DATA NOTE

The genome sequence of the bisetose emerald-bottle, *Bellardia pandia* (Walker, 1849) [version 1; peer review: awaiting peer review]

Steven Falk¹, 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

¹Independent Researcher, Kenilworth, UK

**Abstract**

We present a genome assembly from an individual female *Bellardia pandia* (the bisetose emerald-bottle; Arthropoda; Insecta; Diptera; Calliphoridae). The genome sequence is 617 megabases in span. The majority of the assembly (97.82%) is scaffolded into six chromosomal pseudomolecules, with the X sex chromosome assembled.

**Keywords**

Bellardia pandia, bisetose emerald-bottle, 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: Falk S: Investigation, Resources, Writing – Original Draft Preparation;

Competing interests: No competing interests were disclosed.

Grant information: This work was supported by Wellcome through core funding to the Wellcome Sanger Institute (206194) and the Darwin Tree of Life Discretionary Award (218328).
The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Copyright: © 2021 Falk S 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: Falk S, University of Oxford and Wytham Woods Genome Acquisition Lab, Darwin Tree of Life Barcoding collective et al. The genome sequence of the bisetose emerald-bottle, Bellardia pandia (Walker, 1849) [version 1; peer review: awaiting peer review] Wellcome Open Research 2021, 6:344 https://doi.org/10.12688/wellcomeopenres.17492.1

First published: 14 Dec 2021, 6:344 https://doi.org/10.12688/wellcomeopenres.17492.1
Species taxonomy
Eukaryota; Metazoa; Ecdysozoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Oestroidea; Calliphoridae; Calliphorinae; Bellardia; Bellardia pandia (Meigen, 1822) (NCBI:txid2795671).

Background
The presence of two long posteroventral bristles on the front tibiae of Bellardia pandia (previously known as Onesia biseta) distinguish this medium-sized blowfly from the similar-looking B. viarum and B. vulgaris. All three have green, turquoise or bronze reflections on the tergites, tempered by dusting that produces a tessellated pattern. The thorax is grey-dusted with several darker stripes and has faint metallic-green reflections in some lights. The male genitalia are distinctive and are the most reliable character for separating the three species.

This species occurs locally throughout Great Britain, usually in the vicinity of wetlands. Females are viviparous and larvae are thought to be predators/parasitoids of earthworms.

Genome sequence report
The genome was sequenced from a single female B. pandia (Figure 1) collected from Wytham Woods, Oxfordshire, UK (latitude 51.769, longitude -1.339). A total of 54-fold coverage in Pacific Biosciences single-molecule long reads and 76-fold coverage in 10X Genomics read clouds were generated. Primary assembly contigs were scaffolded with chromosome conformation Hi-C data. Manual assembly curation corrected 24 missing/misjoins, reducing the scaffold number by 14.94%, and increasing the scaffold N50 by 8.33%.

The final assembly has a total length of 617 Mb in 74 sequence scaffolds with a scaffold N50 of 111.8 Mb (Table 1). The majority, 97.82%, 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 1. Genome data for Bellardia pandia, idBelPand1.1.

| Project accession data | Assembly identifier | idBelPand1.1 |
|------------------------|---------------------|--------------|
| Species                | Bellardia pandia    |
| Specimen               | idBelPand1          |
| NCBI taxonomy ID       | 2795671             |
| BioProject             | PRJEB45671          |
| BioSample ID           | SAMEA7746587        |
| Isolate information    | Female, whole organism |

| Raw data accessions    | PacificBiosciences SEQUEL II | ERR6939221, ERR6939222 |
|                       | 10X Genomics Illumina        | ERR6688411-ERR6688414 |
|                       | Hi-C Illumina               | ERR6688410             |

| Genome assembly        | Assembly accession          | GCA_916048285.1 |
|                       | Accession of alternate haplotype | GCA_916050155.1 |
|                       | Span (Mb)                    | 617 |
|                       | Number of contigs            | 135 |
|                       | Contig N50 length (Mb)       | 31.8 |
|                       | Number of scaffolds          | 74 |
|                       | Scaffold N50 length (Mb)     | 111.8 |
|                       | Longest scaffold (Mb)        | 143.1 |

| BUSCO* genome score   | C:99.1%, S:97.9%, D:1.2%, F:0.3%, M:0.6%, 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/idBelPand1.1/dataset/CAJZHU01/busc.

Methods
Sample acquisition and nucleic acid extraction
A female B. pandia (idBelPand1) was collected from Wytham Woods, Oxfordshire, UK (latitude 51.77, longitude -1.331) by Steven Falk, independent researcher, who also identified the specimens. The specimens were collected using a net and snap-frozen on dry ice.

Figure 1. Image of the idBelPand1 specimen used for sequencing, taken during preservation ad processing.
DNA was extracted at the Tree of Life laboratory, Wellcome Sanger Institute. The idBelPand1 sample was weighed and dissected on dry ice with tissue set aside for Hi-C sequencing. Tissue from the whole organism was disrupted using a Nippi Powermasher fitted with a BioMasher pestle. Fragment size analysis of 0.01-0.5 ng of DNA was then performed using an Agilent FemtoPulse. High molecular weight (HMW) DNA was extracted using the Qiagen MagAttract HMW DNA extraction kit. Low molecular weight DNA was removed from a 200-ng aliquot of extracted DNA using 0.8X AMpure XP purification kit.
prior to 10X Chromium sequencing; a minimum of 50 ng DNA was submitted for 10X sequencing. HMW DNA was sheared into an average fragment size between 12-20 kb in a Megaruptor 3 system with speed setting 30. Sheared DNA was purified by solid-phase reversible immobilisation using AMPure PB beads with a 1.8X ratio of beads to sample to remove the shorter fragments and concentrate the DNA sample. The concentration of the sheared and purified DNA was assessed using a Nanodrop spectrophotometer and Qubit Fluorometer and Qubit dsDNA High Sensitivity Assay kit. Fragment size distribution was evaluated by running the sample on the FemtoPulse system.

Figure 3. Genome assembly of *Bellardia pandia*, idBelPand1.1: 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/idBelPand1.1/dataset/CAJZHU01/blob.
Sequencing
Pacific Biosciences HiFi circular consensus and 10X Genomics Chromium read cloud sequencing libraries were constructed according to the manufacturers’ instructions. Sequencing was performed by the Scientific Operations core at the Wellcome Sanger Institute on Pacific Biosciences SEQUEL II and Illumina NovaSeq 6000 instruments. Hi-C data were generated from remaining whole organism tissue using the Arima Hi-C+ kit and sequenced on a NovaSeq 6000 instrument.

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). One round of polishing was

Figure 4. Genome assembly of Bellardia pandia, idBelPand1.1: 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/idBelPand1.1/dataset/CAJZHU01/cumulative.
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 as described previously (Howe et al., 2021). Manual curation (Howe et al., 2021) was performed using HiGlass (Kerpedjiev et al., 2018) and Pretext. The mitochondrial genome was assembled using MitoHiFi (Uliano-Silva et al., 2021), which performed annotation using MitoFinder (Allio et al., 2020). 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.

Ethics/compliance issues
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. 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. 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.

Table 2. Chromosomal pseudomolecules in the genome assembly of Bellardia pandia, idBelPand1.1.

| INSDC accession | Chromosome | Size (Mb) | GC% |
|-----------------|------------|-----------|-----|
| OU696529.1      | 1          | 143.14    | 30.1|
| OU696530.1      | 2          | 113.42    | 30.3|
| OU696531.1      | 3          | 111.82    | 30.0|
| OU696532.1      | 4          | 110.19    | 30.2|
| OU696533.1      | 5          | 105.67    | 30.3|
| OU696534.1      | X          | 19.54     | 33.1|
| OU696535.1      | MT         | 0.02      | 19.0|
| -               | Unplaced   | 13.40     | 31.5|
Table 3. Software tools used.

| Software tool        | Version | Source                                                                 |
|----------------------|---------|------------------------------------------------------------------------|
| Hifiasm              | 0.15    | 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/genome-exome/software/pipelines/latest/advanced/other-pipelines |
| freebayes            | 1.3.1-17-gaa2ace8 | Garrison & Marth, 2012                                |
| MitoHiFi             | 2.0     | Uliano-Silva et al., 2021                                             |
| gEVAL                | N/A     | Chow et al., 2016                                                    |
| HiGlass              | 1.11.6  | Kerpedjiev et al., 2018                                               |
| PretextView          | 0.2.x   | https://github.com/wtsi-hpag/PretextView                              |
| BlobToolKit          | 2.6.2   | Challis et al., 2020                                                  |

Data availability

European Nucleotide Archive: Bellardia pandia (bisetose emerald-bottle). Accession number PRJEB45671: https://www.ebi.ac.uk/ena/browser/view/PRJEB45671.

The genome sequence is released openly for reuse. The B. pandia 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 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.5746938.

Members of the Darwin Tree of Life Barcoding collective are listed here: https://doi.org/10.5281/zenodo.5744972.

Members of the Wellcome Sanger Institute Tree of Life programme are listed here: https://doi.org/10.5281/zenodo.5744840.

Members of the Wellcome Sanger Institute Scientific Operations: DNA Pipelines collective are listed here: https://doi.org/10.5281/zenodo.5746904.

Members of the Tree of Life Core Informatics collective are listed here: https://doi.org/10.5281/zenodo.5743293.

Members of the Darwin Tree of Life Consortium are listed here: https://doi.org/10.5281/zenodo.5638618.

Acknowledgements

The Background section has been adapted from Steven Falk’s flickr album.

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

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

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

Ghurye J, Rhee A, Walenz BP, et al.: Integrating Hi-C Links with Assembly Graphs for Chromosome-Scale Assembly. PLoS Comput Biol. 2019; 15(8): e1007273. PubMed Abstract | Publisher Full Text | Free Full Text

Page 8 of 9
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

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, and Viral Genomes.** Mol Biol Evol. 2021; 38(10): 4647–54. 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; 158(7): 1665–80. PubMed Abstract | Publisher Full Text | Free Full Text

Uliano-Silva M, Nunes JGF, Krasheninnikova K, et al.: **marcelauliano/MitoHiFi:** mitohifi_v2.0. 2021. Publisher Full Text

Kerpedjiev P, Abdennur N, Lekschas F, et al.: **HiGlass: Web-Based Visual Exploration and Analysis of Genome Interaction Maps.** Genome Biol. 2018; 19(1): 125. PubMed Abstract | Publisher Full Text | Free Full Text

Manni M, Berkeley MR, Seppey M, et al.: **BUSCO Update: Novel and Streamlined Workflows along with Broader and Deeper Phylogenetic Coverage for Scoring of Eukaryotic, and Viral Genomes.** Mol Biol Evol. 2021; 38(10): 4647–54. 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; 158(7): 1665–80. PubMed Abstract | Publisher Full Text | Free Full Text

Uliano-Silva M, Nunes JGF, Krasheninnikova K, et al.: **marcelauliano/MitoHiFi:** mitohifi_v2.0. 2021. Publisher Full Text