genom Bioinform.* 2021; **3**(1): lqaa108.
PubMed Abstract | Publisher Full Text | Free Full Text

Challis R, Richards E, Rajan J, *et al.*: **BlobToolKit - interactive quality assessment of genome assemblies.** *G3 (Bethesda).* 2020; **10**(4): 1361–1374.
PubMed Abstract | Publisher Full Text | Free Full Text

Cheng H, Concepcion GT, Feng X, *et al.*: **Haplotype-resolved *de novo* assembly using phased assembly graphs with hifiasm.** *Nat Methods.* 2021; **18**(2): 170–175.
PubMed Abstract | Publisher Full Text | Free Full Text

Denton A, Oatley G, Cornwell C, *et al.*: **Sanger Tree of Life sample homogenisation: PowerMash.** *protocols.io.* 2023a.
Publisher Full Text

Denton A, Yatsenko H, Jay J, *et al.*: **Sanger Tree of Life wet laboratory protocol collection V.1.** *protocols.io.* 2023b.
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

Emmet AM: **The larva of *Oncocera formosa* (Haworth) (Lep .: Pyralidae).** *Entomol's Rec J Var.* 1981; **93**: 191–191.
Reference Source

GBIF Secretariat: ***Rhodophaea formosa* Haworth, 1811.** *GBIF Backbone Taxonomy.* 2024; [Accessed 18 March 2024].
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.* 2021; **10**(1): giaa153.
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, 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

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

Sterling P, Parsons M, Lewington R: **Field Guide to the Micro-moths of Great Britain and Ireland, Second Edition.** London: Bloomsbury Publishing, 2023.
Reference Source

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.
Publisher Full Text

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

Uliano-Silva M, Ferreira JGRN, Krasheninnikova K, *et al.*: **MitoHiFi: a python pipeline for mitochondrial genome assembly from PacBio high fidelity reads.** *BMC Bioinformatics.* 2023; **24**(1): 288.
PubMed Abstract | Publisher Full Text | Free Full Text

Vasimuddin M, Misra S, Li H, *et al.*: **Efficient architecture-aware acceleration of BWA-MEM for multicore systems**. In: *2019 IEEE International Parallel and Distributed Processing Symposium (IPDPS).* IEEE, 2019; 314–324.
Publisher Full Text

Wellcome Sanger Institute: **The genome sequence of the Beautiful Knot-horn moth, *Rhodophaea formosa* Haworth, 1811.** European Nucleotide Archive, [dataset], accession number PRJEB63434, 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 29 August 2024

https://doi.org/10.21956/wellcomeopenres.23762.r93139

✔ **Savarimuthu Ignacimuthu**
St. Xavier's College, Interdisciplinary Research Centre in Biology, Xavier Research Foundation, Manonmaniam Sundaranar University, Thirunelveli, TN, India
Sivasankaran Kuppusamy
Entomology Research Institute, Loyola College, Chennai, Tamil Nadu, India

Authors have sequenced the genome of *Rhodophaea formosa* Haworth, 1811. Totally 616.3 megabases size of genome was found through assembly. Authors have identified 18, 577 protein coding genes and 18,819 gene transcripts through the genome annotations. They have used proper procedure for nucleic acid extraction and appropriate software for assembly and annotations.

Comments on the manuscript

- The first sentence of the Background can be started as "*Rhodophaea formosa* Haworth, 1811..."
- The authors have given the synonyms of the species in the third paragraph of the Background. The synonyms details aren't necessary here.
- The following sentence can be deleted. The details aren't necessary here. "Google Translate gives a literal translation of its Danish name, Elmehalvmøl, as "Elm Half Moth"; the Swedish, Almmolnmott, is "Elm Cloud Moth" and the German Feldulmen-Schmalzünsler is "Field Elm Borer", but the Dutch, Veelkleurigelichtmot, translates into "Multicoloured Light Moth".
- The authors have given the complete form of the genus name *Rhodophaea formosa* throughout the article. The genus name may be given in full for the first time, and then in abbreviated form later on, such as *R. formosa.*
- The mitochondrial genome length wasn't mentioned in the text. Total length of the mitochondrial genome can be given in the text.

Above all, I confirm that 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: Molecular biology

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