DATA NOTE

The genome sequence of the iron prominent, *Notodonta dromedarius* (Linnaeus, 1767) [version 1; peer review: 1 approved, 2 approved with reservations]

Douglas Boyes¹⁺, Peter W.H. Holland², 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, Darwin Tree of Life Consortium

¹ UK Centre for Ecology and Hydrology, Wallingford, Oxfordshire, UK
² Department of Zoology, University of Oxford, Oxford, UK

* Deceased author

**Abstract**

We present a genome assembly from an individual male *Notodonta dromedarius* (iron prominent; Arthropoda; Insecta; Lepidoptera; Notodontidae). The genome sequence is 342 megabases in span. The majority of the assembly, 99.35%, is scaffolded into 31 chromosomal pseudomolecules, with the Z sex chromosome assembled.

**Keywords**

*Notodonta dromedarius*, iron prominent, genome sequence, chromosomal, Lepidoptera

This article is included in the [Tree of Life gateway](https://doi.org/10.12688/wellcomeopenres.17489.1).

### Open Peer Review

**Approval Status**

| 1 | 2 | 3 |
|---|---|---|
| ✔ | ? | ? |

**version 1**

1. **Jason Hill**, Uppsala University, Uppsala, Sweden
2. **Luc Swevers**, National Center for Scientific Research Demokritos, Athens, Greece
3. **Camille Cornet**, University of Neuchatel, Neuchâtel, Switzerland

Any reports and responses or comments on the article can be found at the end of the article.
Species taxonomy
Eukaryota; Metazoa; Ecdysozoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Lepidoptera; Glossata; Ditrysia; Notuoidea; Notodontidae; Notodontinae; Notodonta; *Notodonta dromedarius* (Linnaeus, 1767) (NCBI:txid753204).

Background
*Notodonta dromedarius* (iron prominent) has rust-coloured wing markings that give the moth its common name. The species is widely distributed across Europe and is common throughout the UK; however, abundance has greatly decreased at monitored sites over the past 50 years (Randle et al., 2019). There are two broods of *N. dromedarius* in the south of England flying in May/June and August, but usually a single brood in the north of England and in Scotland (Randle et al., 2019). The moth was one of the first members of the Notodontidae to have the sex pheromone chemical identified (Bestmann et al., 1993). The genome of *N. dromedarius* was sequenced as part of the Darwin Tree of Life Project, a collaborative effort to sequence all of the named eukaryotic species in the Atlantic Archipelago of Britain and Ireland. Here we present a chromosomally complete genome sequence for *N. dromedarius*, based on one male specimen from Wytham Woods, Oxfordshire, UK.

Genome sequence report
The genome was sequenced from a single male *N. dromedarius* (iINotDrom1) collected from Wytham Woods, Oxfordshire, UK (latitude 51.772, longitude -1.338). A total of 77-fold coverage in Pacific Biosciences single-molecule long reads (N50 13 kb) and 112-fold coverage in 10X Genomics read clouds were generated. Primary assembly contigs were scaffolded with chromosome conformation Hi-C data. Manual assembly curation corrected 6 missing/misjoins and removed 56 haplotypic duplications, reducing the assembly length by 0.83% and the scaffold number by 28.57%, and increasing the scaffold N50 by 3.08%.

The final assembly has a total length of 342 Mb in 145 sequence scaffolds with a scaffold N50 of 12.1 Mb (Table 1). Of the assembly sequence, 99.35% was assigned to 31 chromosomal-level scaffolds, representing 30 autosomes (numbered by sequence length), and the Z sex chromosome (Figure 2–Figure 5; Table 2). The assembly has a BUSCO v5.1.2 (Simão et al., 2015) completeness of 98.9% (single 98.6%, duplicated 0.3%) 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 nucleic acid extraction
A single male *N. dromedarius* (iINotDrom1) was collected from Wytham Woods, Oxfordshire, UK (latitude 51.772, longitude -1.338) by Douglas Boyes, UKCEH, using a light trap. The specimen was identified by the same individual and preserved on dry ice.

DNA was extracted from head/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 head/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
Table 1. Genome data for *Notodonta dromedarius*, ilNotDrom1.1.

| Project accession data                  | Assembly identifier | ilNotDrom1.1 |
|----------------------------------------|---------------------|--------------|
| Species                                 | *Notodonta dromedarius* |
| Specimen                                | ilNotDrom1           |
| NCBI taxonomy ID                        | NCBITaxid:753204    |
| BioProject                              | PRJEB42138           |
| BioSample ID                            | SAMEA7520190         |
| Isolate information                     | Male, head/thorax, abdomen |

| Raw data accessions                    |                     |
|----------------------------------------|----------------------|
| Pacific Biosciences SEQUEL II          | ERR6590583           |
| 10X Genomics Illumina                  | ERR6002703-ERR6002706|
| Hi-C Illumina                          | ERR6003044           |
| Illumina PolyA RNA-Seq                 | ERR6286708           |

| Genome assembly                        |                     |
|----------------------------------------|----------------------|
| Assembly accession                     | GCA_905147325.1      |
| Accession of alternate haplotype       | GCA_905147855.1      |
| Span (Mb)                               | 342                  |
| Number of contigs                       | 168                  |
| Contig N50 length (Mb)                  | 10                   |
| Number of scaffolds                     | 146                  |
| Scaffold N50 length (Mb)                | 12                   |
| Longest scaffold (Mb)                   | 15                   |
| BUSCO* genome score                     | C:98.9%,S:98.6%,D:0.3%,F:0.2%,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/ilNotDrom1.1/dataset/CAJHVG01/busc.*

of the integrity of the RNA was done using Agilent RNA 6000 Pico Kit and Eukaryotic Total RNA assay.

Sequencing
Pacific Biosciences HiFi circular consensus and 10X Genomics read cloud DNA sequencing libraries, 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 of the same specimen using the Arima v1 Hi-C kit and sequenced on HiSeq X.

Genome assembly
Assembly was carried out with HiCanu (*Nurk et al.*, 2020); 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 longeralign 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).
assembly was checked for contamination and corrected using the gEVAL system (Chow et al., 2016) as described previously (Howe et al., 2021). Manual curation (Howe et al., 2021) was performed using gEVAL, HiGlass (Kerpedjiev et al., 2018) and Pretex. The mitochondrial genome was assembled using MitoHiFi (Uliano-Silva et al., 2021) and annotated using
Figure 3. Genome assembly of *Notodonta dromedarius*, ilNotDrom1.1: 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/ilNotDrom1.1/dataset/CAJHVG01/blob.

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.

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 [insert specific requirements or guidelines].
Figure 4. Genome assembly of *Notodonta dromedarius*, iINotDrom1.1: cumulative sequence. BlobToolKit cumulative sequence plot. The grey line shows cumulative length for all scaffolds. Coloured lines show cumulative lengths of scaffolds assigned to each phylum using the buscogenes taxrule. An interactive version of this figure is available at https://blobtoolkit.genomehubs.org/view/iINotDrom1.1/dataset/CAJHV0G01/cumulative.

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
Figure 5. Genome assembly of Notodonta dromedarius, iNotDrom1.1: Hi-C contact map. Hi-C contact map of the iNotDrom1.1 assembly, visualised in HiGlass. Chromosomes are shown in order of size from left to right and top to bottom.

Table 2. Chromosomal pseudomolecules in the genome assembly of Notodonta dromedarius, iNotDrom1.1.

| INSDC accession | Chromosome | Size (Mb) | GC%  |
|-----------------|------------|-----------|------|
| LR990159.1      | 1          | 14.52     | 38.7 |
| LR990161.1      | 2          | 14.15     | 38.4 |
| LR990162.1      | 3          | 14.00     | 38.2 |
| LR990163.1      | 4          | 13.87     | 38.6 |
| LR990164.1      | 5          | 13.54     | 37.6 |
| LR990165.1      | 6          | 13.03     | 38.4 |
| LR990166.1      | 7          | 12.72     | 38.4 |
| LR990167.1      | 8          | 12.46     | 37.7 |
| LR990168.1      | 9          | 12.33     | 38.3 |
| LR990169.1      | 10         | 12.21     | 37.6 |
| LR990170.1      | 11         | 12.13     | 38.4 |
| LR990171.1      | 12         | 12.06     | 37.7 |
| LR990172.1      | 13         | 11.99     | 38.2 |
| LR990173.1      | 14         | 11.70     | 38.7 |
| LR990174.1      | 15         | 11.32     | 38.1 |
| LR990159.1      | Z          | 14.29     | 38.2 |
| LR990159.1      | MT         | 0.02      | 19.2 |
| LR990159.1      | Unplaced   | 2.22      | 41.3 |
Table 3. Software tools used.

| Software tool    | Version | Source                                      |
|------------------|---------|---------------------------------------------|
| HiCanu           | 1.0     | Nurk et al., 2020                          |
| 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               |
| 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      | 2.6.2   | Challis et al., 2020                       |

out within this document in respect of all samples acquired for, and supplied to, the Darwin Tree of Life Project. Each transfer 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 dromedarius (iron prominent). Accession number PRJEB42138: https://www.ebi.ac.uk/ena/browser/view/PRJEB42138.

The genome sequence is released openly for reuse. The N. dromedarius 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.

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.5744840.

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.5743293.

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
Bestmann HJ, Kern F, Schäfer D, et al.: (11Z,13Z)-11,13-Hexadecadienal, the Sex Pheromone of Females of Notodonta Dromedarius. Naturwissenschaften. 1993; 80(6): 271–73. 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
Chow W, Brugger K, Caccamo M, et al.: gEVAL - a web-based browser for
evaluating genome assemblies. *Bioinformatics*. 2016; 32(16): 2508–10.

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

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.

Guan D, McCarthy SA, Wood J, et al.: Identifying and Removing Haplotypic Duplication in Primary Genome Assemblies. *Bioinformatics*. 2020; 36(9): 2896–98.

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

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.

Nurk S, Walenz BP, Rhie A, et al.: HiCanu: Accurate Assembly of Segmental Duplications, Satellites, and Allelic Variants from High-Fidelity Long Reads. *Genome Res*. 2020; 30(9): 1291–1305.

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

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; 158(7): 1665–80.

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.

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 20 June 2024

https://doi.org/10.21956/wellcomeopenres.19339.r83095

© 2024 Cornet 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.

Camille Cornet
University of Neuchatel, Neuchâtel, Switzerland

The authors present a reference genome assembly for the moth species Notodonta dromedarius. The methods used to sequence and assemble the genome are sound and well reported in general. The resulting assembly is of high quality. This resource is likely to be very useful for future studies, for example of population genetics of the species and comparative genomics of Lepidoptera. I have only a few comments that I hope might improve this genome note.

Background section:

It would be good to provide some information on the ecology of the species. For example, in which habitat can it be found, what is the host plant of the caterpillar, are adults pollinators?

Methods:

There is an incongruence in this sentence: “DNA was extracted from head/thorax tissue at the Wellcome Sanger Institute (WSI) Scientific Operations core from the whole organism”. Was DNA extracted from the whole organism or just the head and thorax?

Hi-C sequencing yield (number of reads) should be mentioned. Ideally also the percentage of read pairs mapping to a different contig prior to scaffolding. Some steps in the genome assembly could be better explained, for example, how were the 31 chromosomes inferred from the 146 scaffolds? Was it just a matter of size? How was the Z chromosome identified? Simply using depth of coverage from the PacBio reads mapped back to the assembly?

Most importantly, there is no mention of genome annotation (except a small note in the Data availability statement), while RNA sequencing is well described, and annotation is reported online as “complete”. An entire section about genome annotation should be added, including methods used, number of protein-coding genes annotated, etc.
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?
Partly

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

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Comparative genomics, population 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.
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:** Insect molecular biology and biotechnology

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 22 May 2024

https://doi.org/10.21956/wellcomeopenres.19339.r83090

© 2024 Hill J. This is an open access peer review report distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Jason Hill
Uppsala University, Uppsala, Sweden

In this study the authors describe the genome assembly and annotation of the European moth *Notodonta dromedarius*. The methods employed represent the current highest standard for assembly and annotation projects. This represents an additional excellent addition to the knowledge of the genomic architecture of Lepidoptera and will be useful for deeper studies of *Notodonta dromedarius* and for comparative genomics work in broader context.

**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:** Population genomics, comparative genomics, Lepidoptera 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.