






DATA NOTE

# The genome sequence of the Blue-bordered Carpet moth

## *Plemyria rubiginata* (Denis & Schiffermüller) 1775 [version 1;

### peer review: awaiting peer review]

Douglas Boyes <sup>1+</sup>, David C. Lees <sup>2</sup>, Denise C. Wawman <sup>3</sup>,  
University of Oxford and Wytham Woods Genome Acquisition Lab,  
Natural History Museum 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

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

<sup>2</sup>Natural History Museum, London, England, UK

<sup>3</sup>Edward Grey Institute, Department of Biology, University of Oxford, Oxford, England, UK

+ Deceased author

---

**V1** First published: 16 May 2024, 9:271  
<https://doi.org/10.12688/wellcomeopenres.21665.1>

Latest published: 16 May 2024, 9:271  
<https://doi.org/10.12688/wellcomeopenres.21665.1>

---

### Abstract

We present a genome assembly from an individual female *Plemyria rubiginata* (the Blue-bordered Carpet moth; Arthropoda; Insecta; Lepidoptera; Geometridae). The genome sequence is 356.2 megabases in span. Most of the assembly is scaffolded into 30 chromosomal pseudomolecules, including the Z and W sex chromosomes. The mitochondrial genome has also been assembled and is 17.64 kilobases in length.

### Keywords

*Plemyria rubiginata*, Blue-bordered Carpet moth, genome sequence, chromosomal, Lepidoptera

### 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.



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

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

**Author roles:** **Boyes D:** Investigation, Resources; **Lees DC:** Investigation, Resources, Writing – Review & Editing; **Wawman DC:** 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 [206194, <https://doi.org/10.35802/206194>] and the Darwin Tree of Life Discretionary Award [218328, <https://doi.org/10.35802/218328>]. *The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.*

**Copyright:** © 2024 Boyes D *et al.* This is an open access article distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

**How to cite this article:** Boyes D, Lees DC, Wawman DC *et al.* **The genome sequence of the Blue-bordered Carpet moth *Plemyria rubiginata* (Denis & Schiffermüller) 1775 [version 1; peer review: awaiting peer review]** Wellcome Open Research 2024, 9:271 <https://doi.org/10.12688/wellcomeopenres.21665.1>

**First published:** 16 May 2024, 9:271 <https://doi.org/10.12688/wellcomeopenres.21665.1>

## 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; Geometroidea; Geometridae; Larentiinae; *Plemyria*; *Plemyria rubiginata* (Denis & Schiffermüller) 1775 (NCBI:txid934957).

## Background

The Blue-bordered Carpet *Plemyria rubiginata* ([Denis & Schiffermüller], 1775) is a moth in the family Geometridae with a narrow forewing 12–15 mm long (typically 2,126 mm wingspan; Hausmann & Viidalepp, 2012).

Two subspecies have traditionally been recognised in the United Kingdom: *P. rubiginata rubiginata* with a whitish background and blue border, and a large brown blotch on the leading edge of the wing; and ‘subspecies’ *P. rubiginata plumbata* (Curtis, 1837), distinguished by the brown blotch extending to become a full crossband, a darker and brownish border, and a ground colour of greyish-beige (Waring *et al.*, 2017). However, the latter is internationally treated as *f. plumbata* (possibly an example of increasing northwards melanism at higher latitudes and elevations in northern England and Scotland, also in Norway and possibly Japan. By contrast, *P. rubiginata japonica* Inoue, 1955 is treated as validly distinct from the nominotypical subspecies from eastern Siberia and Mongolia to Japan (Hausmann & Viidalepp, 2012). Randle *et al.* (2019) show a continuous distribution, although *P. rubiginata* is more scattered in Scotland and more localised in Ireland. Broadly, the species spans from 0–1600 m across almost the entire Palaearctic from north-western Portugal and Spain to China and Japan and the southern shores of the Mediterranean to northern Scandinavia and north-western Russia, with an isolated population in Turkey (GBIF Secretariat, 2024; BOLD accessed 08/04/2024; Hausmann & Viidalepp, 2012).

Despite not being a species observed in large numbers, the Blue-bordered Carpet has shown a dramatic increase in abundance of 173% since 1970, although no change in distribution (Randle *et al.*, 2019). The moth is univoltine, flying between early June and late September in the UK with some sign of a shift to an earlier modal abundance (Randle *et al.*, 2019); and flying from mid-July at higher elevations in Europe (Hausmann & Viidalepp, 2012).

*P. rubiginata* overwinters as an egg on the twigs of its foodplant, and the greenish larvae with yellow lateral stripes feed among spinings on twig from April to June, preferring alder *Alnus glutinosa* (L.) Gaertn. or blackthorn *Prunus spinosa* L., but are also found on apple *Malus* spp., plum *Prunus* spp., hawthorn *Crataegus monogyna* Jacq. and birch *Betula* spp. (Sterling *et al.*, 2012; Waring *et al.*, 2017) foodplants also utilised by the Nearctic sister species, *P. georgii* Hulst, 1896. (Choi, 1998),

*P. rubiginata* occupies a single BIN cluster on BOLD (08/04/2024), BOLD:AAC1712, with up to 2.77% pairwise divergence within it ( $n = 33$ ), with Russian and Chinese exemplars showing highest divergences (~1.43% to the nearest exemplars). There are eight exemplars DNA barcoded from the UK (08/04/2024), the most northerly being from Lancashire but none so far from Scotland and Ireland, but at present there seems to be no mitochondrial evidence for genetic distinctiveness of northern UK populations. The Nearctic *P. georgii* (BOLD:AAB0097) by contrast is at least 5.65% divergent from it (although a species of *Chloroclystis*, BOLD:AFD5638, approaches *P. rubiginata* more closely at 4.17%).

The morphology and taxonomy of *Plemyria* Hübner, 1825 was reviewed by Choi (1998) and Hausmann & Viidalepp (2012), the genus comprising just the above two species. *Plemyria* belongs to the larentiinae tribe Cidariini. Based on a morphological analysis of adults and immatures, Choi (1997) placed the genus adjacent to *Chloroclystis* Hübner, 1825 and sister to *Thera* Stephens, 1831 + (*Pennithera* Viidalepp, 1980 + *Heterothera* Inoue, 1943). The genus was not, however, included in the molecular larentiine phylogeny of Öunap *et al.* (2016).

The genome will be useful in investigating industrial melanism, stimulating more DNA barcoding, and clarifying the phylogeography of the species and the history of its sister species split, as well as establishing the phylogenetic placement of *Plemyria*.

## Genome sequence report

The genome was sequenced from one female *Plemyria rubiginata* (Figure 1) collected from Wytham Woods, Oxfordshire, UK (51.76, -1.34). A total of 94-fold coverage in Pacific Biosciences single-molecule HiFi long reads was generated. Primary assembly contigs were scaffolded with



**Figure 1.** Photograph of the *Plemyria rubiginata* (iPleRubi1) specimen used for genome sequencing.

chromosome conformation Hi-C data. Manual assembly curation corrected 15 missing joins or mis-joins and removed 8 haplotypic duplications, reducing the assembly length by 1.07% and the scaffold number by 9.80%.

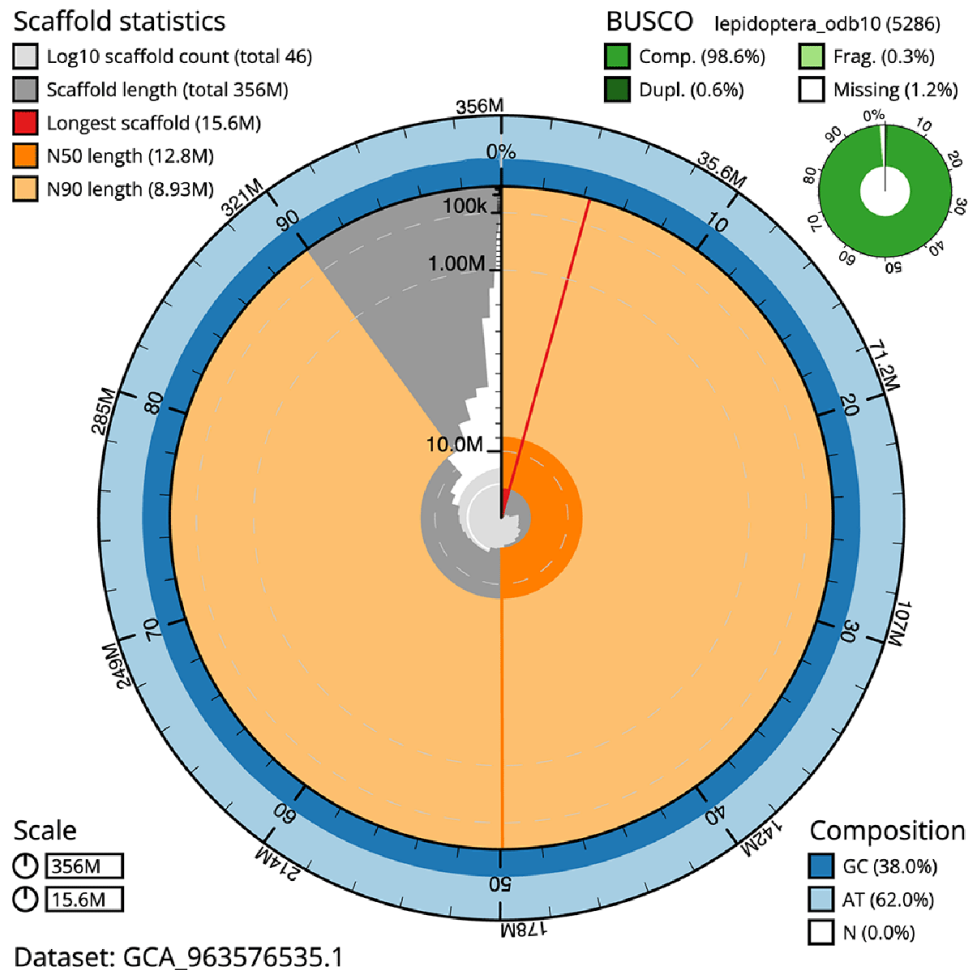
The final assembly has a total length of 356.2 Mb in 45 sequence scaffolds with a scaffold N50 of 12.8 Mb (Table 1). The snail plot in Figure 2 provides a summary of the assembly statistics, while the distribution of assembly scaffolds on

**Table 1. Genome data for *Plemyria rubiginata*, ilPleRubi1.1.**

Project accession data		
Assembly identifier	ilPleRubi1.1	
Species	<i>Plemyria rubiginata</i>	
Specimen	ilPleRubi1	
NCBI taxonomy ID	934957	
BioProject	PRJEB65207	
BioSample ID	SAMEA7701302	
Isolate information	ilPleRubi1, female: whole organism (DNA sequencing) ilPleRubi2: whole organism (Hi-C and RNA sequencing)	
Assembly metrics*		Benchmark
Consensus quality (QV)	63.1	≥ 50
<i>k</i> -mer completeness	100.0%	≥ 95%
BUSCO**	C:98.5%[S:97.9%,D:0.6%], F:0.3%,M:1.2%,n:5,286	C ≥ 95%
Percentage of assembly mapped to chromosomes	99.87%	≥ 95%
Sex chromosomes	ZW	<i>localised homologous pairs</i>
Organelles	Mitochondrial genome: 17.64 kb	<i>complete single alleles</i>
Raw data accessions		
PacificBiosciences Sequel IIe	ERR11867205, ERR11867206	
Hi-C Illumina	ERR11872574	
PolyA RNA-Seq Illumina	ERR12245586	
Genome assembly		
Assembly accession	GCA_963576535.1	
Accession of alternate haplotype	GCA_963576495.1	
Span (Mb)	356.2	
Number of contigs	68	
Contig N50 length (Mb)	11.8	
Number of scaffolds	45	
Scaffold N50 length (Mb)	12.8	
Longest scaffold (Mb)	15.64	

\* 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 v5.4.3. 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/Plemyria\\_rubiginata/dataset/GCA\\_963576535.1/busco](https://blobtoolkit.genomehubs.org/view/Plemyria_rubiginata/dataset/GCA_963576535.1/busco).



**Figure 2. Genome assembly of *Plemyria rubiginata*, iPleRubi1.1: 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 356,234,577 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 (15,641,424 bp, shown in red). Orange and pale-orange arcs show the N50 and N90 scaffold lengths (12,846,343 and 8,926,186 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/Plemyria\\_rubiginata/dataset/GCA\\_963576535.1/snail](https://blobtoolkit.genomehubs.org/view/Plemyria_rubiginata/dataset/GCA_963576535.1/snail).

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 (99.87%) of the assembly sequence was assigned to 30 chromosomal-level scaffolds, representing 28 autosomes and the Z and W sex chromosomes. Chromosome-scale scaffolds confirmed by the Hi-C data are named in order of size (Figure 5; Table 2). While not fully phased, the assembly deposited is of one haplotype. Contigs corresponding to the second haplotype have also been deposited. The mitochondrial genome was also assembled and can be found as a contig within the multifasta file of the genome submission.

The estimated Quality Value (QV) of the final assembly is 63.1 with *k*-mer completeness of 100.0%, and the assembly

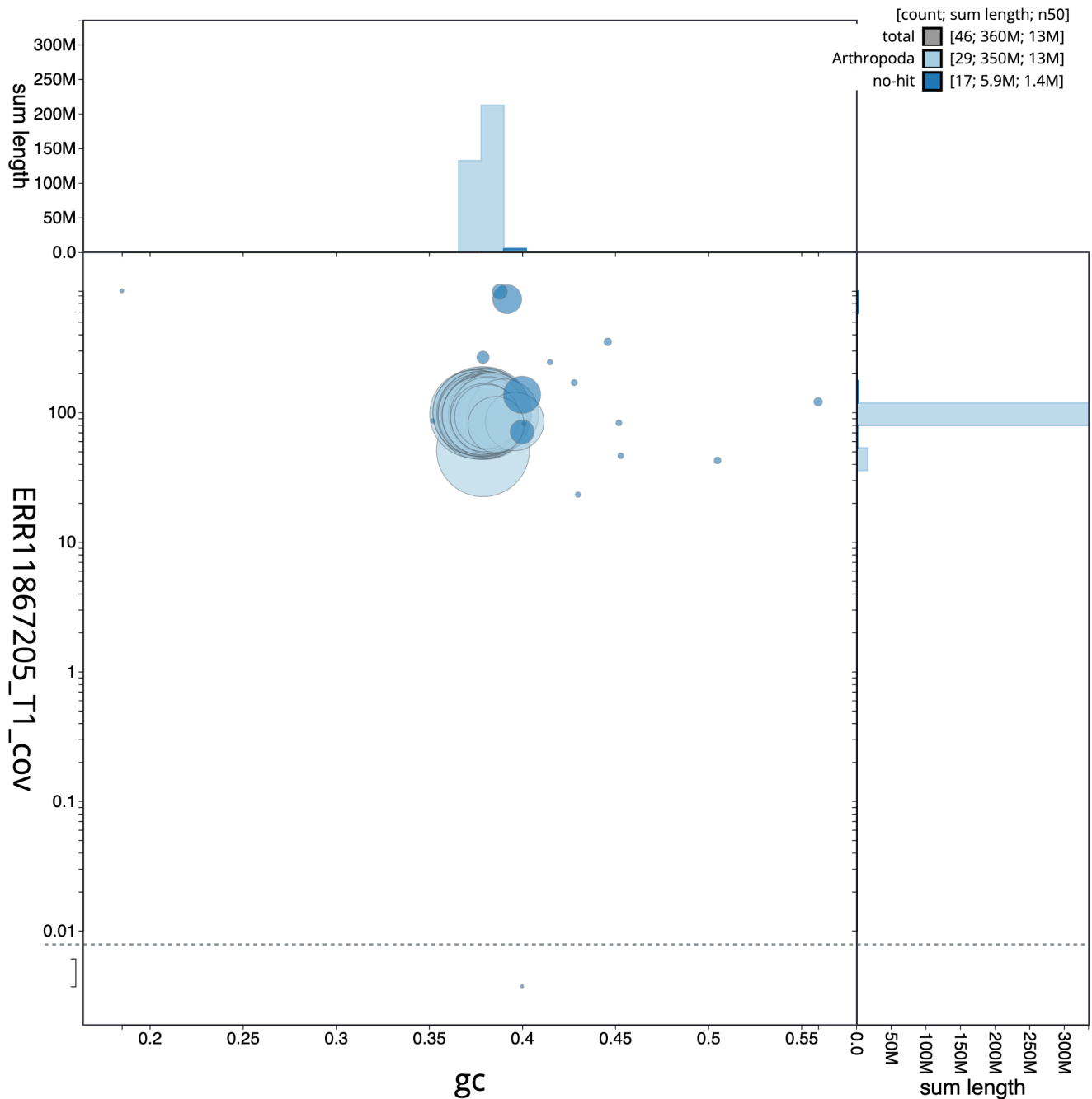
has a BUSCO v completeness of 98.5% (single = 97.9%, duplicated = 0.6%), using the lepidoptera\_odb10 reference set ( $n = 5,286$ ).

Metadata for specimens, BOLD barcode results, spectra estimates, sequencing runs, contaminants and pre-curation assembly statistics are given at <https://links.tol.sanger.ac.uk/species/934957>.

## Methods

### Sample acquisition and nucleic acid extraction

A female *Plemyria rubiginata* (specimen ID Ox000535, ToLID iPleRubi1) was collected from Wytham Woods, Oxfordshire (biological vice-county Berkshire), UK (latitude 51.76, longitude -1.34) on 2020-06-25 using a light trap. The



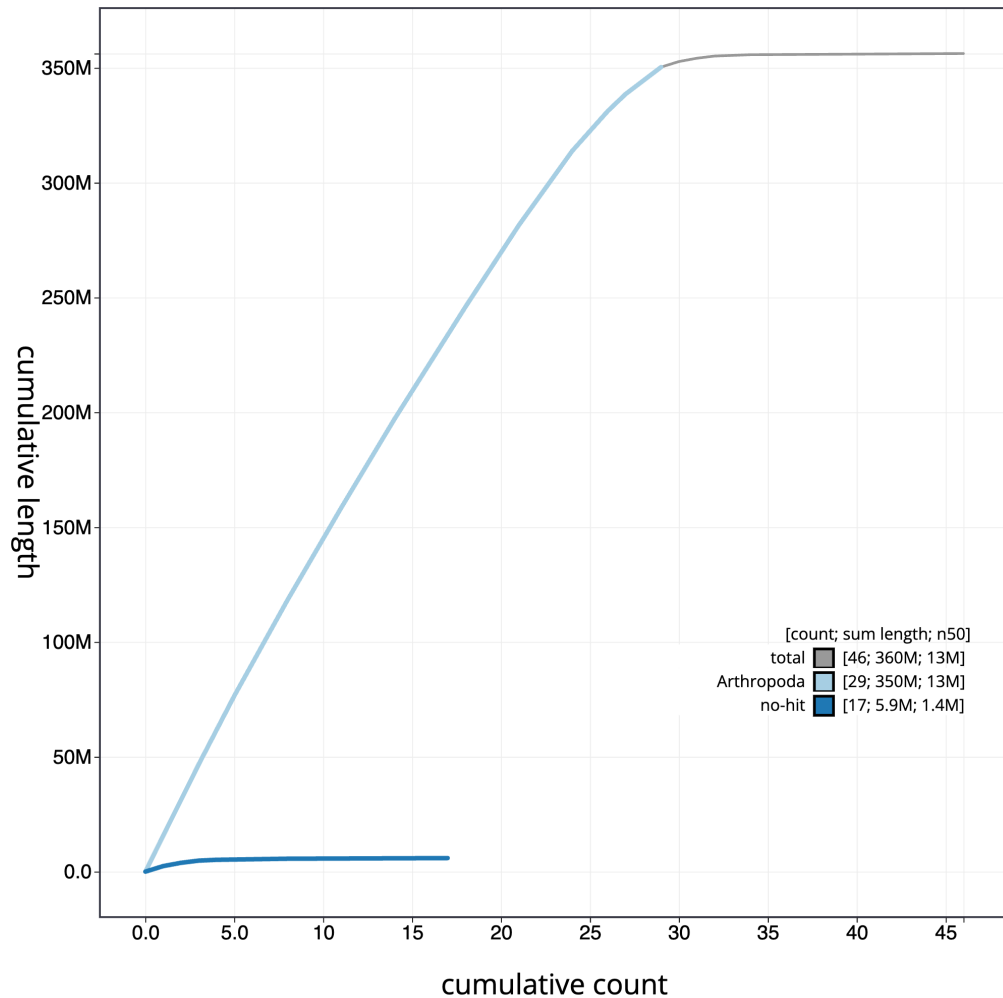
**Figure 3. Genome assembly of *Plemyria rubiginata*, ilPleRubi1.1: BlobToolKit GC-coverage plot.** Sequences are coloured by phylum. Circles are sized in proportion to sequence length. Histograms show the distribution of sequence length sum along each axis. An interactive version of this figure is available at [https://blobtoolkit.genomehubs.org/view/Plemyria\\_rubiginata/dataset/GCA\\_963576535.1/blob](https://blobtoolkit.genomehubs.org/view/Plemyria_rubiginata/dataset/GCA_963576535.1/blob).

specimen was collected and identified by Douglas Boyes (University of Oxford) and preserved on dry ice.

The specimen used for Hi-C and RNA sequencing (specimen ID NHMUK014584832, ToLID ilPleRubi2) was hand-picked in Lucas Wood, High Wycombe, London (latitude 51.63, longitude -0.74) on 2022-06-23. The specimen was collected

and identified by David Lees (Natural History Museum) and preserved by dry-freezing at  $-80^{\circ}\text{C}$ .

The workflow for high molecular weight (HMW) DNA extraction at the Wellcome Sanger Institute (WSI) Tree of Life Core Laboratory includes a sequence of core procedures: sample preparation; sample homogenisation, DNA extraction,



**Figure 4. Genome assembly of *Plemyria rubiginata*, ilPleRubi1.1: 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/Plemyria\\_rubiginata/dataset/GCA\\_963576535.1/cumulative](https://blobtoolkit.genomehubs.org/view/Plemyria_rubiginata/dataset/GCA_963576535.1/cumulative).

fragmentation, and clean-up. In sample preparation, the ilPleRubi1 sample was weighed and dissected 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 with speed setting 30 (Todorovic *et al.*, 2023). Sheared DNA was purified by solid-phase reversible immobilisation (Strickland *et al.*, 2023): in brief, the method employs a 1.8X ratio of AMPure PB beads to sample to eliminate shorter fragments and concentrate the DNA. The concentration of the sheared and purified DNA was assessed using a Nanodrop spectrophotometer and Qubit Fluorometer and Qubit dsDNA High Sensitivity Assay kit. Fragment size distribution was evaluated by running the sample on the FemtoPulse system.

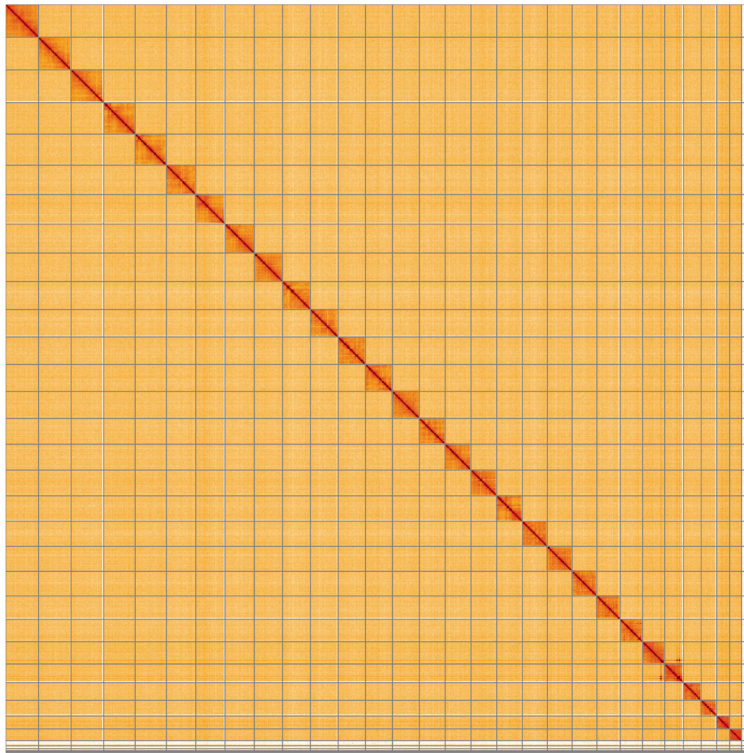
RNA was extracted from tissue of ilPleRubi2 in the Tree of Life Laboratory at the WSI using the RNA Extraction: Automated MagMax™ mirVana protocol (do Amaral *et al.*, 2023). The RNA concentration was assessed using a Nanodrop spectrophotometer and a Qubit Fluorometer using the Qubit RNA Broad-Range Assay kit. Analysis of the integrity of the RNA was done using the Agilent RNA 6000 Pico Kit and Eukaryotic Total RNA assay.

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. Poly(A) RNA-Seq libraries were constructed using the NEB Ultra II RNA Library Prep kit. DNA and RNA sequencing was performed by the Scientific Operations core





**Figure 5. Genome assembly of *Plemyria rubiginata*, ilPleRubi1.1: Hi-C contact map of the ilPleRubi1.1 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=N6e0u10NSg-QOuCMZCLXhQ>.

**Table 2. Chromosomal pseudomolecules in the genome assembly of *Plemyria rubiginata*, ilPleRubi1.**

INSDC accession	Chromosome	Length (Mb)	GC%
OY754906.1	1	15.54	38.0
OY754907.1	2	15.54	38.0
OY754908.1	3	14.99	38.0
OY754909.1	4	14.82	38.0
OY754910.1	5	13.99	37.5
OY754911.1	6	13.97	38.0
OY754912.1	7	13.87	37.5
OY754913.1	8	13.47	38.0
OY754914.1	9	13.31	37.5
OY754915.1	10	13.13	37.5
OY754916.1	11	13.09	37.5
OY754917.1	12	12.85	38.0
OY754918.1	13	12.76	38.0
OY754919.1	14	12.3	37.5

INSDC accession	Chromosome	Length (Mb)	GC%
OY754920.1	15	12.3	37.5
OY754921.1	16	12.25	38.0
OY754922.1	17	12.12	38.5
OY754923.1	18	11.96	38.0
OY754924.1	19	11.83	38.0
OY754925.1	20	11.73	38.0
OY754926.1	21	11.09	38.0
OY754927.1	22	10.6	38.5
OY754928.1	23	10.65	38.0
OY754929.1	24	8.93	39.0
OY754930.1	25	8.44	38.0
OY754931.1	26	7.39	38.0
OY754932.1	27	6.13	39.5
OY754933.1	28	5.64	38.5
OY754934.1	W	2.41	40.0
OY754905.1	Z	15.64	38.0
OY754935.1	MT	0.02	18.5



at the WSI on Pacific Biosciences Sequel IIe (HiFi) and Illumina NovaSeq 6000 (RNA-Seq) instruments. Hi-C data were also generated from tissue of iPLeRubi2 using the Arima2 kit and sequenced on the Illumina NovaSeq 6000 instrument.

### Genome assembly and curation

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 using the TreeVal pipeline (Pointon *et al.*, 2023). Manual curation was performed using JBrowse2 (Diesh *et al.*, 2023), HiGlass (Kerpedjiev *et al.*, 2018) 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.

### Final assembly evaluation

The final assembly was post-processed and evaluated with the three Nextflow (Di Tommaso *et al.*, 2017) DSL2 pipelines “sanger-tol/readmapping” (Surana *et al.*, 2023a), “sanger-tol/genomenote” (Surana *et al.*, 2023b), and “sanger-tol/blobtoolkit” (Muffato *et al.*, 2024). The pipeline sanger-tol/readmapping aligns the Hi-C reads with bwa-mem2 (Vasimuddin *et al.*, 2019) and combines the alignment files with SAMtools (Danecek *et al.*, 2021). The sanger-tol/genomenote pipeline transforms the Hi-C alignments into a contact map with BEDTools (Quinlan & Hall, 2010) and the Cooler tool suite (Abdennur & Mirny, 2020), which is then visualised with HiGlass (Kerpedjiev *et al.*, 2018). It also provides statistics about the assembly with the NCBI datasets (Sayers *et al.*, 2024) report, computes *k*-mer completeness and QV consensus

quality values with FastK and MerquryFK, and a completeness assessment with BUSCO (Manni *et al.*, 2021).

The sanger-tol/blobtoolkit pipeline is a Nextflow port of the previous Snakemake Blobtoolkit pipeline (Challis *et al.*, 2020). It aligns the PacBio reads with SAMtools and minimap2 (Li, 2018) and generates coverage tracks for regions of fixed size. In parallel, it queries the GoT database (Challis *et al.*, 2023) to identify all matching BUSCO lineages to run BUSCO (Manni *et al.*, 2021). For the three domain-level BUSCO lineage, the pipeline aligns the BUSCO genes to the Uniprot Reference Proteomes database (Bateman *et al.*, 2023) with DIAMOND (Buchfink *et al.*, 2021) blastp. The genome is also split into chunks according to the density of the BUSCO genes from the closest taxonomically lineage, and each chunk is aligned to the Uniprot Reference Proteomes database with DIAMOND blastx. Genome sequences that have no hit are then chunked with seqtk and aligned to the NT database with blastn (Altschul *et al.*, 1990). All those outputs are combined with the blobtools suite into a blobdir for visualisation.

All three pipelines were developed using the nf-core tooling (Ewels *et al.*, 2020), use MultiQC (Ewels *et al.*, 2016), and make extensive use of the Conda package manager, the Bioconda initiative (Grüning *et al.*, 2018), the Biocontainers infrastructure (da Veiga Leprevost *et al.*, 2017), and the Docker (Merkel, 2014) and Singularity (Kurtzer *et al.*, 2017) containerisation solutions.

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

### 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

**Table 3. Software tools: versions and sources.**

Software tool	Version	Source
BEDTools	2.30.0	<a href="https://github.com/arg5x/bedtools2">https://github.com/arg5x/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/tolkkit/fasta_windows">https://github.com/tolkkit/fasta_windows</a>
FastK	427104ea91c78c3b8b8b49f1a7d6bbeaa869ba1c	<a href="https://github.com/thegenemyers/FASTK">https://github.com/thegenemyers/FASTK</a>
GoAT CLI	0.2.5	<a href="https://github.com/genomehubs/goat-cli">https://github.com/genomehubs/goat-cli</a>
Hifiasm	0.19.5-r587	<a href="https://github.com/chhy123/hifiasm">https://github.com/chhy123/hifiasm</a>
HiGlass	44086069ee7d4d3f6f3f0012569789ec138f42b84aa44357826c0b6753eb28de	<a href="https://github.com/higlass/higlass">https://github.com/higlass/higlass</a>
MerquryFK	d00d98157618f4e8d1a9190026b19b471055b22e	<a href="https://github.com/thegenemyers/MERQURY.FK">https://github.com/thegenemyers/MERQURY.FK</a>

Software tool	Version	Source
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	<a href="https://github.com/wtsi-hpag/PretextView">https://github.com/wtsi-hpag/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/genomenote	1.1.1	<a href="https://github.com/sanger-tol/genomenote">https://github.com/sanger-tol/genomenote</a>
sanger-tol/readmapping	1.2.1	<a href="https://github.com/sanger-tol/readmapping">https://github.com/sanger-tol/readmapping</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>

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: *Plemyria rubiginata* (blue-bordered carpet). Accession number PRJEB65207; <https://identifiers.org/ena.embl/PRJEB65207> (Wellcome Sanger Institute, 2023).

The genome sequence is released openly for reuse. The *Plemyria rubiginata* genome sequencing initiative is part of the Darwin Tree of Life (DTOL) project. All raw sequence data and the assembly have been deposited in INSDC databases. The genome will be annotated using available RNA-Seq data and presented through the [Ensembl](#) pipeline at the European Bioinformatics Institute. Raw data and assembly accession identifiers are reported in [Table 1](#).

### Author information

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

Members of the Natural History Museum Genome Acquisition Lab are listed here: <https://doi.org/10.5281/zenodo.7139035>.

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](#)
- 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 MJ, 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](#)
- 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](#)
- Buchfink B, Reuter K, Drost HG: **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, Kumar S, Sotero-Caio C, et al.: **Genomes on a Tree (GoAT): a versatile, scalable search engine for genomic and sequencing project metadata across the eukaryotic tree of life [version 1; peer review: 2 approved].** *Wellcome Open Res.* 2023; **8**: 24. [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](#)
- Choi S: **A phylogenetic study on genera of cidariini from the holarctic and the Indo-Australian areas (lepidoptera: geometridae: larentiinae).** *Syst Entomol.* 1997; **22**(4): 287–312. [Publisher Full Text](#)
- Choi SW: **Taxonomy of the genus *Plemyria* hiibner (lepidoptera, geometridae, larentiinae).** *Entomol Fennica.* 1998; **9**: 185–196. [Reference Source](#)
- da Veiga Leprevost F, Grüning BA, Alves Afritos S, et al.: **BioContainers: an open-source and community-driven framework for software standardization.** *Bioinformatics.* 2017; **33**(16): 2580–2582. [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](#)
- 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](#)
- 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, Bates A, Denton A, et al.: **Sanger Tree of Life RNA extraction: automated MagMax™ mirVana.** *protocols.io.* 2023. [Publisher 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](#) | [Publisher Full Text](#)
- 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](#)
- GBIF Secretariat: ***Plemyria rubiginata* (Denis & Schiffermüller) 1775.** *GBIF Backbone Taxonomy.* 2024; [Accessed 17 April 2024]. [Reference Source](#)
- Grüning B, Dale R, Sjödin A, et al.: **Bioconda: sustainable and comprehensive software distribution for the life sciences.** *Nat Methods.* 2018; **15**(7): 475–476. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- 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](#)
- Hausmann A, Viidalepp J: **Geometrid moths of Europe volume 3, larentiinae 1.** Stenstrup: Apollo Books, 2012; **3**. [Reference Source](#)
- 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](#)
- 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](#)
- 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, Seppely M, et al.: **BUSCO update: novel and streamlined workflows along with broader and deeper phylogenetic coverage for scoring of eukaryotic, prokaryotic, and viral genomes.** *Mol Biol Evol.* 2021; **38**(10): 4647–4654. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Merkel D: **Docker: lightweight Linux containers for consistent development and deployment.** *Linux J.* 2014; **2014**(239): 2. [Reference Source](#)
- Muffato M, Butt Z, Challis R, et al.: **sanger-tol/blobtoolkit: v0.3.0 – Poliwig.** 2024. [Publisher Full Text](#)
- Öunap E, Viidalepp J, Truuverk A: **Phylogeny of the subfamily larentiinae (lepidoptera: geometridae): integrating molecular data and traditional classifications.** *Syst Entomol.* 2016; **41**(4): 824–843. [Publisher 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](#)
- Randle Z, Evans-Hill LJ, Parsons MS, et al.: **Atlas of Britain & Ireland's larger moths.** Newbury: NatureBureau, 2019.
- 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](#)
- Sayers EW, Cavanaugh M, Clark K, et al.: **GenBank 2024 update.** *Nucleic Acids Res.* 2024; **52**(D1): D134–D137. [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](#)
- Sterling P, Parsons M, Lewington R: **Field guide to the micro moths of Great Britain and Ireland.** Gillingham, Dorset: British Wildlife Publishing, 2012. [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](#)
- Todorovic M, Sampaio F, Howard C: **Sanger Tree of Life HMW DNA fragmentation: diagenode Megaruptor<sup>3</sup> for PacBio HiFi.** *protocols.io.* 2023. [Publisher Full Text](#)
- Uliano-Silva M, Ferreira JGRN, Krashennnikova 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](#)
- Waring P, Townsend M, Lewington R, et al.: **Field guide to the moths of Great Britain and Ireland: Third Edition.** Bloomsbury Wildlife Guides, 2017. [Reference Source](#)
- Wellcome Sanger Institute: **The genome sequence of the Blue-bordered carpet moth *Plemyria rubiginata* (Denis & Schiffermüller) 1775.** European Nucleotide Archive. [dataset], accession number PRJEB65207, 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](#)