



DATA NOTE

# The genome sequence of the Summer Chafer, *Amphimallon solstitiale* (Linnaeus, 1758) [version 1; peer review: 1 approved, 2 approved with reservations]

Douglas Boyes<sup>1+</sup>, Liam M. Crowley<sup>id</sup><sup>2</sup>, Peter W.H. Holland<sup>2</sup>,  
University of Oxford and Wytham Woods Genome Acquisition Lab,  
Darwin Tree of Life Barcoding collective,  
Wellcome Sanger Institute Tree of Life Management, Samples and Laboratory  
team,  
Wellcome Sanger Institute Scientific Operations: Sequencing Operations,  
Wellcome Sanger Institute Tree of Life Core Informatics team,  
Tree of Life Core Informatics collective, Darwin Tree of Life Consortium

<sup>1</sup>UK Centre for Ecology & Hydrology, Wallingford, England, UK

<sup>2</sup>University of Oxford, Oxford, England, UK

+ Deceased author

**V1** First published: 08 Mar 2024, 9:138  
<https://doi.org/10.12688/wellcomeopenres.21100.1>  
Latest published: 08 Mar 2024, 9:138  
<https://doi.org/10.12688/wellcomeopenres.21100.1>

## Abstract

We present a genome assembly from an individual male *Amphimallon solstitiale* (the Summer Chafer; Arthropoda; Insecta; Coleoptera; Scarabaeidae). The genome sequence is 1,584.1 megabases in span. Most of the assembly is scaffolded into 11 chromosomal pseudomolecules, including the X and Y sex chromosomes. The mitochondrial genome has also been assembled and is 19.29 kilobases in length.

## Keywords

*Amphimallon solstitiale*, Summer Chafer, genome sequence, chromosomal, Coleoptera



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

## Open Peer Review

Approval Status ✓ ? ?

	1	2	3
<b>version 1</b> 08 Mar 2024	✓ <a href="#">view</a>	? <a href="#">view</a>	? <a href="#">view</a>

- Doga Cedden** <sup>id</sup>, University of Göttingen, Göttingen, Germany  
Johann-Friedrich-Blumenbach Institute, Göttingen, Germany
- Phillip Davidson**, Indiana University  
Bloomington, Bloomington, USA
- Changlong Shu** <sup>id</sup>, Chinese Academy of Agricultural Sciences, Beijing, China

Any reports and responses or comments on the article can be found at the end of the article.

**Corresponding author:** Darwin Tree of Life Consortium ([mark.blaxter@sanger.ac.uk](mailto:mark.blaxter@sanger.ac.uk))

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

**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:** © 2024 Boyes D *et al.* This is an open access article distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

**How to cite this article:** Boyes D, Crowley LM, Holland PWH *et al.* **The genome sequence of the Summer Chafer, *Amphimallon solstitiale* (Linnaeus, 1758) [version 1; peer review: 1 approved, 2 approved with reservations]** Wellcome Open Research 2024, 9:138 <https://doi.org/10.12688/wellcomeopenres.21100.1>

**First published:** 08 Mar 2024, 9:138 <https://doi.org/10.12688/wellcomeopenres.21100.1>

## Species taxonomy

Eukaryota; Opisthokonta; Metazoa; Eumetazoa; Bilateria; Protostomia; Ecdysozoa; Panarthropoda; Arthropoda; Mandibulata; Pancrustacea; Hexapoda; Insecta; Dicondylia; Pterygota; Neoptera; Endopterygota; Coleoptera; Polyphaga; Scarabaeiformia; Scarabaeoidea; Scarabaeidae; Melolonthinae; *Amphimallon*; *Amphimallon solstitiale* Linnaeus, 1758 (NCBI:txid360071).

## Background

The tilt of the Earth's axis of rotation means that once each year the northern hemisphere experiences the day with the longest period of daylight, which may fall on either 20 or 21 June. Before the summer solstice the hours of daylight are increasing each day; after the summer solstice the hours of daylight are decreasing. Many species of insect have life-history traits attuned to day length changes, a trait known as photoperiodism (Denlinger *et al.*, 2017). In many cases, there is interplay between photoperiodism and temperature sensitivity, with each contributing to when an insect enters or emerges from diapause or changes behaviour. The cellular and molecular basis of these responses is incompletely understood although progress is being made in studies using Lepidoptera (Denlinger *et al.*, 2017; Kozak *et al.*, 2019); case studies from other insect orders are needed to further an understanding of this important biological phenomenon.

*Amphimallon solstitiale* (order Coleoptera, family Scarabaeidae) is a species of beetle with a life history that seems exquisitely linked to day length in the northern hemisphere. The larvae feed underground on roots and tubers, spending two or three years as a larva before pupating underground in the final spring (UK Beetles, no date). In midsummer, notably around the summer solstice, adults emerge to mate. Males gather at the tops of trees and tall shrubs in late evening, or even above chimney stacks on houses, swarming in numbers ranging from a few individuals to over a hundred. Not every tree is chosen with apparent preference for the tallest (PWHH, personal observation). These impressive aggregations are often seen on the evening of the summer solstice itself and may continue until mid-July. After mating, females descend to the ground to lay eggs in the soil. A female sex pheromone has been identified (Tolasch *et al.*, 2003). The behavioural link to midsummer has given rise to the common name Summer Chafer, European June beetle or the Solstice beetle, and the specific name *solstitiale*.

*A. solstitiale* is found across central and northern Europe including southern regions of England and Wales and southern Scandinavia. Further east, there are abundant records from Russia and scattered records from Georgia, Armenia, Tajikistan, Kyrgyzstan and Uzbekistan (GBIF Secretariat, 2023).

The underground larva can occasionally reach pest proportions in garden lawns and commercial horticulture (Royal Horticultural Society, 2023). Proposed control measures include the use of entomopathogenic nematodes; however, microbiota screening suggests that bacteria in the gut of *A. solstitiale* include species antagonistic to symbiotic bacteria of nematodes (Skowronek *et al.*, 2021).

Here we report a complete genome sequence for the Summer Chafer *Amphimallon solstitiale* determined as part of the Darwin Tree of Life project. The genome sequence of *A. solstitiale* will facilitate research into the molecular control of photoperiodism and contribute to the rapidly growing resources for studying genome evolution in insects.

## Genome sequence report

The genome was sequenced from one male *Amphimallon solstitiale* (Figure 1) collected from Wytham Woods, Oxfordshire, UK (51.77, -1.34). A total of 47-fold coverage in Pacific Biosciences single-molecule HiFi long reads was generated. Primary assembly contigs were scaffolded with chromosome conformation Hi-C data. Manual assembly curation corrected 122 missing joins or mis-joins and removed 25 haplotypic duplications, reducing the assembly length by 1.00% and the scaffold number by 6.10%.

The final assembly has a total length of 1,584.1 Mb in 461 sequence scaffolds with a scaffold N50 of 150.9 Mb (Table 1). The snailplot in Figure 2 provides a summary of the assembly statistics, while the distribution of assembly scaffolds on GC proportion and coverage is shown in Figure 3. The cumulative assembly plot in Figure 4 shows curves for subsets of scaffolds assigned to different phyla. Most (97.69%) of the assembly sequence was assigned to 11 chromosomal-level scaffolds, representing 9 autosomes and the X and Y sex chromosomes. Chromosome-scale scaffolds confirmed by the Hi-C data are named in order of size (Figure 5; Table 2). Chromosomes X and Y were assigned by read coverage statistics and synteny to the genome assembly of *Melolontha melolontha* (GCA\_935421215.2) (Ashworth *et al.*, 2023). While not fully phased, the assembly deposited is of one haplotype. Contigs corresponding to the second haplotype have also been deposited. The mitochondrial genome was also assembled and can be found as a contig within the multifasta file of the genome submission.



**Figure 1.** Photograph of the *Amphimallon solstitiale* (icAmpSols1) specimen used for genome sequencing.

**Table 1. Genome data for *Amphimallon solstitiale*, icAmpSols1.1.**

Project accession data		
Assembly identifier	icAmpSols1.1	
Species	<i>Amphimallon solstitiale</i>	
Specimen	icAmpSols1	
NCBI taxonomy ID	360071	
BioProject	PRJEB61997	
BioSample ID	SAMEA10978893	
Isolate information	icAmpSols1: abdomen (DNA and Hi-C sequencing)	
Assembly metrics*		Benchmark
Consensus quality (QV)	63.1	≥ 50
k-mer completeness	100.0%	≥ 95%
BUSCO**	C:99.2%[S:96.9%,D:2.3%],F:0.5%,M:0.3%,n:2,124	C ≥ 95%
Percentage of assembly mapped to chromosomes	97.69%	≥ 95%
Sex chromosomes	XY	localised homologous pairs
Organelles	Mitochondrial genome: 19.29 kb	complete single alleles
Raw data accessions		
PacificBiosciences SEQUEL II	ERR11458786	
Hi-C Illumina	ERR11439603	
Genome assembly		
Assembly accession	GCA_963170755.1	
Accession of alternate haplotype	GCA_963170735.1	
Span (Mb)	1,584.1	
Number of contigs	1,218	
Contig N50 length (Mb)	3.7	
Number of scaffolds	461	
Scaffold N50 length (Mb)	150.9	
Longest scaffold (Mb)	260.58	

\* Assembly metric benchmarks are adapted from column VGP-2020 of "Table 1: Proposed standards and metrics for defining genome assembly quality" from [Rhie et al. \(2021\)](#).

\*\* BUSCO scores based on the endopterygota\_odb10 BUSCO set using version 5.3.2. 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/CAUJHA01/dataset/CAUJHA01/busco>.

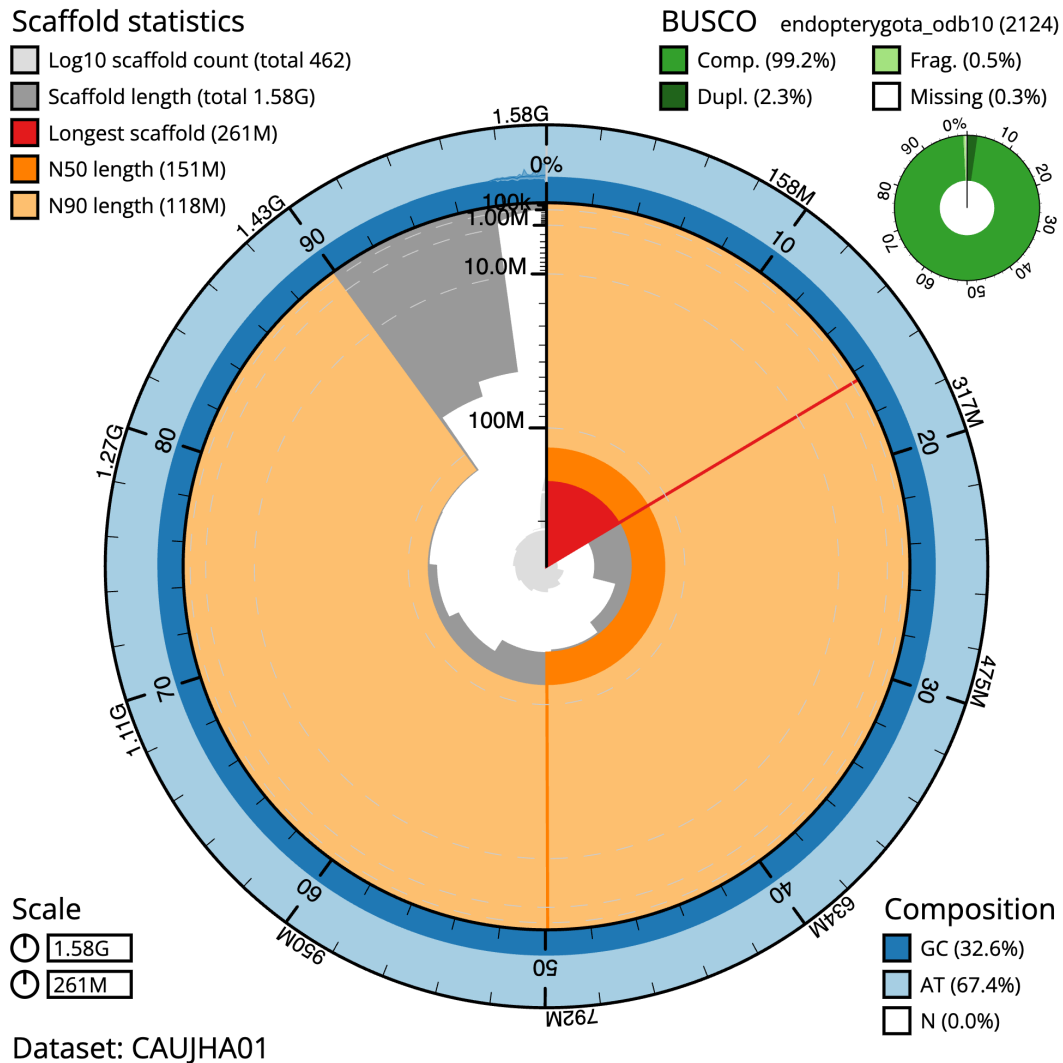
The estimated Quality Value (QV) of the final assembly is 63.1 with k-mer completeness of 100.0%, and the assembly has a BUSCO v5.3.2 completeness of 99.2% (single = 96.9%, duplicated = 2.3%), using the endopterygota\_odb10 reference set (n = 2,124).

Metadata for specimens, barcode results, spectra estimates, sequencing runs, contaminants and pre-curation assembly statistics are given at <https://links.tol.sanger.ac.uk/species/360071>.

## Methods

### Sample acquisition and nucleic acid extraction

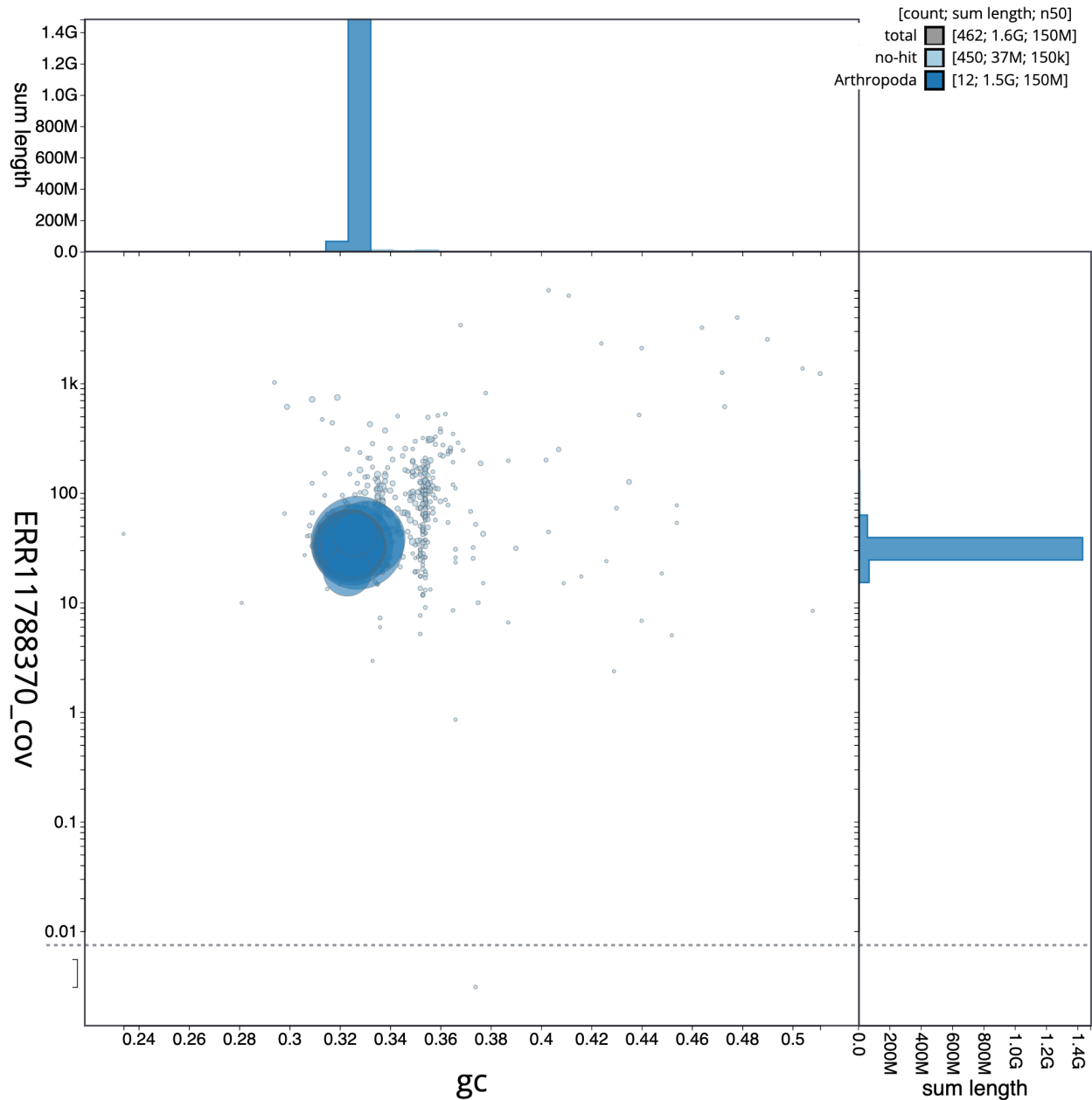
A male *Amphimallon solstitiale* (specimen ID Ox001707, ToLID icAmpSols1) was collected from Wytham Woods, Oxfordshire (biological vice-county Berkshire), UK (latitude 51.77, longitude -1.34) on 2021-07-17 in a light trap. The specimen was collected by Douglas Boyes (University of Oxford) and identified by Liam Crowley (University of Oxford), and then snap-frozen on dry ice.



**Figure 2. Genome assembly of *Amphimallon solstitialis*, icAmpSols1.1: 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 1,584,161,395 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 (260,577,686 bp, shown in red). Orange and pale-orange arcs show the N50 and N90 scaffold lengths (150,921,111 and 118,082,474 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/CAUJHA01/dataset/CAUJHA01/snail>.

The workflow for high molecular weight (HMW) DNA extraction at the Wellcome Sanger Institute (WSI) includes a sequence of core procedures: sample preparation; sample homogenisation, DNA extraction, fragmentation, and clean-up. In sample preparation, the icAmpSols1 sample was weighed and dissected on dry ice (Jay *et al.*, 2023). Tissue from the abdomen was homogenised using a PowerMasher II tissue disruptor (Denton *et al.*, 2023a). HMW DNA was extracted using the

Automated MagAttract v1 protocol (Sheerin *et al.*, 2023). DNA was sheared into an average fragment size of 12–20 kb in a Megaruptor 3 system (Todorovic *et al.*, 2023). Sheared DNA was purified by solid-phase reversible immobilisation (Strickland *et al.*, 2023): in brief, the method employs a 1.8X ratio of AMPure PB beads to sample to eliminate shorter fragments and concentrate the DNA. The concentration of the sheared and purified DNA was assessed using a Nanodrop



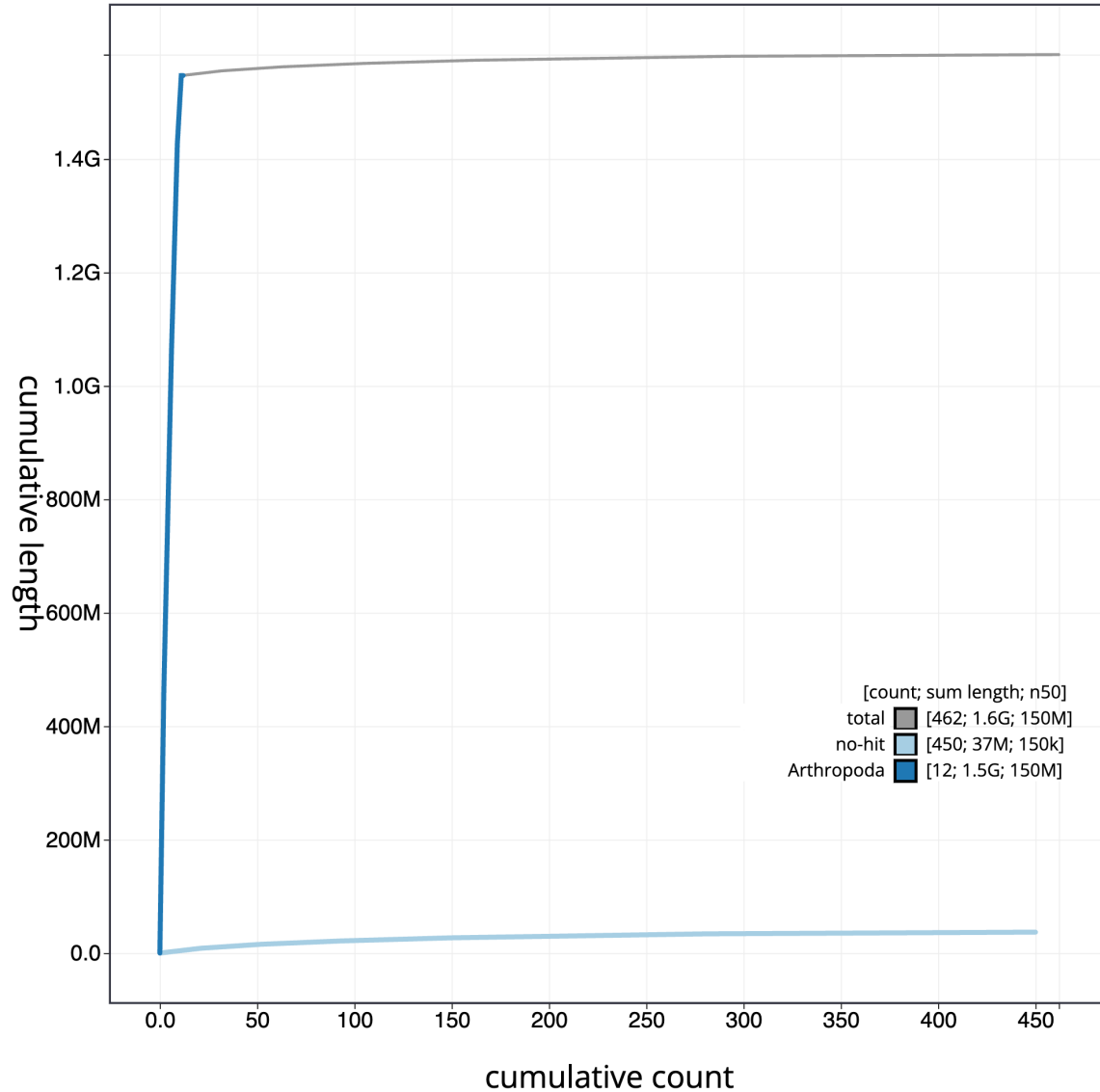
**Figure 3. Genome assembly of *Amphimallon solstitiale*, icAmpSols1.1: BlobToolKit GC-coverage plot.** Sequences are coloured by phylum. Circles are sized in proportion to sequence length. Histograms show the distribution of sequence length sum along each axis. An interactive version of this figure is available at <https://blobtoolkit.genomehubs.org/view/CAUJHA01/dataset/CAUJHA01/blob>.

spectrophotometer and Qubit Fluorometer and Qubit dsDNA High Sensitivity Assay kit. Fragment size distribution was evaluated by running the sample on the FemtoPulse system.

Protocols developed by the WSI Tree of Life laboratory are publicly available on protocols.io (Denton *et al.*, 2023b).

### Sequencing

Pacific Biosciences HiFi circular consensus DNA sequencing libraries were constructed according to the manufacturers' instructions. DNA sequencing was performed by the Scientific Operations core at the WSI on a Pacific Biosciences SEQUEL II instrument. Hi-C data were also generated from



**Figure 4. Genome assembly of *Amphimallon solstitiale*, icAmpSols1.1: BlobToolKit cumulative sequence plot.** The grey line shows cumulative length for all sequences. Coloured lines show cumulative lengths of sequences assigned to each phylum using the buscogenes taxrule. An interactive version of this figure is available at <https://blobtoolkit.genomehubs.org/view/CAUJHA01/dataset/CAUJHA01/cumulative>.

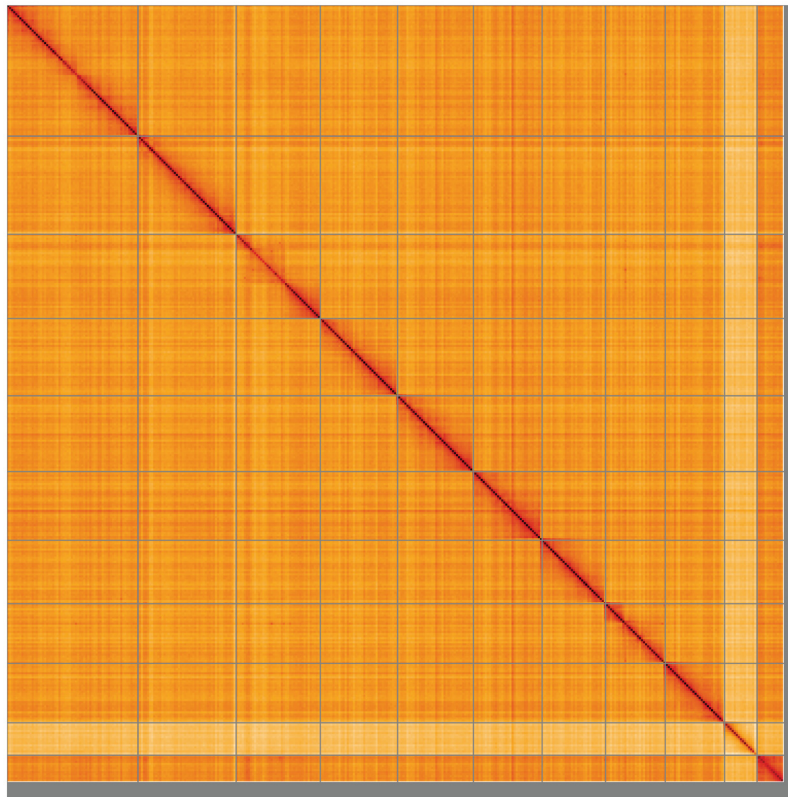
remaining abdomen tissue of icAmpSols1 using the Arima2 kit and sequenced on the Illumina NovaSeq 6000 instrument.

#### Genome assembly, curation and evaluation

Assembly was carried out with Hifiasm (Cheng *et al.*, 2021) and haplotypic duplication was identified and removed with purge\_dups (Guan *et al.*, 2020). The assembly was then scaffolded with Hi-C data (Rao *et al.*, 2014) using YaHS (Zhou *et al.*, 2023). The assembly was checked for contamination and corrected as described previously (Howe *et al.*, 2021).

Manual curation was performed using HiGlass (Kerpedjiev *et al.*, 2018) and PretextView (Harry, 2022). The mitochondrial genome was assembled using MitoHiFi (Uliano-Silva *et al.*, 2023), which runs MitoFinder (Allio *et al.*, 2020) or MITOS (Bernt *et al.*, 2013) and uses these annotations to select the final mitochondrial contig and to ensure the general quality of the sequence.

A Hi-C map of the final assembly was produced using bwa-mem2 (Vasimuddin *et al.*, 2019) in the Cooler file



**Figure 5. Genome assembly of *Amphimallon solstitiale*, icAmpSols1.1: Hi-C contact map of the icAmpSols1.1 assembly, visualised using HiGlass.** Chromosomes are shown in order of size from left to right and top to bottom. An interactive version of this figure may be viewed at <https://genome-note-higlass.tol.sanger.ac.uk/l/?d=bmH40SVZR06IKUSWBLb7PA>.

**Table 2. Chromosomal pseudomolecules in the genome assembly of *Amphimallon solstitiale*, icAmpSols1.**

INSDC accession	Chromosome	Length (Mb)	GC%
OY720675.1	1	260.58	32.5
OY720676.1	2	195.32	32.5
OY720677.1	3	167.66	33.0
OY720678.1	4	153.71	32.5
OY720679.1	5	150.92	32.5
OY720680.1	6	136.69	32.5
OY720681.1	7	126.26	32.5
OY720682.1	8	118.77	32.5
OY720683.1	9	118.08	32.5
OY720684.1	X	64.67	32.5
OY720685.1	Y	54.41	32.5
OY720686.1	MT	0.02	29.5

format (Abdennur & Mirny, 2020). To assess the assembly metrics, the  $k$ -mer completeness and QV consensus quality values were calculated in Merqury (Rhie *et al.*, 2020). This work was done using Nextflow (Di Tommaso *et al.*, 2017) DSL2 pipelines “sanger-tol/readmapping” (Surana *et al.*, 2023a) and “sanger-tol/genomenote” (Surana *et al.*, 2023b). The genome was analysed within the BlobToolKit environment (Challis *et al.*, 2020) and BUSCO scores (Manni *et al.*, 2021; Simão *et al.*, 2015) were calculated.

Table 3 contains a list of relevant software tool versions and sources.

#### Wellcome Sanger Institute – Legal and Governance

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**’, which can be found in full on the Darwin Tree of Life website [here](#). 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.



**Table 3. Software tools: versions and sources.**

Software tool	Version	Source
BlobToolKit	4.2.1	<a href="https://github.com/blobtoolkit/blobtoolkit">https://github.com/blobtoolkit/blobtoolkit</a>
BUSCO	5.3.2	<a href="https://gitlab.com/eziab/busco">https://gitlab.com/eziab/busco</a>
Hifiasm	0.16.1-r375	<a href="https://github.com/chhylp123/hifiasm">https://github.com/chhylp123/hifiasm</a>
HiGlass	1.11.6	<a href="https://github.com/higlass/higlass">https://github.com/higlass/higlass</a>
Merqury	MerquryFK	<a href="https://github.com/thegenemyers/MERQURY.FK">https://github.com/thegenemyers/MERQURY.FK</a>
MitoHiFi	3	<a href="https://github.com/marcelauliano/MitoHiFi">https://github.com/marcelauliano/MitoHiFi</a>
PretextView	0.2	<a href="https://github.com/wtsi-hpag/PretextView">https://github.com/wtsi-hpag/PretextView</a>
purge_dups	1.2.5	<a href="https://github.com/dfguan/purge_dups">https://github.com/dfguan/purge_dups</a>
sanger-tol/genomenote	v1.0	<a href="https://github.com/sanger-tol/genomenote">https://github.com/sanger-tol/genomenote</a>
sanger-tol/readmapping	1.1.0	<a href="https://github.com/sanger-tol/readmapping/tree/1.1.0">https://github.com/sanger-tol/readmapping/tree/1.1.0</a>
YaHS	1.2a.2	<a href="https://github.com/c-zhou/yahs">https://github.com/c-zhou/yahs</a>

Further, the Wellcome Sanger Institute employs a process whereby due diligence is carried out proportionate to the nature of the materials themselves, and the circumstances under which they have been/are to be collected and provided for use. The purpose of this is to address and mitigate any potential legal and/or ethical implications of receipt and use of the materials as part of the research project, and to ensure that in doing so we align with best practice wherever possible. The overarching areas of consideration are:

- Ethical review of provenance and sourcing of the material
- Legality of collection, transfer and use (national and international)

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: *Amphimallon solstitiale* (summer chafer). Accession number PRJEB61997; <https://identifiers.org/ena.embl/PRJEB61997> (Wellcome Sanger Institute, 2023). The genome sequence is released openly for reuse. The *Amphimallon solstitiale* 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 available 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](#).

### Author information

Members of the University of Oxford and Wytham Woods Genome Acquisition Lab are listed here: <https://doi.org/10.5281/zenodo.7125292>.

Members of the Darwin Tree of Life Barcoding collective are listed here: <https://doi.org/10.5281/zenodo.4893703>.

Members of the Wellcome Sanger Institute Tree of Life Management, Samples and Laboratory team are listed here: <https://doi.org/10.5281/zenodo.10066175>.

Members of Wellcome Sanger Institute Scientific Operations: Sequencing Operations are listed here: <https://doi.org/10.5281/zenodo.10043364>.

Members of the Wellcome Sanger Institute Tree of Life Core Informatics team are listed here: <https://doi.org/10.5281/zenodo.10066637>.

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

- Abdennur N, Mirny LA: **Cooler: Scalable storage for Hi-C data and other genomically labeled arrays.** *Bioinformatics.* 2020; **36**(1): 311–316.  
[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](#)
- Ashworth M, Darwin Tree of Life Barcoding collective, Wellcome Sanger Institute Tree of Life programme, et al.: **The genome sequence of a cockchafer, *Melolontha melolontha* (Linnaeus, 1758) [version 1; peer review: awaiting peer review].** *Wellcome Open Res.* 2023; **8**: 222.  
[Publisher Full Text](#)
- Bernt M, Donath A, Jühling F, et al.: **MITOS: Improved *de novo* metazoan mitochondrial genome annotation.** *Mol Phylogenet Evol.* 2013; **69**(2): 313–319.  
[PubMed Abstract](#) | [Publisher 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](#)
- Denlinger DL, Hahn DA, Merlin C, et al.: **Keeping time without a spine: what can the insect clock teach us about seasonal adaptation?** *Philos Trans R Soc Lond B Biol Sci.* 2017; **372**(1734): 20160257.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Denton A, Oatley G, Cornwell C, et al.: **Sanger Tree of Life Sample Homogenisation: PowerMash.** *protocols.io.* 2023a.  
[Publisher Full Text](#)
- Denton A, Yatsenko H, Jay J, et al.: **Sanger Tree of Life Wet Laboratory Protocol Collection V.1.** *protocols.io.* 2023b.  
[Publisher Full Text](#)
- Di Tommaso P, Chatzou M, Floden EW, et al.: **Nextflow enables reproducible computational workflows.** *Nat Biotechnol.* 2017; **35**(4): 316–319.  
[PubMed Abstract](#) | [Publisher Full Text](#)
- GBIF Secretariat: ***Amphimallon solstitiale* (Linnaeus, 1758).** *GBIF Backbone Taxonomy.* 2023; [Accessed 29 January 2024].  
[Reference Source](#)
- 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): g1aa153.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Jay J, Yatsenko H, Narváez-Gómez JP, et al.: **Sanger Tree of Life Sample Preparation: Triage and Dissection.** *protocols.io.* 2023.  
[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](#)
- Kozak GM, Wadsworth CB, Kahne SC, et al.: **Genomic Basis of Circannual Rhythm in the European Corn Borer Moth.** *Curr Biol.* 2019; **29**(20): 3501–3509.e5.  
[PubMed Abstract](#) | [Publisher Full Text](#)
- Manni M, Berkeley MR, Seppely 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 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): 1665–1680.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Rhie A, McCarthy SA, Fedrigo O, et al.: **Towards complete and error-free genome assemblies of all vertebrate species.** *Nature.* 2021; **592**(7856): 737–746.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Rhie A, Walenz BP, Koren S, et al.: **Mercury: Reference-free quality, completeness, and phasing assessment for genome assemblies.** *Genome Biol.* 2020; **21**(1): 245.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Royal Horticultural Society: **Chafer grubs in lawns.** 2023; [Accessed 29 January 2024].  
[Reference Source](#)
- Sheerin E, Sampaio F, Oatley G, et al.: **Sanger Tree of Life HMW DNA Extraction: Automated MagAttract v.1.** *protocols.io.* 2023.  
[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): 3210–3212.  
[PubMed Abstract](#) | [Publisher Full Text](#)
- Skowronek M, Sajnaga E, Kazimierczak W, et al.: **Screening and Molecular Identification of Bacteria from the Midgut of *Amphimallon solstitiale* Larvae Exhibiting Antagonistic Activity against Bacterial Symbionts of Entomopathogenic Nematodes.** *Int J Mol Sci.* 2021; **22**(21): 12005.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Strickland M, Cornwell C, Howard C: **Sanger Tree of Life Fragmented DNA clean up: Manual SPRI.** *protocols.io.* 2023.  
[Publisher Full Text](#)
- Surana P, Muffato M, Qi G: **sanger-tol/readmapping: sanger-tol/readmapping v1.1.0 - Hebridean Black (1.1.0).** Zenodo. 2023a.  
[Publisher Full Text](#)
- Surana P, Muffato M, Sadasivan Baby C: **sanger-tol/genomenote (v1.0.dev).** Zenodo. 2023b.  
[Publisher Full Text](#)
- Todorovic M, Sampaio F, Howard C: **Sanger Tree of Life HMW DNA Fragmentation: Diagenode Megaruptor<sup>3</sup> for PacBio HiFi.** *protocols.io.* 2023.  
[Publisher Full Text](#)
- Tolasch T, Sölter S, Tóth M, et al.: **(R)-Acetoin-Female Sex Pheromone of the Summer Chafer *Amphimallon solstitiale* (L.).** *J Chem Ecol.* 2003; **29**(4): 1045–1050.  
[PubMed Abstract](#) | [Publisher Full Text](#)
- UK Beetles: ***Amphimallon Latreille, 1825.*** [no date]; [Accessed 29 January 2024].  
[Reference Source](#)
- Uliano-Silva M, Ferreira JGRN, Krashenninnikova K, et al.: **MitoHiFi: a python pipeline for mitochondrial genome assembly from PacBio high fidelity reads.** *BMC Bioinformatics.* 2023; **24**(1): 288.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Vasimuddin M, Misra S, Li H, et al.: **Efficient Architecture-Aware Acceleration of BWA-MEM for Multicore Systems.** In: *2019 IEEE International Parallel and Distributed Processing Symposium (IPDPS).* IEEE, 2019; 314–324.  
[Publisher Full Text](#)
- Wellcome Sanger Institute: **The genome sequence of the Summer Chafer, *Amphimallon solstitiale* (Linnaeus, 1758).** European Nucleotide Archive, [dataset], accession number PRJEB61997, 2023.
- Zhou C, McCarthy SA, Durbin R: **YaHS: yet another Hi-C scaffolding tool.** *Bioinformatics.* Edited by C. Alkan, 2023; **39**(1): btac808.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)

# Open Peer Review

Current Peer Review Status:   

---

## Version 1

Reviewer Report 22 May 2024

<https://doi.org/10.21956/wellcomeopenres.23340.r79865>

© 2024 Shu C. 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.



**Changlong Shu** 

Chinese Academy of Agricultural Sciences, Beijing, China

The whole experimental design is very reasonable and the data obtained are also very good, which provides sufficient genetic data for the further study of this insect.

However, the data provided in Figure 2 is too complex. It is recommended to draw statistical charts according to the data category, so that other peers, especially those who are not specialized in bioinformatics research, can quickly and clearly read the information of interest.

**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:** Resource insects and microbiome

**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 22 May 2024

<https://doi.org/10.21956/wellcomeopenres.23340.r83352>

© 2024 Davidson P. 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.



## Phillip Davidson

Department of Biology, Indiana University Bloomington, Bloomington, Indiana, USA

In their genome report, Boyes, Crowley, and Holland present a complete and high quality reference assembly for the Summer Chafer *Amphimallon solstitiale*. Overall, the assembly and report are of superb quality and represent a much needed addition for Melolonthinae genomic resources. The rationale as a model for photoperiodism is well described and the methods are sound. Further, the data are accessible online. I have only a handful of minor revision suggestions that I think will clarify the results and significance of this data set, as listed below.

### Genome Sequence Report:

1. I see the PacBio coverage is stated at 47X coverage, however, please also state the coverage for the HiC data.
2. Figure 1: Please add a scale bar or if not possible, consider substituting a separate image that includes one.
3. The genome of *Amphimallon solstitiale* is remarkably large for a scarab (nearly 1.6Gb), which is very interesting. I think a nice addition to the results would describe what genomic elements compose this enormous expansion (repetitive elements I may guess) and their biological implication. I do not think a formal repetitive element analysis (annotation, classification) is necessary for a data report, but an estimation of what % of the genome is composed of repeats would improve the description of the results.
4. Figure 2: The BUSCO scores mini-circle plot states the middle shade of green is 99.2% and represents the completion score. However, I think this label should be changed to "single copy (96.9%)" score as the middle shade green does not appear to compose 99.2% of the plot. Further, listing the SC, Duplication, Frag, and Missing is more standard as it distinguishes single copy and duplication composition. Then, the complete score of 99.2% can be listed separately or perhaps in the legend.

### Methods:

1. For the clause "snap-frozen on dry ice". Should this instead read "snap-frozen in liquid nitrogen"? Typically dry ice is used for storage and transportation, so I wanted to check.

### Data Availability:

1. When the gene annotations are complete and if possible, it may be nice to update this report with a link to those files.

**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:** Genome assembly and annotation, evolutionary-developmental biology

**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 21 May 2024

<https://doi.org/10.21956/wellcomeopenres.23340.r83369>

© 2024 Cedden D. 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.



**Doga Cedden** 

<sup>1</sup> Agricultural Entomology, Department of Crop Sciences, University of Göttingen, Göttingen, Germany

<sup>2</sup> Department of Evolutionary Developmental Genetics, Johann-Friedrich-Blumenbach Institute, Göttingen, Germany

The data note by Boyes et al. presents a high-quality genome assembly from a male *Amphimallon solstitiale*, commonly known as the summer chafer. Although the beetle is not of high economic importance, the high-quality genome will support studies on its interesting biology.

The quality assessment of the genome sequence is sufficiently detailed. The methods for preparing the specimen, sequencing, and quality control are appropriate and the authors provide enough details for the replication of the study. The genome is made accessible through an appropriate archive service.

The Background section provides a nice context for the presented genome sequence. One minor suggestion from my side would be to combine the third and fourth paragraphs of the Background section into one paragraph.

**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:*** I have expertise in analyzing genomic data from beetle species. Hence, I was able to assess the majority of the manuscript.

**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.**

---