DATA NOTE

REVISED The genome sequence of the harlequin ladybird, *Harmonia axyridis* (Pallas, 1773)

[version 3; peer review: 3 approved, 1 not approved]

Douglas Boyes[1+], Liam M. Crowley[2],
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 & Hydrology, Wallingford, UK
[2]Department of Zoology, University of Oxford, Oxford, UK

[+] Deceased author

v3 **First published:** 08 Nov 2021, **6**:300
https://doi.org/10.12688/wellcomeopenres.17349.1
Second version: 17 Oct 2024, **6**:300
https://doi.org/10.12688/wellcomeopenres.17349.2
Latest published: 28 Nov 2024, **6**:300
https://doi.org/10.12688/wellcomeopenres.17349.3

Abstract

We present a genome assembly from an individual female *Harmonia axyridis* (the harlequin ladybird; Arthropoda; Insecta; Coleoptera; Coccinellidae). The genome sequence is 426 megabases in span. The majority (99.98%) of the assembly is scaffolded into 8 chromosomal pseudomolecules, with the X sex chromosome assembled.

Keywords

Harmonia axyridis, harlequin ladybird, genome sequence, chromosomal

This article is included in the Tree of Life gateway.

Open Peer Review

Approval Status ✓ ✗ ✓ ✓

	1	2	3	4
version 3 (revision) 28 Nov 2024			✓ view	✓ view
version 2 (revision) 17 Oct 2024	✓ view		? view	? view
version 1 08 Nov 2021	? view	✗ view		

1. **Arun Sethuraman**, San Diego State University, San Diego, USA
2. **Kang He**, Zhejiang University, Hangzhou, China
3. **Sudhakar Sivasubramaniam**, Manonmaniam Sundaranar University, Tirunelveli, India

Arun Arumugaperumal, Rajalakshmi Engineering College, Thandalam, Chennai, India

4. **Yu-Hao Huang**, Institute of Zoology, Guangdong Academy of Sciences, Guangzhou, 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;

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).
The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

How to cite this article: Boyes D, Crowley LM, University of Oxford and Wytham Woods Genome Acquisition Lab *et al.* **The genome sequence of the harlequin ladybird, *Harmonia axyridis* (Pallas, 1773) [version 3; peer review: 3 approved, 1 not approved]** Wellcome Open Research 2024, **6**:300 https://doi.org/10.12688/wellcomeopenres.17349.3

First published: 08 Nov 2021, **6**:300 https://doi.org/10.12688/wellcomeopenres.17349.1

REVISED **Amendments from Version 2**

We have corrected the details in the Genome sequence report to give the correct assembly length, scaffold number, scaffold N50 and chromosome number.

Any further responses from the reviewers 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; Harmonia; *Harmonia axyridis* (Pallas, 1773) (NCBI:txid115357).

Background

The harlequin ladybird, *Harmonia axyridis*, is large (5–8 mm), voracious ladybird species widely considered to be one of the world's most invasive insects. Its native range is central and eastern Asia, but it was introduced to North America and Europe as a biocontrol agent. It has spread rapidly and is now established across North, Central and South America, Europe and Africa. Examination of microsatellites has demonstrated that an invasive population in eastern North America acted as the source of those which invaded Europe, South America and South Africa (Lombaert *et al.*, 2010). The Harlequin ladybird was first recorded in the UK in 2003 in south-eastern England. Since its arrival it spread rapidly and is now widespread across the UK, and has been recorded on Ireland, Orkney, Shetland, the Channel Islands, the Isles of Scilly and the Isle of Man. It is a highly polymorphic species with several recognised forms. The colour of the elytra ranges from yellow, orange, red or black, with 0–21 black spots, 4 or 2 red/orange spots. The legs are always brown and the underside is dark with a reddish/brown border. The harlequin ladybird is a generalist, feeding on aphids as well as soft fruit, pollen, nectar and many other soft-bodied insects, including other ladybird larvae. It overwinters as an adult and is often found in buildings where aggregations of adults form. The haemolymph of this species contains high concentrations of isopropyl methoxy pyrazine (Al Abassi *et al.*, 1998) and harmonine (Nagel *et al.*, 2015) and it readily autohaemorrhages when agitated. The defensive secretions have a foul odour and can cause staining. Furthermore, it is also known to bite humans (Ramsey & Losey, 2012), leading to this species' consideration as a minor household pest. The spread of the harlequin ladybird is associated with dramatic declines in other, native ladybird species. This is believed to be driven by *Harmonia axyridis* outcompeting other aphidophagous species as well as intraguild predation (Majerus *et al.*, 2006).

Genome sequence report

The genome was sequenced from one female *H. axyridis* collected from Wytham Woods, Oxfordshire (biological vice-county: Berkshire), UK (latitude 51.772, longitude –1.338) (Figure 1). A total of 53-fold coverage in Pacific Biosciences single-molecule long reads and 93-fold coverage in 10X Genomics read clouds were generated. Primary assembly contigs were scaffolded with chromosome conformation Hi-C data. Manual assembly curation corrected 158 missing/misjoins, reducing the assembly length by 1.32% and the scaffold number by 92.49%, and increasing the scaffold N50 by 56.15%.

The final assembly has a total length of 425.5 Mb in 13 sequence scaffolds with a scaffold N50 of 63.7 Mb (Table 1). Most of the assembly sequence (99.98%) was assigned to 8 chromosomal-level scaffolds, representing 7 autosomes (numbered by sequence length), and the X sex chromosome (Figure 2–Figure 5; Table 2). Some scaffolds remain unplaced due to repetitive content giving an ambiguous Hi-C signal. A large cluster of rDNA sequences was placed on the X chromosome using Hi-C data only. The assembly has a BUSCO v5.1.2 (Manni *et al.*, 2021) completeness of 97.4% (single 95.2%, duplicated 2.3%) using the endopterygota_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, DNA extraction and sequencing

A single female *H. axyridis* was collected from Wytham Woods, Oxfordshire (biological vice-county: Berkshire), UK (latitude 51.772, longitude -1.338) by Douglas Boyes, UKCEH, using a pooter. The sample was identified by the same individual, and preserved on dry ice.

Figure 1. An image of the sequenced specimen, icHarAxyr1, captured immediately prior to processing and preservation.

Table 1. Genome data for *Harmonia axyridis*, icHarAxyr1.1.

Project accession data	
Assembly identifier	icHarAxyr1.1
Species	*Harmonia axyridis*
Specimen	icHarAxyr1
NCBI taxonomy ID	NCBI:txid346838
BioProject	PRJEB45202
BioSample ID	SAMEA7520208
Isolate information	Female, whole organism
Raw data accessions	
PacificBiosciences SEQUEL II	ERR6565943
10X Genomics Illumina	ERR6054990-ERR6054993
Hi-C Illumina	ERR6054994
Genome assembly	
Assembly accession	GCA_914767665.1
Accession of alternate haplotype	GCA_914767675.1
Span (Mb)	426
Number of contigs	186
Contig N50 length (Mb)	22.9
Number of scaffolds	13
Scaffold N50 length (Mb)	63.7
Longest scaffold (Mb)	87.9
BUSCO* genome score	C:97.4%[S:95.2%,D:2.3%],F:0.6%,M:2.0%,n:2,124

*BUSCO scores based on the endopterygota_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/icHarAxyr1.1/dataset/CAJZBN01/busco.

DNA was extracted from the whole organism of *H. axyridis* (icHarAxyr1) at the Wellcome Sanger Institute Scientific Operations core from the whole organism using the Qiagen MagAttract HMW DNA kit, according to the manufacturer's instructions. 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 on Pacific Biosciences SEQUEL II and Illumina HiSeq X instruments. Hi-C data were generated from the whole organism using the Arima v2 Hi-C kit and sequenced on a HiSeq X instrument.

All Wellcome Sanger Tree of Life wet lab protocols are available at DOI: 10.17504/protocols.io.8epv5xxy6g1b/v1.

Genome assembly

Assembly was carried out with Hifiasm (Cheng *et al.*, 2021) using the --primary option; haplotypic duplication was identified and removed with purge_dups (Guan *et al.*, 2020). One round of polishing was performed by aligning 10X Genomics read data to the assembly with LongRanger align, calling variants with Freebayes (Garrison & Marth, 2012). The assembly was then scaffolded with Hi-C data (Rao *et al.*, 2014) using SALSA2 (Ghurye *et al.*, 2019). The assembly was checked

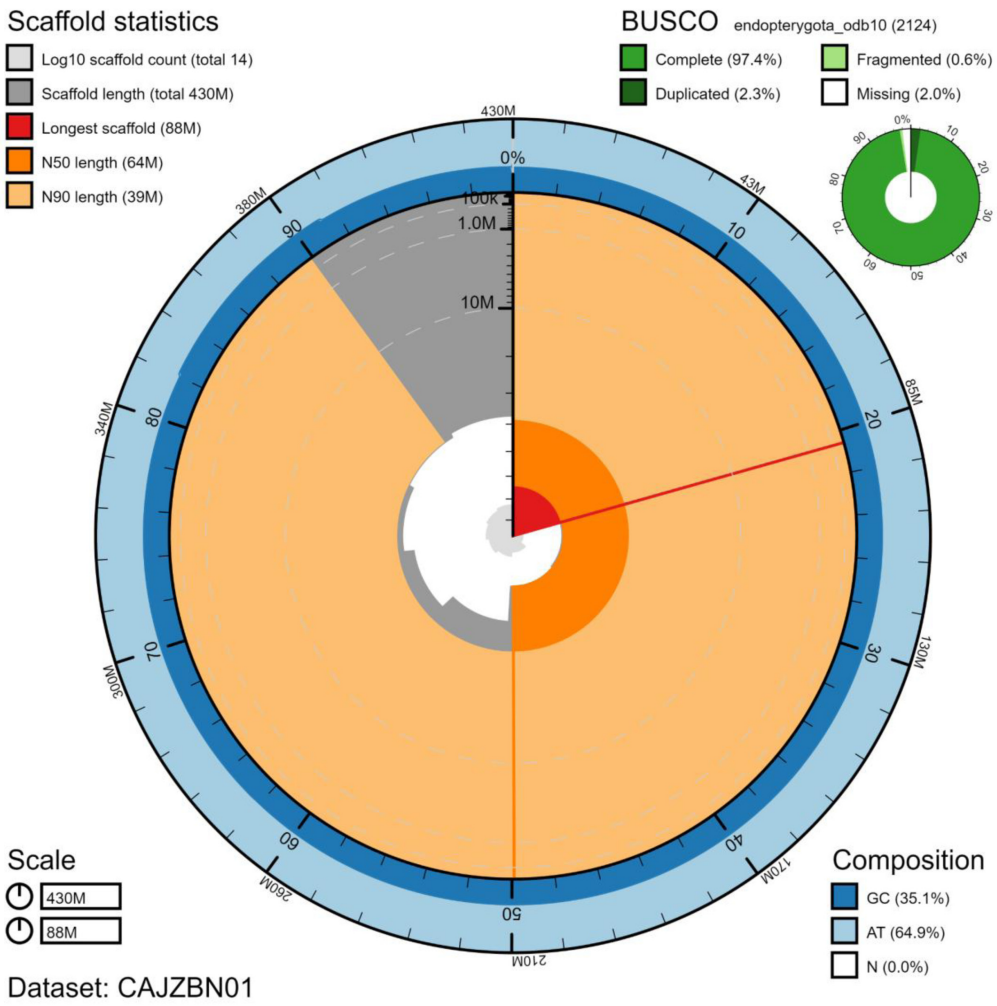

Figure 2. Genome assembly of *Harmonia axyridis*, icHarAxyr1.1: metrics. The BlobToolKit Snailplot shows N50 metrics and BUSCO gene completeness. The main plot is divided into 1,000 size-ordered bins around the circumference with each bin representing 0.1% of the 425,544,856 bp assembly. The distribution of scaffold lengths is shown in dark grey with the plot radius scaled to the longest scaffold present in the assembly (87,845,136 bp, shown in red). Orange and pale-orange arcs show the N50 and N90 scaffold lengths (63,675,256 and 38,596,305 bp), respectively. The pale grey spiral shows the cumulative scaffold count on a log scale with white scale lines showing successive orders of magnitude. The blue and pale-blue area around the outside of the plot shows the distribution of GC, AT and N percentages in the same bins as the inner plot. A summary of complete, fragmented, duplicated and missing BUSCO genes in the endopterygota_odb10 set is shown in the top right. An interactive version of this figure is available at https://blobtoolkit.genomehubs.org/view/icHarAxyr1.1/dataset/CAJZBN01/snail.

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). The assembly pipelines are available at https://pipelines.tol.sanger.ac.uk/pipelines and https://github.com/sanger-tol/. 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

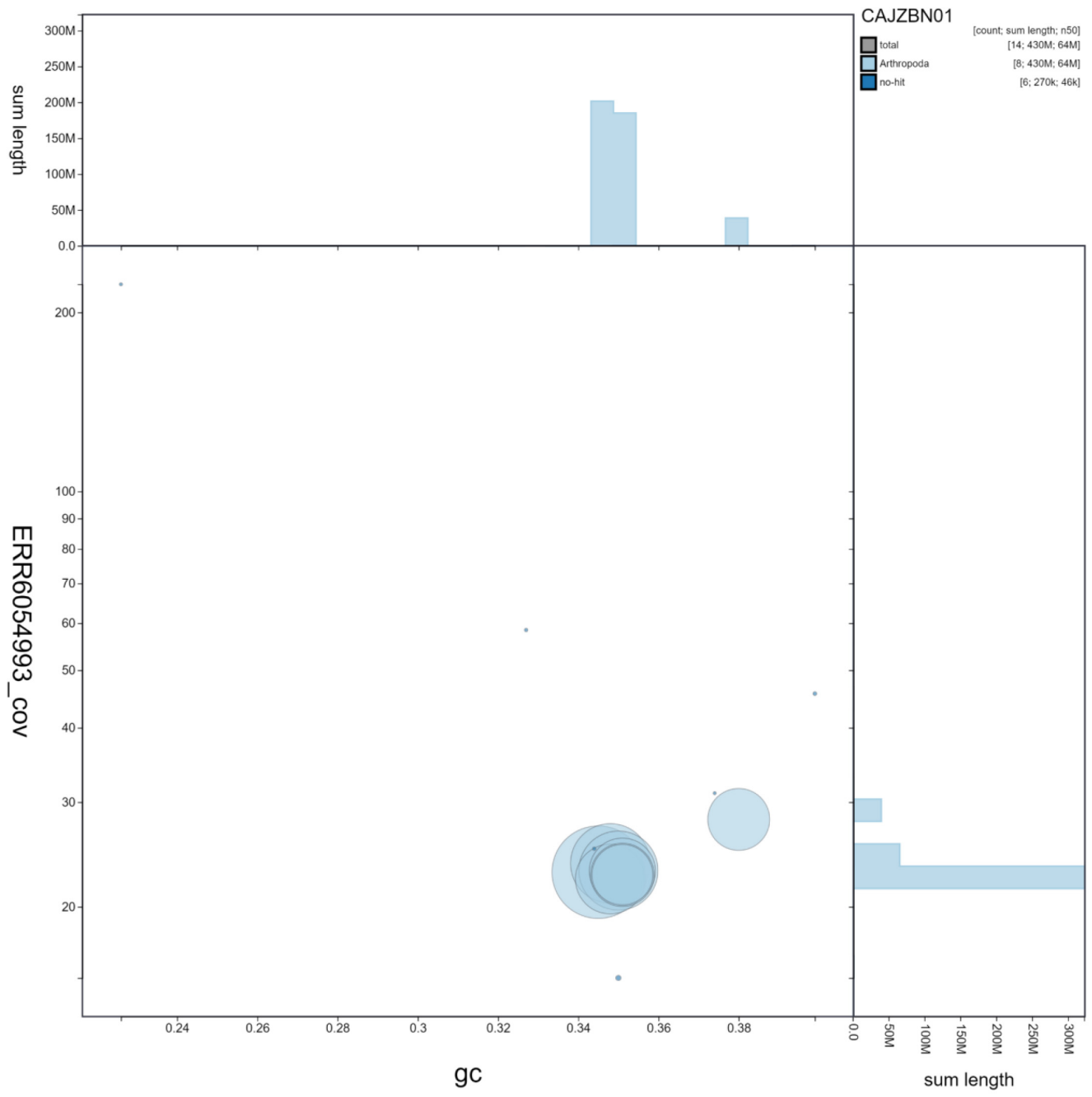

Figure 3. Genome assembly of *Harmonia axyridis*, icHarAxyr1.1: GC coverage. BlobToolKit GC-coverage plot. Scaffolds are coloured by phylum. Circles are sized in proportion to scaffold length Histograms show the distribution of scaffold length sum along each axis. An interactive version of this figure is available at https://blobtoolkit.genomehubs.org/view/icHarAxyr1.1/dataset/CAJZBN01/blob.

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),

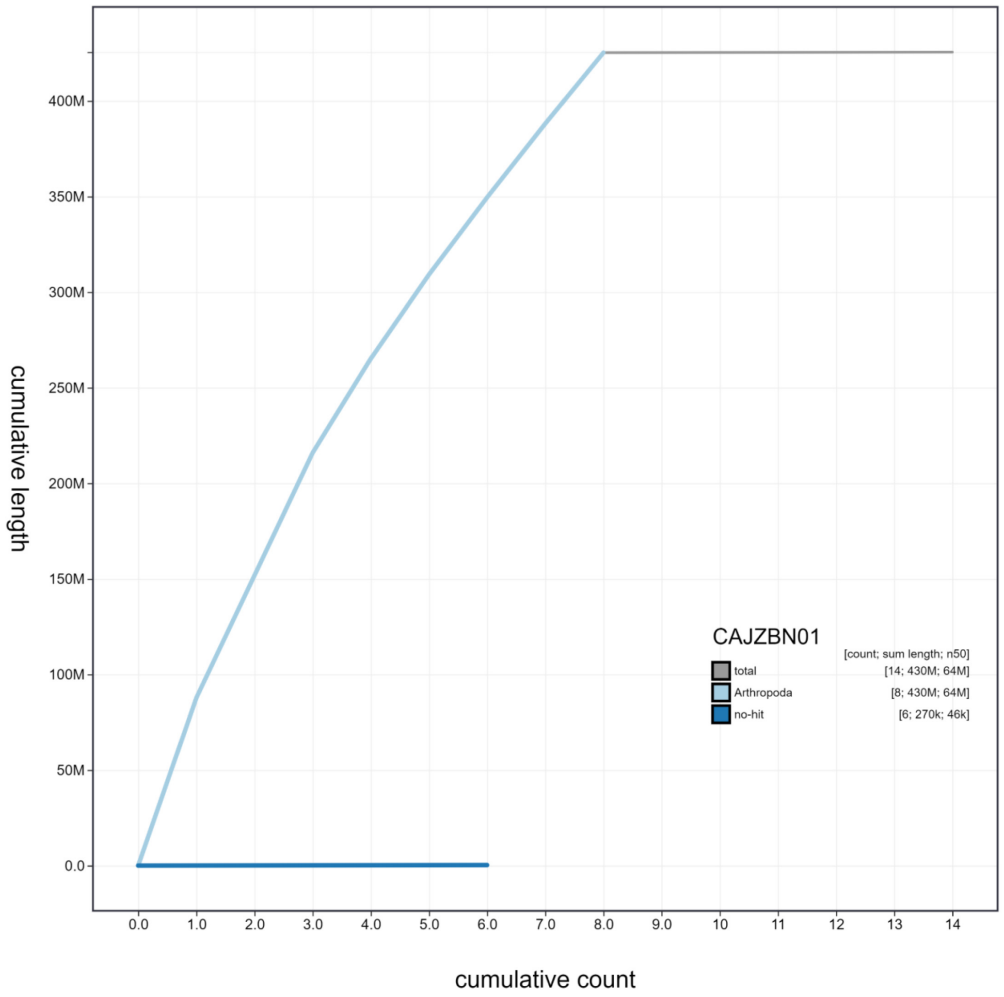

Figure 4. Genome assembly of *Harmonia axyridis*, icHarAxyr1.1: cumulative sequence. BlobToolKit cumulative sequence plot. The grey line shows cumulative length for all scaffolds. Coloured lines show cumulative lengths of scaffolds assigned to each phylum using the buscogenes taxrule. An interactive version of this figure is available at https://blobtoolkit.genomehubs.org/view/icHarAxyr1.1/dataset/CAJZBN01/cumulative.

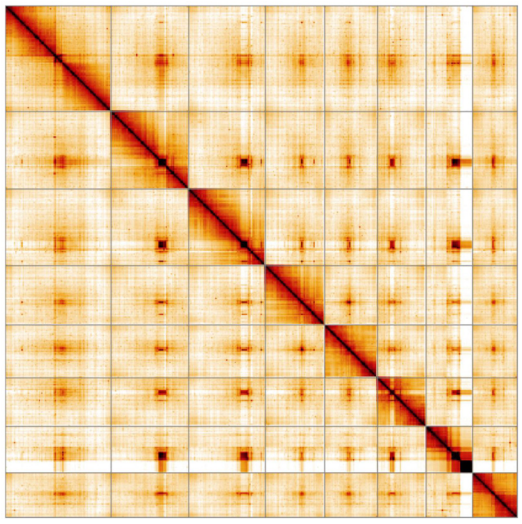

Figure 5. Genome assembly of *Harmonia axyridis*, icHarAxyr1.1: Hi-C contact map. Hi-C contact map of the icHarAxyr1.1 assembly, visualised in HiGlass.

Table 2. Chromosomal pseudomolecules in the genome assembly of *Harmonia axyridis*, icHarAxyr1.1.

INSDC accession	Chromosome	Size (Mb)	GC%
OU611927.1	1	87.85	34.5
OU611928.1	2	64.43	34.8
OU611929.1	3	63.68	35
OU611930.1	4	49.28	34.8
OU611931.1	5	43.98	35.1
OU611932.1	6	40.34	35.1
OU611934.1	7	37.13	35.1
OU611933.1	X	38.60	38
OU611935.1	MT	0.02	22.6
-	Unplaced	0.25	35.7

Table 3. Software tools used.

Software tool	Version	Source
Hifiasm	0.12	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	1.0	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

and in some circumstances other Darwin Tree of Life collaborators.

Data availability

European Nucleotide Archive: Harmonia axyridis (harlequin). Accession number PRJEB45202; https://identifiers.org/ena.embl/PRJEB45202.

The genome sequence is released openly for reuse. The *H. axyridis* 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

Al Abassi S, Birkett MA, Pettersson J, *et al.*: **Ladybird beetle odour identified and found to be responsible for attraction between adults.** *Cell Mol Life Sci.* 1998; **54**(8): 876–79.
Publisher Full Text

Challis R, Richards E, Rajan J, *et al.*: **BlobToolKit - interactive quality assessment of genome assemblies.** *G3 (Bethesda).* 2020; **10**(4): 1361–74.
PubMed Abstract | Publisher Full Text | Free Full Text

Cheng H, Concepcion GT, Feng X, *et al.*: **Haplotype-resolved *de novo* assembly using phased assembly graphs with Hifiasm.** *Nat Methods.* 2021; **18**(2): 170–75.
PubMed Abstract | Publisher Full Text | Free Full Text

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

Garrison E, Marth G: **Haplotype-based variant detection from short-read sequencing.** arXiv: 1207.3907. 2012.
Publisher Full Text

Ghurye J, Rhie A, Walenz BP, *et al.*: **Integrating Hi-C links with assembly graphs for chromosome-scale assembly.** *PLoS Comput Biol.* 2019; **15**(8): e1007273.
PubMed Abstract | Publisher Full Text | Free Full Text

Guan D, McCarthy SA, Wood J, *et al.*: **Identifying and removing haplotypic duplication in primary genome assemblies.** *Bioinformatics.* 2020; **36**(9): 2896–98.
PubMed Abstract | Publisher Full Text | Free Full Text

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

Kerpedjiev P, Abdennur N, Lekschas F, *et al.*: **HiGlass: web-based visual exploration and analysis of genome interaction maps.** *Genome Biol.* 2018; **19**(1): 125.
PubMed Abstract | Publisher Full Text | Free Full Text

Lombaert E, Guillemaud T, Cornuet JM, *et al.*: **Bridgehead effect in the worldwide invasion of the biocontrol harlequin ladybird.** *PLoS One.* 2010; **5**(3): e9743.
PubMed Abstract | Publisher Full Text | Free Full Text

Majerus M, Strawson V, Roy H: **The potential impacts of the arrival of the harlequin ladybird, *Harmonia Axyridis* (Pallas) (coleoptera: coccinellidae), in Britain.** *Ecol Entomol.* 2006; **31**(3): 207–15.
Publisher 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

Nagel NC, Masic A, Schurig U, *et al.*: **Efficient synthesis of (*R*)-Harmonine--the toxic principle of the multicolored asian lady beetle (*Harmonia Axyridis*).** *Org Biomol Chem.* 2015; **13**(18): 5139–46.
PubMed Abstract | Publisher Full Text

Ramsey S, Losey JE: **Why is *Harmonia Axyridis* the culprit in coccinellid biting incidents? An analysis of means, motive, and opportunity.** *American Entomologist.* 2012; **58**(3): 166–70.
Publisher 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

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 3

Reviewer Report 02 December 2024

https://doi.org/10.21956/wellcomeopenres.25898.r113705

✔ **Yu-Hao Huang**
Guangdong Key Laboratory of Animal Conservation and Resource Utilization, Guangdong Public Laboratory of Wild Animal Conservation and Utilization, Institute of Zoology, Guangdong Academy of Sciences, Guangzhou, China

The revised version has addressed my primary concern, and I have no further comments.

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: comparative genomics, evolutionary biology, bioinformatics, phylogenetics, ladybird beetle

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 02 December 2024

https://doi.org/10.21956/wellcomeopenres.25898.r113707

✔ **Sudhakar Sivasubramaniam**
Manonmaniam Sundaranar University, Tirunelveli, Tamil Nadu, India

Arun Arumugaperumal
Department of Biotechnology, Rajalakshmi Engineering College, Thandalam, Chennai, Tamilnadu, 602105, India

The authors have corrected the assembly parameters in text as per Table 1. The data note can be accepted.

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Molecular Biology, Transcriptomics, Genomics, Regeneration biology

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

Version 2

Reviewer Report 25 November 2024

https://doi.org/10.21956/wellcomeopenres.23551.r112729

?

Yu-Hao Huang
Guangdong Key Laboratory of Animal Conservation and Resource Utilization, Guangdong Public Laboratory of Wild Animal Conservation and Utilization, Institute of Zoology, Guangdong Academy of Sciences, Guangzhou, China

The data note presents a high-quality chromosome-level genome assembly of the ladybird beetle *Harmonia axyridis*, demonstrating high completeness and continuity that surpasses other published genomes of this species. This genome assembly will be instrumental for appropriate utilization of this potential biocontrol agent and the mitigation of its impact in introduced regions. The sequencing and bioinformatics methods are sound, and the data is accessible.

However, before approving this version of the manuscript, I have a major concern. The descriptive text of the genome assembly (total length: 249 Mb, scaffold number: 39, scaffold N50: 37.2 Mb, 10 or 9? chromosomal-level scaffolds including 8 autosomes and the X sex chromosome) seems different with Tables 1 and 2 (total length: 426 Mb, scaffold number: 13, scaffold N50: 63.7 Mb, 7 autosomes and the X sex chromosome). Based on my experience of using this genome assembly, the information in the tables is closer to the actual one. Please verify and correct the descriptive text if necessary, or provide an explanation for the discrepancy (e.g., before and after curation).

Additionally, there are at least three previously published genomes of *H. axyridis* (Ando, et al. 2018; Gautier, et al. 2018; Chen, et al. 2021). It is better to mention these genomes in the background to provide readers with a better understanding of the current genome assembly.

References

1. Ando T, Matsuda T, Goto K, Hara K, et al.: Repeated inversions within a pannier intron drive

diversification of intraspecific colour patterns of ladybird beetles.*Nat Commun*. 2018; **9** (1): 3843 PubMed Abstract | Publisher Full Text
2. Gautier M, Yamaguchi J, Foucaud J, Loiseau A, 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
3. Chen M, Mei Y, Chen X, Chen X, et al.: A chromosome-level assembly of the harlequin ladybird Harmonia axyridis as a genomic resource to study beetle and invasion biology.*Mol Ecol Resour*. 2021; **21** (4): 1318-1332 PubMed Abstract | Publisher Full Text

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: comparative genomics, evolutionary biology, bioinformatics, phylogenetics, ladybird beetle

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 07 November 2024

https://doi.org/10.21956/wellcomeopenres.23551.r106346

?

Sudhakar Sivasubramaniam
Manonmaniam Sundaranar University, Tirunelveli, Tamil Nadu, India
Arun Arumugaperumal
Department of Biotechnology, Rajalakshmi Engineering College, Thandalam, Chennai, Tamilnadu, 602105, India

The authors have sequenced the genome of the ladybird beetle, *Harmonia axyridis*. Already a genome sequence for the insect is available [1]. This assembly has a higher N50 value compared

to the previously reported one. It is good to have an improved assembly. But the authors should cite the already published article and mention this. Then only readers can easily understand the difference between the assemblies. Ignoring the already published article may lead to confusion.

Also I have few suggestions,
1. In Background section, the line "The final assembly has a total length of 249 Mb in 39 sequence scaffolds with a scaffold N50 of 37.2 Mb (Table 1)." mentions that scaffold N50 is 37.2 Mb and refers table 1. But in Table 1 scaffold N50 is mentioned as 63.7. I can understand that the latter scaffold is after curation. But readers may get confused. So kindly rephrase the sentence.

2. A scale could have been included in the specimen photograph. If the author has any other phot graph with a scale for this specimen then the photograph can be replaced.

3. Did the authors get mitogenome as a single scaffold? If yes, then that can be mentioned in this data article.

References

1. Chen M, Mei Y, Chen X, Chen X, et al.: A chromosome-level assembly of the harlequin ladybird Harmonia axyridis as a genomic resource to study beetle and invasion biology.*Mol Ecol Resour*. 2021; **21** (4): 1318-1332 PubMed Abstract | Publisher Full Text

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: Molecular Biology, Transcriptomics, Genomics, Regeneration biology

We confirm that we have read this submission and believe that we have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however we have significant reservations, as outlined above.

Reviewer Report 01 November 2024

https://doi.org/10.21956/wellcomeopenres.23551.r105748

Arun Sethuraman
San Diego State University, San Diego, California, USA

I don't have any other comments.

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: Genomics, population genetics, evolution

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.

Version 1

Reviewer Report 31 May 2023

https://doi.org/10.21956/wellcomeopenres.19179.r56659

Kang He
Institute of Insect Science, Zhejiang University, Hangzhou, Zhejiang, China

In the present manuscript, Douglas *et al.* report a chromosome-level genome assembly of harlequin ladybird, I believe this dataset should be valuable for the comparative genomic study of species related. I only have several major concerns on this manuscript.

1. Given that a previous version of the chromosomal level genome assembly has been

published (Chen *et al.*, 2021[1]), a lower BUSCO score of this version for the genome assembly is presented than the previous version. The authors are encouraged to explain the reason.

2. The 'Genome sequence report' mentions that the assembly sequence was assigned to 8 autosomes, including the X and Y sex chromosomes. However, in the method section, it is mentioned that the sequence data was from a single female H. axyridis. How can female's sequence data be assigned to the Y chromosome? Would males have DNA that the females don't have, and if so is this missing information in the assembly? Why not use male for sequencing?

3. The parameters in detail of the softwares for genome assembly are missing in the method part.

References

1. Chen M, Mei Y, Chen X, Chen X, et al.: A chromosome-level assembly of the harlequin ladybird Harmonia axyridis as a genomic resource to study beetle and invasion biology.*Mol Ecol Resour*. 2021; **21** (4): 1318-1332 PubMed Abstract | Publisher Full Text

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

Yes

Are the protocols appropriate and is the work technically sound?

No

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: Insect genomics, Non-conding RNA

I confirm that I have read this submission and believe that I have an appropriate level of expertise to state that I do not consider it to be of an acceptable scientific standard, for reasons outlined above.

Author Response 10 Oct 2024

Tree of Life Team Sanger

Please see our responses to each part of the peer review report below. In the present manuscript, Douglas *et al.* report a chromosome-level genome assembly of harlequin ladybird, I believe this dataset should be valuable for the comparative genomic study of species related. I only have several major concerns on this manuscript.

1. Given that a previous version of the chromosomal level genome assembly has been

published (Chen *et al.*, 2021), a lower BUSCO score of this version for the genome assembly is presented than the previous version. The authors are encouraged to explain the reason.

Response: The BUSCO score is actually **higher** when using the *insecta* dataset, as in Chen et al. (2021):

- Chen et al., 2021 reports 97.5% completeness, 1% fragmented (BUSCO v3, *insecta_odb9*).
- This data note (Boyes et al., 2021) reports 97.4% completeness, 0.6% fragmented (BUSCO v5.1.2, *endopterygota_odb10*). However, the BUSCO score for the assembly presented here is 99% when evaluated with *insecta_odb10* (see link: https://blobtoolkit.genomehubs.org/view/icHarAxyr1.1/dataset/CAJZBN01/busco).

1. The 'Genome sequence report' mentions that the assembly sequence was assigned to 8 autosomes, including the X and Y sex chromosomes. However, in the method section, it is mentioned that the sequence data was from a single female H. axyridis. How can female's sequence data be assigned to the Y chromosome? Would males have DNA that the females don't have, and if so is this missing information in the assembly? Why not use male for sequencing?

Response: This error has been corrected in the data note—no Y chromosome is present in this female specimen. A male was not used for sequencing because Darwin Tree of Life specimens are provided by collectors based on availability during collection events.

1. The parameters in detail of the softwares for genome assembly are missing in the method part.

Response: We have now included links to the protocols that detail how to run all the pipelines used in the Tree of Life genome assembly (see the Tree of Life bioinformatics pipelines at https://pipelines.tol.sanger.ac.uk/pipelines and https://github.com/sanger-tol/).

Is the rationale for creating the dataset(s) clearly described? Yes **Are the protocols appropriate and is the work technically sound?** No *Response*: The genome assembly presented in this data note is the current reference genome for *Harmonia axyridis*, assembled using the standard Darwin Tree of Life pipeline. These methods are openly available, and all scripts can be downloaded (see links provided above). We also provide a link to the wet lab protocol collection at DOI:10.17504/protocols.io.8epv5xxy6g1b/v1. **Are sufficient details of methods and materials provided to allow replication by others?** Partly *Response*: We have now included links to comprehensive resources for replicating the methods used, including both wet lab protocols and bioinformatics pipelines. **Are the datasets clearly presented in a useable and accessible format?** Yes

Competing Interests: No competing interests were disclosed.

Reviewer Report 03 February 2022

https://doi.org/10.21956/wellcomeopenres.19179.r48049

?

Arun Sethuraman
San Diego State University, San Diego, California, USA

The authors have presented a highly contiguous, high quality genome assembly for *Harmonia axyridis* from the UK. I am impressed at the completeness and contiguity of this assembly, and it just might be the most complete genome published for *H. axyridis* thus far. However, I think that this manuscript will benefit from: (1) comparison of assemblies with published ones out there, including Chen *et al*., 2021[1]; (2) sharing all software scripts/pipelines that were utilized in this assembly process (which are not currently available); and (3) additional details of assembly polishing utilized.

Additionally, I acknowledge that this might be beyond the scope of this data note, but provision of a minimal annotation would be of great help to readers.

References

1. Chen M, Mei Y, Chen X, Chen X, et al.: A chromosome-level assembly of the harlequin ladybird Harmonia axyridis as a genomic resource to study beetle and invasion biology.*Mol Ecol Resour*. 2021; **21** (4): 1318-1332 PubMed Abstract | Publisher Full Text

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?
No

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

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Genomics, population genetics, evolution

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.