



DATA NOTE

The genome sequence of the pebble prominent, *Notodonta ziczac* (Linnaeus, 1758) [version 1; peer review: 2 approved]

Douglas Boyes¹⁺,
 University of Oxford and Wytham Woods Genome Acquisition Lab,
 Darwin Tree of Life Barcoding collective,
 Wellcome Sanger Institute Tree of Life programme,
 Wellcome Sanger Institute Scientific Operations: DNA Pipelines collective,
 Tree of Life Core Informatics collective, Thomas Lewin ²,
 Darwin Tree of Life Consortium

¹UK Centre for Ecology and Hydrology, Wallingford, Oxfordshire, UK

²Department of Zoology, University of Oxford, Oxford, UK

+ Deceased author

V1 First published: 25 Mar 2022, 7:111
<https://doi.org/10.12688/wellcomeopenres.17789.1>
 Latest published: 25 Mar 2022, 7:111
<https://doi.org/10.12688/wellcomeopenres.17789.1>

Abstract

We present a genome assembly from an individual male *Notodonta ziczac* (the pebble prominent; Arthropoda; Insecta; Lepidoptera; Notodontidae). The genome sequence is 352 megabases in span. The majority of the assembly (99.66%) is scaffolded into 31 chromosomal pseudomolecules, with the Z sex chromosome assembled. The mitochondrial genome was also assembled, and is 18.3 kilobases in length.

Keywords

Notodonta ziczac, pebble prominent, genome sequence, chromosomal, Lepidoptera



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

Open Peer Review

Approval Status  

	1	2
version 1		
25 Mar 2022	view	view

1. **Akito Y. Kawahara** , University of Florida, Gainesville, USA

2. **Christopher M. Ward**, Australian Wine Research Institute, Adelaide, Australia

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

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

Author roles: **Boyes D:** Investigation, Resources; **Lewin T:** Writing – Original Draft Preparation;

Competing interests: No competing interests were disclosed.

Grant information: This work was supported by Wellcome through core funding to the Wellcome Sanger Institute [206194, <https://doi.org/10.35802/206194>] and the Darwin Tree of Life Discretionary Award [218328, <https://doi.org/10.35802/218328>]. *The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.*

Copyright: © 2022 Boyes D *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: Boyes D, University of Oxford and Wytham Woods Genome Acquisition Lab, Darwin Tree of Life Barcoding collective *et al.* **The genome sequence of the pebble prominent, *Notodonta ziczac* (Linnaeus, 1758) [version 1; peer review: 2 approved]** Wellcome Open Research 2022, 7:111 <https://doi.org/10.12688/wellcomeopenres.17789.1>

First published: 25 Mar 2022, 7:111 <https://doi.org/10.12688/wellcomeopenres.17789.1>

Species taxonomy

Eukaryota; Metazoa; Ecdysozoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Lepidoptera; Glossata; Ditrysia; Noctuoidea; Notodontidae; Notodontinae; Notodonta; *Notodonta ziczac* (Linnaeus, 1758) (NCBI:txid988002).

Background

The pebble prominent (*Notodonta ziczac*) is a moth of the family Notodontidae. A typical specimen's wingspan is 42 to 52 mm; its forewings are primarily ochreous brown, with the apical area carrying a grey, pebble-shaped marking from which its common name is drawn (Skinner & Wilson, 2009). Its species name *ziczac*, from the German zickzack, meaning zigzag, comes from the humps on its caterpillars' sixth, seventh and twelfth segments and the posture it assumes at rest, which creates a zigzag-like pattern (Emmet, 1991; Newman, 1869).

It can be found across the Palearctic region, from North Africa to China, within which the species is sometimes divided into three subspecies (Schintlmeister, 2008). It is widely distributed throughout Britain and Ireland, with observations most frequent in southern England and Wales (Randle *et al.*, 2019). Double-brooded in the south and single-brooded in the north, it usually feeds on willow (*Salix* spp.), and is less commonly observed on poplar (*Populus* spp.) (Skinner & Wilson, 2009). It can be found in almost any kind of habitat containing its host plants, including in urban environments (Schintlmeister, 2008). After overwintering as a pupa, the emergence of its first generation currently peaks in May in the UK; this is a notable advancement of several weeks since observations of the 1970s (Randle *et al.*, 2019; South, 1977). Its abundance in the UK declined sharply from 1970 to 2016, and shows a consistent downward trend (Randle *et al.*, 2019).

Genome sequence report

The genome was sequenced from one male *N. ziczac* (Figure 1) collected from Wytham Woods, Oxfordshire (biological vice-county: Berkshire), UK (latitude 51.772, longitude -1.338). A total of 38-fold coverage in Pacific Biosciences single-molecule long reads and 127-fold coverage in 10X Genomics read clouds were generated. Primary assembly contigs were scaffolded with chromosome conformation Hi-C data. Manual assembly curation corrected 3 missing/misjoins.

The final assembly has a total length of 352 Mb in 31 sequence scaffolds with a scaffold N50 of 12.7 Mb (Table 1). The majority of the assembly sequence (99.99%) 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 v5.2.2 (Manni *et al.*, 2021) completeness of 98.9% (single 98.4%, duplicated 0.5%) using the lepidoptera_odb10 reference set. While not fully phased, the assembly deposited is of one haplotype. Contigs corresponding to the second haplotype have also been deposited.

Methods

Sample acquisition and DNA extraction

A single male *N. ziczac* (ilNotZicz1) was collected from Wytham Woods, Oxfordshire (biological vice-county: Berkshire), UK (latitude 51.772, longitude -1.338) by Douglas Boyes, UKCEH, using a light trap in woodland. The sample was identified by the same individual and preserved on dry ice.

DNA was extracted at the Tree of Life laboratory, Wellcome Sanger Institute. The ilNotZicz1 sample was weighed and dissected on dry ice with tissue set aside for Hi-C sequencing. Thorax tissue was cryogenically disrupted to a fine powder using a Covaris cryoPREP Automated Dry Pulveriser, receiving multiple impacts. Fragment size analysis of 0.01–0.5 ng of DNA was then performed using an Agilent FemtoPulse. High molecular weight (HMW) DNA was extracted using the Qiagen MagAttract HMW DNA extraction kit. Low molecular weight DNA was removed from a 200-ng aliquot of extracted DNA using 0.8X AMPure XP purification kit prior to 10X Chromium sequencing; a minimum of 50 ng DNA was submitted for 10X sequencing. HMW DNA was sheared into an average fragment size between 12–20 kb in a Megaruptor 3 system with speed setting 30. Sheared DNA was purified by solid-phase reversible immobilisation using AMPure PB beads with a 1.8X ratio of beads to sample to remove the shorter fragments and concentrate the DNA sample. The concentration of the



Figure 1. Image of the *Notodonta ziczac* (ilNotZicz1) specimen taken prior to preservation and processing. Specimen shown next to FluidX storage tube, 43.9 mm in length.

Table 1. Genome data for *Notodonta ziczac*, ilNotZicz1.1.

Project accession data	
Assembly identifier	ilNotZicz1.1
Species	<i>Notodonta ziczac</i>
Specimen	ilNotZicz1
NCBI taxonomy ID	NCBI:txid988002
BioProject	PRJEB46845
BioSample ID	SAMEA7746619
Isolate information	Male, thorax (genome assembly), head (Hi-C)
Raw data accessions	
PacificBiosciences SEQUEL II	ERR6939251
10X Genomics Illumina	ERR6688606-ERR6688609
Hi-C Illumina	ERR6688605
Genome assembly	
Assembly accession	GCA_918843915.1
Accession of alternate haplotype	GCA_918843885.1
Span (Mb)	352
Number of contigs	55
Contig N50 length (Mb)	11.7
Number of scaffolds	51
Scaffold N50 length (Mb)	12.7
Longest scaffold (Mb)	14.9
BUSCO* genome score	C:98.9%[S:98.4%,D:0.4%],F:0.3%,M:0.9%,n:5286

*BUSCO scores based on the lepidoptera_odb10 BUSCO set using v5.2.2. C= complete [S= single copy, D=duplicated], F=fragmented, M=missing, n=number of orthologues in comparison. A full set of BUSCO scores is available at <https://blobtoolkit.genomehubs.org/view/ilNotZicz1.1/dataset/CAKKN01.1/busco>.

sheared and purified DNA was assessed using a Nanodrop spectrophotometer and Qubit Fluorometer and Qubit dsDNA High Sensitivity Assay kit. Fragment size distribution was evaluated by running the sample on the FemtoPulse system.

Sequencing

Pacific Biosciences HiFi circular consensus and 10X Genomics Chromium read cloud sequencing libraries were constructed according to the manufacturers' instructions. Sequencing was performed by the Scientific Operations core at the Wellcome Sanger Institute on Pacific Biosciences SEQUEL II (HiFi) and Illumina NovaSeq 6000 (10X) instruments. Hi-C data were

generated from head tissue using the Arima Hi-C+ kit and sequenced on NovaSeq 6000.

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 assembly with longranger align, calling variants with freebayes (Garrison & Marth, 2012). The assembly was then scaffolded with Hi-C data (Rao *et al.*, 2014) using SALSA2 (Ghurye *et al.*, 2019). The assembly was checked for contamination as

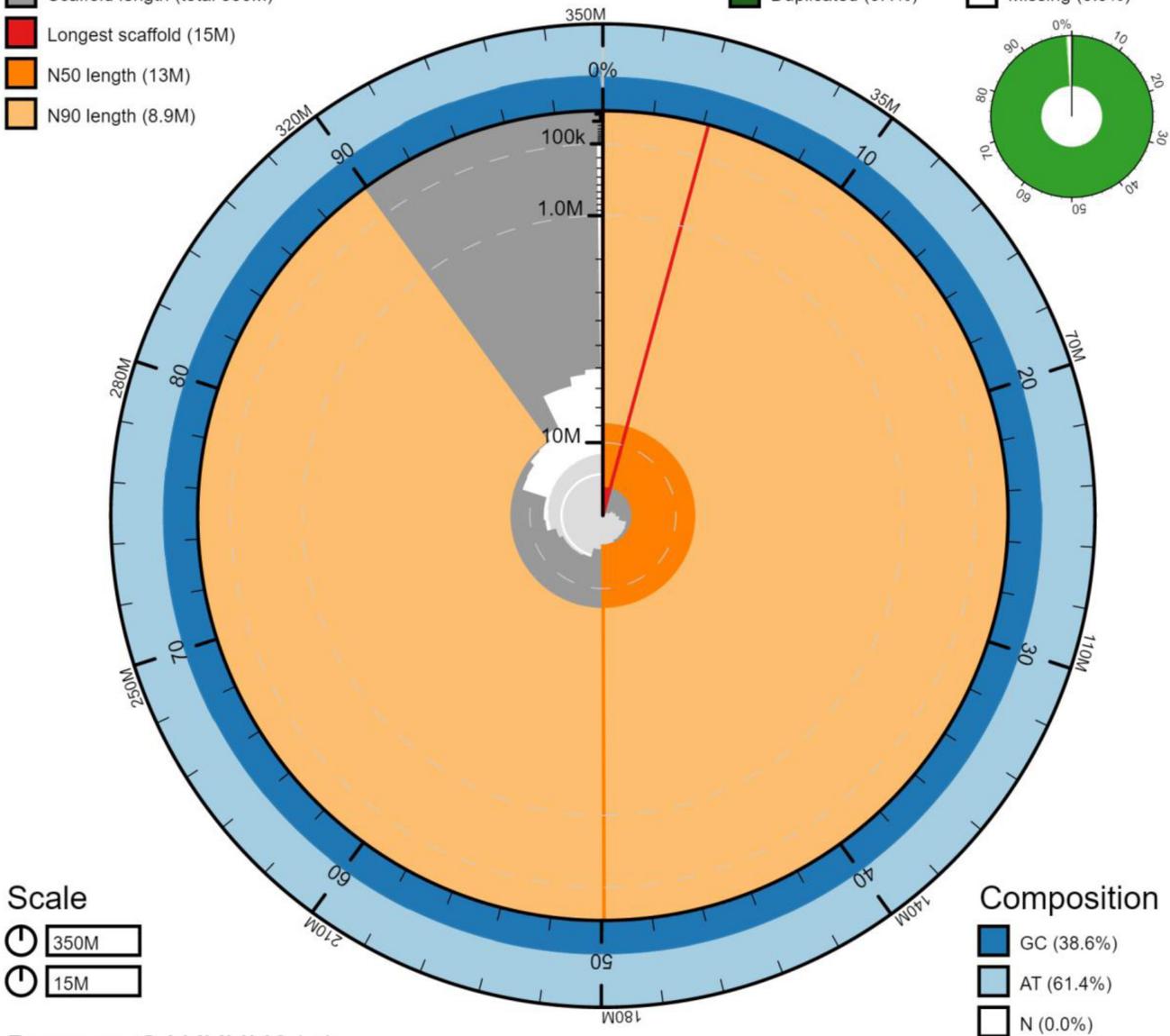
Scaffold statistics

- Log10 scaffold count (total 52)
- Scaffold length (total 350M)
- Longest scaffold (15M)
- N50 length (13M)
- N90 length (8.9M)

BUSCO

lepidoptera_odb10 (5286)

- Complete (98.9%)
- Fragmented (0.3%)
- Duplicated (0.4%)
- Missing (0.9%)



Scale

- 350M
- 15M

Composition

- GC (38.6%)
- AT (61.4%)
- N (0.0%)

Dataset: CAKKNU01.1

Figure 2. Genome assembly of *Notodonta ziczac*, iNotZicz1.1: 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 352,061,436 bp assembly. The distribution of scaffold lengths is shown in dark grey with the plot radius scaled to the longest scaffold present in the assembly (14,880,619 bp, shown in red). Orange and pale-orange arcs show the N50 and N90 scaffold lengths (12,722,079 and 8,864,530 bp), respectively. The pale grey spiral shows the cumulative scaffold count on a log scale with white scale lines showing successive orders of magnitude. The blue and pale-blue area around the outside of the plot shows the distribution of GC, AT and N percentages in the same bins as the inner plot. A summary of complete, fragmented, duplicated and missing BUSCO genes in the lepidoptera_odb10 set is shown in the top right. An interactive version of this figure is available at <https://blobtoolkit.genomehubs.org/view/iNotZicz1.1/dataset/CAKKNU01.1/snail>.

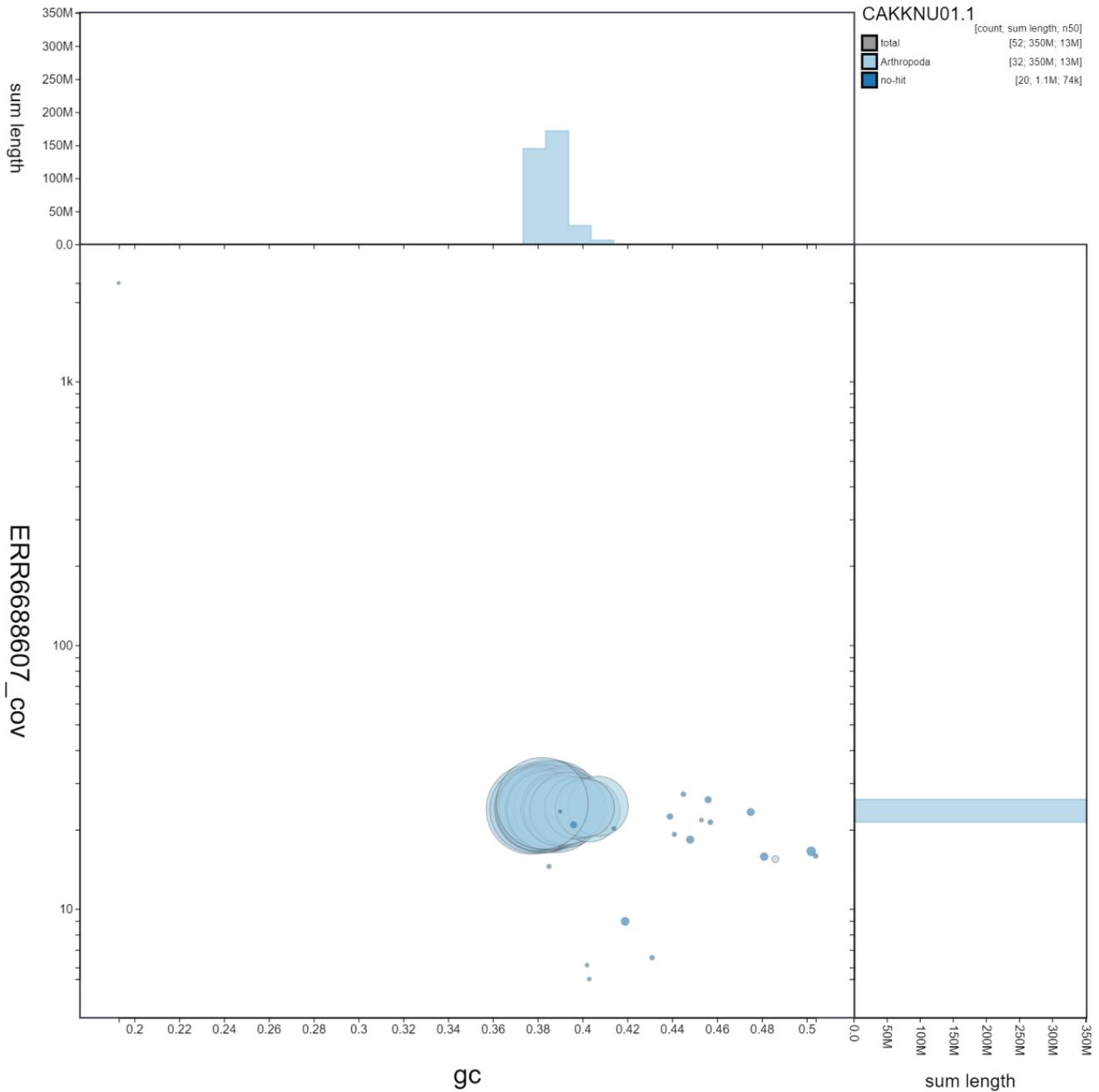


Figure 3. Genome assembly of *Notodonta ziczac*, ilNotZicz1.1: GC coverage. BlobToolKit GC-coverage plot. Scaffolds are coloured by phylum. Circles are sized in proportion to scaffold length. Histograms show the distribution of scaffold length sum along each axis. An interactive version of this figure is available at <https://blobtoolkit.genomehubs.org/view/ilNotZicz1.1/dataset/CAKGNU01.1/blob>.

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

annotation using MitoFinder (Allio *et al.*, 2020). The genome was analysed and BUSCO scores generated within the BlobToolKit environment (Challis *et al.*, 2020). Table 3 contains a list of all software tool versions used, where appropriate.

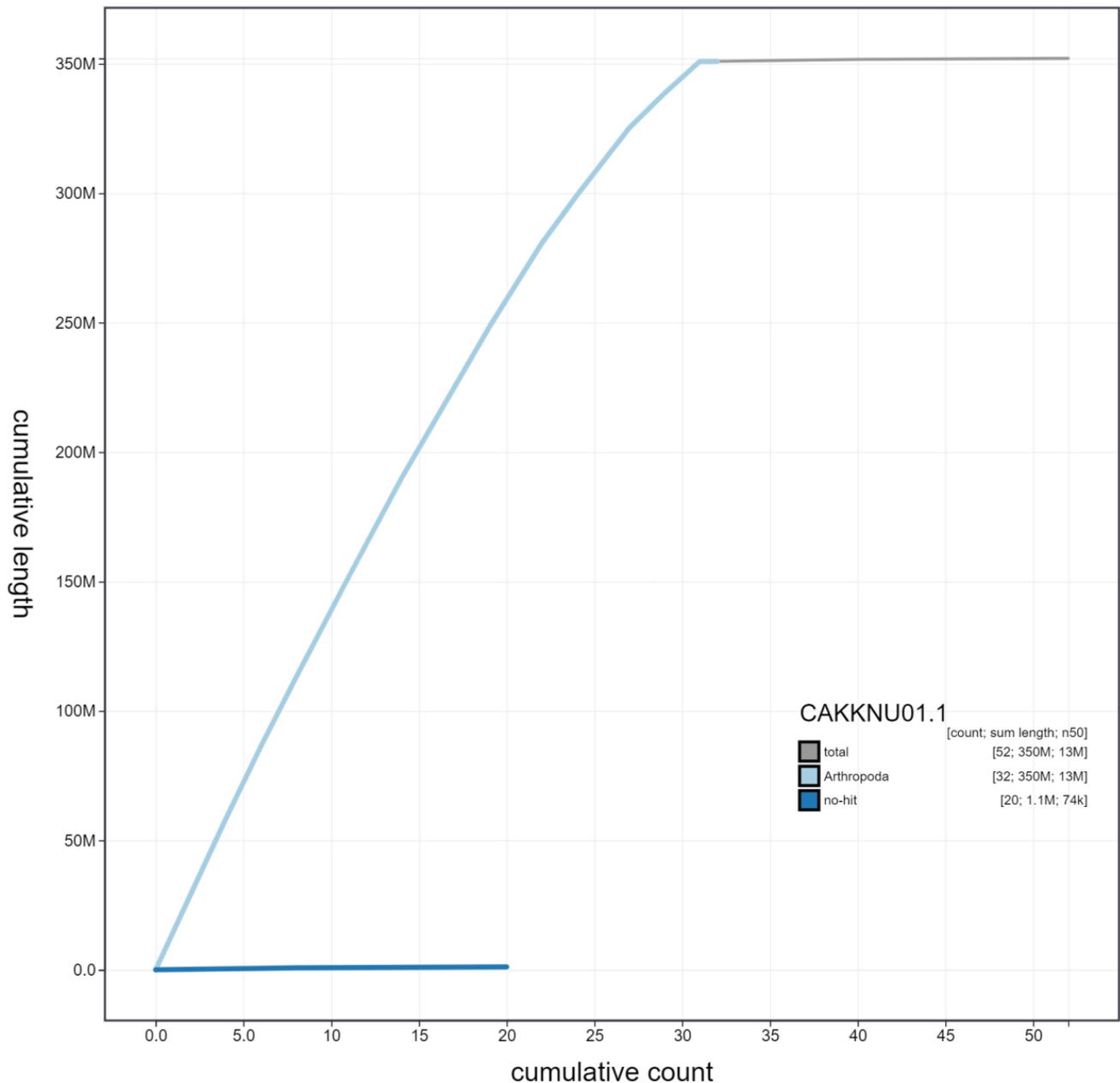


Figure 4. Genome assembly of *Notodonta ziczac*, ilNotZicz1.1: cumulative sequence. BlobToolKit cumulative sequence plot. The grey line shows cumulative length for all scaffolds. Coloured lines show cumulative lengths of scaffolds assigned to each phylum using the buscodegenes taxrule. An interactive version of this figure is available at <https://blobtoolkit.genomehubs.org/view/ilNotZicz1.1/dataset/CAKKNU01.1/cumulative>.

Ethics/compliance issues

The materials that have contributed to this genome note have been supplied by a Darwin Tree of Life Partner. The submission of materials by a Darwin Tree of Life Partner is subject to the [Darwin Tree of Life Project Sampling Code of Practice](#).

By agreeing with and signing up to the Sampling Code of Practice, the Darwin Tree of Life Partner agrees they will meet the legal and ethical requirements and standards set out within this document in respect of all samples acquired for, and supplied to, the Darwin Tree of Life Project. Each transfer

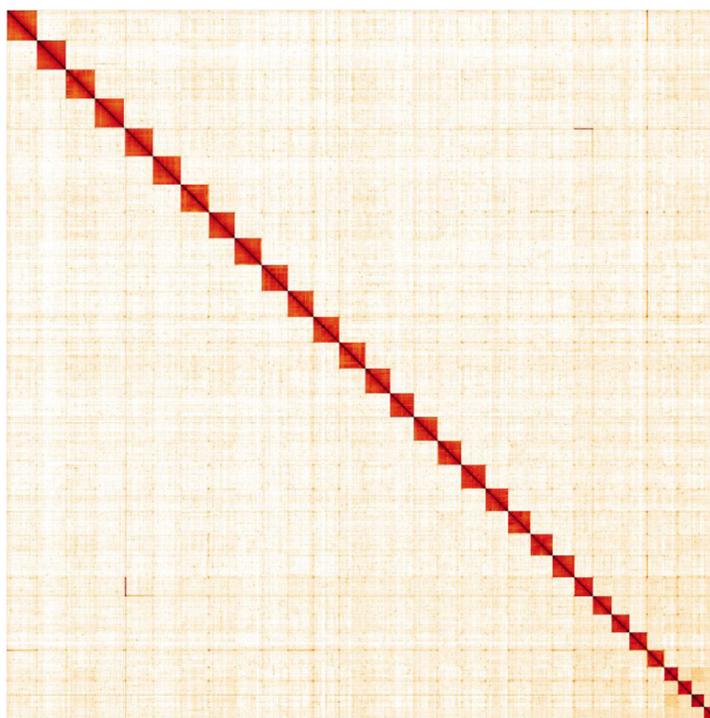


Figure 5. Genome assembly of *Notodonta ziczac*, ilNotZicz1.1: Hi-C contact map. Hi-C contact map of the ilNotZicz1.1 assembly, visualised in HiGlass. Chromosomes are shown in order of size from left to right and top to bottom. An interactive version of this map is available [here](#).

Table 2. Chromosomal pseudomolecules in the genome assembly of *Notodonta ziczac*, ilNotZicz1.1.

INSDC accession	Chromosome	Size (Mb)	GC%
OU974037.1	1	14.88	38.5
OU974039.1	2	14.52	38.8
OU974040.1	3	14.41	38.1
OU974041.1	4	14.05	38.3
OU974042.1	5	13.88	37.7
OU974043.1	6	13.62	38.6
OU974044.1	7	13.07	38.4
OU974045.1	8	13.01	38.1
OU974046.1	9	12.97	37.8
OU974047.1	10	12.92	38.2
OU974048.1	11	12.79	38.0
OU974049.1	12	12.72	38.3
OU974050.1	13	12.42	38.5
OU974051.1	14	11.81	38.0
OU974052.1	15	11.72	38.6

OU974053.1	16	11.70	38.1
OU974054.1	17	11.69	38.4
OU974055.1	18	11.45	39.2
OU974056.1	19	10.88	38.7
OU974057.1	20	10.78	39.2
OU974058.1	21	10.78	38.4
OU974059.1	22	9.26	38.9
OU974060.1	23	9.21	39.0
OU974061.1	24	8.94	39.4
OU974062.1	25	8.86	39.3
OU974063.1	26	8.43	39.2
OU974064.1	27	6.64	40.3
OU974065.1	28	6.64	39.7
OU974066.1	29	6.21	40.7
OU974067.1	30	6.00	40.1
OU974038.1	Z	14.56	38.2
OU974068.1	MT	0.02	19.2
-	Unplaced	1.19	45.2

Table 3. Software tools used.

Software tool	Version	Source
Hifiasm	0.15.3-r339	Cheng et al., 2021
purge_dups	1.2.3	Guan et al., 2020
SALSA	3.0	Ghurye et al., 2019
longranger align	2.2.2	https://support.10xgenomics.com/genome-exome/software/pipelines/latest/advanced/other-pipelines
freebayes	1.3.1-17-gaa2ace8	Garrison & Marth, 2012
MitoHiFi	2.0	Uliano-Silva et al., 2021
HiGlass	1.11.6	Kerpedjiev et al., 2018
PretextView	0.2.x	https://github.com/wtsi-hpag/PretextView
BlobToolKit	3.0.5	Challis et al., 2020

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

Data availability

European Nucleotide Archive: *Notodonta ziczac* (pebble prominent). Accession number [PRJEB46845](#); <https://identifiers.org/ena.embl/PRJEB46845>.

The genome sequence is released openly for reuse. The *N. ziczac* genome sequencing initiative is part of the [Darwin Tree of Life](#) (DTOL) project. All raw sequence data and the assembly have been deposited in INSDC databases. The genome will be annotated and presented through the [Ensembl](#) pipeline at the European Bioinformatics Institute. Raw data and assembly accession identifiers are reported in [Table 1](#).

Author information

Members of the University of Oxford and Wytham Woods Genome Acquisition Lab are listed here: <https://doi.org/10.5281/zenodo.5746938>.

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

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

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

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

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

References

- Allio R, Schomaker-Bastos A, Romiguier J, et al.: **MitoFinder: Efficient Automated Large-Scale Extraction of Mitogenomic Data in Target Enrichment Phylogenomics.** *Mol Ecol Resour.* 2020; **20**(4): 892–905. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Challis R, Richards E, Rajan J, et al.: **BlobToolKit - Interactive Quality Assessment of Genome Assemblies.** *G3 (Bethesda).* 2020; **10**(4): 1361–74. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Cheng H, Concepcion GT, Feng X, et al.: **Haplotype-Resolved de Novo Assembly Using Phased Assembly Graphs with Hifiasm.** *Nat Methods.* 2021; **18**(2): 170–75. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Emmet AM: **The Scientific Names of the British Lepidoptera: Their History and Meaning.** Harley Books. 1991. [Reference Source](#)
- Garrison E, Marth G: **Haplotype-Based Variant Detection from Short-Read Sequencing.** 2012. [Publisher Full Text](#)
- Ghurye J, Rhie A, Walenz BP, et al.: **Integrating Hi-C Links with Assembly Graphs for Chromosome-Scale Assembly.** *PLoS Comput Biol.* 2019; **15**(8): e1007273. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Guan D, McCarthy SA, Wood J, et al.: **Identifying and Removing Haplotypic Duplication in Primary Genome Assemblies.** *Bioinformatics.* 2020; **36**(9): 2896–98. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Howe K, Chow W, Collins J, et al.: **Significantly Improving the Quality of Genome Assemblies through Curation.** *GigaScience.* 2021; **10**(1). [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](#)

Newman E: **An Illustrated Natural History of British Butterflies and Moths.** London: Glaisher. 1869.

[Reference Source](#)

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

[Reference Source](#)

Rao SSP, Huntley MH, Durand NC, *et al.*: **A 3D Map of the Human Genome at Kilobase Resolution Reveals Principles of Chromatin Looping.** *Cell.* 2014; **159**(7): 1665–80.

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

Schintlmeister A: **Notodontidae. Palaearctic Macrolepidoptera. Vol. 1.** Stenstrup: Apollo Books. 2008.

[Reference Source](#)

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

[Reference Source](#)

South R: **The Moths of the British Isles.** F Warne. 1977.

[Reference Source](#)

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

[Publisher Full Text](#)

Open Peer Review

Current Peer Review Status:  

Version 1

Reviewer Report 06 June 2022

<https://doi.org/10.21956/wellcomeopenres.19689.r50553>

© 2022 Ward 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 M. Ward

Australian Wine Research Institute, Adelaide, SA, Australia

The data note by the late Boyes *et al.*, describes the full genome sequence of *Notodonta ziczac* (Linnaeus, 1758). The data note is concise and in line with others of its kind, providing a lucid description of the organism and of the assembly. Therefore, I have no hesitation to support its indexing.

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: Adaptation, insect-hostplant arms race, grapevine genetics

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

Reviewer Report 26 May 2022

<https://doi.org/10.21956/wellcomeopenres.19689.r49487>

© 2022 Kawahara A. 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.



Akito Y. Kawahara 

McGuire Center for Lepidoptera and Biodiversity, Florida Museum of Natural History, University of Florida, Gainesville, FL, USA

This is an important data note by the late Douglas Boyes and others. The species is in steep decline and there is thus an immediate impetus to make this genome assembly public. I see no problems with the way the paper is written or presented, and is in-line with many of the other genome notes. I fully support and approve the indexing of this Data Note.

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 evolution, 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.
