




DATA NOTE

The genome sequence of the Downland Villa bee-fly, *Villa cingulata* (Meigen, 1804) [version 1; peer review: 2 approved, 1 approved with reservations]

Susan C. Taylor¹, Liam M. Crowley ², Sally Luker³, Martin Harvey⁴,
Natural History Museum Genome Acquisition Lab,
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

¹Dipterists Forum, Northchurch, England, UK

²University of Oxford, Oxford, England, UK

³University of Exeter, Penryn, England, UK

⁴UK Centre for Ecology & Hydrology, Wallingford, England, UK

v1 First published: 14 Nov 2023, 8:526
<https://doi.org/10.12688/wellcomeopenres.20123.1>

Latest published: 14 Nov 2023, 8:526
<https://doi.org/10.12688/wellcomeopenres.20123.1>

Abstract

We present a genome assembly from an individual male *Villa cingulata* (the Downland Villa bee-fly; Arthropoda; Insecta; Diptera; Bombyliidae). The genome sequence is 412.6 megabases in span. Most of the assembly is scaffolded into 10 chromosomal pseudomolecules, including the X and Y sex chromosomes. The mitochondrial genome has also been assembled and is 22.43 kilobases in length.

Keywords




Villa cingulata, Downland Villa Bee-fly, genome sequence, chromosomal, Diptera




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


Open Peer Review

Approval Status   

	1	2	3
version 1 14 Nov 2023	 view	 view	 view

1. **Yiyuan Li** , Ningbo University, Ningbo, China

2. **Christophe Klopp** , INRAE, Castanet Tolosan, France

3. **Luciano Cosme** , Yale University, New Haven, USA

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: **Taylor SC:** Investigation, Resources, Writing – Original Draft Preparation, Writing – Review & Editing; **Crowley LM:** Investigation, Resources; **Luker S:** Writing – Original Draft Preparation; **Harvey M:** 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) and the Darwin Tree of Life Discretionary Award (218328).

The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Copyright: © 2023 Taylor SC *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: Taylor SC, Crowley LM, Luker S *et al.* **The genome sequence of the Downland Villa bee-fly, *Villa cingulata* (Meigen, 1804) [version 1; peer review: 2 approved, 1 approved with reservations]** Wellcome Open Research 2023, 8:526 <https://doi.org/10.12688/wellcomeopenres.20123.1>

First published: 14 Nov 2023, 8:526 <https://doi.org/10.12688/wellcomeopenres.20123.1>

Species taxonomy

Eukaryota; Metazoa; Eumetazoa; Bilateria; Protostomia; Ecdysozoa; Panarthropoda; Arthropoda; Mandibulata; Pancrustacea; Hexapoda; Insecta; Dicondylia; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Asiloidea; Bombyliidae; Anthracinae; Villini; *Villa*; *Villa cingulata* (Meigen, 1804) (NCBI:txid2753613).

Background

The genus *Villa* Lioy 1864 comprises more than 270 described species globally (Evenhuis, 1999; Evenhuis & Greathead, 2015; GBIF Secretariat, 2022; Hull, 1973) of short-tongued bee-flies (Diptera: Bombyliidae). *Villa* spp. are considered parasites of Lepidopteran species, notably of Noctuid moths (Lepidoptera, Noctuidae). As with other bee-flies, while in flight, females flick single, dust-covered eggs in the direction of their hosts, which then develop inside the larvae/pupae (Falck, 2009; Stubbs & Drake, 2014). *Villa cingulata* (Meigen, 1804) is commonly known as the Downland Villa; it is one of three *Villa* species known from the UK, the other two being *V. modesta* (Meigen, 1820) and *V. venusta* (Meigen, 1820). *V. venusta* is presumed extinct in the UK (Drake, 2017).

A photographic identification guide to the UK *Villa* spp. has been compiled for the Soldierflies and Allies Recording Scheme (Harvey, 2022), and a detailed description and key is provided in Stubbs and Drake (2014). All three UK species are medium-to-large flies with a characteristic overall appearance of a fairly broad, flat abdomen, clear wings with dark bases, and varying degrees and arrangements of golden hairs and dark scales (Harvey, 2022; Stubbs & Drake, 2014). Differences in abdominal banding and extent of darkened wings are particularly useful in separating the species, with all three exhibiting varying degrees of sexual dimorphism.

While the three UK *Villa* species are very similar in appearance, clear differences in the associated habitats and current distributions of each are currently observed, with *V. cingulata* showing a preference for inland grasslands. Until quite recently *V. cingulata* was very predominantly an insect of calcareous meadows and chalk downland with short turf and/or minimal vegetation cover (Harvey, 2022; Stubbs & Drake, 2014). However, since 2013 it has also been found in neutral hay meadow habitats (e.g. Chandler, 2015) and in the last few years has also colonised some acid grasslands, e.g. in the New Forest (Soldierflies and Allies Recording Scheme, 2023). Some of these recent sites have tall grassland swards, but paths and patches of bare ground are usually present as well.

Historically recorded, albeit in very low numbers, from several southern UK counties, where it was described as occurring on shrubby, chalky slopes with tall vegetation (Drake, 2017; Stubbs & Drake, 2014; Verrall, 1909). A period of about 60 years with no records of *V. cingulata* led to concerns that the species might be extinct in the UK; however, in 2000, the species was rediscovered in the Cotswolds, with additional records from nearby areas

subsequently reported (Stubbs & Drake (2014) and references therein). The threat status for *V. cingulata* was assessed as “Least Concern” by Drake (2017), and at that time it was given a rarity status of Nationally Rare. However, that assessment was based largely on data prior to 2016, and since then the number of records per year has greatly increased, as has the number of occupied ten-kilometre squares, and the species would no longer qualify as Nationally Rare based on current data. Since 2000 it has been recorded from 21 vice-counties in southern England and south-east Wales (Soldierflies and Allies Recording Scheme, 2023).

Adults of *V. cingulata* are found in summer, from June to August (Harvey, 2022; Stubbs & Drake, 2014). Adult flies can be observed visiting umbellifers (Fam. Apiaceae), such as Wild Parsnip and Hogweed, while females will often rest on warm ground, gathering dust on their abdomens, which subsequently adheres to their eggs to enable effective dispersal via flicking (Falck, 2009). The host of *V. cingulata* in the UK remains unknown (e.g. Stubbs & Drake, 2014). Males have been observed to be territorial/defensive to conspecific males (Falck, 2009).

The specimen was collected in July 2021. On the continent *V. cingulata* is found throughout Europe and the Middle East (Evenhuis & Greathead, 2015).

The genome of *V. cingulata* 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 *Villa cingulata*, based on one male specimen from Holtspur Bottom Nature Reserve, UK.

Genome sequence report

The genome was sequenced from one male *Villa cingulata* (Figure 1) collected from Holtspur Bottom Nature reserve, England, UK (51.61, -0.67). A total of 54-fold coverage



Figure 1. Photograph of the *Villa cingulata* (idVilCing2) specimen used for genome sequencing.

in Pacific Biosciences single-molecule HiFi long reads was generated. Primary assembly contigs were scaffolded with chromosome conformation Hi-C data. Manual assembly curation removed contaminating scaffolds and corrected 5 missing joins or mis-joins and removed one haplotypic duplication, reducing the assembly length by 47.54% and the scaffold number by 99.35%, and increasing the scaffold N50 by 221.12%.

The final assembly has a total length of 412.6 Mb in 45 sequence scaffolds with a scaffold N50 of 45.3 Mb (Table 1).

The snailplot in Figure 2 represents a summary of the assembly statistics, while the distribution of assembly scaffolds on GC proportion and coverage is shown in Figure 3. The cumulative assembly plot in Figure 4 shows curves for subsets of scaffolds assigned to different phyla. Most (99.29%) of the assembly sequence was assigned to 10 chromosomal-level scaffolds, representing 8 autosomes and the X and Y sex chromosome. Chromosome-scale scaffolds confirmed by the Hi-C data are named in order of size (Figure 5, Table 2). The Y chromosome is not scaffolded due

Table 1. Genome data for *Villa cingulata*, idVilCing2.1.

Project accession data		
Assembly identifier	idVilCing2.1	
Species	<i>Villa cingulata</i>	
Specimen	idVilCing2	
NCBI taxonomy ID	2753613	
BioProject	PRJEB57894	
BioSample ID	SAMEA14448141	
Isolate information	idVilCing2, male: thorax (DNA sequencing) idVilCing1, female: head and thorax (Hi-C scaffolding)	
Assembly metrics*		Benchmark
Consensus quality (QV)	66.9	≥ 50
k-mer completeness	100%	≥ 95%
BUSCO**	C:94.5%[S:93.5%,D:1.0%], F:1.1%,M:4.4%,n:3,285	C ≥ 95%
Percentage of assembly mapped to chromosomes	99.29%	≥ 95%
Sex chromosomes	X and Y chromosomes	localised homologous pairs
Organelles	Mitochondrial genome assembled	complete single alleles
Raw data accessions		
PacificBiosciences SEQUEL II	ERR10662019	
Hi-C Illumina	ERR10614877	
Genome assembly		
Assembly accession	GCA_951394055.1	
Accession of alternate haplotype	GCA_951394075.1	
Span (Mb)	412.6	
Number of contigs	152	
Contig N50 length (Mb)	5.8	
Number of scaffolds	45	
Scaffold N50 length (Mb)	45.3	
Longest scaffold (Mb)	65.2	

* 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 diptera_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/Villa%20cingulata/dataset/CATOAS01/busco>.

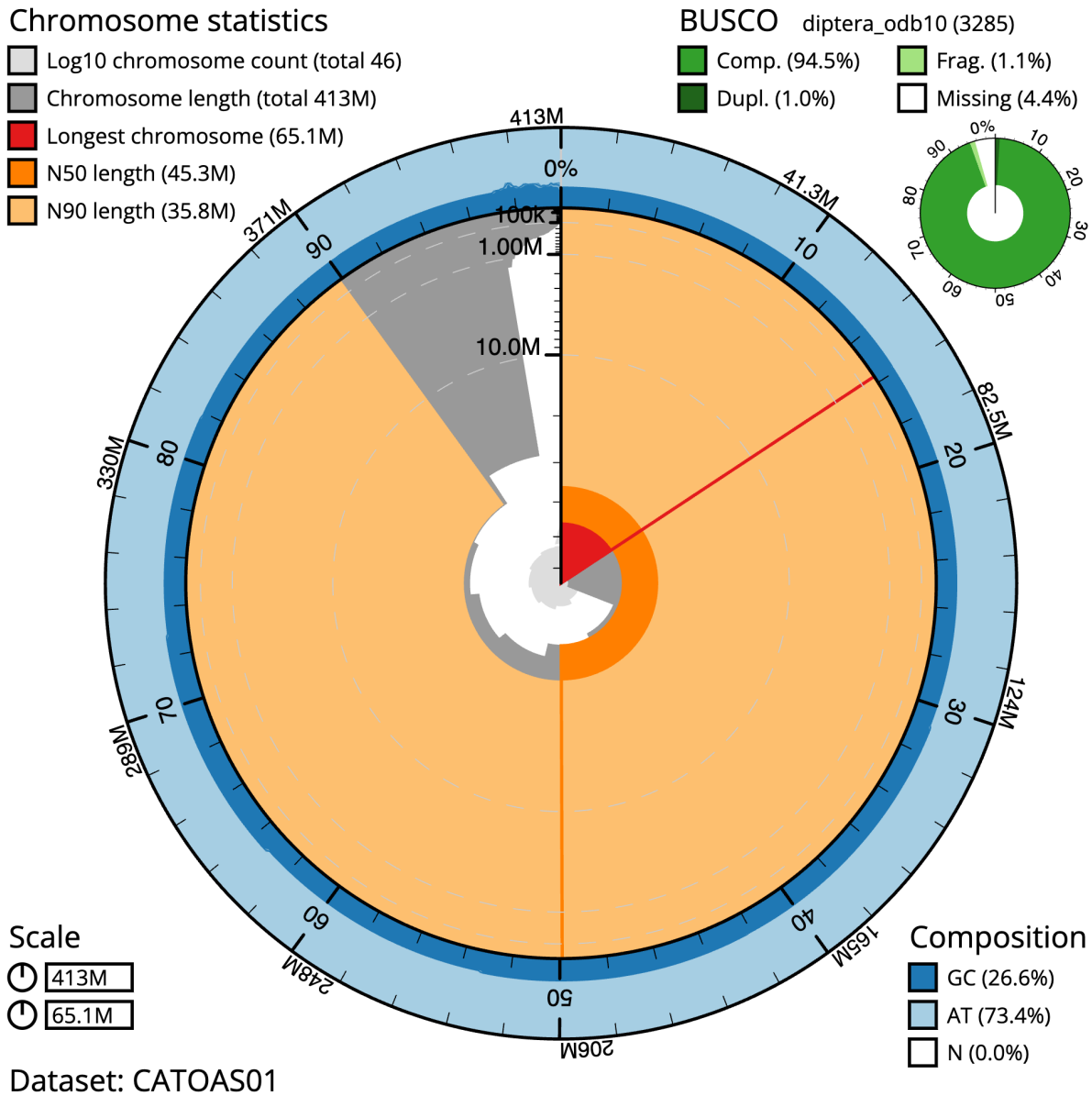


Figure 2. Genome assembly of *Villa cingulata*, idVilCing2.1: metrics. The BlobToolKit Snailplot shows N50 metrics and BUSCO gene completeness. The main plot is divided into 1,000 size-ordered bins around the circumference with each bin representing 0.1% of the 412,610,433 bp assembly. The distribution of scaffold lengths is shown in dark grey with the plot radius scaled to the longest scaffold present in the assembly (65,149,795 bp, shown in red). Orange and pale-orange arcs show the N50 and N90 scaffold lengths (45,252,737 and 35,763,846 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 diptera_odb10 set is shown in the top right. An interactive version of this figure is available at <https://blobtoolkit.genomehubs.org/view/idVilCing2.1/dataset/CATOAS01/snail>.

to lack of Hi-C signal, as the Hi-C data appears to be from a female sample whereas the PacBio data used for *de novo* assembly is from a male. (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.

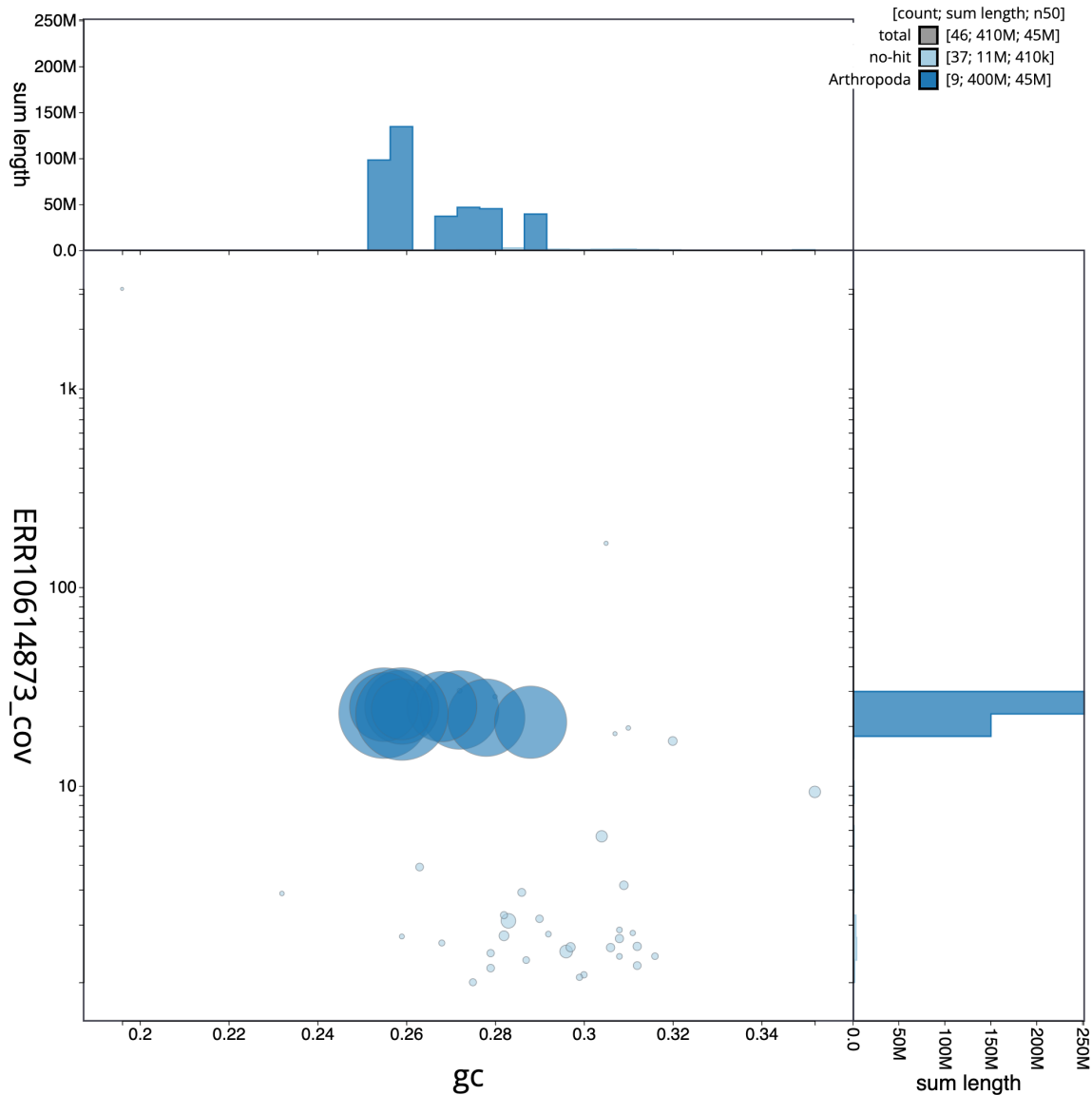


Figure 3. Genome assembly of *Villa cingulata*, idVilCing2.1: BlobToolKit GC-coverage plot. Scaffolds are coloured by phylum. Circles are sized in proportion to scaffold length. Histograms show the distribution of scaffold length sum along each axis. An interactive version of this figure is available at <https://blobtoolkit.genomehubs.org/view/idVilCing2.1/dataset/CATOAS01/blob>.

The estimated Quality Value (QV) of the final assembly is 66.9 with k -mer completeness of 100%, and the assembly has a BUSCO v5.3.2 completeness of 94.5% (single = 93.5%, duplicated = 1.0%), using the diptera_odb10 reference set ($n = 3,285$).

Metadata for specimens, spectral estimates, sequencing runs, contaminants and pre-curation assembly statistics can be found at <https://links.tol.sanger.ac.uk/species/2753613>.

Methods

Sample acquisition and nucleic acid extraction

A male *Villa cingulata* (specimen ID NHMUK014543879, individual idVilCing2) was netted in Holtspur Bottom

Nature Reserve, England, UK (latitude 51.61, longitude -0.67) on 2021-02-13. The specimen was collected and identified by Sue Taylor (Dipterists Forum) and preserved by freezing at -80°C . The specimen used for Hi-C sequencing (specimen ID Ox000478, ToLID idVilCing1) was netted in Wytham Woods, Oxfordshire (biological vice-county Berkshire), UK (latitude 51.77, longitude -1.34) on 2020-06-15. The specimen was collected and identified by Liam Crowley (University of Oxford) and snap-frozen on dry ice.

DNA was extracted at the Tree of Life laboratory, Wellcome Sanger Institute (WSI). The idVilCing2 sample was weighed and dissected on dry ice with tissue set aside for Hi-C sequencing. Thorax tissue was disrupted using a Nippi Powermasher fitted

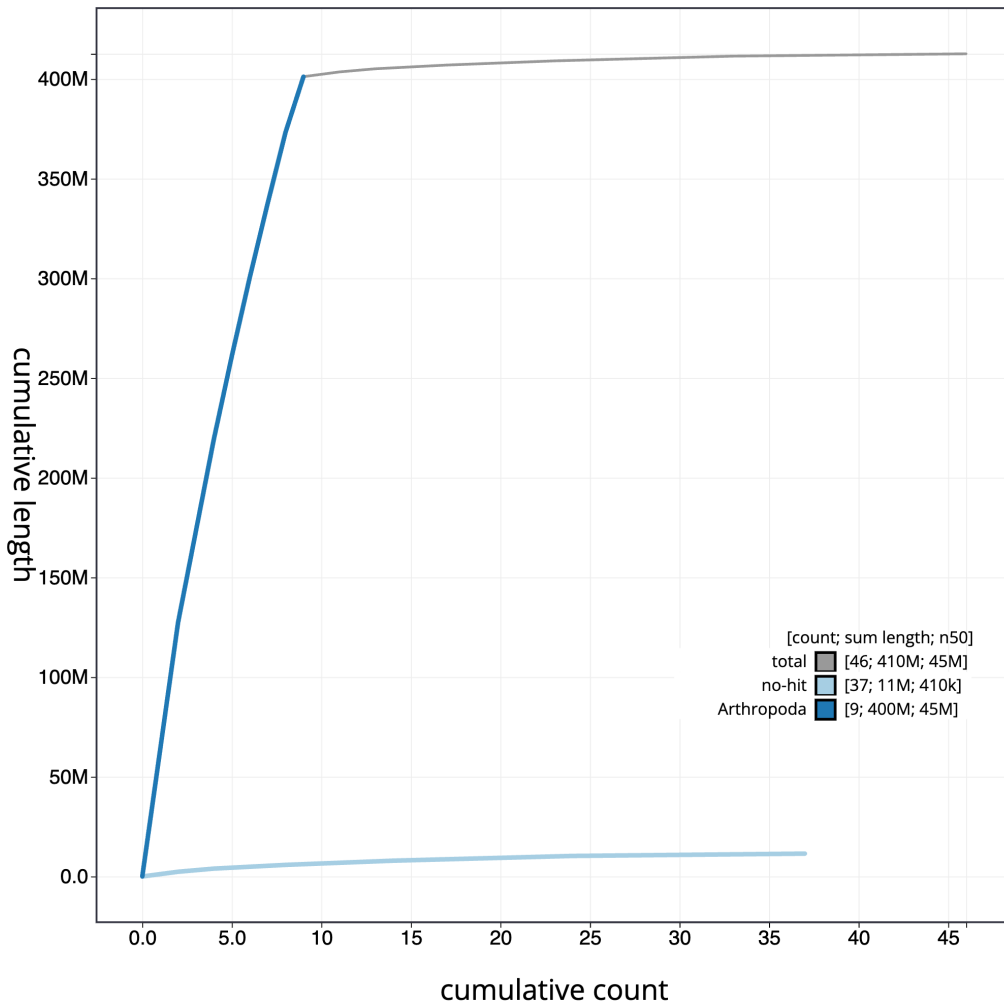


Figure 4. Genome assembly of *Villa cingulata*, idVilCing2.1: BlobToolKit cumulative sequence plot. The grey line shows cumulative length for all scaffolds. Coloured lines show cumulative lengths of scaffolds assigned to each phylum using the buscodegenes taxrule. An interactive version of this figure is available at <https://blobtoolkit.genomehubs.org/view/idVilCing2.1/dataset/CATOAS01/cumulative>.

with a BioMasher pestle. High molecular weight (HMW) DNA was extracted using the Qiagen MagAttract HMW DNA extraction kit. HMW DNA was sheared into an average fragment size of 12–20 kb in a Megaruptor 3 system with speed setting 30. Sheared DNA was purified by solid-phase reversible immobilisation using AMPure PB beads with a 1.8X ratio of beads to sample to remove the 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 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 (HiFi) instrument. Hi-C data were also generated from head and thorax tissue of idVilCing1 using the Arima2 kit and sequenced on the HiSeq X Ten instrument.

Genome assembly, curation and evaluation

Assembly was carried out with Hifiasm (Cheng *et al.*, 2021) and haplotypic duplication was identified and removed with purge_dups (Guan *et al.*, 2020). The assembly was then scaffolded with Hi-C data (Rao *et al.*, 2014) using YaHS (Zhou *et al.*, 2023). The assembly was checked for contamination and corrected as described previously (Howe *et al.*, 2021). Manual curation was performed using HiGlass (Kerpedjiev *et al.*, 2018) and 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.

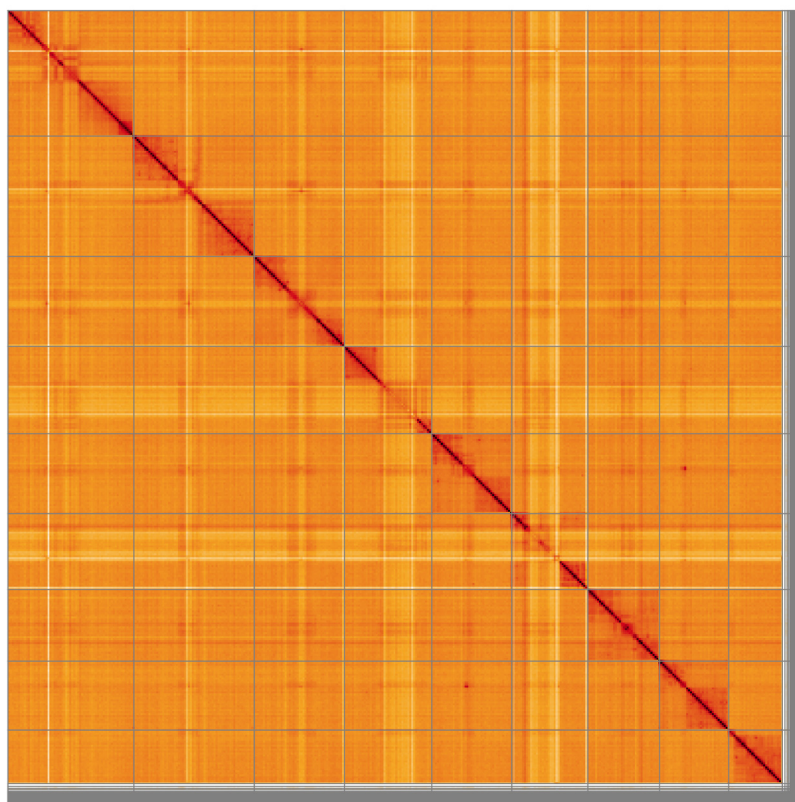


Figure 5. Genome assembly of *Villa cingulata*, idVilCing2.1: Hi-C contact map of the idVilCing2.1 assembly, visualised using HiGlass. Chromosomes are shown in order of size from left to right and top to bottom. An interactive version of this figure may be viewed at <https://genome-note-higlass.tol.sanger.ac.uk/l/?d=5XAgkaluToGo3fqjAOJUjw>.

Table 2. Chromosomal pseudomolecules in the genome assembly of *Villa cingulata*, idVilCing2.

INSDC accession	Chromosome	Length (Mb)	GC%
OX596018.1	1	62.38	25.5
OX596019.1	2	46.73	27.0
OX596020.1	3	45.25	28.0
OX596021.1	4	41.52	26.0
OX596022.1	5	39.4	29.0
OX596023.1	6	37.12	27.0
OX596024.1	7	35.76	25.5
OX596025.1	8	27.82	26.0
OX596017.1	X	65.15	26.0
OX596026.1	Y	1.37	28.5
OX596027.1	MT	0.02	19.5

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.

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

Table 3. Software tools: versions and sources.

Software tool	Version	Source
BlobToolKit	4.1.7	https://github.com/blobtoolkit/blobtoolkit
BUSCO	5.3.2	https://gitlab.com/ezlab/busco
Hifiasm	0.16.1	https://github.com/chhylp123/hifiasm
HiGlass	1.11.6	https://github.com/higlass/higlass
Mercury	MercuryFK	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	1.1a.2	https://github.com/c-zhou/yahs

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: *Villa cingulata*. Accession number PRJEB57894; <https://identifiers.org/ena.embl/PRJEB57894>. (Wellcome Sanger Institute, 2023) The genome

sequence is released openly for reuse. The *Villa cingulata* genome sequencing initiative is part of the Darwin Tree of Life (DToL) project. All raw sequence data and the assembly have been deposited in INSDC databases. The genome will be annotated using available RNA-Seq data and presented through the [Ensembl](#) pipeline at the European Bioinformatics Institute. Raw data and assembly accession identifiers are reported in [Table 1](#).

Author information

Members of the Natural History Museum Genome Acquisition Lab are listed here: <https://doi.org/10.5281/zenodo.4790042>.

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](#)
- 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](#)
- Chandler PJ: **Diptera recording at Bushy Park, Middlesex.** *Dipterists Digest.* 2015; **22**(1): 69–110.
[Reference Source](#)
- 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](#)
- Drake CM: **A review of the status of Larger Brachycera flies of Great Britain - Species Status No.29.** Natural England Commissioned Reports, Number 192. 2017.
[Reference Source](#)
- Evenhuis NL: **World catalog of bee flies (Diptera: Bombyliidae).** 1999; [Accessed 13 September 2023].
[Reference Source](#)
- Evenhuis NL, Greathead DJ: **World catalog of bee flies (Diptera: Bombyliidae).** 2015; [Accessed 13 September 2023].
- Falck M: **The Norwegian species of *Villa* Lioy 1864 (Diptera: Bombyliidae).** *Nor J Entomol.* 2009; **56**: 120–130.
[Reference Source](#)
- GBIF Secretariat: ***Villa* Lioy, 1864.** *GBIF Backbone Taxonomy - checklist dataset.* 2022; [Accessed 13 September 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](#)
- Harvey M: **Identifying soldierflies and allies: bee-flies in genus *Villa*.** 2022.
[Reference Source](#)
- Howe K, Chow W, Collins J, et al.: **Significantly improving the quality of genome assemblies through curation.** *GigaScience.* Oxford University Press, 2021; **10**(1): g1aa153.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Hull FM: **Bee flies of the world: the genera of the family Bombyliidae.** Washington: Smithsonian Institution Press, 1973.
[Publisher Full Text](#)
- Kerpedjiev P, Abdennur N, Lekschas F, et al.: **HiGlass: web-based visual exploration and analysis of genome interaction maps.** *Genome Biol.* 2018; **19**(1): 125.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Manni M, Berkeley MR, Seppy M, et al.: **BUSCO update: Novel and streamlined workflows along with broader and deeper phylogenetic coverage for scoring of eukaryotic, prokaryotic, and viral genomes.** *Mol Biol Evol.* 2021; **38**(10): 4647–4654.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Rao SSP, Huntley MH, Durand NC, et al.: **A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping.** *Cell.* 2014; **159**(7): 1665–1680.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Rhie A, McCarthy SA, Fedrigo O, et al.: **Towards complete and error-free genome assemblies of all vertebrate species.** *Nature.* 2021; **592**(7856): 737–746.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Rhie A, Walenz BP, Koren S, et al.: **Mercury: Reference-free quality, completeness, and phasing assessment for genome assemblies.** *Genome Biol.* 2020; **21**(1): 245.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- 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](#)
- Soldierflies and Allies Recording Scheme: **Recording scheme data held on iRecord.** *iRecord.org.uk.* 2023; [Accessed 13 September 2023].
- Stubbs A, Drake M: **British Soldierflies and Their Allies. 2nd Edition.** British Entomological and Natural History Society, 2014.
[Reference Source](#)
- 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, Krashenninnikova K, et al.: **MitoHiFi: a python pipeline for mitochondrial genome assembly from PacBio high fidelity reads.** *BMC Bioinformatics.* 2023; **24**(1): 288.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Vasimuddin 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](#)
- Verrall GH: **Stratiomyidae and succeeding families of the Diptera Brachycera of Great Britain** *British flies Volume 5.* London: Gurney and Jackson, 1909.
[Reference Source](#)
- Wellcome Sanger Institute: **The genome sequence of the Downland *Villa* bee-fly, *Villa cingulata* (Meigen, 1804).** European Nucleotide Archive. [dataset], accession number PRJEB57894, 2023.
- 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 06 August 2024

<https://doi.org/10.21956/wellcomeopenres.22285.r89384>

© 2024 Cosme L. 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.



Luciano Cosme 

Yale University, New Haven, CT, USA

This manuscript presents the genome assembly of the *Villa cingulata*, commonly known as the Downland Villa bee-fly. The genome sequence is 412.6 megabases, scaffolded into 10 chromosomal pseudomolecules, including the X and Y sex chromosomes. The mitochondrial genome has also been assembled, with a length of 22.43 kilobases.

Replication Concerns: The manuscript lacks sufficient detail on how to replicate the findings using the raw data. Although the authors mention the sequencing methods and tools used, they do not provide a detailed workflow or pipeline for processing the raw data to achieve the same genome assembly results.

Recommendation for Authors: Include a supplementary section or a dedicated repository that outlines a step-by-step protocol for processing the raw data, including all software parameters, commands, and scripts used. This will enable other researchers to replicate the analysis and verify the results. A detailed methods section will enhance transparency and reproducibility, which are crucial for scientific research.

Does the project cover genome annotation?

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?

Partly

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Genomics and Population genetics

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 05 August 2024

<https://doi.org/10.21956/wellcomeopenres.22285.r89387>

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



Christophe Klopp 

INRAE, Castanet Tolosan, France

The article presents read sets and a principal genome assembly for *Villa cingulata* species. The assembly contigs were built with hifiasm which is the most commonly used assembler for PacBio HiFi reads. The contigs were scaffolded using Hi-C reads with YaHS. The assembly is of high quality. The authors should check the presence of telomers at both ends of their chromosomes.

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: Transcriptome and genome assembly and annotation

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 01 August 2024

<https://doi.org/10.21956/wellcomeopenres.22285.r89382>

© 2024 Li 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.



Yiyuan Li 

Ningbo University, Ningbo, Zhejiang, China

In this paper, the authors assembled the genome of a male Downland Villa bee-fly, *Villa cingulata*. They assembled the 10 chromosomes including the X and Y chromosomes. The genome has been manually curated with high quality. The authors have also included the picture of the specimen they sequenced with detailed information on the collecting time and location. The manuscript is well written with very useful details of the species and its genome. I only have two minor comments:

1. "the three UK Villa species are very similar in appearance" Would it be possible to distinguish the three species using the COI gene? If so, as the authors already assembled and annotated its mitochondrial genome, it would be helpful to search for the COI gene on the Barcode of Life Database (<https://www.boldsystems.org/>) or NCBI NT database (<https://blast.ncbi.nlm.nih.gov/Blast.cgi>) to confirm the taxonomy of the sequenced specimen.
2. Hifiasm has a few parameters during assembly. It would be helpful to also include the parameters they use for assembling. For example, did the authors use --primary and HiC options (-h1 and -h2) for their assembly?

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

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.
