DATA NOTE

The genome sequence of the ringlet, *Aphantopus hyperantus* Linnaeus 1758 [version 1; peer review: 2 approved with reservations]

Dan Mead[1,2], Ilik Saccheri[3], Carl J. Yung[3], Konrad Lohse[4], Carla Lohse[4], Philip Ashmole[5], Michelle Smith[1], Craig Corton[1], Karen Oliver[1], Jason Skelton[1], Emma Betteridge[1], Michael A. Quail[1], Jale Dolucan[1,6], Shane A. McCarthy[1,7], Kerstin Howe[1], Jonathan Wood[1], James Torrance[1], Alan Tracey[1], Sam Whiteford[3], Richard Challis[1], Richard Durbin[1,7], Mark Blaxter[1]

[1]Wellcome Sanger Institute, Wellcome Genome Campus, Hinxton, Cambridgeshire, CB10 1SA, UK
[2]Owlstone Medical, Cambridge Science Park, Cambridge, CB4 0GJ, UK
[3]Institute of Integrative Biology, University of Liverpool, Liverpool, L69 7ZB, UK
[4]Institute of Evolutionary Biology, University of Edinburgh, Edinburgh, EH9 3JT, UK
[5]c/o Borders Forest Trust, Monteviot Nurseries, Ancrum, Jedburgh, TD8 6TU, UK
[6]Achilles Therapeutics plc, London, W6 8PW, UK
[7]Department of Genetics, University of Cambridge, Cambridge, CB2 3EH, UK

v1 **First published:** 29 Jun 2021, **6**:165
https://doi.org/10.12688/wellcomeopenres.16983.1
Latest published: 29 Jun 2021, **6**:165
https://doi.org/10.12688/wellcomeopenres.16983.1

Abstract

We present a genome assembly based on an individual female *Aphantopus hyperantus*, also known as *Maniola hyperantus* (the ringlet butterfly; Arthropoda; Insecta; Lepidoptera, Nymphalidae), scaffolded using data from a second, unrelated specimen. The genome sequence is 411 megabases in span. The majority of the assembly is scaffolded into 29 chromosomal pseudomolecules, including the Z sex chromosome.

Keywords

Aphantopus hyperantus, Maniola hyperantus, ringlet butterfly, genome sequence, chromosomal

This article is included in the Tree of Life gateway.

Open Peer Review

Reviewer Status ? ?

	Invited Reviewers	
	1	**2**
version 1 29 Jun 2021	? report	? report

1. **Brian A. Counterman**, Auburn University, Auburn, USA

2. **James Walters**, University of Kansas, Lawrence, USA

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

Corresponding author: Mark Blaxter (mark.blaxter@sanger.ac.uk)

Author roles: Mead D: Conceptualization, Investigation, Writing – Review & Editing; **Saccheri I**: Formal Analysis, Investigation, Resources, Writing – Review & Editing; **Yung CJ**: Investigation, Resources, Writing – Review & Editing; **Lohse K**: Formal Analysis, Investigation, Resources, Writing – Review & Editing; **Lohse C**: Resources, Writing – Review & Editing; **Ashmole P**: Formal Analysis, Investigation, Resources, Writing – Review & Editing; **Smith M**: Formal Analysis, Investigation, Methodology, Writing – Review & Editing; **Corton C**: Formal Analysis, Investigation, Methodology, Writing – Review & Editing; **Oliver K**: Formal Analysis, Investigation, Methodology, Supervision, Writing – Review & Editing; **Skelton J**: Formal Analysis, Investigation, Methodology, Writing – Review & Editing; **Betteridge E**: Formal Analysis, Investigation, Methodology, Writing – Review & Editing; **Quail MA**: Data Curation, Funding Acquisition, Investigation, Writing – Review & Editing; **Dolucan J**: Formal Analysis, Investigation, Methodology, Writing – Review & Editing; **McCarthy SA**: Data Curation, Formal Analysis, Investigation, Software, Writing – Review & Editing; **Howe K**: Data Curation, Formal Analysis, Investigation, Software, Supervision, Validation, Visualization, Writing – Review & Editing; **Wood J**: Data Curation, Formal Analysis, Investigation, Software; **Torrance J**: Data Curation, Formal Analysis, Investigation, Software, Supervision, Validation, Writing – Review & Editing; **Tracey A**: Data Curation, Formal Analysis, Investigation, Software, Validation, Writing – Review & Editing; **Whiteford S**: Formal Analysis, Investigation, Methodology, Visualization, Writing – Review & Editing; **Challis R**: Data Curation, Formal Analysis, Investigation, Software, Validation, Writing – Review & Editing; **Durbin R**: Conceptualization, Data Curation, Formal Analysis, Investigation, Software, Validation, Visualization, Writing – Review & Editing; **Blaxter M**: Conceptualization, Data Curation, Formal Analysis, Investigation, Software, Validation, Visualization, 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). SMcC and RD were supported by Wellcome grant 207492. IS was supported by NERC grants NE/N015711/1 and NE/N015797/1. Carrifran Wildwood has support from the Borders Forest Trust and the John Muir Trust, and by donations of time and money from a wide family of generous supporters.
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: Mead D, Saccheri I, Yung CJ *et al.* **The genome sequence of the ringlet, *Aphantopus hyperantus* Linnaeus 1758 [version 1; peer review: 2 approved with reservations]** Wellcome Open Research 2021, **6**:165
https://doi.org/10.12688/wellcomeopenres.16983.1

First published: 29 Jun 2021, **6**:165 https://doi.org/10.12688/wellcomeopenres.16983.1

Species taxonomy

Eukaryota; Metazoa; Arthropoda; Insecta; Lepidoptera; Nymphalidae; Satyrinae; Aphantopus *Aphantopus hyperantus* Linnaeus 1758 (also known as *Maniola hyperantus*) (NCBI txid: 111886).

Introduction

The ringlet, *Aphantopus hyperantus* Linnaeus 1758, also known as *Maniola hyperantus*, is a common butterfly of the northern Palaearctic, found from Ireland to Korea. It is strictly univoltine and overwintering larvae feed on a variety of coarse grasses. In Britain and Ireland it is found south of the Great Glen. It is known as *fàinneag* in Scottish Gaelic and *gweirlöyn y glaw* in Welsh. The fore and hindwings are marked by eyespots, and variation in presence, size and shape of these is under genetic control (Ford, 1945). A dwarf form is found at high elevations in Kerry in Ireland (Huggins, 1960). It is not endangered in Britain and Ireland, with increases in both abundance and distribution measured over the last decade, suggesting that it may be resilient to climate change (Fox *et al.*, 2015). *A. hyperantus* has 29 chromosome pairs (Federley, 2010), and the female is heterogametic (WZ).

Genome sequence report

The genome was sequenced using long-read Pacific Biosciences SEQUEL I platform from a single female *A. hyperantus* collected in Wiltshire (Figure 1A, B). A total of 45-fold coverage in Pacific Biosciences single-molecule long reads (N50 11 kb)

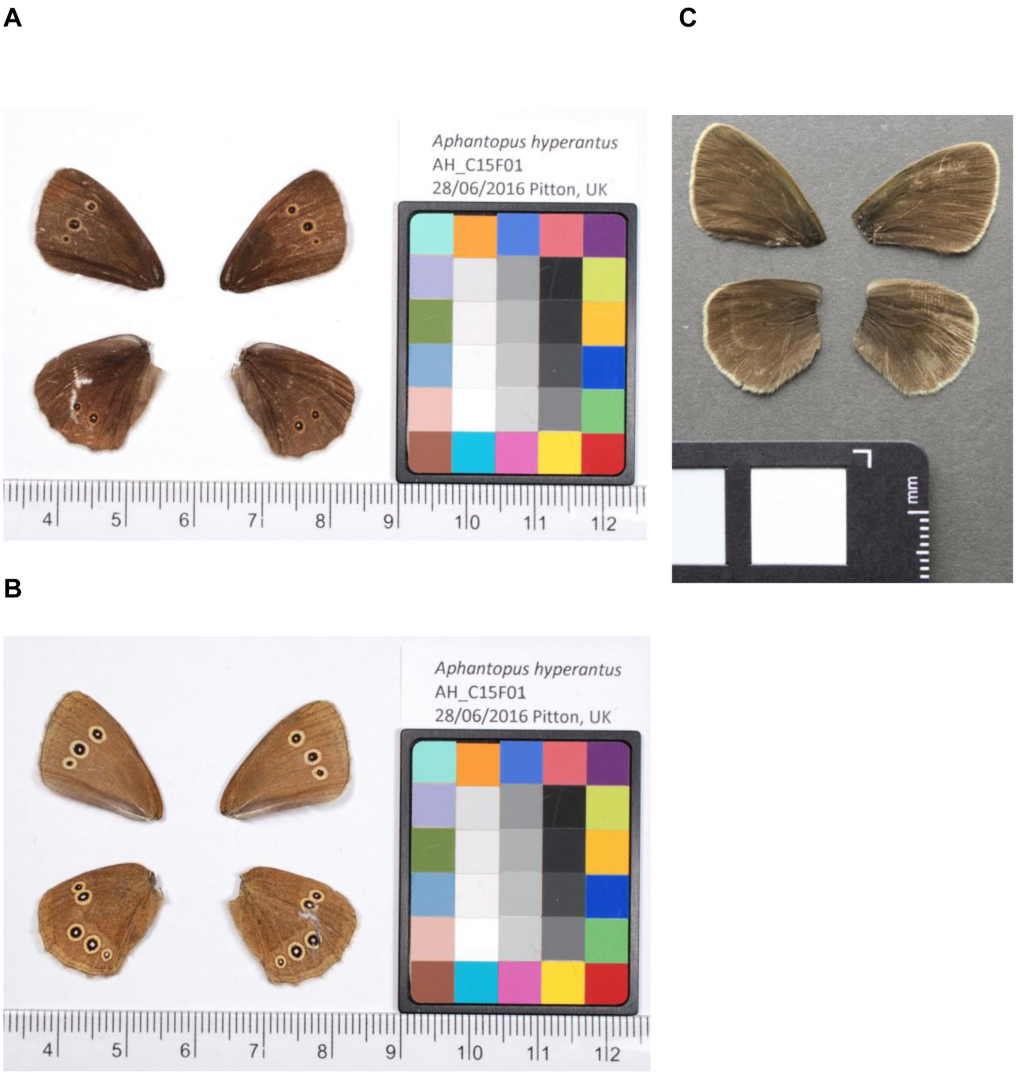

Figure 1. Fore and hind wings of *Aphantopus hyperantus* specimens from which the genome was sequenced. (**A**) Dorsal surface view of wings from specimen AH_C15F01 from Pitton, used to generate Pacific Biosciences and 10X genomics data. (**B**) Ventral surface view of wings from specimen AH_C15F01 from Pitton, used to generate Pacific Biosciences and 10X genomics data. (**C**) Dorsal surface view of wings from specimen UK_AH_1241 from Carrifran Wildwood, Scotland) used to generate Hi-C data. The Carrifran specimen had no visible wingspots.

and 61-fold coverage in 10X Genomics read clouds (from molecules with an estimated N50 of 16 kb) were generated. Primary assembly contigs were scaffolded with chromosome conformation HiC data, generated from a second individual from Southern Scotland (Figure 1C). The final assembly has a total length of 408 Mb in 87 sequence scaffolds with a scaffold N50 of 15.2 Mb (Table 1). The majority, 99.2%, of the assembly sequence was assigned to 29 chromosomal-level scaffolds representing 28 autosomes and the Z sex chromosome (Figure 2–Figure 6; Table 2). These chromosomal scaffolds were mapped to presumed ancestral chromosomal units through comparison to the genome of the diamondback moth *Plutella xylostella*, which is believed to display the ancestral chromosome number (Baxter *et al.*, 2011) (Table 2; Figure 7). *A. hyperantus* has fewer chromosomes than does *P. xylostella*, and two iAntHyp1 scaffolds have dual ancestral chromosome origins (submitted iAphHyp1 chromosome 7 is a fusion of *P. xylostella* chromosomes 20 and 28, and iAphHyp1 chromosome 9 is a fusion of *P. xylostella* chromosomes 16 and 30). iAphHyp1 chromosome 24 has matches to *P. xylostella* chromosome 26 over most of its length, but a short segment at one end has matches to *P. xylostella* chromosome 5. Since only a small section of *P. xylostella* chromosome 5 is translocated, in contrast to the other fusion events which involve entire chromosomes, we refrain from labelling iAphHyp1 chromosome 24 a fusion of two chromosomes.The assembly has a BUSCO

Table 1. Genome data for *Aphantopus hyperantus* iAphHyp1.

Project accession data	
Assembly identifier	iAphHyp1.2
Species	*Aphantopus hyperantus*
Specimen	For main assembly: AH_C15F01/ilAphHyp1 For Hi-C: DTOL8084307/UK_AH_1241/ilAphHyp4
NCBI taxonomy ID	NCBI:txid111886
BioProject	PRJEB36756
Biosample ID	SAMEA994723
Isolate information	female adult (main assembly)
Raw data accessions	
PacificBiosciences SEQUEL I	ERX3338833, ERX3338832, ERX3338741, ERX3338740, ERX3338739, ERX3338738
10X Genomics Illumina	ERX3341582, ERX3341581, ERX3341580, ERX3341579
Hi-C Illumina	ERX5605665-ERX5605667
Genome assembly	
Assembly accession	GCA_902806685.2
Accession of alternate haplotype	GCA_902806615.1
Span (Mb)	408
Number of contigs	472
Contig N50 length (Mb)	2
Number of scaffolds	87
Scaffold N50 length (Mb)	15
Longest scaffold (Mb)	19
BUSCO* genome score	C:98.6%[S:97.5%,D:1.1%],F:0.7%,M:0.7%,n:1013

* BUSCO scores based on the arthropoda_odb10 BUSCO set using v5.0.0. 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/Maniola%20hyperantus/dataset/CADCXM02/busco.

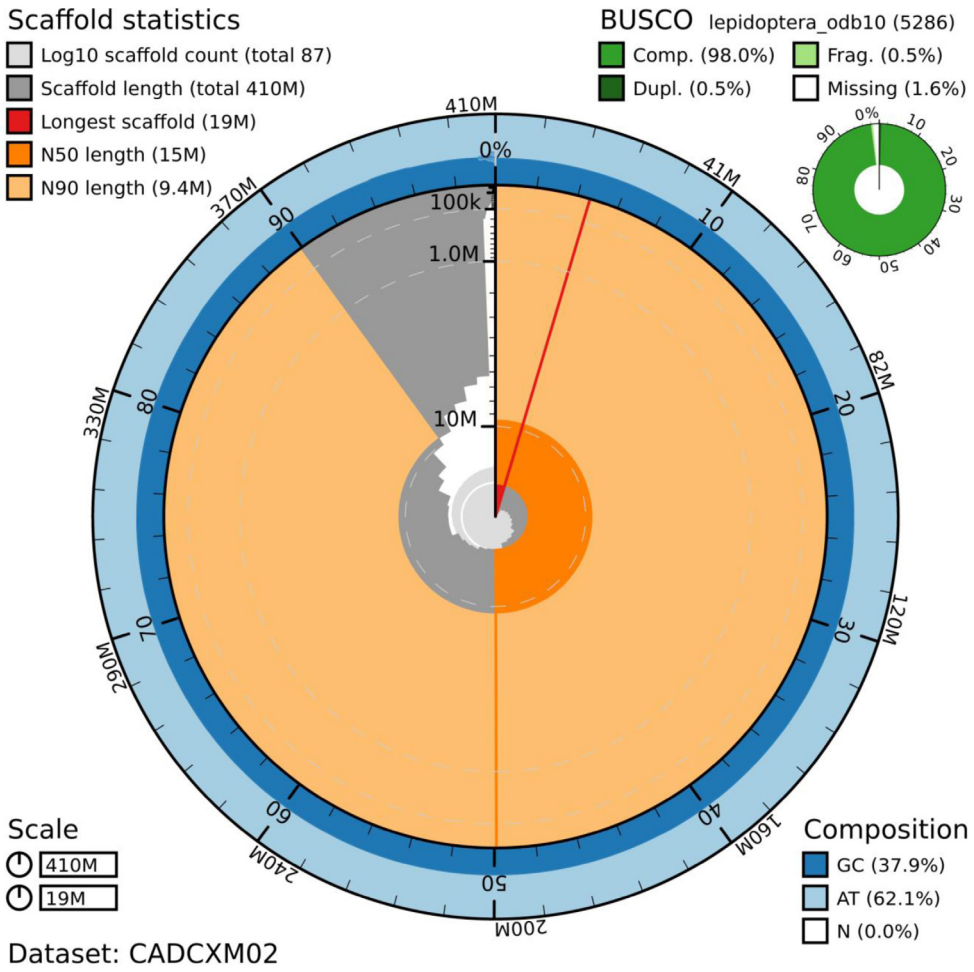

Figure 2. Genome assembly of *Aphantopus hyperantus* iAphHyp1.2. BlobToolKit Snailplot. Snailplot showing N50 metrics and BUSCO scores for the Lepidoptera set of orthologues. Interactive version available at https://blobtoolkit.genomehubs.org/view/Maniola%20hyperantus/dataset/CADCXM02/snail.

v5.0.0 (Simão *et al.*, 2015) completeness of 98.6% using the arthropoda_odb10 reference set. While not fully phased, the assembly deposited is of one haplotype, and secondary haplotype contigs have also been deposited.

Methods

The ringlet specimen AH_C15F01 was collected from Pitton, Wiltshire, South West England, England, SP5 1ED, United Kingdom (latitude 51.0788, longitude -1.7087; Gridref SU205311) on 28/06/2016. Specimen UK_AH_1241 (Figure 1) was collected from Carrifran Wildwood, Scotland (latitude 55.3908° N, longitude -3.3279). Sample AH_C15F01 was supplied as DNA extracted from the abdomen using a phenol-chloroform protocol. DNA was extracted from the Hi-C specimen (iAphHyp4) using a modified magnetic attraction bead method extraction from thoracic tissue. Pacific Biosciences CLR long read and 10X Genomics 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 I and Illumina HiSeq X instruments. HiC data were generated using the Arima v1.0 kit and sequenced on HiSeq X.

Assembly was carried out using Falcon-unzip (falcon-kit 1.1.2) (Chin *et al.*, 2016), haplotypic duplication was identified and removed with purge_dups (Guan *et al.*, 2020) and

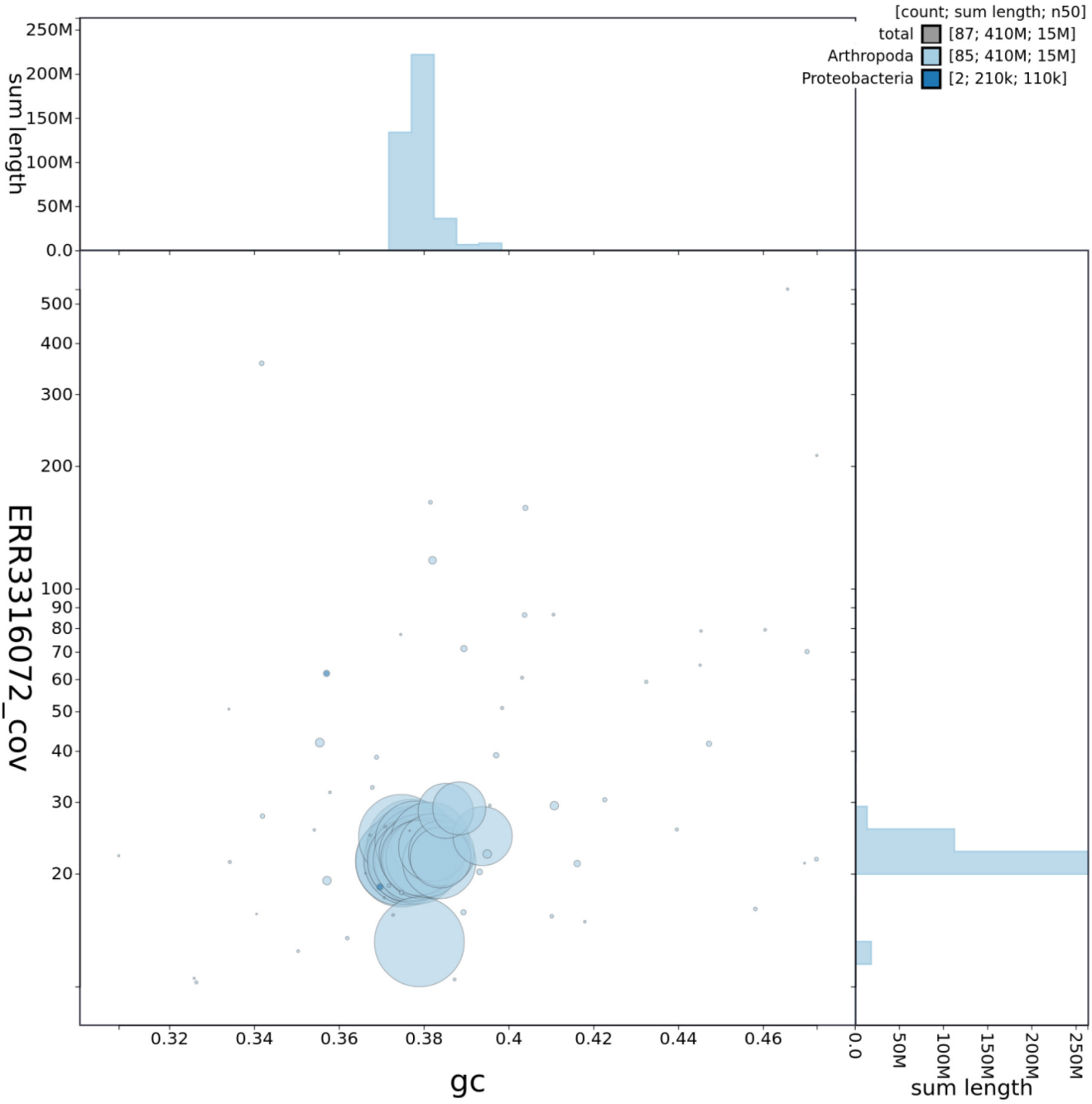

Figure 3. Genome assembly of Aphantopus hyperantus iAphHyp1.2. BlobToolKit GC-coverage plot. Interactive version available at https://blobtoolkit.genomehubs.org/view/iAphHyp1_1/dataset/iAphHyp1_1/blob?plotShape=circle.

a first round of scaffolding carried out with 10X Genomics read clouds using scaff10x. Scaffolding with Hi-C data was carried out using SALSA2. The Hi-C scaffolded assembly was polished with arrow using the PacBio data, then polished with the 10X Genomics Illumina data by aligning to the assembly with longranger align, calling variants with freebayes ((Garrison & Marth, 2012)) and applying homozygous non-reference edits using bcftools consensus. Hi-C data were visualized in

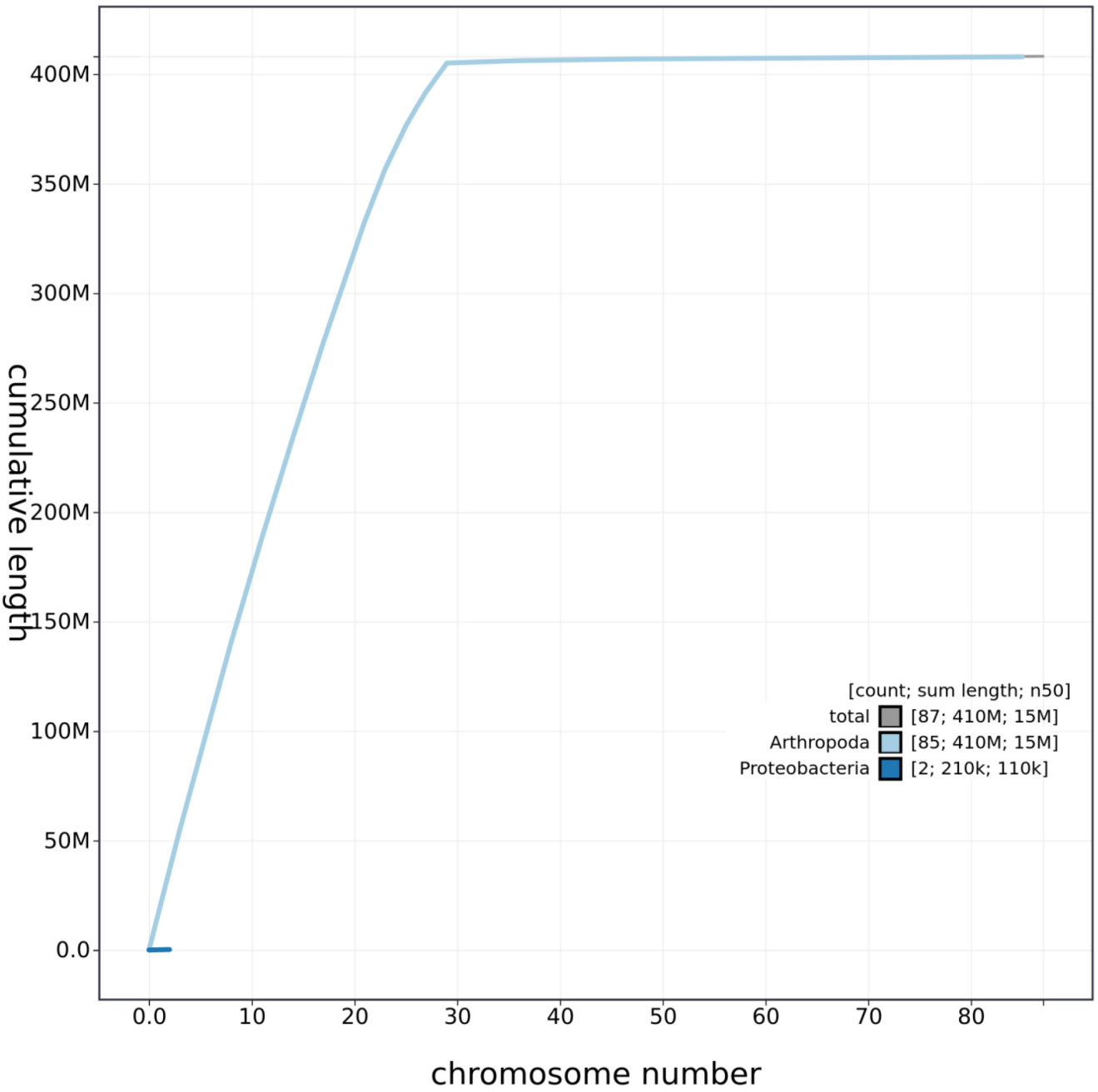

Figure 4. Genome assembly of *Aphantopus hyperantus* iAphHyp1.2. Cumulative sequence plot. Interactive version available at https://blobtoolkit.genomehubs.org/view/iAphHyp1_1/dataset/iAphHyp1_1/cumulative.

pretext and HiGlass. Two rounds of Illumina polishing were applied. The assembly was checked for contamination and manually corrected using the gEVAL system (Chow *et al.*, 2016). During manual curation, 143 breaks, 231 joins and 40 removals of erroneously duplicated sequence regions were made. These changes were visualised using circos plots (Krzywinski *et al.*, 2009) (Figure 5, Figure 6). The genome was analysed and BUSCO scores generated within the BlobToolKit environment

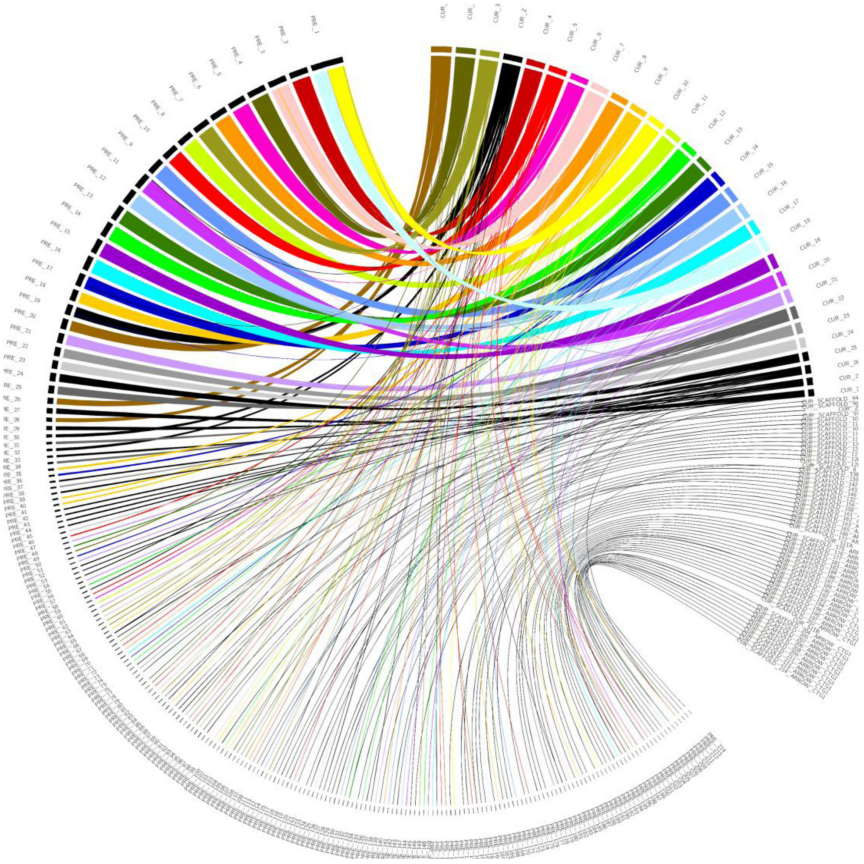

Figure 5. Genome assembly of *Aphantopus hyperantus* iAphHyp1. Curation improves the assembly. Circos plot (Krzywinski *et al.* 2009) comparison of the precurated assembly (left) to the curated (right) in circos, illustrating the changes made during curation. Scaffolds are linked through shared nucleotide sequence identity based on filtered nucmer results (100% identity, min length 100 kb, adjusted for small contigs < 100 kb to an alignment of >30% of their length). The curated assembly has joined large scaffolds, placed short ones and removed remaining haplotypic duplicate segments, leaving 29 large (chromosomal) scaffolds (totalling 405.0 Mb; 99.2% of the assembly) and 58 smaller, unplaced ones (3.1 Mb; 0.8%).

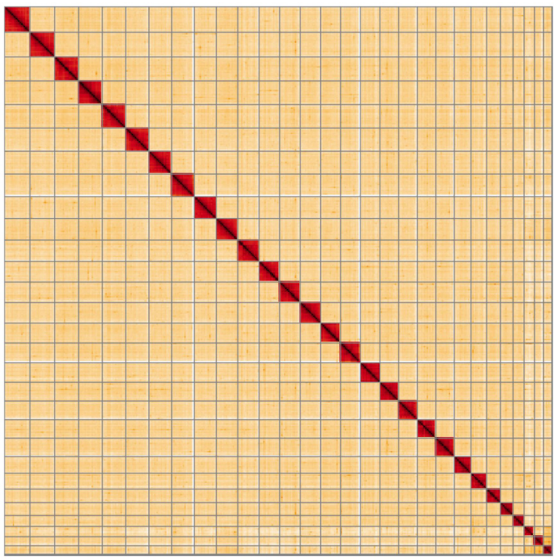

Figure 6. Genome assembly of *Aphantopus hyperantus* iAphHyp1.2. Hi-C contact map of 29 chromosomal scaffolds. Hi-C contact map of the *A. hyperantus* assembly, visualized in HiGlass.

Table 2. Chromosomal pseudomolecules in the genome assembly of *Aphantopus hyperantus* iAphHyp1.2.

INSDC accession	Chromosome numbered by size	Chromosomes numbered by orthology to *Plutella xylostella**	Size (Mb)
LR761647.1	Chromosome 1	Chr04	18.856
LR761648.1	Chromosome 2	Chr05	18.294
LR761649.1	Chromosome 3	Chr10	17.757
LR761651.1	Chromosome 4	Chr15	17.319
LR761652.1	Chromosome 5	Chr22	17.143
LR761653.1	Chromosome 6	Chr13	16.908
LR761654.1	Chromosome 7	Chr28/Chr20	16.699
LR761655.1	Chromosome 8	Chr12	16.178
LR761656.1	Chromosome 9	Chr30/Chr16	15.949
LR761657.1	Chromosome 10	Chr06	15.781
LR761658.1	Chromosome 11	Chr23	15.241
LR761659.1	Chromosome 12	Chr09	15.230
LR761660.1	Chromosome 13	Chr17	15.130
LR761661.1	Chromosome 14	Chr29	14.784
LR761662.1	Chromosome 15	Chr08	14.633
LR761663.1	Chromosome 16	Chr18	14.440
LR761664.1	Chromosome 17	Chr21	13.807
LR761665.1	Chromosome 18	Chr25	13.796
LR761666.1	Chromosome 19	Chr03	13.765
LR761667.1	Chromosome 20	Chr19	13.590
LR761668.1	Chromosome 21	Chr07	12.342
LR761669.2	Chromosome 22	Chr14	11.670
LR761670.1	Chromosome 23	Chr27	10.071
LR761671.1	Chromosome 24	Chr26**	9.443
LR761672.1	Chromosome 25	Chr31	8.138
LR761673.1	Chromosome 26	Chr24	7.670
LR761674.1	Chromosome 27	Chr02	6.681
LR761675.1	Chromosome 28	Chr11	6.197
LR761650.1	Chromosome Z	Chr1z	17.540
n/a	unplaced		3.085

* Chr# as in *P. xylostella*; where two Chr# are given, the *A. hyperantus* chromosome is likely to be a recent, simple fusion of two *P. xylostella* chromosomes.

** LR761671.1, iAphHyp1 chromosome 24, shares BUSCO marker genes with *P. xylostella* chromosome 26 over most of its length, but a short segment at one end shares BUSCO loci with *P. xylostella* chromosome 5. See text for discussion.

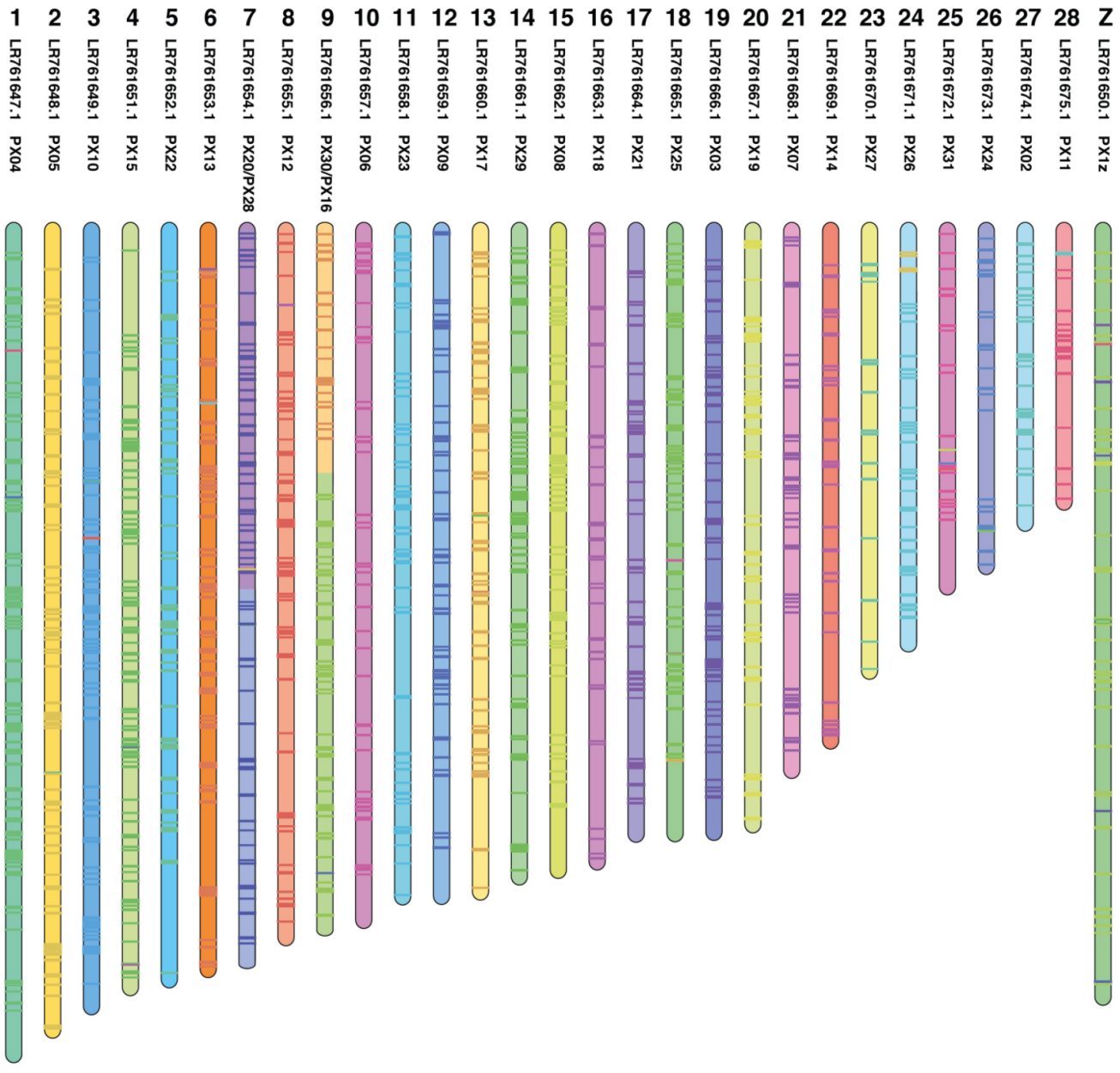

Figure 7. Karyotype map of the *A. hyperantus* iAphHyp1 assembly mapped to the karyotype of *Plutella xylostella*. Visualised with BUSCO_karyotyping (https://github.com/swomics/BUSCO_karyotyping/). High-resolution PDF version with identification of specific BUSCO assignments available at https://github.com/swomics/BUSCO_karyotyping/blob/master/Px_output.pdf.

(Challis *et al.*, 2020) (Figure 2–Figure 4). Table 3 contains a list of all software tool versions used, where appropriate. Chromosomal pseudomolecules were analysed for synteny compared to the *P. xylostella* genome sequence (Baxter *et al.*, 2011), which is believed to display the ancestral lepidopteran karyotype (n=31), using a custom script that exploits conservation of BUSCO (Simão *et al.*, 2015) marker genes between chromosomes in different species (Whiteford *et al.*, 2019)

Table 3. Software tools used.

Software tool	Version	Source
Falcon-unzip	falcon-kit 1.2.2	(Chin *et al.*, 2016)
purge_dups	1.0.0	(Guan *et al.*, 2020)
SALSA2	2.2	(Ghurye *et al.*, 2018)
scaff10x	4.2	https://github.com/wtsi-hpag/Scaff10X
arrow	GenomicConsensus 2.3.3	https://github.com/PacificBiosciences/GenomicConsensus
longranger align	2.2.2	https://support.10xgenomics.com/genome-exome/software/pipelines/latest/advanced/other-pipelines
freebayes	v1.1.0-3-g961e5f3	(Garrison & Marth, 2012)
bcftools consensus	1.9	http://samtools.github.io/bcftools/bcftools.html
gEVAL	2016	(Chow *et al.*, 2016)
HiGlass	1.11.6	(Kerpedjiev *et al.*, 2018)
PretextView	0.0.4	https://github.com/wtsi-hpag/PretextView
BlobToolKit	2.5	(Challis *et al.*, 2020)
nucmer	4	http://mummer.sourceforge.net/
dot	release02/2020	https://github.com/dnanexus/dot
circos		(Krzywinski *et al.*, 2009)
pretext	0.1.0	https://github.com/wtsi-hpag/PretextView
BUSCO	3.0.2	(Simão *et al.*, 2015) https://busco-archive.ezlab.org/v3/

(available at https://github.com/swomics/BUSCO_karyotyping/) (Figure 7).

Data availability

Underlying data

European Nucleotide Archive: Aphantopus hyperantus (ringlet butterfly) genome assembly, iAphHyp1. Accession number PRJEB36756.

The genome sequence is released openly for reuse. The *A. hyperantus* genome sequencing initiative is part of the Wellcome Sanger Institute's "25 genomes for 25 years" project. It is also part of the DNA Zoo Project and the Darwin Tree of Life (DToL) project. The wings of the sequenced specimens have been preserved and will be submitted to the Natural History Museum London for long term curation. All raw data and the assembly have been deposited in the ENA. 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.

Acknowledgements

We thank Mike Stratton and Julia Wilson for their continuing support for the 25 genomes for 25 years project. We thank Simon Martin for the photograph of UK_AH_1241. We thank Chloe Leech for administrative support.

References

Baxter SW, Davey JW, Johnston JS, *et al.*: **Linkage Mapping and Comparative Genomics Using next-Generation RAD Sequencing of a Non-Model Organism.** *PloS One.* 2011; **6**(4): e19315.
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

Chin CS, Peluso P, Sedlazeck FJ, *et al.*: **Phased Diploid Genome Assembly with Single-Molecule Real-Time Sequencing.** *Nat Methods.* 2016; **13**(12): 1050–54.
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 FINNLÄN-DISCHER LEPIDOPTEREN: I.**

RHOPALOCERA. *Hereditas.* 2010; **24**(4): 397–464.
Publisher Full Text

Ford EB: **Butterflies (Collins New Naturalist Library, Book 1).** London: Collins. 1945.

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 [q-bio.GN].* arXiv. 2012.
Reference Source

Ghurye J, Rhie A, Walenz BP, *et al.*: **Integrating Hi-C Links with Assembly Graphs for Chromosome-Scale Assembly.** *bioRxiv.* 2018.
Publisher Full Text

Guan D, McCarthy SA, Wood J, *et al.*: **Identifying and Removing Haplotypic Duplication in Primary Genome Assemblies.** *Bioinformatics.* 2020; **36**(9): 2896–2898.
PubMed Abstract | Publisher Full Text | Free Full Text

Huggins HC: **A Naturalist in the Kingdom of Kerry.** *Proceedings & Transactions of the South London Entomological & Natural History Society.* 1960; **1959**: 176–83.

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

Krzywinski M, Schein J, Birol I, *et al.*: **Circos: An Information Aesthetic for Comparative Genomics.** *Genome Res.* 2009; **19**(9): 1639–45.
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

Whiteford S, van't Hof AE, Krishna R, *et al.*: **Recovering Individual Haplotypes and a Contiguous Genome Assembly from Pooled Long Read Sequencing of the Diamondback Moth (Lepidoptera: Plutellidae).** *bioRxiv.* 2019.
Publisher Full Text

Open Peer Review

Current Peer Review Status: ? ?

Version 1

Reviewer Report 15 September 2021

https://doi.org/10.21956/wellcomeopenres.18751.r45788

?

James Walters
Department of Ecology and Evolutionary Biology, University of Kansas, Lawrence, KS, USA

This manuscript reports a genome assembly for *Aphantopus hyperantus*, presenting a very high quality data resource. The reported assembly is highly contiguous, with assembly to chromosome, which is still currently notable among lepidopteran genomes despite the increasingly common use of HiC data to achieve chromosomal-length assemblies. A variety of quality metrics are applied to demonstrate the lack of contamination, completeness of assembly, and conservation of synteny to other Lepidoptera. The methods are succinctly but clearly and sufficiently described, and the tabulation of software versions is also an important inclusion for the sake of reproducibility.

While this manuscript and the reported results appear generally robust and well-executed, I still have a couple of points of concern:

1. Figure 3 would benefit from a further explanation. It is not a plot typically encountered in genome analysis, though if I understand it correctly, it seems it should be. However, since the axes are not explicitly explained, nor the significance of bubble size, I am left to surmise that this is assembly coverage on Y versus GC content on X, with bubbles representing scaffolds and sized proportional to length.
   Furthermore, it might be worth pointing out or confirming that the one bubble at half coverage relative to all others is the Z, as confirmed via comparison to Plutella, which I assume to be the case.

2. Are there any other data or analyses that could be employed to corroborate the genome size and karyotype? Could you apply kmer methods to the Illumina data to estimate genome size? Are there any published karyotypes for the species? There may not be, but if so, it would be a nice confirmation that the number of reconstructed chromosomes corresponds to expectations from direct observation.

3. I am struck that there is no analysis or discussion of any potential W chromosome sequence in this assembly. Since the data generated was from a female, there is presumably sequencing reads from the entire W chromosome contained within this data set. Some consideration of the W chromosome seems merited. Was it intentionally excluded during

the assembly process? Do you expect it to be found, in pieces, among the unassigned contigs? Was it included in the “secondary” haplotype contigs?

4. The motivation for generating this data set is never explained. Why sequence this particular species? There is some passing reference to the “25 genomes in 25 years”, but that isn’t really any kind of particular motivation for this species. The introduction gives some interesting natural history of the species, but doesn’t offer any indication as to why obtaining the genome of this species is of any particular interest.

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: Lepidopteran evolutionary genomics

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.

Reviewer Report 26 July 2021

https://doi.org/10.21956/wellcomeopenres.18751.r44878

?

Brian A. Counterman
Department of Biological Sciences, Auburn University, Auburn, AL, USA

The manuscript “The genome sequence of the ringlet, *Aphantopus hyperantus* Linnaeus 1758” presents the genome assembly of the ringlet butterfly. The authors provide a comprehensive analysis of the assembly, showing that 99.2% of the genome sequence is included in 29 scaffolds that correspond to the 29 expected chromosomes. Further, they show that the assembly includes 98.6% of the BUSCO gene set, which is impressive when compared to other lepidoptera genomes. The voucher images of the sequenced specimens are much appreciated so that the genome can

be connected to specific phenotypes. The paper is clearly written, and I am able to clearly follow the methods used to produce the genome assembly. The table that details each of the software packages is also appreciated. Overall, from the details provided it appears the assembly and analyses were performed correctly. The genome is sure to be a valuable resource for lepidopterists and future genomic studies.
Following are a few suggestions or concerns:

1. Will the voucher specimens (e.g. wings) be deposited into a curated collection (e.g. museum collection) that will be accessible for others to access?

2. I am not convinced that Figures 3, 4, and 6 need to be included in the paper. These figures are not individually referred to or directly discussed in the text. Figure 3 is difficult to interpret without further explanation in the figure caption. Is each bubble a separate chromosome? Is there a point the authors are trying to illustrate about GC content with this figure? If so, please state the purpose in the main text. If not, I would consider removing Figure 3. Figure 4 also needs further explanation in the caption. For example, what is the Proteobacteria that is shown? And what do the values reflect in the brackets? Figure 6 is referred to as a Circos plot on page 7, which I do not believe is correct. Again, more explanation of this figure is needed in the caption, and discussion of its relevance is needed in the main text, or the authors should removing the figure.

3. There is inconsistency on when an in-text citation is given for software packages. For example, on page 5 in-text citations are given for Falcon-unzip and purge_dups, but on page 6 no in-text citations are given for Salsa2 or arrow. I would urge the authors to add the in-text citations for all software packages when they are mentioned.

4. It is unclear what is meant by “polishing” the genome. This is mentioned a few places throughout the manuscript, but it is unclear what edits or changes to the genome are involved at each of these “polishing” steps. Some brief further details of what each “polishing” step entails would be useful for the reader.

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: evolutionary genomics; 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.