DATA NOTE

The genome sequence of the two-spot ladybird, *Adalia bipunctata* (Linnaeus, 1758) [version 1; peer review: awaiting peer review]

Wellcome Sanger Institute Tree of Life programme, Wellcome Sanger Institute Scientific Operations: DNA Pipelines collective, Tree of Life Core Informatics collective, Zoe Goate¹, Darwin Tree of Life Consortium

¹Wellcome Sanger Institute, Hinxton, UK

**Abstract**

We present a genome assembly from an individual male *Adalia bipunctata* (the two-spot ladybird; Arthropoda; Insecta; Coleoptera; Coccinellidae). The genome sequence is 475 megabases in span. Most of the assembly (94.87%) is scaffolded into 11 chromosomal pseudomolecules, with the X and Y sex chromosomes assembled. The complete mitochondrial genome was also assembled and is 21.2 kilobases in length. Gene annotation of this assembly in Ensembl identified 13,611 protein coding genes.

**Keywords**

*Adalia bipunctata*, two-spot ladybird, chromosomal, Coleoptera

This article is included in the Tree of Life gateway.
Corresponding author: Darwin Tree of Life Consortium (mark.blaxter@sanger.ac.uk)

Author roles: Goate Z: 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 Wellcome Sanger Institute Tree of Life programme 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: Wellcome Sanger Institute Tree of Life programme, Wellcome Sanger Institute Scientific Operations: DNA Pipelines collective, Tree of Life Core Informatics collective et al. The genome sequence of the two-spot ladybird, Adalia bipunctata (Linnaeus, 1758) [version 1; peer review: awaiting peer review] Wellcome Open Research 2022, 7:288

First published: 30 Nov 2022, 7:288 https://doi.org/10.12688/wellcomeopenres.18610.1
**Species taxonomy**
Eukaryota; Metazoa; Ecdysozoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Coleoptera; Polyphaga; Cucujiformia; Coccinellidae; Coccinellinae; Coccinelini; Adalia; *Adalia bipunctata* (Linnaeus, 1758) (NCBI:txid7084).

**Background**
The two-spot ladybird, *Adalia bipunctata* (Linnaeus, 1758) is a Holarctic species native to Europe, Central Asia and North America. *A. bipunctata* was once the second most common ladybird in the US, but the invasion of the predatory Asian species, the harlequin ladybird, *Harmonia axyridis* has seen a rapid decline in the two-spot population over the last decade (Kenis *et al.*, 2020). This widespread species occupies a variety of habitats, from deciduous or coniferous woodlands to orchards and crops. In temperate regions, adults appear in March and are known to overwinter in large groups along with other common species in among loose bark, leaf-litter and outhouses. Both adult and larval forms of *A. bipunctata* are voracious aphidophagous hunters, making them suitable biocontrol agents against aphids in agricultural systems (Riddick, 2017). Two-spots exhibit complex polymorphism with typical morphs conspicuously marked with vivid red elytra and a large black spot in the middle of each (Figure 1), whilst melanic morphs display a black elytra with red spots (Rutkowski *et al.*, 2019).

The two-spot ladybird is a classic model for population genetics studies, and a complete genome assembly of *A. bipunctata* may help to characterise the genetic diversity underpinning phenotypic polymorphisms among populations across different environments (Gautier *et al.*, 2018).

We present a complete genome assembly for *A. bipunctata* as part of the Darwin Tree of Life project, which aims to sequence the genomes of 70,000 species of eukaryotic organisms in Britain and Ireland.

![Figure 1. An *Adalia bipunctata* image (Photograph from www.entomart.be CC-BY).](image)

**Genome sequence report**
The genome was sequenced from an individual male *A. bipunctata* (icAdaBipu1) purchased live from Dragonfly, Essex, UK. A total of 48-fold coverage in Pacific Biosciences single-molecule HiFi long reads and 80-fold coverage in 10X Genomics read clouds were generated. Primary assembly contigs were scaffolded with chromosome conformation Hi-C data. Manual assembly curation corrected 72 missing/misjoins and removed 17 haplotypic duplications, reducing the assembly size by 2.45% and the scaffold number by 34.08%, and increasing the scaffold N50 by 105.95%. The final assembly has a total length of 475 Mb in 118 sequence scaffolds with a scaffold N50 of 45.9 Mb (Table 1). Most of the assembly

<table>
<thead>
<tr>
<th>Table 1. Genome data for <em>A. bipunctata</em>, icAdaBipu1.</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Project accession data</strong></td>
</tr>
<tr>
<td>Assembly identifier</td>
</tr>
<tr>
<td>Species</td>
</tr>
<tr>
<td>Specimen</td>
</tr>
<tr>
<td>NCBI taxonomy ID</td>
</tr>
<tr>
<td>BioProject</td>
</tr>
<tr>
<td>BioSample ID</td>
</tr>
<tr>
<td>Isolate information</td>
</tr>
<tr>
<td><strong>Raw data accessions</strong></td>
</tr>
<tr>
<td>PacificBiosciences SEQUEL II</td>
</tr>
<tr>
<td>10X Genomics Illumina</td>
</tr>
<tr>
<td>Hi-C Illumina</td>
</tr>
<tr>
<td><strong>Genome assembly</strong></td>
</tr>
<tr>
<td>Assembly accession</td>
</tr>
<tr>
<td>Accession of alternate haplotype</td>
</tr>
<tr>
<td>Span (Mb)</td>
</tr>
<tr>
<td>Number of contigs</td>
</tr>
<tr>
<td>Contig N50 length (Mb)</td>
</tr>
<tr>
<td>Number of scaffolds</td>
</tr>
<tr>
<td>Scaffold N50 length (Mb)</td>
</tr>
<tr>
<td>Longest scaffold (Mb)</td>
</tr>
<tr>
<td>BUSCO* genome score</td>
</tr>
<tr>
<td><strong>Genome annotation</strong></td>
</tr>
<tr>
<td>Number of protein-coding genes</td>
</tr>
</tbody>
</table>

*BUSCO scores based on the endopterygotaodb10 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.genomeshubs.org/view/icAdaBipu1.1/dataset/CAJUZD01/busco.*
sequence (94.87%) was assigned to 11 chromosomal-level scaffolds, representing 9 autosomes (numbered by sequence length), and the X and Y sex chromosomes (Figure 2–Figure 5; Table 2).

The assembly has a BUSCO v5.2.2 (Manni et al., 2021) completeness of 97.6% (single 96.2%, duplicated 1.4%) using the endopterygota_odb10 reference set (n = 2,124).

While not fully phased, the assembly deposited is of one haplotype. Contigs corresponding to the second haplotype have also been deposited.

Genome annotation report
The A. bipunctata genome was annotated using the Ensembl rapid annotation pipeline (Table 1; https://rapid.ensembl.org/Adalia_bipunctata_GCA_910592335.1/Info/Index). The resulting annotation includes 26,646 gene transcripts from 13,611 protein-coding genes and 3,277 non-coding genes.

Methods
Sample acquisition and DNA extraction
One male A. bipunctata specimen (icAdaBipu1), purchased live from Dragonfly, Essex, UK, was used for this genome assembly. The specimen was preserved on dry ice. DNA was

![Figure 2. Genome assembly of A. bipunctata, icAdaBipu1.1: metrics.](https://blobtoolkit.genomehubs.org/view/icAdaBipu1.1/dataset/CAJUZD01/snail)
extracted at the Tree of Life laboratory, Wellcome Sanger Institute. The icAdaBipul sample was weighed and dissected on dry ice with tissue set aside for Hi-C sequencing. Whole tissue was disrupted using a Nippi Powermasher fitted with a BioMasher pestle. 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 of 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
Figure 4. Genome assembly of *A. bipunctata*, icAdaBipu1.1: 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/icAdaBipu1.1/dataset/CAJUZD01/cumulative.

Figure 5. Genome assembly of *A. bipunctata*, icAdaBipu1.1: Hi-C contact map. Hi-C contact map of the icAdaBipu1.1 assembly, visualised in HiGlass. Chromosomes are arranged in size order from left to right and top to bottom. An interactive version of this image can be viewed at https://genome-note-higlass.tol.sanger.ac.uk/l/?d=HVP-dSN3RVK7oemLsXTA.
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 HiSeq X (10X) instruments. Hi-C data were generated in the Tree of Life laboratory from remaining tissue of icAdaBipu1 using the Arima v2 kit and sequenced on a HiSeq X instrument.

Genome assembly
Assembly of PacBio reads was carried out with Hiatusm (Cheng et al., 2021) and haplotypic duplication was identified and removed with purge_dups (Guan et al., 2020). One round of short-read 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 SALS2A (Ghurye et al., 2019). The assembly was checked for contamination as described previously (Howe et al., 2021). Manual curation was performed using and gEVAL (Chow et al., 2016), HiGlass (Kerpedjiev et al., 2018) and Pretext (Harry, 2022). 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 were generated within the BlobToolKit environment (Challis et al., 2020). Table 3 contains a list of all software tool versions used, where appropriate.

Genome annotation
The icAdaBipu1 genome was annotated using the Ensembl rapid annotation pipeline (Aken et al., 2016) (Table 1; https://rapid.ensembl.org/Adalia_bipunctata_GCA_910592335.1/Info/Index). The resulting annotation includes 26,646 transcribed mRNAs from 13,611 protein coding and 3,277 non-coding genes.

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: Adalia bipunctata (2-spot ladybird), Accession number PRJEB45127, https://identifiers.org/ena.emb|PRJEB45127 (Wellcome Sanger Institute, 2022).

The genome sequence is released openly for reuse. The A. bipunctata genome sequencing initiative is part of the Darwin Tree of Life (DTol) project. All raw sequence data

Table 2. Chromosomal pseudomolecules in the genome assembly of A. bipunctata, icAdaBipu1.1.

<table>
<thead>
<tr>
<th>INSDC accession</th>
<th>Chromosome</th>
<th>Size (Mb)</th>
<th>GC%</th>
</tr>
</thead>
<tbody>
<tr>
<td>OU342948.1</td>
<td>1</td>
<td>76.34</td>
<td>36.1</td>
</tr>
<tr>
<td>OU342949.1</td>
<td>2</td>
<td>74.05</td>
<td>36.0</td>
</tr>
<tr>
<td>OU342950.1</td>
<td>3</td>
<td>55.99</td>
<td>36.0</td>
</tr>
<tr>
<td>OU342951.1</td>
<td>4</td>
<td>45.87</td>
<td>36.2</td>
</tr>
<tr>
<td>OU342952.1</td>
<td>5</td>
<td>43.01</td>
<td>36.2</td>
</tr>
<tr>
<td>OU342953.1</td>
<td>6</td>
<td>36.65</td>
<td>36.2</td>
</tr>
<tr>
<td>OU342954.1</td>
<td>7</td>
<td>30.9</td>
<td>36.3</td>
</tr>
<tr>
<td>OU342955.1</td>
<td>8</td>
<td>30.76</td>
<td>36.2</td>
</tr>
<tr>
<td>OU342956.1</td>
<td>9</td>
<td>27.12</td>
<td>36.7</td>
</tr>
<tr>
<td>OU342957.1</td>
<td>X</td>
<td>27.05</td>
<td>36.2</td>
</tr>
<tr>
<td>OU342958.1</td>
<td>Y</td>
<td>1.75</td>
<td>36.7</td>
</tr>
<tr>
<td>OU342959.1</td>
<td>MT</td>
<td>0.02</td>
<td>20.3</td>
</tr>
<tr>
<td></td>
<td>-</td>
<td>25.78</td>
<td>31.2</td>
</tr>
</tbody>
</table>

Table 3. Software tools used.

<table>
<thead>
<tr>
<th>Software tool</th>
<th>Version</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>BlobToolKit</td>
<td>3.4.0</td>
<td>(Challis et al., 2020)</td>
</tr>
<tr>
<td>freebayes</td>
<td>1.3.1-17-gaa2ace8</td>
<td>(Garrison &amp; Marth, 2012)</td>
</tr>
<tr>
<td>gEVAL</td>
<td>N/A</td>
<td>(Chow et al., 2016)</td>
</tr>
<tr>
<td>HiGlass</td>
<td>1.11.6</td>
<td>(Kerpedjiev et al., 2018)</td>
</tr>
<tr>
<td>longranger align</td>
<td>2.2.2</td>
<td><a href="https://support.10xgenomics.com/">https://support.10xgenomics.com/</a> genome-exome/software/pipelines/latest/advanced/other-pipelines</td>
</tr>
<tr>
<td>MitoHiFi</td>
<td>1.0</td>
<td>(Uliano-Silva et al., 2021)</td>
</tr>
<tr>
<td>PretextView</td>
<td>0.1.x</td>
<td><a href="https://github.com/wtsi-hpag/PretextView">https://github.com/wtsi-hpag/PretextView</a></td>
</tr>
<tr>
<td>purge_dups</td>
<td>1.2.3</td>
<td>(Guan et al., 2020)</td>
</tr>
<tr>
<td>SALSA2</td>
<td>2.2</td>
<td>(Ghurye et al., 2019)</td>
</tr>
</tbody>
</table>
and the assembly have been deposited in INSDC databases. Raw data and assembly accession identifiers are reported in Table 1.

Author information
Members of the Wellcome Sanger Institute Tree of Life programme are listed here: https://doi.org/10.5281/zenodo.4783585.


Members of the Tree of Life Core Informatics collective are listed here: https://doi.org/10.5281/zenodo.5013541.

Members of the Darwin Tree of Life Consortium are listed here: https://doi.org/10.5281/zenodo.4783558.

References


Harry E: PretextView (Paired Read TEXTure Viewer): A desktop application for viewing pretext contact maps. 2022; (Accessed: 19 October 2022). Reference Source


Wellcome Sanger Institute: The genome sequence of the 2-spot ladybird, Adalia bipunctata (Linnaeus, 1758). European Nucleotide Archive. [Dataset]. 2022. Reference Source