DATA NOTE

The genome sequence of the large tortoiseshell, *Nymphalis polychloros* (Linnaeus, 1758) [version 1; peer review: 2 approved]

Konrad Lohse[1], Dominik Laetsch[1], Roger Vila[2],
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

[1]Institute for Evolutionary Biology, University of Edinburgh, Edinburgh, UK
[2]Institut de Biologia Evolutiva (CSIC - Universitat Pompeu Fabra), Barcelona, Spain

v1 **First published:** 16 Sep 2021, **6**:238
https://doi.org/10.12688/wellcomeopenres.17196.1
Latest published: 16 Sep 2021, **6**:238
https://doi.org/10.12688/wellcomeopenres.17196.1

Abstract

We present a genome assembly from an individual female *Nymphalis polychloros* (the large tortoiseshell; Arthropoda; Insecta; Lepidoptera; Nymphalidae). The genome sequence is 398 megabases in span. The majority of the assembly is scaffolded into 32 chromosomal pseudomolecules, with the W and Z sex chromosome assembled.

Keywords

Nymphalis polychloros, large tortoiseshell, genome sequence, chromosomal

This article is included in the Tree of Life gateway.

Open Peer Review

Approval Status ✓ ✓

	1	2
version 1 16 Sep 2021	✓ view	✓ view

1. **Mathieu Joron**, Université de Montpellier, Montpellier, France
   Quentin Rougemont, Université de Montpellier, Montpellier, France
2. **Jiasheng Hao**, Anhui Normal University, Wuhu, China

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: Lohse K: Investigation, Resources, Writing – Original Draft Preparation, Writing – Review & Editing; **Laetsch D**: Investigation, Resources, Writing – Original Draft Preparation, Writing – Review & Editing; **Vila R**: Writing – Original Draft Preparation, Writing – Review & Editing;

Competing interests: No competing interests were disclosed.

Grant information: This work was supported by Wellcome through core funding to the Wellcome Sanger Institute (206194) and the Darwin Tree of Life Discretionary Award (218328). Fieldwork in Spain was supported by an ERC grant (ModelGenom Land 757648). RV is supported by grant PID2019-107078GB-I00/AEI/10.13039/501100011033 from the Spanish Government.
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: Lohse K, Laetsch D, Vila R *et al.* **The genome sequence of the large tortoiseshell, *Nymphalis polychloros* (Linnaeus, 1758) [version 1; peer review: 2 approved]** Wellcome Open Research 2021, **6**:238 https://doi.org/10.12688/wellcomeopenres.17196.1

First published: 16 Sep 2021, **6**:238 https://doi.org/10.12688/wellcomeopenres.17196.1

Species taxonomy

Eukaryota; Metazoa; Ecdysozoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Lepidoptera; Glossata; Ditrysia; Papilionoidea; Nymphalidae; Nymphalinae; Nymphalis; Nymphalis; *Nymphalis polychloros* (Linnaeus, 1758) (NCBI: txid171594).

Introduction

The large tortoiseshell, also known as the black-legged tortoiseshell or elm nymphalid, is a widespread but rare butterfly in woodlands across continental Europe, North Africa and Central Asia. Once common in England and Wales, *N. polychloros* went extinct in Southern Britain in the 1960s for unknown reasons and is currently classified as 'vulnerable' in several European countries (Maes *et al.*, 2020). It is listed as Least Concern in the IUCN Red List Category (Europe) (van Swaay *et al.*, 2010). However, recent sightings of a breeding colony in Dorset in 2021 suggest that this species is once again resident in the UK. It is morphologically very close to both the small tortoiseshell, *Aglais urticae*, and the scarce tortoiseshell, *N. xanthomelas*, in adult appearance. The species uses a wide variety of host plants such as *Pyrus*, *Prunus*, *Salix*, *Ulmus*, *Crataegus*, and others. It is univoltine and overwinters as an adult. (Lorković, 1941) reported a karyotype of 31 chromosomes and the genome size estimated for its relative, *Aglais io*, is 363.5 Mb (Mackintosh *et al.*, 2019).

Genome sequence report

The genome was sequenced from a single female *N. polychloros* (Figure 1) to 36-fold coverage in Pacific Biosciences single-molecule long reads and 84-fold coverage in 10X Genomics read clouds. Primary assembly contigs were scaffolded with chromosome conformation Hi-C data. Manual assembly curation corrected two missing/misjoins, reducing the scaffold number by 5.31%. The final assembly has a total length of 398 Mb in 38 sequence scaffolds with a scaffold N50 of 14 Mb (Table 1). Of the assembly sequence, 100% was assigned to 32 chromosomal-level scaffolds, representing 30 autosomes (numbered by sequence length), and the W and Z sex chromosome (Figure 2–Figure 5; Table 2). The assembly has a BUSCO v5.1.2 (Simão *et al.*, 2015) completeness of 98.8% 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

The female *N. polychloros* specimen SC_NP_345 was collected using a net from Somiedo, Brana de Mumian, Asturias, Spain

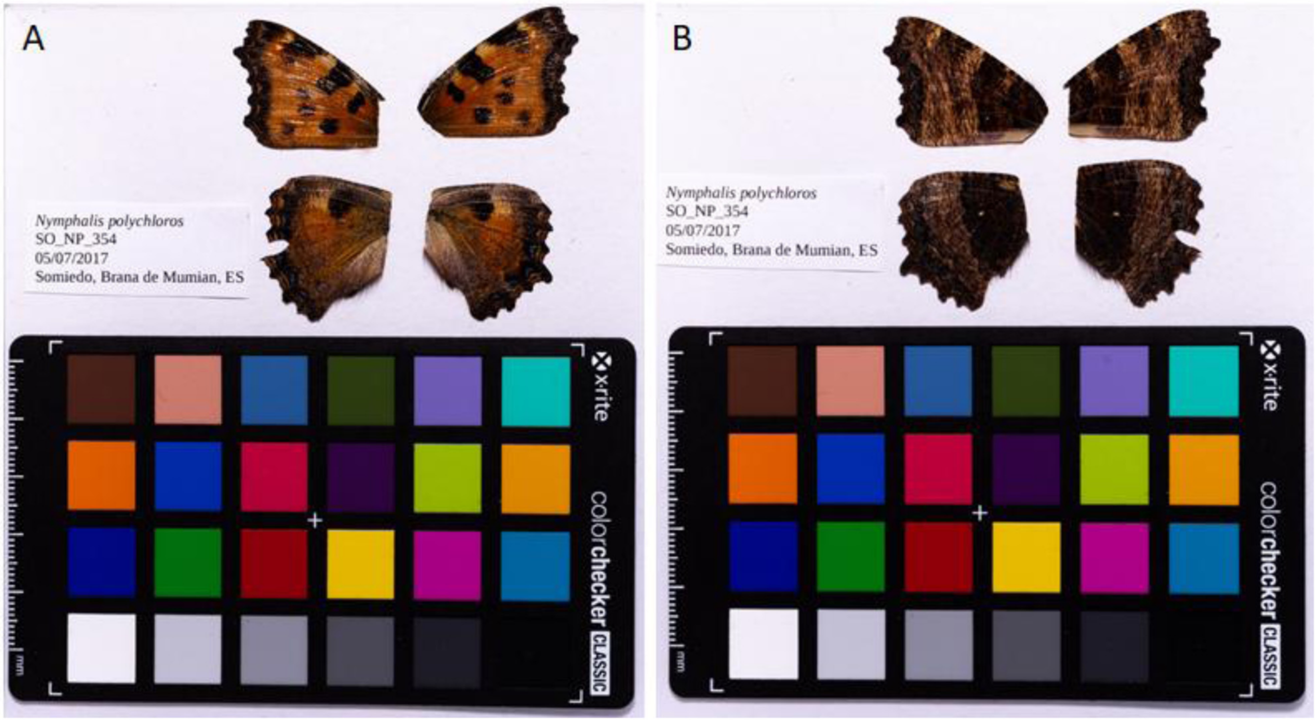

Figure 1. Fore and hind wings of *Nymphalis polychloros* specimen from which the genome was sequenced. (**A**) Dorsal surface view of wings from specimen SO_NP_354 (ilNymPoly1) from Somiedo, Spain used to generate Pacific Biosciences and 10X genomics data. (**B**) Ventral surface view of wings from specimen SO_NP_354 from Somiedo, Spain, used to generate Pacific Biosciences and 10X genomics data.

Table 1. Genome data for *Nymphalis polychloros*, ilNymPoly1.1.

Project accession data	
Assembly identifier	ilNymPoly1.1
Species	*Nymphalis polychloros*
Specimen	ilNymPoly1
NCBI taxonomy ID	NCBI:txid171594
BioProject	PRJEB43012
BioSample ID	SAMEA7523140
Isolate information	Female, whole organism
Raw data accessions	
PacificBiosciences SEQUEL II	ERR6590585
10X Genomics Illumina	ERR6054433-ERR6054436
Hi-C Illumina	ERR6054437
RNAseq PolyA Illumina	ERR6286714
Genome assembly	
Assembly accession	GCA_905220585.1
Accession of alternate haplotype	GCA_905220575.1
Span (Mb)	398
Number of contigs	45
Contig N50 length (Mb)	14
Number of scaffolds	38
Scaffold N50 length (Mb)	14
Longest scaffold (Mb)	17
BUSCO* genome score	C:98.8%[S:98.6%,D:0.2%],F:0.3%,M:0.8%,n:5286

*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/ilNymPoly1.1/dataset/CAJNAJ01/busco.

(latitude 43.0679, longitude -6.239918) by Konrad Lohse, University of Edinburgh. Permissions for field sampling were granted by the Gobierno del Principado de Asturias (014252). The specimen was snap-frozen from live in liquid nitrogen.

DNA was extracted from thorax 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. RNA was extracted (also from thorax tissue) in the Tree of Life Laboratory at the WSI using TRIzol (Invitrogen), according to the manufacturer's instructions. RNA was then eluted in 50 μl 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.

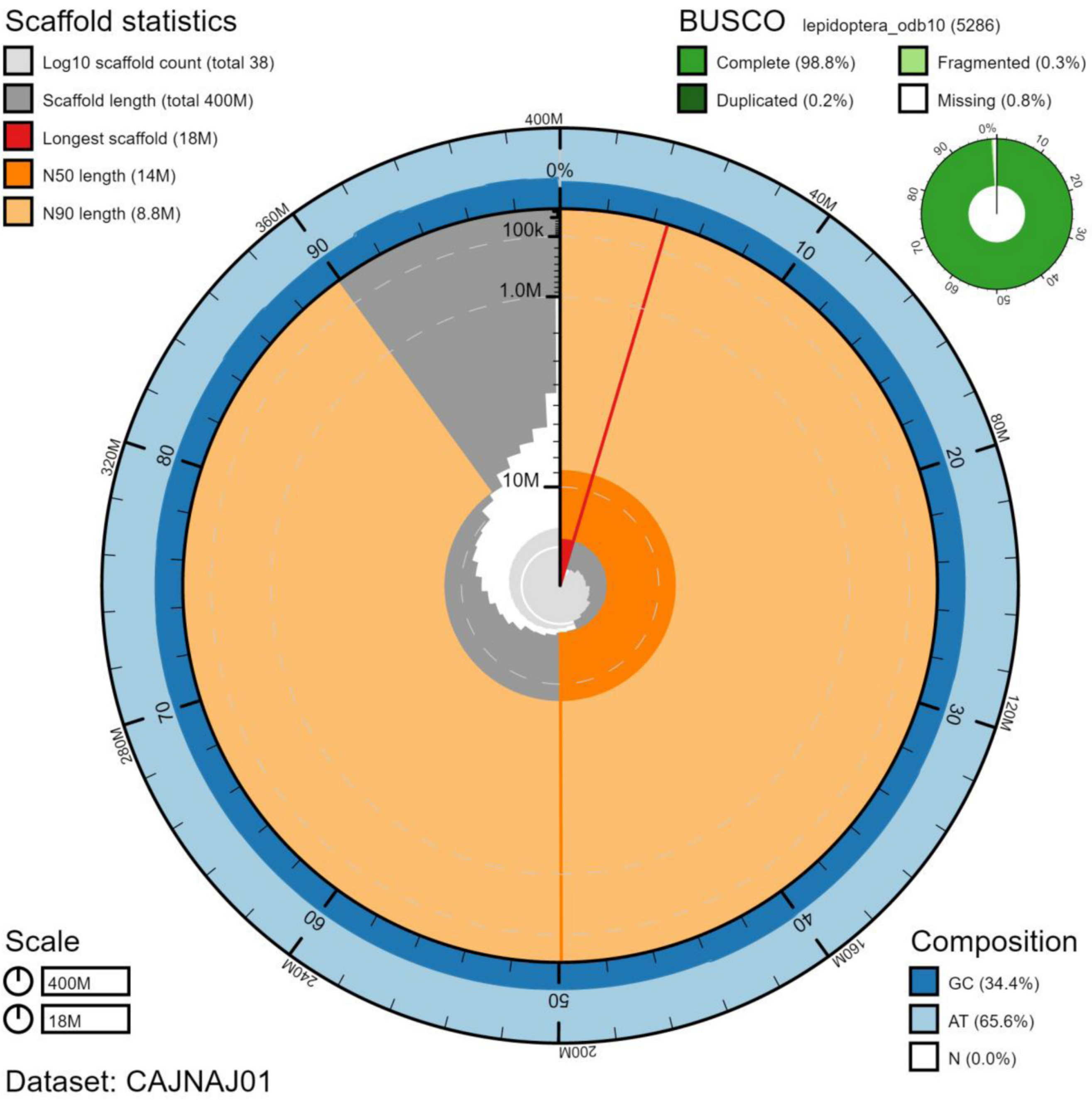

Figure 2. Genome assembly of *Nymphalis polychloros*, ilNymPoly1.1: metrics. The BlobToolKit Snailplot shows N50 metrics and BUSCO gene completeness. An interactive version of this figure is available at https://blobtoolkit.genomehubs.org/view/ilNymPoly1.1/dataset/CAJNAJ01/snail.

Pacific Biosciences HiFi circular consensus and 10X Genomics read cloud DNA sequencing libraries, in addition to PolyA RNA-Seq libraries, were constructed according to the manufacturers' instructions. DNA and RNA sequencing was performed by the Scientific Operations core at the WSI on Pacific Biosciences SEQUEL II (HiFi), Illumina HiSeq X (10X) and Illumina HiSeq 4000 (RNA-Seq) instruments. Hi-C data were generated from abdomen tissue using the Arima v2.0 kit and sequenced on Illumina NovaSeq.

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

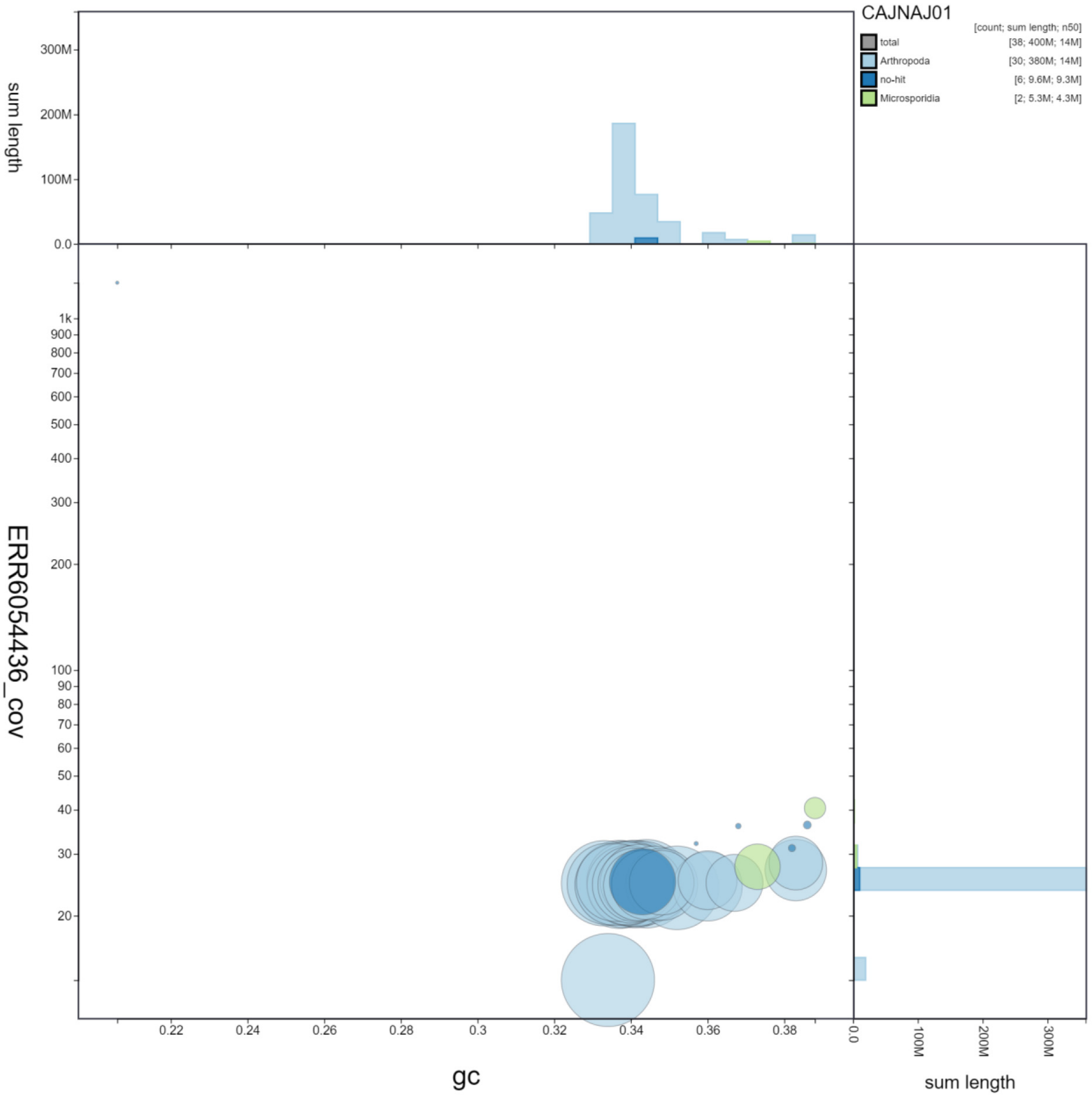

Figure 3. Genome assembly of *Nymphalis polychloros*, ilNymPoly1.1: GC coverage. BlobToolKit GC-coverage plot. Chromosomes are coloured by phylum. Circles are sized in proportion to chromosome length Histograms show the distribution of chromosome length sum along each axis. An interactive version of this figure is available at https://blobtoolkit.genomehubs.org/view/ilNymPoly1.1/dataset/CAJNAJ01/blob.

purge_dups (Guan *et al.*, 2020). One round of polishing was 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 and corrected using the gEVAL system (Chow *et al.*, 2016) as described previously (Howe *et al.*, 2021). Manual curation was performed using

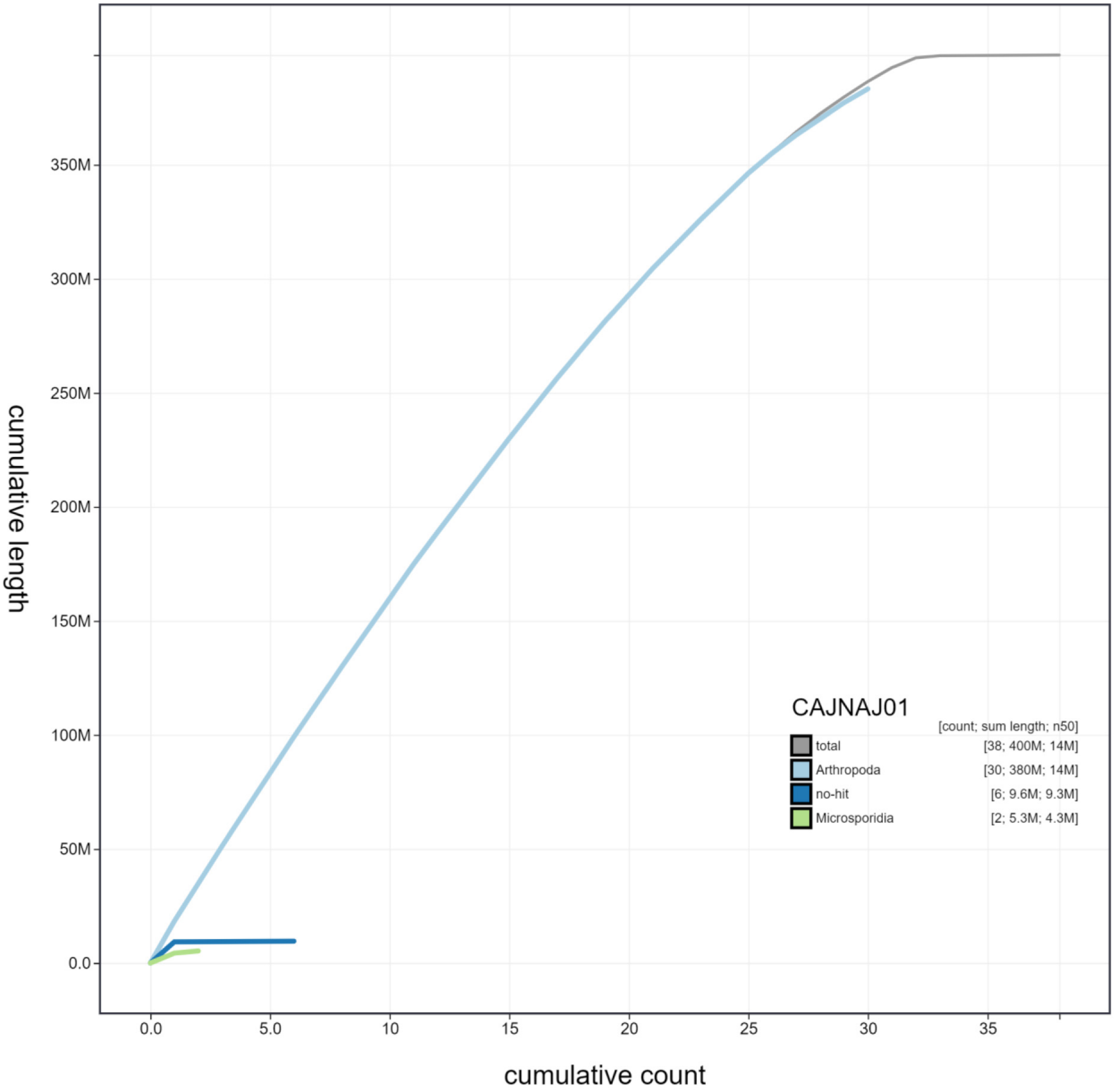

Figure 4. Genome assembly of *Nymphalis polychloros*, ilNymPoly1.1: cumulative sequence. BlobToolKit cumulative sequence plot. The grey line shows cumulative length for all chromosomes. Coloured lines show cumulative lengths of chromosomes assigned to each phylum using the buscogenes taxrule. An interactive version of this figure is available at https://blobtoolkit.genomehubs.org/view/ilNymPoly1.1/dataset/CAJNAJ01/cumulative.

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

The materials that have contributed to this genome note were supplied by a Tree of Life collaborator. 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

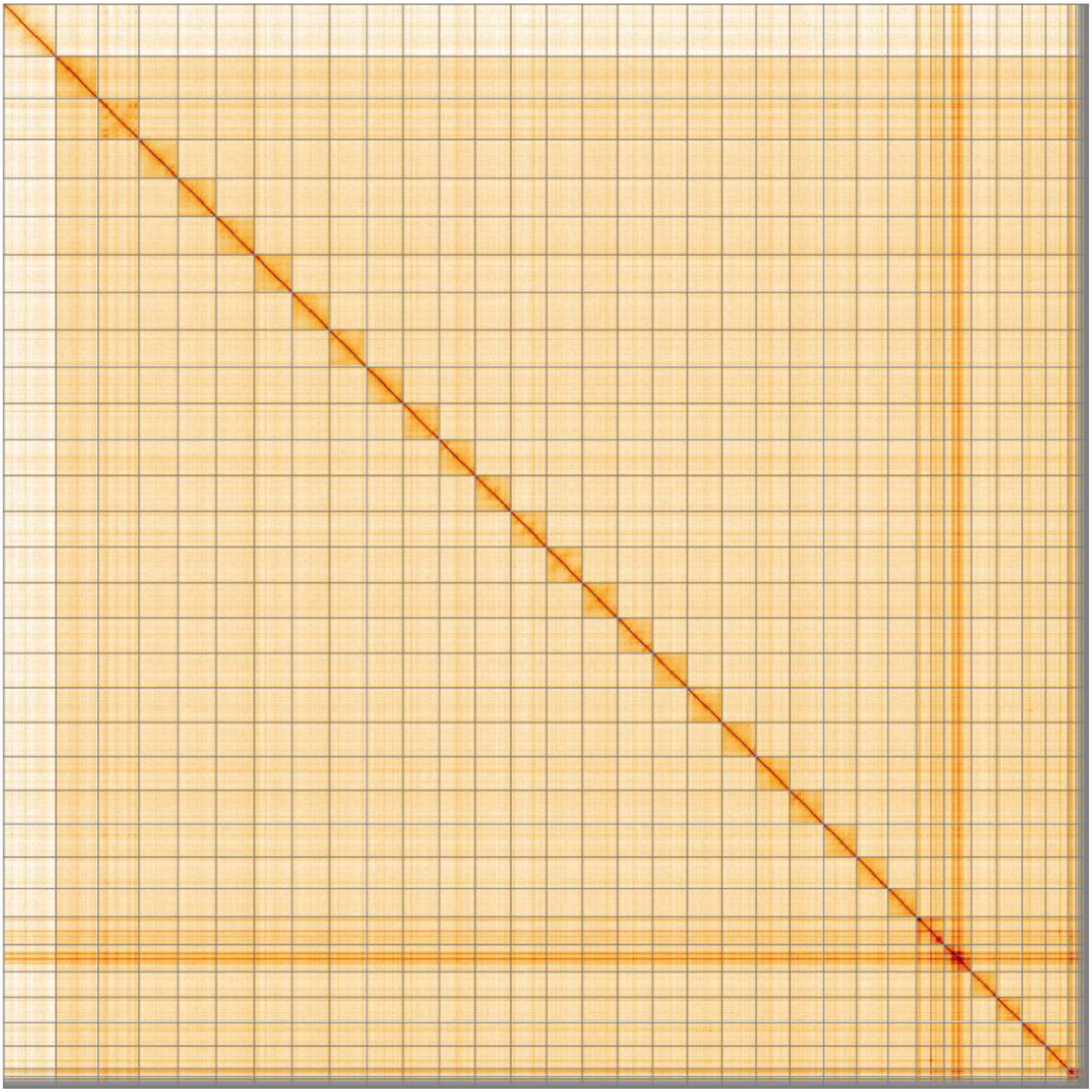

Figure 5. Genome assembly of *Nymphalis polychloros*, ilNymPoly1.1: Hi-C contact map. Hi-C contact map of the ilNymPoly1.1 assembly, visualised in HiGlass.

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 undertaken according to a Research Collaboration Agreement or Material Transfer Agreement entered into by the Tree of Life collaborator, Genome Research Limited (operating as the Wellcome Sanger Institute) and in some circumstances other Tree of Life collaborators.

Table 2. **Chromosomal pseudomolecules in the genome assembly of *Nymphalis polychloros*, ilNymPoly1.1.**

INSDC accession	Chromosome	Size (Mb)	GC%
HG992242.1	1	16.56	34.4
HG992243.1	2	16.45	33.7
HG992244.1	3	16.03	34.1
HG992245.1	4	15.91	33.8
HG992246.1	5	15.83	34.1
HG992247.1	6	15.48	34
HG992248.1	7	15.41	33.3
HG992249.1	8	15.04	34.3
HG992250.1	9	14.99	34.1
HG992251.1	10	14.77	35.2
HG992252.1	11	14.06	33.4
HG992253.1	12	13.93	33.6
HG992254.1	13	13.74	33.9
HG992255.1	14	13.53	33.9
HG992256.1	15	13.45	33.9
HG992257.1	16	12.92	33.6
HG992258.1	17	12.55	34
HG992259.1	18	12.34	34.2
HG992260.1	19	11.88	34.6
HG992261.1	20	11.42	34.1
HG992262.1	21	10.92	34.8
HG992263.1	22	10.46	34.2
HG992264.1	23	10.26	34.4
HG992265.1	24	10.09	36
HG992266.1	25	9.27	34.3
HG992267.1	26	8.82	34.8
HG992268.1	27	7.95	38.3
HG992269.1	28	7.29	36
HG992270.1	29	6.82	36.7
HG992271.1	30	6.08	38.3
HG992272.1	W	4.33	37.3
HG992241.1	Z	18.34	33.4
HG992273.1	MT	0.02	20.3
-	Unplaced	1.22	38.6

Table 3. **Software tools used.**

Software tool	Version	Source
Hifiasm	0.12	Cheng *et al.*, 2021
purge_dups	1.2.3	Guan *et al.*, 2020
longranger	2.2.2	https://support.10xgenomics.com/genome-exome/software/pipelines/latest/advanced/other-pipelines
freebayes	1.3.1-17-gaa2ace8	Garrison & Marth, 2012
SALSA2	2.2	Ghurye *et al.*, 2019
MitoHiFi	1.0	Uliano-Silva *et al.*, 2021
gEVAL	N/A	Chow *et al.*, 2016
HiGlass	1.11.6	Kerpedjiev *et al.*, 2018
PretextView	0.1.x	https://github.com/wtsi-hpag/PretextView
BlobToolKit	2.6.1	Challis *et al.*, 2020

Data availability

European Nucleotide Archive: Nymphalis polychloros (large tortoiseshell). Accession number PRJEB42956; https://identifiers.org/ena.embl:PRJEB42956.

The genome sequence is released openly for reuse. The *N. polychloros* 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 the 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.

Acknowledgements

Members of the Darwin Tree of Life Barcoding collective are listed here: https://doi.org/10.5281/zenodo.4893704.

Members of the Wellcome Sanger Institute Tree of Life programme collective are listed here: https://doi.org/10.5281/zenodo.4783586.

Members of Wellcome Sanger Institute Scientific Operations: DNA Pipelines collective are listed here: https://doi.org/10.5281/zenodo.4790456.

Members of the Tree of Life Core Informatics collective are listed here: https://doi.org/10.5281/zenodo.5013542.

Members of the Darwin Tree of Life Consortium are listed here: https://doi.org/10.5281/zenodo.4783559.

References

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

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

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

Lorković Z: **Die Chromosomenzahlen in der Spermatogenese der Tagfalter.** *Chromosoma.* 1941; **2**(1): 155–91.
Publisher Full Text

Mackintosh A, Laetsch DR, Hayward A, *et al.*: **The Determinants of Genetic Diversity in Butterflies.** *Nat Commun.* 2019; **10**(1): 3466.
PubMed Abstract | Publisher Full Text | Free Full Text

Maes D, Wiemers M, Verovnik R, *et al.*: **National Checklists and Red Lists for European Butterflies.** Research Institute for Nature and Forest (INBO), 2020.
Publisher Full Text

Rao SS, 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

Simão FA, Waterhouse RM, Ioannidis P, *et al.*: **BUSCO: Assessing Genome Assembly and Annotation Completeness with Single-Copy Orthologs.** *Bioinformatics.* 2015; **31**(19): 3210–12.
PubMed Abstract | Publisher Full Text

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

van Swaay C, Wynhoff I, Verovnik R, *et al.*: **Nymphalis polychloros. The IUCN Red List of Threatened Species**. 2010; e.T174359A7057445.
Reference Source

Open Peer Review

Current Peer Review Status: ✓ ✓

Version 1

Reviewer Report 25 April 2022

https://doi.org/10.21956/wellcomeopenres.19000.r49892

✓ **Jiasheng Hao**
Laboratory of Molecular Evolution and Biodiversity, Anhui Normal University, Wuhu, China

1. In the last sentence of the manuscript's Introduction, "(Lorković, 1941) reported a karyotype of 31 chromosomes and the genome size estimated for its relative, *Aglais io*, is 363.5Mb (Mackintosh *et al.*, 2019)."We see that the genome size of *Nymphalis polychloros* has not been reported in previous studies. It is suggested that the authors should increase the analysis of genome size prediction based on GCE or other tools in the Methods, and this is a good guide for subsequent genome assembly or haplotypic selection.

2. In the part of "Genome sequence report", the W and Z sex chromosomes were mentioned to be assembled in this work. However, how to identify sex chromosomes in the Methods was not stated clearly, and the authors are suggested to clearly introduce the method for distinguishing sex chromosomes in this study.

3. The results of genome assembly probably depend on the setting of parameters used in the software, thus the authors are suggested to add the main parameters of tools in Table 3.

4. For some quality control analysis in the Methods, can the authors describe the changes in data volume before and after filtering in the text? For example, "haplotypic duplication was identified and removed with purge_dups (Guan *et al.*, 2020)." "Manual curation was performed using Manual curation was performed using gEVAL, HiGlass (Kerpedjiev *et al.*, 2018) and Pretext."

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: the phylogeny and evolutionary history of some butterfly groups.

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 18 October 2021

https://doi.org/10.21956/wellcomeopenres.19000.r45905

✓

Mathieu Joron

Centre d'Ecologie Fonctionnelle et Evolutive (CEFE), CNRS, Université de Montpellier, Montpellier, France

Quentin Rougemont

Centre d'Ecologie Fonctionnelle et Evolutive (CEFE), CNRS, Université de Montpellier, Montpellier, France

This data note by Lohse and colleagues reports the genome sequence of a butterfly in the large family Nymphalidae, well known for harbouring other iconic butterfly models for metapopulation dynamics or for the study of wing coloration. This adds a reference genome for a group of butterflies with interesting biology such as adult overwintering and long distance migration.

The genome presented is of very high quality. I see nothing problematic with the assembly which offers excellent contiguity. My comments are minor but may reflect the perspective of someone involved in assembling similar genomes.

It may make no difference, but wondered why the length was initially estimated to the 363Mb of the peacock butterfly *Aglais io*, while the comma butterfly *Polygonia c-album* appears more closely related, and has a published, slightly higher genome size estimated to 373Mb (Celorio-Mancera *et al.* 2021 Genome Biol. Evol. 13:evab054)[1]

Table 2 reports 1.22Mb of unplaced scaffolds, and the text reports 38 sequence scaffolds. Yet the authors report that 100% of the assembly was assigned to 32 chromosomal-level scaffolds. I was wondering what constitutes the difference between those statistics (large heterozygous tracts/haplotypes?)

Hi-C data was generated from abdomen tissue. Since this is a wild and therefore mated female, the abdomen is likely to contain recombinant gametes from one or several unknown males. I assume that this is unlikely to introduce large errors for scaffolding, but wondered whether the authors could perhaps give a few words on this possible issue, since it is a question that frequently arises when using abdominal female tissue from wild-caught individuals.

This assembly is made from the DNA of a wild-caught specimen. Therefore it would be interesting to provide details on its observed heterozygosity. Dealing with heterozygosity is a recurrent issue in genome assembly.

Similarly, chosen parameter values would be good to provide for all packages and softwares (such as hifiasm, freebase, etc), perhaps as an additional column for table 3. Those values are essential for reproducibility, but also very useful for people assembling similar genomes.

It would be interesting to provide details of the improvements allowed by the different steps (for instance the polishing step, by giving stats before and after). Again this would be useful for other users and generally for assembly of similar genomes. This could take the form of a table.

The note presents the generation of RNAseq data, which is great, but the data is not (yet) used for annotation. I wondered why include this in the methods if it is actually not analysed.

Status/justification: This species the large tortoiseshell is a fairly common though elusive butterfly in its predominantly continental European range. I understand that the Darwin tree of life effort is motivated by sequencing “British taxa” and the status of this species in Britain may have influenced its position on the priority list. However, the aim of a reference genome probably goes beyond that. Abundance and conservation status of taxa are very variable depending on how far from the range margin one stands! From a broader perspective, the large tortoiseshell is a forest species with a broad European distribution. It has a relatively poorly known ecology compared to closely related species. And interesting question marks remain regarding the origins and status of its genetic structure (vicariance, speciation?). *Nymphalis* as a genus also has unclear relationships with other genera such as *Polygonia* and *Kaniska*. Perhaps a reference genome could stimulate interesting research on those aspects which could make a better "justification" for sequencing it than the recent sighting of a colony in Dorset where the species is teetering on its range margin.

References

1. Celorio-Mancera M, Rastas P, Steward R, Nylin S, et al.: Chromosome Level Assembly of the Comma Butterfly (Polygonia c-album). *Genome Biology and Evolution*. 2021; **13** (5). Publisher Full Text

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?

Partly

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

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Population genomics

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