DATA NOTE

The genome sequence of the hoverfly, *Brachypalpoides lentus* Meigen, 1822 (Diptera: Syrphidae)

[version 1; peer review: 2 approved]

Ryan Mitchell[1], Katie J. Woodcock[2],
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]Independent researcher, Sligo, County Sligo, Ireland
[2]Wellcome Sanger Institute, Hinxton, England, UK

v1 **First published:** 08 Sep 2025, **10**:485
https://doi.org/10.12688/wellcomeopenres.24852.1
Latest published: 08 Sep 2025, **10**:485
https://doi.org/10.12688/wellcomeopenres.24852.1

Abstract

We present a genome assembly from an individual male *Brachypalpoides lentus* (hoverfly; Arthropoda; Insecta; Diptera; Syrphidae). The assembly contains two haplotypes with total lengths of 1 315.37 megabases and 1 154.75 megabases. Most of haplotype 1 (95.66%) is scaffolded into 6 chromosomal pseudomolecules, including the X and Y sex chromosomes. Haplotype 2 was assembled to scaffold level. The mitochondrial genome has also been assembled, with a length of 15.88 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.

Keywords

Brachypalpoides lentus; hoverfly; genome sequence; chromosomal; Diptera

This article is included in the Tree of Life gateway.

Open Peer Review

Approval Status ✓ ✓

	1	2
version 1 08 Sep 2025	✓ view	✓ view

1. **Lore Esselens**, Royal Museum for Central Africa, Tervuren, Belgium
2. **Mihajla Djan**, University of Novi Sad, Novi Sad, Serbia

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: Mitchell R: Investigation, Resources; **Woodcock KJ**: Writing – Original Draft Preparation;

Competing interests: No competing interests were disclosed.

Grant information: This work was supported by Wellcome through core funding to the Wellcome Sanger Institute (220540) 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: Mitchell R, Woodcock KJ, Natural History Museum Genome Acquisition Lab *et al.* **The genome sequence of the hoverfly, *Brachypalpoides lentus* Meigen, 1822 (Diptera: Syrphidae) [version 1; peer review: 2 approved]** Wellcome Open Research 2025, **10**:485 https://doi.org/10.12688/wellcomeopenres.24852.1

First published: 08 Sep 2025, **10**:485 https://doi.org/10.12688/wellcomeopenres.24852.1

Species taxonomy

Eukaryota; Opisthokonta; Metazoa; Eumetazoa; Bilateria; Protostomia; Ecdysozoa; Panarthropoda; Arthropoda; Mandibulata; Pancrustacea; Hexapoda; Insecta; Dicondylia; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Eremoneura; Cyclorrhapha; Aschiza; Syrphoidea; Syrphidae; Eristalinae; Xylotini; *Brachypalpoides*; *Brachypalpoides lentus* Meigen, 1822 (NCBI:txid224232)

Background

Brachypalpoides lentus (Meigen 1822) is a hoverfly species found throughout north-western Europe (Ball & Morris, 2015a; van Veen, 2010). In the UK it is most abundant within southern forest habitats, particularly those rich in beech and oak trees, though its range extends as far north as central Scotland (Ball & Morris, 2000). Adults are most commonly observed between April and August, peaking in numbers during June (Stubbs & Falk, 2002).

The species is visually distinctive, with a black body and a bright red band across a large portion of the abdomen and entirely black legs (Stubbs & Falk, 2002). It is believed to mimic sawfly and spider hunting wasp species, which share similar abdominal markings such as *Macrophya annulata* and *Priocnemis perturbator* (Ball & Morris, 2015b; van Veen, 2010). *B. lentus* could be confused with *Xylota segnis* and *Xylota tarda*, which differ by having an orange/yellow abdominal band and partially yellow legs. Notably, *B. lentus* also displaya behaviours similar to that of *Xylota* species, with individuals scuttling restlessly over leaves, feeding on honeydew and pollen grains (Ball & Morris, 2015a; van Veen, 2010). Adults have been sighted visiting hawthorn and ranunculus flowers (Ball & Morris, 2000). Females have been observed flying low into hollows at the base of living trees, a behaviour believed to be linked to egg-laying (Stubbs & Falk, 2002). Saprophagous *B. lentus* larvae can be found within the damp, rotten heartwood of large deciduous trees, often in live trees with exposed decay at the roots (Ball & Morris, 2015b; Rotheray, 1993). The completed genome sequence for *Brachypalpoides lentus* will provide a valuable tool to further our knowledge of this hoverfly. The assembly was produced using the Tree of Life pipeline from a specimen collected in Township: Sheringham, England, United Kingdom (Figure 1), as part of the Darwin Tree of Life project.

Methods

Sample acquisition and DNA barcoding

The specimen used for genome sequencing was an adult male *Brachypalpoides lentus* (specimen ID NHMUK015059424, ToLID idBraLent1; Figure 1), collected from Sheringham, England, United Kingdom (latitude 52.94, longitude 1.21) on 2022-07-07. The specimen was collected and identified by Ryan Mitchell. Sample metadata were collected in line with the Darwin Tree of Life project standards described by Lawniczak *et al.* (2022).

The initial identification was verified by an additional DNA barcoding process according to the framework developed by Twyford *et al.* (2024). A small sample was dissected from the specimen and stored in ethanol, while the remaining parts were shipped on dry ice to the Wellcome Sanger Institute (WSI) (see the protocol). The tissue was lysed, the COI marker region was amplified by PCR, and amplicons were sequenced and compared to the BOLD database, confirming the species identification (Crowley *et al.*, 2023). Following whole genome sequence generation, the relevant DNA barcode region was also used alongside the initial barcoding data for sample tracking at the WSI (Twyford *et al.*, 2024). The standard operating procedures for Darwin Tree of Life barcoding are available on protocols.io.

Figure 1. Photograph of the *Brachypalpoides lentus* (idBraLent1) 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 idBraLent1 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 in the WSI Scientific Operations core 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 manual 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 21.4 ng/μL and a yield of 1 005.80 ng, with a fragment size of 13.8 kb. The 260/280 spectrophotometric ratio was 1.88, and the 260/230 ratio was 1.78.

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 idBraLent1 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 and 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.

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 in Hi-C phasing mode (Cheng *et al.*, 2021; Cheng *et al.*, 2022), producing two haplotypes. Hi-C reads (Rao *et al.*, 2014) were mapped to the primary contigs using bwa-mem2 (Vasimuddin *et al.*, 2019). Contigs were further scaffolded with Hi-C data in YaHS (Zhou *et al.*, 2023), using 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), which runs MitoFinder (Allio *et al.*, 2020) and uses these annotations to select the final mitochondrial contig and to ensure the general quality of the sequence.

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 21 breaks and 186 joins. The curation process is documented 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 both 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 *Brachypalpoides lentus* specimen generated 111.90 Gb (gigabases) from 11.16 million reads, which were used to assemble the genome. GenomeScope2.0 analysis estimated the haploid genome size at 1 116.53 Mb, with a heterozygosity of 0.75% and repeat content of 42.34% (Figure 2). These estimates guided expectations for the assembly. Based on the estimated genome size, the sequencing data provided approximately 97× coverage. Hi-C sequencing produced 132.46 Gb from 877.25 million reads, which were used to scaffold the assembly. Table 1 summarises the specimen and sequencing details.

Assembly statistics

The genome was assembled into two haplotypes using Hi-C phasing. Haplotype 1 was curated to chromosome level, while

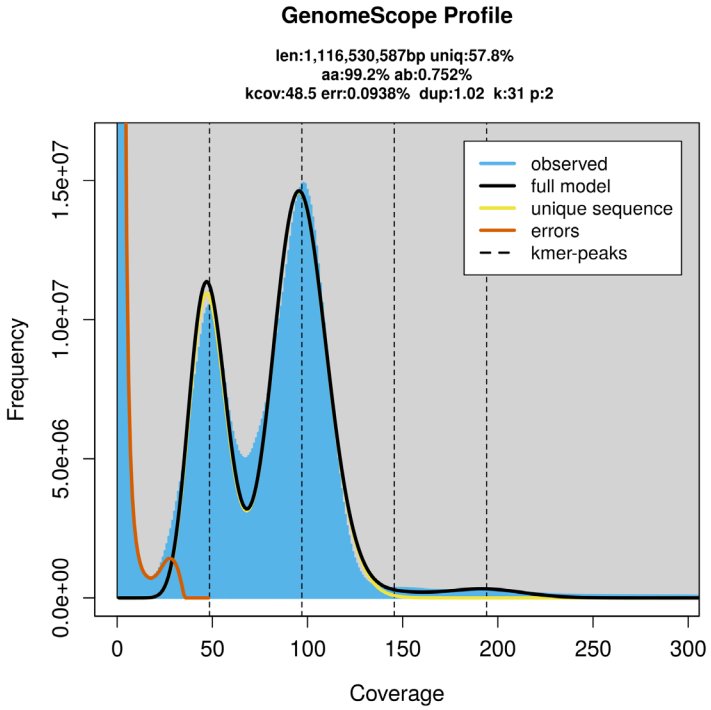

Figure 2. Frequency distribution of *k*-mers generated using GenomeScope2. The plot shows observed and modelled *k*-mer spectra, providing estimates of genome size, heterozygosity, and repeat content based on unassembled sequencing reads.

Table 1. Specimen and sequencing data for BioProject PRJEB77346.

Platform	PacBio HiFi	Hi-C
ToLID	idBraLent1	idBraLent1
Specimen ID	NHMUK015059424	NHMUK015059424
BioSample (source individual)	SAMEA112964232	SAMEA112964232
BioSample (tissue)	SAMEA112975397	SAMEA112975397
Tissue	whole organism	whole organism
Instrument	Sequel IIe	Illumina NovaSeq 6000
Run accessions	ERR13362618; ERR13362619	ERR13350340
Read count total	11.16 million	877.25 million
Base count total	111.90 Gb	132.46 Gb

haplotype 2 was assembled to scaffold level. The final assembly has a total length of 1 315.37 Mb in 813 scaffolds, with 547 gaps, and a scaffold N50 of 357.9 Mb (Table 2).

Most of the assembly sequence (95.66%) was assigned to 6 chromosomal-level scaffolds, representing 4 autosomes and the X and Y sex chromosomes. These chromosome-level scaffolds, confirmed by Hi-C data, are named according to size (Figure 3; Table 3). The order and orientation of contigs is uncertain in the following regions: Chromosome 2 (182–220 Mb); Chromosome 3 (42.5–52.5 Mb); Chromosome 4 (45.5–52 Mb); Chromosome Y (0–15 Mb and 25.5 Mb to the end).

Table 2. Genome assembly statistics.

Metric	Haplotype 1	Haplotype 2
Assembly name	idBraLent1.hap1.1	idBraLent1.hap2.1
Assembly accession	GCA_965362245.1	GCA_965362235.1
Assembly level	chromosome	scaffold
Span (Mb)	1 315.37	1 154.75
Number of chromosomes	6	N/A
Number of contigs	1 360	750
Contig N50	4.53 Mb	4.73 Mb
Number of scaffolds	813	317
Scaffold N50	357.9 Mb	344.07 Mb
Longest scaffold length (Mb)	383.99	N/A
Sex chromosomes	X and Y	N/A
Organelles	Mitochondrion: 15.88 kb	N/A

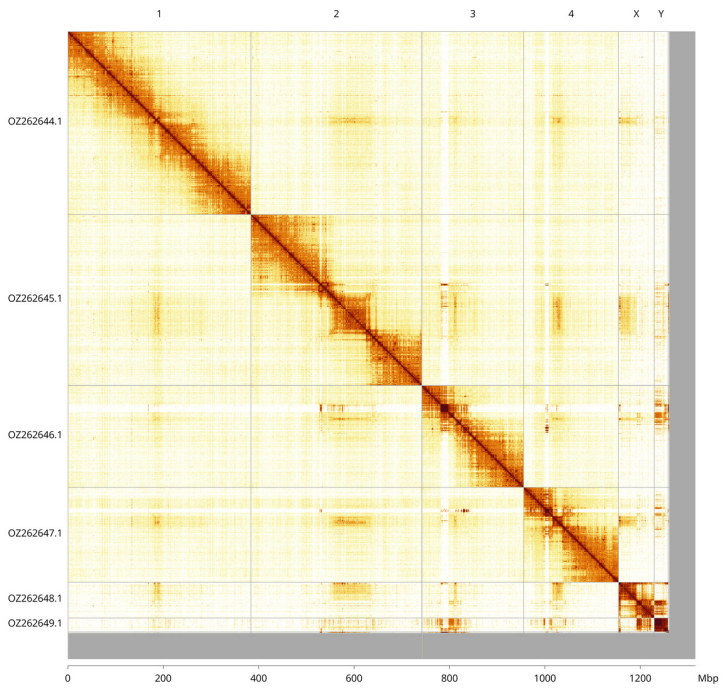

Figure 3. Hi-C contact map of the *Brachypalpoides lentus* genome assembly. Assembled chromosomes are shown in order of size and labelled along the axes, with a megabase scale shown below. The plot was generated using PretextSnapshot.

The mitochondrial genome was also assembled. This sequence is included as a contig in the multifasta file of the genome submission and as a standalone record.

For haplotype 1, the estimated QV is 62.8, and for haplotype 2, 64.2. When the two haplotypes are combined, the assembly achieves an estimated QV of 63.4. The *k*-mer completeness is 87.50% for haplotype 1, 82.86% for haplotype 2, and 99.27% for the combined haplotypes (Figure 4).

BUSCO analysis using the diptera_odb10 reference set (n = 3 285) identified 96.7% of the expected gene set (single = 94.9%, duplicated = 1.9%) for haplotype 1. The snail plot in Figure 5 summarises the scaffold length distribution and other assembly statistics for haplotype 1. The blob plot in Figure 6 shows the distribution of scaffolds by GC proportion and coverage for haplotype 1.

Table 4 lists the assembly metric benchmarks adapted from Rhie *et al.* (2021) the Earth BioGenome Project Report on Assembly Standards September 2024. The EBP metric, calculated for the haplotype 1, is **6.C.Q62**, 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

Table 3. Chromosomal pseudomolecules in the haplotype 1 genome assembly of *Brachypalpoides lentus* idBraLent1.

INSDC accession	Molecule	Length (Mb)	GC%
OZ262644.1	1	383.99	40.50
OZ262645.1	2	358.14	40.50
OZ262646.1	3	213.45	40
OZ262647.1	4	198.78	40.50
OZ262648.1	X	74.80	39
OZ262649.1	Y	29.15	38.50

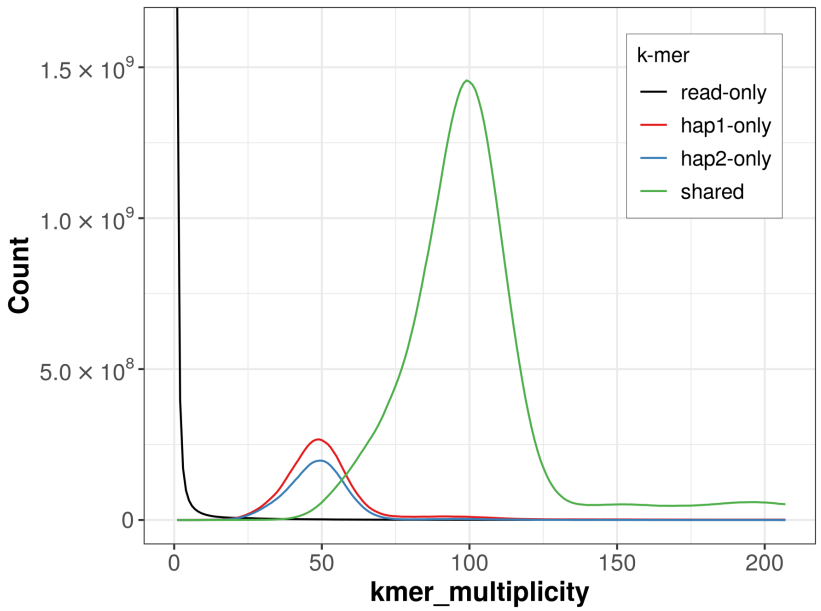

Figure 4. Evaluation of *k*-mer completeness using MerquryFK. This plot illustrates the recovery of *k*-mers from the original read data in the final assemblies. The horizontal axis represents *k*-mer multiplicity, and the vertical axis shows the number of *k*-mers. The black curve represents *k*-mers 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.

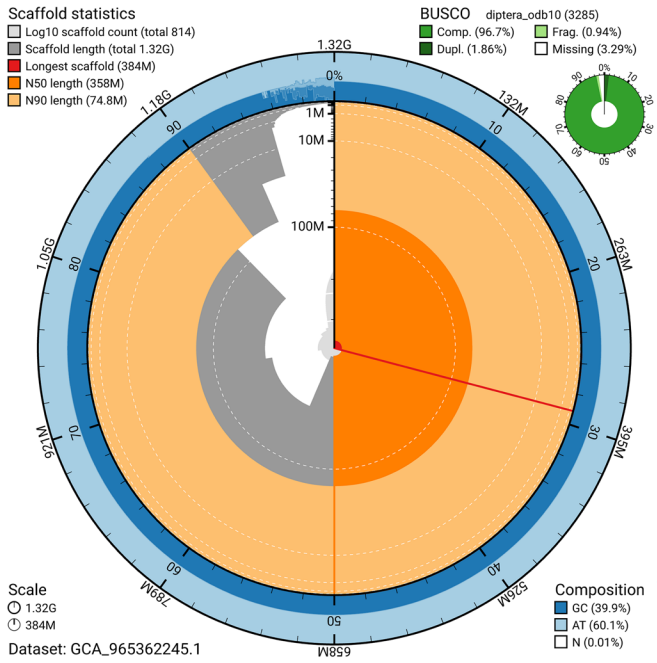

Figure 5. Assembly metrics for idBraLent1.hap1.1. The BlobToolKit snail plot provides an overview of assembly metrics and BUSCO gene completeness. The circumference represents the length of the whole genome sequence, and the main plot is divided into 1 000 bins around the circumference. The outermost 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 set is presented at the top right. An interactive version of this figure can be accessed on the BlobToolKit viewer.

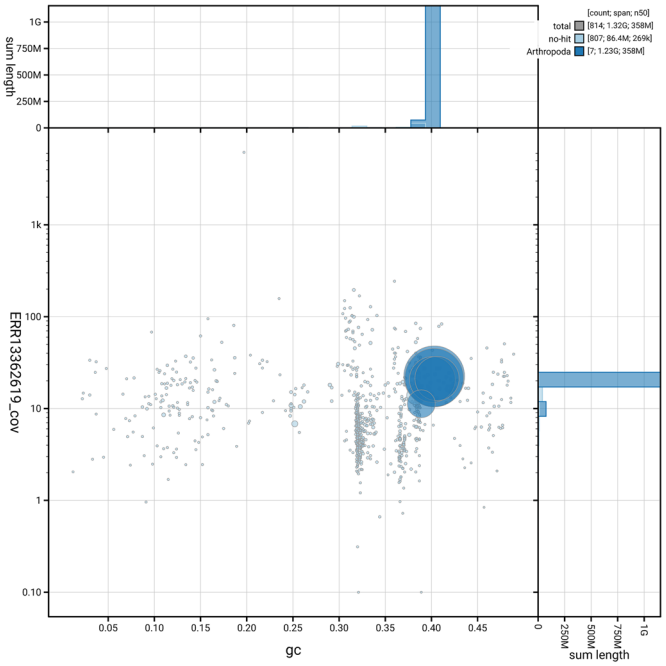

Figure 6. BlobToolKit GC-coverage plot for idBraLent1.hap1.1. Blob plot showing sequence coverage (vertical axis) and GC content (horizontal axis). The circles represent scaffolds, with the size proportional to scaffold length and the colour representing phylum membership. The histograms along the axes 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 *Brachypalpoides lentus* assembly.

Measure	Value	Benchmark
EBP summary (haplotype 1)	6.C.Q62	6.C.Q40
Contig N50 length	4.53 Mb	≥ 1 Mb
Scaffold N50 length	357.90 Mb	= chromosome N50
Consensus quality (QV)	Haplotype 1: 62.8; haplotype 2: 64.2; combined: 63.4	≥ 40
k-mer completeness	Haplotype 1: 87.50%; Haplotype 2: 82.86%; combined: 99.27%	≥ 95%
BUSCO	C:96.7% [S:94.9%; D:1.9%]; F:0.9%; M:2.3%; n:3 285	S > 90%; D < 5%
Percentage of assembly assigned to chromosomes	95.66%	≥ 90%

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 diligence is carried out proportionate to the nature of the materials themselves, and the circumstances under which they have been/are to be collected and provided for use. The purpose of this is to address and mitigate any potential legal and/or ethical implications of receipt and use of the materials as part of the research project, and to ensure that in doing so we align with best practice wherever possible. The overarching areas of consideration are:

- Ethical review of provenance and sourcing of the material
- Legality of collection, transfer and use (national and international)

Each transfer of samples is further undertaken according to a Research Collaboration Agreement or Material Transfer Agreement entered into by the Darwin Tree of Life Partner, Genome Research Limited (operating as the Wellcome Sanger Institute), and in some circumstances, other Darwin Tree of Life collaborators.

Data availability

European Nucleotide Archive: Brachypalpoides lentus. Accession number PRJEB77346. The genome sequence is released openly for reuse. The *Brachypalpoides lentus* genome sequencing initiative is part of the Darwin Tree of Life Project (PRJEB40665) and the 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 Darwin Tree of Life Barcoding collective
- 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

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	4.4.5	https://github.com/blobtoolkit/blobtoolkit
BUSCO	5.7.1	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	N/A	https://github.com/sanger-tol/PretextSnapshot
PretextView	0.2.5	https://github.com/sanger-tol/PretextView
samtools	1.21	https://github.com/samtools/samtools
sanger-tol/ascc	0.1.0	https://github.com/sanger-tol/ascc
sanger-tol/blobtoolkit	v0.7.1	https://github.com/sanger-tol/blobtoolkit
Seqtk	1.3	https://github.com/lh3/seqtk
Singularity	3.9.0	https://github.com/sylabs/singularity
TreeVal	1.2.0	https://github.com/sanger-tol/treeval
YaHS	1.2a.2	https://github.com/c-zhou/yahs

References

Allio R, Schomaker-Bastos A, Romiguier J, *et al.*: **MitoFinder: efficient automated large-scale extraction of mitogenomic data in target enrichment phylogenomics.** *Mol Ecol Resour.* 2020; **20**(4): 892–905.
PubMed Abstract | Publisher Full Text | Free Full Text

Altschul SF, Gish W, Miller W, *et al.*: **Basic Local Alignment Search Tool.** *J Mol Biol.* 1990; **215**(3): 403–410.
PubMed Abstract | Publisher Full Text

Ball S, Morris R: **Britain's hoverflies**. Princeton, New Jersey, USA: Princeton

University Press, 2015a.

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

Ball SG, Morris RKA: **A provisional Atlas of British hoverflies (Diptera, Syrphidae)**. Biological Records Centre, Centre for Ecology; Hydrology, 2000.
Reference Source

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

Cheng H, Jarvis ED, Fedrigo O, *et al.*: **Haplotype-resolved assembly of diploid genomes without parental data.** *Nat Biotechnol.* 2022; **40**(9): 1332–35.
PubMed Abstract | Publisher Full Text | Free Full Text

Crowley L, Allen H, Barnes I, *et al.*: **A sampling strategy for genome sequencing the British terrestrial arthropod fauna [version 1; peer review: 2 approved].** *Wellcome Open Res.* 2023; **8**: 123.
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

Lawniczak MKN, Davey RP, Rajan J, *et al.*: **Specimen and sample metadata standards for biodiversity genomics: a proposal from the Darwin Tree of Life project [version 1; peer review: 2 approved with reservations].** *Wellcome Open Res.* 2022; **7**: 187.
Publisher Full Text

Li H: **Minimap2: pairwise alignment for nucleotide sequences.** *Bioinformatics.* 2018; **34**(18): 3094–3100.
PubMed Abstract | Publisher Full Text | Free Full Text

Manni M, Berkeley MR, 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

Rotheray GE: **Colour guide to hoverfly larvae (Diptera, Syrphidae) in Britain and Europe.** *Dipterological Digest.* 1993; **9**: 1–156.
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

Stubbs AE, Falk SJ: **British hoverflies: an illustrated identification guide**. Reading, Berkshire: British Entomological; Natural History Society, 2002.
Reference Source

Twyford AD, Beasley J, Barnes I, *et al.*: **A DNA barcoding framework for taxonomic verification in the Darwin Tree of Life Project [version 1; peer review: 2 approved].** *Wellcome Open Res.* 2024; **9**: 339.
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

van Veen MP: **Hoverflies of Northwest Europe: identification keys to the syrphidae**. Leiden, The Netherlands: BRILL, 2010.
Reference Source

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

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 14 October 2025

https://doi.org/10.21956/wellcomeopenres.27376.r133662

✔ **Mihajla Djan**
University of Novi Sad, Novi Sad, Novi Sad, Serbia

The Data Note “The genome sequence of the hoverfly, *Brachypalpoides lentus* Meigen, 1822 (Diptera: Syrphidae)” gives the first genome assembly of the species and it is produced within the Darwin Tree of Life Project. Using Hi-C and PacBio HiFi long-read sequencing, authors generated a chromosome-level assembly comprising two haplotypes. The mitochondrial genome was also assembled. Authors used a wide spectrum of quality assessment tools in order to confirm accuracy of obtained genome sequence. All sequencing data and assembly have been deposited in the European Nucleotide Archive (PRJEB77346) and are openly available. The genome assembled may be used in further genomic studies of the species, as well as in evolutionary genomic research, taxonomy and molecular ecology analyses not only focused on this species, but including research of all hoverflies, as very important pollinator group.

The rationale for conducting the study is well justified in the Background section. Ecological importance of the species is highlighted, as well as their behavioral characteristics. The justification could be stronger if authors highlighted the potential for comparative genomics with other hoverfly genomes, but it is a minor comment and does not affect the clarity of the rationale. One typo is noticed in part “B. lentus also displaya behaviours”.

Used protocols are in the alignment with all recommendations and adopted strategies in generating high-quality reference genome. Those include Hi-C, PacBio HiFi long-read sequencing and wide quality control. The assembly showed 6 chromosomes, including X and Y. Methods section provided full information for DNA extraction, library preparation, and assembly workflows, providing study reproducibility.

The *Brachypalpoides lentus* genome is deposited in the European Nucleotide Archive, Accession number PRJEB77346 and the genome sequence is released openly for reuse. All raw sequence data and the assembly have been deposited in INSDC databases. Assembly metrics and QC plots are clearly presented in Tables and Figures 2–6. The presentation meets open-data standards. Enough information is provided for study replication and for downstream use of generated data.

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, conservation genetic, molecular taxonomy, phylogenetics

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 October 2025

https://doi.org/10.21956/wellcomeopenres.27376.r133660

Lore Esselens
Biology, Royal Museum for Central Africa, Tervuren, Belgium

Mitchell and Cockwood present a clear and well-written manuscript describing the first high-quality genome assembly of *Brachypalpoides lentus*, produced within the Darwin Tree of Life project. While the paper follows the standard DToL genome note format, it represents substantial work and careful attention to quality at every stage. The assembly spans 1 315.37 Mb (95.66%) with haplotype 1 scaffolded into six chromosomes, including the X and Y sex chromosomes, and haplotype 2 assembled to scaffold level. The manuscript provides comprehensive methodological detail, including PacBio HiFi sequencing, Hi-C scaffolding, detailed assembly curation, and multiple quality assessments such as BUSCO and Merqury. The data are clearly presented resulting in a valuable genomic resource to support future research on *Brachypalpoides lentus* and hoverfly biology more broadly. The manuscript would be ready for indexing after only very minor revisions.

In the Keywords section: *Brachypalpoides lentus* should be in italic.

In the Background section: “Notably, *B. lentus* also displaya” should be ‘displays’.

In the Assembly statistics section: the sentence “Table 4 lists the assembly metric benchmarks adapted from Rhie et al. (2021) the Earth BioGenome Project Report on Assembly Standards (September 2024).” seems to miss a connector such as “and” between the two sources. I suggest rephrasing to something like: “Table 4 lists the assembly metric benchmarks adapted from Rhie et al. (2021) and the Earth BioGenome Project Report on Assembly Standards (September 2024).” for clarity.

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: Evolutionary genomics

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