DATA NOTE

REVISED The genome sequence of a hoverfly *Eristalinus aeneus* (Scopoli, 1763) [version 2; peer review: 4 approved]

Olga Sivell[1], Chris Raper[1], Ryan Mitchell[2], Duncan Sivell[1],
Natural History Museum Genome Acquisition Lab,
Darwin Tree of Life Barcoding collective,
Wellcome Sanger Institute Tree of Life Management, Samples and Laboratory team,
Wellcome Sanger Institute Scientific Operations: Sequencing Operations,
Wellcome Sanger Institute Tree of Life Core Informatics team,
Tree of Life Core Informatics collective, Darwin Tree of Life Consortium

[1]Natural History Museum, London, England, UK
[2]Independent researcher, Sligo Town, County Sligo, Ireland

v2 **First published:** 19 Feb 2024, **9**:69
https://doi.org/10.12688/wellcomeopenres.20636.1
Latest published: 24 May 2024, **9**:69
https://doi.org/10.12688/wellcomeopenres.20636.2

Abstract

We present a genome assembly from an individual female *Eristalinus aeneus* (a hoverfly; Arthropoda; Insecta; Diptera; Syrphidae). The genome sequence is 495.4 megabases in span. Most of the assembly is scaffolded into 6 chromosomal pseudomolecules. The mitochondrial genome has also been assembled and is 15.97 kilobases in length.

Keywords

Eristalinus aeneus, hoverfly, genome sequence, chromosomal, Diptera

This article is included in the Tree of Life gateway.

Open Peer Review

Approval Status ✓ ✓ ✓ ✓

	1	2	3	4
version 2 (revision) 24 May 2024	✓ view		✓ view	✓ view
version 1 19 Feb 2024	? view	✓ view		

1. **Ljiljana Šašić Zorić**, BioSense Institute, University of Novi Sad, Novi Sad, Serbia
2. **Chufei Tang**, Jiangsu Academy of Agricultural Sciences, Nanjing, China
3. **Claus Vogl**, Vetmeduni Vienna, Veterinärplatz, Austria
4. **Andrzej Grzywacz**, Nicolaus Copernicus University in Toruń, Lwowska, Poland

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: Sivell O: Investigation, Resources, Writing – Original Draft Preparation, Writing – Review & Editing; **Raper C**: Investigation, Resources; **Mitchell R**: Investigation, Resources; **Sivell D**: 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 [206194, https://doi.org/10.35802/206194] and the Darwin Tree of Life Discretionary Award [218328, https://doi.org/10.35802/218328].
The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

How to cite this article: Sivell O, Raper C, Mitchell R *et al.* **The genome sequence of a hoverfly *Eristalinus aeneus* (Scopoli, 1763) [version 2; peer review: 4 approved]** Wellcome Open Research 2024, **9**:69 https://doi.org/10.12688/wellcomeopenres.20636.2

First published: 19 Feb 2024, **9**:69 https://doi.org/10.12688/wellcomeopenres.20636.1

REVISED **Amendments from Version 1**

The number of scaffolds has been changed from 197 to 198 in this version.

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

Species taxonomy

Eukaryota; Metazoa; Eumetazoa; Bilateria; Protostomia; Ecdysozoa; Panarthropoda; Arthropoda; Mandibulata; Pancrustacea; Hexapoda; Insecta; Dicondylia; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Eremoneura; Cyclorrhapha; Aschiza; Syrphoidea; Syrphidae; Eristalinae; Eristalini; *Eristalinus*; *Eristalinus aeneus* (Scopoli, 1763) (NCBI:txid2017640).

Background

The genus *Eristalinus* is represented in Europe by four species, only two of which: *E. aeneus* (Scopoli, 1763) and *E. sepulchralis* (Linnaeus, 1758) occur in Britain (Ball & Morris, 2015; Chandler, 2023; Speight, 2017). Both *E. sepulchralis* and *E. aeneus* are dark, shiny and slightly bronze in appearance. They both have spotted eyes and a loop in the wing vein R_{4+5} that distinguishes them from other British hoverflies. They can be separated based on the extent of hairs on the eye: in *E. sepulchralis* the whole eye is hairy while *E. aeneus* does not have hairs on the lower part of the eye (Ball & Morris, 2015; Falk *et al.*, 2023; Stubbs & Falk, 2002; van Veen, 2014). In Europe the females of these species can be separated by the appearance of the frons and vertex: in *E. aeneus* the area around the ocellar triangle and often also along the upper margins of the compound eyes is dark and lustrous, while in *E. sepulchralis* it is dull or slightly shiny and uniformly grey dusted (Kahanpää, 2022). The males can be identified by the width of the gap between the eyes: these are clearly separated in *E. sepulchralis* but holoptic in *E. aeneus* (Kahanpää, 2022; van Veen, 2014). Male genitalia are distinct (Pérez-Banón *et al.*, 2003).

Eristalinus aeneus is a widespread and cosmopolitan species with a Holarctic, Oriental, Afrotropical and Australasian distribution that includes Hawaii, Mauritius, Bermuda and the Gilbert and Ellis Islands. In the Afrotropics it extends south to Tanzania and in the Nearctic it reaches California and Texas (Speight, 2017).

Through most of its range *E. aeneus* can be found near ponds, slow-moving rivers, streams, irrigation ditches and coastal lagoons. In southern Europe the larvae of *E. aeneus* were found in association with animal dung and sewage farms (Speight, 2017).

In the northern limits of its range, including Britain and Ireland, this species is almost exclusively coastal where the larvae live in rotting seaweed in brackish waters or in rock pools (Ball & Morris, 2015; Stubbs & Falk, 2002). According to Hartley (1961), there are two generations per year. *Eristalinus aeneus* is widely distributed around the British coastline but is more frequently recorded in the south (Ball *et al.*, 2011). The flight period is generally from March to November, peaking in July and August, although adults can be encountered in any month of the year (Ball *et al.*, 2011). This hoverfly overwinters as an adult and has been found hibernating in buildings (Hartley, 1961).

The adults visit a variety of flowers, mainly yellow composites, white umbellifers, Aster, hoary alison *Berteroa incana*, rock-rose *Cistus*, oregano *Origanum*, creeping willow *Salix repens*, ragworts and groundsels *Senecio*, and dandelions *Taraxacum* (Speight, 2017) as well as other plants, such as *Thymelaea velutina*, endemic to the Balearic Islands (De La Bandera & Traveset, 2006). This species was also found to be an effective pollinator of various crops, for example mango *Mangifera indica*, watermelon *Citrullus lanatus*, onion *Allium cepa*, chickpea *Cicer arietinum*, celery *Apium graveolens* and fennel *Foeniculum vulgare* (Latif *et al.*, 2019; Saeed *et al.*, 2008; Sánchez *et al.*, 2022a; Sánchez *et al.*, 2022b; Sánchez *et al.*, 2022c).

The third instar larva and puparium of *E. aeneus* were described by Hartley (1961). Pérez-Bañón *et al.* (2003) also published a description of the third instar larva and puparium with SEM images of the anterior larval and pupal spiracles and a key to the puparia of European *Eristalinus* species. The intra-puparial development of *Eristalinus aeneus* was researched by Campoy *et al.* (2020).

Pérez-Bañón *et al.* (2003) also published molecular data (mitochondrial COI and nuclear 28S rDNA) for all the European species and *E. dubiosus* (Curran, 1939) from Kenya, as well as a few other species from the tribe Eristalini. The proposed phylogeny based on their findings (molecular and male genitalia) divides species into two clades (genera/subgenera), placing *E. aeneus* together with *E. sepulchralis* in *Eristalinus* Rondani, 1845, while the Mediterranean species *E. taeniops* (Wiedemann, 1818) and *E. megacephalus* (Rossi, 1794) belong in *Eristalodes* Mik, 1897. However, further studies are needed to resolve their taxonomic status (Pérez-Banón *et al.*, 2003).

Here we present a high-quality genome of *E. aeneus.* It was sequenced based on one female specimen from Mullion, The Lizard National Nature Reserve, Cornwall, England. The genome of the other British *Eristalinus* species, *E. sepulchralis*, was published by Falk *et al.* (2023). Both these genomes will aid research on the taxonomy and phylogeny of *Eristalinus* and related species. The genomes have been generated 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.

Genome sequence report

The genome was sequenced from a female *Eristalinus aeneus* (Figure 1) collected from Mullion, England (50.02, –5.24). A total of 56-fold coverage in Pacific Biosciences single-molecule

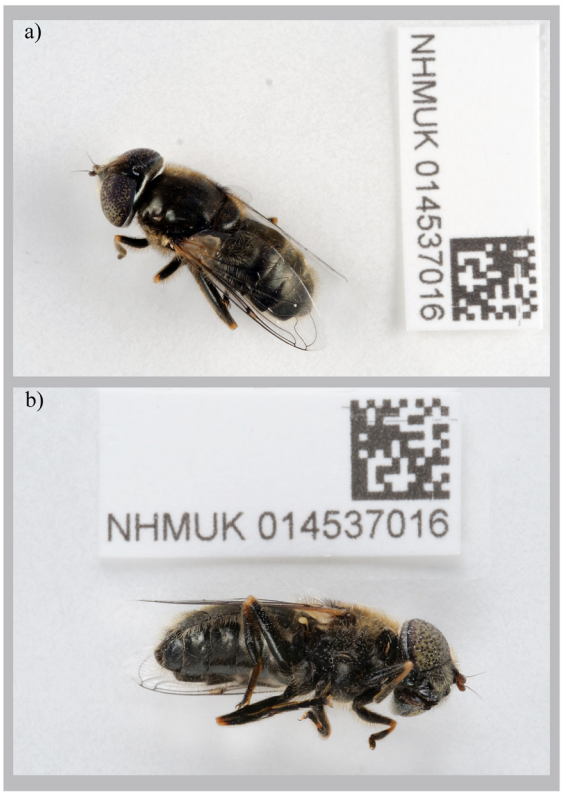

Figure 1. Photographs of the *Eristalinus aeneus* (idEriAene1, NHMUK014537016) specimen used for genome sequencing **a**) dorsal view, **b**) latero-ventral view.

HiFi long reads was generated. Primary assembly contigs were scaffolded with chromosome conformation Hi-C data. Manual assembly curation corrected 85 missing joins or mis-joins and removed one haplotypic duplication, reducing the scaffold number by 17.15%, and increasing the scaffold N50 by 9.65%.

The final assembly has a total length of 495.4 Mb in 198 sequence scaffolds with a scaffold N50 of 85.8 Mb (Table 1). The snail plot in Figure 2 provides 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 (97.3%) of the assembly sequence was assigned to 6 chromosomal-level scaffolds, representing 6 autosomes. Chromosome-scale scaffolds confirmed by the Hi-C data are named in order of size (Figure 5; Table 2). While not fully phased, the assembly deposited is of one haplotype. Contigs corresponding to the second haplotype have also been deposited. The mitochondrial genome was also assembled and can be found as a contig within the multifasta file of the genome submission.

The estimated Quality Value (QV) of the final assembly is 61.2 with *k*-mer completeness of 100.0%, and the assembly has a BUSCO v5.3.2 completeness of 96.6% (single = 96.1%, duplicated = 0.5%), using the diptera_odb10 reference set (n = 3,285).

Metadata for specimens, barcode results, spectra estimates, sequencing runs, contaminants and pre-curation assembly statistics are given at https://links.tol.sanger.ac.uk/species/2017640.

Methods

Sample acquisition and nucleic acid extraction

A female *Eristalinus aeneus* (specimen ID NHMUK014537016, ToLID idEriAene1) was collected from Mullion, Cornwall, England. The Lizard National Nature Reserve, England, UK (latitude 50.02, longitude –5.24) on 2021-07-01 using an aerial net. The specimen was collected by Olga Sivell and Chris Raper (Natural History Museum) and identified by Ryan Mitchell (Oxford University Museum of Natural History) and dry frozen at –80 °C.

The workflow for high molecular weight (HMW) DNA extraction at the Wellcome Sanger Institute (WSI) includes a sequence of core procedures: sample preparation; sample homogenisation, DNA extraction, fragmentation, and clean-up. For sample preparation, the idEriAene1 sample was weighed and dissected on dry ice (Jay *et al.*, 2023). Tissue from the whole organism was homogenised using a PowerMasher II tissue disruptor (Denton *et al.*, 2023a). HMW DNA was extracted using the Automated MagAttract v2 protocol (Oatley *et al.*, 2023a). The DNA was sheared into an average fragment size of 12–20 kb in a Megaruptor 3 system with speed setting 31 (Bates *et al.*, 2023). Sheared DNA was purified by solid-phase reversible immobilisation (Oatley *et al.*, 2023b): in brief, the method employs a 1.8X ratio of AMPure PB beads to sample to eliminate shorter fragments and concentrate the DNA. 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.

Protocols developed by the WSI Tree of Life core laboratory have been deposited on protocols.io (Denton *et al.*, 2023b).

Sequencing

Pacific Biosciences HiFi circular consensus DNA sequencing libraries were constructed according to the manufacturers' instructions. DNA sequencing was performed by the Scientific Operations core at the WSI on a Pacific Biosciences SEQUEL II instrument. Hi-C data were also generated from remaining tissue of idEriAene1 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

Table 1. Genome data for *Eristalinus aeneus*, idEriAene1.1.

Project accession data		
Assembly identifier	idEriAene1.1	
Species	*Eristalinus aeneus*	
Specimen	idEriAene1	
NCBI taxonomy ID	2017640	
BioProject	PRJEB62161	
BioSample ID	SAMEA112221990	
Isolate information	idEriAene1, female: whole organism (DNA sequencing and Hi-C sequencing)	
Assembly metrics*		***Benchmark***
Consensus quality (QV)	61.2	*≥ 50*
k-mer completeness	100.0%	*≥ 95%*
BUSCO**	C:96.6%[S:96.1%,D:0.5%], F:0.8%,M:2.6%,n:3,285	*C ≥ 95%*
Percentage of assembly mapped to chromosomes	97.3%	*≥ 95%*
Sex chromosomes	Not identified	*localised homologous pairs*
Organelles	Mitochondrial genome: 15.97 kb	*complete single alleles*
Raw data accessions		
PacificBiosciences SEQUEL II	ERR11458808	
Hi-C Illumina	ERR11468733	
Genome assembly		
Assembly accession	GCA_955652365.1	
Accession of alternate haplotype	GCA_955652355.1	
Span (Mb)	495.4	
Number of contigs	448	
Contig N50 length (Mb)	5.3	
Number of scaffolds	198	
Scaffold N50 length (Mb)	85.8	
Longest scaffold (Mb)	131.69	

* 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 version 5.3.2. 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/idEriAene1_1/dataset/idEriAene1_1/busco.

and corrected as described previously (Howe *et al.*, 2021). Manual curation was performed using gHiGlass (Kerpedjiev *et al.*, 2018) and PretextView (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.

A Hi-C map for the final assembly was produced using bwa-mem2 (Vasimuddin *et al.*, 2019) in the Cooler file format (Abdennur & Mirny, 2020). To assess the assembly metrics,

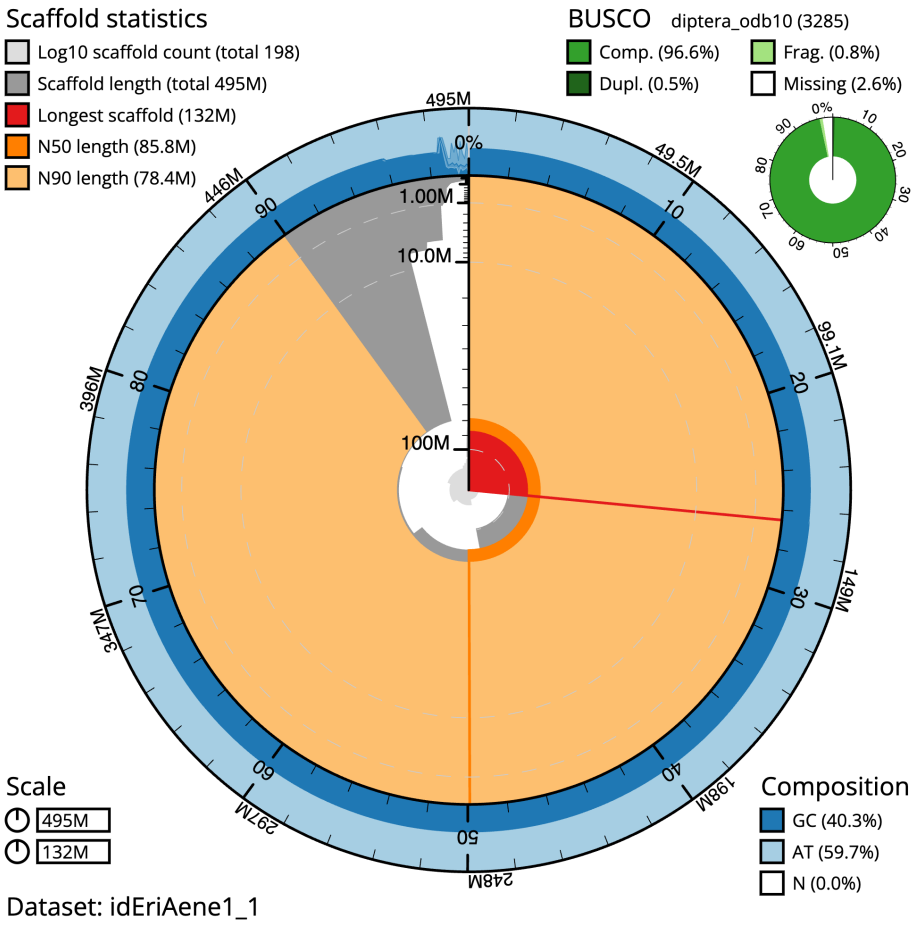

Figure 2. Genome assembly of *Eristalinus aeneus*, idEriAene1.1: metrics. The BlobToolKit snailplot shows N50 metrics and BUSCO gene completeness. The main plot is divided into 1,000 size-ordered bins around the circumference with each bin representing 0.1% of the 495,389,070 bp assembly. The distribution of scaffold lengths is shown in dark grey with the plot radius scaled to the longest scaffold present in the assembly (131,691,179 bp, shown in red). Orange and pale-orange arcs show the N50 and N90 scaffold lengths (85,846,282 and 78,373,126 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/idEriAene1_1/dataset/idEriAene1_1/snail.

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 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

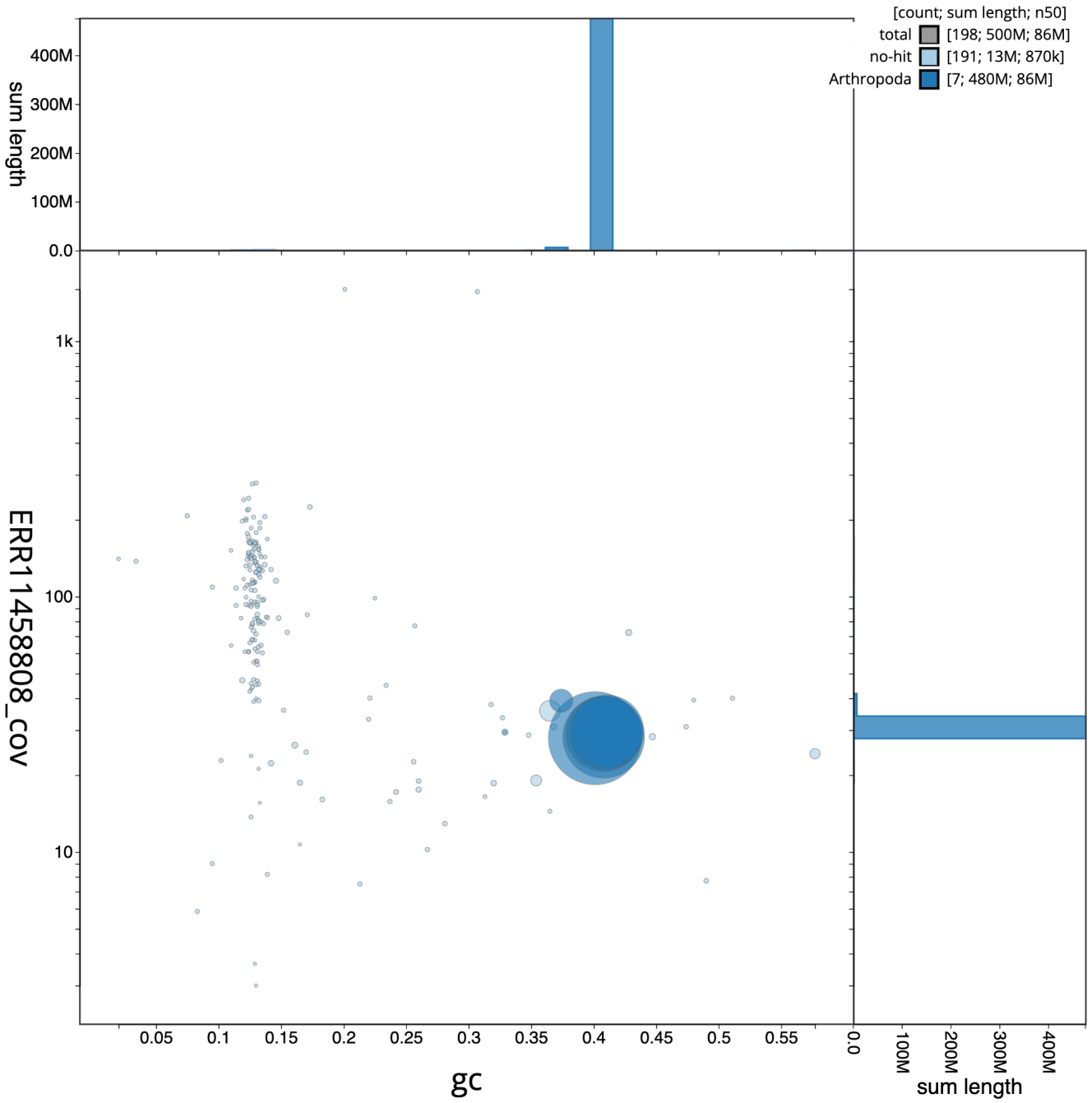

Figure 3. Genome assembly of *Eristalinus aeneus*, idEriAene1.1: BlobToolKit GC-coverage plot. Scaffolds are coloured by phylum. Circles are sized in proportion to scaffold length. Histograms show the distribution of scaffold length sum along each axis. An interactive version of this figure is available at https://blobtoolkit.genomehubs.org/view/idEriAene1_1/dataset/idEriAene1_1/blob.

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.

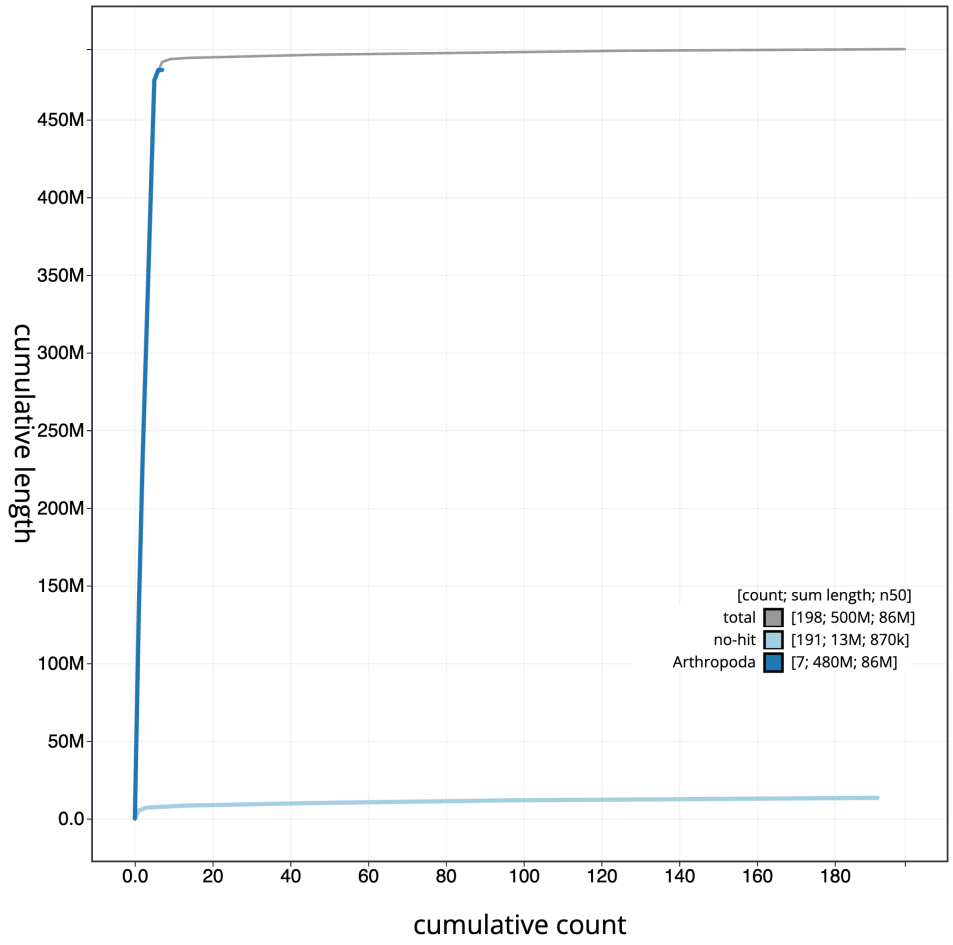

Figure 4. Genome assembly of *Eristalinus aeneus*, idEriAene1.1: BlobToolKit cumulative sequence plot. The grey line shows cumulative length for all scaffolds. Coloured lines show cumulative lengths of scaffolds assigned to each phylum using the buscogenes taxrule. An interactive version of this figure is available at https://blobtoolkit.genomehubs.org/view/idEriAene1_1/dataset/idEriAene1_1/cumulative.

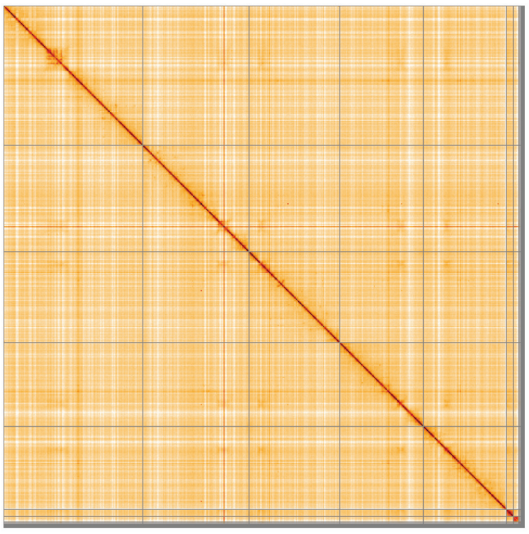

Figure 5. Genome assembly of *Eristalinus aeneus*, idEriAene1.1: Hi-C contact map of the idEriAene1.1 assembly, visualised using HiGlass. Chromosomes are shown in order of size from left to right and top to bottom. An interactive version of this figure may be viewed at https://genome-note-higlass.tol.sanger.ac.uk/l/?d=TFGREwXKStWHgFtrW8UVWQ.

Table 2. Chromosomal pseudomolecules in the genome assembly of *Eristalinus aeneus*, idEriAene1.

INSDC accession	Chromosome	Length (Mb)	GC%
OY019130.1	1	131.69	40.0
OY019131.1	2	100.31	41.0
OY019132.1	3	85.85	41.0
OY019133.1	4	79.05	41.0
OY019134.1	5	78.37	41.0
OY019135.1	6	6.56	37.5
OY019136.1	MT	0.02	20.0

Table 3. Software tools: versions and sources.

Software tool	Version	Source
BlobToolKit	4.2.1	https://github.com/blobtoolkit/blobtoolkit
BUSCO	5.3.2	https://gitlab.com/ezlab/busco
Hifiasm	0.16.1-r375	https://github.com/chhylp123/hifiasm
HiGlass	1.11.6	https://github.com/higlass/higlass
Merqury	MerquryFK	https://github.com/thegenemyers/MERQURY.FK
MitoHiFi	3	https://github.com/marcelauliano/MitoHiFi
PretextView	0.2	https://github.com/wtsi-hpag/PretextView
purge_dups	1.2.5	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	1.2a.2	https://github.com/c-zhou/yahs

Data availability

European Nucleotide Archive: *Eristalinus aeneus*. Accession number PRJEB62161; https://identifiers.org/ena.embl/PRJEB62161 (Wellcome Sanger Institute, 2023). The genome sequence is released openly for reuse. The *Eristalinus aeneus* 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.7139035.

Members of the Darwin Tree of Life Barcoding collective are listed here: https://doi.org/10.5281/zenodo.4893703.

Members of the Wellcome Sanger Institute Tree of Life Management, Samples and Laboratory team are listed here: https://doi.org/10.5281/zenodo.10066175.

Members of Wellcome Sanger Institute Scientific Operations: Sequencing Operations are listed here: https://doi.org/10.5281/zenodo.10043364.

Members of the Wellcome Sanger Institute Tree of Life Core Informatics team are listed here: https://doi.org/10.5281/zenodo.10066637.

Members of the Tree of Life Core Informatics collective are listed here: https://doi.org/10.5281/zenodo.5013541.

Members of the Darwin Tree of Life Consortium are listed here: https://doi.org/10.5281/zenodo.4783558.

Acknowledgements

Thank you to Natural England, the National Trust, Cornwall Wildlife Trust, RSPB, and the Woodland Trust for permission to collect at the Lizard NNR.

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

Ball SG, Morris RKA, Rotheray GE, *et al.*: **Atlas of the hoverflies of great Britain (Diptera, Syrphidae).** Wallingford: Biological Records Centre, 2011.
Reference Source

Ball S, Morris R: **Britain's hoverflies: a field guide - revised and updated second edition.** Princeton, NJ: Princeton University Press, 2015.
Reference Source

Bates A, Clayton-Lucey I, Howard C: **Sanger Tree of Life HMW DNA fragmentation: diagenode Megaruptor®3 for LI PacBio.** *protocols.io.* 2023.
Publisher 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

Campoy A, Pérez-Bañón C, Rojo S: **Intra-puparial development in the hoverflies *Eristalinus aeneus* and *Eristalis tenax* (Diptera: Syrphidae).** *J Morphol.* 2020; **281**(11): 1436–1445.
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 P: **Checklist of diptera of the British isles.** 2023.
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

De La Bandera MC, Traveset A: **Breeding system and spatial variation in the pollination biology of the heterocarpic *Thymelaea velutina* (Thymelaeaceae).** *Plant Syst Evol.* 2006; **257**(1–2): 9–23.
Publisher Full Text

Denton A, Oatley G, Cornwell C, *et al.*: **Sanger Tree of Life sample homogenisation: PowerMash.** *Protocols.io.* 2023a.
Publisher Full Text

Denton A, Yatsenko H, Jay J, *et al.*: **Sanger Tree of Life wet laboratory protocol collection.** *Protocols.io.* 2023b.
Publisher 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

Falk S, Sivell O, Sivell D, *et al.*: **The genome sequence of the small spotty-eyed dronefly, *Eristalinus sepulchralis* (Linnaeus, 1758) [version 1; peer review: awaiting peer review].** *Wellcome Open Res.* 2023; **8**: 269.
Publisher Full Text

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

Hartley JC: **A taxonomic account of the larvae of some British syrphidae.** *Proc Zool Soc Lond.* 1961; **136**(4): 505–73.
Publisher Full Text

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

Jay J, Yatsenko H, Narváez-Gómez JP, *et al.*: **Sanger Tree of Life sample preparation: triage and dissection.** *Protocols.io.* 2023.
Publisher Full Text

Kahanpää J: **A new character for identification of females of *Eristalinus aeneus* (Scopoli, 1763) and *E. sepulchralis* (Linnaeus, 1758) (Diptera: Syrphidae).** *Sahlbergia.* 2022; **27**(2): 11–13.
Reference Source

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

Latif A, Malik SA, Saeed S, *et al.*: **Diversity of pollinators and their role in the pollination biology of chickpea, *Cicer arietinum* L. (Fabaceae).** *J Asia Pac Entomol.* 2019; **22**(2): 597–601.
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

Oatley G, Denton A, Howard C: **Sanger Tree of Life HMW DNA extraction: automated MagAttract v.2.** *Protocols.Io.* 2023a.
Publisher Full Text

Oatley G, Sampaio F, Howard C: **Sanger Tree of Life fragmented DNA clean up: automated SPRI.** *protocols.io.* 2023b; [Accessed 21 November 2023].
Publisher Full Text

Pérez-Banón C, Rojo S, Ståhls G, *et al.*: **Taxonomy of European *Eristalinus* (Diptera: Syrphidae) based on larval morphology and molecular data.** *Eur J Entomol.* 2003; **100**(3): 417–428.
Publisher 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

Saeed S, Sajjad A, Kwon O, *et al.*: **Fidelity of Hymenoptera and Diptera pollinators in onion (*Allium cepa* L.) pollination.** *Entomol Res.* 2008; **38**(4): 276–280.
Publisher Full Text

Sánchez M, Belliure B, Montserrat M, *et al.*: **Pollination by the hoverfly *Eristalinus aeneus* (Diptera: Syrphidae) in two hybrid seed crops: celery and fennel (Apiaceae).** *J Agric Sci.* 2022a; **160**(3–4): 1–48.
Publisher Full Text

Sánchez M, Velásquez Y, González M, *et al.*: **Activity and foraging behaviour of the hoverfly *Eristalinus aeneus* (Scopoli, 1763) in protected cultivation of mango (*Mangifera indica* L.).** *Bull Entomol Res.* 2022b; **112**(1): 101–109.
PubMed Abstract | Publisher Full Text

Sánchez M, Velásquez Y, González M, *et al.*: **Pollination effectiveness of the hoverfly *Eristalinus aeneus* (Scopoli, 1763) in diploid and triploid associated watermelon crop.** *Insects.* 2022c; **13**(11): 1021.
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

Speight MCD: **Species accounts of European Syrphidae (Diptera).** In: *Syrph the Net the database of European Syrphidae.* Dublin: Syrph the Net publications, 65. 2017.

Stubbs AE, Falk SJ: **British hoverflies: an illustrated identification guide.** British Entomological and Natural History Society, 2002.
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, 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

van Veen MP: **Hoverflies of Northwest Europe: identification keys to the Syrphidae.** KNNV Publishing. Utrecht, Netherlands: BRILL, 2014.
Reference Source

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

Wellcome Sanger Institute: **The genome sequence of a hoverfly, *Eristalinus aeneus* (Scopoli, 1763).** European Nucleotide Archive. [dataset], accession number PRJEB62161, 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 2

Reviewer Report 08 August 2024

https://doi.org/10.21956/wellcomeopenres.24731.r82090

✔ **Andrzej Grzywacz**
Nicolaus Copernicus University in Toruń, Lwowska, Toruń, Poland

Well written report of *Eristalinus aeneus* genome. Contribution is satisfactory and contains all the necessary information. Introduction covers most important aspects of *Eristalinus aeneus* life history with the most recent references. The genome assembly is of high quality. A brief explanation of taxonomic issues of *Eristalinus aeneus* justify obtaining the genome of this species.

Only a few minor comments listed below:

- Should "Aster" be in italics?
- Please merge "A female *Eristalinus aeneus* (specimen ID NHMUK014537016, ToLID idEriAene1) was collected from Mullion, Cornwall, England. The Lizard National Nature Reserve, England, UK (latitude 50.02, longitude –5.24) on 2021-07-01 using an aerial net." into a single sentence.
- "Hi-C data were also generated from remaining tissue of idEriAene1..." please be more specific.

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: Entomology, Systematics, Taxonomy

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

Reviewer Report 06 June 2024

https://doi.org/10.21956/wellcomeopenres.24731.r82085

Claus Vogl
Vetmeduni Vienna, Veterinärplatz, Wien, Austria

The article presents the genome of the hoverfly *Eristalinus aeneus*. The species seems identified cleanly. I am not sure that an insect with such a large distribution range is really one species and not subdivided into many phenotypically similar cryptic species. But only data sets, such as this one will help to answer this question. In any case, the location where the sequenced specimen was collected was given accurately.

Generally, the authors left little to criticise. On page 3, the following change may clarify the meaning: "In the Afrotropics it extends south to Tanzania and in the Nearctic it reaches California and Texas ()" to "In the Afrotropics it extends south to Tanzania and in the Nearctic south to California and Texas ()".
And parentheses may help here: "... ragworts and groundsels (Senecio), ..."

Otherwise the methods seem state of the art, the quality criteria seem to have been met, and the genome appears to be sound. I did not check if I could actually download the genome, but everything seems OK.

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: Statistics, population genetics and genomics. I have collaborated on fly genome projects.

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 28 May 2024

https://doi.org/10.21956/wellcomeopenres.24731.r84476

Ljiljana Šašić Zorić
BioSense Institute, University of Novi Sad, Novi Sad, Serbia

Do not have any further comments.

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: Population genetics and molecular taxonomy

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

Version 1

Reviewer Report 21 May 2024

https://doi.org/10.21956/wellcomeopenres.22839.r80211

Chufei Tang
Jiangsu Academy of Agricultural Sciences, Nanjing, China

This article presents the genome assembly of a cosmopolitan species, *Eristalinus aeneus*, known for its fascinating habits such as potential salt tolerance and pollination. The rationale for creating the dataset is well-explained and highlights its importance. The protocols are suitable and the work is technically sound. Sufficient details of methods and materials are provided to allow replication by others, and the datasets are presented in a clear and accessible format.
One minor suggestion: the introduction could be better organized - there are eight paragraphs, some of which are quite short. It is suggested to integrate information about adults and larvae together. For example, the study of larvae in the 6th paragraph could be incorporated into the discussion of adult morphology in the first paragraph. Maybe the introduction could be divided as follows.
1) The species range, as discussed in the first, second, and fourth paragraphs, pertains to the representativeness of the species.
2) The morphology, including the information in the first and potential additional details from the references cited in the 6th paragraph.
3) The habits and habitats, maybe first adults then larvae, including the information from the third, fourth and fifth paragraphs.
4) The progress on the molecular data, including this species and the sister species mentioned in the last paragraph.
5) The aim of the paper.

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: taxonomy, phylogeny

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 13 May 2024

https://doi.org/10.21956/wellcomeopenres.22839.r80217

?

Ljiljana Šašić Zorić
BioSense Institute, University of Novi Sad, Novi Sad, Serbia

"The genome sequence of a hoverfly, *Eristalinus aeneus* (Scopoli, 1763)" is presented in the form of a technical report on the genome assembly of hoverfly species *Eristalinus aeneus*. The article is an important contribution to further research on hoverflies (Diptera, Syrphidae). It describes technical aspects of genome assembly clearly and concisely. The applied approach includes PacBio and Hi-C data. The methodology is suitable and overall, very well described. I have noticed only few small inconsistencies:

- No comma is needed in the title;
- I am not familiar with the word "nan" which is listed in the keywords. It is mentioned only at this point in the text. Please check if the spelling is correct;
- Please check the total number of scaffolds. In the text and table 1 it is 197 but based on figures 2 – 4 it should be 198;
- Since abbreviation WSI was introduced for the Wellcome Sanger Institute, please use it consistently in the text.

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: Population genetics and molecular taxonomy

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