DATA NOTE

The genome sequence of the 24-spot ladybird, *Subcoccinella vigintiquattuorpunctata* (Linnaeus, 1758) (Coleoptera: Coccinellidae)

[version 1; peer review: 3 approved]

Liam M. Crowley[1], Maxwell V. L. Barclay[2], Matthew N. Smith[3], Peter M. J. Brown[4], Helen E. Roy[5,6],
University of Oxford and Wytham Woods Genome Acquisition Lab,
Natural History Museum Genome Acquisition Lab,
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]University of Oxford, Oxford, England, UK
[2]Natural History Museum, London, England, UK
[3]Independent researcher, Winnersh, England, UK
[4]Applied Ecology Research Group, School of Life Sciences, Anglia Ruskin University Faculty of Arts Humanities and Social Sciences, Cambridge, England, UK
[5]UK Centre for Ecology & Hydrology, Crowmarsh Gifford, England, UK
[6]University of Exeter Centre for Ecology and Conservation, Penryn, England, UK

v1 **First published:** 23 Dec 2025, **10**:698
https://doi.org/10.12688/wellcomeopenres.25409.1
Latest published: 23 Dec 2025, **10**:698
https://doi.org/10.12688/wellcomeopenres.25409.1

Abstract

We present a genome assembly from an individual female *Subcoccinella vigintiquattuorpunctata* (24-spot ladybird; Arthropoda; Insecta; Coleoptera; Coccinellidae). The genome sequence has a total length of 532.03 megabases. Most of the assembly (97.41%) is scaffolded into 15 chromosomal pseudomolecules, including the X sex chromosome. The mitochondrial genome has also been assembled, with a length of 18.91 kilobases. This assembly was generated as part of the Darwin Tree of Life project, which produces reference genomes for eukaryotic species found in Britain and Ireland.

Open Peer Review

Approval Status ✓ ✓ ✓

	1	2	3
version 1 23 Dec 2025	✓ view	✓ view	✓ view

1. **Panagiotis Ioannidis**, Foundation for Research & Technology - Hellas, Crete, Greece

2. **Andres Mariano Alonso**, Universidad Nacional de San Martín (CONICET-UNSAM), Chascomús, Argentina

Keywords
Subcoccinella vigintiquattuorpunctata; 24-spot ladybird; genome sequence; chromosomal; Coleoptera

This article is included in the Tree of Life gateway.

3. **Mario Stanke**, University of Greifswald, Greifswald, 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: Crowley LM: Investigation, Resources; **Barclay MVL**: Investigation, Resources; **Smith MN**: Investigation, Resources; **Brown PMJ**: Writing – Original Draft Preparation, Writing – Review & Editing; **Roy HE**: Writing – Original Draft Preparation, Writing – Review & Editing;

Competing interests: No competing interests were disclosed.

Grant information: This work was supported by Wellcome through core funding to the Wellcome Sanger Institute (220540) and the Darwin Tree of Life Discretionary Award [218328, https://doi.org/10.35802/218328]. HER is supported by the Natural Environment Research Council (NE/Y006208/1) as part of the NC-UK pro¬gram within UK Centre for Ecology & Hydrology delivering national capability.
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: Crowley LM, Barclay MVL, Smith MN *et al.* **The genome sequence of the 24-spot ladybird, *Subcoccinella vigintiquattuorpunctata* (Linnaeus, 1758) (Coleoptera: Coccinellidae) [version 1; peer review: 3 approved]** Wellcome Open Research 2025, **10**:698 https://doi.org/10.12688/wellcomeopenres.25409.1

First published: 23 Dec 2025, **10**:698 https://doi.org/10.12688/wellcomeopenres.25409.1

Species taxonomy

Eukaryota; Opisthokonta; Metazoa; Eumetazoa; Bilateria; Protostomia; Ecdysozoa; Panarthropoda; Arthropoda; Mandibulata; Pancrustacea; Hexapoda; Insecta; Dicondylia; Pterygota; Neoptera; Endopterygota; Coleoptera; Polyphaga; Cucujiformia; Coccinelloidea; Coccinellidae; Epilachninae; Epilachnini; *Subcoccinella*; *Subcoccinella vigintiquattuorpunctata* (Linnaeus, 1758) (NCBI:txid295815)

Background

The 24-spot ladybird *Subcoccinella vigintiquattuorpunctata* (Linnaeus, 1758) is a small ladybird, 3–4 mm in length, and is red with a variable number (0 to 24, most commonly 20) of black spots. Melanic individuals occur extremely rarely. Like most members of the ladybird tribe Epilachnini, the upper body is somewhat hairy. The larva of the 24-spot is somewhat unusual in appearance compared to that of most ladybird species, having a yellow-green short stubby body with thick spiny bristles and short legs. The 24-spot ladybird is one of only two herbivorous ladybird species resident in the UK and the only one regarded as native (Roy & Brown, 2018). As with most ladybirds, the larval and adult stages have similar diets. Unlike many other epilachnids, the 24-spot ladybird is not restricted to a single family of plants, and feeds on diverse plants, including grasses, campions, plantains and legumes (Majerus, 2016). It is not regarded as a plant pest in the UK, although it can cause damage to alfalfa crops and other legumes in parts of its range (Baris, 2022). At least in some populations, a high proportion of adult 24-spot ladybirds are wingless (Baldwin, 1990), thus dispersal and consequent expansion of range may be limited in comparison to many ladybird species.

The 24-spot ladybird is a species of grassland that has a southerly UK distribution and here it is often found in coastal regions (Roy & Brown, 2018). On a global scale it is very widespread across Europe and Asia and was introduced in parts of North America in the 1970s (Wheeler Jr. & Henry, 1981). Like most ladybirds in northerly regions, it is generally inactive (in adult form) during winter. However, it sometimes becomes active if winter weather conditions are mild, as it has a year-round food source, unlike most predatory ladybirds which feed on pest insects that are not in abundance during winter (Majerus, 1994). Conspicuous ladybirds often have strong chemical defences to deter predators, hence their aposematic warning colouration. The 24-spot ladybird, unusually, was formerly a species in which no alkaloids were detected, implying reduced chemical defence. However, an alkaloid acting as a deterrent to ants was later identified (Wang *et al.*, 1996).

Here we present a chromosomally complete genome sequence for *S. vigintiquattuorpunctata*, based on a specimen collected from Wytham Great Wood, United Kingdom (Figure 1).

Methods

Sample acquisition

The specimen used for genome sequencing was an adult female *Subcoccinella vigintiquattuorpunctata* (specimen ID Ox000275, ToLID icSubVigi2; Figure 1), collected from Wytham Woods, Oxfordshire, UK (latitude 51.778, longitude –1.326) on 2019-09-17. The specimen was collected and identified by Liam Crowley. A different specimen was used for Hi-C sequencing (specimen ID NHMUK014444477, ToLID icSubVigi4). It was collected from Hever Castle, England, UK (latitude 51.188, longitude 0.12) on 2020-08-27. The specimen was collected and identified by Maxwell Barclay. Another specimen was used for RNA sequencing (specimen ID NHMUK014444771, ToLID icSubVigi3). It was collected from Hartslock, England, UK (latitude 51.5113, longitude –1.1122) on 2020-08-20. The specimen was collected and identified by Matt Smith.

Figure 1. Photograph of the *Subcoccinella vigintiquattuorpunctata* (icSubVigi2) specimen used for genome sequencing.

Nucleic acid extraction

Protocols for high molecular weight (HMW) DNA extraction developed at the Wellcome Sanger Institute (WSI) Tree of Life Core Laboratory are available on protocols.io (Howard *et al.*, 2025). The icSubVigi2 sample was weighed and triaged to determine the appropriate extraction protocol. Tissue from the whole organism was homogenised by powermashing using a PowerMasher II tissue disruptor. HMW DNA was extracted using the Automated MagAttract v2 protocol. DNA was sheared into an average fragment size of 12–20 kb following the Megaruptor®3 for LI PacBio protocol. Sheared DNA was purified by automated SPRI (solid-phase reversible immobilisation). The concentration of the sheared and purified DNA was assessed using a Nanodrop spectrophotometer and Qubit Fluorometer using the Qubit dsDNA High Sensitivity Assay kit. Fragment size distribution was evaluated by running the sample on the FemtoPulse system. For this sample, the final post-shearing DNA had a Qubit concentration of 11.2 ng/μL and a yield of 1 456.00 ng.

RNA was extracted from whole organism tissue of icSubVigi3 in the Tree of Life Laboratory at the WSI using the RNA Extraction: Automated MagMax™ *mir*Vana protocol. The RNA concentration was assessed using a Nanodrop spectrophotometer and a Qubit Fluorometer using the Qubit RNA Broad-Range Assay kit. Analysis of the integrity of the RNA was done using the Agilent RNA 6000 Pico Kit and Eukaryotic Total RNA assay.

PacBio HiFi library preparation and sequencing

Library preparation and sequencing were performed at the WSI Scientific Operations core. Libraries were prepared using the SMRTbell Prep Kit 3.0 (Pacific Biosciences, California, USA), following the manufacturer's instructions. The kit includes reagents for end repair/A-tailing, adapter ligation, post-ligation SMRTbell bead clean-up, and nuclease treatment. Size selection and clean-up were performed using diluted AMPure PB beads (Pacific Biosciences). DNA concentration was quantified using a Qubit Fluorometer v4.0 (ThermoFisher Scientific) and the Qubit 1X dsDNA HS assay kit. Final library fragment size was assessed with the Agilent Femto Pulse Automated Pulsed Field CE Instrument (Agilent Technologies) using the gDNA 55 kb BAC analysis kit.

The sample was sequenced using the Sequel IIe system (Pacific Biosciences, California, USA). The concentration of the library loaded onto the Sequel IIe was in the range 40–135 pM. The SMRT link software, a PacBio web-based end-to-end workflow manager, was used to set-up and monitor the run, and to perform primary and secondary analysis of the data upon completion.

Hi-C

Sample preparation and crosslinking

The Hi-C sample was prepared from 20–50 mg of frozen whole organism tissue of the icSubVigi4 sample using the Arima-HiC v2 kit (Arima Genomics). Following the manufacturer's instructions, tissue was fixed and DNA crosslinked using TC buffer to a final formaldehyde concentration of 2%. The tissue was homogenised using the Diagnocine Power Masher-II. Crosslinked DNA was digested with a restriction enzyme master mix, biotinylated, and ligated. Clean-up was performed with SPRISelect beads before library preparation. DNA concentration was measured with the Qubit Fluorometer (Thermo Fisher Scientific) and Qubit HS Assay Kit. The biotinylation percentage was estimated using the Arima-HiC v2 QC beads.

Hi-C library preparation and sequencing

Biotinylated DNA constructs were fragmented using a Covaris E220 sonicator and size selected to 400–600 bp using SPRISelect beads. DNA was enriched with Arima-HiC v2 kit Enrichment beads. End repair, A-tailing, and adapter ligation were carried out with the NEBNext Ultra II DNA Library Prep Kit (New England Biolabs), following a modified protocol where library preparation occurs while DNA remains bound to the Enrichment beads. Library amplification was performed using KAPA HiFi HotStart mix and a custom Unique Dual Index (UDI) barcode set (Integrated DNA Technologies). Depending on sample concentration and biotinylation percentage determined at the crosslinking stage, libraries were amplified with 10–16 PCR cycles. Post-PCR clean-up was performed with SPRISelect beads. Libraries were quantified using the AccuClear Ultra High Sensitivity dsDNA Standards Assay Kit (Biotium) and a FLUOstar Omega plate reader (BMG Labtech).

Prior to sequencing, libraries were normalised to 10 ng/μL. Normalised libraries were quantified again to create equimolar and/or weighted 2.8 nM pools. Pool concentrations were checked using the Agilent 4200 TapeStation (Agilent) with High Sensitivity D500 reagents before sequencing. Sequencing was performed using paired-end 150 bp reads on the Illumina NovaSeq 6000.

RNA library preparation and sequencing

Libraries were prepared using the NEBNext® Ultra™ II Directional RNA Library Prep Kit for Illumina (New England Biolabs), following the manufacturer's instructions. Poly(A) mRNA in the total RNA solution was isolated using oligo(dT) beads, converted to cDNA, and uniquely indexed; 14 PCR cycles were performed. Libraries were size-selected to produce fragments between 100–300 bp. Libraries were quantified, normalised, pooled to a final concentration of 2.8 nM, and diluted to 150 pM for loading. Sequencing was carried out on the Illumina NovaSeq 6000, generating paired-end reads.

Genome assembly

Prior to assembly of the PacBio HiFi reads, a database of *k*-mer counts (k = 31) was generated from the filtered reads using FastK. GenomeScope2 (Ranallo-Benavidez *et al.*, 2020) was used to analyse the *k*-mer frequency distributions, providing estimates of genome size, heterozygosity, and repeat content.

The HiFi reads were assembled using Hifiasm (Cheng *et al.*, 2021) with the --primary option. The Hi-C reads (Rao *et al.*, 2014) were mapped to the primary contigs using bwa-mem2 (Vasimuddin *et al.*, 2019), and the contigs were scaffolded in YaHS (Zhou *et al.*, 2023) with the --break option for handling potential misassemblies. The scaffolded assemblies were evaluated using Gfastats (Formenti *et al.*, 2022), BUSCO (Manni *et al.*, 2021) and MERQURY.FK (Rhie *et al.*, 2020).

The mitochondrial genome was assembled using MitoHiFi (Uliano-Silva *et al.*, 2023).

Assembly curation

The assembly was decontaminated using the Assembly Screen for Cobionts and Contaminants (ASCC) pipeline. TreeVal was used to generate the flat files and maps for use in curation. Manual curation was conducted primarily in PretextView and HiGlass (Kerpedjiev *et al.*, 2018). Scaffolds were visually inspected and corrected as described by Howe *et al.* (2021). Manual corrections included 26 breaks and 114 joins. This reduced the scaffold count by 3.4% and reduced the total assembly length by 1.4%. The curation process is described at https://gitlab.com/wtsi-grit/rapid-curation. PretextSnapshot was used to generate a Hi-C contact map of the final assembly.

Assembly quality assessment

The Merqury.FK tool (Rhie *et al.*, 2020) was run in a Singularity container (Kurtzer *et al.*, 2017) to evaluate *k*-mer completeness and assembly quality for the primary and alternate haplotypes using the *k*-mer databases (k = 31) computed prior to genome assembly. The analysis outputs included assembly QV scores and completeness statistics.

The genome was analysed using the BlobToolKit pipeline, a Nextflow implementation of the earlier Snakemake version (Challis *et al.*, 2020). The pipeline aligns PacBio reads using minimap2 (Li, 2018) and SAMtools (Danecek *et al.*, 2021) to generate coverage tracks. It runs BUSCO (Manni *et al.*, 2021) using lineages identified from the NCBI Taxonomy (Schoch *et al.*, 2020). For the three domain-level lineages, BUSCO genes are aligned to the UniProt Reference Proteomes database (Bateman *et al.*, 2023) using DIAMOND blastp (Buchfink *et al.*, 2021). The genome 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 with DIAMOND blastx. Sequences without hits are chunked using seqtk and aligned to the NT database with blastn (Altschul *et al.*, 1990). The BlobToolKit suite consolidates all outputs into a blobdir for visualisation. The BlobToolKit pipeline was developed using nf-core tooling (Ewels *et al.*, 2020) and MultiQC (Ewels *et al.*, 2016), with containerisation through Docker (Merkel, 2014) and Singularity (Kurtzer *et al.*, 2017).

Genome sequence report

Sequence data

PacBio sequencing of the *Subcoccinella vigintiquattuorpunctata* specimen generated 16.66 Gb (gigabases) from 1.47 million reads, which were used to assemble the genome. GenomeScope2.0 analysis estimated the haploid genome size at 569.68 Mb, with a heterozygosity of 1.70% and repeat content of 41.05% (Figure 2). These estimates guided expectations for the assembly. Based on the estimated genome size, the sequencing data provided approximately 28× coverage. Hi-C sequencing produced 56.89 Gb from 376.78 million reads, which were used to scaffold the assembly. RNA sequencing data were also generated and are available in public sequence repositories. Table 1 summarises the specimen and sequencing details.

Assembly statistics

The primary haplotype was assembled, and contigs corresponding to an alternate haplotype were also deposited in INSDC databases. The final assembly has a total length of 532.03 Mb in 256 scaffolds, with 129 gaps, and a scaffold N50 of 40.0 Mb (Table 2).

Most of the assembly sequence (97.41%) was assigned to 15 chromosomal-level scaffolds, representing 14 autosomes and the X sex chromosome. These chromosome-level scaffolds, confirmed by Hi-C data, are named according to size (Figure 3; Table 3). There are inversions between haplotypes in the following regions: Chromosome 5 between approximatley 15.6–23.7 Mb, and Chromosome 9 between approximately 21.9–33.1 Mb.

The mitochondrial genome was also assembled (length 18.91 kb, OZ297803.1). This sequence is included as a contig in the multifasta file of the genome submission and as a standalone record.

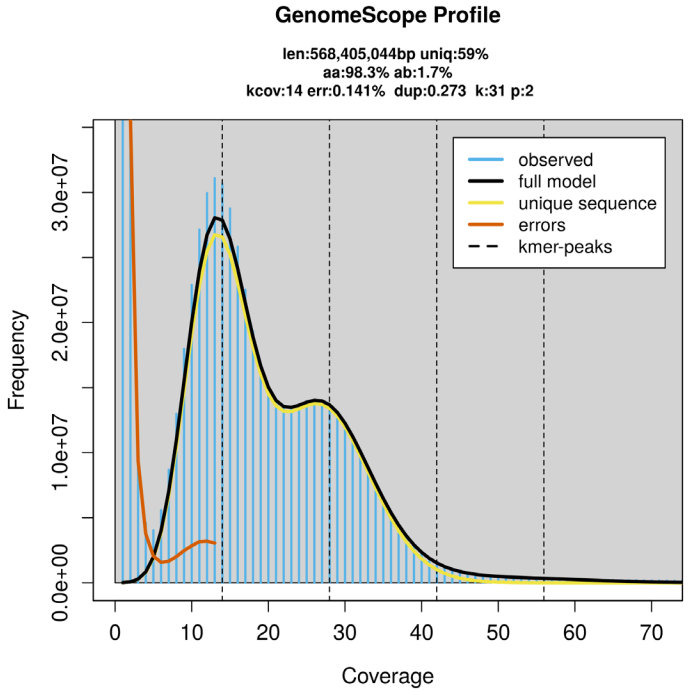

Figure 2. Frequency distribution of *k*-mers generated using GenomeScope2. The plot shows observed and modelled *k*-mer spectra, providing estimates of genome size, heterozygosity, and repeat content based on unassembled sequencing reads.

Table 1. **Specimen and sequencing data for BioProject PRJEB65189.**

Platform	PacBio HiFi	Hi-C	RNA-seq
ToLID	icSubVigi2	icSubVigi4	icSubVigi3
Specimen ID	Ox000275	NHMUK014444477	NHMUK014444771
BioSample (source individual)	SAMEA7520324	SAMEA9065876	SAMEA8534321
BioSample (tissue)	SAMEA7520396	SAMEA9065969	SAMEA8534325
Tissue	whole organism	whole organism	whole organism
Instrument	Sequel IIe	Illumina NovaSeq 6000	Illumina NovaSeq 6000
Run accessions	ERR11867194	ERR11872539	ERR11872538
Read count total	1.47 million	376.78 million	69.38 million
Base count total	16.66 Gb	56.89 Gb	10.48 Gb

Table 2. **Genome assembly statistics.**

Assembly name	icSubVigi2.1
Assembly accession	GCA_965783985.1
Alternate haplotype accession	GCA_965789425.1
Assembly level	chromosome
Span (Mb)	532.03
Number of chromosomes	15
Number of contigs	385
Contig N50	7.29 Mb
Number of scaffolds	256
Scaffold N50	40.0 Mb
Sex chromosomes	X
Organelles	Mitochondrion: 18.91 kb

Assembly quality metrics

The combined primary and alternate assemblies achieve an estimated QV of 53.3. The *k*-mer completeness is 70.76% for the primary assembly, 68.11% for the alternate haplotype, and 99.06% for the combined assemblies (Figure 4).

BUSCO v.6.0.0 analysis using the endopterygota_odb10 reference set (n = 2 124) identified 98.7% of the expected gene set (single = 97.4%, duplicated = 1.3%). The snail plot in Figure 5 summarises the scaffold length distribution and other assembly statistics for the primary assembly. The blob plot in Figure 6 shows the distribution of scaffolds by GC proportion and coverage.

Table 4 lists the assembly metric benchmarks adapted from Rhie *et al.* (2021) and the Earth BioGenome Project Report on Assembly Standards September 2024. The EBP metric, calculated for the primary assembly, is **6.C.Q53**, meeting the recommended reference standard.

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. 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

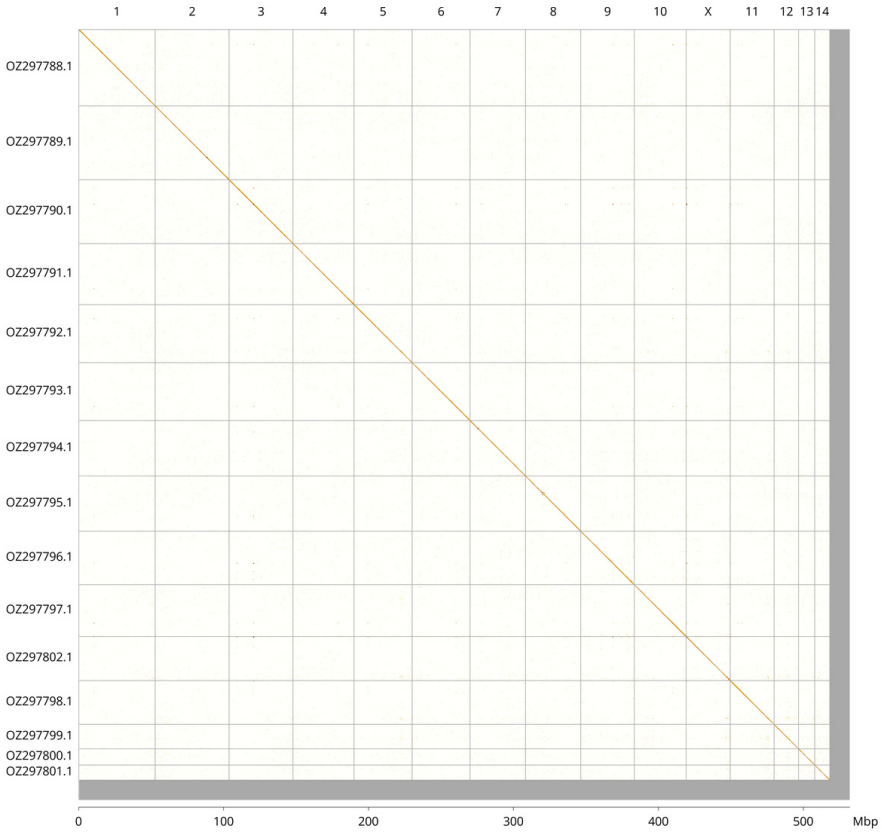

Figure 3. Hi-C contact map of the *Subcoccinella vigintiquattuorpunctata* genome assembly. Assembled chromosomes are shown in order of size and labelled along the axes, with a megabase scale shown below. The plot was generated using PretextSnapshot.

Table 3. Chromosomal pseudomolecules in the primary genome assembly of *Subcoccinella vigintiquattuorpunctata* icSubVigi2.

INSDC accession	Molecule	Length (Mb)	GC%
OZ297788.1	1	52.85	32
OZ297789.1	2	51.03	32
OZ297790.1	3	44.07	32
OZ297791.1	4	42.09	32
OZ297792.1	5	40.09	32
OZ297793.1	6	40	32
OZ297794.1	7	38.23	32.50
OZ297795.1	8	38.17	32
OZ297796.1	9	36.82	32.50
OZ297797.1	10	35.81	32.50
OZ297798.1	11	30.32	32.50
OZ297799.1	12	16.75	32.50
OZ297800.1	13	11.22	33
OZ297801.1	14	10.34	32
OZ297802.1	X	30.45	32.50

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,

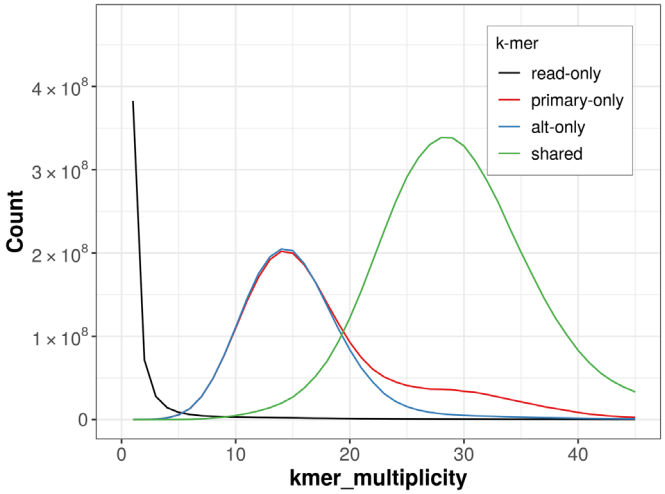

Figure 4. Evaluation of *k*-mer completeness using MerquryFK. This plot illustrates the recovery of *k*-mers from the original read data in the final assemblies. The horizontal axis represents *k*-mer multiplicity, and the vertical axis shows the number of *k*-mers. The black curve represents *k*-mers that appear in the reads but are not assembled. The green curve corresponds to *k*-mers shared by both haplotypes, and the red and blue curves show *k*-mers found only in one of the haplotypes.

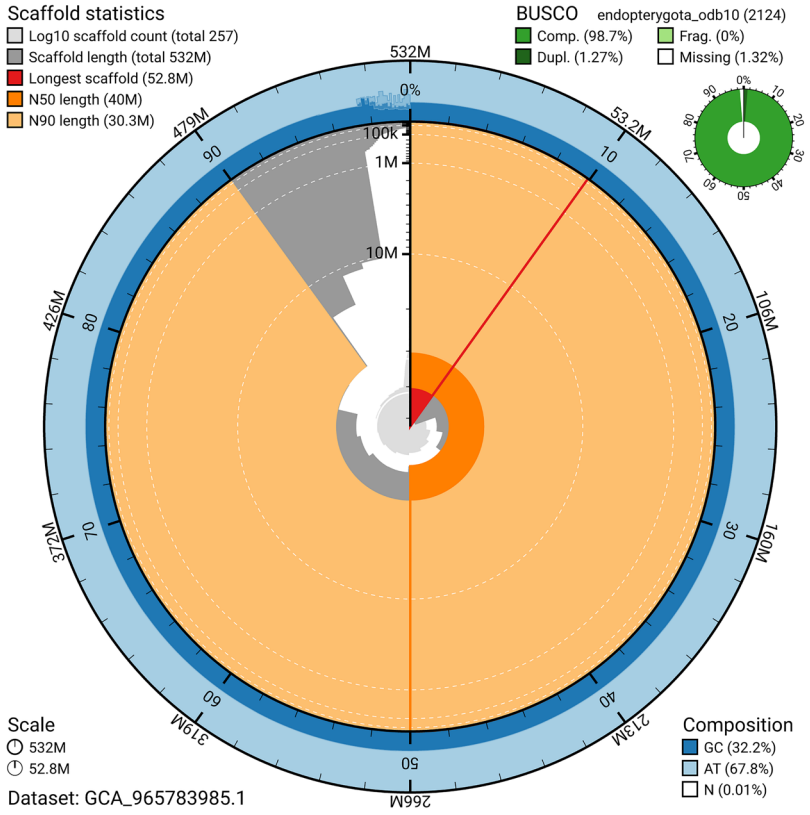

Figure 5. Assembly metrics for icSubVigi2.1. The BlobToolKit snail plot provides an overview of assembly metrics and BUSCO gene completeness. The circumference represents the length of the whole genome sequence, and the main plot is divided into 1 000 bins around the circumference. The outermost blue tracks display the distribution of GC, AT, and N percentages across the bins. Scaffolds are arranged clockwise from longest to shortest and are depicted in dark grey. The longest scaffold is indicated by the red arc, and the deeper orange and pale orange arcs represent the N50 and N90 lengths. A light grey spiral at the centre shows the cumulative scaffold count on a logarithmic scale. A summary of complete, fragmented, duplicated, and missing BUSCO genes in the endopterygota_odb10 set is presented at the top right. An interactive version of this figure can be accessed on the BlobToolKit viewer.

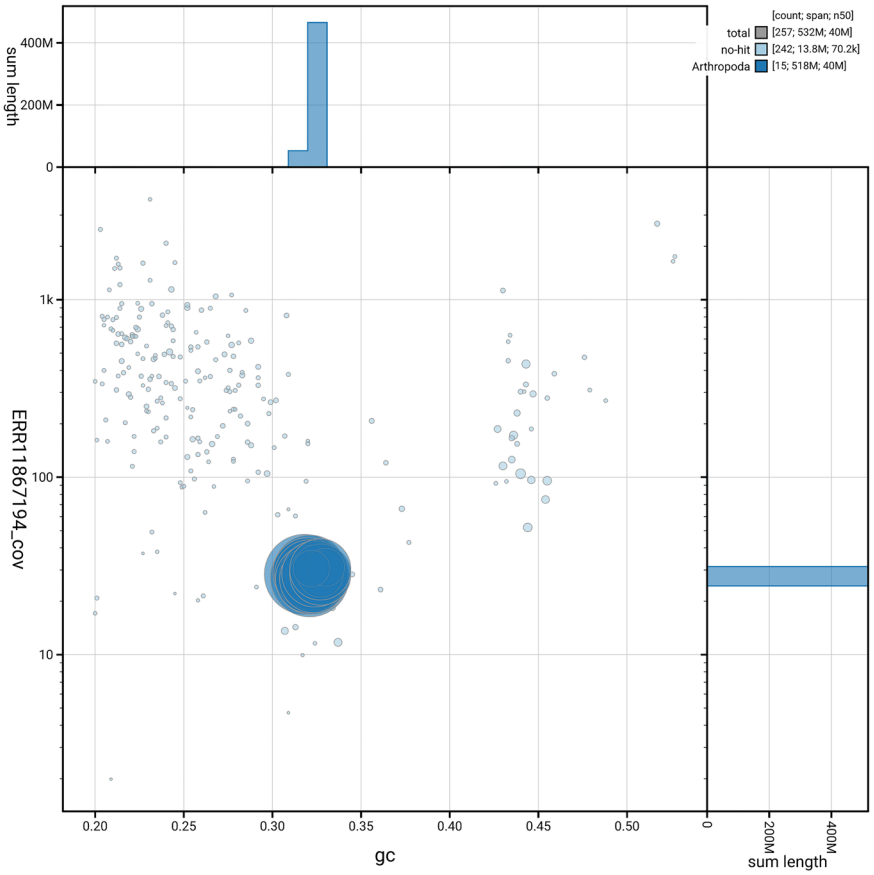

Figure 6. BlobToolKit GC-coverage plot for icSubVigi2.1. Blob plot showing sequence coverage (vertical axis) and GC content (horizontal axis). The circles represent scaffolds, with the size proportional to scaffold length and the colour representing phylum membership. The histograms along the axes display the total length of sequences distributed across different levels of coverage and GC content. An interactive version of this figure is available on the BlobToolKit viewer.

Table 4. Earth Biogenome Project summary metrics for the *Subcoccinella vigintiquattuorpunctata* assembly.

Measure	Value	Benchmark
EBP summary (primary)	6.C.Q53	6.C.Q40
Contig N50 length	7.29 Mb	≥ 1 Mb
Scaffold N50 length	40 Mb	= chromosome N50
Consensus quality (QV)	Primary: 53.7; alternate: 53.2; combined: 53.3	≥ 40
k-mer completeness	Primary: 70.76%; alternate: 68.11%; combined: 99.06%	≥ 95%
BUSCO	C:98.7% [S:97.4%; D:1.3%]; F:0.0%; M:1.3%; n:2 124	S > 90%; D < 5%
Percentage of assembly assigned to chromosomes	97.41%	≥ 90%

Genome Research Limited (operating as the Wellcome Sanger Institute), and in some circumstances, other Darwin Tree of Life collaborators.

Data availability

European Nucleotide Archive: Subcoccinella vigintiquattuorpunctata (24-pointed ladybird beetle). Accession number

PRJEB65189. The genome sequence is released openly for reuse. The *Subcoccinella vigintiquattuorpunctata* genome sequencing initiative is part of the Darwin Tree of Life Project (PRJEB40665) and Sanger Institute Tree of Life Programme (PRJEB43745). 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 and Table 2.

Production code used in genome assembly at the WSI Tree of Life is available at https://github.com/sanger-tol. Table 5 lists software versions used in this study.

Author information

Contributors are listed at the following links:

- Members of the Natural History Museum Genome Acquisition Lab
- Members of the University of Oxford and Wytham Woods Genome Acquisition Lab
- Members of the Wellcome Sanger Institute Tree of Life Management, Samples and Laboratory team
- Members of Wellcome Sanger Institute Scientific Operations – Sequencing Operations
- Members of the Wellcome Sanger Institute Tree of Life Core Informatics team
- Members of the Tree of Life Core Informatics collective
- Members of the Darwin Tree of Life Consortium

Acknowledgements

HER and PB would like to thank all the inspiring volunteer recorders who contribute their fascinating observations to the UK Ladybird Survey www.ladybird-survey.org.

Table 5. Software versions and sources.

Software	Version	Source
BEDTools	2.30.0	https://github.com/arq5x/bedtools2
BLAST	2.14.0	ftp://ftp.ncbi.nlm.nih.gov/blast/executables/blast+/
BlobToolKit	4.4.6	https://github.com/blobtoolkit/blobtoolkit
BUSCO	6.0.0	https://gitlab.com/ezlab/busco
bwa-mem2	2.2.1	https://github.com/bwa-mem2/bwa-mem2
Cooler	0.8.11	https://github.com/open2c/cooler
DIAMOND	2.1.8	https://github.com/bbuchfink/diamond
fasta_windows	0.2.4	https://github.com/tolkit/fasta_windows
FastK	1.1	https://github.com/thegenemyers/FASTK
GenomeScope2.0	2.0.1	https://github.com/tbenavi1/genomescope2.0
Gfastats	1.3.6	https://github.com/vgl-hub/gfastats
GoaT CLI	0.2.5	https://github.com/genomehubs/goat-cli
Hifiasm	0.19.8-r603	https://github.com/chhylp123/hifiasm
HiGlass	1.13.4	https://github.com/higlass/higlass
MerquryFK	1.1.2	https://github.com/thegenemyers/MERQURY.FK
Minimap2	2.28-r1209	https://github.com/lh3/minimap2
MitoHiFi	3	https://github.com/marcelauliano/MitoHiFi
MultiQC	1.14; 1.17 and 1.18	https://github.com/MultiQC/MultiQC
Nextflow	24.10.4	https://github.com/nextflow-io/nextflow
PretextSnapshot	0.0.5	https://github.com/sanger-tol/PretextSnapshot
PretextView	1.0.3	https://github.com/sanger-tol/PretextView

Software	Version	Source
samtools	1.21	https://github.com/samtools/samtools
sanger-tol/ascc	0.1.0	https://github.com/sanger-tol/ascc
sanger-tol/blobtoolkit	v0.9.0	https://github.com/sanger-tol/blobtoolkit
sanger-tol/curationpretext	1.4.2	https://github.com/sanger-tol/curationpretext
Seqtk	1.3	https://github.com/lh3/seqtk
Singularity	3.9.0	https://github.com/sylabs/singularity
TreeVal	1.4.0	https://github.com/sanger-tol/treeval
YaHS	1.2a.2	https://github.com/c-zhou/yahs

References

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

Baldwin AJ: **Further biological observations on *Subcoccinella vigintiquattuorpunctata* (L.) (Col., Coccinellidae).** *Entomologist's Monthly Magazine.* 1990; **126**(1516–1519): 223–29.
Reference Source

Baris A: **Age-stage, two-sex life tables of the 24 spot ladybird (*Subcoccinella vigintiquatuorpunctata* (Coleoptera: Coccinellidae)).** *Brazilian Archives of Biology and Technology.* 2022; **65**(1).
Publisher Full Text

Bateman A, Martin MJ, 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 HG: **Sensitive protein alignments at Tree-of-Life scale using DIAMOND.** *Nat 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

Ewels P, Magnusson M, Lundin S, *et al.*: **MultiQC: summarize analysis results for multiple tools and samples in a single report.** *Bioinformatics.* 2016; **32**(19): 3047–3048.
PubMed Abstract | Publisher Full Text | Free Full Text

Ewels PA, Peltzer A, Fillinger S, *et al.*: **The nf-core framework for community-curated bioinformatics pipelines.** *Nat Biotechnol.* 2020; **38**(3): 276–278.
PubMed Abstract | Publisher Full Text

Formenti G, Abueg L, Brajuka A, *et al.*: **Gfastats: conversion, evaluation and manipulation of genome sequences using assembly graphs.** *Bioinformatics.* 2022; **38**(17): 4214–4216.
PubMed Abstract | Publisher Full Text | Free Full Text

Howard C, Denton A, Jackson B, *et al.*: **On the path to reference genomes for all biodiversity: lessons learned and laboratory protocols created in the Sanger Tree of Life core laboratory over the first 2000 species.** *bioRxiv.* 2025.
Publisher Full Text

Howe K, Chow W, Collins J, *et al.*: **Significantly improving the quality of genome assemblies through curation.** *GigaScience.* 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

Kurtzer GM, Sochat V, Bauer MW: **Singularity: scientific containers for mobility of compute.** *PLoS One.* 2017; **12**(5): e0177459.
PubMed Abstract | Publisher Full Text | 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

Majerus MEN: **Ladybirds**. London: HarperCollins, 1994.
Reference Source

Majerus MEN: **A natural history of ladybird beetles**. Cambridge: Cambridge University Press, 2016.
Publisher 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

Merkel D: **Docker: lightweight Linux containers for consistent development and deployment.** *Linux J.* 2014; **2014**(239): 2.
Reference Source

Ranallo-Benavidez TR, Jaron KS, Schatz MC: **GenomeScope 2.0 and Smudgeplot for reference-free profiling of polyploid genomes.** *Nat Commun.* 2020; **11**(1): 1432.
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.*: **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

Roy HE, Brown PMJ: **Field guide to the ladybirds of Great Britain and Ireland**. London: Bloomsbury Publishing, 2018.
Reference Source

Schoch CL, Ciufo S, Domrachev M, *et al.*: **NCBI Taxonomy: a comprehensive update on curation, resources and tools.** *Database (Oxford).* 2020; **2020**: baaa062.
PubMed Abstract | Publisher Full Text | Free Full Text

Uliano-Silva M, Ferreira JGRN, Krasheninnikova K, *et al.*: **MitoHiFi: a python pipeline for mitochondrial genome assembly from PacBio high fidelity reads.** *BMC Bioinformatics.* 2023; **24**(1): 288.
PubMed Abstract | Publisher Full Text | Free Full Text

Vasimuddin M, Misra S, Li H, *et al.*: **Efficient architecture-aware acceleration of BWA-MEM for multicore systems.** In: *2019 IEEE International Parallel and Distributed Processing Symposium (IPDPS).* IEEE, 2019; 314–324.
Publisher Full Text

Wang SF, Braekman JC, Daloze D, *et al.*: **Nα-quinaldyl-l-arginine·HCl, a**

new defensive alkaloid from *Subcoccinella-24-punctata* (Coleoptera, Coccinellidae). *Experientia.* 1996; **52**: 628–630.
Publisher Full Text

Wheeler AG Jr, Henry TJ: **Seasonal history and habits of the European alfalfa beetle, *Subcoccinella vigintiquatuorpunctata* (L.) (Coleoptera: Coccinellidae).** *The Coleopterists Bulletin.* 1981; **35**(2): 197–203.
Reference Source

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 18 February 2026

https://doi.org/10.21956/wellcomeopenres.27995.r146128

✔ **Mario Stanke**
University of Greifswald, Greifswald, Germany

The paper introduces the genome assembly of the 24-spot ladybird (ladybug). The generated sequencing data includes PacBio HiFi reads, Hi-C data for scaffolding, and whole-organism RNA-Seq from Illumina. The paper largely follows a standardized template for Data Notes on genome assemblies. The results are useful, and plausible to me.

The presented assembly comfortably satisfies the current thresholds formulated by the EBP standards committee: the N50 exceeds 1 million bp, most of the sequence is assigned to chromosomes, and the estimated per-base error rate is below 1/10000.

Although the genome was not annotated, the RNA-Seq is a good resource to either create or assess a genome annotation in the future. Regarding the Hi-C contact map in Figure 3, no specific off-diagonal points are visible, only a barely visible background hue. It would be helpful to add a sentence explicitly describing the nature of this off-diagonal signal.

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: bioinformatics method development

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

Reviewer Report 11 February 2026

https://doi.org/10.21956/wellcomeopenres.27995.r146131

Andres Mariano Alonso
Universidad Nacional de San Martín (CONICET-UNSAM), Chascomús, Argentina

Crowley et al. describe the genome assembly of the 24-spot ladybird (*S. vigintiquattuorpunctata*). The metrics for the presented assembly are adequate, although the authors previously presented an assembly for a similar organism with better performance (10.12688/wellcomeopenres.17346.1). Could these differences be related to genome size or complexity? In this sense, the standard pipeline implemented here could be improved and adapted for the analyzed organism. I suggest the authors add a sentence discussing this in the main text.
The major issue I observed is related to the contact map (Figure 3). The authors describe features in the main text, but these are not clearly observable in the figure, particularly the TADs. The authors need to revise the presented information and provide a better figure so that the claimed scaffolding is visible to the readers.

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: Computational Biology, Bioinformatics

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 21 January 2026

https://doi.org/10.21956/wellcomeopenres.27995.r143212

Panagiotis Ioannidis
Foundation for Research & Technology - Hellas, Crete, Greece

This paper describes the sequencing and assembly of the genome of the beetle Subcoccinella vigitiquattuorpunctata.

The genome sequencing methodology is the standard one used in all the DToL projects (HiFi + Hi-C + RNAseq). The Hifiasm genome assembler is the tool producing the best genome assemblies from Hifi data. QC is done also using the state of the art (BUSCO + Merqury + Blobplots). The resulting genone assembly is one with a good BUSCO score, QV score and k-mer completeness.

However, the total number of scaffolds is relatively high (n = 256), and even the vast majority of the genome sequence is assigned to one of the 15 chromosomal level scaffolds, I've seen much better assemblies. This is very obvious in the Blob plot where there are two "clouds" of small scaffolds which on average have higher coverage than the chromosomal scaffolds and they also have a different GC content (both higher and lower than the 32-34% for the chromosomal scaffolds). The authors must comment on the origin of these extra scaffolds, and why are there so many of them?

Also, the Hi-C contact map (figure 3) only contains the diagonal. How is this possible? Why are there no TADs? I'm suspecting some kind of technical error in the processing of the Hi-C data. Most probably this figure needs to be re-plotted and replaced.

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: insect genomics, bioinformatics

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.