



DATA NOTE

The genome sequence of the Brick, *Agrochola circellaris* (Hufnagel, 1766)

[version 1; peer review: 2 approved]

Douglas Boyes ¹⁺,

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, Asia Hoile²,
Darwin Tree of Life Consortium

¹UK Centre for Ecology and Hydrology, Wallingford, Oxfordshire, UK²Department of Biology, University of Oxford, Oxford, UK

+ Deceased author

V1 First published: 31 Jan 2023, 8:44
<https://doi.org/10.12688/wellcomeopenres.18894.1>

Latest published: 31 Jan 2023, 8:44
<https://doi.org/10.12688/wellcomeopenres.18894.1>

Abstract

We present a genome assembly from an individual male *Agrochola circellaris* (the Brick; Arthropoda; Insecta; Lepidoptera; Noctuidae). The genome sequence is 572 megabases in span. Most of the assembly is scaffolded into 30 chromosomal pseudomolecules, including the Z sex chromosome. The mitochondrial genome has also been assembled and is 15.5 kilobases in length. Gene annotation of this assembly on Ensembl has identified 18,319 protein coding genes.

Keywords

Agrochola circellaris, the Brick, genome sequence, chromosomal, Lepidoptera



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

Open Peer Review

Approval Status

	1	2
version 1 31 Jan 2023	 view	 view

1. **Sivasankaran Kuppusamy** , Loyola College, Chennai, India

2. **Vahap Eldem** , Istanbul University, Istanbul, Turkey

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; **Hoile A:** Writing – Original Draft Preparation;

Competing interests: No competing interests were disclosed.

Grant information: This work was supported by Wellcome through core funding to the Wellcome Sanger Institute (206194, <https://doi.org/10.35802/206194>) and the Darwin Tree of Life Discretionary Award (218328, <https://doi.org/10.35802/218328>). *The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.*

Copyright: © 2023 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, University of Oxford and Wytham Woods Genome Acquisition Lab, Darwin Tree of Life Barcoding collective *et al.* **The genome sequence of the Brick, *Agrochola circellaris* (Hufnagel, 1766) [version 1; peer review: 2 approved]** Wellcome Open Research 2023, 8:44 <https://doi.org/10.12688/wellcomeopenres.18894.1>

First published: 31 Jan 2023, 8:44 <https://doi.org/10.12688/wellcomeopenres.18894.1>

Species taxonomy

Eukaryota; Metazoa; Ecdysozoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Lepidoptera; Glossata; Ditrysia; Noctuoidea; Noctuidae; Xyleninae; *Agrochola*; *Agrochola circellaris* (Hufnagel, 1766)(NCBI:txid987866).

Background

Agrochola circellaris (the Brick), is a species of moth belonging to the large and taxonomically controversial family Noctuidae within the superfamily Noctuoidea (Regier *et al.*, 2017). The phylogeny of the Noctuoidea has been difficult to resolve, although considerable progress has been made using a sample of protein-coding genes (Yang *et al.*, 2019).

A. circellaris is a relatively common moth in the UK, and is distributed throughout most of Europe, Asia Minor and Armenia. In the UK, the moth can be found in woodland, parkland and gardens, with the nocturnal adult on the wing between late August and early November (Usher & Keiller, 1998). It can be identified by a distinctive dark spot inside the reniform stigma or ‘kidney mark’, with a diffuse band linking it to the leading edge of the wing. The forewings vary in colour between individuals, ranging from yellow to reddish brown (Chinery, 1991); the genetic basis of the colour variation is unknown. The common name, ‘the Brick’, refers to the terracotta colour observed in many individuals.

The eggs overwinter and the larvae feed on flower heads and developing seeds of deciduous trees such as ash, poplar, sallow and wych elm. The larvae develop rapidly on this nutritious diet and, unusually, once fully developed they then spend several weeks inside a cocoon before pupation (Butterfly Conservation, 2006). The species has been used in studies to assess the relation between the number of seeds damaged by *A. circellaris* caterpillars and stand density of field elm trees in Poland (Bak-Badowska *et al.*, 2017).

The genome of *A. circellaris* was sequenced as part of the Darwin Tree of Life Project, a collaborative effort to sequence all named eukaryotic species in the Atlantic Archipelago of Britain and Ireland. Here we present a chromosomally complete genome sequence for *A. circellaris*, based on one male specimen from Wytham Woods, Oxfordshire, UK.

Genome sequence report

The genome was sequenced from one male *Agrochola circellaris* (Figure 1) collected from Wytham Woods (latitude 51.77, longitude -1.34). A total of 33-fold coverage in Pacific Biosciences single-molecule HiFi long reads and 73-fold coverage in 10X Genomics read clouds were generated. Primary assembly contigs were scaffolded with chromosome conformation Hi-C data. Manual assembly curation corrected four missing or mis-joins, reducing the scaffold number by 5.56%.

The final assembly has a total length of 572.1 Mb in 34 sequence scaffolds with a scaffold N50 of 19.7 Mb (Table 1). Most (99.99%) of the assembly sequence was assigned to



Figure 1. Photograph of the *Agrochola circellaris* (ilAgrCirc1) specimen used for genome sequencing.

30 chromosomal-level scaffolds, representing 29 autosomes and the Z sex chromosome. Chromosome-scale scaffolds confirmed by the Hi-C data have been named in order of size (Figure 2–Figure 5; Table 2). The assembly has a BUSCO v5.3.2 (Manni *et al.*, 2021) completeness of 99.2% (single 98.5%, duplicated 0.6%) using the lepidoptera_odb10 reference set ($n = 5,286$). While not fully phased, the assembly deposited is of one haplotype. Contigs corresponding to the second haplotype have also been deposited.

Genome annotation report

The *A. circellaris* GCA_914767755.1 genome was annotated using the Ensembl rapid annotation pipeline (Table 1; https://rapid.ensembl.org/Agrochola_circellaris_GCA_914767755.1/). The resulting annotation includes 18,499 transcribed mRNAs from 18,319 protein-coding genes.

Methods

Sample acquisition and nucleic acid extraction

Two *A. circellaris* specimens (ilAgrCirc1 and ilAgrCirc2) were collected in Wytham Woods, Oxfordshire (biological vice-county: Berkshire), UK (latitude 51.77, longitude -1.34) on 8 October 2020, using a light trap. The specimens were collected and identified by Douglas Boyes (University of Oxford) and snap-frozen on dry ice.

DNA was extracted at the Tree of Life laboratory, Wellcome Sanger Institute (WSI). The ilAgrCirc1 sample was weighed and dissected on dry ice with head tissue set aside for Hi-C sequencing. Thorax tissue was disrupted using a Nippi Powermasher fitted with a BioMasher pestle. High molecular weight (HMW) DNA was extracted using the Qiagen MagAttract HMW DNA extraction kit. Low molecular weight DNA was removed from a 20 ng aliquot of extracted DNA using 0.8X AMPure XP purification kit prior to 10X Chromium sequencing; a minimum of 50 ng DNA was submitted for 10X sequencing. HMW DNA was sheared into an average fragment size of 12–20 kb in a Megaruptor 3 system with speed setting 30. Sheared DNA was purified by solid-phase reversible immobilisation using AMPure PB beads with a 1.8X ratio of beads to sample to remove the

Table 1. Genome data for *Agrochola circellaris*, ilAgrCirc1.1.

Project accession data		
Assembly identifier	ilAgrCirc1.1	
Species	<i>Agrochola circellaris</i>	
Specimen	ilAgrCirc1	
NCBI taxonomy ID	987866	
BioProject	PRJEB46316	
BioSample ID	SAMEA8603201	
Isolate information	male; ilAgrCirc1: thorax (PacBio and 10X sequencing); head (Hi-C unknown sex; ilAgrCirc2: abdomen (RNA-Seq)	
Assembly metrics*		Benchmark
Consensus quality (QV)	59.5	≥ 50
<i>k</i> -mer completeness	100%	≥ 95%
BUSCO**	C:99.2%[S:98.5%,D:0.6%], F:0.2%,M:0.7%,n:5286	C ≥ 95%
Percentage of assembly mapped to chromosomes	99.99%	≥ 95%
Sex chromosomes	ZZ	<i>localised homologous pairs</i>
Organelles	Mitochondrial genome assembled	<i>complete single alleles</i>
Raw data accessions		
PacificBiosciences SEQUEL II	ERR6808001	
10X Genomics Illumina	ERR6688511-ERR6688514	
Hi-C Illumina	ERR6688510	
PolyA RNA-Seq Illumina	ERR9435003	
Genome assembly		
Assembly accession	GCA_914767755.1	
<i>Accession of alternate haplotype</i>	GCA_914767685.1	
Span (Mb)	572.1	
Number of contigs	39	
Contig N50 length (Mb)	19.7	
Number of scaffolds	34	
Scaffold N50 length (Mb)	19.7	
Longest scaffold (Mb)	33.0	
Genome annotation		
Number of protein-coding genes	18,319	
Number of gene transcripts	18,499	

* 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 lepidoptera_odb10 BUSCO set using v 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/ilAgrCirc1.1/dataset/CAJZBG01.1/busco>.

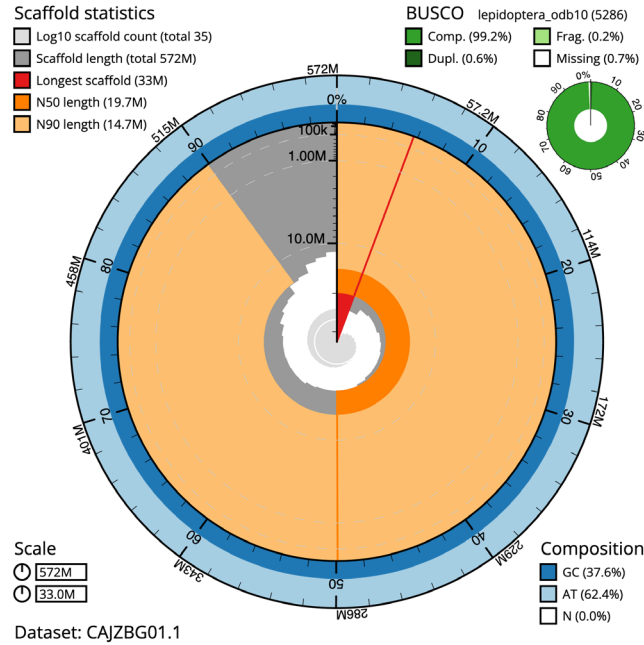


Figure 2. Genome assembly of *Agrochola circellaris*, ilAgrCirc1.1: metrics. The BlobToolKit Snailplot shows N50 metrics and BUSCO gene completeness. The main plot is divided into 1,000 size-ordered bins around the circumference with each bin representing 0.1% of the 572,162,832 bp assembly. The distribution of sequence lengths is shown in dark grey with the plot radius scaled to the longest sequence present in the assembly (32,976,333 bp, shown in red). Orange and pale-orange arcs show the N50 and N90 sequence lengths (19,711,079 and 14,656,473 bp), respectively. The pale grey spiral shows the cumulative sequence 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 lepidoptera_odb10 set is shown in the top right. An interactive version of this figure is available at <https://blobtoolkit.genomehubs.org/view/ilAgrCirc1.1/dataset/CAJZBG01.1/snail>.

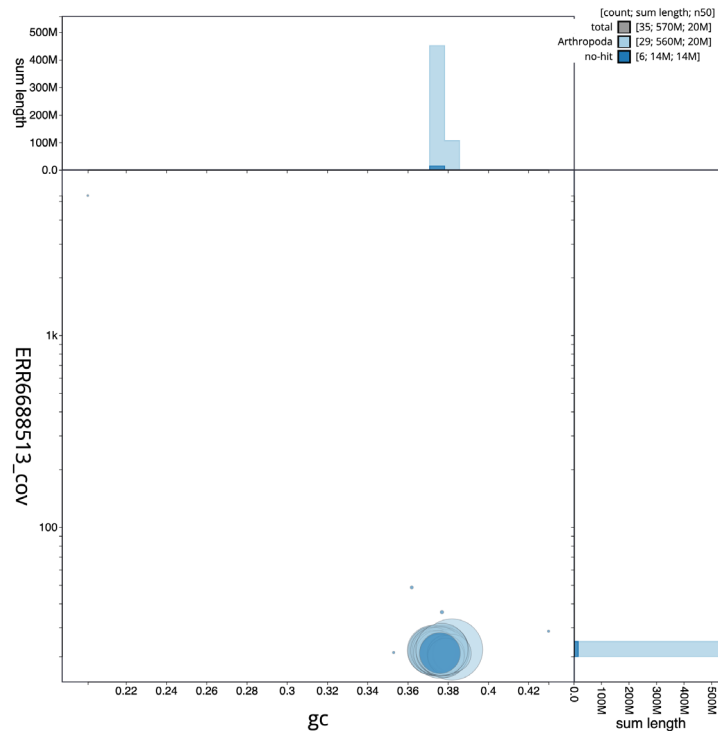


Figure 3. Genome assembly of *Agrochola circellaris*, ilAgrCirc1.1: GC coverage. BlobToolKit GC-coverage plot. Scaffolds are coloured by phylum. Circles are sized in proportion to scaffold length. Histograms show the distribution of scaffold length sum along each axis. An interactive version of this figure is available at <https://blobtoolkit.genomehubs.org/view/ilAgrCirc1.1/dataset/CAJZBG01.1/blob>.

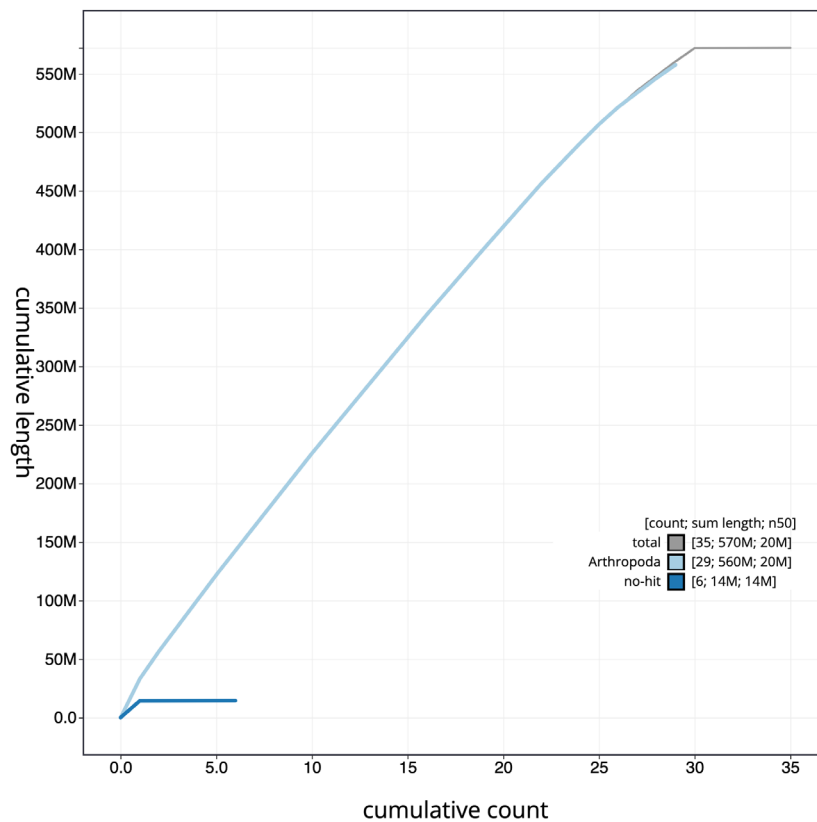


Figure 4. Genome assembly of *Agrochola circellaris*, ilAgrCirc1.1: cumulative sequence. BlobToolKit cumulative sequence plot. The grey line shows cumulative length for all scaffolds. Coloured lines show cumulative lengths of scaffolds assigned to each phylum using the buscogenes taxrule. An interactive version of this figure is available at <https://blobtoolkit.genomehubs.org/view/ilAgrCirc1.1/dataset/CAJZBG01.1/cumulative>.

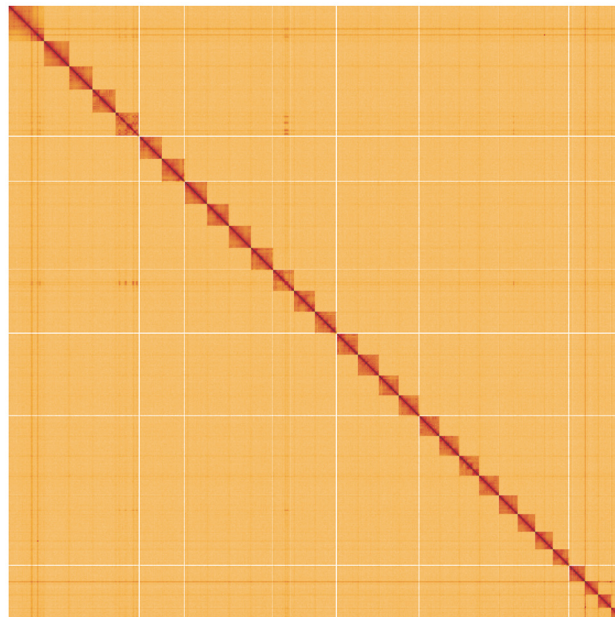


Figure 5. Genome assembly of *Agrochola circellaris*, ilAgrCirc1.1: Hi-C contact map. Hi-C contact map of the ilAgrCirc1.1 assembly, visualised using HiGlass. Chromosomes are shown in order of size from left to right and top to bottom. An interactive version of this figure may be viewed at https://genome-note-higlass.tol.sanger.ac.uk/#!/?d=ZSISneTzRIK_n3QbgCsDxg.

shorter fragments and concentrate the DNA sample. The concentration of the sheared and purified DNA was assessed using a Nanodrop spectrophotometer and Qubit Fluorometer and Qubit dsDNA High Sensitivity Assay kit. Fragment size distribution was evaluated by running the sample on the FemtoPulse system.

RNA was extracted from abdomen tissue of iAgrCirc2 in the Tree of Life Laboratory at the WSI using TRIzol, according

to the manufacturer's instructions. RNA was then eluted in 50 µl RNase-free water and its concentration 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.

Sequencing

Pacific Biosciences HiFi circular consensus and 10X Genomics read cloud DNA sequencing libraries were constructed according to the manufacturers' instructions. Poly(A) RNA-Seq libraries were constructed using the NEB Ultra II RNA Library Prep kit. DNA and RNA sequencing were performed by the Scientific Operations core at the WSI on Pacific Biosciences SEQUEL II (HiFi), Illumina HiSeq 4000 (RNA-Seq) and Illumina NovaSeq 6000 (10X) instruments. Hi-C data were also generated from head tissue of iAgrCirc1 using the Arima v2 kit and sequenced on the Illumina NovaSeq 6000 instrument.

Genome assembly

Assembly was carried out with Hifiasm (Cheng *et al.*, 2021) and haplotypic duplication was identified and removed with purge_dups (Guan *et al.*, 2020). One round of polishing was performed by aligning 10X Genomics read data to the assembly with Long Ranger ALIGN, calling variants with freebayes (Garrison & Marth, 2012). The assembly was then scaffolded with Hi-C data (Rao *et al.*, 2014) using SALSA2 (Ghurye *et al.*, 2019). The assembly was checked for contamination as described previously (Howe *et al.*, 2021). Manual curation was performed using HiGlass (Kerpedjiev *et al.*, 2018) and PretextView (Harry, 2022). The mitochondrial genome was assembled using MitoHiFi (Uliano-Silva *et al.*, 2022), which performed annotation using MitoFinder (Allio *et al.*, 2020). 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.

Table 2. Chromosomal pseudomolecules in the genome assembly of *Agrochola circellaris*, iAgrCirc1.

INSDC accession	Chromosome	Size (Mb)	GC%
OU611839.1	1	32.98	38.2
OU611841.1	2	22.02	37.5
OU611842.1	3	21.65	37.8
OU611843.1	4	21.65	37.6
OU611844.1	5	21.14	37.3
OU611845.1	6	20.91	37.2
OU611846.1	7	20.69	37.8
OU611847.1	8	20.61	37.3
OU611848.1	9	20.51	37.6
OU611849.1	10	20.18	37.7
OU611850.1	11	19.8	37.8
OU611851.1	12	19.76	37.6
OU611852.1	13	19.71	37.2
OU611853.1	14	19.69	37.4
OU611854.1	15	19.42	37.4
OU611855.1	16	18.94	37.5
OU611856.1	17	18.87	37.7
OU611857.1	18	18.78	37.9
OU611858.1	19	18.66	37.6
OU611859.1	20	18.61	37.3
OU611860.1	21	18.28	37.9
OU611861.1	22	17.31	37.4
OU611862.1	23	16.69	37.6
OU611863.1	24	16.39	37.5
OU611864.1	25	14.66	37.6
OU611865.1	26	14.3	37.6
OU611866.1	27	12.68	38.2
OU611867.1	28	12.1	37.9
OU611868.1	29	11.39	37.9
OU611840.1	Z	23.61	37.7
OU611869.1	MT	0.02	20.3
-	-	0.18	37.7

Table 3. Software tools and versions used.

Software tool	Version	Source
BlobToolKit	3.4.0	Challis <i>et al.</i> , 2020
freebayes	1.3.1-17-gaa2ace8	Garrison & Marth, 2012
Hifiasm	0.15.1	Cheng <i>et al.</i> , 2021
HiGlass	1.11.6	Kerpedjiev <i>et al.</i> , 2018
Long Ranger ALIGN	2.2.2	https://support.10xgenomics.com/genome-exome/software/pipelines/latest/advanced/other-pipelines
MitoHiFi	2	Uliano-Silva <i>et al.</i> , 2022
PretextView	0.2	Harry, 2022
purge_dups	1.2.3	Guan <i>et al.</i> , 2020
SALSA	2.2	Ghurye <i>et al.</i> , 2019

Genome annotation

The BRAKER2 pipeline (Brůna *et al.*, 2021) was used in the default protein mode to generate annotation for the *Agrochola circellaris* assembly (GCA_914767755.1) in Ensembl Rapid Release.

Ethics/compliance issues

The materials that have contributed to this genome note have been supplied by a Darwin Tree of Life Partner. The submission of materials by a Darwin Tree of Life Partner is subject to the [Darwin Tree of Life Project Sampling Code of Practice](#). By agreeing with and signing up to the Sampling Code of Practice, the Darwin Tree of Life Partner agrees they will meet the legal and ethical requirements and standards set out within this document in respect of all samples acquired for, and supplied to, the Darwin Tree of Life Project. Each transfer of samples is further undertaken according to a Research Collaboration Agreement or Material Transfer Agreement entered into by the Darwin Tree of Life Partner, Genome Research Limited (operating as the Wellcome Sanger Institute), and in some circumstances other Darwin Tree of Life collaborators.

Data availability

European Nucleotide Archive: *Agrochola circellaris* (the brick). Accession number [PRJEB46316](#), <https://identifiers.org/ena.embl/PRJEB46316> (Wellcome Sanger Institute, 2021).

The genome sequence is released openly for reuse. The *Agrochola circellaris* genome sequencing initiative is part of the Darwin Tree of Life (DToL) project. All raw sequence data and the assembly have been deposited in INSDC databases. Raw data and assembly accession identifiers are reported in [Table 1](#).

Author information

Members of the University of Oxford and Wytham Woods Genome Acquisition Lab are listed here: <https://doi.org/10.5281/zenodo.4789928>.

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 programme are listed here: <https://doi.org/10.5281/zenodo.4783585>.

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

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

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

References

- 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](#)
- Bąk-Badowska J, Żeber-Dzikowska I, Chmielewski J: **The impact of brick (*Agrochola circellaris* Hufn.) and owl moths (Lepidoptera: Noctuidae) on the health of seeds of field elms (*Ulmus minor* Mill.) in the landscape parks of the Świętokrzyskie Province.** *Ochrona Srodowiska i Zasobów Naturalnych.* 2017; **28**(2): 41–45. [Publisher Full Text](#)
- Brůna T, Hoff KJ, Lomsadze A, *et al.*: **BRAKER2: Automatic eukaryotic genome annotation with GeneMark-EP+ and AUGUSTUS supported by a protein database.** *NAR Genom Bioinform.* 2021; **3**(1): lqaa108. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Butterfly Conservation: **The Brick, *Agrochola circellaris*.** 2006; (Accessed: 15 January 2023).
- 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](#)
- Chinery M: **Collins Guide to the Insects of Britain and Western Europe.** London: Collins, 1991.
- Garrison E, Marth G: **Haplotype-based variant detection from short-read sequencing.** 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): 2896–2898. [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): gjaa153. [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](#)
- Manni M, Berkeley MR, Seppy M, *et al.*: **BUSCO Update: Novel and Streamlined Workflows along with Broader and Deeper Phylogenetic Coverage for Scoring of Eukaryotic, Prokaryotic, and Viral Genomes.** *Mol Biol Evol.* 2021; **38**(10): 4647–4654. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Rao SS, Huntley MH, Durand NC, *et al.*: **A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping.** *Cell.* 2014; **159**(7): 1665–1680. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Regier JC, Mitter C, Mitter K, *et al.*: **Further progress on the phylogeny of Noctuoidea (Insecta: Lepidoptera) using an expanded gene sample.** *Systematic Entomology.* 2017; **42**(1): 82–93. [Publisher 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](#)
- Uliano-Silva M, Ferreira JG, Krasheninnikova K, *et al.*: **MitoHiFi: a python**

pipeline for mitochondrial genome assembly from PacBio High Fidelity reads. *bioRxiv*. [Preprint]. 2022.
[Publisher Full Text](#)

Usher M, Keiller S: **The macrolepidoptera of farm woodlands: determinants of diversity and community structure.** *Biodiversity and Conservation*. 1998; **7**: 725–748.
[Publisher Full Text](#)

Wellcome Sanger Institute: **The genome sequence of the Brick, *Agrochola circellaris* (Hufnagel, 1766).** European Nucleotide Archive. [Dataset]. 2021.
<https://identifiers.org/ena.embl/PRJEB46316>

Yang ZH, Yang TT, Liu Y, *et al.*: **The complete mitochondrial genome of *Sinna extrema* (Lepidoptera: Nolidae) and its implications for the phylogenetic relationships of Noctuoidea species.** *Int J Biol Macromol*. 2019; **137**: 317–326.
[PubMed Abstract](#) | [Publisher Full Text](#)

Open Peer Review

Current Peer Review Status:  

Version 1

Reviewer Report 30 September 2025

<https://doi.org/10.21956/wellcomeopenres.20950.r132257>

© 2025 Eldem V. 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.



Vahap Eldem 

Istanbul University, Istanbul, Turkey

I recommend acceptance of this data note following correction of one minor error. The manuscript successfully fulfills its intended purpose of documenting a high-quality genome assembly for *Agrochola circellaris* as part of the Darwin Tree of Life project. The assembly demonstrates excellent technical quality with a BUSCO completeness score of 99.2%, scaffold N50 of 19.7 Mb, and 99.99% of sequence assigned to chromosomal scaffolds. The methodological documentation is thorough and transparent, including all necessary software versions, protocols, and data accession numbers to ensure reproducibility. All raw sequencing data and assembled genomes have been properly deposited in public repositories with clearly reported accession identifiers.

The scope of the manuscript is appropriate for a data note format. Rather than presenting novel biological insights or extensive comparative analyses, it focuses appropriately on documenting the generation and quality characteristics of a genomic resource that will be valuable to the research community. The assembly metrics meet or exceed the stated benchmarks adapted from current genome assembly standards. One correction is required before indexing: on page 3, line 4 of the Background section, "Noctuoidae" is incorrectly typed and should read "Noctuidae." This error was correctly identified by the previous reviewer. Additionally, while not essential for acceptance, the manuscript could be marginally strengthened by briefly mentioning expected heterozygosity levels given that an alternate haplotype assembly is referenced, and by explicitly noting that the absence of chromosome W is expected for the male ZZ specimen used. I have a minor methodological question regarding tissue selection. The study used thorax and head tissues from specimen iAgrCirc1 for genomic sequencing and Hi-C respectively, while specimen iAgrCirc2 contributed only abdominal tissue for RNA-Seq. Could the authors briefly clarify the rationale for this tissue selection strategy? I'm curious whether this was driven by practical constraints (e.g., tissue quality, specimen condition) or specific experimental considerations. While this doesn't affect the quality of the reported genome assembly, a brief explanation would be informative for readers planning similar projects.

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: Mult,-omics of non-model organism

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 12 February 2024

<https://doi.org/10.21956/wellcomeopenres.20950.r72330>

© 2024 Kuppusamy S. This is an open access peer review report distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.



Sivasankaran Kuppusamy 

Loyola College, Chennai, India

I'd like to appreciate the authors for putting together the whole genome sequence of *Agrochola circellaris*. The authors sequenced over 572 million base pairs for the genome assembly using 34 scaffolds. The authors found 18,499 gene transcripts and 18,319 protein-coding genes during genome assembly.

Page no 3 second paragraph under the subheading "Background". Line no 4. The phylogeny of the Noctuoidae.....The family name Noctuidae was incorrectly typed. It should be Noctuidae not Noctuoidae.

Overall, the manuscript was well prepared.

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 Entomology

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.
