DATA NOTE

The genome sequence of Svensson’s copper underwing,
Amphipyra berbera Rungs, 1949 [version 1; peer review: 3
approved]

Douglas Boyes¹†, Liam M. Crowley, Peter W.H. Holland,¹²,
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

¹UK Centre for Ecology & Hydrology, Wallingford, UK
²Department of Zoology, University of Oxford, Oxford, UK
†Deceased author

First published: 18 Nov 2021, 6:314
https://doi.org/10.12688/wellcomeopenres.17350.1
Latest published: 18 Nov 2021, 6:314
https://doi.org/10.12688/wellcomeopenres.17350.1

Abstract
We present a genome assembly from an individual male Amphipyra
berbera (Svensson’s copper underwing; Arthropoda; Insecta;
Lepidoptera; Noctuidae). The genome sequence is 582 megabases in
span. The majority (99.97%) of the assembly is scaffolded into 31
chromosomal pseudomolecules, with the Z sex chromosome
assembled.

Keywords
Amphipyra berbera, Svensson’s copper underwing, genome
sequence, chromosomal

This article is included in the Tree of Life
gateway.

Open Peer Review

Approval Status ✓ ✓ ✓

version 1
18 Nov 2021

1. Sina Beier, MRC Toxicology Unit,
Cambridge, UK

2. Sivasankaran Kuppusamy, Loyola
College, Chennai, India

3. Shiqi Luo, China Agricultural University,
Beijing, China

Any reports and responses or comments on the
article can be found at the end of the article.
Corresponding author: Darwin Tree of Life Consortium (mark.blaxter@sanger.ac.uk)

Author roles: Boyes D: Investigation, Resources; Crowley LM: Writing – Original Draft Preparation, Writing – Review & Editing; Holland PWH: Supervision, 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).

Copyright: © 2021 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, Crowley LM, Holland PWH et al. The genome sequence of Svensson's copper underwing, Amphipyra berbera Rungs, 1949 [version 1; peer review: 3 approved] Wellcome Open Research 2021, 6:314 https://doi.org/10.12688/wellcomeopenres.17350.1

First published: 18 Nov 2021, 6:314 https://doi.org/10.12688/wellcomeopenres.17350.1
Species taxonomy
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Lepidoptera; Glossata; Ditrysia; Noctuoidea; Noctuidae; Amphipyra (Rungs, 1949) (NCBI:txid987877).

Background
Amphipyra berbera (Svensson’s copper underwing) is a large noctuid moth with broad, brown forewings patterned with pale zigzags and hindwings suffused with copper brown. The moth has been recorded across much of Eurasia and North Africa; in the UK it is common across England and Wales with scattered records from Scotland. Svensson’s copper underwing is very similar morphologically to A. pyramidea (copper underwing) and was initially considered a subspecies A. pyramidea berbera (Rungs, 1949) until recognised as a separate species in 1968 (Fletcher, 1968). Adults can be distinguished by the extent of the copper colouration on the underside of the hindwing and larvae can be separated by the colour of the dorsal point on abdominal segment eight. The status of A. berbera as a distinct species is supported by mitochondrial COI barcode data (Chen et al., 2013). Larvae of A. berbera feed on the leaves of several deciduous trees, frequently oak (Quercus). In the UK, adults fly from June to September and are attracted to sweet substances including tree sap. The overwintering stage is as an ovum and pupation occurs underground.

Genome sequence report
The genome was sequenced from one male A. berbera (Figure 1) collected from Wytham Woods, Oxfordshire (biological vice-county: Berkshire), UK (latitude 51.772, longitude -1.338). A total of 41-fold coverage in Pacific Biosciences single-molecule long reads and 71-fold coverage in 10X Genomics read clouds were generated. Primary assembly contigs were scaffolded with chromosome conformation Hi-C data. Manual assembly curation corrected 8 missing/misjoins, reducing the assembly length by 0.01% and the scaffold number by 15.38%.

Table 1. Genome data for Amphipyra berbera, ilAmpBerb1.1.

| Project accession data          |                   |
|--------------------------------|-------------------|
| Assembly identifier            | ilAmpBerb1.1      |
| Species                        | Amphipyra berbera |
| Specimen                       | ilAmpBerb1        |
| NCBI taxonomy ID               | NCBI:txid987877   |
| BioProject                     | PRJEB45130        |
| BioSample ID                   | SAMEA7701493      |
| Isolate information            | Male, abdomen (genome assembly), head/thorax (HI-C) |

| Raw data accessions            |                   |
|--------------------------------|-------------------|
| PacificBiosciences SEQUEL II   | ERR6436378        |
| 10X Genomics Illumina          | ERR6054818-ERR6054821 |
| HI-C Illumina                  | ERR6054817        |

| Genome assembly                |                   |
|--------------------------------|-------------------|
| Assembly accession             | GCA_910594945.1   |
| Accession of alternate haplotype| GCA_910595045.1   |
| Span (Mb)                      | 582               |
| Number of contigs              | 39                |
| Contig N50 length (Mb)         | 19.9              |
| Number of scaffolds            | 33                |
| Scaffold N50 length (Mb)       | 20.1              |
| Longest scaffold (Mb)          | 23.5              |
| BUSCO* genome score            | 98.9%(S:98.5%,D:0.5%), F:0.3%,M:0.8%,n:5286 |

*BUSCO scores based on the lepidoptera_odb10 BUSCO set using v5.1.2. C= complete [S= single copy, D=duplicated], F=fragmented, M=missing, n=number of orthologues in comparison. A full set of BUSCO scores is available at https://blobtoolkit.genomehubs.org/view/ilAmpBerb1.1r
dataset/CAjvcgo1/busco.
UKCEH, using a light trap. 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 ilAmpBerb1 sample was weighed and dissected on dry ice with tissue set aside for Hi-C sequencing. Abdomen 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 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 read cloud DNA 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 and Illumina HiSeq X instruments. Hi-C data were generated from head/thorax tissue using the Arima v2 Hi-C kit and sequenced on an Illumina NovaSeq 6000 instrument.

**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

---

**Figure 3. Genome assembly of Amphipyra berbera, ilAmpBerb1.1: 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/ilAmpBerb1.1/dataset/CAJVCG01/blob.
performed by aligning 10X Genomics read data to the assembly with longranger align, calling variants with freebayes (Garrison & Marth, 2012). The assembly was then scaffolded with Hi-C data (Rao et al., 2014) using SALSA2 (Ghurye et al., 2019). The assembly was checked for contamination and corrected using the gEVAL system (Chow et al., 2016) as described previously (Howe et al., 2021). Manual curation (Howe et al., 2021) was performed using gEVAL, HiGlass (Kerpedjiev et al., 2018) and Pretext. The mitochondrial genome was assembled using MitohiFi (Uliano-Silva et al., 2021). The genome was analysed and BUSCO scores generated within the BlobToolKit environment (Challis et al., 2020). Table 3 contains a list of all software tool versions used, where appropriate.

Figure 4. Genome assembly of Amphipyra berbera, ilAmpBerb1.1: 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/ilAmpBerb1.1/dataset/CAJVCG01/cumulative.
Table 2. Chromosomal pseudomolecules in the genome assembly of *Amphipyra berbera*, ilAmpBerb1.1.

| INSDC accession | Chromosome | Size (Mb) | GC% |
|-----------------|------------|-----------|-----|
| OU343122.1      | 1          | 23.51     | 38  |
| OU343123.1      | 2          | 22.83     | 38.3|
| OU343124.1      | 3          | 22.62     | 38.3|
| OU343125.1      | 4          | 22.01     | 38.5|
| OU343126.1      | 5          | 21.94     | 38.3|
| OU343127.1      | 6          | 21.88     | 38  |
| OU343128.1      | 7          | 21.87     | 38.2|
| OU343129.1      | 8          | 21.69     | 38  |
| OU343130.1      | 9          | 21.69     | 38.1|
| OU343131.1      | 10         | 21.37     | 38.1|
| OU343132.1      | 11         | 21.16     | 38.1|
| OU343133.1      | 12         | 20.95     | 38.3|
| OU343134.1      | 13         | 20.13     | 38.4|
| OU343135.1      | 14         | 20.09     | 38.5|
| OU343136.1      | 15         | 19.93     | 38.5|
| OU343137.1      | 16         | 19.30     | 38.7|
| OU343121.1      | Z          | 26.79     | 37.8|
| OU343152.1      | MT         | 0.02      | 21  |
|                 | -          | 0.12      | 48.3|

Figure 5. Genome assembly of *Amphipyra berbera*, ilAmpBerb1.1: Hi-C contact map. Hi-C contact map of the ilAmpBerb1.1 assembly, visualised in HiGlass.
**Table 3. Software tools used.**

| Software tool     | Version | Source                                      |
|-------------------|---------|---------------------------------------------|
| Hifiasm           | 0.15    | (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          | 2.0     | (Uliano-Silva et al., 2021)                |
| gEVAL             | N/A     | Chow et al., 2016                          |
| HiGlass           | 1.11.6  | (Kerpedjiev et al., 2018)                  |
| PretextView       | 0.2.x   | https://github.com/wtsi-hpag/PretextView   |
| BlobToolKit       | 2.6.2   | Challis et al., 2020                       |

**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 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: Amphipyra berbera (Svensson’s copper underwing). Accession number PRJEB45130; https://identifiers.org/ena.embl/PRJEB45130.

The genome sequence is released openly for reuse. The *A. berbera* 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.4789929.

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**

Challis R, Richards E, Rajan J, et al.: BlobToolKit - Interactive Quality Assessment of Genome Assemblies. G3 (Bethesda). 2020; 10(4): 1361–74.

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.
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

Fletcher DS: Amphipyra Pyramidea (Linnaeus) and A. Berbera Rungs (Lep., Noctuidae), Two Species Confused. Entomologist's Gazette. 1968; 19: 91–106.

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

Ghurye J, Rhie A, Walenz BP, et al.: Integrating Hi-C Links with Assembly Graphs for Chromosome-Scale Assembly. PLoS Comput Biol. 2019; 15(8): e1007273. PubMed Abstract | Publisher Full Text | Free Full Text

Guan D, McCarthy SA, Wood J, et al.: Identifying and Removing Haplotypic Duplication in Primary Genome Assemblies. Bioinformatics. 2020; 36(9): 2896–98. PubMed Abstract | Publisher Full Text | Free Full Text

Howe K, Chow W, Collins J, et al.: Significantly Improving the Quality of Genome Assemblies through Curation. Gigascience. 2021; 10(1): giaa153. PubMed Abstract | Publisher Full Text | Free Full Text

Kerpedjiev P, Abdennur N, Lekschas E, 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, Seppey 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

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

Rungs C: Notes de Lépidoptérologie Marocaine (XV): Noctuidæ Nouvelles. Bulletin de La Société Des Sciences Naturelles Du Maroc. 1949; (25–27): 328–32.

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 23 October 2024

https://doi.org/10.21956/wellcomeopenres.19180.r103652

© 2024 Luo S. 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.

Shiqi Luo
Department of Entomology, College of Plant Protection, China Agricultural University, Beijing, China

This paper presents the genome assembly of Svensson's copper underwing at the chromosomal level. The sequencing and bioinformatics analysis methods used were sound, resulting in a high-quality assembly.

The paper briefly describes the genome assembly and provides useful information on how to access and download the original sequencing data and genome assembly.

While the software tools used are listed, it would be helpful to include the main parameters utilized with these tools.

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: insect functional genomics, non-coding RNAs

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 23 October 2024

https://doi.org/10.21956/wellcomeopenres.19180.r103659

© 2024 Kuppusamy S. 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.

Sivasankaran Kuppusamy
Loyola College, Chennai, Tamil Nadu, India

Chromosome level genome assembly was carried out from Amphipyra berbera Rungs, 1949. Appropriate software was used for the assembly and annotation. The comprehensive data will be useful for the phylogenomic study of lepidopteran moths

Comments on the manuscript:

The taxonomic details “(Svensson's copper underwing; Arthropoda; Insecta; Lepidoptera; Noctuidae)” of the species need not be mentioned in the abstract. Mitochondrial genome size is not given in the abstract. It can be given in the abstract. In the background first paragraph first sentence can be started as Amphipyra berbera Rungs, 1949. Through genome assembly authors would have received the protein-coding genes and gene transcripts. The protein-coding genes and gene transcripts haven't been included in the information in the text and table. Explain. Mitochondrial genome sequence length 15,345 bp can be included in the text.

Above all, I confirm that the manuscript meets the necessary scientific standard and is suitable for 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: Phylogenetic analysis of superfamily Noctuoidea moths using mitochondrial
Sina Beier
MRC Toxicology Unit, Cambridge, UK

This study presents the chromosome level assembly of the genome of the Svenson's copper underwing.

It was assembled following the state-of-the-art pipeline developed and used by the Darwin Tree of Life project including long and short read sequencing. This resulted in a high quality assembly which is based on the assembly of the log reads and polished with the short reads, leading to a great resolution of haplotype-resolved and nearly complete chromosomes.

The data is presented in a concise way and leaves me to only ask for the possibility to add the W chromosome which is currently missing in a later state of the project.

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: bioinformatics, genome assembly, genome assembly of 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.