




DATA NOTE

The genome sequence of the common red soldier beetle, *Rhagonycha fulva* (Scopoli, 1763) [version 1; peer review: 2 approved]

Liam M. Crowley ¹,
 University of Oxford and Wytham Woods Genome Acquisition Lab,
 Natural History Museum 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

¹Department of Zoology, University of Oxford, Oxford, OX1 3SZ, UK

V1 First published: 22 Sep 2021, 6:243
<https://doi.org/10.12688/wellcomeopenres.17198.1>
 Latest published: 22 Sep 2021, 6:243
<https://doi.org/10.12688/wellcomeopenres.17198.1>

Abstract

We present a genome assembly from an individual female *Rhagonycha fulva* (the common red soldier beetle; Arthropoda; Insecta; Coleoptera; Cantharidae). The genome sequence is 425 megabases in span. The majority of the assembly is scaffolded into seven chromosomal pseudomolecules, with the X sex chromosome assembled.

Keywords





Rhagonycha fulva, common red soldier beetle, genome sequence, chromosomal



This article is included in the [Tree of Life gateway](#).

Open Peer Review

Approval Status  

	1	2
version 1 22 Sep 2021	 view	 view
1. Ladislav Bocak  ,	Biodiversity & Molecular Evolution Olomouc, Olomouc, Czech Republic	
2. Yi-Ming Weng  ,	University of Wisconsin Madison, Madison, USA	

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: Crowley LM: Investigation, Resources, 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.

Copyright: © 2021 Crowley LM *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: Crowley LM, University of Oxford and Wytham Woods Genome Acquisition Lab, Natural History Museum Genome Acquisition Lab *et al.* **The genome sequence of the common red soldier beetle, *Rhagonycha fulva* (Scopoli, 1763) [version 1; peer review: 2 approved]** Wellcome Open Research 2021, 6:243 <https://doi.org/10.12688/wellcomeopenres.17198.1>

First published: 22 Sep 2021, 6:243 <https://doi.org/10.12688/wellcomeopenres.17198.1>

Species taxonomy

Eukaryota; Metazoa; Ecdysozoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Coleoptera; Polyphaga; Elateriformia; Elateroidea; Cantharidae; Cantharinae; Rhagonycha; *Rhagonycha fulva* (Scopoli, 1763) (NCBI:txid100007).

Introduction

The common red soldier beetle, *Rhagonycha fulva*, is the most abundant and widespread soldier beetle (Coleoptera: Cantharidae) in the UK. They can be found in a variety of habitats, where adults are frequently encountered on the flowers of umbellifers (Apiaceae), thistles (Asteraceae) and ragwort (*Senecio jacobaea*) throughout the summer. It can be particularly abundant on the flowers of common hogweed, *Heracleum sphondylium* (Grace & Nelson, 1981), and their association with flowers indicates this species' potential role as an important pollinator. Adults are predatory on small insects, but also feed extensively on floral resources. They are diurnal and fly readily, males in particular are highly mobile (Rodwell *et al.*, 2018). Mating occurs over a prolonged period of time, meaning female-male pairs are often encountered in copulation. Eggs are laid into the soil and the larvae are predatory, hunting amongst the leaf litter.

Adults can be easily recognised by the extensive reddish colour of the entire body with black tips to the elytra and black tarsi, antennae and palps.

Genome sequence report

The genome was sequenced from one female *R. fulva* collected from Wytham farm, Oxfordshire (biological vice-county: Berkshire), UK (latitude 51.779, longitude -1.317). A total of 41-fold coverage in Pacific Biosciences single-molecule long reads and 103-fold coverage in 10X Genomics read clouds were generated. Primary assembly contigs were scaffolded with chromosome conformation Hi-C data. Manual assembly curation corrected 73 missing/misjoins and removed 12 haplotypic duplications, reducing the assembly length by 1.54% and the scaffold number by 84.62%, and increasing the scaffold N50 by 238.56%. The final assembly has a total length of 425 Mb in 13 sequence scaffolds with a scaffold N50 of 116 Mb (Table 1). The majority, 99.97%, of the assembly sequence was assigned to seven chromosomal-level scaffolds, representing six autosomes (numbered by sequence length), and the X sex chromosome (Figure 1–Figure 4; Table 2). The assembly has a BUSCO v5.1.2 (Simão *et al.*, 2015) completeness of 98.9% 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.

Chromosome 1 contains a largely homogeneous, heterochromatic block at 24.45–95.85 Mb (Figure 4), in accordance with existing karyotyping (see Figure 3 of (James & Angus, 2007)). This block consists of numerous scaffolds with high repeat content that can be localised to chromosome 1, but their order and

Table 1. Genome data for *Rhagonycha fulva*, icRhaFulv1.1.

Project accession data	
Assembly identifier	icRhaFulv1.1
Species	<i>Rhagonycha fulva</i>
Specimen	icRhaFulv1
NCBI taxonomy ID	NCBI:txid411101
BioProject	PRJEB43742
BioSample ID	SAMEA7520319
Isolate information	Female, whole organism
Raw data accessions	
PacificBiosciences SEQUEL II	ERR6606788
10X Genomics Illumina	ERR6054565-ERR6054568
Hi-C Illumina	ERR6054569
RNaseq PolyA Illumina	ERR6286718
Genome assembly	
Assembly accession	GCA_905340355.1
Accession of alternate haplotype	GCA_905340395.1
Span (Mb)	425
Number of contigs	82
Contig N50 length (Mb)	21
Number of scaffolds	13
Scaffold N50 length (Mb)	116
Longest scaffold (Mb)	125
BUSCO* genome score	C:98.9%[S:96.0%,D:2.9%],F:0.6%,M:0.6%,n:2124

*BUSCO scores based on the coccidia_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/icRhaFulv1.1/dataset/CAJPIC01.1/busco>.

orientation is unsure. It is likely that the assembly overexpands this region due to difficulties in identifying and removing haplotypic duplications.

Methods

A single female *R. fulva* was collected from Wytham farm, Oxfordshire (biological vice-county: Berkshire), UK (latitude 51.779, longitude -1.317) by Liam Crowley, University of Oxford, and snap-frozen on dry ice using a CoolRack. A second specimen of unknown sex, icRhaFulv4, was collected from Wigmore Park, Luton, UK (latitude 51.88378, longitude -0.36861422) by Olga Sivell, Natural History Museum, and snap-frozen on dry ice.

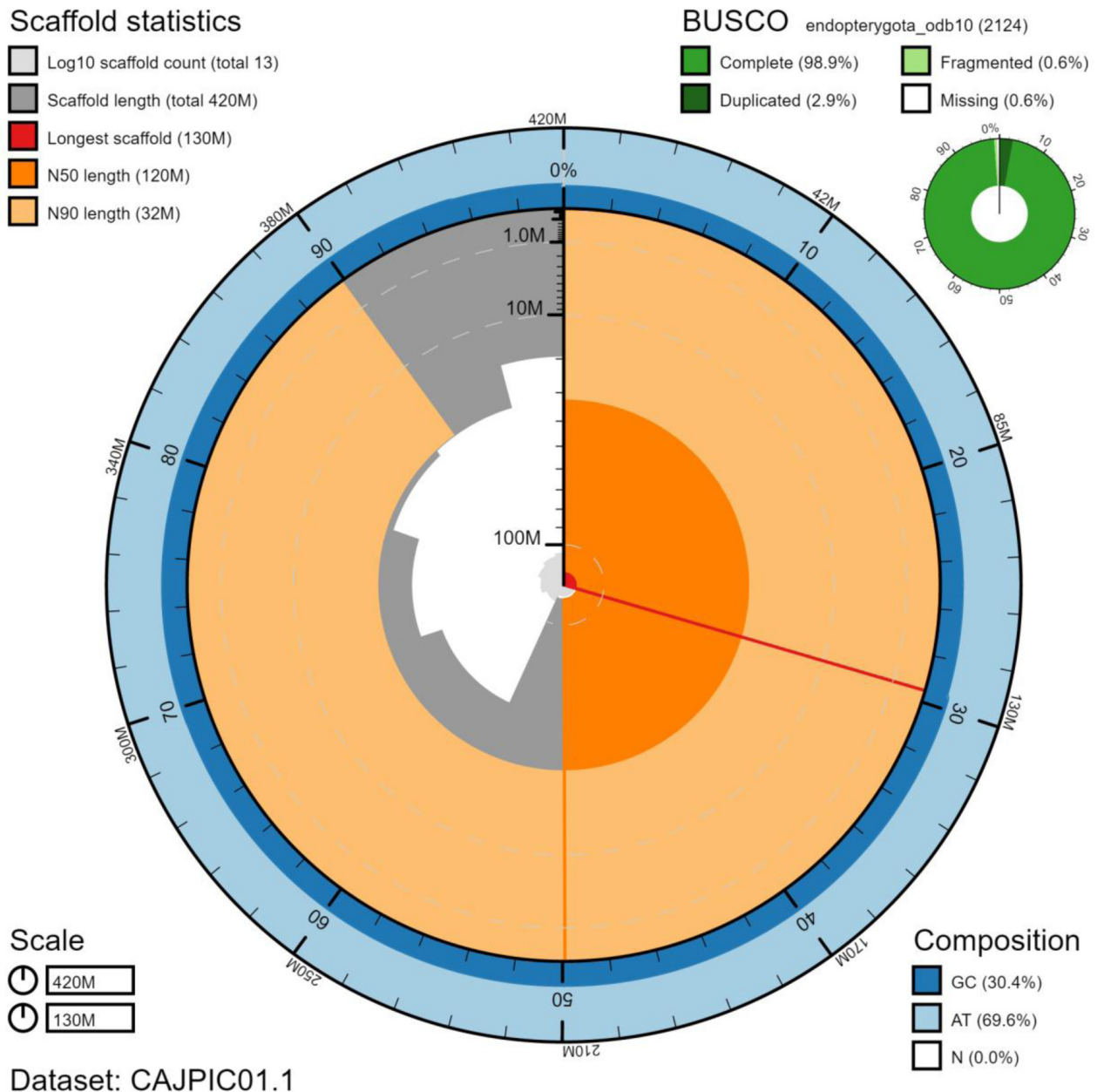


Figure 1. Genome assembly of *Rhagonycha fulva*, icRhaFulv1.1: metrics. The BlobToolKit Snailplot shows N50 metrics and BUSCO gene completeness. An interactive version of this figure is available at <https://blobtoolkit.genomehubs.org/view/icRhaFulv1.1/dataset/CAJPIC01.1/snail>.

DNA was extracted from the whole organism of icRhaFulv1 at the Wellcome Sanger Institute (WSI) Scientific Operations core from the whole organism using the Qiagen MagAttract HMW DNA kit, according to the manufacturer's instructions. RNA (from the whole organism of icRhaFulv4) was extracted in the Tree of Life Laboratory at the WSI using TRIzol, according to the manufacturer's instructions. RNA was then eluted in 50 μ l RNase-free water and its concentration RNA assessed using a Nanodrop spectrophotometer and Qubit Fluorometer

using the Qubit RNA Broad-Range (BR) Assay kit. Analysis of the integrity of the RNA was done using Agilent RNA 6000 Pico Kit and Eukaryotic Total RNA assay.

Pacific Biosciences HiFi circular consensus and 10X Genomics read cloud DNA sequencing libraries, in addition to PolyA RNA-Seq libraries, were constructed according to the manufacturers' instructions. Sequencing was performed by the Scientific Operations core at the WSI on Pacific

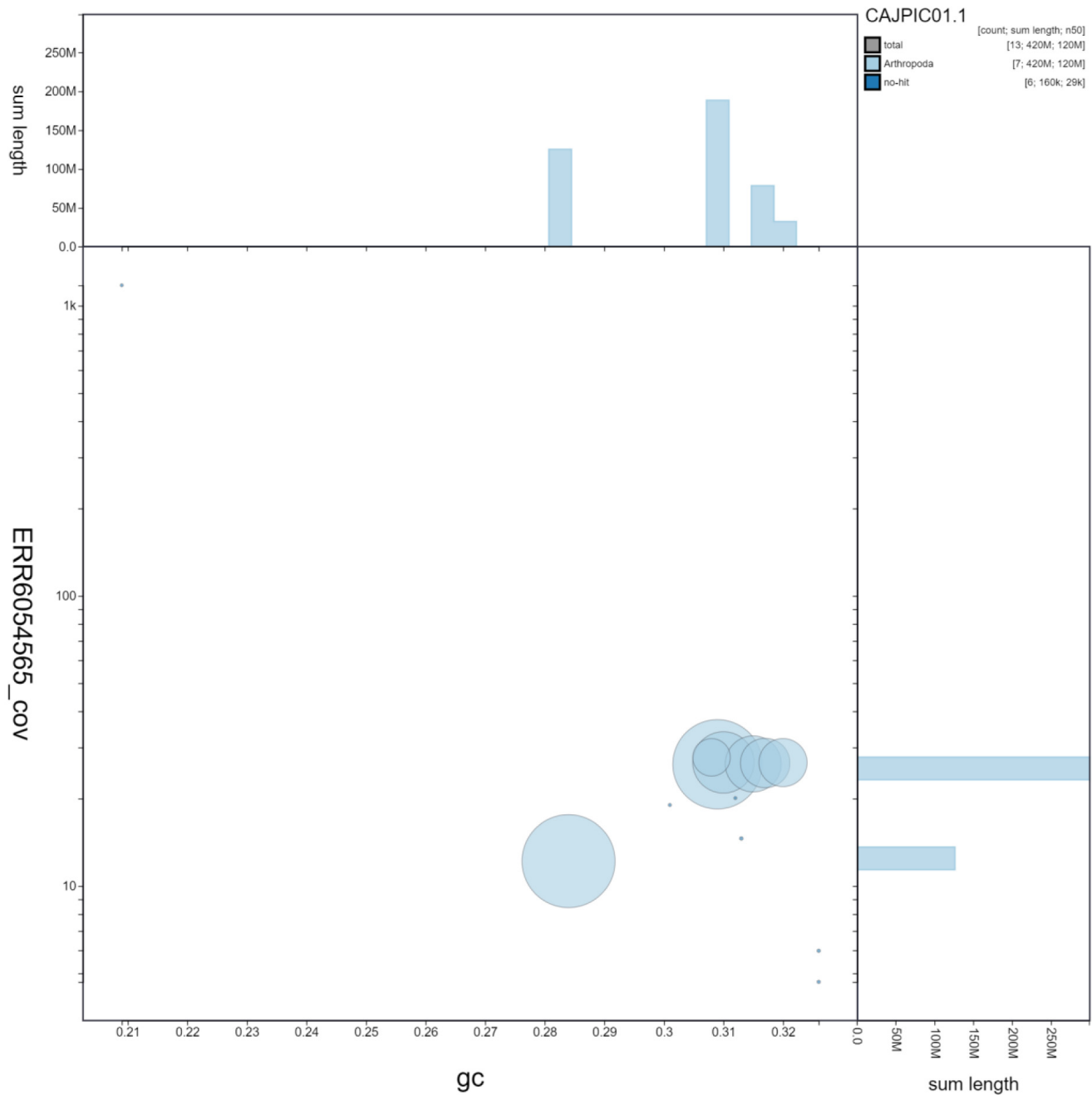


Figure 2. Genome assembly of *Rhagonycha fulva*, icRhaFulv1.1: GC coverage. BlobToolKit GC-coverage plot. Scaffolds are coloured by phylum. Circles are sized in proportion to scaffold length. Histograms show the distribution of scaffold length sum along each axis. An interactive version of this figure is available at <https://blobtoolkit.genomehubs.org/view/icRhaFulv1.1/dataset/CAJPIC01.1/blob>.

Biosciences SEQUEL II (HiFi), Illumina HiSeq X (10X) and Illumina HiSeq 4000 (RNA-Seq) instruments. Hi-C data were generated using the Arima v2 Hi-C kit and sequenced on a HiSeq X instrument.

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 performed by aligning 10X Genomics read data to the

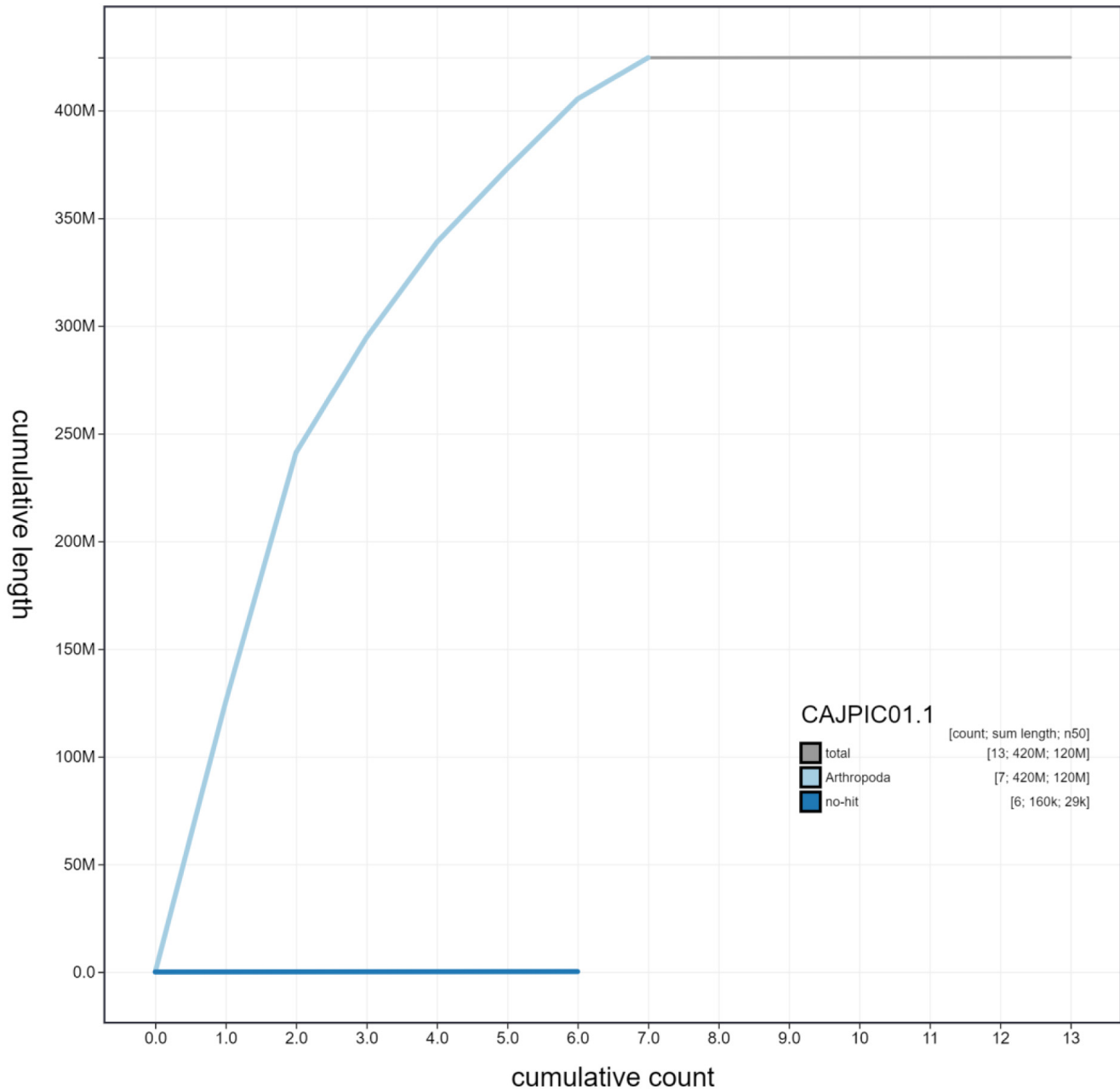


Figure 3. Genome assembly of *Rhagonycha fulva*, icRhaFulv1.1: cumulative sequence. BlobToolKit cumulative sequence plot. The grey line shows cumulative length for all scaffolds. Coloured lines show cumulative lengths of scaffolds assigned to each phylum using the buscogenes taxrule. An interactive version of this figure is available at <https://blobtoolkit.genomehubs.org/view/icRhaFulv1.1/dataset/CAJPIC01.1/cumulative>.

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 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.

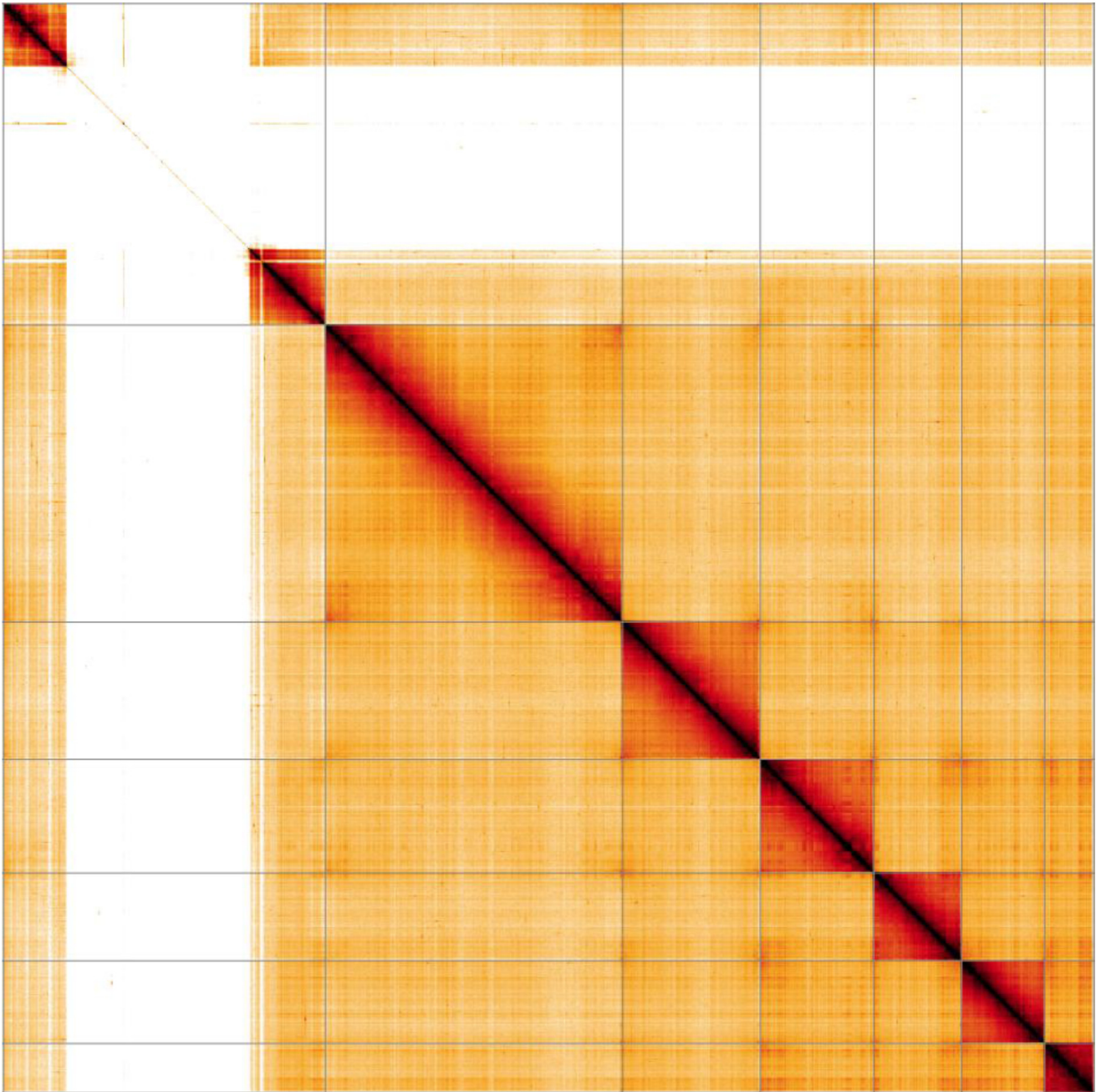


Figure 4. Genome assembly of *Rhagonycha fulva*, icRhaFulv1.1: Hi-C contact map. Hi-C contact map of the icRhaFulv1.1 assembly, visualised in HiGlass.

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

Table 2. Chromosomal pseudomolecules in the genome assembly of *Rhagoxycha fulva*, icRhaFulv1.1.

INSDC accession	Chromosome	Size (Mb)	GC%
HG996554.1	1	125.35	28.4
HG996555.1	2	115.75	30.9
HG996556.1	3	53.51	31
HG996557.1	4	44.31	31.5
HG996558.1	5	34.19	31.7
HG996559.1	6	32.27	32
HG996560.1	X	19.14	30.8
HG996561.1	MT	0.02	20.9
-	Unplaced	0.14	31.7

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	0.1.x	https://github.com/wtsi-hpag/PretextView
BlobToolKit	2.6.1	Challis et al., 2020

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: *Rhagoxycha fulva* (common red soldier beetle). Accession number PRJEB43742; <https://identifiers.org/ena.embl:PRJEB43742>.

The genome sequence is released openly for reuse. The *R. fulva* genome sequencing initiative is part of the Darwin Tree of Life (DTOL) project. All raw sequence data and the assembly have been deposited in INSDC databases. The genome will be annotated using the RNA-Seq data and presented through the Ensembl pipeline at the European Bioinformatics Institute. Raw data and assembly accession identifiers are reported in Table 1.

Acknowledgements

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 Natural History Museum Genome Acquisition Lab are listed here: <https://doi.org/10.5281/zenodo.4790043>.

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. [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. [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](#)

Grace J, Nelson M: **Insects and Their Pollen Loads at a Hybrid *Heracleum* Site.** *New Phytologist.* 1981; **87**(2): 413–23. [Publisher 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): gjaa153.

[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)

James LV, Angus RB: **A Chromosomal Investigation of Some British Cantharidae (Coleoptera).** *Genetica.* 2007; **130**(3): 293–300.

[PubMed Abstract](#) | [Publisher 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](#)

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](#)

Rodwell LE, Day JJ, Foster CW, *et al.*: **Daily Survival and Dispersal of Adult *Rhagonycha Fulva* (Coleoptera: Cantharidae) in a Wooded Agricultural Landscape.** *Eur J Entomol.* 2018; **115**: 432–36.

[Publisher Full Text](#)

Simão FA, Waterhouse RM, Ioannidis P, *et al.*: **BUSCO: Assessing Genome Assembly and Annotation Completeness with Single-Copy Orthologs.** *Bioinformatics.* 2015; **31**(19): 3210–12.

[PubMed Abstract](#) | [Publisher Full Text](#)

Uliano-Silva M, Ferreira Nunes JG, Krasheninnikova K, *et al.*: **marcelauliano/MitoHiFi: mitohifi_v2.0.** 2021.

[Publisher Full Text](#)

Open Peer Review

Current Peer Review Status:  

Version 1

Reviewer Report 09 August 2022

<https://doi.org/10.21956/wellcomeopenres.19002.r51568>

© 2022 Weng Y. 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.



Yi-Ming Weng 

University of Wisconsin Madison, Madison, WI, USA

Upon saying that the sufficient details of methods and materials were provided, I do want to learn more about the process of the assembly polishing. The author provided the aligning and variance calling tools for polishing, and I expected that the output of freebayes is a vcf file storing the variances. However, I can't directly depict the process of how the assembly was polished with the vcf file. It might be better to add this information in the manner of repeatability.

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 evolution and population genomics

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

Reviewer Report 18 May 2022

<https://doi.org/10.21956/wellcomeopenres.19002.r50422>

© 2022 Bocak L. 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.



Ladislav Bocak 

CATRIN, Biodiversity & Molecular Evolution Olomouc, Olomouc, Czech Republic

The data note requires minor revision. The information on the species biology is sufficient, but I would recommend mentioning the distribution of the species as the researchers from other countries can use data and it is valuable to indicate where the species occurs. Geographic origin is mentioned twice with all coordinates, etc. If this is not a part of the report template, modify the text.

The description of applied methods is detailed, data access information is complete.

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: phylogenetics, phylogenomics

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.
