DATA NOTE

# The genome sequence of the Mottled Pug, *Eupithecia exiguata* (Hübner, 1813) [version 1; peer review: 1 approved, 2 approved with reservations]

Douglas Boyes[1+], Owen T. Lewis[2],
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

[1]UK Centre for Ecology & Hydrology, Wallingford, England, UK
[2]University of Oxford, Oxford, England, UK

[+] Deceased author

**v1** **First published:** 19 Feb 2024, **9**:65
https://doi.org/10.12688/wellcomeopenres.20637.1
**Latest published:** 19 Feb 2024, **9**:65
https://doi.org/10.12688/wellcomeopenres.20637.1

## Abstract

We present a genome assembly from an individual male *Eupithecia exiguata* (the Mottled Pug; Arthropoda; Insecta; Lepidoptera; Geometridae). The genome sequence is 372.9 megabases in span. Most of the assembly is scaffolded into 31 chromosomal pseudomolecules, including the Z sex chromosome. The mitochondrial genome has also been assembled and is 16.39 kilobases in length. Gene annotation of this assembly on Ensembl identified 11,194 protein coding genes.

## Keywords

Eupithecia exiguata, Mottled Pug, genome sequence, chromosomal, Lepidoptera

This article is included in the Tree of Life gateway.

**Open Peer Review**

**Approval Status** ? ? ✓

| | 1 | 2 | 3 |
|---|---|---|---|
| **version 1** 19 Feb 2024 | ? view | ? view | ✓ view |

1. **Marco Gerdol**, University of Trieste, Trieste, Italy
2. **Jaakko Pohjoismäki**, University of Eastern Finland, Joensuu, Finland
3. **Saskia Wutke**, University or Eastern Finland, Joensuu, Finland

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; **Lewis OT**: 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.*



**How to cite this article:** Boyes D, Lewis OT, University of Oxford and Wytham Woods Genome Acquisition Lab *et al.* **The genome sequence of the Mottled Pug, *Eupithecia exiguata* (Hübner, 1813) [version 1; peer review: 1 approved, 2 approved with reservations]** Wellcome Open Research 2024, **9**:65 https://doi.org/10.12688/wellcomeopenres.20637.1

**First published:** 19 Feb 2024, **9**:65 https://doi.org/10.12688/wellcomeopenres.20637.1

## Species taxonomy

Eukaryota; Metazoa; Eumetazoa; Bilateria; Protostomia; Ecdysozoa; Panarthropoda; Arthropoda; Mandibulata; Pancrustacea; Hexapoda; Insecta; Dicondylia; Pterygota; Neoptera; Endopterygota; Amphiesmenoptera; Lepidoptera; Glossata; Neo lepidoptera;Heteroneura; Ditrysia; Obtectomera; Geometroidea; Geometridae; Larentiinae; *Eupithecia; Eupithecia exiguata* (Hübner, 1813) (NCBI:txid934847).

## Background

The Mottled Pug (*Eupithecia exiguata*) is a small geometrid moth. Its larvae feed on Hawthorn, Blackthorn and other shrubs (Henwood *et al.*, 2020; Waring *et al.*, 2017). It occurs in woodland, hedgerow and garden habitats and is common and widespread across much of England and Wales. It is also widespread in Ireland but there are fewer records from Scotland, where it is spreading; its distribution overall has increased markedly since 1970 (Randle *et al.*, 2019). Globally, The Mottled Pug occurs across Europe and Asia to the Pacific coast of Russia and China (GBIF Secretariat, 2023).

The genome of the mottled pug, *Eupithecia exiguata*, 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 *Eupithecia exiguata*, based on one male specimen from Wytham Woods, Oxfordshire, UK.

## Genome sequence report

The genome was sequenced from one male *Eupithecia exiguata* (Figure 1) collected from Wytham Woods, Oxfordshire, UK (51.77, –1.32). A total of 56-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 20 missing joins or mis-joins and removed one haplotypic duplication, reducing the scaffold number by 15.38%.


**Figure 1. Photograph of the *Eupithecia exiguata* (ilEupExig1) specimen used for genome sequencing.**

The final assembly has a total length of 372.9 Mb in 32 sequence scaffolds with a scaffold N50 of 13.1 Mb (Table 1). Most (99.99%) of the assembly sequence was assigned to 31 chromosomal-level scaffolds, representing 30 autosomes and the Z 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 66.5 with *k*-mer completeness of 100%, and the assembly has a BUSCO v5.3.2 completeness of 97.9% (single = 97.4%, duplicated = 0.5%), using the lepidoptera_odb10 reference set ($n$ = 5,286).

Metadata for specimens, spectral estimates, sequencing runs, contaminants and pre-curation assembly statistics can be found at https://links.tol.sanger.ac.uk/species/934847.

## Genome annotation report

The *Eupithecia exiguata* genome assembly (GCA_947086465.1) was annotated using the Ensembl rapid annotation pipeline (Table 1; https://rapid.ensembl.org/Eupithecia_exiguata_GCA_947086465.1/Info/Index). The resulting annotation includes 19,529 transcribed mRNAs from 11,194 protein-coding and 1,243 non-coding genes.

## Methods

### Sample acquisition and nucleic acid extraction

A male *Eupithecia exiguata* (specimen ID Ox001895, individual ilEupExig1) was collected from Wytham Woods, Oxfordshire, UK (latitude 51.77, longitude –1.32) on 2021-05-28 using a light trap. The specimen was collected and identified by Douglas Boyes (University of Oxford) and preserved on dry ice.

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; HMW DNA fragmentation; and fragmented DNA clean-up. The sample was prepared for DNA extraction at the WSI Tree of Life laboratory: the ilEupExig1 sample was weighed and dissected on dry ice with tissue set aside for Hi-C sequencing (https://dx.doi.org/10.17504/protocols.io.x54v9prmqg3e/v1). Tissue from the whole organism was disrupted using a Nippi Powermasher fitted with a BioMasher pestle (https://dx.doi.org/10.17504/protocols.io.5qpvo3r19v4o/v1). DNA was extracted at the WSI Scientific Operations core using the Qiagen MagAttract HMW DNA kit, according to the manufacturer's instructions.

**Table 1. Genome data for *Eupithecia exiguata*, ilEupExig1.1.**

| **Project accession data** | | |
|---|---|---|
| Assembly identifier | ilEupExig1.1 | |
| Species | *Eupithecia exiguata* | |
| Specimen | ilEupExig1 | |
| NCBI taxonomy ID | 934847 | |
| BioProject | PRJEB55723 | |
| BioSample ID | SAMEA10979157 | |
| Isolate information | ilEupExig1, male: whole organism (DNA sequencing and Hi-C scaffolding) | |
| **Assembly metrics*** | | ***Benchmark*** |
| Consensus quality (QV) | 66.5 | *≥ 50* |
| *k*-mer completeness | 100% | *≥ 95%* |
| BUSCO** | C:97.9%[S:97.4%,D:0.5%], F:0.6%,M:1.5%,n:5,286 | *C ≥ 95%* |
| Percentage of assembly mapped to chromosomes | 99.99% | *≥ 95%* |
| Sex chromosomes | Z chromosome | *localised homologous pairs* |
| Organelles | Mitochondrial genome assembly | *complete single alleles* |
| **Raw data accessions** | | |
| PacificBiosciences SEQUEL II | ERR10168716 | |
| Hi-C Illumina | ERR10149546 | |
| **Genome assembly** | | |
| Assembly accession | GCA_947086465.1 | |
| *Accession of alternate haplotype* | GCA_947086475.1 | |
| Span (Mb) | 372.9 | |
| Number of contigs | 118 | |
| Contig N50 length (Mb) | 5.4 | |
| Number of scaffolds | 32 | |
| Scaffold N50 length (Mb) | 13.1 | |
| Longest scaffold (Mb) | 19.9 | |
| **Genome annotation** | | |
| Number of protein-coding genes | 11,194 | |
| Number of non-coding genes | 1,243 | |
| Number of gene transcripts | 19,529 | |

* 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.* (Rhie *et al.*, 2021).

** BUSCO scores based on the lepidoptera_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/Eupithecia%20exiguata/dataset/CAMTYU01/busco.

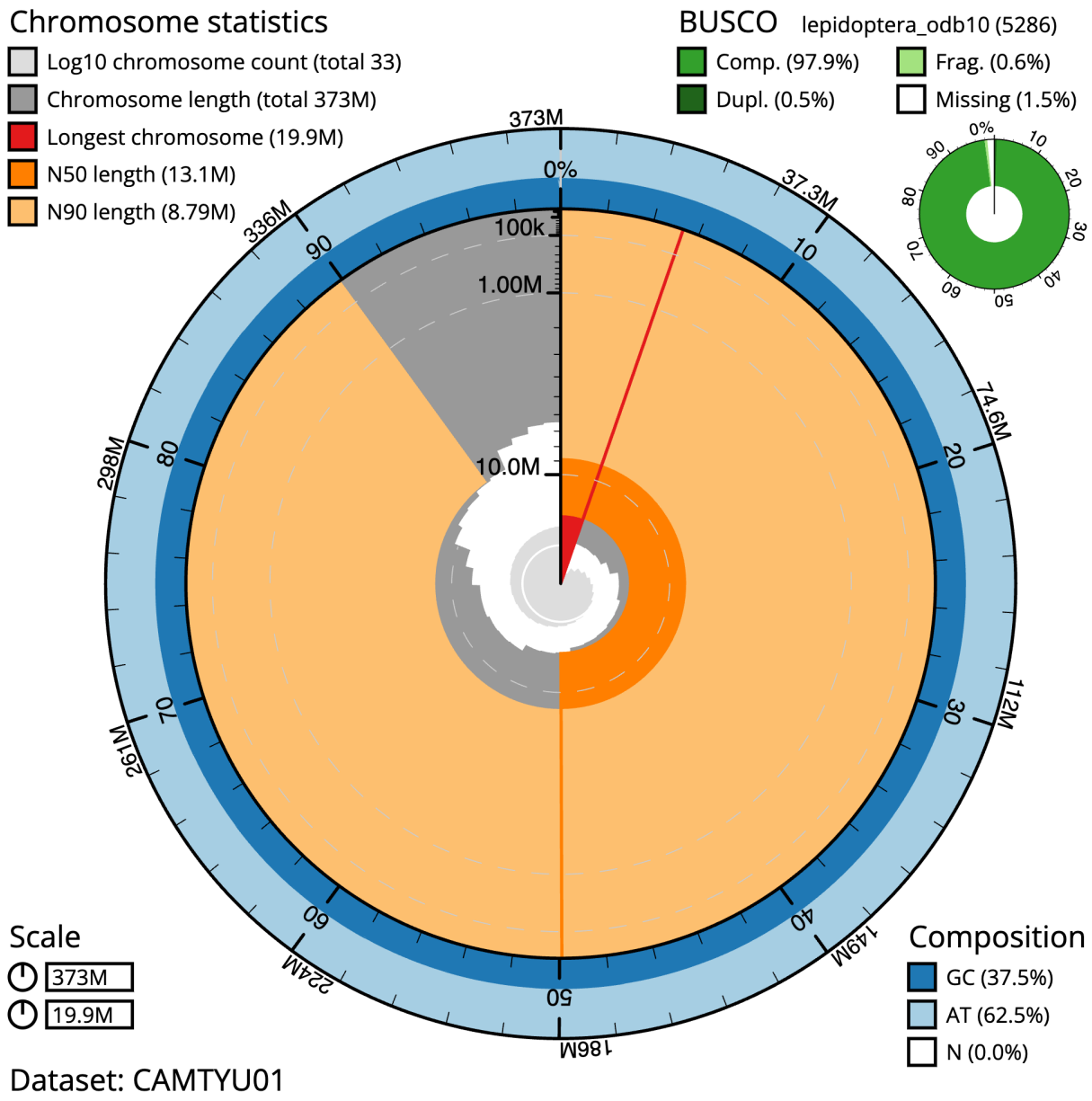


**Figure 2. Genome assembly of *Eupithecia exiguata*, ilEupExig1.1: metrics.** The BlobToolKit Snailplot shows N50 metrics and BUSCO gene completeness. The main plot is divided into 1,000 size-ordered bins around the circumference with each bin representing 0.1% of the 372,887,712 bp assembly. The distribution of sequence lengths is shown in dark grey with the plot radius scaled to the longest sequence present in the assembly (19,876,806 bp, shown in red). Orange and pale-orange arcs show the N50 and N90 sequence lengths (13,143,396 and 8,794,744 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/Eupithecia%20exiguata/dataset/CAMTYU01/snail.

Protocols developed in the Tree of Life laboratory are publicly available on protocols.io (https://dx.doi.org/10.17504/protocols.io.8epv5xxy6g1b/v1).

## 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 remaining tissue of ilEupExig1 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

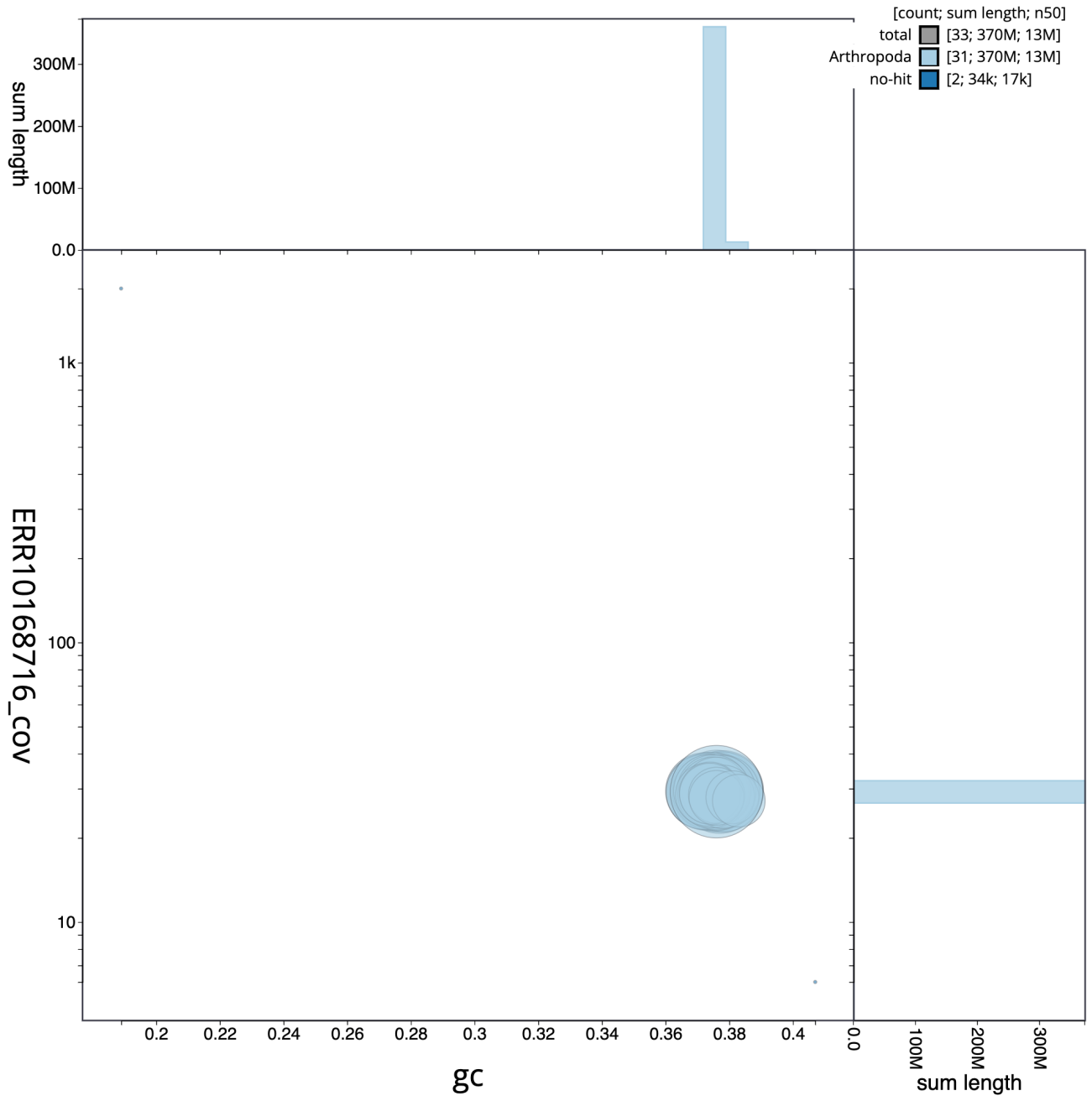


**Figure 3. Genome assembly of *Eupithecia exiguata*, ilEupExig1.1: BlobToolKit GC-coverage plot.** Scaffolds are coloured by phylum. Circles are sized in proportion to scaffold length. Histograms show the distribution of scaffold length sum along each axis. An interactive version of this figure is available at https://blobtoolkit.genomehubs.org/view/Eupithecia%20exiguata/dataset/CAMTYU01/blob.

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 Pretext (Harry, 2022). The mitochondrial genome was assembled using MitoHiFi (Uliano-Silva *et al.*, 2022), 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 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,

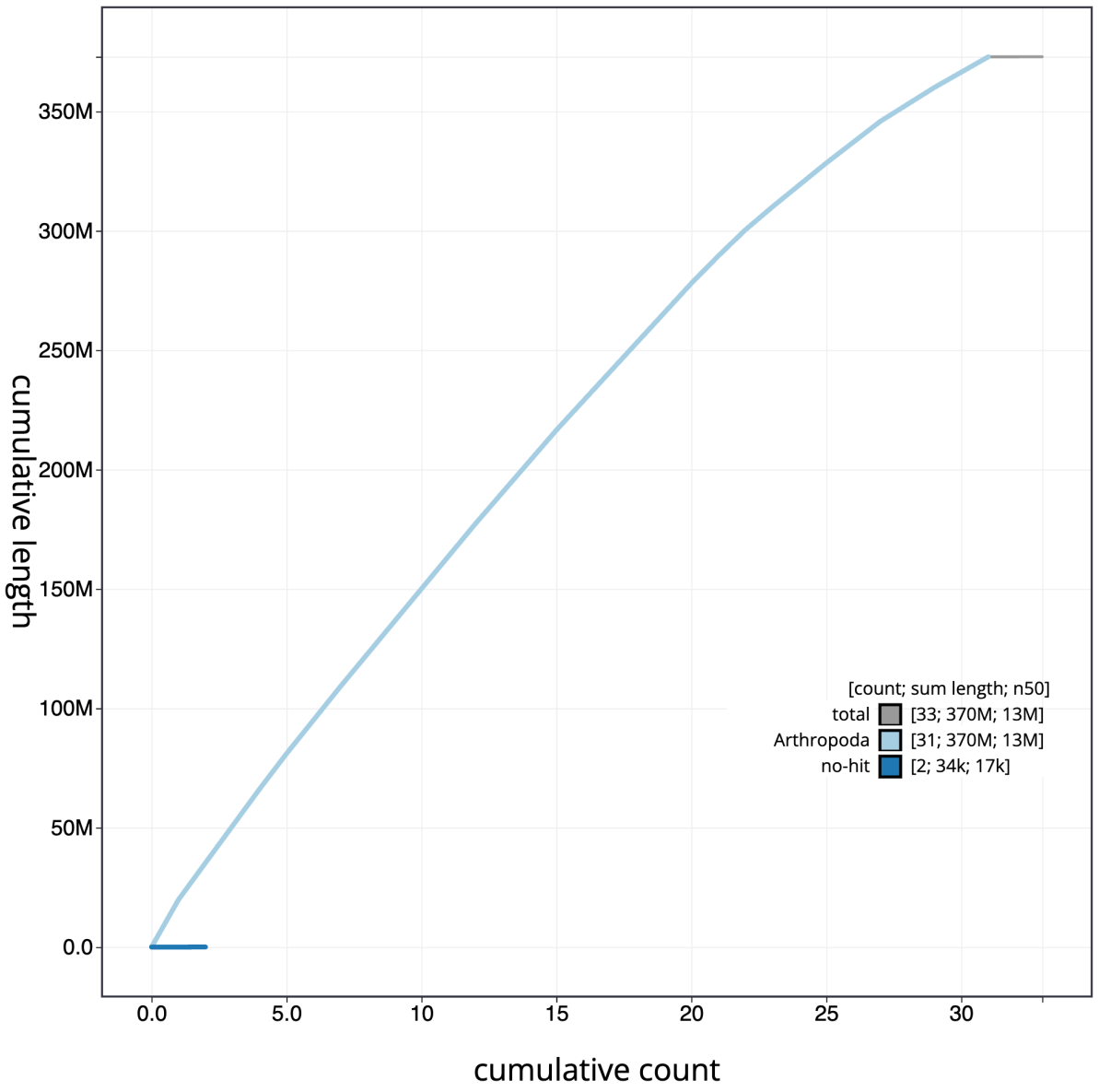


**Figure 4. Genome assembly of *Eupithecia exiguata*, ilEupExig1.1: BlobToolKit cumulative sequence plot.** The grey line shows cumulative length for all scaffolds. Coloured lines show cumulative lengths of scaffolds assigned to each phylum using the buscogenes taxrule. An interactive version of this figure is available at https://blobtoolkit.genomehubs.org/view/Eupithecia%20exiguata/dataset/CAMTYU01/cumulative.

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 Ensembl gene annotation system (Aken *et al.*, 2016) was used to generate annotation for the *Eupithecia exiguata* assembly (GCA_947086465.1). Annotation was created primarily through alignment of transcriptomic data to the genome,

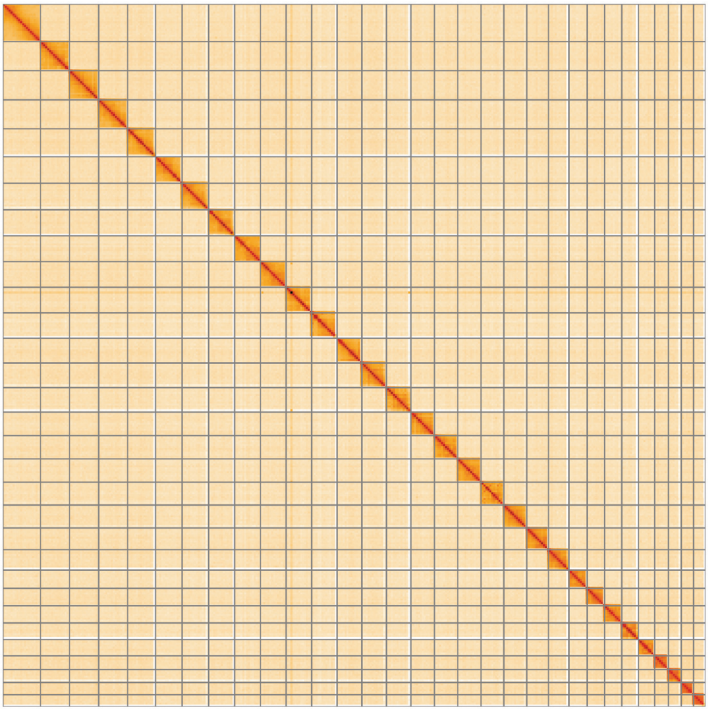

**Figure 5. Genome assembly of *Eupithecia exiguata*, ilEupExig1.1: Hi-C contact map of the ilEupExig1.1 assembly, visualised using HiGlass.** Chromosomes are shown in order of size from left to right and top to bottom. An interactive version of this figure may be viewed at https://genome-note-higlass.tol.sanger.ac.uk/l/?d=PURr_tyLQ6Obof4vztyYVA.

**Table 2. Chromosomal pseudomolecules in the genome assembly of *Eupithecia exiguata*, ilEupExig1.**

| INSDC accession | Chromosome | Length (Mb) | GC% |
|---|---|---|---|
| OX352227.1 | 1 | 15.5 | 37.5 |
| OX352228.1 | 2 | 15.45 | 37.5 |
| OX352229.1 | 3 | 15.33 | 38.0 |
| OX352230.1 | 4 | 14.83 | 38.0 |
| OX352231.1 | 5 | 14.09 | 37.5 |
| OX352232.1 | 6 | 14.08 | 37.5 |
| OX352233.1 | 7 | 13.77 | 37.5 |
| OX352234.1 | 8 | 13.74 | 37.5 |
| OX352235.1 | 9 | 13.61 | 37.5 |
| OX352236.1 | 10 | 13.55 | 37.5 |
| OX352237.1 | 11 | 13.48 | 37.0 |
| OX352238.1 | 12 | 13.14 | 37.5 |
| OX352239.1 | 13 | 13.06 | 37.5 |
| OX352240.1 | 14 | 13.01 | 37.0 |
| OX352241.1 | 15 | 12.44 | 37.5 |
| OX352242.1 | 16 | 12.38 | 38.0 |
| OX352243.1 | 17 | 12.33 | 37.5 |
| OX352244.1 | 18 | 12.16 | 37.5 |
| OX352245.1 | 19 | 12.15 | 37.5 |
| OX352246.1 | 20 | 11.5 | 37.5 |
| OX352247.1 | 21 | 10.88 | 37.5 |
| OX352248.1 | 22 | 9.62 | 37.5 |
| OX352249.1 | 23 | 9.26 | 38.0 |
| OX352250.1 | 24 | 9.14 | 37.5 |
| OX352251.1 | 25 | 8.79 | 37.5 |
| OX352252.1 | 26 | 8.64 | 37.5 |
| OX352253.1 | 27 | 7.28 | 37.5 |
| OX352254.1 | 28 | 6.9 | 37.5 |
| OX352255.1 | 29 | 6.48 | 38.0 |
| OX352256.1 | 30 | 6.38 | 38.5 |
| OX352226.1 | Z | 19.88 | 37.5 |
| OX352257.1 | MT | 0.02 | 19.0 |

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

| Software tool | Version | Source |
|---|---|---|
| BlobToolKit | 4.0.7 | https://github.com/blobtoolkit/blobtoolkit |
| BUSCO | 5.3.2 | https://gitlab.com/ezlab/busco |
| Hifiasm | 0.16.1-r375 | https://github.com/chhylp123/hifiasm |
| HiGlass | 1.11.6 | https://github.com/higlass/higlass |
| 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 | yahs-1.1.91eebc2 | https://github.com/c-zhou/yahs |

with gap filling via protein-to-genome alignments of a select set of proteins from UniProt (UniProt Consortium, 2019).

### 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 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: *Eupithecia exiguata* (mottled pug). Accession number PRJEB55723; https://identifiers.org/ena.embl/PRJEB55723 (Wellcome Sanger Institute, 2022). The genome sequence is released openly for reuse. The *Eupithecia exiguata* 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.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

Aken BL, Ayling S, Barrell D, *et al.*: **The Ensembl gene annotation system.** *Database (Oxford).* 2016; **2016**: baw093.
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

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

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: ***Eupithecia exiguata* (Hübner).** *GBIF Backbone Taxonomy.* [Preprint], 2023; (Accessed: 2 November 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

Henwood B, Sterling P, Lewington R: **Field Guide to the Caterpillars of Great Britain and Ireland.** London: Bloomsbury, 2020.
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

Randle Z, Evans-Hill L, Parsons MS, *et al.*: **Atlas of Britain & Ireland's Larger Moths.** Newbury: NatureBureau, 2019.
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].
Publisher Full Text

Uliano-Silva M, Ferreira JGRN, Krasheninnikova K, *et al.*: **MitoHiFi: a python pipeline for mitochondrial genome assembly from PacBio High Fidelity reads.** *bioRxiv.* 2022.
Publisher Full Text

UniProt Consortium: **UniProt: a worldwide hub of protein knowledge.** *Nucleic Acids Res.* 2019; **47**(D1): D506–D515.
PubMed Abstract | Publisher Full Text | Free Full Text

Vasimuddin Md, 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

Waring P, Townsend M, Lewington R: **Field Guide to the Moths of Great Britain and Ireland: Third Edition.** Bloomsbury Wildlife Guides, 2017.
Reference Source

Wellcome Sanger Institute: **The genome sequence of the Mottled Pug, Eupithecia exiguata (Hübner, 1813).** European Nucleotide Archive. [dataset], accession number PRJEB55723, 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: ? ? ✓

**Version 1**

Reviewer Report 15 July 2024

https://doi.org/10.21956/wellcomeopenres.22840.r82099



✓ **Saskia Wutke**
University or Eastern Finland, Joensuu, Finland

This data note reports the high-quality genome of the geometrid moth *Eupithecia exiguata* (Mottled Pug). Genome sequencing, assembly and annotation is based PacBio-HiFi and Hi-C data. Overall, the dataset is presented clearly and concisely. The methods are state-of-the-art and follow the usual DToL procedures. Nevertheless, as with other DToL genome reports, I would wish for more details in the method section, particularly concerning the software settings and the protein set used for genome annotation. Likewise, the background could be more elaborate with some explanation about its relevance. It seems like the background section of DToL genome reports gets shorter and shorter.
Still, the genome provides a valuable resource for evolutionary and comparative genomic research.

**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?**
Partly

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

***Competing Interests:*** No competing interests were disclosed.

***Reviewer Expertise:*** Genome sequencing, genomics, phylogenomics, Hymenoptera

**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 16 May 2024

https://doi.org/10.21956/wellcomeopenres.22840.r78189



?

**Jaakko Pohjoismäki**
University of Eastern Finland, Joensuu, Finland

The manuscript by Boyes, Lewis & the DToL consortium presents the reference-level genome assembly for the geometrid moth *Eupithecia exiguata*. The genome quality and assembly corresponds to the standards of previous reference genome releases and has been obtained using the standard DToL pipeline.
Unfortunately the genome is again from a male and thus lacks the W sex chromosome. I think that more effort should be put into obtaining the heterogametic sex for the reference genomes, as sequencing another reference genome from the same species is unlikely or at least low priority. Also this species is very common and can be also found resting on walls etc. during days, so obtaining a female is not very difficult. Otherwise I have only minor suggestions:

The background is quite sparingly written. *Eupithecia* is a very species-rich genus, with diverse host plant associations and habitat requirements. Some more information about the diversity in the UK would be useful.  Also as some (or most) species can be difficult to determine, it would be important to explain how *exiguata* can be differentiated from other similar species. Despite these minor shortcomings, this high quality genome assembly is a nice addition to the growing number of available lepidopteran reference genomes.

**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:*** mitochondrial DNA, biodiversity genomics, taxonomy

**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 11 April 2024

https://doi.org/10.21956/wellcomeopenres.22840.r78208



?

**Marco Gerdol**
University of Trieste, Trieste, Italy

This manuscript reports a high-quality genome assembly of the Mottled Pug Eupithecia exiguata, using the standardized format of the Darwin Tree of Life genome reports. This resource was obtained through the application of gold standard methodologies and the metrics of the assembly itself are excellent. Compared with other similar papers, I also appreciate the presence of a brief gene annotation summary, created with the Ensembl rapid pipeline. There are, however, a few points that would benefit from some additional clarification:

- It is unclear whether the number of pseudo chromosomal scaffolds does actually match the expected number of chromosomes from previous cytogenetic studies (if available).
- The authors mentioned that genome annotation was aided by transcriptomic data, but it is unclear whether this data was generated within the frame of this study, or it was previously available. In both cases, the materials and methods section should specify the accession IDs of RNA-seq of the datasets used for this purpose.
- The data availability section should include a link to the Ensembl rapid webpage with the genome annotation.
- It would be useful also to briefly mentioned the observed level of heterozygosity in the results section.

**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?**
Partly

***Competing Interests:*** No competing interests were disclosed.

***Reviewer Expertise:*** Invertebrate genomics

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