The genome sequence of the drone fly, *Eristalis tenax* (Linnaeus, 1758) [version 1; peer review: awaiting peer review]

William Hawkes\(^1\), Karl Wotton\(^1\),

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

\(^1\)Centre for Ecology and Conservation, University of Exeter, Penryn, UK

**Abstract**

We present a genome assembly from an individual female *Eristalis tenax* (the drone fly; Arthropoda; Insecta; Diptera; Syrphidae). The genome sequence is 487 megabases in span. The majority of the assembly (96.50%) is scaffolded into six chromosomal pseudomolecules, with the X sex chromosome assembled.

**Keywords**

Eristalis tenax, drone fly, genome sequence, chromosomal, Diptera

This article is included in the Tree of Life gateway.
Corresponding author: Darwin Tree of Life Consortium (mark.blaxter@sanger.ac.uk)

Author roles: Hawkes W: Investigation, Resources, Writing – Original Draft Preparation, Writing – Review & Editing; Wotton K: Investigation, Supervision, 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) 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: © 2021 Hawkes W 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: Hawkes W, Wotton K, University of Oxford and Wytham Woods Genome Acquisition Lab et al. The genome sequence of the drone fly, Eristalis tenax (Linnaeus, 1758) [version 1; peer review: awaiting peer review] Wellcome Open Research 2021, 6:307 https://doi.org/10.12688/wellcomeopenres.17357.1

First published: 10 Nov 2021, 6:307 https://doi.org/10.12688/wellcomeopenres.17357.1
Species taxonomy
Eukaryota; Metazoa; Ecdysozoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Syrphoidea; Syrphidae; Eristalinae; Eristalini; Eristalis; *Eristalis tenax* (Linnaeus 1758) (NCBI:txid198635).

Background
The dronefly, *Eristalis tenax* (Figure 1), is perhaps the most widespread of the hoverflies, with a cosmopolitan distribution. It is common in the United Kingdom and Ireland, where it can be found in the springtime following the emergence of females from overwintering in sheltered cavities in caves or buildings (Ball & Morris, 2013). They are at their most numerous in the UK during the late summer and autumn as the local population is augmented by migratory influxes from mainland Europe. Large southward migrations have been observed during the autumn in the UK, Europe and North America (Aubert et al., 1976; Owen, 1956; Shannon, 1926). *E. tenax* is a fairly large hoverfly separated from others in the *Eristalis* genus by the presence of a broad, black longitudinal face-stripe, longitudinal dark stripes on their eyes, and a distinctly enlarged and curved hind tibia. Like their sister species *Eristalis pertinax*, *E. tenax* is a visual, acoustic and behavioural mimic of the honeybee, *Apis mellifera* (Golding et al., 2001; Moore & Hassall, 2016). Males are territorial, aggressively defending patches of flowers from conspecifics and other flying insects (Wellington & Fitzpatrick, 1981). Their larvae, colloquially known as rat-tailed maggots, live in a wide array of organically rich pools, feeding on decaying organic matter. They therefore play an important ecological role in terms of decomposition (Hurtado et al., 2008). In addition, *E. tenax* is an important pollinator, visiting a wide range of crops and wild plants, while its role as a migratory pollinator may be particularly important for geographically isolated plant populations (Doyle et al., 2020; Pérez-Bañón et al., 2007; Rader et al., 2020). *E. tenax* can be reared and has been the subject of numerous investigations into its biology, including studies on population structure, flight, mimicry, vision and behaviour (Francuski & Milankov, 2015; Golding et al., 2001; Lunau et al., 2018; Nicholas et al., 2018; Straw et al., 2006). This is the first production of a high quality *Eristalis tenax* genome, and we believe that the sequence described here, generated as part of the Darwin Tree of Life project, will further aid understanding of the biology and ecology of this hoverfly.

Genome sequence report
The genome was sequenced from a single female *E. tenax* collected from Wytham Great Wood, Oxfordshire, UK (latitude 51.769, longitude -1.33). A total of 36-fold coverage in Pacific Biosciences single-molecule long reads (N50 12 kb) and 60-fold coverage in 10X Genomics read clouds (from molecules with an estimated N50 of 60 kb) were generated. Primary assembly contigs were scaffolded with chromosome conformation Hi-C data. Manual assembly curation corrected 71 missing/misjoins and removed 11 haplotypic duplications, reducing the assembly length by 1.92% and the scaffold number by 31.28%, and increasing the scaffold N50 by 72.23%.

The final assembly has a total length of 487 Mb in 157 sequence scaffolds with a scaffold N50 of 77.1 Mb (Table 1). The majority, 96.50%, of the assembly sequence was assigned to 6 chromosomal-level scaffolds, representing 5 autosomes (numbered by sequence length), and the X sex chromosome (Figure 2–Figure 5; Table 2). The assembly has a BUSCO (Simão et al., 2015) completeness of 96.6% using the diptera_odb10 reference set. While not fully phased, the assembly deposited is of one haplotype. Contigs corresponding to the second haplotype have also been deposited.

Methods
Sample acquisition and nucleic acid extraction
A female (idEriTena2) and a male (idEriTena3) *E. tenax* were collected from Wytham Great Wood, Oxfordshire, UK (latitude 51.769, longitude -1.33) by Will Hawkes, University of Exeter, who also identified the sample. The samples were collected using a net, snap-frozen on dry ice and stored in a CoolRack.
DNA was extracted from the head/thorax of idEriTena2 at the Wellcome Sanger Institute Scientific Operations core from the whole organism using the Qiagen MagAttract HMW DNA kit, according to the manufacturer’s instructions. RNA was extracted from head/thorax tissue of idEriPert3 in the Tree of Life Laboratory at the Wellcome Sanger Institute using TRIzol (Invitrogen), according to the manufacturer’s instructions. RNA was then eluted in 50 μl RNAse-free water and its concentration assessed using a Nanodrop spectrophotometer and Qubit Fluorometer using the Qubit RNA Broad-Range (BR) Assay kit. Analysis of the integrity of the RNA was done using Agilent RNA 6000 Pico Kit and Eukaryotic Total RNA assay.

### Sequencing

Pacific Biosciences HiFi circular consensus and 10X Genomics Chromium read cloud sequencing libraries were constructed according to the manufacturers’ instructions. Poly(A) RNA-Seq libraries were constructed using the NEB Ultra II RNA Library Prep kit. Sequencing was performed by the Scientific Operations core at the Wellcome Sanger Institute on Pacific Biosciences SEQUEL II (HiFi), Illumina HiSeq X (10X) and Illumina HiSeq 4000 (RNA-Seq) instruments. Hi-C data were generated from the abdomen tissue of idEriTena2 using the Arima v1 Hi-C kit and sequenced on HiSeq X.

| Project accession data |
|------------------------|
| Assembly identifier    | idEriTena2.1          |
| Species                | Eristalis tenax       |
| Specimen               | idEriTena2 (genome assembly, Hi-C); idEriTena3 (RNA-Seq) |
| NCBI taxonomy ID       | 219539                |
| BioProject             | PRJEB43032            |
| BioSample ID           | SAMEA7520160          |
| Isolate information    | Female, head/thorax (genome assembly) |

| Raw data accessions    |
|------------------------|
| PacificBiosciences SEQUEL II | ERR6565938 |
| 10X Genomics Illumina  | ERR6054456-ERR6054459 |
| Hi-C Illumina          | ERR6054460          |
| Illumina polyA RNA-Seq| ERR6054461          |

| Genome assembly        |
|------------------------|
| Assembly accession     | GCA_905231855.1      |
| Accession of alternate haplotype | GCA_905231845.1 |
| Span (Mb)               | 487                  |
| Number of contigs       | 290                  |
| Contig N50 length (Mb)  | 6.65                 |
| Number of scaffolds     | 157                  |
| Scaffold N50 length (Mb)| 77.07                |
| Longest scaffold (Mb)   | 141                  |

| BUSCO* genome score    |
|------------------------|
| C:96.6[%;S:95.9%,D:0.7%],F:0.8%, M:2.6%,n:3285 |

*BUSCO scores based on the diptera_odb10 BUSCO set using v5.1.2. 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/idEriTena2.1/dataset/CAJNIU01/busc.
Figure 2. Genome assembly of *Eristalis tenax*, idEriTena2.1: 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 487,019,564 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 (140,698,496 bp, shown in red). Orange and pale-orange arcs show the N50 and N90 scaffold lengths (77,068,984 and 65,930,618 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/idEriTena2.1/dataset/CAJNIU01/snail.
Genome assembly

Assembly was carried out with Hifiasm (Cheng et al., 2021); haplotypic duplication was identified and removed with purge_dups (Guan et al., 2020) with the -e flag. One round of polishing was performed by aligning 10X Genomics read data to the assembly with longranger align, calling variants with freebayes (Garrison & Marth, 2012). The assembly was then scaffolded with Hi-C data (Rao et al., 2014) using SALSA2 (Ghurye et al., 2019). The assembly was checked for contamination and corrected using the gEVAL system (Chow et al., 2016) as described previously (Howe et al., 2021). Manual curation was performed using gEVAL, HiGlass.

Figure 3. Genome assembly of *Eristalis tenax*, idEriTena2.1: GC coverage. 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/idEriTena2.1/dataset/CAJNIU01/blob.
Figure 4. Genome assembly of *Eristalis tenax*, idEriTena2.1: cumulative sequence. 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/idEriTena2.1/dataset/CAJNIU01/cumulative. (Kerpedjiev et al., 2018) and Pretext. The mitochondrial genome was assembled using MitoHiFi (Uliano-Silva et al., 2021). The genome was analysed and BUSCO scores generated within the BlobToolKit environment (Challis et al., 2020). Table 3 contains a list of all software tool versions used, where appropriate.
Figure 5. Genome assembly of *Eristalis tenax*, idEriTena2.1: Hi-C contact map. Hi-C contact map of the idEriTena2.1 assembly, visualised in HiGlass.

Table 2. Chromosomal pseudomolecules in the genome assembly of *Eristalis tenax*, idEriTena2.

| INSDC accession | Chromosome | Size (Mb) | GC% |
|-----------------|------------|-----------|-----|
| HG993124.1      | 1          | 140.7     | 41  |
| HG993125.1      | 2          | 97.83     | 41.9|
| HG993126.1      | 3          | 77.07     | 42.7|
| HG993127.1      | 4          | 66.31     | 43.1|
| HG993128.1      | 5          | 65.93     | 42.7|
| HG993129.1      | X          | 16.18     | 42.7|
| HG993130.1      | MT         | 0.02      | 18.8|
| -               | Unplaced   | 22.99     | 42.8|
**Table 3. Software tools used.**

| Software tool        | Version     | Source                                      |
|----------------------|-------------|---------------------------------------------|
| Hifiasm              | 0.12        | Cheng *et al.*, 2021                       |
| purge_dups           | 1.2.3       | Guan *et al.*, 2020                        |
| SALSA2               | 2.2         | Ghurye *et al.*, 2019                      |
| longranger align     | 2.2.2       | https://support.10xgenomics.com/.../other-pipelines |
| freebayes            | 1.3.1-17-gaa2ace8 | Garrison & Marth, 2012                  |
| MitoHiFi             | 1           | Uliano-Silva *et al.*, 2021                |
| gEVAL                | N/A         | Chow *et al.*, 2016                       |
| HiGlass              | 1.11.6      | Kerpedjiev *et al.*, 2018                  |
| PretextView          | 0.1.x       | https://github.com/wtsi-hpag/PretextView   |
| BlobToolKit          | 2.6.2       | Challis *et al.*, 2020                     |

**Data availability**

European Nucleotide Archive: Eristalis tenax (the dronefly). Accession number PRJEB43032; https://identifiers.org/ena.embl/PRJEB43032.

The genome sequence is released openly for reuse. The *E. pertinax* 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 the 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 University of Oxford and Wytham Woods Genome Acquisition Lab are listed here: https://doi.org/10.5281/zenodo.4789929.

Members of the Darwin Tree of Life Barcoding collective are listed here: https://doi.org/10.5281/zenodo.4893704.

Members of the Wellcome Sanger Institute Tree of Life programme are listed here: https://doi.org/10.5281/zenodo.5377053.

Members of Wellcome Sanger Institute Scientific Operations: DNA Pipelines collective are listed here: https://doi.org/10.5281/zenodo.4790456.

Members of the Tree of Life Core Informatics collective are listed here: https://doi.org/10.5281/zenodo.5013542.

Members of the Darwin Tree of Life Consortium are listed here: https://doi.org/10.5281/zenodo.4783559.

**References**

Aubert J, Aubert Jj, Goeldlin P: DOUZE ANS DE CAPTURES SYSTEMATIQUES DE SYRPHIDES (DIPTERES) AU COL DE BRETOLET (ALPES VALAISANNES). 1976. Reference Source

Ball S, Morris R: Britain’s Hoverflies: A Field Guide (Second Edi). WILDGuides. 2013.

Challis R, Richards E, Rajan J, *et al.*: BlobToolKit - Interactive Quality Assessment of Genome Assemblies. *G3 (Bethesda)*. 2020; 10(4): 1361-74. Published 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): 170-75. Published Abstract | Publisher Full Text | Free Full Text

Chow W, Brugger K, Caccamo M, *et al.*: gEVAL - a web-based browser for evaluating genome assemblies. *Bioinformatics*. 2016; 32(16): 2508-10. PubMed Abstract | Publisher Full Text | Free Full Text

Doyle T, Hawkes WLS, Massy R, *et al.*: Pollination by Hoverflies in the Anthropocene. *Proc Biol Sci*. 2020; 287(1927): 20200008. PubMed Abstract | Publisher Full Text | Free Full Text

Francuski L, Milankov V: Assessing Spatial Population Structure and Heterogeneity in the Dronefly. *J Zool.*. 2015; 297(4): 286-300. Publisher Full Text

Garrison E, Marth G: Haplotype-Based Variant Detection from Short-Read Sequencing. *arXiv*: 1207.3907, 2012. Reference Source

Ghurye J, Rhie A, Wolenz BP, *et al.*: Integrating Hi-C Links with Assembly Graphs for Chromosome-Scale Assembly. *PloS Comput Biol*. 2019; 15(8):...
Golding YC, Ennos AR, Edmunds M: Similarity in Flight Behaviour between the Honeybee Apis Mellifera (Hymenoptera: Apidae) and Its Presumed Mimic, the Dronefly Eristalis Tenax (Diptera: Syrphidae). J Exp Biol. 2001; 204(Pt 1): 139–45.

Guan D, McCarthy SA, Wood J, et al.: Identifying and Removing Haplotypic Duplication in Primary Genome Assemblies. Bioinformatics. 2020; 36(9): 2896–98.

Howe K, Chow W, Collins J, et al.: Significantly Improving the Quality of Genome Assemblies through Curation. Gigascience. 2021; 10(1): giaa153.

Hurtado P, Pérez-Bañón C, Gladis T, et al.: Biology of Saprophagous Hoverflies (Diptera, Syrphidae) and Its Role in Degrading of Pig Slurry. In XXIII International Congress of Entomology, Durban (South Africa). 2008.

Kerpedjiev P, Abdennur N, Lekschas F, et al.: HiGlass: Web-Based Visual Exploration and Analysis of Genome Interaction Maps. Genome Biol. 2018; 10(1): giaa153.

Lunau K, An L, Donda M, et al.: Limitations of Learning in the Proboscis Reflex of the Flower Visiting Syrphid Fly Eristalis Tenax. PLoS One. 2018; 13(3): e0194167.

Moore CD, Hassall C: A Bee or Not a Bee: An Experimental Test of Acoustic Mimicry by Hoverflies. Behavioral Ecology: Official Journal of the International Society for Behavioral Ecology. 2016; 27(6): 1767–74.

Nicholas S, Thysellius M, Holden M, et al.: Rearing and Long-Term Maintenance of Eristalis Tenax Hoverflies for Research Studies. J Vis Exp. 2018; (135): 57711.

Owen DF: A Migration of Insects at Spurn Point, Yorkshire. Entomologist’s Monthly Magazine. 1956; 92: 43–44.

Pérez-Bañón C, Petanidou T, Angeles Marcos-García Ma: Pollination in Small Islands by Occasional Visitors: The Case of Daucus Carota Subsp. Commutatus (Apiaceae) in the Columbretes Archipelago, Spain. Plant Ecology. 2007; 192(1): 153–51.

Rader R, Cunningham SA, Howlett BG, et al.: Non-Bee Insects as Visitors and Pollinators of Crops: Biology, Ecology, and Management. Annu Rev Entomol. 2020; 65: 391–407.

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): 1665–80.

Shannon Hj: A Preliminary Report on the Seasonal Migrations of Insects. J New York Entomol S. 1926; 34(2): 199–205.

Simão FA, Waterhouse RM, Ioannidis P, et al.: BUSCO: Assessing Genome Assembly and Annotation Completeness with Single-Copy Orthologs. Bioinformatics. 2015; 31(19): 3210–12.

Straw AD, Warrant EJ, O’Carroll DC: A “bright zone” in male hoverfly (Eristalis tenax) eyes and associated faster motion detection and increased contrast sensitivity. J Exp Biol. 2006; 209(Pt 21): 4339–54.

Ulmano-Silva M, Nunes JGF, Krasheninnikova K, et al.: marcelauliano/MitoHiFi: mitohifi_v2.0. 2021.

Wellington WG, Fitzpatrick SM: TERRITORIALITY IN THE DRONE FLY, Eristalis Tenax (Diptera: Syrphidae). The Canadian Entomologist. 1981; 113(8): 695–704.

Owen DF, a migration of insects at spurn point, yorkshire. Entomologist’s Monthly Magazine. 1956; 92: 43–44.

Pérez-Bañón C, Petanidou T, Angeles Marcos-García Ma: Pollination in small islands by occasional visitors: the case of daucus carota subsp. commutatus (apiaceae) in the columbretes archipelago, Spain. Plant ecology. 2007; 192(1): 153–51.