DATA NOTE

The genome sequence of the double-striped pug, *Gymnoscelis rufifasciata* (Haworth, 1809) [version 1; peer review: 3 approved]

Douglas Boyes[1+],
University of Oxford and Wytham Woods Genome Acquisition Lab,
Darwin Tree of Life Barcoding collective,
Wellcome Sanger Institute Tree of Life programme,
Wellcome Sanger Institute Scientific Operations: DNA Pipelines collective,
Tree of Life Core Informatics collective, Thomas Lewin[2],
Darwin Tree of Life Consortium

[1]UK Centre for Ecology and Hydrology, Wallingford, Oxfordshire, UK
[2]Department of Zoology, University of Oxford, Oxford, UK

[+] Deceased author

v1 **First published:** 13 Apr 2022, **7**:135
https://doi.org/10.12688/wellcomeopenres.17790.1
Latest published: 13 Apr 2022, **7**:135
https://doi.org/10.12688/wellcomeopenres.17790.1

Abstract

We present a genome assembly from an individual female *Gymnoscelis rufifasciata* (the double-striped pug; Arthropoda; Insecta; Lepidoptera; Geometridae). The genome sequence is 352 megabases in span. The majority of the assembly (99.82%) is scaffolded into 32 chromosomal pseudomolecules, with the W and Z sex chromosomes assembled. The mitochondrial genome was also assembled, and is 15.4 kilobases in length.

Keywords

Gymnoscelis rufifasciata, double-striped pug, genome sequence, chromosomal, Lepidoptera

This article is included in the Tree of Life gateway.

Open Peer Review

Approval Status ✓ ✓ ✓

	1	2	3
version 1 13 Apr 2022	✓ view	✓ view	✓ view

1. **Henry North**, University of Cambridge, Cambridge, UK
2. **Simon H Martin**, University of Edinburgh, Edinburgh, UK
3. **Christopher B Cunningham**, University of Georgia, Athens, USA

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

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

Author roles: Boyes D: Investigation, Resources; **Lewin T**: 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.

How to cite this article: Boyes D, University of Oxford and Wytham Woods Genome Acquisition Lab, Darwin Tree of Life Barcoding collective *et al.* **The genome sequence of the double-striped pug, *Gymnoscelis rufifasciata* (Haworth, 1809) [version 1; peer review: 3 approved]** Wellcome Open Research 2022, **7**:135 https://doi.org/10.12688/wellcomeopenres.17790.1

First published: 13 Apr 2022, **7**:135 https://doi.org/10.12688/wellcomeopenres.17790.1

Species taxonomy

Eukaryota; Metazoa; Ecdysozoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Lepidoptera; Glossata; Ditrysia; Geometroidea; Geometridae; Larentiinae; Gymnoscelis; *Gymnoscelis rufifasciata* (Haworth, 1809) (NCBI: txid934940).

Background

The double-striped pug (*Gymnoscelis rufifasciata*), named after the two dark bands across its wings, is a moth of the family Geometridae. A typical adult's wingspan is 15 to 19 mm, with forewings varying from light brown to dark reddish-brown, and the intensity of its stripes also variable (Riley & Prior, 2003; Skinner & Wilson, 2009). A double brooded species, adults are first on the wing in the UK from March to May, although they have been observed as early as January in the mildest winters, and then again from July to August (Randle *et al.*, 2019); they are frequently found in light traps throughout these periods.

Gymnoscelis rufifasciata is common across western and central Europe, and while it has long been common in southern England, observations in northern England and Scotland have increased dramatically since the year 2000 (Fox *et al.*, 2021), likely caused by the increasing suitability of the climate due to rising average temperatures. Indeed, *G. rufifasciata* appears to be thriving in the current climate, and its abundance in Britain was reported to have increased 220% from 1986 to 2016 (Randle *et al.*, 2019); its generalist habitat usage, which includes gardens, wasteland, heathland and hedgerows (Riley & Prior, 2003; Randle *et al.*, 2019) is likely to have contributed to its recent success. Its larvae are polyphagous, and favoured food plants include gorse (*Ulex*), holly (*Ilex*) and heather (*Calluna*) (Riley & Prior, 2003; Skinner & Wilson, 2009).

Genome sequence report

The genome was sequenced from one female *G. rufifasciata* (Figure 1) collected from Wytham Woods, Oxfordshire (biological vice-county: Berkshire), UK (latitude 51.765, longitude -1.327). A total of 58-fold coverage in Pacific Biosciences single-molecule long reads and 102-fold coverage in 10X Genomics read clouds were generated. Primary assembly contigs were scaffolded with chromosome conformation Hi-C data. Manual assembly curation corrected 11 missing/misjoins and removed 3 haplotypic duplications, reducing the assembly size by 0.54% and the scaffold number by 19.05%.

Figure 1. Image of the *Gymnoscelis rufifasciata* (ilGymRufi1) specimen taken prior to preservation and processing. Specimen shown next to FluidX storage tube, 43.9 mm in length.

The final assembly has a total length of 462 Mb in 51 sequence scaffolds with a scaffold N50 of 15.6 Mb (Table 1). The majority of the assembly sequence (99.82%) was assigned to 32 chromosomal-level scaffolds, representing 30 autosomes

Table 1. Genome data for *Gymnoscelis rufifasciata*, ilGymRufi1.1.

Project accession data	
Assembly identifier	ilGymRufi1.1
Species	*Gymnoscelis rufifasciata*
Specimen	ilGymRufi1 (genome assembly); ilGymRufi2 (Hi-C)
NCBI taxonomy ID	NCBI:txid934940
BioProject	PRJEB48374
BioSample ID	SAMEA7519910
Isolate information	Female, whole organism (ilGymRufi1 genome assembly), unknown sex, whole organism (Hi-C, ilGymRufi2)
Raw data accessions	
PacificBiosciences SEQUEL II	ERR7221644
10X Genomics Illumina	ERR7220453-ERR7220456
Hi-C Illumina	ERR7220457
Genome assembly	
Assembly accession	GCA_929108375.1
Accession of alternate haplotype	GCA_929108405.1
Span (Mb)	462
Number of contigs	62
Contig N50 length (Mb)	15.6
Number of scaffolds	51
Scaffold N50 length (Mb)	15.6
Longest scaffold (Mb)	18.8
BUSCO* genome score	C:98.1%[S:97.7%,D:0.5%], F:0.5%,M:1.3%,n:5286

*BUSCO scores based on the lepidoptera_odb10 BUSCO set using v5.2.2. 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/ilGymRufi1.1/dataset/CAKMYF01/busco.

(numbered by sequence length), and the W and Z sex chromosomes (Figure 2–Figure 5; Table 2). The assembly has a BUSCO v5.2.2 (Manni *et al.*, 2021) completeness of 98.1% (single 97.7%, duplicated 0.5%) using the lepidoptera_odb10 reference set. While not fully phased, the assembly deposited is of one haplotype. Contigs corresponding to the second haplotype have also been deposited.

Methods

Sample acquisition, DNA extraction and sequencing

A single female *G. rufifasciata* (ilGymRufi1) and a second *G. rufifasciata* of unknown sex (ilGymRufi2) were collected from Wytham Woods, Oxfordshire (biological vice-county: Berkshire), UK (ilGymRufi1: latitude 51.765, longitude -1.327; ilGymRufi2: latitude 51.772, longitude -1.338) by Douglas Boyes, UKCEH, using a light trap in woodland. The sample was identified by the same individual, and preserved on dry ice.

DNA was extracted from whole organism tissue at the Wellcome Sanger Institute (WSI) Scientific Operations core from the whole organism using the Qiagen MagAttract HMW DNA kit, according to the manufacturer's instructions. Pacific Biosciences HiFi circular consensus and 10X Genomics Chromium read cloud sequencing libraries were constructed according to the manufacturers' instructions. Sequencing was performed by the Scientific Operations core at the Wellcome Sanger Institute on Pacific Biosciences SEQUEL II (HiFi) and Illumina HiSeq X (10X) instruments. Hi-C data were generated from head tissue of ilGymRufi2 using the Arima Hi-C+ kit and sequenced on an Illumina NovaSeq 6000 instrument.

Genome assembly

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

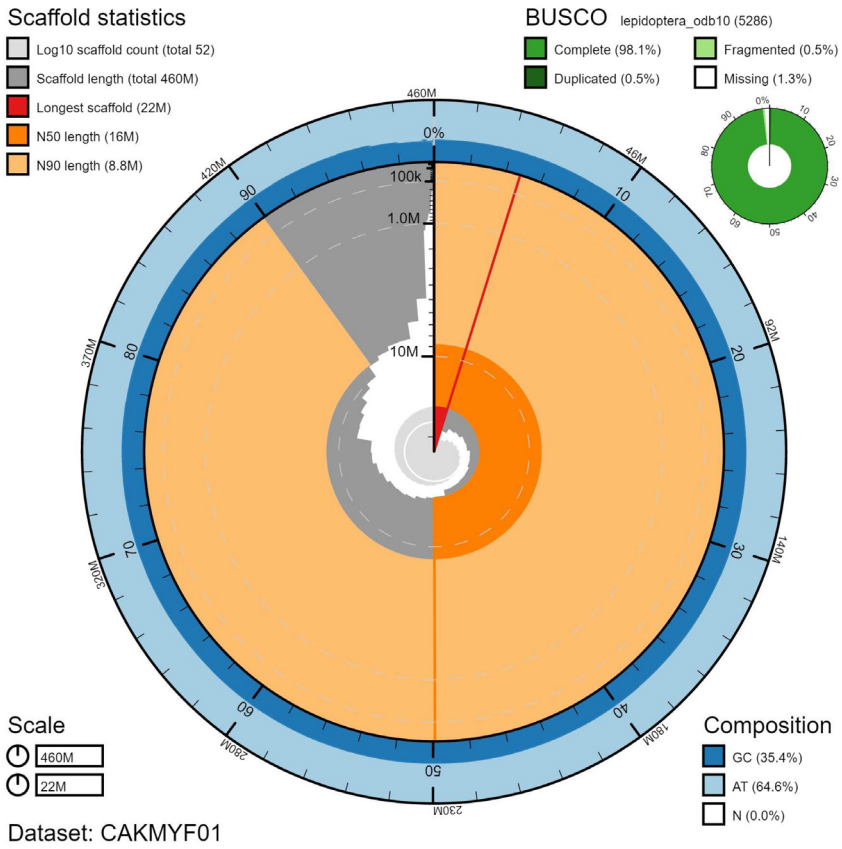

Figure 2. Genome assembly of *Gymnoscelis rufifasciata*, ilGymRufi1.1: metrics. The BlobToolKit Snailplot shows N50 metrics and BUSCO gene completeness. The main plot is divided into 1,000 size-ordered bins around the circumference with each bin representing 0.1% of the 462,009,964 bp assembly. The distribution of scaffold lengths is shown in dark grey with the plot radius scaled to the longest scaffold present in the assembly (22,195,239 bp, shown in red). Orange and pale-orange arcs show the N50 and N90 scaffold ome lengths (15,602,950 and 8,767,853 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/ilGymRufi1.1/dataset/CAKMYF01/snail.

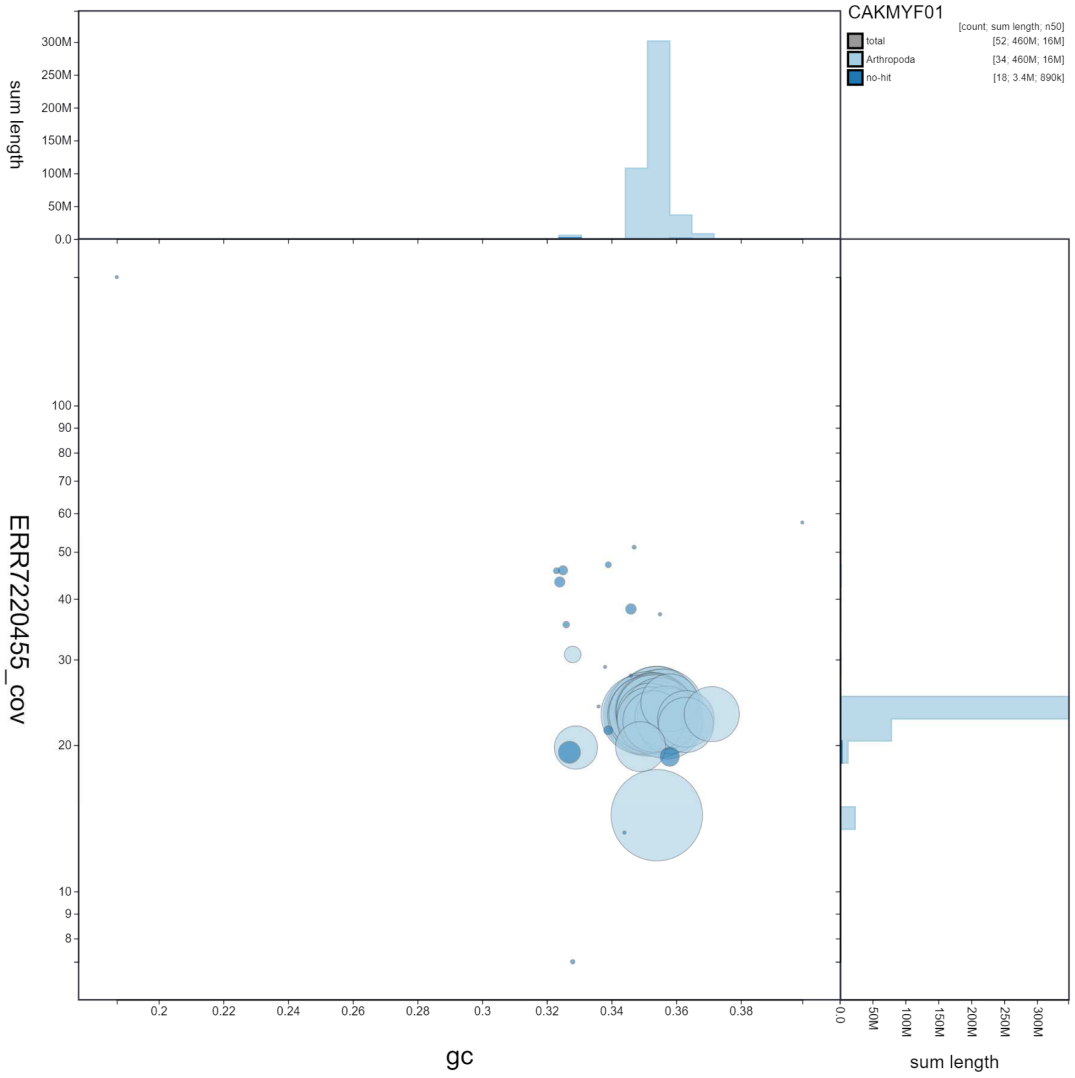

Figure 3. Genome assembly of *Gymnoscelis rufifasciata*, ilGymRufi1.1: GC coverage. BlobToolKit GC-coverage plot. Scaffolds are coloured by phylum. Circles are sized in proportion to scaffold length Histograms show the distribution of scaffold length sum along each axis. An interactive version of this figure is available at https://blobtoolkit.genomehubs.org/view/ilGymRufi1.1/dataset/CAKMYF01/blob.

performed by aligning 10X Genomics read data to the assembly with longranger align, calling variants with freebayes (Garrison & Marth, 2012). The assembly was then scaffolded with Hi-C data (Rao *et al.*, 2014) using SALSA2 (Ghurye *et al.*, 2019). The assembly was checked for contamination as described previously (Howe *et al.*, 2021). Manual curation (Howe *et al.*, 2021) was performed using HiGlass (Kerpedjiev *et al.*, 2018) and Pretext. The mitochondrial genome was assembled using MitoHiFi (Uliano-Silva *et al.*, 2021), which performs annotation using MitoFinder (Allio *et al.*, 2020). The genome was analysed and BUSCO scores generated within the BlobToolKit environment (Challis *et al.*, 2020). Table 3 contains a list of all software tool versions used, where appropriate.

Ethics/compliance issues

The materials that have contributed to this genome note have been supplied by a Darwin Tree of Life Partner. The submission of materials by a Darwin Tree of Life Partner is subject to the Darwin Tree of Life Project Sampling Code of Practice. By agreeing with and signing up to the Sampling Code of Practice, the Darwin Tree of Life Partner agrees they will meet the legal and ethical requirements and standards set out within this document in respect of all samples acquired for, and supplied to, the Darwin Tree of Life Project. 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

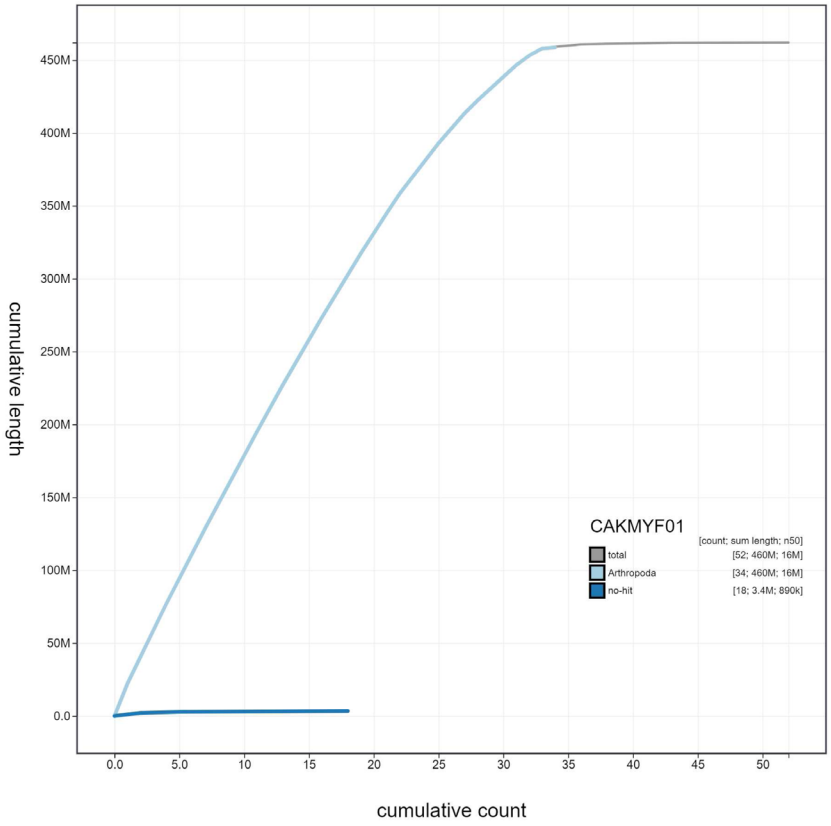

Figure 4. Genome assembly of *Gymnoscelis rufifasciata*, ilGymRufi1.1: cumulative sequence. BlobToolKit cumulative sequence plot. The grey line shows cumulative length for all scaffolds. Coloured lines show cumulative lengths of scaffolds assigned to each phylum using the buscogenes taxrule. An interactive version of this figure is available at https://blobtoolkit.genomehubs.org/view/ilGymRufi1.1/dataset/CAKMYF01/cumulative.

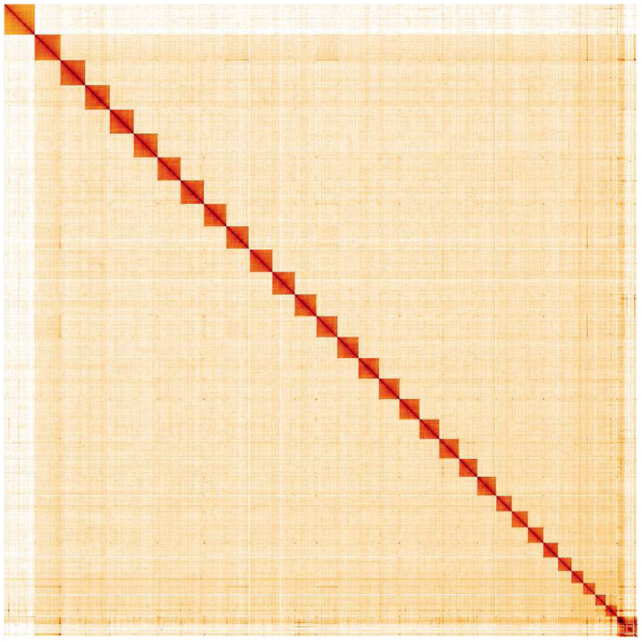

Figure 5. Genome assembly of *Gymnoscelis rufifasciata*, ilGymRufi1.1: Hi-C contact map. Hi-C contact map of the ilGymRufi1.1 assembly, visualised in HiGlass. Chromosomes are shown in order of size from left to right and top to bottom. An interactive version of this map is available here.

Table 2. Chromosomal pseudomolecules in the genome assembly of *Gymnoscelis rufifasciata*, ilGymRufi1.1.

INSDC accession	Chromosome	Size (Mb)	GC%
OV815302.1	1	18.83	35.1
OV815303.1	2	18.12	35.4
OV815304.1	3	17.86	35.4
OV815305.1	4	17.46	35.5
OV815306.1	5	17.34	35.0
OV815307.1	6	16.92	35.3
OV815308.1	7	16.84	35.6
OV815309.1	8	16.61	35.1
OV815310.1	9	16.56	34.9
OV815311.1	10	16.46	35.3
OV815312.1	11	16.42	35.1
OV815313.1	12	15.73	35.4
OV815314.1	13	15.60	35.3
OV815315.1	14	15.38	35.1
OV815316.1	15	15.01	35.3
OV815317.1	16	14.84	35.5
OV815318.1	17	14.76	35.3
OV815320.1	18	14.39	35.7
OV815321.1	19	14.02	35.4
OV815322.1	20	13.58	35.5
OV815323.1	21	13.48	35.5
OV815324.1	22	11.75	35.2
OV815325.1	23	11.66	35.2
OV815326.1	24	11.20	35.8
OV815327.1	25	10.47	35.7
OV815328.1	26	9.95	35.3
OV815329.1	27	8.77	35.8
OV815330.1	28	8.40	36.3
OV815331.1	29	8.01	36.3
OV815332.1	30	7.98	37.1
OV815319.1	W	6.55	34.9
OV815301.1	Z	22.20	35.4
OV815333.1	MT	0.02	18.9
-	Unplaced	8.85	33.2

Table 3. Software tools used.

Software tool	Version	Source
Hifiasm	0.15.3	Cheng *et al.*, 2021
purge_dups	1.2.3	Guan *et al.*, 2020
SALSA	2.2	Ghurye *et al.*, 2019
longranger align	2.2.2	https://support.10xgenomics.com/genome-exome/software/pipelines/latest/advanced/other-pipelines
freebayes	1.3.1-17-gaa2ace8	Garrison & Marth, 2012
MitoHiFi	2.0	Uliano-Silva *et al.*, 2021
HiGlass	1.11.6	Kerpedjiev *et al.*, 2018
PretextView	0.2.x	https://github.com/wtsi-hpag/PretextView
BlobToolKit	3.0.5	Challis *et al.*, 2020

(operating as the Wellcome Sanger Institute), and in some circumstances other Darwin Tree of Life collaborators.

Data availability

European Nucleotide Archive: Gymnoscelis rufifasciata (double-striped pug). Accession number PRJEB48374; https://identifiers.org/ena.embl/PRJEB48374.

The genome sequence is released openly for reuse. The *G. rufifasciata* 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 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.5746938.

Members of the Darwin Tree of Life Barcoding collective are listed here: https://doi.org/10.5281/zenodo.5744972.

Members of the Wellcome Sanger Institute Tree of Life programme are listed here: https://doi.org/10.5281/zenodo.6125027.

Members of Wellcome Sanger Institute Scientific Operations: DNA Pipelines collective are listed here: https://doi.org/10.5281/zenodo.5746904.

Members of the Tree of Life Core Informatics collective are listed here: https://doi.org/10.5281/zenodo.6125046.

Members of the Darwin Tree of Life Consortium are listed here: https://doi.org/10.5281/zenodo.5638618.

References

Allio R, Schomaker-Bastos A, Romiguier J, *et al.*: **MitoFinder: Efficient Automated Large-Scale Extraction of Mitogenomic Data in Target Enrichment Phylogenomics.** *Mol Ecol Resour.* 2020; **20**(4): 892–905.
PubMed Abstract | Publisher Full Text | Free Full Text

Challis R, Richards E, Rajan J, *et al.*: **BlobToolKit - Interactive Quality Assessment of Genome Assemblies.** *G3 (Bethesda).* 2020; **10**(4): 1361–74.
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): 170–75.
PubMed Abstract | Publisher Full Text | Free Full Text

Fox R, Dennis EB, Harrower CA, *et al.*: **The State of Britain's Larger Moths 2021**. 2021 March, 44.
Reference Source

Garrison E, Marth G: **Haplotype-Based Variant Detection from Short-Read Sequencing.** arXiv: 1207.3907, 2012.
Publisher Full Text

Ghurye J, Rhie A, Walenz BP, *et al.*: **Integrating Hi-C Links with Assembly Graphs for Chromosome-Scale Assembly.** *PLoS Comput Biol.* 2019; **15**(8): e1007273.
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): 2896–98.
PubMed Abstract | Publisher Full Text | Free Full Text

Howe K, Chow W, Collins J, *et al.*: **Significantly Improving the Quality of Genome Assemblies through Curation.** *GigaScience.* 2021; **10**(1): giaa153.
PubMed Abstract | Publisher Full Text | Free Full Text

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): 4647–54.
PubMed Abstract | Publisher Full Text | Free Full Text

Randle Z, Evans-Hill LJ, Parsons MS, *et al.*: **Atlas of Britain and Ireland's Larger Moths**. Pisces Publications, Newbury. 2019.
Reference Source

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): 1665–80.
PubMed Abstract | Publisher Full Text | Free Full Text

Riley A, Prior G: **British and Irish Pug Moths – a Guide to their Identification and Biology**. 2003.
Publisher Full Text

Skinner B, Wilson D: **Colour Identification Guide to the Moths of the British Isles**. 2009.
Publisher Full Text

Uliano-Silva M, Nunes JGF, Krasheninnikova K, *et al.*: **marcelauliano/MitoHiFi: mitohifi_v2.0**. 2021.
Publisher Full Text

Open Peer Review

Current Peer Review Status: ✔ ✔ ✔

Version 1

Reviewer Report 07 August 2023

https://doi.org/10.21956/wellcomeopenres.19690.r62475

✔ **Christopher B Cunningham**
University of Georgia, Athens, Georgia, USA

I think this is a good general resource for the genomics community. The methods seem sound. The article needs a few more details to be reasonably reproducible. Justification or rationale for the work is completely absent.

Abstract. There is no justification within the abstract, just results. At least a sentence that this was part of a large survey of insects is needed. And one sentence for its possible value.

Keywords. Usually, keywords are required to be in alphabetical order, but check with journal policies.

Background. There is no motivation for this study in the introduction. Nothing wrong is said, but the read has no idea why the work was done. Its perfectly fine that the work was done in the context of a large survey, but tell the reader that. Is its expansion north in the UK a problem or likely to become one? The paper is not about its ecology or morphology, but about its genetics. That is not mentioned once.

Genome Sequence Report. The BUSCO reference should be in the methods, not results section.

Figure 2/3/4/5 titles. These should be informative and tell the reader what you would like them to understand about the data; e.g., Fig 3. Little contamination was found in the final genome assembly.

Figure 4. has very little information and can easily be summarized as a sentence in the results. The same is true of Fig 5

Methods. I assume that the Hi-C library was constructed according to the manufacturer's specifications?

All parameter setting if they were different from default of each piece of bioinformatic software

needs to be specified. Just put a sentence to begin that everything was default unless specified otherwise. What assessment was done to ensure that default was adequate?

Table 3 is of little to no value and contains much duplicated information. Just add the version numbers of software used inline at the appropriate places in the methods section.

Is the Ethics statement needed? It just says the provider of the sample should meet some standard, but does not actually say that they did in this case. Either drop the statement because there are no legal requirements for insect research beyond basic humane treatment or actually say the standard was met.

Add something to the report that contextualize the research for the reader. Is this species the first of a taxon to be assembled, is it an outgroup to some established model species, is it now possible to investigate some interesting aspect of the species biology, etc? This does not to be extensive or highly directed, but there is currently no justification of this work outside of the Data Availability Statement. Even Data Note should contain rationale for the work.

Is the rationale for creating the dataset(s) clearly described?
No

Are the protocols appropriate and is the work technically sound?
Yes

Are sufficient details of methods and materials provided to allow replication by others?
No

Are the datasets clearly presented in a useable and accessible format?
Yes

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Genome assembly & Annotation, Insects, Molecular Genetics

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

Reviewer Report 04 August 2023

https://doi.org/10.21956/wellcomeopenres.19690.r60442

Simon H Martin

Institute of Evolutionary Biology, University of Edinburgh, Edinburgh, UK

This note describes a highly complete chromosomal genome assembly for the moth Gymnoscelis rufifasciata.

The review report from asks whether the rationale for creating the dataset is clearly described. Although no justification is provided within this note, the justification for the DToL project is clearly described elsewhere, and the assemblies, including the one reported here, are already being used as resources in other studies, demonstrating their value.

The introduction is entirely focused on the UK population. This is understandable given the origin of the sequenced specimen, but it would be helpful to a least mention the full extent of the species' range.

One important issue to address is that the abstract says the assembly length is 352 Mb but the main text and figures show that it is 462 Mb.

Is the rationale for creating the dataset(s) clearly described?
Partly

Are the protocols appropriate and is the work technically sound?
Yes

Are sufficient details of methods and materials provided to allow replication by others?
Yes

Are the datasets clearly presented in a useable and accessible format?
Yes

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Evolutionary genomics, population genetics

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

Reviewer Report 02 August 2023

https://doi.org/10.21956/wellcomeopenres.19690.r62472

Henry North
University of Cambridge, Cambridge, UK

Boyes *et al.* present a reference genome of *Gymnoscelis rufifasciata*, including both sex chromosomes and the mitochondrial genome. The quality of this assembly is high by all measures of completeness and contiguity. The methods are reported clearly and in detail.

These data will prove valuable in broader research efforts aimed at understanding chromosomal evolution during the radiation of the Lepidoptera, which requires assemblies from species that span the rich diversity of this order. Pugs such as this species are a familiar occurrence in British light traps and, as the authors point out, *G. rufifasciata* is noteworthy for its increased abundance over the past half-century. This assembly permits an investigation of relative population size change deeper in time — when we know what the climate looked like but have no census population size estimates — using coalescent-based demographic inference. The assembly will also be of use in research aimed at understanding the evolution of lepidopteran polyphagy using comparative genomics approaches.

This publicly available data is a testament to the efforts of the Tree of Life team, and to Douglas' contribution to the study and public appreciation of the full diversity of the Lepidoptera.

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: Population genomics, speciation, adaptation

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