Check for updates

DATA NOTE

The genome sequence of a click beetle, *Agrypnus murinus* (Linnaeus, 1758) [version 1; peer review: 2 approved]

Liam M. Crowley [iD]1, Jennifer Sudworth [iD]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

1University of Oxford, Oxford, England, UK
2Independent researcher, Bristol, England, UK

Abstract

We present a genome assembly from an individual male *Agrypnus murinus* (a click beetle; Arthropoda; Insecta; Coleoptera; Elateridae). The genome sequence is 1,578.5 megabases in span. Most of the assembly is scaffolded into 9 chromosomal pseudomolecules, including the X sex chromosome. The mitochondrial genome has also been assembled and is 18.23 kilobases in length. Gene annotation of this assembly on Ensembl identified 42,204 protein coding genes.

Keywords

Agrypnus murinus, a click beetle, genome sequence, chromosomal, Coleoptera

This article is included in the Tree of Life gateway.

Open Peer Review

Approval Status ✔ ✔

	1	2
version 1
18 Aug 2023	✔
view	✔
view

1. **R Axel W Wiberg**, Stockholm University, Stockholm, Sweden

2. **Yuhao Huang** [iD], Sun Yat-sen University, Guangzhou, China
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: Crowley LM: Investigation, Resources, Writing – Review & Editing; **Sudworth J**: Writing – Original Draft Preparation;

How to cite this article: Crowley LM, Sudworth J, University of Oxford and Wytham Woods Genome Acquisition Lab *et al.* **The genome sequence of a click beetle, *Agrypnus murinus* (Linnaeus, 1758) [version 1; peer review: 2 approved]** Wellcome Open Research 2023, **8**:355 https://doi.org/10.12688/wellcomeopenres.19847.1

First published: 18 Aug 2023, **8**:355 https://doi.org/10.12688/wellcomeopenres.19847.1

Species taxonomy

Eukaryota; Metazoa; Eumetazoa; Bilateria; Protostomia; Ecdyso-zoa; Panarthropoda; Arthropoda; Mandibulata; Pancrustacea; Hexapoda; Insecta; Dicondylia; Pterygota; Neoptera; Endop-terygota; Coleoptera; Polyphaga; Elateriformia; Elateroidea; Elateridae; Agrypninae; Agrypnini; *Agrypnus*; *Agrypnus murinus* (Linnaeus, 1758) (NCBI:txid195168).

Background

Agrypnus murinus (Order: Coleoptera, Family: Elateridae) is a species of click beetle with a Palearctic distribution, with records throughout western Europe, and parts of Asia (GBIF Secretariat, 2022). Adults grow up to 17 mm in length, and have a distinctive dense, dark brown-grey, scale-like pubescence covering the fused elytra.

Agrypnus murinus inhabits scrubland and grasslands across the UK. It is more commonly found across mid- and southern England and Wales, although there are records of it as far north as Yorkshire (NBN Atlas Partnership, 2021). It is widespread across Europe, except in the far northern regions. Eggs are laid in soil during late summer and during the autumn *A. murinus* emerge in their larval form, known as wireworms. The larvae feed on the roots of plants and other insects, and can cause sig-nificant damage to crops such as potatoes, and are therefore considered a significant agricultural pest (Dettner & Beran, 2000).

The genome sequence of this beetle will be important in future studies involving environmental DNA traces or investigat-ing invertebrate communities for analysis of agricultural soils and pest control. The genome of *A. murinus* has now been sequenced as part of the Darwin Tree of Life Project, a col-laborative 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. murinus*, based on one male specimen from Wytham Woods, Oxfordshire, UK.

Genome sequence report

The genome was sequenced from one male *Agrypnus muri-nus* (Figure 1) collected from Wytham Woods, Oxfordshire,

Figure 1. Photograph of the *Agrypnus murinus* (icAgrMuri1) specimen used for genome sequencing.

UK (51.77, –1.34). A total of 28-fold coverage in Pacific Bio-sciences single-molecule HiFi long reads and 50-fold coverage in 10X Genomics read clouds were generated. Primary assem-bly contigs were scaffolded with chromosome conformation Hi-C data. Manual assembly curation corrected 405 missing joins or mis-joins and removed 10 haplotypic duplications, reduc-ing the assembly length by 0.12% and the scaffold number by 57.18%, and increasing the scaffold N50 by 7.32%.

The final assembly has a total length of 1578.5 Mb in 161 sequence scaffolds with a scaffold N50 of 177.3 Mb (Table 1). Most (99.59%) of the assembly sequence was assigned to 9 chro-mosomal-level scaffolds, representing 8 autosomes and the X sex chromosome. Chromosome-scale scaffolds confirmed by the Hi-C data are named in order of size (Figure 2–Figure 5; Table 2). 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.

The estimated Quality Value (QV) of the final assembly is 55.5 with k-mer completeness of 99.99%, and the assembly has a BUSCO v5.3.2 completeness of 99.2% (single = 96.8%, duplicated = 2.4%), using the endopterygota_odb10 reference set ($n = 2,124$).

Metadata for specimens, spectral estimates, sequencing runs, contaminants and pre-curation assembly statistics can be found at https://links.tol.sanger.ac.uk/species/195168.

Genome annotation report

The *Agrypnus murinus* genome assembly (GCA_929113105.1) was annotated using the Ensembl rapid annotation pipeline (Table 1; https://rapid.ensembl.org/Agrypnus_murinus_GCA_929113105.1/Info/Index). The resulting annotation includes 42,570 transcribed mRNAs from 42,204 protein-coding genes.

Methods

Sample acquisition and nucleic acid extraction

A male *Agrypnus murinus* (icAgrMuri1) was collected from Wytham Woods, Oxfordshire (biological vice-county Berkshire), UK (latitude 51.77, longitude –1.34) on 2020-06-20 by netting. Liam Crowley (University of Oxford) collected and identified the specimen. The specimen was snap-frozen on dry ice.

DNA was extracted at the Tree of Life laboratory, Wellcome Sanger Institute (WSI). The icAgrMuri1 sample was weighed and dissected on dry ice with tissue set aside for Hi-C sequenc-ing. Abdomen tissue was disrupted using a Nippi Power-masher fitted with a BioMasher pestleHigh 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 the 0.8X AMpure XP purification kit prior to 10X Chromium sequenc-ing; a minimum of 50 ng DNA was submitted for 10X sequenc-ing. 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

Table 1. Genome data for *Agrypnus murinus*, icAgrMuri1.1.

Project accession data	
Assembly identifier	icAgrMuri1.1
Species	*Agrypnus murinus*
Specimen	icAgrMuri1
NCBI taxonomy ID	195168
BioProject	PRJEB46297
BioSample ID	SAMEA7701277
Isolate information	icAgrMuri1, male; abdomen (DNA sequencing), head and thorax (Hi-C scaffolding)

Assembly metrics*		Benchmark
Consensus quality (QV)	55.5	≥ 50
k-mer completeness	99.99%	≥ 95%
BUSCO**	C:99.2%[S:96.8%,D:2.4%],F:0.5%,M:0.4%,n:2,124	C ≥ 95%
Percentage of assembly mapped to chromosomes	99.59%	≥ 95%
Sex chromosomes	X chromosome	localised homologous pairs
Organelles	Mitochondrial genome assembled	complete single alleles

Raw data accessions	
PacificBiosciences SEQUEL II	ERR6807991, ERR6897496
10X Genomics Illumina	ERR6363338–ERR6363345
Hi-C Illumina	ERR6363337

Genome assembly	
Assembly accession	GCA_929113105.1
Accession of alternate haplotype	GCA_929114945.1
Span (Mb)	1,578.5
Number of contigs	986
Contig N50 length (Mb)	4.5
Number of scaffolds	161
Scaffold N50 length (Mb)	177.3
Longest scaffold (Mb)	255.3

Genome annotation	
Number of protein-coding genes	42,204
Number of gene transcripts	42,570

* 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 v5.3.2. C = complete [S = single copy, D = duplicated], F = fragmented, M = missing, n = number of orthologues in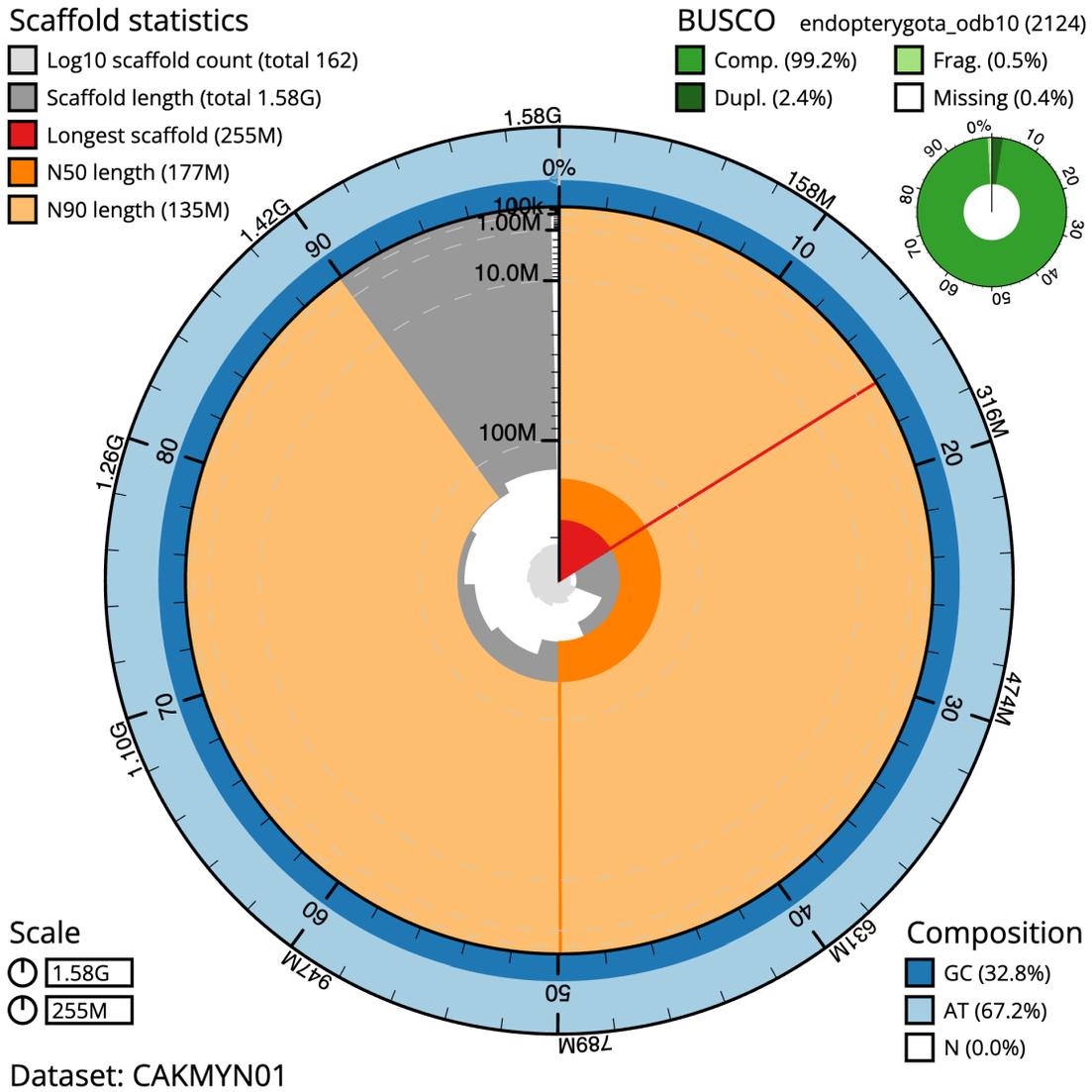 comparison. A full set of BUSCO scores is available at https://blobtoolkit.genomehubs.org/view/icAgrMuri1.1/dataset/CAKMYN01/busco.

Figure 2. Genome assembly of *Agrypnus murinus*, icAgrMuri1.1: 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 1,578,523,859 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 (255,265,360 bp, shown in red). Orange and pale-orange arcs show the N50 and N90 scaffold lengths (177,293,933 and 135,112,989 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/icAgrMuri1.1/dataset/CAKMYN01/snail.

using AMPure PB beads with a 1.8X ratio of beads to sample to remove the shorter fragments and concentrate the DNA sample. The concentration of the sheared and purified DNA was assessed using a Nanodrop spectrophotometer and Qubit Fluorometer and Qubit dsDNA High Sensitivity Assay kit. Fragment size distribution was evaluated by running the sample on the FemtoPulse system.

Sequencing
Pacific Biosciences HiFi circular consensus and 10X Genomics read cloud DNA sequencing libraries were constructed according to the manufacturers' instructions. DNA sequencing was performed by the Scientific Operations core at the WSI on Pacific Biosciences SEQUEL II (HiFi) and Illumina NovaSeq 6000 (10X) instruments. Hi-C data were also generated from
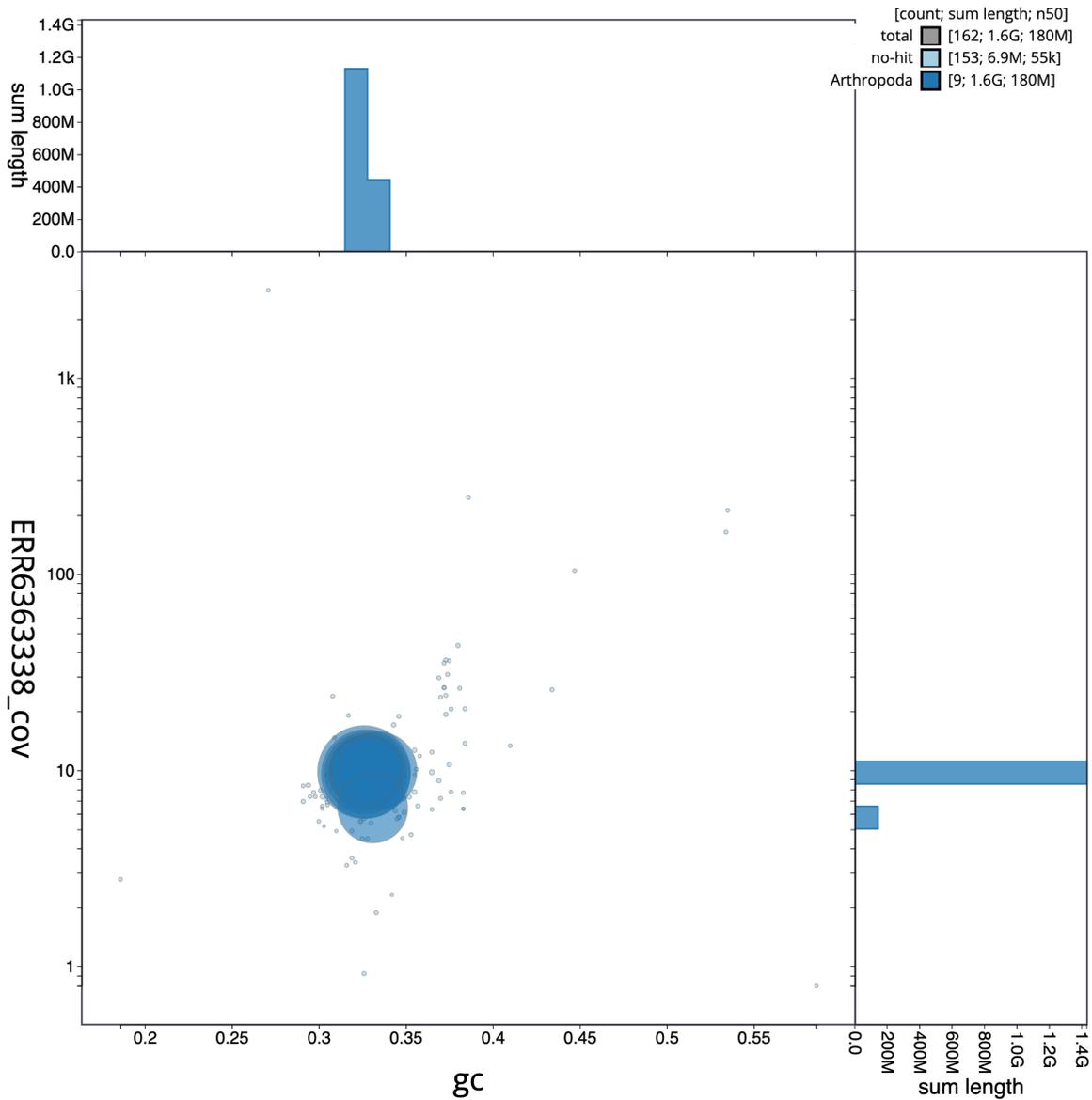
Figure 3. Genome assembly of *Agrypnus murinus*, icAgrMuri1.1: BlobToolKit GC-coverage plot. Scaffolds are coloured by phylum. Circles are sized in proportion to scaffold length. Histograms show the distribution of scaffold length sum along each axis. An interactive version of this figure is available at https://blobtoolkit.genomehubs.org/view/icAgrMuri1.1/dataset/CAKMYN01/blob.

head and thorax tissue of icAgrMuri1 using the Arimav2 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). 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 YaHS (Zhou *et al.*, 2023). 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 (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.
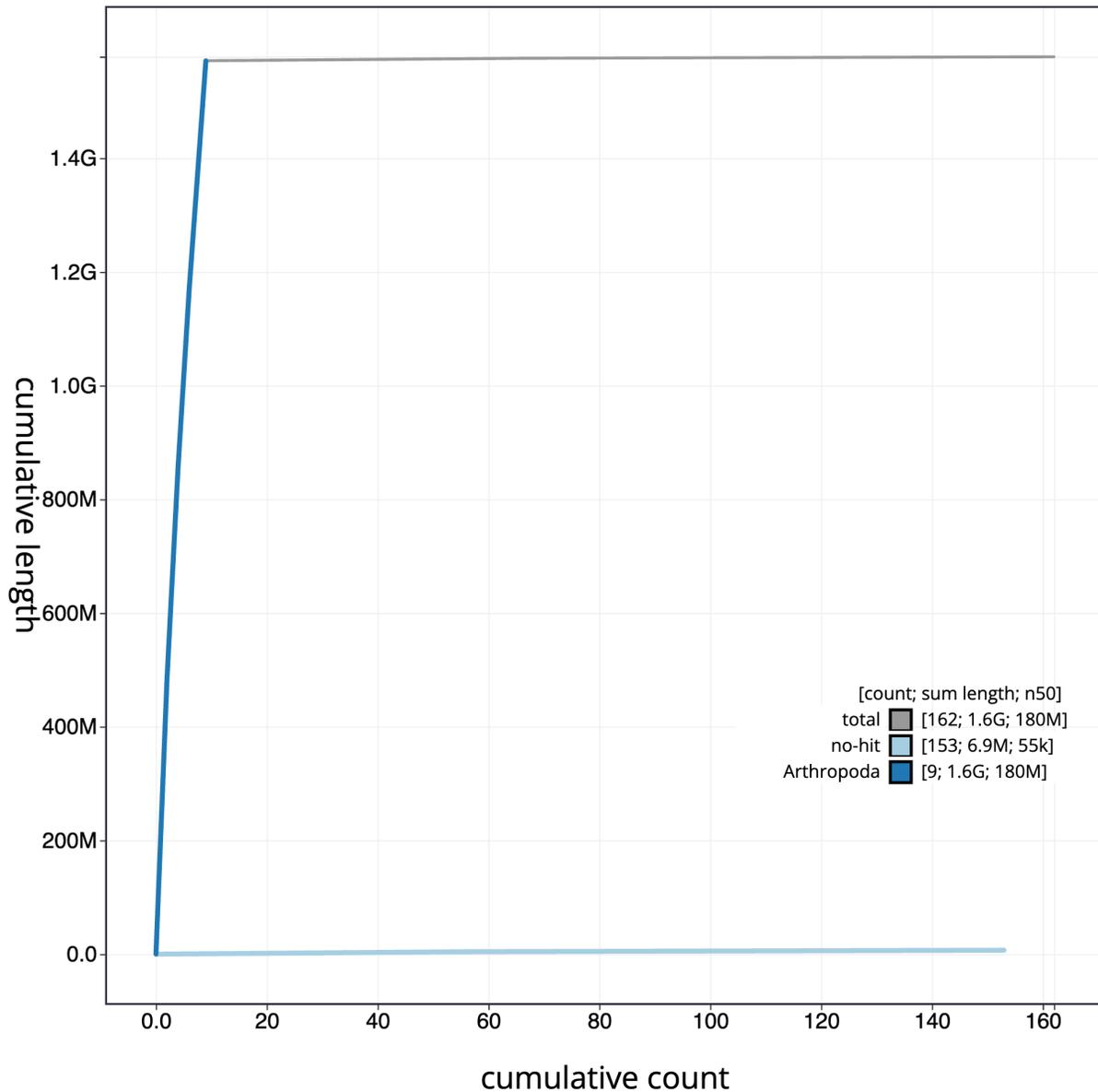
Figure 4. Genome assembly of *Agrypnus murinus*, icAgrMuri1.1: BlobToolKit cumulative sequence plot. The grey line shows cumulative length for all scaffolds. Coloured lines show cumulative lengths of scaffolds assigned to each phylum using the buscogenes taxrule. An interactive version of this figure is available at https://blobtoolkit.genomehubs.org/view/icAgrMuri1.1/dataset/CAKMYN01/cumulative.

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

Genome annotation
The BRAKER2 pipeline (Brůna *et al.*, 2021) was used in the default protein mode to generate annotation for the *Agrypnus*
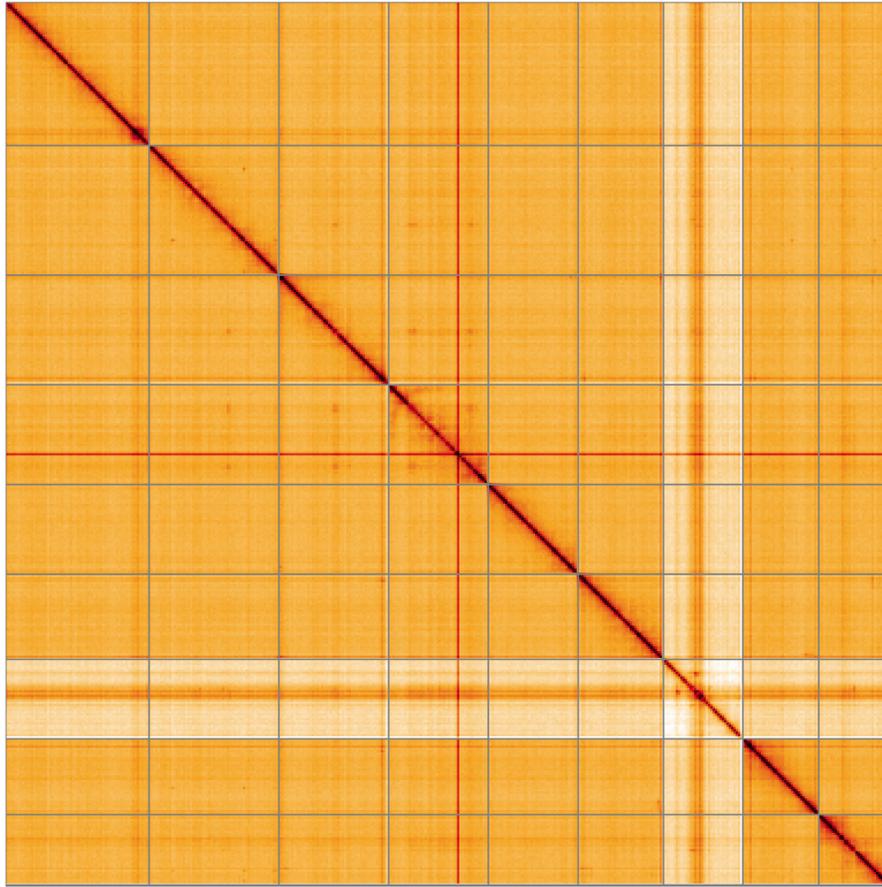
Figure 5. Genome assembly of *Agrypnus murinus*, icAgrMuri1.1: Hi-C contact map of the icAgrMuri1.1 assembly, visualised using HiGlass. Chromosomes are shown in order of size from left to right and top to bottom. An interactive version of this figure may be viewed at https://genome-note-higlass.tol.sanger.ac.uk/l/?d=EbECSrE6ShqDN7sF_hOHsA.

Table 2. Chromosomal pseudomolecules in the genome assembly of *Agrypnus murinus*, icAgrMuri1.

INSDC accession	Chromosome	Length (Mb)	GC%
OV816024.1	1	255.27	32.5
OV816025.1	2	230.79	32.5
OV816026.1	3	195.33	32.5
OV816027.1	4	177.29	33.5
OV816028.1	5	159.83	32.5
OV816029.1	6	151.69	32.5
OV816030.1	7	135.11	32.5
OV816031.1	8	125.14	33.0
OV816023.1	X	141.13	33.0
OV816032.1	MT	0.02	27.0

murinus assembly (GCA_929113105.1) in Ensembl Rapid Release.

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.

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

Table 3. Software tools: versions and sources.

Software tool	Version	Source
BlobToolKit	4.1.5	https://github.com/blobtoolkit/blobtoolkit
BUSCO	5.3.2	https://gitlab.com/ezlab/busco
FreeBayes	1.3.1-17-gaa2ace8	https://github.com/freebayes/freebayes
gEVAL	N/A	https://geval.org.uk/
Hifiasm	0.15.3	https://github.com/chhylp123/hifiasm
HiGlass	1.11.6	https://github.com/higlass/higlass
Long Ranger ALIGN	2.2.2	https://support.10xgenomics.com/genome-exome/software/pipelines/latest/advanced/other-pipelines
Merqury	MerquryFK	https://github.com/thegenemyers/MERQURY.FK
MitoHiFi	2	https://github.com/marcelauliano/MitoHiFi
PretextView	0.2	https://github.com/wtsi-hpag/PretextView
purge_dups	1.2.3	https://github.com/dfguan/purge_dups
sanger-tol/genomenote	v1.0	https://github.com/sanger-tol/genomenote
sanger-tol/readmapping	1.1.0	https://github.com/sanger-tol/readmapping/tree/1.1.0
YaHS	1.0	https://github.com/c-zhou/yahs

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: *Agrypnus murinus*. Accession number PRJEB46297; https://identifiers.org/ena.embl/PRJEB46297. (Wellcome Sanger Institute, 2022)

The genome sequence is released openly for reuse. The *Agrypnus murinus* 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

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**
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**

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**

Challis R, Richards E, Rajan J, *et al.*: **BlobToolKit - Interactive Quality Assessment of Genome Assemblies.** *G3 (Bethesda).* 2020; **10**(4): 1361–1374.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

Cheng H, Concepcion GT, Feng X, *et al.*: **Haplotype-resolved *de novo* assembly using phased assembly graphs with hifiasm.** *Nat Methods.* 2021; **18**(2): 170–175.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

Chow W, Brugger K, Caccamo M, *et al.*: **gEVAL - a web-based browser for evaluating genome assemblies.** *Bioinformatics.* 2016; **32**(16): 2508–2510.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

Dettner K, Beran A: **Chemical defense of the fetid smelling click beetle *Agrypnus murinus* (Coleoptera: Elateridae).** *Entomologica Generalis.* 2000; **25**(1): 27–32.
Reference Source

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**

Garrison E, Marth G: **Haplotype-based variant detection from short-read sequencing.** 2012; [Accessed 26 July 2023].
Publisher Full Text

GBIF Secretariat: ***Agrypnus murinus* (Linnaeus, 1758).** *GBIF Backbone Taxonomy.* 2022; [Accessed 17 May 2023].
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): 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**

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

NBN Atlas Partnership: ***Agrypnus murinus* (Linnaeus, 1758).** NBN Atlas, 2021; [Accessed 17 May 2023].
Reference Source

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.*: **Merqury: Reference-free quality, completeness, and phasing assessment for genome assemblies.** *Genome Biol.* 2020; **21**(1): 245.
PubMed Abstract | **Publisher Full Text** | **Free 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**

Surana P, Muffato M, Qi G: **sanger-tol/readmapping: sanger-tol/readmapping v1.1.0 - Hebridean Black (1.1.0).** *Zenodo.* 2023a; [Accessed 21 July 2023].
Publisher Full Text

Surana P, Muffato M, Sadasivan Baby C: **sanger-tol/genomenote (v1.0.dev).** *Zenodo.* 2023b; [Accessed 21 July 2023].
Reference Source

Uliano-Silva M, Ferreira JGRN, Krasheninnikova 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 a click beetle, *Agrypnus murinus* (Linnaeus, 1758).** European Nucleotide Archive. [dataset], accession number PRJEB46297, 2022.

Zhou C, McCarthy SA, Durbin R: **YaHS: yet another Hi-C scaffolding tool.** *Bioinformatics.* 2023; **39**(1): btac808.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

Open Peer Review

Current Peer Review Status: ✓ ✓

✓   **Yuhao Huang** [iD]

[1] Sun Yat-sen University, Guangzhou, China
[2] Institute of Zoology, Guangdong Academy of Sciences, Guangzhou, Guangdong, China

The data note presents a high-quality, chromosome-level genome assembly of the click beetle *Agrypnus murinus*, with a completeness of 99.2%. This genome assembly will be valuable for understanding this agricultural pest and for research into genetic diversity.

The rationale for creating the dataset is well-justified. The sequencing and assembly methods employed are both reasonable and reproducible. Additionally, the raw data, assembly, and annotations are accessible.

I approve of the manuscript in its current form. However, there are a few comments that could enhance the manuscript and the dataset for user convenience:

1. Although the link to the annotations is provided in the genome annotation report section, it is recommended to also include this link in the data availability section to improve visibility.

2. In the mitochondrial genome assembly method, "MitoFinder or MITOS" was used for the annotation. Please clarify which tool was used for this species. In addition, it would be beneficial to provide the mitochondrial genome annotation as well.

3. The predicted gene count is 42,204, which seems unusually high based on my experience with beetle genomes, unless a large-scale duplication has occurred in this species. This high number might result from fragmented gene predictions by BRAKER2, particularly if only the protein mode was utilized. I have noticed transcriptomic sequencing data under BioProject accession PRJEB46297. A more refined version of gene predictions using transcriptomic data and the complete Ensembl pipeline would be advantageous for users of this genome.

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, 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 11 June 2024

https://doi.org/10.21956/wellcomeopenres.21980.r85302

✔ **R Axel W Wiberg**
Stockholm University, Stockholm, Sweden

The Data Note describes the genome assembly of beetle species *Agrypnus murinus* using the DTOL standardised pipeline. Overall I think the manuscript is very good.

The rationale for creating the dataset is well described. Although I think there could be some important additions here (see below).

It is technically sound with the methods/protocols well described and referenced.

The references to the DTOL nextflow pipelines and other software used ensure that replication would be straightforward.

The results and datasets are mostly clearly presented (see below). The genome sequence and the annotation are easily available.

I think the article is valid in it's current form, but do have a few additional comments:

1) In the 1st paragraph of the "Background" section the authors write:

"Adults grow up to 17 mm in length.."

There is often a sexual dimorphism in insects, I would like to see sizes for males and females separately, or if there is no sexual dimorphism for that to be stated.

2) Coleopterans vary a lot in their sex-determination systems and sex-chromosome karyology. While most are XY, XO systems are common in several groups. For this species, and especially because the specimen sequenced here is male, it would be really useful to note in the text that, while the karyology of *Agrypnus murinus* seems unknown at the moment, other close relatives are XO. Indeed, among the Elateridae, XO systems reach nearly 70% of examined species. Thus, a Y chromosome is not expected in the assembly. Adding this information would add context to those using the genome, increase it the generality of the note.

See the very useful database here: http://coleoguy.github.io/karyotypes/
and the citations here:
[ref 1]
[ref 2]

In addition, the data note mentions 8 autosomes and the X chromosome in the assembly. Is the general karyology of the species well known? Are 8 autosomes expected, I'm not sure if there is information on the full karyotype of the species?

3) The ~42,000 genes seems on the high end from my experience. In another Coleopteran genome that I was involved in, annotation was quite fragmented at least in part due to the high number of repetitive elements. I think it would be useful to include in this Data Note (and perhaps in DTOL's standardised pipelines in general) a simple RepeatModeler and RepeatMasker analysis to get at least a rough picture of the repeat content.

RepeatModeler: https://www.repeatmasker.org/RepeatModeler/

RepeatMasker: https://www.repeatmasker.org/

It is in general a useful thing for a genome assembly to have as it helps in all sorts of down-stream applications.

4) Figure 5 would benefit from some kind of x- and y- axis labels at least for the largest scaffolds. Which is the X chromosome? One can work it out from table 2, but axis labels would be better I think.

5) As far as I can tell, the mtGenome is present as a contig in the assembly file, but the annotation files do not include the mtGenome. Would it be possible to integrate the MITOS annotation somehow? Or provide it separately? This would make the resource much more accessible and useful.

References
1. Reference Source
2. Reference Source

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

Partly

Are the protocols appropriate and is the work technically sound?
Yes

Are sufficient details of methods and materials provided to allow replication by others?
Yes

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

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: evolutionary biology, comparative and population genomics, sexual selection and sexual conflict

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.