



DATA NOTE

REVISED The genome sequence of the Tufted Button, *Acleris cristana* (Denis & Schiffermüller, 1775)

[version 2; peer review: 2 approved]

Douglas Boyes¹⁺, Clare Boyes^{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

¹UK Centre for Ecology & Hydrology, Wallingford, England, UK

²Independent researcher, Welshpool, Wales, UK

+ Deceased author

v2 First published: 07 Jun 2023, 8:236
<https://doi.org/10.12688/wellcomeopenres.19508.1>
Latest published: 27 Aug 2026, 8:236
<https://doi.org/10.12688/wellcomeopenres.19508.2>

Abstract

We present a genome assembly from an individual female *Acleris cristana* (the Tufted Button; Arthropoda; Insecta; Lepidoptera; Tortricidae). The genome sequence is 562.6 megabases in span. Most of the assembly is scaffolded into 31 chromosomal pseudomolecules, including the W and Z sex chromosomes. The mitochondrial genome has also been assembled and is 16.1 kilobases in length. Gene annotation of this assembly on Ensembl identified 12,598 protein coding genes.

Keywords

Acleris cristana, tufted button, genome sequence, chromosomal, Lepidoptera



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

Open Peer Review

Approval Status ✓✓

	1	2
version 2 (revision) 27 Aug 2026		✓ view
version 1 07 Jun 2023	✓ view	↑ ?

1. **Feng Zhang** ^{id}, Nanjing Agricultural University, Nanjing, China
2. **Yannis Schoeneberg** ^{id}, Trier University, Tier, Germany

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; Boyes C: 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: © 2026 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, Boyes C, University of Oxford and Wytham Woods Genome Acquisition Lab *et al.* **The genome sequence of the Tufted Button, *Acleris cristana* (Denis & Schiffermüller, 1775) [version 2; peer review: 2 approved]** Wellcome Open Research 2026, 8:236 <https://doi.org/10.12688/wellcomeopenres.19508.2>

First published: 07 Jun 2023, 8:236 <https://doi.org/10.12688/wellcomeopenres.19508.1>

REVISED Amendments from Version 1

We have added detail to many of the methods and listed more software versions and sources in Table 3.

We have replaced the chromosome map in Figure 5 with a labelled version.

We have updated the MerquryFK results using a newer version of the software.

Any further responses from the reviewers can be found at the end of the article

Species taxonomy

Eukaryota; Opisthokonta; Metazoa; Eumetazoa; Bilateria; Protostomia; Ecdysozoa; Panarthropoda; Arthropoda; Mandibulata; Pancrustacea; Hexapoda; Insecta; Dicondylia; Pterygota; Neoptera; Endopterygota; Amphiesmenoptera; Lepidoptera; Glossata; Neolepidoptera; Heteroneura; Ditrysia; Apoditrysia; Tortricodea; Tortricidae; Tortricinae; Tortricini; *Acleris*; *Acleris cristana* (Denis & Schiffermüller, 1775) (NCBI:txid758705).

Background

Acleris cristana (the Tufted Button) is a micro-moth in the family Tortricidae. The species has a southerly distribution in Britain and is found throughout mainland Europe. There are also a few records from Japan (GBIF Secretariat, 2023).

A. cristana is probably the most variable species amongst British Lepidoptera with over 130 named forms (Sterling & Parsons, 2023). It is thought that the several of the genes influencing colours of individual pattern elements found on the wings of *A. cristana* segregate independently, which has resulted in numerous forms with very little gradation between them. In contrast, the closely related species, *A. hastiana*, which is also very variable, demonstrates many intermediate forms making it difficult to separate out the named forms (Hancock *et al.*, 2015). In *A. cristana*, although the forewing colour varies significantly, the moth almost always has a distinctive tuft of raised scales in the centre of the forewing, giving rise to the common name of ‘Tufted Button’.

The adult moth flies at dusk and in the UK is on the wing from September to mid-April. However, this includes a period of dormancy, from late autumn until early spring, after which the moth awakens to mate. Eggs are laid singly, or in small groups, on twigs of trees in the Rosaceae family, usually blackthorn (*Prunus spinosa*). The larvae can be found in rolled leaf edges, and later instars are found in spun leaves. The larvae pupate between June to August either in folded leaves, or on the ground in leaf litter (Emmet, 2010). The moth occasionally comes to light but can also be found by beating shrubs in the day.

A genome sequence from *A. cristana* will be useful for research into colour variation in moths, and more generally for comparative studies across the Lepidoptera. The genome of *A. cristana* 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. cristana* based on one female specimen from Wytham Woods, Oxfordshire, UK.

Genome sequence report

The genome was sequenced from one female *Acleris cristana* (Figure 1) collected from Wytham Woods, Oxfordshire, UK (51.77, -1.34). A total of 37-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 53 missing joins or mis-joins and removed six haplotypic duplications, reducing the assembly length by 0.4% and the scaffold number by 36.62%, and increasing the scaffold N50 by 10.95%.



Figure 1. Photograph of the *Acleris cristana* (ilAclCris2) specimen used for genome sequencing.

The final assembly has a total length of 562.6 Mb in 45 sequence scaffolds with a scaffold N50 of 17.8 Mb (Table 1). Most (99.72%) of the assembly sequence was assigned to 31 chromosomal-level scaffolds, representing 29 autosomes and the Z and W sex chromosomes. The W and Z chromosomes are similar in size, and are large chromosomes, in line with the cytogenetic findings of (Šichová *et al.*, 2013). The Z chromosome was identified based on alignment with *Acleris emargana* (GCA_927399475.2) which was assembled from a male sample (Z chromosome only). Chromosome-scale scaffolds confirmed by the Hi-C data are named in order of size (Figure 2–Figure 5; Table 2). The primary haplotype was assembled, and contigs corresponding to an alternate haplotype were also deposited in INSDC databases. The mitochondrial genome was also assembled and can be found as a contig within the multifasta file of the genome submission.

Table 1. Genome data for *Acleris cristana*, ilAclCris2.1.

Project accession data	
Assembly identifier	ilAclCris2.1
Species	<i>Acleris cristana</i>
Specimen	ilAclCris2
NCBI taxonomy ID	758705
BioProject	PRJEB58659
BioSample ID	SAMEA8603216
Isolate information	ilAclCris2, female; whole organism (PacBio sequencing) ilAclCris1; whole organism (Hi-C scaffolding)

Table 1. *Continued*

Assembly metrics*		Benchmark
Consensus quality (QV)	Primary: 65.9; alternate: 65.4; combined: 65.6	≥ 40
<i>k</i> -mer completeness	Primary: 87.39%; alternate: 73.18%; combined: 99.52%	$\geq 90\%$
BUSCO**	C:98.3%[S:97.6%,D:0.8%], F:0.5%,M:1.2%,n:5,286	S > 90%; D < 5%
Percentage of assembly mapped to chromosomes	99.72%	$\geq 90\%$
Sex chromosomes	Z and W chromosomes	<i>localised homologous pairs</i>
Organelles	Mitochondrial genome assembled	<i>complete single alleles</i>
Raw data accessions		
PacificBiosciences SEQUEL II	ERR10753928	
Hi-C Illumina	ERR10742410	
Genome assembly		
Assembly accession	GCA_948252455.1	
<i>Accession of alternate haplotype</i>	GCA_948250105.1	
Span (Mb)	562.6	
Number of contigs	156	
Contig N50 length (Mb)	6.5	
Number of scaffolds	45	
Scaffold N50 length (Mb)	17.8	
Longest scaffold (Mb)	62.1	
Genome annotation		
Number of protein-coding genes	12,598	
Number of non-coding genes	1,631	
Number of gene transcripts	22,144	

*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 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/ilAcIris2.1/dataset/CAOCPX01/busco>.

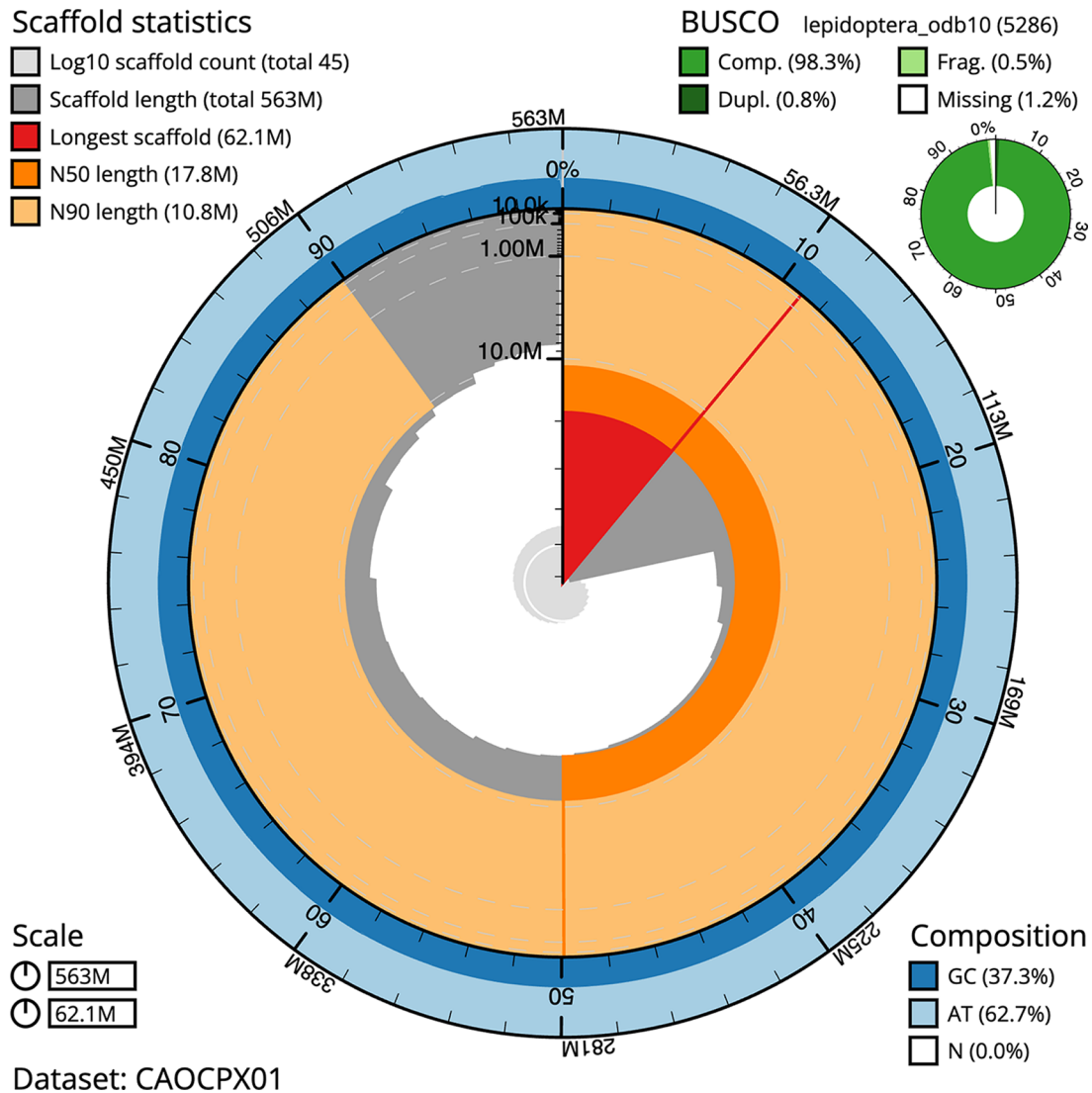


Figure 2. Genome assembly of *Acleris cristana*, ilAclCris2.1: metrics. The BlobToolKit Snailplot shows N50 metrics and BUSCO gene completeness. The main plot is divided into 1,000 size-ordered bins around the circumference with each bin representing 0.1% of the 562,572,194 bp assembly. The distribution of scaffold lengths is shown in dark grey with the plot radius scaled to the longest scaffold present in the assembly (62,083,584 bp, shown in red). Orange and pale-orange arcs show the N50 and N90 scaffold lengths (17,811,303 and 10,842,494 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 lepidoptera_odb10 set is shown in the top right. An interactive version of this figure is available at <https://blobtoolkit.genomehubs.org/view/ilAclCris2.1/dataset/CAOCPX01/snail>.

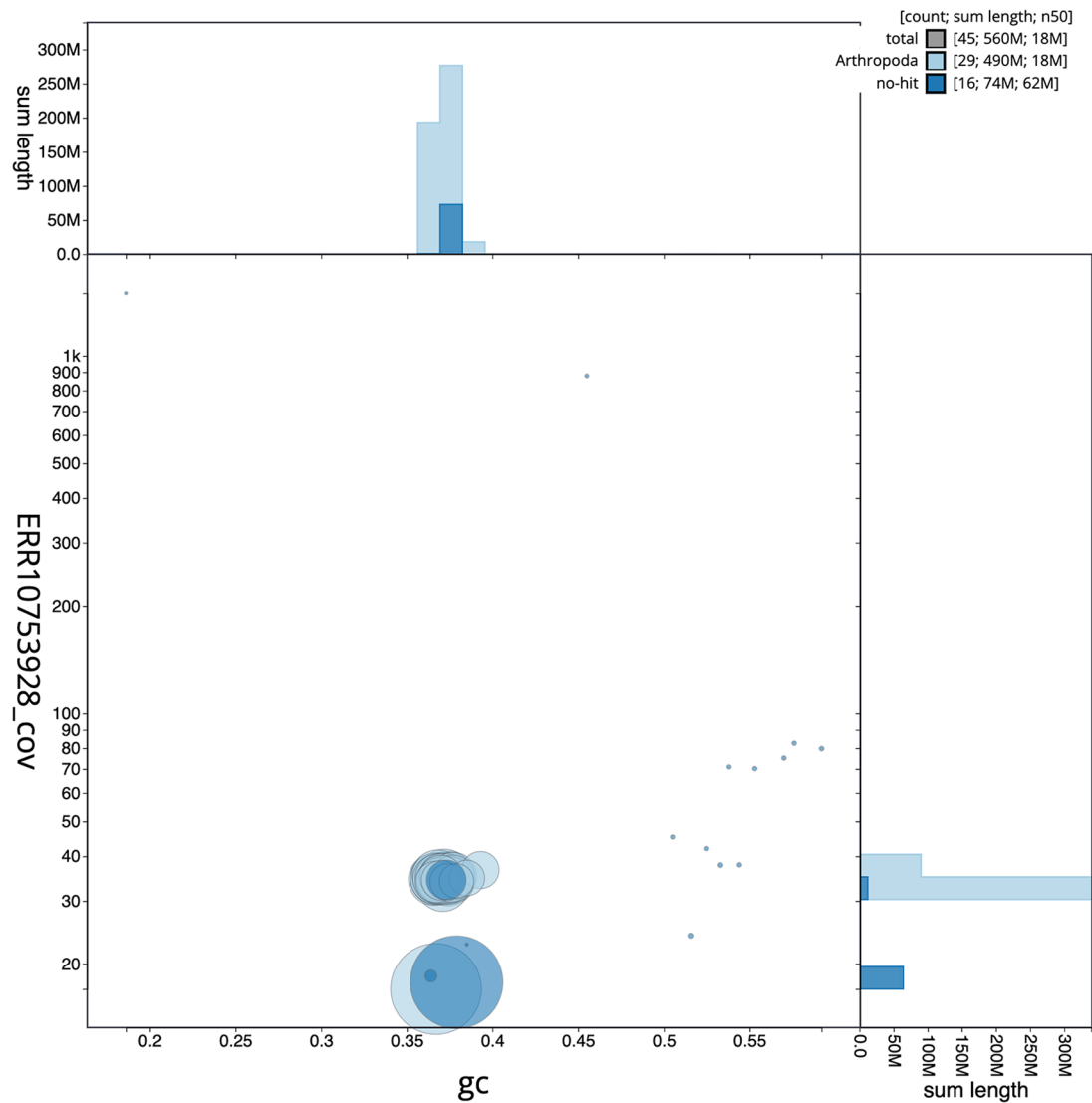


Figure 3. Genome assembly of *Acleris cristana*, ilAclCris2.1: BlobToolKit GC-coverage plot. Scaffolds are coloured by phylum. Circles are sized in proportion to scaffold length. Histograms show the distribution of scaffold length sum along each axis. An interactive version of this figure is available at <https://blobtoolkit.genomehubs.org/view/ilAclCris2.1/dataset/CAOCPX01/blob>.

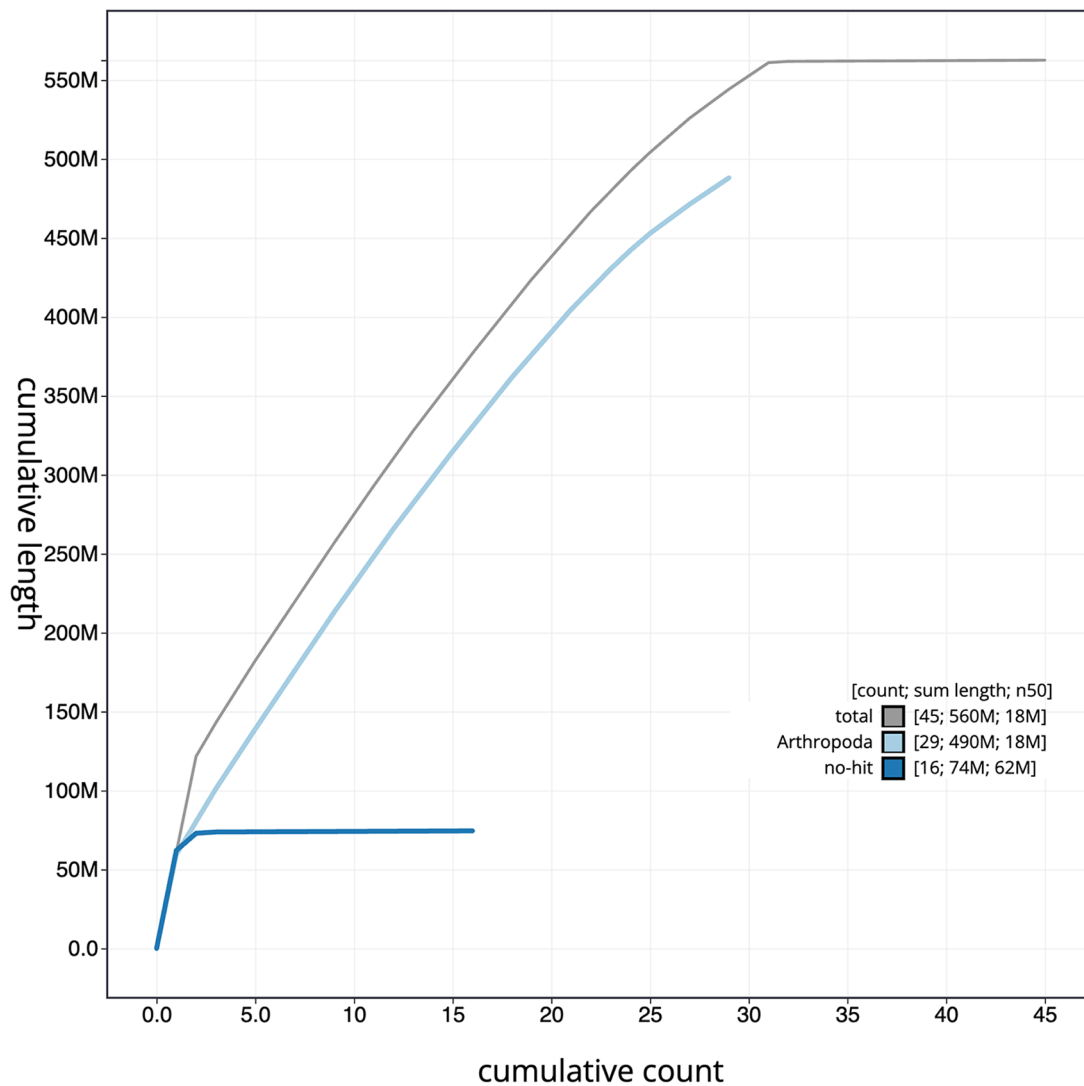


Figure 4. Genome assembly of *Acleris cristana*, ilAclCris2.1: BlobToolKit cumulative sequence plot. The grey line shows cumulative length for all scaffolds. Coloured lines show cumulative lengths of scaffolds assigned to each phylum using the buscogenes taxrule. An interactive version of this figure is available at <https://blobtoolkit.genomehubs.org/view/ilAclCris2.1/dataset/CAOCPX01/cumulative>.

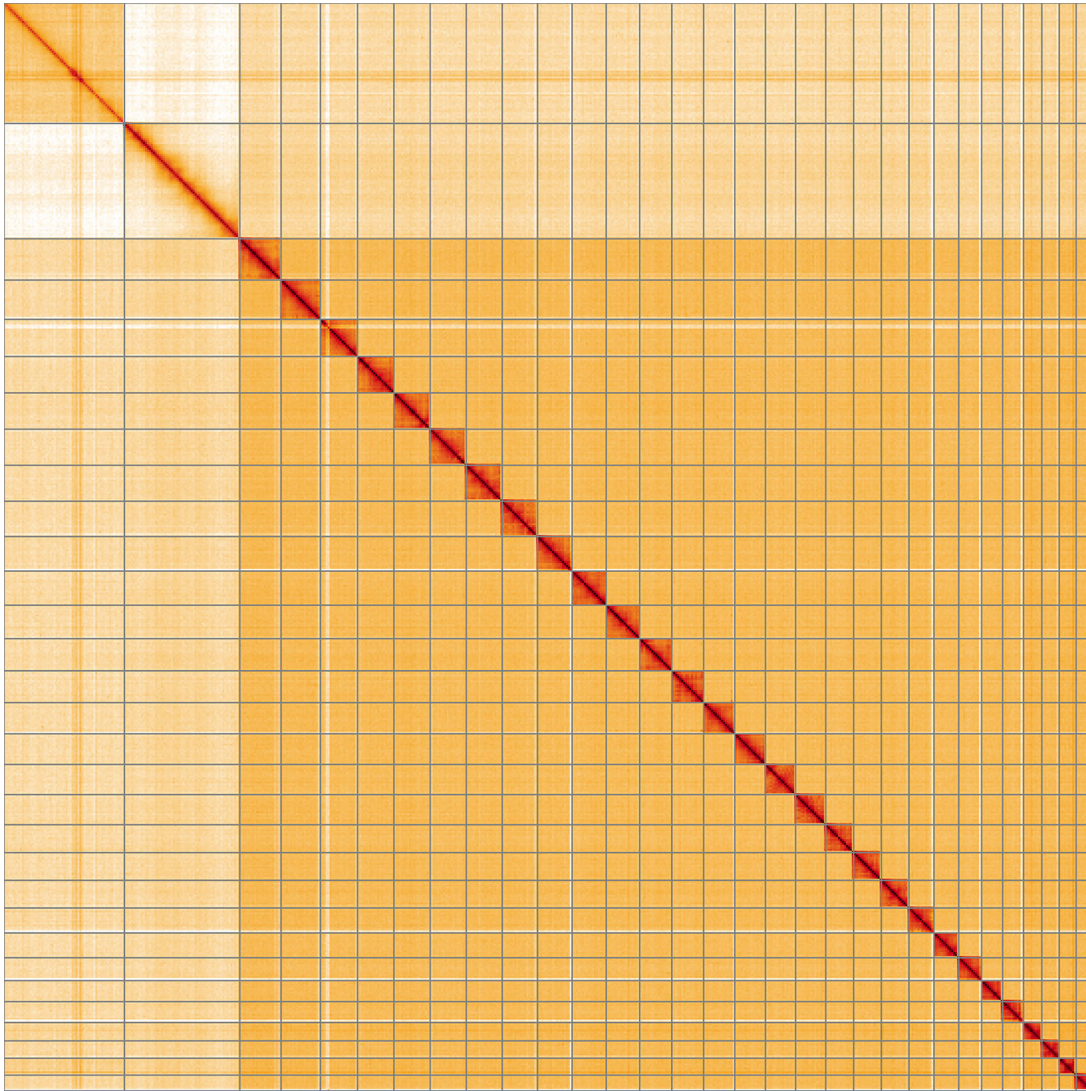


Figure 5. Genome assembly of *Acleris cristana*, ilAclCris2.1: Hi-C contact map of the ilAclCris2.1 assembly, visualised using PretextSnapshot. Chromosomes are shown in order of size from left to right and top to bottom. An interactive version of this figure may be viewed in HiGlass at <https://genome-note-higlass.tol.sanger.ac.uk/l/?d=DY0q7m1NSh2aFyNPJT93pw>.

Table 2. Chromosomal pseudomolecules in the genome assembly of *Acleris cristana*, ilAclCris2.

INSDC accession	Name	Length (Mb)	GC%
OX411773.1	1	21.28	37.5
OX411774.1	2	20.25	37
OX411775.1	3	19.17	37
OX411776.1	4	18.76	36.5
OX411777.1	5	18.72	37.5
OX411778.1	6	18.6	36.5
OX411779.1	7	18.55	36.5
OX411780.1	8	18.19	37
OX411781.1	9	17.81	37
OX411782.1	10	17.62	37
OX411783.1	11	17.24	37
OX411784.1	12	16.61	37.5
OX411785.1	13	16.38	37
OX411786.1	14	16.05	37.5
OX411787.1	15	15.83	37.5
OX411788.1	16	15.59	37
OX411789.1	17	15.49	37.5
OX411790.1	18	14.4	37
OX411791.1	19	14.32	37
OX411792.1	20	14.26	37.5
OX411793.1	21	12.86	37.5
OX411794.1	22	12.76	37
OX411795.1	23	11.82	36.5
OX411796.1	24	10.84	37.5
OX411797.1	25	10.78	37
OX411798.1	26	9.42	39.5
OX411799.1	27	9	37.5
OX411800.1	28	8.7	38.5
OX411801.1	29	8.09	38
OX411771.1	W	62.08	38
OX411772.1	Z	59.52	36.5
OX411802.1	MT	0.02	19

The combined primary and alternate assemblies achieve an estimated QV of 65.6. The *k*-mer completeness is 87.39% for the primary assembly, 73.18% for the alternate haplotype, and 99.52% for the combined assemblies. BUSCO v.5.3.2 analysis using the lepidoptera_odb10 reference set ($n = 5\,286$) identified 98.3% of the expected gene set (single = 97.6%, duplicated = 0.8%).

Genome annotation report

The *A. cristana* genome assembly (GCA_948252455.1) was annotated using the Ensembl Genebuild annotation pipeline. The resulting annotation includes 22,144 transcribed mRNAs from 12,598 protein-coding and 1,631 non-coding genes. The average transcript length is 16,287.92 bp, with an average of 1.56 coding transcripts per gene and 7.23 exons per transcript. Further details of this annotation are available from the [Ensembl annotation page](#).

Methods

Sample acquisition and nucleic acid extraction

Two *Acleris cristana* specimens (specimen number Ox000993 and Ox000832, individuals ilAclCris2 and ilAclCris1) were collected from Wytham Woods, Oxfordshire (biological vice-county Berkshire), UK (latitude 51.77, longitude -1.34) on 2020-11-21 and 2020-08-01 respectively. The specimens were taken from the woodland habitat by Douglas Boyes (University of Oxford) using a light trap. The specimens were identified by the collector and snap-frozen on dry ice.

The sample was prepared for DNA extraction in the Tree of Life laboratory, Wellcome Sanger Institute (WSI). The ilAclCris2 specimen was weighed and dissected on dry ice. Whole organism tissue was disrupted using a Nippi Powermasher fitted with a BioMasher pestle. DNA was extracted at the WSI Scientific Operations core using the Qiagen MagAttract HMW DNA kit, according to the manufacturer's instructions.

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 Pacific Biosciences SEQUEL II (HiFi) instrument. Hi-C data were also generated from whole organism tissue of ilAclCris1 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 as described previously (Howe *et al.*, 2021). Manual curation was performed using HiGlass (Kerpedjiev *et al.*, 2018) and Pretext (Harry, 2022). The curation process is described at <https://github.com/sanger-tol/curation-resources>. PretextSnapshot was used to generate a Hi-C contact map of the final assembly.

The mitochondrial genome was assembled using MitoHiFi (Uliano-Silva *et al.*, 2022), which runs MitoFinder (Allio *et al.*, 2020) 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, the *k*-mer completeness and QV consensus quality values were calculated in MerquryFK (Rhie *et al.*, 2020).

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. In the BlobToolKit pipeline, PacBio reads were aligned to the assembly using minimap2 (Li, 2018) and processed with SAMtools (Danecek *et al.*, 2021) to generate coverage tracks. The pipeline runs BUSCO (Manni *et al.*, 2021) using lineages identified from the NCBI Taxonomy (Schoch *et al.*, 2020). For the three basal BUSCO lineages, eukaryota, bacteria and archaea, BUSCO genes are aligned to the UniProt Reference Proteomes database (Bateman *et al.*, 2023) using DIAMOND BLASTp (Buchfink *et al.*, 2021). The genome assembly is divided into chunks based on the density of BUSCO genes from the closest taxonomic lineage, and each chunk is aligned to the UniProt Reference Proteomes database using DIAMOND BLASTx. Sequences without DIAMOND BLASTx hits are extracted using seqtk and aligned to the NT database using BLASTn (Altschul *et al.*, 1990). The outputs are consolidated into a BlobDir for visualisation in BlobToolKit.

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

Table 3. Software tools: versions and sources.

Software tool	Version	Source
BLAST	2.14.0	ftp://ftp.ncbi.nlm.nih.gov/blast/executables/blast/
BlobToolKit	4.0.7	https://github.com/blobtoolkit/blobtoolkit
BUSCO	5.3.2	https://gitlab.com/ezlab/busco
bwa-mem2	2.2.1	https://github.com/bwa-mem2/bwa-mem2

Table 3. *Continued*

Software tool	Version	Source
DIAMOND	2.0.15	https://github.com/bbuchfink/diamond
GenomeScope2.0	2.0.1	https://github.com/tbenavi1/genomescope2.0
Hifiasm	0.16.1-r375	https://github.com/chhylp123/hifiasm
HiGlass	1.11.6	https://github.com/higlass/higlass
MerquryFK	1.1.0-c1	https://github.com/thegenemyers/MERQURY.FK
MitoHiFi	2	https://github.com/marcelauliano/MitoHiFi
PretextView	0.25	https://github.com/wtsi-hpag/PretextView
PretextSnapshot	0.0.6	https://github.com/sanger-tol/PretextSnapshot
purge_dups	1.2.3	https://github.com/dfguan/purge_dups
samtools	1.15.1	https://github.com/samtools/samtools
YaHS	1.2a	https://github.com/c-zhou/yahs

Genome annotation

The Ensembl gene annotation system (Aken *et al.*, 2016) was used to generate annotation for the *Acleris cristana* assembly (GCA_948252455.1) at the European Bioinformatics Institute. Annotation was created primarily through alignment of transcriptomic data to the genome, 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: *Acleris cristana* (tufted button). Accession number PRJEB58659; <https://identifiers.org/ena.embl/PRJEB58659>. (Wellcome Sanger Institute, 2023).

The genome sequence is released openly for reuse. The *Acleris cristana* 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](#)
- 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](#)
- Altschul SF, Gish W, Miller W, *et al.*: **Basic Local Alignment Search Tool**. *J Mol Biol*. 1990; **215**(3): 403–410.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Bateman A, Martin M-J, Orchard S, *et al.*: **UniProt: The Universal Protein Knowledgebase in 2023**. *Nucleic Acids Res* 2023; **51**(D1): D523–D531.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Buchfink B, Reuter K, Drost H-G: **Sensitive protein alignments at tree-of-life scale using DIAMOND**. *Nature Methods*. 2021; **18**(4): 366–368.
[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](#)
- Danecek P, Bonfield JK, Liddle J, *et al.*: **Twelve years of SAMtools and BCFtools**. *GigaScience*. 2021; **10**(2): giab008.
[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](#)
- Emmett AM: *A field guide to the smaller British Lepidoptera*. 2nd ed. Reading: British Entomological and Natural History Society; 2010.
- GBIF Secretariat: **Acleris cristana (Denis & Schiffermüller, 1775), GBIF Backbone Taxonomy**. 2023; (Accessed: 9 February 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](#)
- Hancock F, Bland KP, Razowski J: *The Moths and Butterflies of Great Britain and Ireland*. Volume 5 (Part 2). Leiden: Brill; 2015.
- 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**: 30143029.
[PubMed Abstract](#) | [Free Full Text](#)
- Li H: **Minimap2: Pairwise alignment for nucleotide sequences**. *Bioinformatics*. 2018; **34**(18): 3094–3100.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free 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 K, *et al.*: **Mercury: Reference-free quality, completeness, and phasing assessment for genome assemblies**. *Genome Biol*. 2020; **21**.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Schoch CL, Ciufo S, Domrachev M, *et al.*: **NCBI taxonomy: A comprehensive update on curation, resources and tools**. *Database*. 2020; **2020**: baaa062.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Šichová J, Nguyen P, Daliková M, *et al.*: **Chromosomal Evolution in Tortricid Moths: Conserved Karyotypes with Diverged Features**. *PLoS One*. 2013; **8**.
[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](#)
- Sterling P, Parsons M: *Field Guide to the Micro-moths of Great Britain and Ireland*. London: Bloomsbury; 2023.
- Uliano-Silva M, Ferreira JGRN, Krashenninnikova K, *et al.*: **MitoHiFi: a python pipeline for mitochondrial genome assembly from PacBio High Fidelity reads**. *bioRxiv*. [Preprint]. 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 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 Tufted Buton, *Acleris cristana* (Denis & Schiffermüller, 1775)**. European Nucleotide Archive, [dataset]. accession number PRJEB58659, 2023.
- Zhou C, McCarthy SA, Durbin R: **YaHS: yet another Hi-C scaffolding tool**. *Bioinformatics*. Edited by Alkan C, 2023; **39**(1): btac808.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)

Open Peer Review

Current Peer Review Status:  

Version 2

Reviewer Report 02 September 2026

<https://doi.org/10.21956/wellcomeopenres.30151.r165884>

© 2026 Schoeneberg Y. This is an open access peer review report distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.



Yannis Schoeneberg 

Trier University, Trier, Germany

In their revision, the authors have significantly improved the Methods section and addressed all of my previously raised concerns. I would like to commend the authors on the outstanding quality of their genome assembly, which will provide a valuable resource for future studies. I have no further comments.

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: genome assembly; comparative genomics; phylogenomics

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

Version 1

Reviewer Report 13 May 2024

<https://doi.org/10.21956/wellcomeopenres.21610.r82252>

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



Yannis Schoeneberg 

Trier University, Trier, Germany

In this article the authors describe how they sequenced and assembled the genome of *Acleris*

cristana, a moth of the family Tortricidae. The genome was assembled to chromosome level, annotated and the sex chromosomes were identified.

The authors assembled a high quality genome assembly, which will be an important resource for the study of this species and beyond. The assembly strategy is appropriate, and the quality metrics indicate an assembly of outstanding quality. However, I found the methods to be unreproducible, as the section contained too little detail and important information was absent. Apart from that, I only have minor suggestions and would like to congratulate the authors on their genome assembly of exceptional quality.

Best wishes
Yannis Schöneberg

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

No - The method section seems to be incomplete and often lacks detail necessary to make this publication reproducible. I advise the authors to thoroughly check that all steps were included in the methods (e.g. BlobtoolKit lacks a description how the input data was generated, the identification of sex chromosomes is missing entirely). Furthermore, a very brief explanation why each step was performed would greatly enhance the understandability for readers not being experts in genome assembly.

Detailed Comments:

Background:

- I think it would be necessary to cite a reference for the statement that there are several genes influencing the colours of individual pattern elements.

Genome sequence report:

- I assume it is the case, but it would be nice for the reader if the authors specifically mention that the number of chromosome-level scaffolds matches the expected karyotype.
- I'm not sure what is meant with "not fully phased". This should be described in more detail.

Genome annotation report

- The rapid annotation pipeline seems to be a standardized pipeline. However, that doesn't guarantee it generates reliable annotations. I haven't tried this pipeline, but to my experience many standardized pipelines (e.g. NCBI RefSeq) perform poorly on arthropod genomes, as they are usually designed for vertebrates. I suggest providing the reader with quality metrics for the annotation, such as BUSCO score, or Annotation Edit Distance (AED), or else. Otherwise, the annotation's quality is unclear.

Methods:

In general, I think the methods should have more detail and the authors should thoroughly check that each step is properly described (e.g. sex chromosome identification is missing). I think that short explanations why a certain step was conducted could help the reader understand the workflow. As a couple of programs and work steps were missing from the text and table 3, I advise the authors to check whether all relevant programs and work steps are listed.

Genome assembly, curation and evaluation:

- Hifiasm can be run with HIFI data alone or combined with the Hi-C data. It would be good to briefly mention which data was used.
- This might be nit picking, but purge_dups is also part of the Hifiasm workflow. For clarity, I think it should be mentioned whether it was run as part of Hifiasm, standalone, or both.
- As much as I understand, Pretext is not a single program but a prefix used for several programs (e.g. PretextView or PretextMap). These programs seem to be bundled in the PretextSuite, but for clarity I think it would be best to mention and cite each program individually and briefly describe what it was used for.
- A precise description of how the HiC contact map was generated is missing. As I'm not aware that bwa-mem2 outputs cooler files, I assume the authors used the cooler program suite to generate these files, and the contact maps. If this is the case, the program is also missing from table 3.
- For the mitochondrial genome assembly it should be specifically stated whether MitoFinder or MITOS was used.
- BlobtoolKit requires mapped reads and a taxonomic annotation of the scaffolds as input files. The information which programs and data were used to generate the input files is missing from the text and table 3.
- It was mentioned in the genome sequence report, but the BUSCO set used for QC of the assembly should be named again in the methods.
- The MerquyFK workflow is different and has more possibilities than the original Merquy workflow. Please describe exactly what was done and name and cite the programs utilized (e.g. FastK).
- The methods lack a description of how the sex chromosomes were identified.

Table 3:

- Depending on what was used for the mitochondrial assembly, MitoFinder or MITOS is missing from the table.
- For Merquy, I would suggest to call it MerquyFK in the Software tool column and the text, otherwise there might be confusion. I assume the version used is v1.0.0 as this is the tag of the only release on the Github repository.
- bwa-mem2 is missing from the table.
- Nextflow is missing from the table.
- Ensembl Annotation Pipeline is missing from the table.
- As much as I understand the author's workflow, they used several of the programs bundled in the Pretext-suite, not just PretextView. Therefore, I think it would be best to cite each of the programs individually.

Genome annotation

- I think this section should generally be described in more detail. I suggest to include a very short description of all work steps and citation of the programs utilized by the pipeline.
- It's unclear which training/reference data was used for the annotation. I suggest giving accessions for the transcriptomic and protein dataset if possible.

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?

No

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

Yes

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: genome assembly; comparative genomics; phylogenomics

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

Reviewer Report 10 October 2023

<https://doi.org/10.21956/wellcomeopenres.21610.r67857>

© 2023 Zhang F. 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.



Feng Zhang 

Department of Entomology, College of Plant Protection, Nanjing Agricultural University, Nanjing, China

The authors describe the first genome assembly of the Tufted Button *Acleris cristana*, a common but poorly genetic studied moth distributes in Europe. Combining HiFi long reads and Hi-C data from an adult female moth, the authors established a chromosomal-level assembly of ~562.6Mb. The genome contains 29 autosomes and two sex chromosomes (Z and W). Several metrics suggest that genome has a high completeness and contiguity.

There is a discrepancy in the paper as to whether this represents a single or two female samples (line 1 in Methods says two females were collected, the Genome sequence report says one female for sequencing). The authors need to clarify the reason why they only used a single female instead of two.

In Genome sequence report, the authors mentioned W and Z chromosomes are large without giving size of sex chromosomes. It could be clearer if the authors provide size information for W and Z chromosomes to explain why they can be classified as "large chromosome".

Minor revisions:

1. In Background, delete "the" in front of "family Tortricidae".

2. In Background, change “The species” to “This species”.

3. In Methods, add the name of “collector”.

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Insect genomics, Phylogenomics

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