DATA NOTE

The genome sequence of the two-spot ladybird, *Adalia bipunctata* (Linnaeus, 1758) [version 1; peer review: 2 approved, 3 approved with reservations]

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.

---

**Open Peer Review**

| Approval Status | 1 | 2 | 3 | 4 | 5 |
|-----------------|---|---|---|---|---|
| version 1       | ? | view | view | view | view |

1. Gregor Bucher, Georg-August-Universität Göttingen, Göttingen, Germany
2. Hermes E. Escalona, CSIRO, Canberra, Australia
3. Fahad Alqahtani, King Abdulaziz City for Science and Technology, Riyadh, Saudi Arabia
4. Themistoklis Giannoulis, University of Thessaly, Larissa, Greece
5. Shota Nomura, National Institute for Basic Biology, Okazaki, Japan

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; Coleoptera; Polyphaga; Cucujiformia; Coccinellidae; Coccinellinae; Coccinellini; 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.

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 1. Genome data for A. bipunctata, icAdaBipu1.

| Project accession data | Assembly identifier | icAdaBipu1.1 |
|------------------------|---------------------|--------------|
| Species                | Adalia bipunctata    |
| Specimen               | icAdaBipu1          |
| NCBI taxonomy ID       | 7084                |
| BioProject             | PRJEB45109          |
| BioSample ID           | SAMEA9089055        |
| Isolate information    | Male, whole organism|

Raw data accessions
- PacificBiosciences SEQUEL II: ERR7015066
- 10X Genomics Illumina: ERR6842405–ERR6842408
- Hi-C Illumina: ERR9866423–ERR9866427

Genome assembly
- Assembly accession: GCA_910592335.1
- Accession of alternate haplotype: GCA_910591895.1
- Span (Mb): 475
- Number of contigs: 212
- Contig N50 length (Mb): 18.3
- Number of scaffolds: 118
- Scaffold N50 length (Mb): 45.9
- Longest scaffold (Mb): 76.3
- BUSCO* genome score: C:97.6%, S:96.1%, D:1.4%, F:0.8%, M:1.6%, n:2124

Genome annotation
- Number of protein-coding genes: 13,611

*BUSCO scores based on the endopterygota_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/icAdaBipu1.1/dataset/CAJUZD01/busco.

Figure 1. An Adalia bipunctata image (Photograph from www.entomart.be CC-BY).
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.** 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 475,288,177 bp assembly. The distribution of chromosome lengths is shown in dark grey with the plot radius scaled to the longest chromosome present in the assembly (76,338,783 bp, shown in red). Orange and pale-orange arcs show the N50 and N90 chromosome lengths (45,869,983 and 27,047,566 bp), respectively. The pale grey spiral shows the cumulative chromosome 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 endopterygota_odb10 set is shown in the top right. An interactive version of this figure is available at 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 3. Genome assembly of A. bipunctata, icAdaBipul1.1: 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/icAdaBipul1.1/dataset/CAJUZD01/blob.
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 taxonomy. 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-dSN3RK37oemLsXTA.
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 icAda Bipu1 using the Arima v2 kit and sequenced on a HiSeq X instrument.

Genome assembly
Assembly of PacBio reads was carried out with HiFiasm (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 SALSA2 (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.

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

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

Table 3. Software tools used.

| Software tool | Version       | Source                                      |
|---------------|---------------|---------------------------------------------|
| BlobToolKit   | 3.4.0         | (Challis et al., 2020)                      |
| freebayes     | 1.3.1-17-gaa2ace8 | (Garrison & Marth, 2012)                  |
| gEVAL         | N/A           | (Chow et al., 2016)                        |
| HiFiasm       | 0.12          | (Cheng et al., 2021)                       |
| HiGlass       | 1.11.6        | (Kerpedjiev et al., 2018)                  |
| longranger    | align         |                                             |
|               | 2.2.2         | https://support.10xgenomics.com/             |
|               |               | genome-exome/software/pipelines/latest/     |
|               |               | advanced/other-pipelines                    |
| MitoHiFi      | 1.0           | (Uliano-Silva et al., 2021)                 |
| PretextView   | 0.1.x         | https://github.com/wtsi-hpag/PretextView    |
| purge_dups    | 1.2.3         | (Guan et al., 2020)                        |
| SALSA2        | 2.2           | (Ghurye et al., 2019)                      |

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/enalps: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
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 Wellcome Sanger Institute Scientific Operations: DNA Pipelines collective are listed here: https://doi.org/10.5281/zenodo.4790455.

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.478358.

References
Aken BL, Ayling S, Barrett D, et al.: The Ensembl gene annotation system. Database (Oxford). 2016; 2016: baw093. PubMed Abstract | Publisher Full Text | Free Full Text

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): 1361–1374. 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): 170–175. 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): 2508–2510. PubMed Abstract | Publisher Full Text | Free Full Text

Garrison E, Marth G: Haplotype-based variant detection from short-read sequencing. 2012. Publisher Full Text

Gautier M, Yamaguchi J, Foucaud J, et al.: The Genomic Basis of Color Pattern Polymorphism in the Harlequin Ladybird. Curr Biol. 2018; 28(20): 3296–3302.e7. PubMed Abstract | Publisher Full Text | Free Full Text

Ghurye J, Rhte 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): 2896–2898. PubMed Abstract | Publisher Full Text | Free Full Text

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

Howe K, Chow W, Collins J, et al.: Significantly improving the quality of genome assemblies through curation. GigaScience. Oxford University Press, 2021; 10(1): giaa153. PubMed Abstract | Publisher Full Text | Free Full Text

Kenis M, Nacambo S, Van Vaendere J, et al.: Long Term Monitoring in Switzerland Reveals that Adalia bipunctata Strongly Declines in Response to Harmonia axyridis Invasion. Insects. 2020; 11(12): 883. PubMed Abstract | Publisher Full Text | Free Full Text

Kerpedjiev P, Abdennour N, Lekshas 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

Mannii 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): 4647–4654. PubMed Abstract | Publisher Full Text | Free Full Text

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; 159(7): 1665–80. PubMed Abstract | Publisher Full Text | Free Full Text

Riddick EW: Identification of Conditions for Successful Aphid Control by Ladybirds in Greenhouses. Insects. 2017; 8(2): 38. PubMed Abstract | Publisher Full Text | Free Full Text

Rutkowski R, Gwiazdowska A, Ceryngier P, et al.: Microsatellite Polymorphism Suggests High Genetic Diversity but Disrupted Gene Flow in the Two-Spot Ladybird Adalia bipunctata (Linnaeus, 1758) (Coleoptera: Coccinellidae) Populations from Diverse Environments. Annales Zoologici. 2019; 69(2): 477–486. Publisher Full Text

Uliano-Silva M, Nunes JGF, Krascheninnikova K, et al.: MitoHIFI. 2021; (Accessed: 19 October 2022). Reference Source

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

Current Peer Review Status: ? ✔️ ? ✔️ ✔️

Version 1

Reviewer Report 27 September 2023

https://doi.org/10.21956/wellcomeopenres.20637.r65980

© 2023 Nomura 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.

Shota Nomura
Division of Evolutionary Developmental Biology, National Institute for Basic Biology, Okazaki, Japan

This study performed the chromosome-level genome assembly of the two-spot ladybird, *Adalia bipunctata*. Authors assembled the contigs using PacBio single-molecule HiFi long reads and scaffolded them using Hi-C data. As a result, the authors obtained 118 scaffolds with 45.9 Mb of N50 and 97.6% of BUSCO score. Among the assembly, 94.87% were assigned to 9 autosomes and X and Y sex chromosomes, which was consistent with the number of chromosomes observed in *A. bipunctata*.

Authors found high quality chromosome-level assembly, and the methods of analyses were appropriate and well explained. I consider this manuscript is worthy for publication. I note below a few minor questions.

1. How many ladybird species have chromosome-level genome sequences been published? I know the harlequin ladybird has been well studied and that genome assembly has been published, but has chromosome-level assembly been done for other ladybird species? Also, does the genome compositions or number of chromosomes in two-spot ladybird are different from other ladybirds that have chromosome-level genome assembly? This information would be useful to readers who use the assembled genome sequences and would be better to be explained in the Background section.

2. How did authors distinguish X and Y sex chromosomes from autosomes? If authors obtained male and female reads, they can be distinguished based on differences in reads mapping between the sexes. Did the authors use such a method? In any case, it would be better to be explained the methods of distinguish them.

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: genomic, ecological evolution, RNA-seq, speciation, Insect, genitalia

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 15 September 2023

https://doi.org/10.21956/wellcomeopenres.20637.r65975

© 2023 Giannoulis T. 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.

Themistoklis Giannoulis
Laboratory of Genetics, Faculty of Animal Science, University of Thessaly, Larissa, Greece

This is a very well presented research, the results are clearly described and the introduction has the proper length. The authors could provide more information about the bioinformatics processes (maybe as a supplementary file), p.ex from the arguments they used in freebayes etc. Also, since there was only 1 specimen, how were the variants called? Is there a reference genome already available?

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: Animal Genetics and 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 12 September 2023

https://doi.org/10.21956/wellcomeopenres.20637.r65976

© 2023 Alqahtani F. 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.

Fahad Alqahtani
National Center for Agricultural Technology, National Center for Agricultural Technology, King Abdulaziz City for Science and Technology, Riyadh, Saudi Arabia

The authors successfully sequenced the two-spot ladybird, Adalia bipunctata, in their study titled "The genome sequence of the two-spot ladybird, Adalia bipunctata (Linnaeus, 1758)". They employed three sequencing technologies: PacificBiosciences SEQUEL II, 10X Genomics Illumina, and Hi-C Illumina. The average coverage achieved was 48X for Pacific Biosciences single-molecule HiFi long reads and 80X for 10X Genomics.

To assemble the two-spot ladybird PacBio reads, the authors utilized Hifiasm, followed by polishing the PacBio contigs using the longranger align tool with 10X Genomics reads. Additionally, they performed scaffolding with Hi-C data using SALSA2. Overall, the genome assembly consisted of 118 sequence scaffolds, representing nine chromosomes and the sex chromosomes (X and Y). The total length of these sequences amounted to 475,288,177 base pairs. The quality of the assembly was assessed using BUSCO and yielded high scores.

However, there are some minor comments that need to be addressed. Firstly, in the method section, it would be beneficial to mention the closely related species used by the MitoHiFi/Mitofinder tool for guiding the annotation of the mitochondrial genome. Secondly, in the Genome Assembly section, the term "Pretext" should be corrected to "PreTextView".

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

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 12 September 2023

https://doi.org/10.21956/wellcomeopenres.20637.r65973

© 2023 Escalona H. 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.

Hermes E. Escalona
Australian National Insect Collection, CSIRO, Canberra, Australia

This is another great addition by DTOL. I am currently performing phylogenomic and comparative genomic research on ladybirds and other beetles and this resource is invaluable. I think that the manuscript is impeccable with some minor comments, see below.

- PI notice that *Harmonia axyridis* was introduced for biocontrol of Aphids worldwide. Certainly, it is an invasive species, but it was initially introduced. If possible please clarify in the Background section.

- The Genome sequence report mention 10X Genomics and Hi-C data, will it be possible to clarify in which order they were used?

- I found that approx 800 sequences from the cds annotation file have identical residues e.g. ENSABPT00005022599.1 and ENSABPT00005022609.1. Will it be possible to add a note explaining this and may be also in the annotation README release? I found a higher (<3000) number of sequences with identical residues in the ladybirds icCocSept1.1 and icHarAxyr1 generated by DTOL.

Please add a few more details on the Manual assembly curation so the methods are clear and reproducible. If not all the details are necessary, if possible at least please provide major guidelines.

**Is the rationale for creating the dataset(s) clearly described?**

Yes

**Are the protocols appropriate and is the work technically sound?**
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: Entomology, insect morphology, beetle taxonomy and systematics and molecular evolution in particular phylogenomics and comparative 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.

Reviewer Report 26 May 2023

https://doi.org/10.21956/wellcomeopenres.20637.r56737

© 2023 Bucher G. 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.

Gregor Bucher
Johann-Friedrich-Blumenbach-Institute for Zoology and Anthropology, Dept. of Evolutionary Developmental Genetics, Georg-August-Universitat Gottingen, Göttingen, Lower Saxony, Germany

The Tree of Life Programme team presents its efforts to sequence the genome of the two-spot ladybird Adalia bipunctata. This species is an important predator of aphids and is currently challenged by invasive ladybirds such that research on this animal is of importance.

PacBio long reads and 10X genomics read clouds were combined with Hi-C data to sequence and assemble the genome. Based on the high coverage and the combination of techniques, the result is an excellent genome sequence with chromosome level assembly including both sex chromosomes and a BUSCO score of 97.6%. A number of analyses such as GC content were done to characterize the genome further. The annotation revealed a small protein coding gene set (appr. 13,600 genes), which seems very surprising. In summary, a very valuable high quality genome sequence of an insect of importance is provided to the community and will surely be used in many future studies.

While this clearly is a valuable resource, I think that the paper could do more to inform the reader beyond the fact that an excellent sequence was provided. First, in its current form, the paper simply describes the procedure and results of the sequencing and some rather superficial analyses. I would have appreciated some more effort interpreting and putting the results into context. Some non-comprehensive suggestions: Does the number of scaffolded chromosomes
coincide with the cytological data? How does the genome size and GC content relate to other insects and beetles? Any property of the genome that deviates from expectations? Similarly, the results of the analyses shown in the figures are not well explained and not interpreted at all leaving it to the reader to make sense of those findings (see comments below).

My second concern is the annotation: A well annotated gene set is a very important aspect to make the resource useful for others. In that respect, I remain a bit unsure with respect to the comparably low number of genes. If true, this would represent a surprising case of gene loss, which should be highlighted. However, it could as well be an erroneous number due to a annotation issues. Actually, it appears that not much emphasis has been put into that topic: The annotation procedure is described rather briefly, information on quality control or the optimization of annotation parameters are not given and it appears that no transcriptomic data nor comparative annotation across species have been used to enhance the annotation. In summary, some more efforts to extract a good gene set would make the resource even more valuable unless such efforts are underway by groups dedicated to that challenge.

Finally, the presentation and description of some data could be improved in order to help understanding the analyses.

The following aspects have remained unclear to me or were hard to read in the current form:

- I remained a bit unsure, how the chromosomes were assembled from the contigs - I guess this was based on Hi-C data?

- **Figure 2** was hard to read for me. It may be more obvious for genomics aficionados - to help the "users" one could make the figure more intuitive. Some suggestions:
  - mark the extend of individual scaffolds with lines (similar to the red line marking the largest scaffold) and number them in order to relate them to the other data
  - mark the X and Y chromosomes
  - the use of colors for indicating the length of scaffolds was difficult to interpret and it leads to some weird depictions: The scaffold covering the N50 line seems to belong at the same time to N50 (orange) and larger scaffolds (grey). Does that scaffold really exactly represent the N50 value? Same with N90. Suggestion for an alternative: Mark N50/N90 simply with dotted circles instead of colors and the length of the scaffolds with a line instead of a color.
  - visualized on that scale the GC content seems to be essentially constant – hence, there is not much information that can be gained from the blue circles outside. Does the GC content really not vary along the chromosomes (e.g. in centromeric regions)? I would have loved to know, which chromosomes have the higher AT content (the smaller chromosomes? sex chromosomes?)

- **Figure 3**: I have remained unsure what actually was plotted here. The 118 scaffolds? I did not understand how you map a scaffold to a phylum? Based on homology of the contained protein coding genes? Isn’t it trivial that all scaffolds map to arthropods given that beetles belong to that clade? What do the axes represent? (what is gc coverage – do you mean gc
content? what is the axis “ERR6842....”? An interesting aspect appeared to be that there are scaffolds with clearly lower GC content - but what does that mean? Do these represent the centromers? Or the sex chromosomes? I have to confess that I have remained unsure what I can actually learn from that figure. Either the results are better described and the results interpreted or instead of that plot the GC content is plotted along each chromosome, which might be more informative.

○ **Figure 4:** same issue: what is plotted, how do you map scaffolds to phyla, what do I learn?

○ **Figure 5:** Please explain better, what is shown here and mark the chromosome scaffolds and name them (same numbering as in Fig. 2), mark the sex chromosomes. Any curious patterns worth mentioning?

**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?**
Partly

**Competing Interests:** No competing interests were disclosed.

**Reviewer Expertise:** Evolution, gene function, development, functional 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.