



DATA NOTE

# The genome sequence of the Sandy Carpet moth, *Perizoma flavofasciatum* (Thunberg, 1792)

[version 1; peer review: awaiting peer review]

Douglas Boyes<sup>1+</sup>, Jonathan Davis<sup>2</sup>,  
University of Oxford and Wytham Woods Genome Acquisition Lab,  
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

<sup>1</sup>UK Centre for Ecology & Hydrology, Wallingford, England, UK

<sup>2</sup>Independent researcher, Lanark, Scotland, UK

+ Deceased author

**V1** First published: 03 Feb 2025, 10:40  
<https://doi.org/10.12688/wellcomeopenres.23612.1>

Latest published: 03 Feb 2025, 10:40  
<https://doi.org/10.12688/wellcomeopenres.23612.1>

## Abstract

We present a genome assembly from a male specimen of *Perizoma flavofasciatum* (Sandy Carpet; Arthropoda; Insecta; Lepidoptera; Geometridae). The genome sequence has a total length of 369.30 megabases. Most of the assembly (99.88%) is scaffolded into 30 chromosomal pseudomolecules, including the Z sex chromosome. The mitochondrial genome has also been assembled and is 16.61 kilobases in length. Gene annotation of this assembly on Ensembl identified 11,915 protein-coding genes.

## Keywords

*Perizoma flavofasciatum*, Sandy Carpet moth, 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](mailto:mark.blaxter@sanger.ac.uk))

**Author roles:** **Boyes D:** Investigation, Resources; **Davis J:** Investigation, Resources;

**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:** © 2025 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, Davis J, University of Oxford and Wytham Woods Genome Acquisition Lab *et al.* **The genome sequence of the Sandy Carpet moth, *Perizoma flavofasciatum* (Thunberg, 1792) [version 1; peer review: awaiting peer review]** Wellcome Open Research 2025, 10:40 <https://doi.org/10.12688/wellcomeopenres.23612.1>

**First published:** 03 Feb 2025, 10:40 <https://doi.org/10.12688/wellcomeopenres.23612.1>

## Species taxonomy

Eukaryota; Opisthokonta; Metazoa; Eumetazoa; Bilateria; Protostomia; Ecdysozoa; Panarthropoda; Arthropoda; Mandibulata; Pancrustacea; Hexapoda; Insecta; Dicondylia; Pterygota; Neoptera; Endopterygota; Amphimesenoptera; Lepidoptera; Glossata; Neolepidoptera; Heteroneura; Ditrysia; Obtectomera; Geometroidea; Geometridae; Larentiinae; *Perizoma*; *Perizoma flavofasciatum* (Thunberg, 1792) (NCBI:txid934819)

## Background

*Perizoma flavofasciatum* (family Geometridae), commonly known as the Sandy Carpet moth, is widely distributed across Europe and extends into parts of Asia. Its range includes most European countries and stretches eastwards through the Palearctic region to the Urals and the Altai Mountains (GBIF Secretariat, 2023). This species inhabits various environments across its range, such as meadow valleys, floodplains, waterside areas, bushy meadows, and gardens. In mountainous regions like the Alps, it can be found at elevations up to 1,500 metres.

*Perizoma flavofasciatum* has a wingspan of 26–32 mm (Kimber, 2025). The species is characterised by the sandy brown cross-lines on a white ground colour and the two interneural blotches connecting the median and subterminal fasciae at about halfway between costa and dorsum (British Lepidoptera, 2025).

In the UK, it is fairly common across Britain, particularly in the south of England (GBIF Secretariat, 2023) and is listed as “least concern” in the macro-moth status review (Fox *et al.*, 2019). The adult flies from dusk onwards in June and July (Kimber, 2025), and inhabits woodland, commons, chalky ground and other dry areas. The larvae feed on the seed-pods of campions (*Silene spp.*) (Waring *et al.*, 2017).

We present a chromosomal-level genome sequence for *Perizoma flavofasciatum*, based on a male specimen from Wytham Woods, Berkshire, United Kingdom (Figure 1). This was sequenced as part of the Darwin Tree of Life project.

## Genome sequence report

The genome of *Perizoma flavofasciatum* (Figure 1) was sequenced using Pacific Biosciences single-molecule HiFi long reads, generating a total of 25.84 Gb (gigabases) from 2.00 million reads, providing an estimated 66-fold coverage. Primary assembly contigs were scaffolded with chromosome conformation Hi-C data, which produced 91.65 Gb from 606.97 million reads. Specimen and sequencing details are summarised in Table 1.

Assembly errors were corrected by manual curation, including three missing joins or mis-joins. The final assembly has a total length of 369.30 Mb in 39 sequence scaffolds, with three gaps. The scaffold N50 is 13.4 Mb (Table 2).

The snail plot in Figure 2 provides a summary of the assembly statistics, indicating the distribution of scaffold lengths and other assembly metrics. Figure 3 shows the distribution of



**Figure 1.** Photograph of the *Perizoma flavofasciatum* (ilPerFlao1) specimen used for genome sequencing.

scaffolds by GC proportion and coverage. Figure 4 presents a cumulative assembly plot, with separate curves representing different scaffold subsets assigned to various phyla, illustrating the completeness of the assembly.

Most of the assembly sequence (99.88%) was assigned to 30 chromosomal-level scaffolds, representing 29 autosomes and the Z sex chromosome. These chromosome-level scaffolds, confirmed by the Hi-C data, are named in order of size (Figure 5; Table 3). During manual curation the Z chromosome was identified based on synteny with *Eulithis prunata* (GCA\_918843925.1) (Boyes *et al.*, 2023).

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, and as a separate fasta file.

The final assembly has a Quality Value (QV) of 65.5. The *k*-mer completeness value for the primary assembly was estimated as 88.65%, for the alternate haplotype was 82.01%, and of the combined assemblies was 99.57%. BUSCO (v5.4.3) analysis using the lepidoptera\_odb10 reference set ( $n = 5,286$ ) indicated a completeness score of 98.2% (single = 97.8%, duplicated = 0.4%). The assembly achieves the EBP reference standard of 6.7.65. Other quality metrics are given in Table 2.

## Genome annotation report

The *Perizoma flavofasciatum* genome assembly (GCA\_958496245.1) was annotated at the European Bioinformatics Institute (EBI) on Ensembl Rapid Release. The resulting annotation includes 22,268 transcribed mRNAs from 11,915 protein-coding and 1,605 non-coding genes (Table 2; [https://rapid.ensembl.org/Perizoma\\_flavofasciatum\\_GCA\\_958496245.1/Info/Index](https://rapid.ensembl.org/Perizoma_flavofasciatum_GCA_958496245.1/Info/Index)). The average transcript length is 14,574.44. There

**Table 1. Specimen and sequencing data for *Perizoma flavofasciatum*.**

Project information			
Study title	Perizoma flavofasciatum (sandy carpet)		
Umbrella BioProject	PRJEB63430		
Species	<i>Perizoma flavofasciatum</i>		
BioSample	SAMEA7701445		
NCBI taxonomy ID	934819		
Specimen information			
Technology	ToLID	BioSample accession	Organism part
PacBio long read sequencing	ilPerFlao1	SAMEA7701609	Whole organism
Hi-C sequencing	ilPerFlao2	SAMEA112360821	thorax
RNA sequencing	ilPerFlao2	SAMEA112360822	abdomen
Sequencing information			
Platform	Run accession	Read count	Base count (Gb)
Hi-C Illumina NovaSeq 6000	ERR11606315	6.07e+08	91.65
PacBio Sequel IIE	ERR11593799	2.00e+06	25.84
RNA Illumina NovaSeq 6000	ERR11837494	7.68e+07	11.59

are 1.65 coding transcripts per gene and 7.73 exons per transcript.

## Methods

### Sample acquisition and DNA barcoding

An adult male specimen of *Perizoma flavofasciatum* (specimen ID Ox000581, ToLID ilPerFlao1) was collected from Wytham Woods, Berkshire, United Kingdom (latitude 51.77, longitude -1.34) on 2020-07-05, using a light trap. The specimen was collected and identified by Douglas Boyes (University of Oxford) and on dry ice.

The specimen used for Hi-C and RNA sequencing (specimen ID SAN00002560, ToLID ilPerFlao2) was collected from Little Sparta, Dunsyre, Pentland Hills, South Lanarkshire, Scotland (latitude 55.72, longitude -3.50) on 2022-06-17. The specimen was collected and identified by Jo Davis (independent researcher).

### Nucleic acid extraction

The workflow for high molecular weight (HMW) DNA extraction at the Wellcome Sanger Institute (WSI) Tree of Life Core Laboratory includes a sequence of procedures: sample preparation and homogenisation, DNA extraction, fragmentation and purification. Detailed protocols are available on protocols.io (Denton *et al.*, 2023b). The ilPerFlao1 sample was prepared for DNA extraction by weighing and dissecting it on

dry ice (Jay *et al.*, 2023). Tissue from the whole organism was homogenised using a PowerMasher II tissue disruptor (Denton *et al.*, 2023a). HMW DNA was extracted using the Automated MagAttract v1 protocol (Sheerin *et al.*, 2023). DNA was sheared into an average fragment size of 12–20 kb in a Megaruptor 3 system (Todorovic *et al.*, 2023). Sheared DNA was purified by solid-phase reversible immobilisation, using AMPure PB beads to eliminate shorter fragments and concentrate the DNA (Strickland *et al.*, 2023). The concentration of the sheared and purified DNA was assessed using a Nanodrop spectrophotometer and a Qubit Fluorometer using the Qubit dsDNA High Sensitivity Assay kit. The fragment size distribution was evaluated by running the sample on the FemtoPulse system.

RNA was extracted from abdomen tissue of ilPerFlao2 in the Tree of Life Laboratory at the WSI using the RNA Extraction: Automated MagMax™ mirVana protocol (do Amaral *et al.*, 2023). The RNA concentration was assessed using a Nanodrop spectrophotometer and a Qubit Fluorometer using the 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.

### Hi-C preparation

Thorax tissue of the ilPerFlao2 sample was processed at the WSI Scientific Operations core, using the Arima-HiC

**Table 2. Genome assembly data for *Perizoma flavofasciatum*, ilPerFlao1.1.**

Genome assembly		
Assembly name	ilPerFlao1.1	
Assembly accession	GCA_958496245.1	
Accession of alternate haplotype	GCA_958496225.1	
Span (Mb)	369.30	
Number of contigs	43	
Number of scaffolds	39	
Longest scaffold (Mb)	20.74	
Assembly metrics*		Benchmark
Contig N50 length (Mb)	13.4	≥ 1 Mb
Scaffold N50 length (Mb)	13.4	= chromosome N50
Consensus quality (QV)	65.5	≥ 40
k-mer completeness	Primary: 88.65%; alternate: 82.01%; combined: 99.57%	≥ 95%
BUSCO v5.4.3 lineage: lepidoptera_odb10	C:98.2%[S:97.8%,D:0.4%], F:0.4%,M:1.4%,n:5,286	S > 90%, D < 5%
Percentage of assembly mapped to chromosomes	99.88%	≥ 90%
Sex chromosomes	Z	localised homologous pairs
Organelles	Mitochondrial genome: 16.61 kb	complete single alleles
Genome annotation of assembly GCA_958496245.1 at Ensembl		
Number of protein-coding genes	11,915	
Number of non-coding genes	1,605	
Number of gene transcripts	22,268	

\* Assembly metric benchmarks are adapted from [Rhie et al. \(2021\)](#) and the Earth BioGenome Project Report on Assembly Standards [September 2024](#).

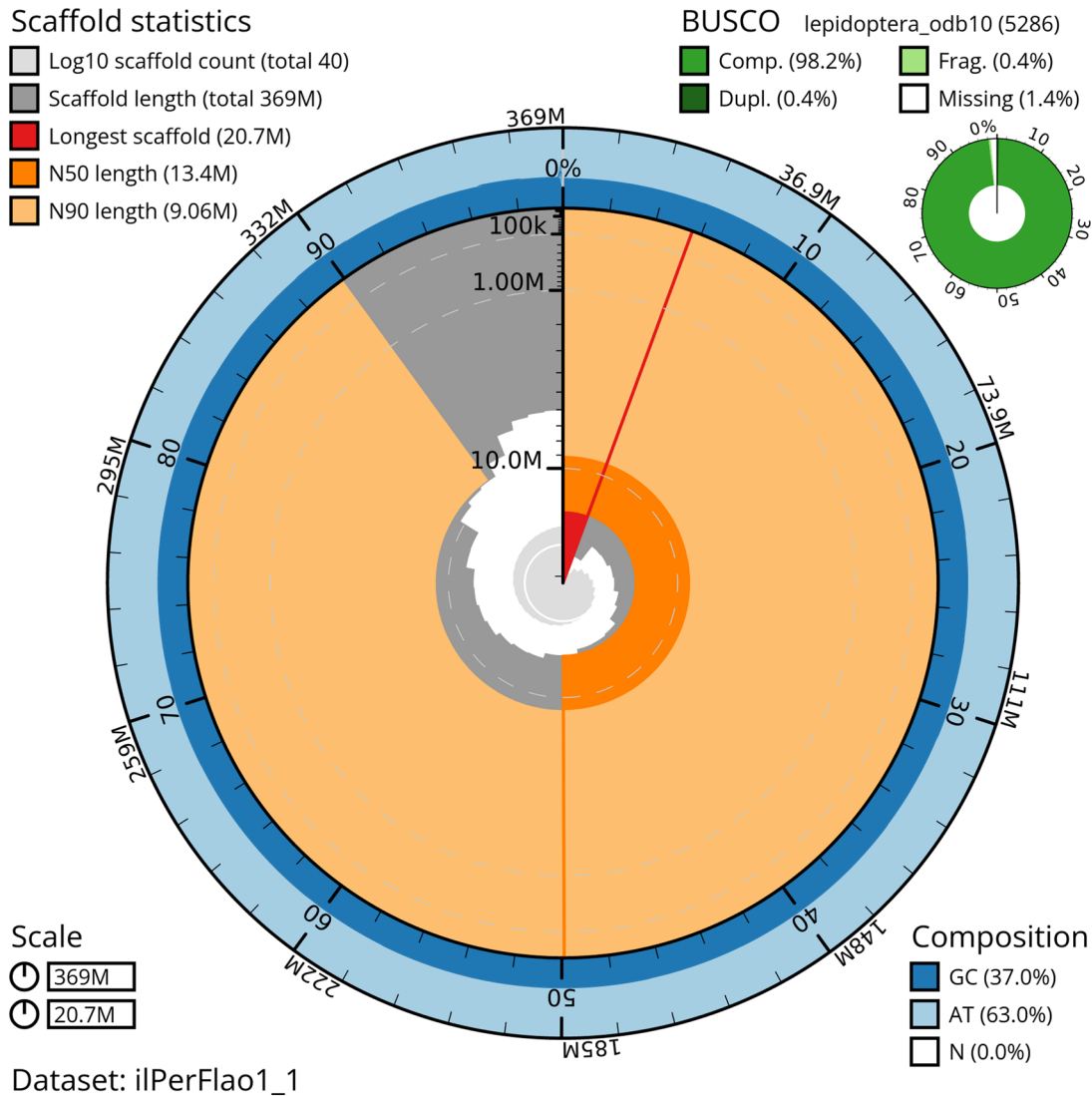
\*\* BUSCO scores based on the lepidoptera\_odb10 BUSCO set using version 5.4.3. C = complete [S = single copy, D = duplicated], F = fragmented, M = missing, n = number of orthologues in comparison.

v2 kit. Tissue (stored at  $-80^{\circ}\text{C}$ ) was fixed, and the DNA crosslinked using a TC buffer with 22% formaldehyde. After crosslinking, the tissue was homogenised using the Diagnocine Power Masher-II and BioMasher-II tubes and pestles. Following the kit manufacturer's instructions, crosslinked DNA was digested using a restriction enzyme master mix. The 5'-overhangs were then filled in and labelled with biotinylated nucleotides and proximally ligated. An overnight incubation was carried out for enzymes to digest remaining proteins and for crosslinks to reverse. A clean up was performed with SPRIselect beads prior to library preparation.

## Library preparation and sequencing

### *PacBio HiFi*

Libraries were prepared using the PacBio Express Template Preparation Kit v2.0 (Pacific Biosciences, California, USA) as per the manufacturer's instructions. The kit includes the reagents required for removal of single-strand overhangs, DNA damage repair, end repair/A-tailing, adapter ligation, and nuclease treatment. Library preparation also included a library purification step using AMPure PB beads (Pacific Biosciences, California, USA) and size selection step to remove templates shorter than 3 kb using AMPure PB modified SPRI. DNA



**Figure 2. Genome assembly of *Perizoma flavofasciatum*, ilPerFlao1.1: metrics.** 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 is available at [https://blobtoolkit.genomehubs.org/view/ilPerFlao1\\_1/dataset/ilPerFlao1\\_1/snail](https://blobtoolkit.genomehubs.org/view/ilPerFlao1_1/dataset/ilPerFlao1_1/snail).

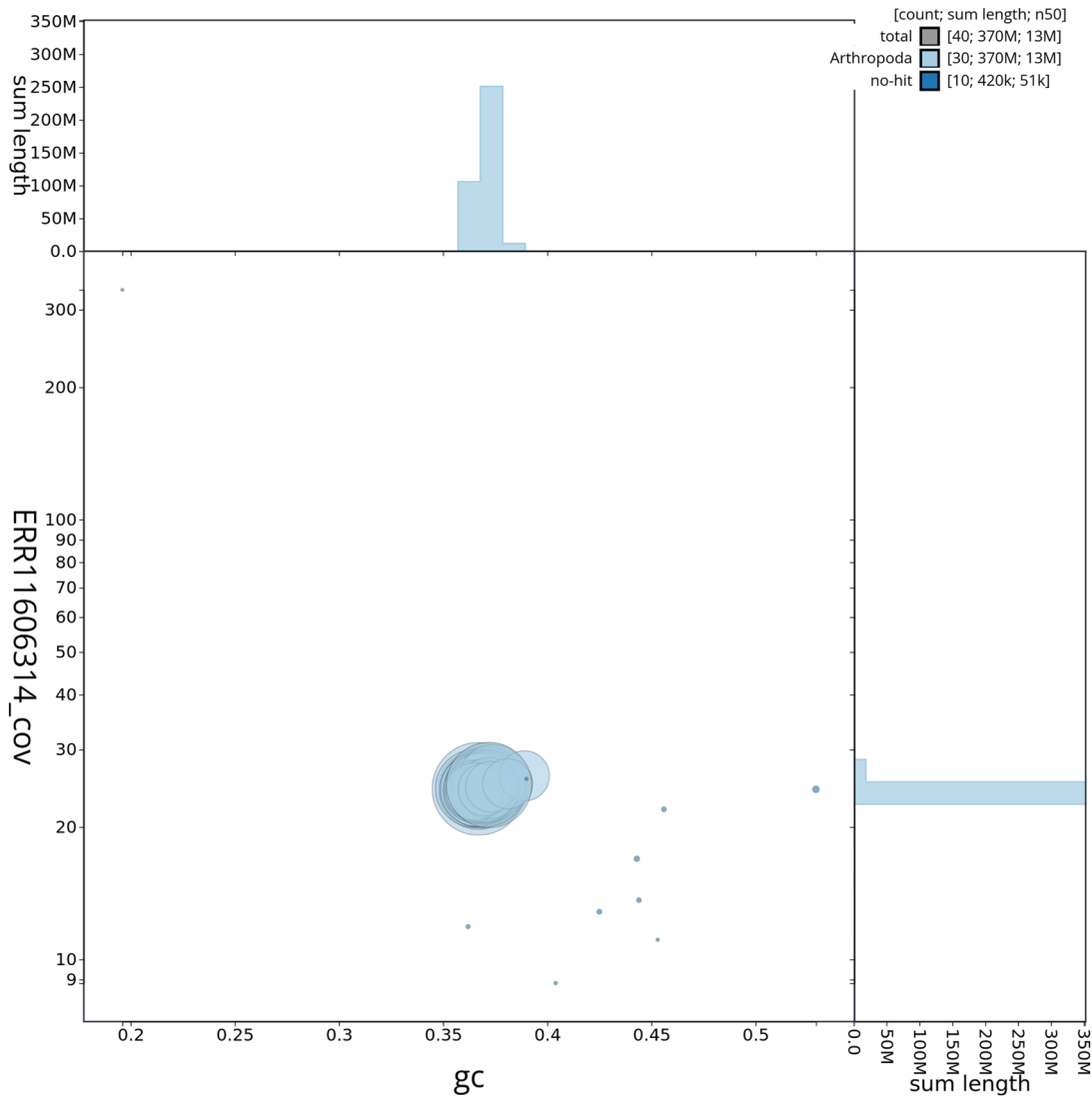
concentration was quantified using the Qubit Fluorometer v2.0 (Thermo Fisher Scientific) and Qubit HS Assay Kit and the final library fragment size analysis was carried out using the Agilent Femto Pulse Automated Pulsed Field CE Instrument (Agilent Technologies).

Samples were 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. The SMRT link software, a PacBio web-based end-to-end workflow

manager, was used to set-up and monitor the run, as well as perform primary and secondary analysis of the data upon completion.

#### Hi-C data

For Hi-C library preparation, DNA was fragmented to a size of 400 to 600 bp using a Covaris E220 sonicator. The DNA was then enriched, barcoded, and amplified using the NEBNext Ultra II DNA Library Prep Kit following manufacturers' instructions. The Hi-C sequencing was performed



**Figure 3. Genome assembly of *Perizoma flavofasciatum*, ilPerFlao1.1: BlobToolKit GC-coverage plot showing sequence 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 at [https://blobtoolkit.genomehubs.org/view/ilPerFlao1\\_1/dataset/ilPerFlao1\\_1/blob](https://blobtoolkit.genomehubs.org/view/ilPerFlao1_1/dataset/ilPerFlao1_1/blob).

using paired-end sequencing with a read length of 150 bp on an Illumina NovaSeq 6000 instrument.

#### RNA

Poly(A) RNA-Seq libraries were constructed using the NEB Ultra II RNA Library Prep kit, following the manufacturer's instructions. RNA sequencing was performed on the Illumina NovaSeq 6000 instrument.

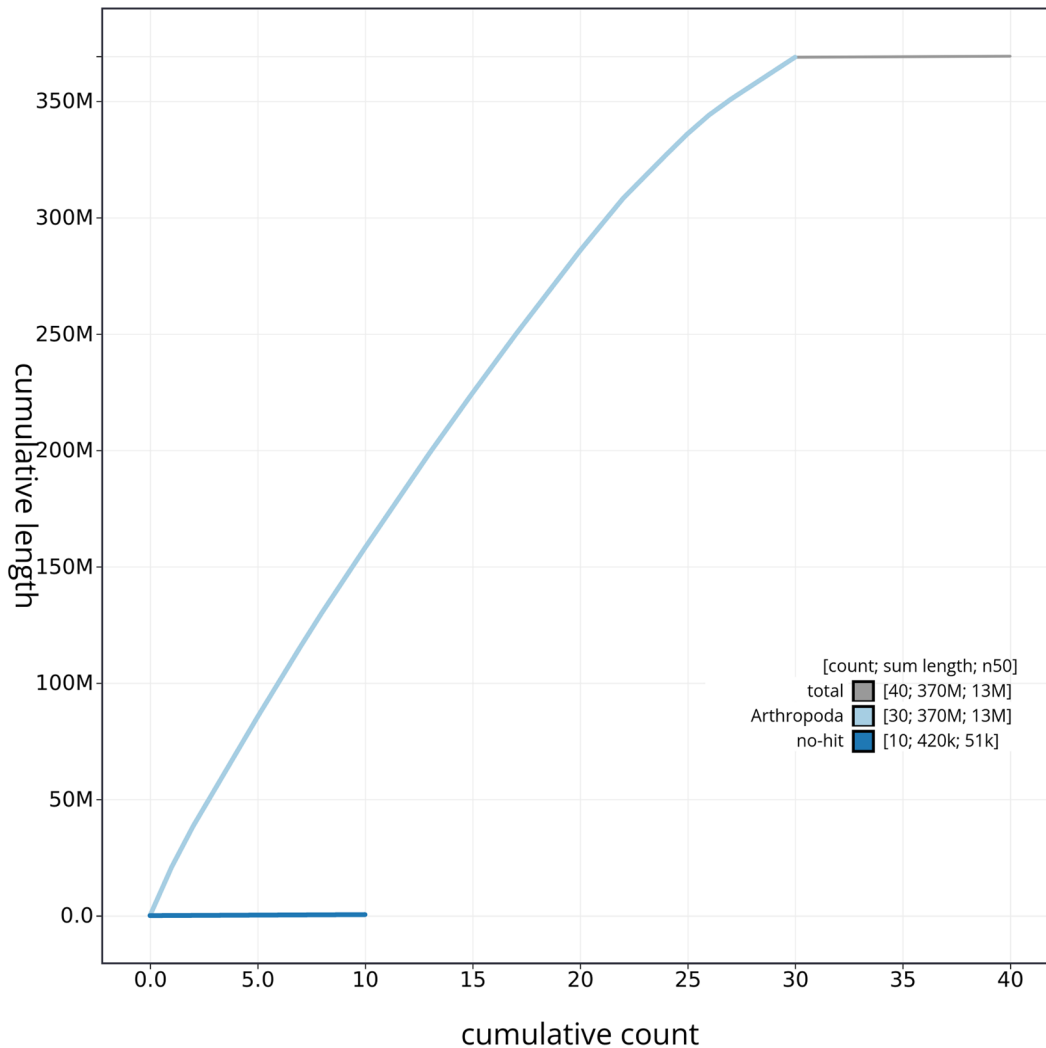
#### Genome assembly, curation and evaluation

##### Assembly

The HiFi reads were first assembled using Hifiasm (Cheng *et al.*, 2021) with the --primary option. Haplotypic duplications

were identified and removed using purge\_dups (Guan *et al.*, 2020). The Hi-C reads were mapped to the primary contigs using bwa-mem2 (Vasimuddin *et al.*, 2019). The contigs were further scaffolded using the provided Hi-C data (Rao *et al.*, 2014) in YaHS (Zhou *et al.*, 2023) using the --break option for handling potential misassemblies. The scaffolded assemblies were evaluated using Gfastats (Formenti *et al.*, 2022), BUSCO (Manni *et al.*, 2021) and MERQURY.FK (Rhie *et al.*, 2020).

The mitochondrial genome was assembled using MitoHiFi (Uliano-Silva *et al.*, 2023), which runs MitoFinder (Allio *et al.*, 2020) and uses these annotations to select the final



**Figure 4. Genome assembly of *Perizoma flavofasciatum* ilPerFlao1.1: 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 busco genes taxruler. An interactive version of this figure is available at [https://blobtoolkit.genomehubs.org/view/ilPerFlao1\\_1/dataset/ilPerFlao1\\_1/cumulative](https://blobtoolkit.genomehubs.org/view/ilPerFlao1_1/dataset/ilPerFlao1_1/cumulative).

mitochondrial contig and to ensure the general quality of the sequence.

#### Assembly curation

The assembly was decontaminated using the Assembly Screen for Cobionts and Contaminants (ASCC) pipeline (article in preparation). Flat files and maps used in curation were generated in TreeVal (Pointon *et al.*, 2023). Manual curation was primarily conducted using PretextView (Harry, 2022), with additional insights provided by JBrowse2 (Diesh *et al.*, 2023) and HiGlass (Kerpedjiev *et al.*, 2018). Scaffolds were visually inspected and corrected as described by Howe *et al.* (2021). Any identified contamination, missed joins, and misjoins were corrected, and duplicate sequences were tagged

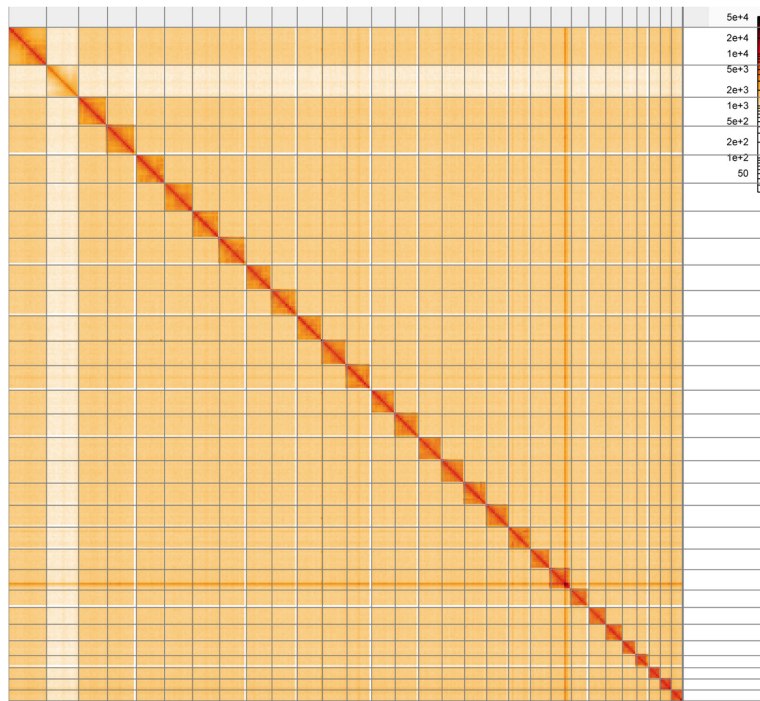
and removed. The sex chromosome was identified by synteny analysis. The curation process is documented at <https://gitlab.com/wtsi-grit/rapid-curation> (article in preparation).

#### Assembly quality assessment

The Merqury.FK tool (Rhie *et al.*, 2020) was used to evaluate  $k$ -mer completeness and assembly quality for the primary and alternate haplotypes using the  $k$ -mer databases ( $k = 31$ ) that were pre-computed prior to genome assembly. The analysis outputs included assembly QV scores and completeness statistics.

A Hi-C contact map was produced for the final, public version of the assembly. The Hi-C reads were aligned using bwa-mem2 (Vasimuddin *et al.*, 2019) and the alignment files





**Figure 5. Genome assembly of *Perizoma flavofasciatum* ilPerFlao1.1: Hi-C contact map of the ilPerFlao1.1 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/?d=KJLbAfRRWP4E1ra\\_QFAQ](https://genome-note-higlass.tol.sanger.ac.uk/?d=KJLbAfRRWP4E1ra_QFAQ).

**Table 3. Chromosomal pseudomolecules in the genome assembly of *Perizoma flavofasciatum*, ilPerFlao1.**

INSDC accession	Name	Length (Mb)	GC%
OY292419.1	1	20.74	36.5
OY292421.1	2	15.81	37.0
OY292422.1	3	15.75	37.0
OY292423.1	4	15.46	37.0
OY292424.1	5	15.35	36.5
OY292425.1	6	14.89	37.0
OY292426.1	7	14.63	36.5
OY292427.1	8	13.98	37.0
OY292428.1	9	13.86	37.0
OY292429.1	10	13.83	36.5
OY292430.1	11	13.42	37.0
OY292431.1	12	13.42	37.0
OY292432.1	13	12.96	37.0
OY292433.1	14	12.9	36.5

INSDC accession	Name	Length (Mb)	GC%
OY292434.1	15	12.72	37.0
OY292435.1	16	12.32	37.0
OY292436.1	17	12.15	37.0
OY292437.1	18	11.99	37.0
OY292438.1	19	11.94	37.0
OY292439.1	20	11.24	36.5
OY292440.1	21	11.2	37.5
OY292441.1	22	9.54	37.0
OY292442.1	23	9.2	36.5
OY292443.1	24	9.06	37.0
OY292444.1	25	8.06	36.5
OY292445.1	26	6.7	37.0
OY292446.1	27	6.17	37.5
OY292447.1	28	6.03	39.0
OY292448.1	29	5.97	38.0
OY292420.1	Z	17.6	37.0
OY292449.1	MT	0.02	19.5

were combined using SAMtools (Danecek *et al.*, 2021). The Hi-C alignments were converted into a contact map using BEDTools (Quinlan & Hall, 2010) and the Cooler tool suite (Abdennur & Mirny, 2020). The contact map is visualised in HiGlass (Kerpedjiev *et al.*, 2018).

The genome was analysed within the BlobToolKit environment (Challis *et al.*, 2020) and BUSCO scores (Manni *et al.*, 2021) were calculated.

Table 4 contains a list of relevant software tool versions and sources.

#### Genome annotation

The Ensembl Genebuild annotation system (Aken *et al.*, 2016) was used to generate annotation for the *Perizoma flavofasciatum* assembly (GCA\_958496245.1) in Ensembl Rapid Release at the EBI. Annotation was created primarily through alignment of transcriptomic data to the genome, with gap

**Table 4. Software tools: versions and sources.**

Software tool	Version	Source
BEDTools	2.30.0	<a href="https://github.com/arq5x/bedtools2">https://github.com/arq5x/bedtools2</a>
BLAST	2.14.0	<a href="ftp://ftp.ncbi.nlm.nih.gov/blast/executables/blast/">ftp://ftp.ncbi.nlm.nih.gov/blast/executables/blast/</a>
BlobToolKit	4.3.7	<a href="https://github.com/blobtoolkit/blobtoolkit">https://github.com/blobtoolkit/blobtoolkit</a>
BUSCO	5.4.3 and 5.5.0	<a href="https://gitlab.com/ezlab/busco">https://gitlab.com/ezlab/busco</a>
bwa-mem2	2.2.1	<a href="https://github.com/bwa-mem2/bwa-mem2">https://github.com/bwa-mem2/bwa-mem2</a>
Cooler	0.8.11	<a href="https://github.com/open2c/cooler">https://github.com/open2c/cooler</a>
DIAMOND	2.1.8	<a href="https://github.com/bbuchfink/diamond">https://github.com/bbuchfink/diamond</a>
fasta_windows	0.2.4	<a href="https://github.com/tolkit/fasta_windows">https://github.com/tolkit/fasta_windows</a>
FastK	427104ea91c78c3b8b8b49f1a7d6bbeaa869ba1c	<a href="https://github.com/thegenemyers/FASTK">https://github.com/thegenemyers/FASTK</a>
Gfastats	1.3.6	<a href="https://github.com/vgl-hub/gfastats">https://github.com/vgl-hub/gfastats</a>
Goat CLI	0.2.5	<a href="https://github.com/genomehubs/goat-cli">https://github.com/genomehubs/goat-cli</a>
Hifiasm	0.19.8-r587	<a href="https://github.com/chhylp123/hifiasm">https://github.com/chhylp123/hifiasm</a>
HiGlass	44086069ee7d4d3f6f3f0012569789ec138f42b84aa44357826c0b6753eb28de	<a href="https://github.com/higlass/higlass">https://github.com/higlass/higlass</a>
Mercury.FK	d00d98157618f4e8d1a9190026b19b471055b22e	<a href="https://github.com/thegenemyers/MERQURY.FK">https://github.com/thegenemyers/MERQURY.FK</a>
MitoHiFi	3	<a href="https://github.com/marcelauliano/MitoHiFi">https://github.com/marcelauliano/MitoHiFi</a>
MultiQC	1.14, 1.17, and 1.18	<a href="https://github.com/MultiQC/MultiQC">https://github.com/MultiQC/MultiQC</a>
NCBI Datasets	15.12.0	<a href="https://github.com/ncbi/datasets">https://github.com/ncbi/datasets</a>
Nextflow	23.04.0-5857	<a href="https://github.com/nextflow-io/nextflow">https://github.com/nextflow-io/nextflow</a>
PretextView	0.2.5	<a href="https://github.com/sanger-tol/PretextView">https://github.com/sanger-tol/PretextView</a>
purge_dups	1.2.5	<a href="https://github.com/dfguan/purge_dups">https://github.com/dfguan/purge_dups</a>
samtools	1.16.1, 1.17, and 1.18	<a href="https://github.com/samtools/samtools">https://github.com/samtools/samtools</a>
sanger-tol/ascc	-	<a href="https://github.com/sanger-tol/ascc">https://github.com/sanger-tol/ascc</a>
sanger-tol/blobtoolkit	0.6.0	<a href="https://github.com/sanger-tol/blobtoolkit">https://github.com/sanger-tol/blobtoolkit</a>
Seqtk	1.3	<a href="https://github.com/lh3/seqtk">https://github.com/lh3/seqtk</a>
Singularity	3.9.0	<a href="https://github.com/sylabs/singularity">https://github.com/sylabs/singularity</a>
TreeVal	1.0.0	<a href="https://github.com/sanger-tol/treeval">https://github.com/sanger-tol/treeval</a>
YaHS	1.2a.2	<a href="https://github.com/c-zhou/yahs">https://github.com/c-zhou/yahs</a>

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.

### Data availability

European Nucleotide Archive: *Perizoma flavofasciatum* (sandy carpet). Accession number PRJEB63430; <https://identifiers.org/ena.embl/PRJEB63430>. The genome sequence is released openly for reuse. The *Perizoma flavofasciatum* 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](#) and [Table 2](#).

### Author information

Members of the University of Oxford and Wytham Woods Genome Acquisition Lab are listed here: <https://doi.org/10.5281/zenodo.12157525>.

Members of the Darwin Tree of Life Barcoding collective are listed here: <https://doi.org/10.5281/zenodo.12158331>.

Members of the Wellcome Sanger Institute Tree of Life Management, Samples and Laboratory team are listed here: <https://doi.org/10.5281/zenodo.12162482>.

Members of Wellcome Sanger Institute Scientific Operations: Sequencing Operations are listed here: <https://doi.org/10.5281/zenodo.12165051>.

Members of the Wellcome Sanger Institute Tree of Life Core Informatics team are listed here: <https://doi.org/10.5281/zenodo.12160324>.

Members of the Tree of Life Core Informatics collective are listed here: <https://doi.org/10.5281/zenodo.12205391>.

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](#)
- Boyes D, Holland PWH, University of Oxford and Wytham Woods Genome Acquisition Lab, et al.: **The genome sequence of the Phoenix, *Eulithis prunata* (Linnaeus, 1758) [version 1; peer review: 2 approved, 1 approved with reservations]**. *Wellcome Open Res*. 2023; **8**: 185. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- British Lepidoptera: ***Perizoma flavofasciata* (Sandy Carpet)**. 2025; [Accessed 13 January 2025]. [Reference Source](#)
- 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](#)
- Danecek P, Bonfield JK, Liddle J, et al.: **Twelve years of SAMtools and BCFtools**. *GigaScience*. 2021; **10**(2): gjab008. [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](#)
- Diesh C, Stevens GJ, Xie P, et al.: **JBrowse 2: a modular genome browser with**

views of synteny and structural variation. *Genome Biol.* 2023; **24**(1): 74.

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

do Amaral RJV, Denton A, Yatsenko H, *et al.*: **Sanger Tree of Life RNA extraction: automated MagMax™ mirVana.** *protocols.io.* 2023.

[PubMed Abstract](#)

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](#)

Fox R, Parsons MS, Harrower CA: **A review of the status of the macro-moths of Great Britain.** Dorset, UK: Butterfly Conservation, 2019.

[Reference Source](#)

GBIF Secretariat: ***Perizoma flavofasciata* (Thunberg, 1792).** *GBIF Backbone Taxonomy*, 2023; [Accessed 13 January 2025].

[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.

[Reference Source](#)

Howe K, Chow W, Collins J, *et al.*: **Significantly improving the quality of genome assemblies through curation.** *GigaScience.* 2021; **10**(1): g1aa153.

[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.

[PubMed Abstract](#)

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](#)

Kimber I: **Sandy Carpet *Perizoma flavofasciata* (Thunberg, 1792).** *UKMoths*, 2025; [Accessed 13 January 2025].

[Reference Source](#)

Manni M, Berkeley MR, Seppy 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 Eval.* 2021; **38**(10): 4647–4654.

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

Pointon DL, Eagles W, Sims Y, *et al.*: **sanger-tol/treeval v1.0.0 – Ancient**

**Atlantis.** 2023.

[Publisher Full Text](#)

Quinlan AR, Hall IM: **BEDTools: a flexible suite of utilities for comparing genomic features.** *Bioinformatics.* 2010; **26**(6): 841–842.

[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.*: **Mercury: reference-free quality, completeness, and phasing assessment for genome assemblies.** *Genome Biol.* 2020; **21**(1): 245.

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

Sheerin E, Sampaio F, Oatley G, *et al.*: **Sanger Tree of Life HMW DNA extraction: automated MagAttract v.1.** *protocols.io.* 2023.

[Publisher Full Text](#)

Strickland M, Cornwell C, Howard C: **Sanger Tree of Life fragmented DNA clean up: manual SPRI.** *protocols.io.* 2023.

[Publisher Full Text](#)

Todorovic M, Sampaio F, Howard C: **Sanger Tree of Life HMW DNA fragmentation: diagenode Megaruptor®3 for PacBio HiFi.** *protocols.io.* 2023.

[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](#)

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 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](#)

Waring P, Townsend M, Lewington R: **Field guide to the Moths of Great Britain and Ireland: third edition.** Bloomsbury Wildlife Guides, 2017.

[Reference Source](#)

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](#)