DATA NOTE

The genome sequence of Trimmer's Mining Bee, *Andrena trimmerana* (Kirby, 1802)

[version 1; peer review: 2 approved, 1 approved with reservations]

Ellen Baker[1], Liam M. Crowley[1], Steven Falk[2],
University of Oxford and Wytham Woods 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]University of Oxford, Oxford, England, UK
[2]Independent researcher, Kenilworth, England, UK

v1 **First published:** 22 Nov 2024, **9**:691
https://doi.org/10.12688/wellcomeopenres.23370.1
Latest published: 22 Nov 2024, **9**:691
https://doi.org/10.12688/wellcomeopenres.23370.1

Abstract

We present a genome assembly from afemale specimen of *Andrena trimmerana* (Trimmer's Mining Bee; Arthropoda; Insecta; Hymenoptera; Andrenidae). The genome sequence has a total length of 399.10 megabases. Most of the assembly (86.18%) is scaffolded into 5 chromosomal pseudomolecules. The mitochondrial genome has also been assembled and is 19.77 kilobases in length. Gene annotation of this assembly on Ensembl identified 10,570 protein-coding genes.

Keywords

Andrena trimmerana, Trimmer's Mining Bee, genome sequence, chromosomal, Hymenoptera

This article is included in the Tree of Life gateway.

Open Peer Review

Approval Status ✓ ✓ ?

	1	2	3
version 1 22 Nov 2024	✓ view	✓ view	? view

1. **Rodolpho S. T. Menezes**, State University of Santa Cruz, Ilhéus, Brazil
2. **Liang Lu**, Hebei Normal University, Shijiazhuang, China
3. **Josephine R Paris**, University of Exeter Library, Devon, UK

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: Baker E: Investigation, Resources; **Crowley LM**: Investigation, Resources; **Falk S**: Investigation, Resources;

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: Baker E, Crowley LM, Falk S *et al.* **The genome sequence of Trimmer's Mining Bee, *Andrena trimmerana* (Kirby, 1802) [version 1; peer review: 2 approved, 1 approved with reservations]** Wellcome Open Research 2024, **9**:691 https://doi.org/10.12688/wellcomeopenres.23370.1

First published: 22 Nov 2024, **9**:691 https://doi.org/10.12688/wellcomeopenres.23370.1

Species taxonomy

Eukaryota; Opisthokonta; Metazoa; Eumetazoa; Bilateria; Protostomia; Ecdysozoa; Panarthropoda; Arthropoda; Mandibulata; Pancrustacea; Hexapoda; Insecta; Dicondylia; Pterygota; Neoptera; Endopterygota; Hymenoptera; Apocrita; Aculeata; Apoidea; Anthophila; Andrenidae; Andreninae; *Andrena*; *Hoplandrena; Andrena trimmerana* (Kirby, 1802) (NCBI: txid1431430).

Background

Andrena trimmerana, or Trimmer's Mining Bee, is a solitary bee species within the family Andrenidae. It is widely distributed across western Europe, including Britain and Ireland, and extends into parts of western Asia and North Africa (Else, 2023; GBIF Secretariat, 2024). In the UK, this species is found from East Kent to Cornwall and northwards to the Midlands, with isolated records in the Lake District. It is generally absent from Scotland (NBN Atlas Partnership, 2024). However, some recorders aggregate sightings of *A. trimmerana* and *A. scotica* due to the difficulty in distinguishing them (Else, 2023).

Andrena trimmerana inhabits a variety of environments such as coastal cliffs, open woodlands, heathlands, chalk grasslands, fens, commons, and gardens (Else, 2023). The species prefers well-drained soils and typically nests individually in soil banks and bare slopes, rather than forming large nesting aggregations. The species is polylectic and visits willow, dandelion, gorse, bramble, blackthorn, alexanders, and rhododendron (Else, 2023).

The genus *Andrena* is taxonomically challenging because of the morphological similarity across species and the variation between broods in bivoltine species (Wood *et al.*, 2022). *A. trimmerana* typifies this problem. In addition to being difficult to separate from *A. scotica*, it it is bivoltine, with morphologically distinct spring and summer broods, particularly in males. The broods have previously been classified as different species and there is a possibility that they may be (Else, 2023).

An extensive DNA barcode survey of wild urban bees in France showed that *Andrena trimmerana*, *A. carantonica*, *A. scotica* and *A. spinigera* share a Barcode Index Number (BIN), suggesting the need for taxonomic revision to establish the status of these species (Villalta *et al.*, 2021). Documenting the genome sequence of *A. trimmerana* within the Darwin Tree of Life project supports accurate species identification and improves understanding of solitary bee ecology, aiding pollinator conservation across the UK and Ireland.

Genome sequence report

The genome of *Andrena trimmerana* (Figure 1) was sequenced using Pacific Biosciences single-molecule HiFi long reads, generating a total of 23.95 Gb (gigabases) from 2.30 million reads, providing an estimated 61-fold coverage. Primary assembly contigs were scaffolded with chromosome conformation Hi-C data, which produced 107.78 Gb from 713.76 million reads. Specimen and sequencing details are summarised in Table 1.

Figure 1. Photograph of the *Andrena trimmerana* (iyAndTrim1) specimen used for genome sequencing.

Assembly errors were corrected by manual curation, including 15 missing joins or mis-joins and haplotypic duplications. This reduced the scaffold number by 5.06%, and increased the scaffold N50 by 48.51%. The final assembly has a total length of 399.10 Mb in 224 sequence scaffolds, with 27 gaps, and a scaffold N50 of 91.9 Mb (Table 2). The snail plot in Figure 2 provides a summary of the assembly statistics, indicating the distribution of scaffold lengths and other assembly metrics. Figure 3 shows the distribution of scaffolds by GC proportion and coverage. Figure 4 presents a cumulative assembly plot, with separate curves representing different scaffold subsets assigned to various phyla, illustrating the completeness of the assembly.

Most of the assembly sequence (86.18%) was assigned to 5 chromosomal-level scaffolds. These chromosome-level scaffolds, confirmed by the Hi-C data, are named in order of size (Figure 5; Table 3).

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 65.2 with *k*-mer completeness of 100.0%, and the assembly has a BUSCO v5.3.2 completeness of 96.9% (single = 96.5%, duplicated = 0.4%), using the hymenoptera_odb10 reference set (n = 5,991). The assembly achieves the EBP reference standard of 6.7.65.2. Other quality metrics are given in Table 2.

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/1431430.

Genome annotation report

The *Andrena trimmerana* genome assembly (GCA_951215215.1) was annotated at the European Bioinformatics Institute (EBI) on Ensembl Rapid Release. The resulting annotation includes

Table 1. Specimen and sequencing data for *Andrena trimmerana*.

Project information			
Study title	Andrena trimmerana (trimmer's mining bee)		
Umbrella BioProject	PRJEB56904		
Species	*Andrena trimmerana*		
BioSample	SAMEA10107041		
NCBI taxonomy ID	1431430		
Specimen information			
Technology	**ToLID**	**BioSample accession**	**Organism part**
PacBio long read sequencing	iyAndTrim1	SAMEA10200733	thorax
Hi-C sequencing	iyAndTrim2	SAMEA10201024	head
RNA sequencing	iyAndTrim1	SAMEA10200734	abdomen
Sequencing information			
Platform	**Run accession**	**Read count**	**Base count (Gb)**
Hi-C Illumina NovaSeq 6000	ERR10395991	7.14e+08	107.78
PacBio Sequel IIe	ERR10439742	2.30e+06	23.95
RNA Illumina NovaSeq 6000	ERR10395992	7.77e+07	11.74

18,433 transcribed mRNAs from 10,570 protein-coding and 1,621 non-coding genes (Table 2; https://rapid.ensembl.org/Andrena_trimmerana_GCA_951215215.1/Info/Index). The average transcript length is 12,564.09. There are 1.51 coding transcripts per gene and 6.57 exons per transcript.

Methods

Sample acquisition and DNA barcoding

A female adult *Andrena trimmerana* (specimen ID Ox001117, ToLID iyAndTrim1) was collected from Wytham Woods, Oxfordshire, UK (latitude 51.78, longitude –1.32) on 2021-04-01 by netting. The specimen was collected by Ellen Baker (University of Oxford), identified by Liam Crowley (University of Oxford) and then preserved on dry ice.

The specimen used for Hi-C sequencing (specimen ID Ox001294, ToLID iyAndTrim2) was an adult specimen, netted in the same location on 2021-04-23. The specimen was collected and identified by Steven Falk (independent researcher) and preserved on dry ice.

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 specimens and stored in ethanol, while the remaining parts were shipped on dry ice to the Wellcome Sanger Institute (WSI). 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 have been deposited on protocols.io (Beasley *et al.*, 2023).

Nucleic acid extraction

The workflow for high molecular weight (HMW) DNA extraction at the Wellcome Sanger Institute (WSI) Tree of Life Core Laboratory includes a sequence of procedures: sample preparation and homogenisation, DNA extraction, fragmentation and purification. Detailed protocols are available on protocols.io (Denton *et al.*, 2023b). In sample preparation, the iyAndTrim1 sample was weighed and dissected on dry ice (Jay *et al.*, 2023). Tissue from the thorax was homogenised using a PowerMasher II tissue disruptor (Denton *et al.*, 2023a). HMW DNA was extracted in the WSI Scientific Operations core using the Automated MagAttract v2 protocol (Oatley *et al.*, 2023). The DNA was sheared into an average fragment size of 12–20 kb in a Megaruptor 3 system (Bates *et al.*, 2023). Sheared DNA was purified by solid-phase reversible immobilisation, using AMPure PB beads to sample to eliminate shorter fragments and concentrate the DNA (Strickland *et al.*, 2023). The concentration of the sheared and purified DNA

Table 2. Genome assembly data for *Andrena trimmerana*, iyAndTrim1.1.

Genome assembly		
Assembly name	iyAndTrim1.1	
Assembly accession	GCA_951215215.1	
Accession of alternate haplotype	*GCA_951212825.1*	
Span (Mb)	399.10	
Number of contigs	252	
Number of scaffolds	224	
Longest scaffold (Mb)	142.9	
Assembly metrics*		***Benchmark***
Contig N50 length (Mb)	17.8	*≥ 1 Mb*
Scaffold N50 length (Mb)	91.9	*= chromosome N50*
Consensus quality (QV)	65.2	*≥ 40*
k-mer completeness	100.0%	*≥ 95%*
BUSCO v5.4.3 lineage: hymenoptera_odb10	C:96.9%[S:96.5%,D:0.4%], F:0.6%,M:2.5%,n:5991	*S > 90%, D < 5%*
Percentage of assembly mapped to chromosomes	86.18%	*≥ 90%*
Sex chromosomes	None	*localised homologous pairs*
Organelles	Mitochondrial genome: 19.77 kb	*complete single alleles*
Genome annotation of assembly GCA_951215215.1 at Ensembl		
Number of protein-coding genes	10,570	
Number of non-coding genes	1,621	
Number of gene transcripts	18,433	

* Assembly metric benchmarks are adapted from Rhie *et al.* (2021) and the Earth BioGenome Project Report on Assembly Standards September 2024.

** BUSCO scores based on the hymenoptera_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/Andrena%20trimmerana/dataset/iyAndTrim1_1/busco.

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.

RNA was extracted from abdomen tissue of iyAndTrim1 in the Tree of Life Laboratory at the WSI using the RNA Extraction: Automated MagMax™ *mir*Vana protocol (do Amaral *et al.*, 2023). The RNA concentration was assessed using a Nanodrop spectrophotometer and a Qubit Fluorometer using the Qubit RNA Broad-Range Assay kit. Analysis of the integrity of the RNA was done using the Agilent RNA 6000 Pico Kit and Eukaryotic Total RNA assay.

Sequencing

Pacific Biosciences HiFi circular consensus DNA 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. DNA and RNA sequencing was performed by the Scientific Operations core at the WSI on Pacific Biosciences Sequel IIe (HiFi) and Illumina NovaSeq 6000 (RNA-Seq) instruments.

Hi-C data were generated from frozen whole organism tissue of iyAndTrim2, using the Arima-HiC v2 kit. The tissue was fixed with a TC buffer containing formaldehyde, resulting in crosslinked DNA. The crosslinked DNA was digested

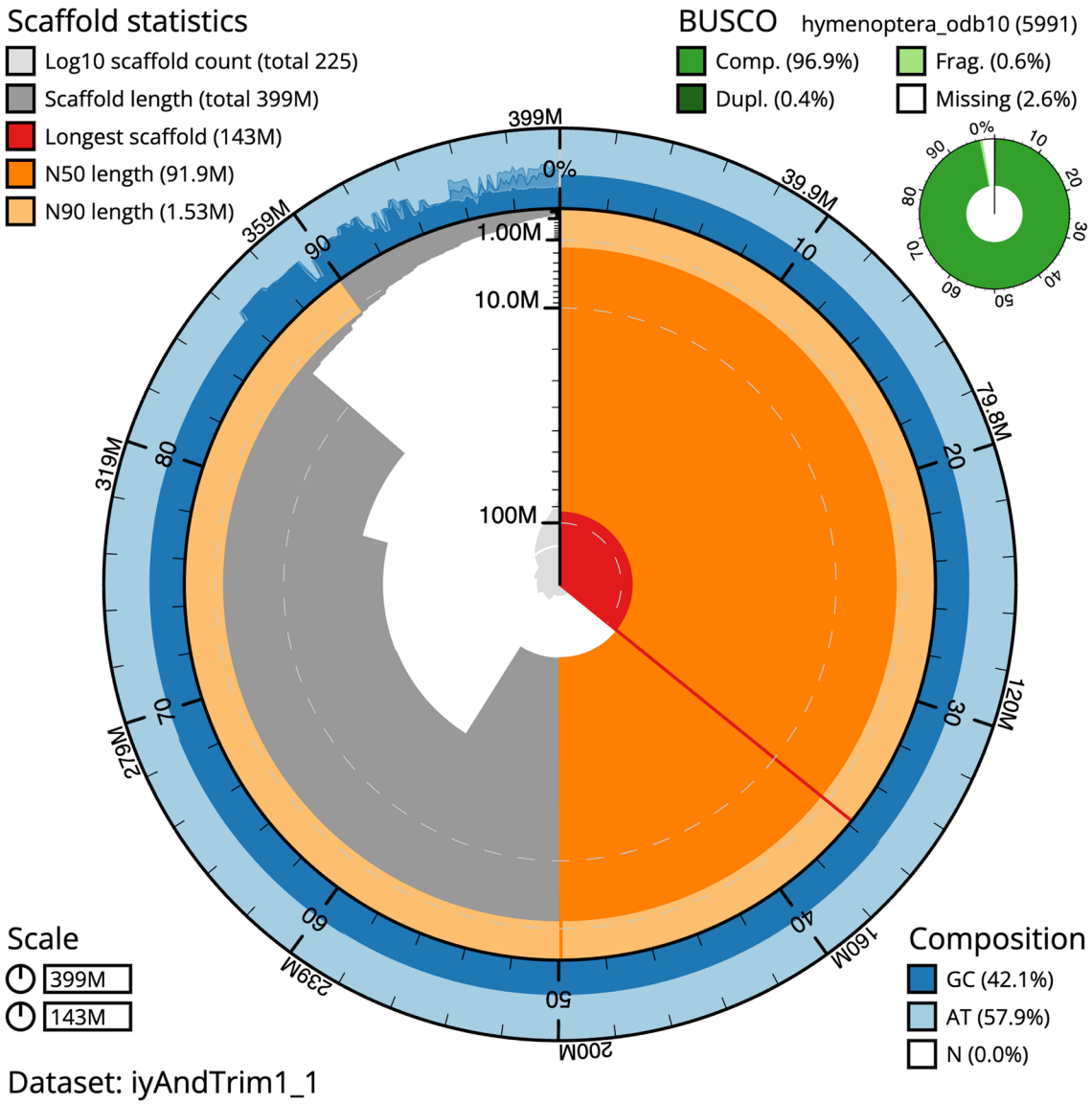

Figure 2. Genome assembly of *Andrena trimmerana*, iyAndTrim1.1: metrics. The BlobToolKit snail plot shows N50 metrics and BUSCO gene completeness. The main plot is divided into 1,000 bins around the circumference with each bin representing 0.1% of the 399,083,795 bp assembly. The distribution of scaffold lengths is shown in dark grey with the plot radius scaled to the longest scaffold present in the assembly (142,895,829 bp, shown in red). Orange and pale-orange arcs show the N50 and N90 scaffold lengths (91,870,870 and 1,525,000 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 hymenoptera_odb10 set is shown in the top right. An interactive version of this figure is available at https://blobtoolkit.genomehubs.org/view/Andrena%20trimmerana/dataset/iyAndTrim1_1/snail.

with a restriction enzyme master mix. The resulting 5'-overhangs were filled in and labelled with a biotinylated nucleotide. The biotinylated DNA was then fragmented, enriched, barcoded, and amplified using the NEBNext Ultra II DNA Library Prep Kit. Hi-C sequencing was performed on an Illumina NovaSeq 6000 instrument, using paired-end sequencing with a read length of 150 bp.

Genome assembly and curation

The HiFi reads were first assembled using Hifiasm (Cheng *et al.*, 2021) with the --primary option. Haplotypic duplications were identified and removed using purge_dups (Guan *et al.*, 2020). The Hi-C reads were mapped to the primary contigs using bwa-mem2 (Vasimuddin *et al.*, 2019). The contigs were further scaffolded using the provided Hi-C data (Rao *et al.*, 2014)

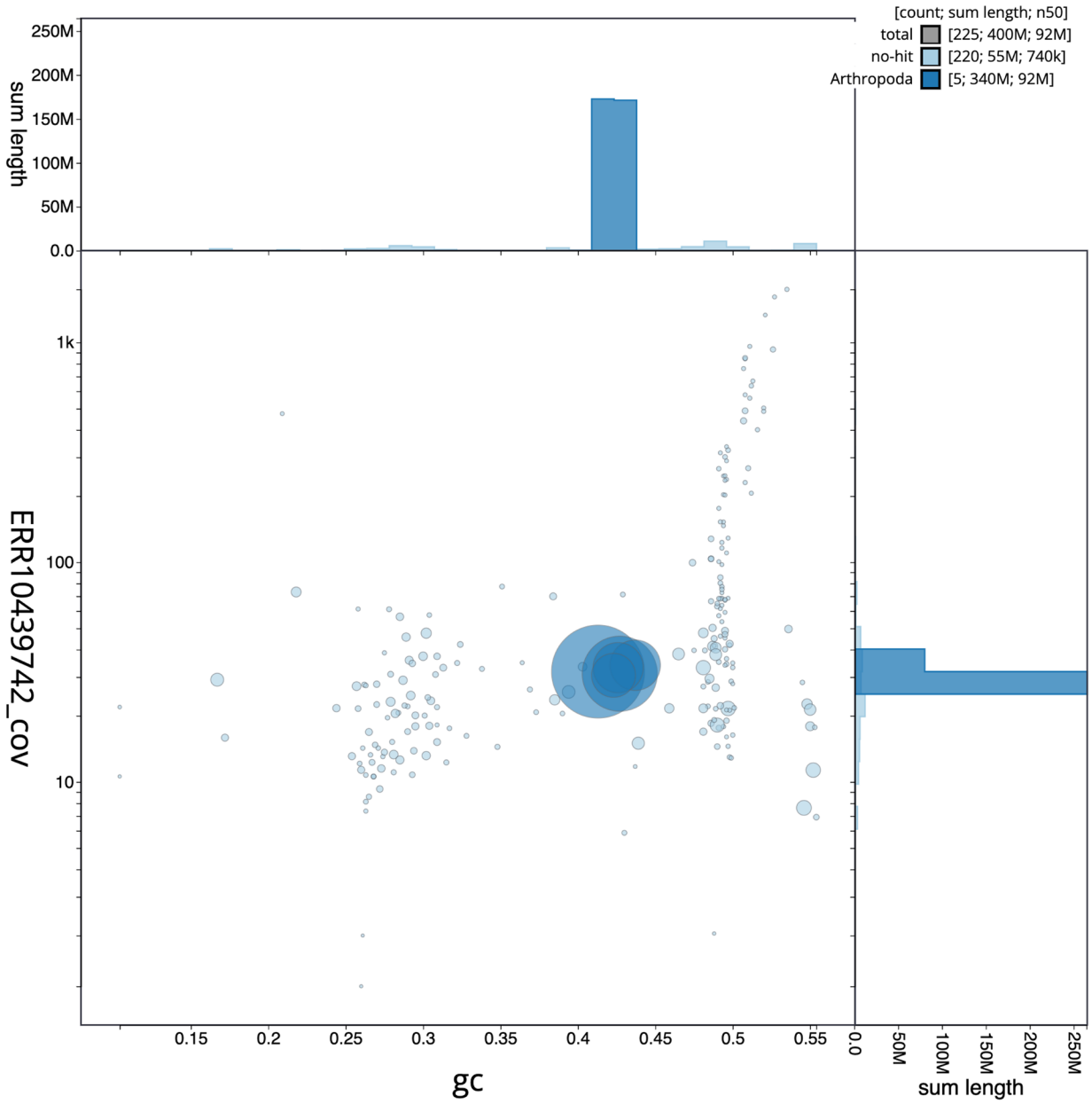

Figure 3. Genome assembly of *Andrena trimmerana*, iyAndTrim1.1: BlobToolKit GC-coverage plot. Sequences are coloured by phylum. Circles are sized in proportion to sequence length. Histograms show the distribution of sequence length sum along each axis. An interactive version of this figure is available at https://blobtoolkit.genomehubs.org/view/Andrena%20trimmerana/dataset/iyAndTrim1_1/blob.

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.

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 (Kerpedjiev *et al.*, 2018) and PretextView (Harry, 2022).

Evaluation of final assembly

A Hi-C map for the final assembly was produced using bwa-mem2 (Vasimuddin *et al.*, 2019) in the Cooler file format

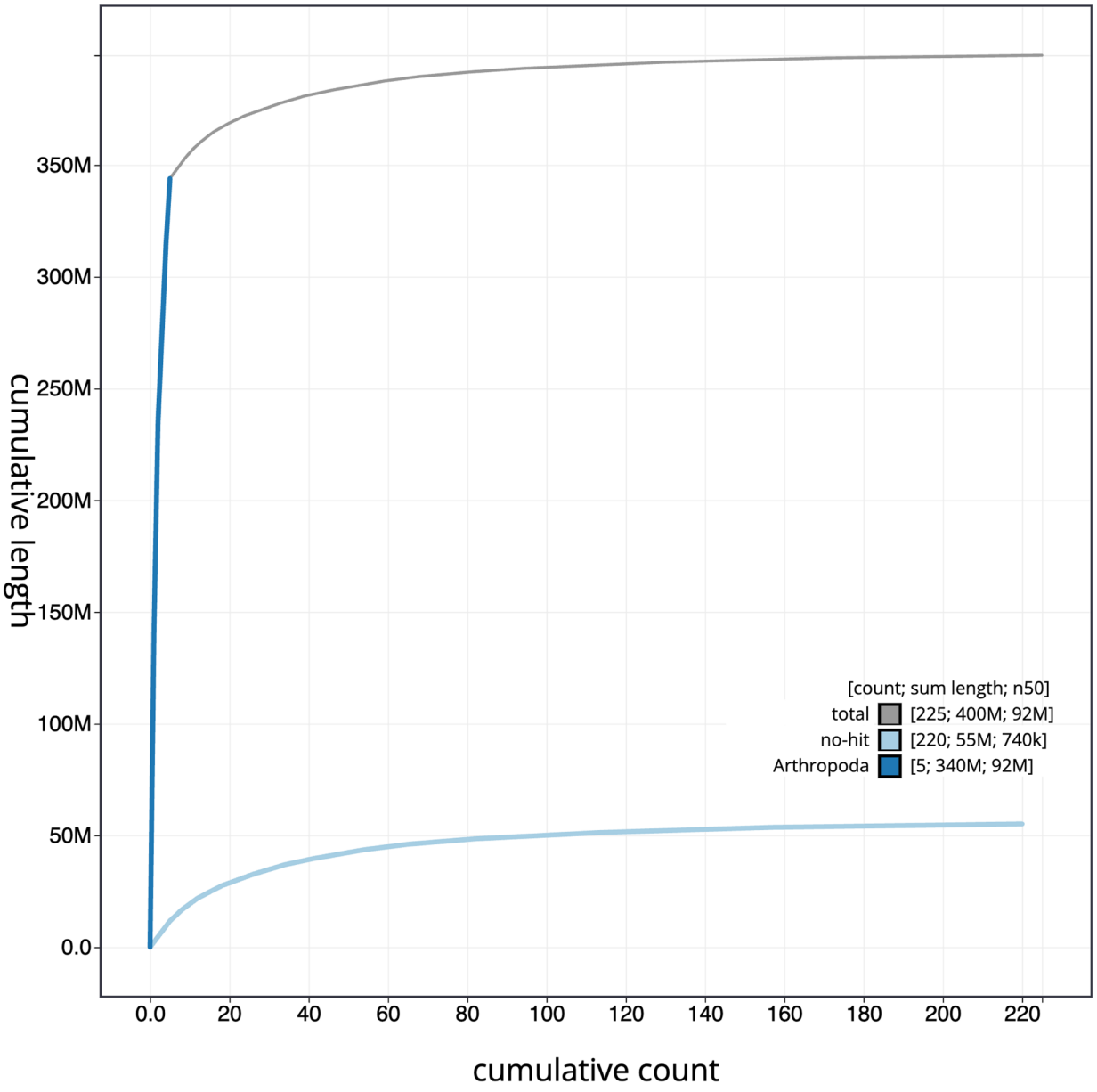

Figure 4. Genome assembly of *Andrena trimmerana* iyAndTrim1.1: BlobToolKit cumulative sequence plot. The grey line shows cumulative length for all sequences. Coloured lines show cumulative lengths of sequences assigned to each phylum using the buscogenes taxrule. An interactive version of this figure is available at https://blobtoolkit.genomehubs.org/view/Andrena%20trimmerana/dataset/iyAndTrim1_1/cumulative.

(Abdennur & Mirny, 2020). To assess the assembly metrics, 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 4 contains a list of relevant software tool versions and sources.

Genome annotation

The Ensembl Genebuild annotation system (Aken *et al.*, 2016) was used to generate annotation for the *Andrena trimmerana* assembly (GCA_951215215.1) in Ensembl Rapid Release at the EBI. Annotation was created primarily through alignment of transcriptomic data to the genome, with gap filling via

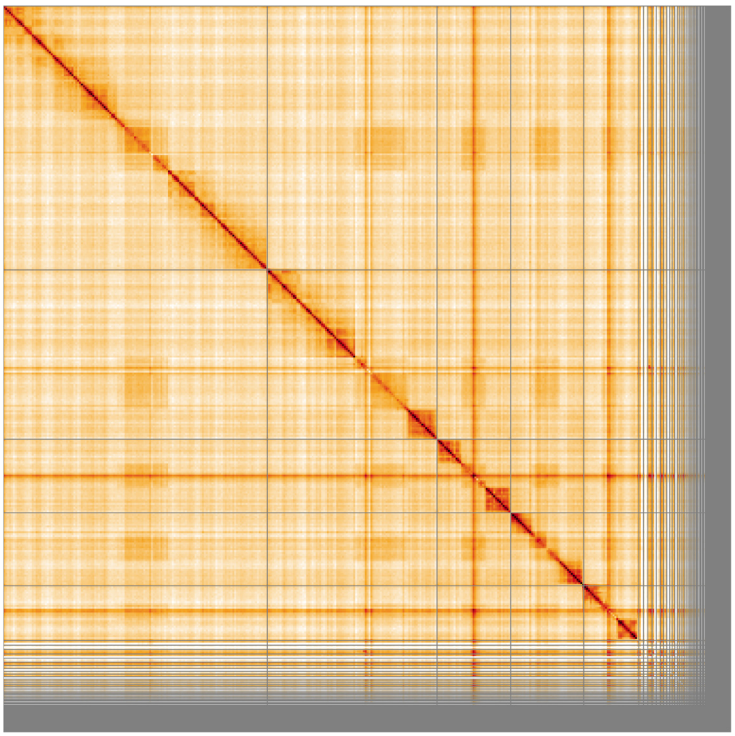

Figure 5. Genome assembly of *Andrena trimmerana*, iyAndTrim1.1: Hi-C contact map of the iyAndTrim1.1 assembly, visualised using HiGlass. Chromosomes are shown in order of size from left to right and top to bottom. An interactive version of this figure may be viewed at https://genome-note-higlass.tol.sanger.ac.uk/l/?d=LwthMvtpQ6aOYoK0J8qqsQ.

Table 3. Chromosomal pseudomolecules in the genome assembly of *Andrena trimmerana*, iyAndTrim1.

INSDC accession	Name	Length (Mb)	GC%
OX578059.1	1	142.9	41.5
OX578060.1	2	91.87	42.5
OX578061.1	3	39.86	43.5
OX578062.1	4	39.55	42.5
OX578063.1	5	29.78	42.5
OX578064.1	MT	0.02	21.0

protein-to-genome alignments of a select set of proteins from UniProt (UniProt Consortium, 2019).

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 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

Table 4. 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
bwa-mem2	2.2.1	https://github.com/bwa-mem2/bwa-mem2
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	2	https://github.com/marcelauliano/MitoHiFi
PretextView	0.2	https://github.com/sanger-tol/PretextView
purge_dups	1.2.3	https://github.com/dfguan/purge_dups
sanger-tol/ascc	-	https://github.com/sanger-tol/ascc
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	yahs-1.1.91eebc2	https://github.com/c-zhou/yahs

Institute), and in some circumstances other Darwin Tree of Life collaborators.

Data availability

European Nucleotide Archive: *Andrena trimmerana* (trimmer's mining bee). Accession number PRJEB56904; https://identifiers.org/ena.embl/PRJEB56904. The genome sequence is released openly for reuse. The *Andrena trimmerana* 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. 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.7125292.

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.

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

Aken BL, Ayling S, Barrell D, *et al.*: **The Ensembl gene annotation system.** *Database (Oxford).* 2016; **2016**: baw093.
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

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

Beasley J, Uhl R, Forrest LL, *et al.*: **DNA barcoding SOPs for the Darwin Tree of Life project.** *protocols.io.* 2023; [Accessed 25 June 2024].
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

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

Chow W, Brugger K, Caccamo M, *et al.*: **gEVAL — a web-based browser for evaluating genome assemblies.** *Bioinformatics.* 2016; **32**(16): 2508–2510.
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

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 V.1.** *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

do Amaral RJV, Bates A, Denton A, *et al.*: **Sanger Tree of Life RNA extraction: automated MagMax™ mirVana.** *protocols.io.* 2023.
Publisher Full Text

Else GR: ***Andrena trimmerana* (Kirby,1802)**. *Bees Wasps & Ants Recording Society*. 2023; [Accessed 8 September 2024].
Reference Source

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

GBIF Secretariat: ***Andrena* (*Hoplandrena*) *trimmerana* (Kirby, 1802).** *GBIF Backbone Taxonomy*. 2024; [Accessed 8 September 2024].
Reference Source

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.
Reference Source

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

Jay J, Yatsenko H, Narváez-Gómez JP, *et al.*: **Sanger Tree of Life sample preparation: triage and dissection.** *protocols.io.* 2023.
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, prokaryotic, and viral genomes.** *Mol Biol Evol.* 2021; **38**(10): 4647–4654.
PubMed Abstract | Publisher Full Text | Free Full Text

NBN Atlas Partnership: ***Andrena trimmerana* (Kirby, 1802) Trimmer's Mining Bee**. *NBN Atlas*. 2024; [Accessed 30 October 2024].
Reference Source

Oatley G, Denton A, Howard C: **Sanger Tree of Life HMW DNA extraction: automated MagAttract v.2.** *protocols.io.* 2023.
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

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

Strickland M, Cornwell C, Howard C: **Sanger Tree of Life fragmented DNA clean up: manual SPRI.** *protocols.io.* 2023.
Publisher Full Text

Surana P, Muffato M, Qi G: **sanger-tol/readmapping: sanger-tol/readmapping v1.1.0 - Hebridean Black (1.1.0).** *Zenodo.* 2023a.
Publisher Full Text

Surana P, Muffato M, Sadasivan Baby C: **sanger-tol/genomenote (v1.0.dev).** *Zenodo.* 2023b.
Publisher Full Text

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

UniProt Consortium: **UniProt: a worldwide hub of protein knowledge.** *Nucleic Acids Res.* 2019; **47**(D1): D506–D515.
PubMed Abstract | Publisher Full Text | Free Full Text

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

Villalta I, Ledet R, Baude M, *et al.*: **A DNA barcode-based survey of wild urban bees in the Loire Valley, France.** *Sci Rep.* 2021; **11**(1): 4770.
PubMed Abstract | Publisher Full Text | Free Full Text

Wood TJ, Hogan J, Edwards M, *et al.*: ***Andrena scotica* Perkins is the valid name for the widespread European taxon previously referred to as *Andrena carantonica* Pérez (Hymenoptera: Andrenidae).** *Brit J Entomol Nat Hist.* 2022; **35**(4): 393–408.
Reference Source

Zhou C, McCarthy SA, Durbin R: **YaHS: yet another Hi-C scaffolding tool.** *Bioinformatics.* 2023; **39**(1): btac808.
PubMed Abstract | Publisher Full Text | Free Full Text

Open Peer Review

Current Peer Review Status: ✔ ✔ ?

Version 1

Reviewer Report 13 November 2025

https://doi.org/10.21956/wellcomeopenres.25765.r136969

?

Josephine R Paris
University of Exeter Library, Devon, UK

This is a well-written Data Note describing the genome assembly and annotation of the Trimmer's Mining Bee. The Background information provides thorough information on the biology and ecology of the species. The genome assembly lab and bioinformatics protocols are standard. Data can be found on the ENA and Ensembl Rapid Release.

I have only a few minor comments:

Abstract: there's a space missing: "afemale specimen" should be "a female specimen"

Background: please include the latin names: willow, dandelion, gorse, bramble, blackthorn, alexanders, and rhododendron

Report: please explain why only 86.18% of the assembly was assigned to 5 chromosomes. Were there difficulties in the curation? Was the data coverage too low? This will help others experiencing similar issues

Is the rationale for creating the dataset(s) clearly described?
Yes

Are the protocols appropriate and is the work technically sound?
Partly

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 genomics of non-model organisms, including genome assembly and annotation

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.

Reviewer Report 08 November 2025

https://doi.org/10.21956/wellcomeopenres.25765.r136974

Liang Lu
Hebei Normal University, Shijiazhuang, Hebei, China

The manuscript presents a high-quality HiFi genome assembly of *Andrena trimmerana* (Kirby, 1802), complemented by Hi-C based chromosomal scaffolding. Structurally, it follows the established format of previous Wellcome Open Research genome reports, offering comprehensive background on taxonomy, biology, and specimen collection. The inclusion of clear diagnostic images enhances confidence in species identification. The methodologies for sample preparation, DNA extraction, sequencing, assembly, and annotation are consistent with the journal's prior publications and appear technically sound.

However, I would encourage the authors to consider consolidating future genome reports—perhaps by publishing multiple assemblies or a thematic package within a single manuscript or a limited series. Much of the content in these reports is standardized and could be efficiently presented in a more aggregated format.

One tiny omission is the lack of detail regarding tissue sampling. Table 1 indicates that sequencing was performed on the three tagmata using different technologies, but it remains unclear whether entire tagmata were used or if specific tissues (e.g., muscle) were selectively processed. Furthermore, the manuscript does not address how intestinal contents or potential intracellular parasites and symbionts were handled. Clarification is needed on whether post-assembly decontamination procedures were sufficient to remove non-host sequences originating from these sources.

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: phylogenomics, entomology

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 30 December 2024

https://doi.org/10.21956/wellcomeopenres.25765.r114036

Rodolpho S. T. Menezes
State University of Santa Cruz, Ilhéus, Brazil

The data note by Baker et al. presents a genome assembly for Trimmer's Mining Bee, *Andrena trimmerana* (Kirby, 1802), with a total sequence length of 399.10 megabases. This genome provides valuable insights for comparative genomics, as well as evolutionary and conservation studies. The methodology is comprehensive and well-documented, while the figures are clearly constructed, contributing to a well-written data note.
However, I recommend that the authors include more detailed information about the mitogenome. Additionally, a brief comparison of key descriptive statistics with the recent genome assembly of the Small Scabious Mining Bee, *Andrena marginata*, would enhance the note by providing context and broader relevance.

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: Phylogenomics, Cytogenetics, Phylogeography.

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.