DATA NOTE

The genome sequence of the small tortoiseshell butterfly, *Aglais urticae* (Linnaeus, 1758) [version 1; peer review: 1 approved, 1 approved with reservations]

Gertjan Bisschop¹, Sam Ebdon ¹, Konrad Lohse 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

¹Institute of Evolutionary Biology, University of Edinburgh, Edinburgh, UK

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

V1 First published: 14 Sep 2021, 6:233
<https://doi.org/10.12688/wellcomeopenres.17197.1>
 Latest published: 14 Sep 2021, 6:233
<https://doi.org/10.12688/wellcomeopenres.17197.1>

Abstract

We present a genome assembly from an individual female *Aglais urticae* (also known as *Nymphalis urticae*; the small tortoiseshell; Arthropoda; Insecta; Lepidoptera; Nymphalidae). The genome sequence is 384 megabases in span. The majority of the assembly is scaffolded into 32 chromosomal pseudomolecules, with the W and Z sex chromosome assembled.

Keywords

Aglais urticae, small tortoiseshell butterfly, genome sequence, chromosomal

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

Open Peer Review

Approval Status ? ✓

	1	2
version 1	?	✓
14 Sep 2021	view	view

1. **Christopher W. Wheat** , Stockholm University, Stockholm, Sweden
2. **Reuben Nowell** , University of Oxford, Oxford, UK

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: **Bisschop G:** Investigation, Resources, Writing – Original Draft Preparation, Writing – Review & Editing; **Ebdon S:** Investigation, Resources, Visualization, Writing – Original Draft Preparation, Writing – Review & Editing; **Lohse K:** Investigation, Resources, Writing – Original Draft Preparation, Writing – Review & Editing; **Vila R:** Investigation, Writing – Original Draft Preparation, Writing – Review & Editing;

Competing interests: No competing interests were disclosed.

Grant information: This work was supported by the Wellcome Trust through core funding to the Wellcome Sanger Institute (206194) and the Darwin Tree of Life Discretionary Award (218328). GB and KL are supported by an ERC grant (ModelGenom Land 757648), SE is supported by an EastBio studentship from the Biotechnology and Biological Sciences Research Council UK.

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

Copyright: © 2021 Bisschop G *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: Bisschop G, Ebdon S, Lohse K *et al.* **The genome sequence of the small tortoiseshell butterfly, *Aglais urticae* (Linnaeus, 1758) [version 1; peer review: 1 approved, 1 approved with reservations]** Wellcome Open Research 2021, 6:233 <https://doi.org/10.12688/wellcomeopenres.17197.1>

First published: 14 Sep 2021, 6:233 <https://doi.org/10.12688/wellcomeopenres.17197.1>

Species taxonomy

Eukaryota; Metazoa; Ecdysozoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Lepidoptera; Glossata; Ditrysia; Papilionoidea; Nymphalidae; Nymphalinae; Nymphalis; Aglais; *Aglais urticae* (also known as *Nymphalis urticae*) (Linnaeus, 1758) (NCBI:txid111881).

Introduction

Aglais urticae (also known as *Nymphalis urticae*), the small tortoiseshell, is a widespread butterfly found in temperate regions from western Europe to Japan. Occasionally, individuals are reported from Eastern North America. Known as *slige an t-sligeanaich bheag* in Scottish Gaelic, it is ubiquitous in the British Isles despite facing declines in population size over the past 50 years (Fox *et al.*, 2015). Adults can be seen on the wing from spring to autumn or overwintering in sheds and outhouses. As indicated by its name, the caterpillars can be seen feeding on nettles (*Urtica dioica* and *U. urens*) over two

generations in the summer (with the exception of the Scottish populations, which regularly have just one annual generation). Variation in wing morphology has led to suggestions of a suite of subspecies across the range although evidence for evolutionary lineages using mitochondrial markers is inconclusive (Vandewoestijne *et al.*, 2004). This species is listed as Least Concern in the IUCN Red List (Europe) (van Swaay *et al.*, 2010). *A. urticae* has 31 pairs of chromosomes (Beliajeff, 1930) and the female is heterogametic (WZ).

Genome sequence report

The genome was sequenced from one female *A. urticae* (ilAglUrti1) collected from Carrifran Wildwood, Dumfries and Galloway, Scotland (latitude 55.400132, longitude -3.3352); Hi-C data were obtained from a second female *A. urticae* (ilAglUrti2) collected from Falkland, Fife, Scotland (latitude 56.25567, longitude -3.210498) (Figure 1). A total of 58-fold coverage in Pacific Biosciences single-molecule long

Figure 1. Fore and hind wings of *Aglais urticae* specimens from which the genome was sequenced. (A) Dorsal surface view of wings from specimen ilAglUrti1 (SC_AU_1387) from Carrifran Wildwood, Scotland used to generate Pacific Biosciences and 10X genomics data. (B) Ventral surface view of wings from specimen ilAglUrti1 (SC_AU_1387) from Carrifran Wildwood, Scotland used to generate Pacific Biosciences and 10X genomics data. (C) Dorsal surface view of wings from specimen ilAglUrti2 (SC_AU_1351) from Falkland, Scotland used to generate Hi-C data. (D) Ventral surface view of wings from specimen ilAglUrti2 (SC_AU_1351) from Falkland, Scotland used to generate Hi-C data.

reads (N50 15 kb) and 92-fold coverage in 10X Genomics read clouds (from molecules with an estimated N50 of 41 kb) were generated. Primary assembly contigs were scaffolded with chromosome conformation Hi-C data. Manual assembly curation corrected 12 missing/misjoins and removed one haplotypic duplication, reducing the assembly length by 0.01% and the scaffold number by 12.82%, and increasing the scaffold N50 by 0.15%. The final assembly has a total length of 393 Mb in 35 sequence scaffolds with a scaffold N50 of 13.17 Mb (Table 1). The assembly sequence 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 (Simão *et al.*, 2015) v5.1.2 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 first female *A. urticae* sample, ilAglUrti1 (SC_AU_1387), was collected from Carrifran Wildwood, Dumfries and Galloway, Scotland, UK (latitude 55.400132, longitude -3.3352) by Konrad Lohse. The second female *A. urticae*, ilAglUrti2 (SC_AU_1351), was collected from Falkland, Fife, Scotland, UK (latitude 56.25567, longitude -3.210498) by Gertjan Bishop, University of Edinburgh. Both specimens were captured using a net and snap-frozen in liquid nitrogen.

DNA was extracted from the whole organism of ilAglUrti1 at the Wellcome Sanger Institute (WSI) Scientific Operations core using the Qiagen MagAttract HMW DNA kit, according

Table 1. Genome data for *Aglais urticae*, ilAglUrti1.1.

Project accession data	
Assembly identifier	ilAglUrti1.1
Species	<i>Aglais urticae</i>
Specimen	ilAglUrti1, SC_AU_1387 (genome assembly); ilAglUrti2, SC_AU_1351 (Hi-C)
NCBI taxonomy ID	NCBI:txid111881
BioProject	PRJEB42112
BioSample ID	SAMEA7523286
Isolate information	Female, whole organisms
Raw data accessions	
PacificBiosciences SEQUEL II	ERR6590580
10X Genomics Illumina	ERR6002566-ERR6002569
Hi-C Illumina	ERR6002570-ERR6002573
Illumina PolyA RNA-Seq	ERR6286701
Genome assembly	
Assembly accession	GCA_905147175.1
Accession of alternate haplotype	GCA_905147055.1
Span (Mb)	393
Number of contigs	44
Contig N50 length (Mb)	13
Number of scaffolds	35
Scaffold N50 length (Mb)	13
Longest scaffold (Mb)	17
BUSCO* genome score	C:98.8%[S:98.6%,D:0.1%],F:0.3%,M:0.9%,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/ilAglUrti1.1/dataset/CAJHUP01/busco>.

Figure 2. Genome assembly of *Aglais urticae*, iAglUrti1.1: 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/iAglUrti1.1/dataset/CAJHUP01/snail>.

to the manufacturer's instructions. RNA (also from the whole organism) was extracted in the Tree of Life Laboratory at the WSI using TRIzol, 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.

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 the whole organism of iAglUrti2 using the Arima v1.0 kit and sequenced on HiSeq X.

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 3. Genome assembly of *Aglais urticae*, ilAglUrti1.1: 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/ilAglUrti1.1/dataset/CAJHUP01/blob>.

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 gEVAL, HiGlass (Kerpedjiev *et al.*, 2018) and Pretext. The mitochondrial genome was assembled using

Figure 4. Genome assembly of *Aglais urticae*, iAglUrti1.1: 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/iAglUrti1.1/dataset/CAJHUP01/cumulative>.

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,

Figure 5. Genome assembly of *Aglais urticae*, ilAglUrti1.1: Hi-C contact map. Hi-C contact map of the ilAglUrti1.1 assembly, visualised in HiGlass.

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).

Table 2. Chromosomal pseudomolecules in the genome assembly of *Aglais urticae*, ilAglUrt1.1.

INSDC accession	Chromosome	Size (Mb)	GC%
LR989983.1	1	16.96	33.6
LR989984.1	2	15.45	32.8
LR989985.1	3	15.24	33.2
LR989986.1	4	14.77	32.8
LR989987.1	5	14.58	32.8
LR989988.1	6	14.44	33.2
LR989989.1	7	14.35	33.3
LR989990.1	8	13.79	32.9
LR989991.1	9	13.78	33
LR989992.1	10	13.73	33.4
LR989993.1	11	13.37	33.2
LR989994.1	12	13.26	32.8
LR989995.1	13	13.17	32.7
LR989997.1	14	12.91	33
LR989998.1	15	12.84	33.3
LR989999.1	16	12.53	32.8
LR990000.1	17	12.32	34.1
LR990001.1	18	12.15	33.3
LR990002.1	19	12.13	33.8
LR990003.1	20	11.20	33.3
LR990004.1	21	10.56	33.6
LR990005.1	22	10.00	34
LR990006.1	23	9.86	33.7
LR990007.1	24	9.83	34.4
LR990008.1	25	9.65	33.4
LR990009.1	26	9.20	37.5
LR990010.1	27	8.23	34.8
LR990011.1	28	8.19	34
LR990012.1	29	7.46	35.3
LR990013.1	30	6.48	36.9
LR989996.1	W	13.15	36
LR989982.1	Z	17.13	32.4
LR990014.1	MT	0.02	19.5
	Unplaced	0.09	47.4

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
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	1.0	Uliano-Silva et al., 2021
gEVAL	2016	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

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.

Data availability

European Nucleotide Archive: *Aglais urticae* (small tortoise-shell). Accession number PRJEB42112; <https://identifiers.org/ena.embl:PRJEB42112>.

The genome sequence is released openly for reuse. The *A. urticae* 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.5377053>.

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

- Beliajeff NK: **Die Chromosomenkomplexe Und Ihre Beziehung Zur Phylogenie Bei Den Lepidopteren**. *Zeitschrift Für Induktive Abstammungs- Und Vererbungslehre*. 1930; **54**: 369–399.
[Publisher 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): 170–75.
[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](#)

Fox R, Brereton TM, Asher J, *et al.*: **The State of the UK's Butterflies 2015.** 2015.

[Reference Source](#)

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): g1aa153.

[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](#)

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](#)

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, Krashenninnikova K, *et al.*: **marcelauliano/MitoHiFi: mitohifi_v2.0.** 2021.

[Publisher Full Text](#)

Vandewoestijne S, Baguette M, Brakefield PM, *et al.*: **Phylogeography of *Aglaia urticae* (Lepidoptera) Based on DNA Sequences of the Mitochondrial COI Gene and Control Region.** *Mol Phylogenet Evol.* 2004; **31**(2): 630–46.

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

van Swaay C, Wynhoff J, Verovnik R, *et al.*: ***Aglaia urticae*. The IUCN Red List of Threatened Species.** 2010; e.T174463A7076762.

[Reference Source](#)

Open Peer Review

Current Peer Review Status:

Version 1

Reviewer Report 07 October 2021

<https://doi.org/10.21956/wellcomeopenres.19001.r45869>

© 2021 Nowell R. 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.

Reuben Nowell

Department of Zoology, University of Oxford, Oxford, UK

The manuscript of Bishop *et al.* presents the high quality genome data and assembly for the small tortoiseshell butterfly *Aglaia urticae*.

It is concise and easy to read, with clear links to the raw datasets and the final assembly. I particularly like the use of interactive data visualisations. Overall I am convinced as to the high quality of the work and I'm sure the data will be a great resource to the community.

I have a few small comments for improvements:

1. I think it would be useful to see a bit more detail given for some of the methods. For example, it is stated "Manual assembly curation corrected 12 missing/misjoins and removed one haplotypic duplication", but there are no details (beyond the tools used) as to what these manual steps involved. Similarly, it is not clear exactly how polishing was performed, beyond the fact that the FreeBayes program was used. It would be helpful to the community to include the actual program commands and/or parameters and flags etc. that were executed for these steps. Perhaps these could be added to Table 3, which already helpfully provides the versioned software used? I don't think every last detail is required - just any non-standard steps (e.g., FreeBayes is a variant-calling tool, so it's unclear to me from the manuscript how this was used to polish).
2. Figure 1 is a bit blurry, I can only just read the text. The same for Figure 5 although perhaps this is less important (no text).
3. Figure 2 and Figure 5 could also benefit from a better explanation of what they are showing in the legend.
4. Any idea what those two high GC, high(ish) coverage small scaffolds are on the blobplot?

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: Evolutionary genomics, eukaryote genome assembly

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 15 September 2021

<https://doi.org/10.21956/wellcomeopenres.19001.r45866>

© 2021 Wheat C. 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.

Christopher W. Wheat

Department of Zoology, Faculty of Science, Stockholm University, Stockholm, Sweden

The paper reports a chromosome assembly of a butterfly genome, with extensive documentation as to its quality. The reporting of datasets and methods used is very clear. This will serve as an excellent resource to the community.

However, I find several problems with this genome assembly report. First, BUSCO results are perhaps best reported as the % of single copy complete and duplicated, as these two values are much more informative than the single completeness value. Reporting a single completeness value as an assessment of genome assembly quality can be very misleading, as completeness can be very high along with a high level of duplicated genes, indicating a high % of non-haploid genome assembly content. Stated another way, when reporting gene completeness in the context of genome assembly, one is inherently working within the framework of reporting the completeness of a haploid genome. Thus, simply reporting completeness is inappropriate, as it can mask the non-haploid assembly quality. Instead, one should either primarily report single copy complete results or completeness along with duplicate level. I hope that going forward all Darwin Tree of Life will correct this issue, as it currently plagues many genomic reports and degrades the insight provided by BUSCO scores. Second, the figures should be revised a bit. For Figure 2, it would be good to have a brief description of what this figure is actually depicting. Certainly, one can read the values, but understanding what is being shown would be most helpful,

even in a standard fashion that can be generic and used for all genomes shown. For Figure 5, it would greatly help to have the assigned chromosomes indicated along one axis, as I'm left wondering what the chromosomal group in the middle is likely to be (should be W no?).

My final comment is that while all of the data files and software used are clearly reported, none of the bioinformatic commands used for the assembly were reported. I find this rather unfortunate and a missed opportunity. While these may be default implementations of simple command line operations, the reporting of these would help standardize method deployment within the genomics community and thus I strongly suggest that the Darwin Tree of Life consortium refer to versioned command line operations used for their assemblies.

Minor comments:

- Figure 1 should be in higher resolution, as this is an important image as voucher information for the specimen.

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: Genome assembly, assessment and publication, with a focus upon Lepidoptera.

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.