



DATA NOTE

The genome sequence of the grizzled skipper, *Pyrgus malvae* (Linnaeus, 1758) [version 1; peer review: 2 approved, 1 approved with reservations]

Alex Hayward¹, Roger Vila ², Konrad Lohse ³, Dominik Laetsch ³, 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, Darwin Tree of Life Consortium

¹College of Life and Environmental Sciences, Department of Biosciences, University of Exeter, Penryn, UK

²Institut de Biologia Evolutiva (CSIC - Universitat Pompeu Fabra), Barcelona, Spain

³Institute of Evolutionary Biology, University of Edinburgh, Edinburgh, UK

V1 First published: 29 Mar 2022, 7:114
<https://doi.org/10.12688/wellcomeopenres.17806.1>
 Latest published: 29 Mar 2022, 7:114
<https://doi.org/10.12688/wellcomeopenres.17806.1>

Abstract

We present a genome assembly from an individual male *Pyrgus malvae* (the grizzled skipper; Arthropoda; Insecta; Lepidoptera; Hesperiiidae). The genome sequence is 725 megabases in span. The majority (99.97%) of the assembly is scaffolded into 31 chromosomal pseudomolecules, with the Z sex chromosome assembled.

Keywords

Pyrgus malvae, grizzled skipper, genome sequence, chromosomal, Lepidoptera



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

Open Peer Review

Approval Status ? ✓ ✓

	1	2	3
version 1	?	✓	✓
29 Mar 2022	view	view	view

1. **Jeffrey M. Marcus** , University of Manitoba, Winnipeg, Canada
2. **Marko Mutanen** , University of Oulu, Oulu, Finland
3. **Pável Matos-Maraví** , Institute of Entomology, České Budějovice, České Budějovice, Czech Republic

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: **Hayward A:** Investigation, Resources; **Vila R:** Investigation, Resources, Writing – Original Draft Preparation, Writing – Review & Editing; **Lohse K:** Investigation, Resources, Writing – Original Draft Preparation, Writing – Review & Editing; **Laetsch D:** Investigation, Resources, 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) and the Darwin Tree of Life Discretionary Award (218328). Fieldwork was supported by an ERC grant (ModelGenom Land 757648) to KL. AH is supported by a Biotechnology and Biological Sciences Research Council (BBSRC) David Phillips Fellowship (BB/N020146/1). RV was supported by the Spanish government through grant PID2019-107078GB-I00/ MCIN/AEI/ 10.13039/501100011033.

The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Copyright: © 2022 Hayward A *et al.* This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

How to cite this article: Hayward A, Vila R, Lohse K *et al.* **The genome sequence of the grizzled skipper, *Pyrgus malvae* (Linnaeus, 1758) [version 1; peer review: 2 approved, 1 approved with reservations]** Wellcome Open Research 2022, 7:114 <https://doi.org/10.12688/wellcomeopenres.17806.1>

First published: 29 Mar 2022, 7:114 <https://doi.org/10.12688/wellcomeopenres.17806.1>

Species taxonomy

Eukaryota; Metazoa; Ecdysozoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Lepidoptera; Glossata; Ditrysia; Papilionoidea; Hesperidae; Pyrginae; *Pyrgus*; *Pyrgus malvae* (Linnaeus, 1758) (NCBI:txid218760).

Background

The grizzled skipper, *Pyrgus malvae*, is a small butterfly, characteristic of chalk downland and woodland clearings, and other grassland habitats. Not to be confused with the term ‘grisly’ (i.e. extremely unpleasant or gruesome), *P. malvae* gets its common name from the tufts of long grey hair that cover its body and inner wings. Its wings bear a striking black and white checkerboard pattern, with alternating black and white stripes on the wing fringes and antennae. Notoriously difficult to follow, *P. malvae* has a fast and darting, low flight pattern. *Pyrgus malvae* is found throughout Europe, except for northern Scandinavia, several Mediterranean Islands and Iberia, southern France and Italy (where it is replaced by its sister species *P. malvoides*), with a range that extends eastwards across temperate Asia to Northern China and Korea (Tolman & Lewington, 2008). In the UK the species is found mainly in central and southern England, with a patchy distribution in Wales and the southwest. *P. malvae* typically exists in small populations (<100 adults) that are thought to form metapopulations across its range (Asher *et al.*, 2001).

In the UK, *P. malvae* typically emerges in April and flies until June, although the date of first emergence is advancing, and in warm years may occur as early as March. It is univoltine in northern Europe and at higher altitudes, but is bivoltine elsewhere, and in the north it may be bivoltine when weather conditions are particularly favourable (Asher *et al.*, 2001).

Pyrgus malvae larvae feed on a variety of host plants in the Rosaceae family, particularly agrimony (*Agrimonia eupatoria*), creeping cinquefoil (*Potentilla reptans*) and wild strawberry (*Fragaria vesca*) (Asher *et al.*, 2001). When fully-grown, the larva constructs a cocoon at the base of low vegetation, where it overwinters as a pupa. Adults feed on a wide variety of nectar sources, including Bird’s foot trefoil (*Lotus corniculatus*), bugle (*Ajuga reptans*), buttercup (*Ranunculus* species), daisy (*Bellis perennis*), and dandelion (*Taraxacum officinale*). Males are territorial, and exhibit either perching or patrolling behaviour according to habitat type (Brereton *et al.*, 1998), and have two scent organs: the forewing costal fold and tibial tufts composed of specialised setae on the hind leg, which appear to be used to waft pheromones towards the female during courtship (Hernández-Roldán *et al.*, 2014). Eggs are laid singly on the leaf underside of larval host plants, with the majority deposited on short vegetation in locations with a favourably warm microclimate and/or elevated nutritional content (Brereton *et al.*, 1998).

Populations of *P. malvae* in the UK have declined markedly in the twentieth century (Brereton *et al.*, 1998) and the species is a conservation priority in the UK (Brig, 2007). Encouragingly, *P. malvae* appears to be positively associated with grazed

vegetation, and implementing grazing in habitat restoration regimes may offer a means to help reverse population declines (Wallis De Vries & Raemakers, 2001). At European level this species is listed as Least Concern in the IUCN Red List (Van Swaay *et al.*, 2010). *Pyrgus malvae* has been reported as having 33 (Bigger, 1960; England) and 31 (Federley, 1938; Finland) chromosome pairs. The assembly described herein contains 31 chromosome pairs.

Genome sequence report

The genome was sequenced from a single male *P. malvae* (Figure 1) collected from Suatu, Cluj County, Romania (latitude 46.7648, longitude 23.9845). A total of 69-fold coverage in Pacific Biosciences single-molecule circular consensus (HiFi) long reads and 49-fold coverage in 10X Genomics read clouds were generated. Primary assembly contigs were scaffolded with chromosome conformation Hi-C data. Manual assembly curation corrected 2 missing/misjoins and removed 1 haplotypic duplication, reducing the assembly length by 0.01% and the scaffold number by 7.69%.

The final assembly has a total length of 725 Mb in 36 sequence scaffolds with a scaffold N50 of 27.0 Mb (Table 1). The majority, 99.97%, of assembly sequence was assigned to 31 chromosomal-level scaffolds, representing 30 autosomes (numbered by sequence length), and the Z sex chromosome (Figure 2–Figure 5; Table 2). The assembly has a BUSCO v5.1.2 (Manni *et al.*, 2021) completeness of 98.8% (single 98.3%, duplicated 0.4%) using the lepidoptera_odb10 reference set (n=5286). While not fully phased, the assembly deposited is of one haplotype. Contigs corresponding to the second haplotype have also been deposited.

Genome annotation report

The ilPyrMalv3.1 genome has been annotated using the Ensembl rapid annotation pipeline (Table 1; https://rapid.ensembl.org/Pyrgus_malvae_GCA_911387765.1/). The resulting annotation includes 23,484 transcribed mRNAs from 12,096 protein-coding and 2,976 non-coding genes. There are 1.66 coding transcripts per gene and 7.83 exons per transcript.

Methods

Sample acquisition and nucleic acid extraction

A male *P. malvae* specimen (ilPyrMalv3, male, genome assembly) was collected from Suatu, Cluj County, Romania (latitude 46.7648, longitude 23.9845) using a net by Konrad Lohse, Alex Hayward Dominik Laetsch and Roger Vila, who also identified the sample. A further two specimens (ilPyrMalv2, unknown sex, Hi-C; ilPyrMalv1, RNA-Seq) were collected from Baci, Cluj County, Romania (latitude 46.8, longitude 23.5) using a net and were identified by the same team. All samples were snap-frozen at -80°C.

DNA was extracted from the whole organism of ilPyrMalv3 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. RNA (from the whole organism of ilPyrMalv1) was extracted



Figure 1. Fore and hind wings of the *Pyrgus malvae* specimen from which the genome was sequenced. Top: Dorsal (left) and ventral (right) surface view of wings from specimen RO_PM_973 (ilPyrMalv3) from Suatu, Romania, used to generate Pacific Biosciences and 10X genomics data. Bottom: Dorsal (left) and ventral (right) surface view of wings from specimen RO_PM_838 (ilPyrMalv2) from Cluj-Napoca, Romania, used to generate Hi-C data.

Table 1. Genome data for *Pyrgus malvae*, ilPyrMalv3.1.

Project accession data	
Assembly identifier	ilPyrMalv3.1
Species	<i>Pyrgus malvae</i>
Specimen	ilPyrMalv3 (genome assembly); ilPyrMalv2 (Hi-C); ilPyrMalv1 (RNA-Seq)
NCBI taxonomy ID	NCBI:txid111923
BioProject	PRJEB46857
BioSample ID	SAMEA7523296
Isolate information	Male, whole organism (ilPyrMalv3); unknown sex, whole organisms (ilPyrMalv1, ilPyrMalv2)
Raw data accessions	
PacificBiosciences SEQUEL II	ERR6606794-ERR6606796
10X Genomics Illumina	ERR6363273-ERR6363276
Hi-C Illumina	ERR6363278
Illumina polyA RNA-Seq	ERR6363277
Genome assembly	
Assembly accession	GCA_911387765.1
Accession of alternate haplotype	GCA_911387725.2
Span (Mb)	725
Number of contigs	41
Contig N50 length (Mb)	26.0
Number of scaffolds	36
Scaffold N50 length (Mb)	27.0
Longest scaffold (Mb)	33.2
BUSCO* genome score	C:98.8%[S:98.3%,D:0.4%],F:0.2%,M:1.0%,n:5286
Genome annotation	
Number of protein-coding genes	12,096
Average length of coding sequence (bp)	1,534.63
Average number of exons per transcript	7.83
Average exon size (bp)	207.85
Average intron size (bp)	2,914.74

*BUSCO scores based on the lepidoptera_odb10 BUSCO set using v5.1.2. 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/ilPyrMalv3.1/dataset/CAJVQT01/busco>.

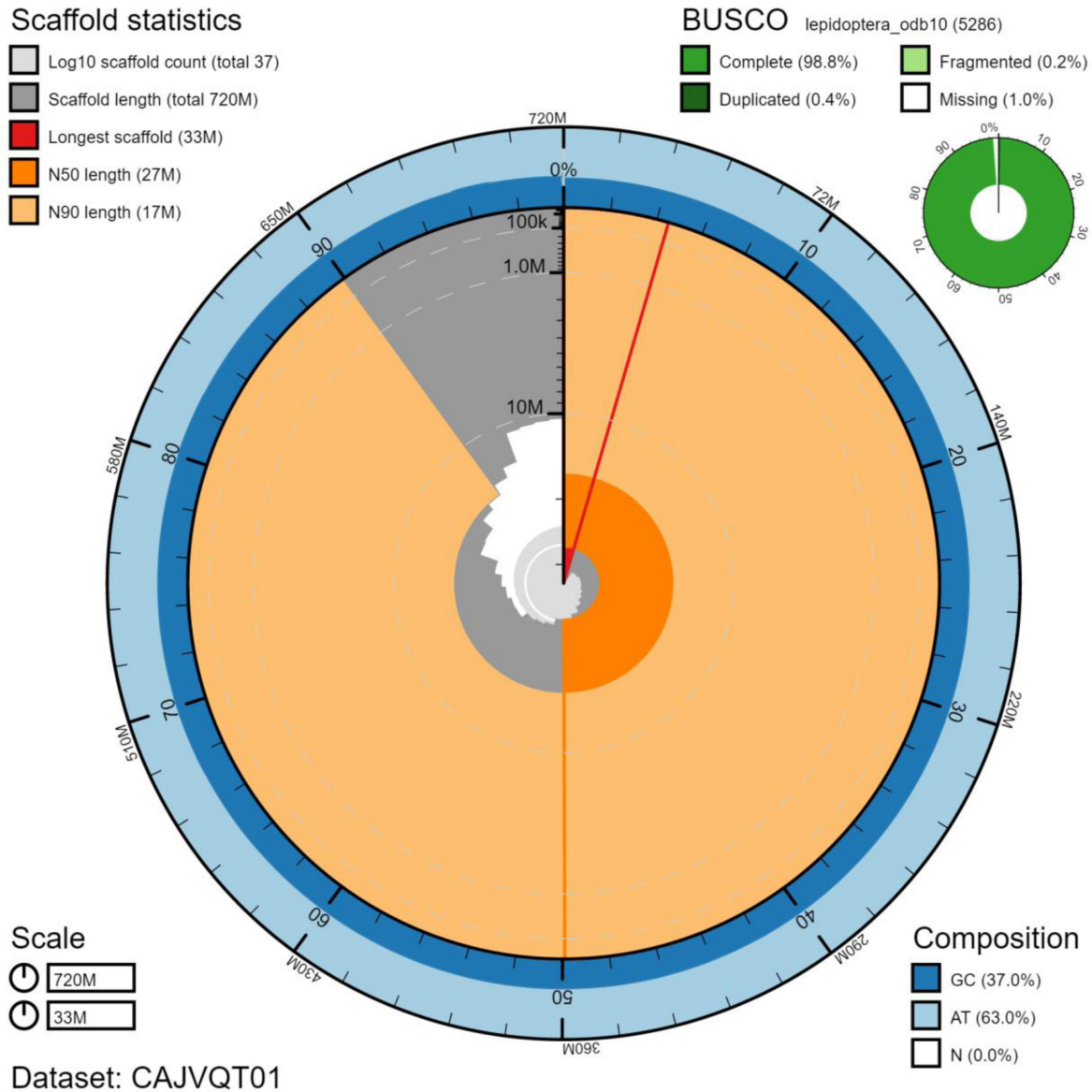


Figure 2. Genome assembly of *Pyrgus malvae*, iIPyrMalv3.1: metrics. The BlobToolKit Snailplot shows N50 metrics and BUSCO gene completeness. The main plot is divided into 1,000 size-ordered bins around the circumference with each bin representing 0.1% of the 724,649,524 bp assembly. The distribution of chromosome lengths is shown in dark grey with the plot radius scaled to the longest chromosome present in the assembly (33,217,309 bp, shown in red). Orange and pale-orange arcs show the N50 and N90 chromosome lengths (26,976,370 and 16,663,010 bp), respectively. The pale grey spiral shows the cumulative chromosome 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/iIPyrMalv3.1/dataset/CAJVQT01/snail>.

in the Tree of Life Laboratory at the WSI using TRIzol, according to the manufacturer's instructions. RNA was then eluted in 50 μ l RNase-free water and its concentration RNA assessed using a Nanodrop spectrophotometer and Qubit Fluorometer using the Qubit RNA Broad-Range (BR) Assay kit. Analysis of the integrity of the RNA was done using Agilent RNA 6000 Pico Kit and Eukaryotic Total RNA assay.

Sequencing

Pacific Biosciences HiFi circular consensus and 10X Genomics read cloud 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 at the WSI on Pacific Biosciences

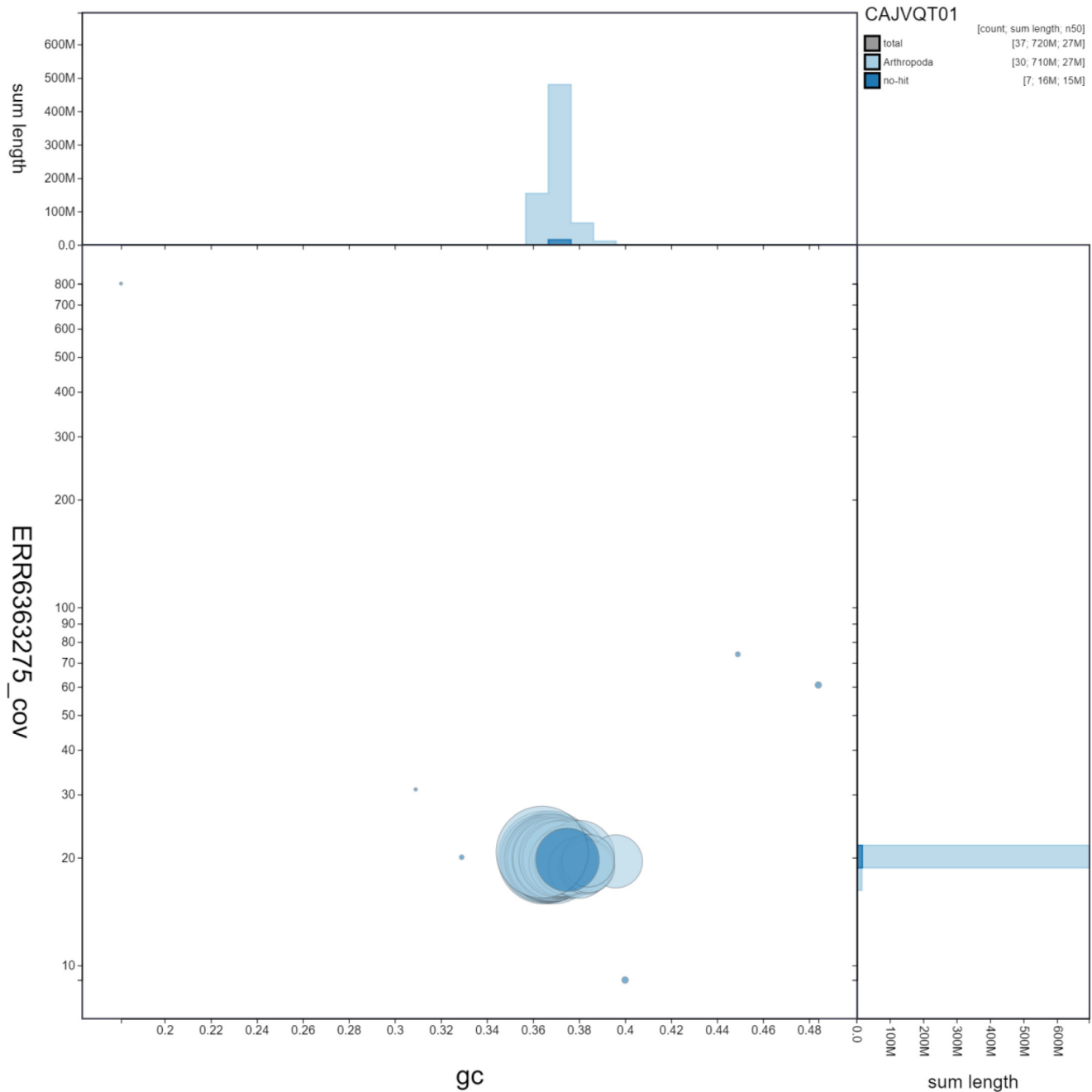


Figure 3. Genome assembly of *Pyrgus malvae*, ilPyrMalv3.1: GC coverage. BlobToolKit GC-coverage plot. Scaffolds are coloured by phylum. Circles are sized in proportion to scaffold length. Histograms show the distribution of scaffold length sum along each axis. An interactive version of this figure is available at <https://blobtoolkit.genomehubs.org/view/ilPyrMalv3.1/dataset/CAJVQT01/blob>.

SEQUEL II (HiFi), Illumina HiSeq X (10X) and Illumina HiSeq 4000 (RNA-Seq) instruments. Hi-C data were also generated from whole organism tissue of ilPyrMalv2 using the Arima v1 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 performed by aligning 10X Genomics read data to the

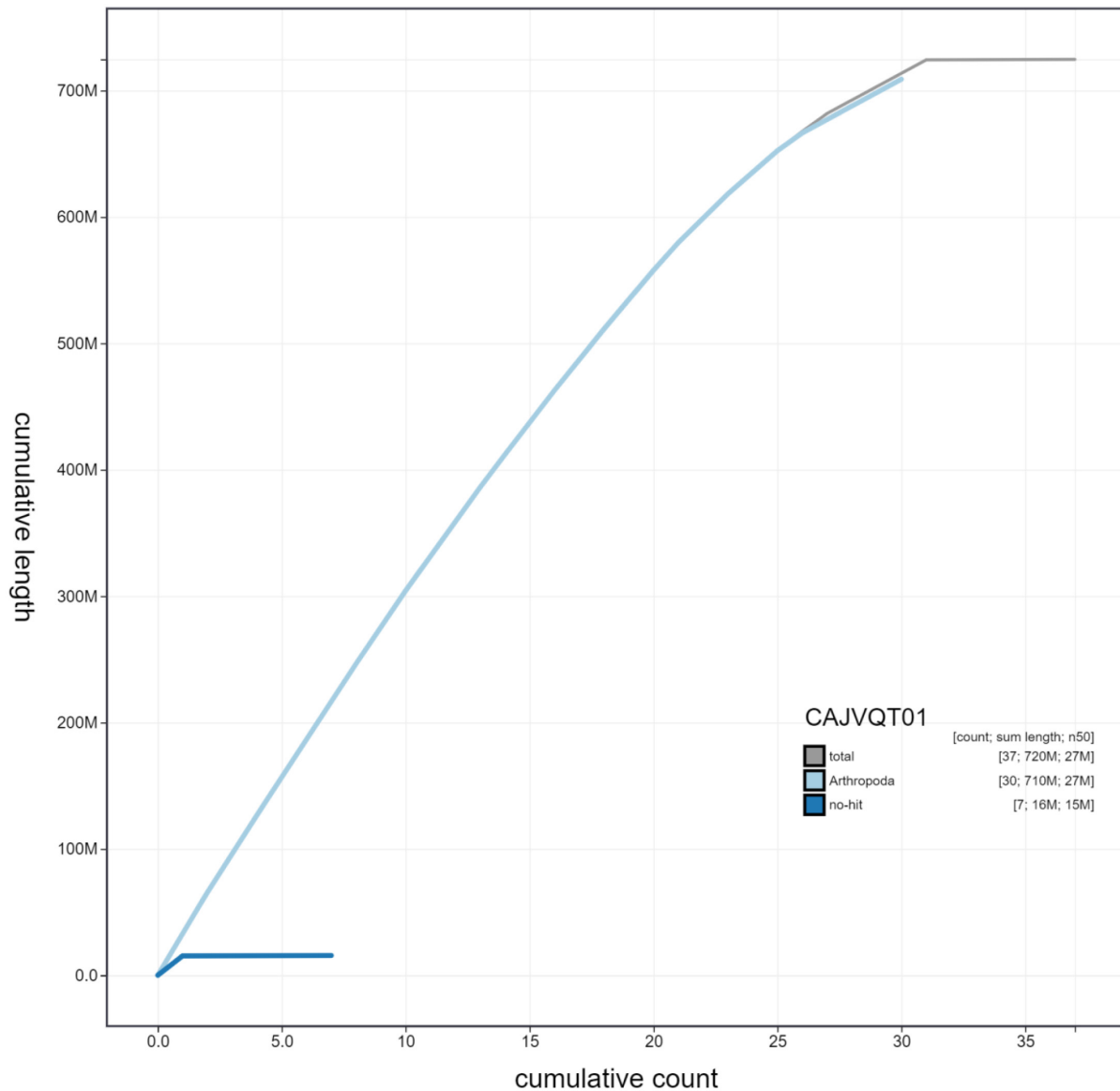


Figure 4. Genome assembly of *Pyrgus malvae*, ilPyrMalv3.1: cumulative sequence. BlobToolKit cumulative sequence plot. The grey line shows cumulative length for all scaffolds. Coloured lines show cumulative lengths of scaffolds assigned to each phylum using the buscogenes taxrule. An interactive version of this figure is available at <https://blobtoolkit.genomehubs.org/view/ilPyrMalv3.1/dataset/CAJVQT01/cumulative>.

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 and corrected using the gEVAL system (Chow *et al.*, 2016) as described previously (Howe *et al.*, 2021). Manual curation (Howe *et al.*, 2021) was performed using gEVAL,

HiGlass (Kerpedjiev *et al.*, 2018) and Pretext. The mitochondrial genome was assembled using MitoHiFi (Uliano-Silva *et al.*, 2021), which performed 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.

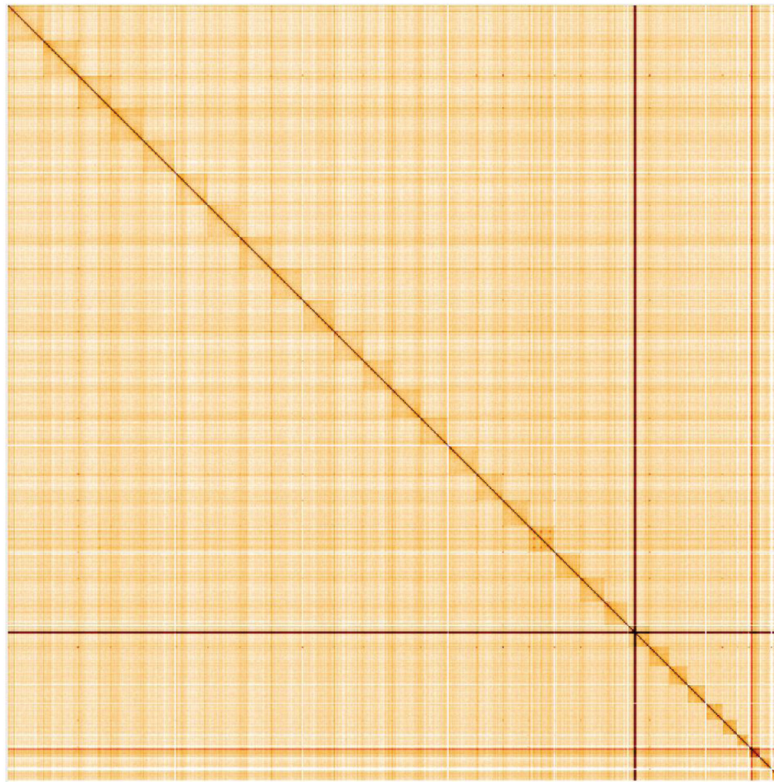


Figure 5. Genome assembly of *Pyrgus malvae*, ilPyrMalv3.1: Hi-C contact map. Hi-C contact map of the ilPyrMalv3.1 assembly, visualised in HiGlass. Chromosomes are shown in size order from left to right and top to bottom. The interactive Hi-C map can be viewed [here](#).

Table 2. Chromosomal pseudomolecules in the genome assembly of *Pyrgus malvae*, ilPyrMalv3.1.

INSDC accession	Chromosome	Size (Mb)	GC%
OU426946.1	1	33.22	36.5
OU426948.1	2	30.61	36.9
OU426949.1	3	30.48	36.8
OU426950.1	4	30.18	36.6
OU426951.1	5	30.06	36.5
OU426952.1	6	29.97	36.8
OU426953.1	7	29.61	36.7
OU426954.1	8	29.23	36.7
OU426955.1	9	28.53	36.8
OU426956.1	10	27.64	36.6
OU426957.1	11	27.02	36.7
OU426958.1	12	26.98	36.9
OU426959.1	13	25.95	36.9
OU426960.1	14	25.59	36.9
OU426961.1	15	25.29	36.9

INSDC accession	Chromosome	Size (Mb)	GC%
OU426962.1	16	24.25	37.1
OU426963.1	17	24.17	37.5
OU426964.1	18	23.61	37.2
OU426965.1	19	22.98	37.1
OU426966.1	20	21.93	37.2
OU426967.1	21	19.36	38.0
OU426968.1	22	19.17	37.5
OU426969.1	23	17.50	37.3
OU426970.1	24	16.66	37.5
OU426971.1	25	15.31	37.5
OU426972.1	26	14.10	38.0
OU426974.1	27	10.67	39.6
OU426973.1	28	10.82	38.3
OU426975.1	29	10.54	38.4
OU426976.1	30	10.42	38.4
OU426947.1	Z	32.47	36.4
OU426977.1	MT	0.02	18.4
-	Unplaced	0.34	41.9

Table 3. Software tools used.

Software tool	Version	Source
Hifiasm	0.15.1	Cheng et al., 2021
purge_dups	1.2.3	Guan et al., 2020
SALSA2	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	Uliano-Silva et al., 2021
gEVAL	N/A	Chow et al., 2016
HiGlass	1.11.6	Kerpedjiev et al., 2018
PretextView	0.2.x	https://github.com/wtsi-hpag/PretextView
BlobToolKit	2.6.4	Challis et al., 2020

Genome annotation

The Ensembl gene annotation system ([Aken et al., 2016](#)) was used to generate annotation for the *Pyrgus malvae* assembly ([GCA_911387765.1](#)). Annotation was created primarily through alignment of transcriptomic data to the genome, with gap filling via protein-to-genome alignments of a select set of proteins from UniProt ([UniProt Consortium, 2019](#)).

Data availability

European Nucleotide Archive: *Pyrgus malvae* (grizzled skipper). Accession number [PRJEB45665](#); <https://identifiers.org/ena.embl/PRJEB45665>.

The genome sequence is released openly for reuse. The *P. malvae* genome sequencing initiative is part of the [Darwin Tree of Life](#) (DTOL) project. All raw sequence data and the assembly have been deposited in INSDC databases. Raw data and assembly accession identifiers are reported in [Table 1](#).

Author information

Members of the 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

- 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](#)
- Asher J, Warren M, Fox R, et al.: **The Millennium Atlas of Butterflies in Britain and Ireland**. Oxford University Press. 2001. [Reference Source](#)
- Bigger TRL: **Chromosome Numbers of Lepidoptera. Part I**. *Entomologist's Gazette*. 1960; **11**: 149–52.
- Brereton TM, Bourn NAD, Warren MS: **Grizzled Skipper Action Plan**. *Butterfly Conserv*. 1998. [Publisher Full Text](#)
- Brig: **Report on the Species and Habitat Review (Report by the Biodiversity Reporting and Information Group (BRIG) to the UK Standing Committee)**. JNCC Peterborough. 2007. [Reference Source](#)
- 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](#)
- Chow W, Brugger K, Caccamo M, et al.: **gEVAL — a Web-Based Browser for Evaluating Genome Assemblies**. *Bioinformatics*. 2016; **32**(16): 2508–10. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Federley H: **Chromosomenzahlen Finnland-Discher Lepidopteren**. *Hereditas*.

1938; **24**(4): 397–464.

[Publisher Full Text](#)

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](#)

Hernández-Roldán JL, Bofill R, Dapporto L, *et al.*: **Morphological and Chemical Analysis of Male Scent Organs in the Butterfly Genus *Pyrgus* (Lepidoptera: Hesperiiidae)**. *Org Divers Evol.* 2014; **14**(3): 269–78.

[Publisher Full Text](#)

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

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

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, 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): 4647–54.

[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): 1665–80.

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

Tolman T, Lewington R: **Collins Butterfly Guide**. HarperCollins UK. 2008.

[Reference Source](#)

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

[Publisher Full Text](#)

UniProt Consortium: **UniProt: A Worldwide Hub of Protein Knowledge**. *Nucleic Acids Res.* 2019; **47**(D1): D506–15.

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

Van Swaay C, Wynhoff I, Verovnik R, *et al.*: **IUCN Red List of Threatened Species: *Pyrgus Malvae***. *IUCN Red List of Threatened Species.* 2010.

[Reference Source](#)

Wallis De Vries MF, Raemakers I: **Does Extensive Grazing Benefit Butterflies in Coastal Dunes?** *Restor Ecol.* 2001; **9**(2): 179–88.

[Publisher Full Text](#)

Open Peer Review

Current Peer Review Status: ? ✓ ✓

Version 1

Reviewer Report 31 May 2022

<https://doi.org/10.21956/wellcomeopenres.19712.r49569>

© 2022 Matos-Maraví P. This is an open access peer review report 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.



Pável Matos-Maraví 

Biology Centre of the Czech Academy of Sciences, Institute of Entomology, České Budějovice, České Budějovice, Czech Republic

Hayward *et al.* report the genome sequence of *Pyrgus malvae*. The article describes well the relevance of having an annotated reference genome for the species. The methods are clearly described and allow the reproducibility of the experiments. Perhaps of interest for people interested in generating further reference genomes of Hesperidae, it would be informative to state the amount of high molecular weight DNA used for each HiFi circular consensus and 10X Genomics read cloud sequencing libraries.

The authors ensured that the data is open and that the figures and tables describe the output in a very informative way.

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: Lepidoptera, Phylogenetics, Macroevolution

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 21 April 2022

<https://doi.org/10.21956/wellcomeopenres.19712.r49571>

© 2022 Mutanen M. This is an open access peer review report 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.



Marko Mutanen 

Ecology and Genetics Research Unit, University of Oulu, Oulu, Finland

This work published a genome for a butterfly species representing family HesperIIDae. High-quality genomes like the one generated here are of great value for people working on genomics of butterflies and other taxa. I was carefully reading the manuscript through and found it very well prepared. As far as I can see, there are no questions concerning the validity of the methods and conducted analyses, and the quality of this work appears very good to me. I am, however, not very familiar with several of the analytical tools used in this study. Despite this, I am fully confident that this data merits being published in Wellcome Open Research journal and that there are no major (nor even minor) flaws in this paper. I am therefore happy to recommend its publication in its present form with two very minor reservations:

1. In the Genome annotation report section, the authors state that 2,976 non-coding genes were annotated. What are non-coding genes given that by definition genes are coding regions of genome?
2. I think a comma is lacking after the name Hayward on page 3.

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: Biodiversity genomics, DNA barcoding, Insect taxonomy

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 20 April 2022

<https://doi.org/10.21956/wellcomeopenres.19712.r49570>

© 2022 Marcus J. This is an open access peer review report 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.



Jeffrey M. Marcus 

Department of Biological Sciences, University of Manitoba, Winnipeg, MB, Canada

This manuscript describes the sequencing and genome assembly of the European grizzled skipper *Pyrgus malvae*. There are some conservation concerns in parts of the range of this species, making collecting its genome sequence justifiable. The work described is competent and the writing is clear. Most of the data is clearly presented in a useable and accessible format. The one exception is that it is not clear where to find the assembled sequences for the second haplotype. I can find only one set of accessions in the text of the manuscript, presumably for the first haplotype. Where the second haplotype can be found should be made clear.

Minor Concerns:

- p. 3 Background: First 4 sentences of the background are without in-text citations. Where does this information come from?
- p. 3 Similarly in 3rd background paragraph, no citations for: "When fully-grown, the larva constructs a cocoon at the base of low vegetation, where it overwinters as a pupa. Adults feed on a wide variety of nectar sources, including Bird's foot trefoil (*Lotus corniculatus*), bugle (*Ajuga reptans*), buttercup (*Ranunculus* species), daisy (*Bellis perennis*), and dandelion (*Taraxacum officinale*)."
- p. 3 suggested reword "At the European level this species is listed as Least Concern in the IUCN Red List..."
- p. 3 "While not fully phased, the assembly deposited is of one haplotype. Contigs corresponding to the second haplotype have also been deposited." Where is the second haplotype deposited? I can only find set of INSDC accession numbers in Table 2, which I presume are associated with the first haplotype.

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?

Partly

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: genomics, Lepidoptera, phylogenetics, genetics, evo-devo

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, however I have significant reservations, as outlined above.
