






DATA NOTE

# The genome sequence of a chalcid wasp, *Gastracanthus pulcherrimus* (Westwood, 1833) [version 1; peer review: 2 approved]

Gavin R. Broad <sup>1</sup>, Chris Fletcher <sup>1</sup>, Inez Januszczak <sup>1</sup>,  
Natural History Museum 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

<sup>1</sup>Natural History Museum, London, England, UK

**V1** First published: 12 Oct 2023, 8:440  
<https://doi.org/10.12688/wellcomeopenres.19764.1>  
Latest published: 12 Oct 2023, 8:440  
<https://doi.org/10.12688/wellcomeopenres.19764.1>

## Abstract

We present a genome assembly from an individual female *Gastracanthus pulcherrimus* (a chalcid wasp; Arthropoda; Insecta; Hymenoptera; Pteromalidae). The genome sequence is 1,010.0 megabases in span. Most of the assembly is scaffolded into 5 chromosomal pseudomolecules. The mitochondrial genome has also been assembled and is 24.4 kilobases in length.

## Keywords



*Gastracanthus pulcherrimus*, chalcid wasp, genome sequence, chromosomal, Hymenoptera



This article is included in the [Tree of Life gateway](#).

## Open Peer Review

Approval Status  

	1	2
<b>version 1</b> 12 Oct 2023	 <a href="#">view</a>	 <a href="#">view</a>

1. **Jonathan Berenguer Uhuad Koch** ,  
USDA-ARS Pollinating Insect Research Unit,  
Maryland, USA
2. **Daniel Garcia-Souto** ,  
Universidade de  
Vigo (Ringgold ID: 16784), Vigo, Spain

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](mailto:mark.blaxter@sanger.ac.uk))

**Author roles:** **Broad GR:** Investigation, Resources, Writing – Original Draft Preparation, Writing – Review & Editing; **Fletcher C:** Investigation, Resources; **Januszczyk I:** 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) 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:** © 2023 Broad GR *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:** Broad GR, Fletcher C, Januszczyk I *et al.* **The genome sequence of a chalcid wasp, *Gastracanthus pulcherrimus* (Westwood, 1833) [version 1; peer review: 2 approved]** Wellcome Open Research 2023, 8:440 <https://doi.org/10.12688/wellcomeopenres.19764.1>

**First published:** 12 Oct 2023, 8:440 <https://doi.org/10.12688/wellcomeopenres.19764.1>



DATA NOTE

# The genome sequence of a chalcid wasp, *Gastracanthus pulcherrimus* (Westwood, 1833)

Gavin R. Broad <sup>1</sup>, Chris Fletcher <sup>1</sup>, Inez Januszczak<sup>1</sup>,  
Natural History Museum 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

<sup>1</sup>Natural History Museum, London, England, UK

---

**v1** First published: N/A, N/A: N/A N/A  
Latest published: N/A, N/A: N/A N/A

---

## Abstract

We present a genome assembly from an individual female *Gastracanthus pulcherrimus* (a chalcid wasp; Arthropoda; Insecta; Hymenoptera; Pteromalidae). The genome sequence is 1,010.0 megabases in span. Most of the assembly is scaffolded into 5 chromosomal pseudomolecules. The mitochondrial genome has also been assembled and is 24.4 kilobases in length.

## Keywords

*Gastracanthus pulcherrimus*, chalcid wasp, genome sequence, chromosomal, Hymenoptera



This article is included in the [Tree of Life gateway](#).

## Open Peer Review

**Approval Status** AWAITING PEER REVIEW

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](mailto:mark.blaxter@sanger.ac.uk))

**Author roles:** **Broad GR:** Investigation, Resources, Writing – Original Draft Preparation, Writing – Review & Editing; **Fletcher C:** Investigation, Resources; **Januszczyk I:** 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) 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:** © 2023 Broad GR *et al.* This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

**How to cite this article:** Broad GR, Fletcher C, Januszczyk I *et al.* **The genome sequence of a chalcid wasp, *Gastracanthus pulcherrimus* (Westwood, 1833)** Wellcome Open Research , : <https://doi.org/>

**First published:** N/A, N/A: N/A N/A

## Species taxonomy

Eukaryota; Metazoa; Eumetazoa; Bilateria; Protostomia; Ecdysozoa; Panarthropoda; Arthropoda; Mandibulata; Pancrustacea; Hexapoda; Insecta; Dicondylia; Pterygota; Neoptera; Endopterygota; Hymenoptera; Apocrita; Proctotrupomorpha; Chalcidoidea; Pteromalidae; Pteromalinae; *Gastracanthus*; *Gastracanthus pulcherrimus* (Westwood, 1833) (NCBI: txid2922068).

## Background

With females measuring around 5 mm in length, *Gastracanthus pulcherrimus* is a relatively large species of the chalcid wasp family Pteromalidae. It is also a striking species, slender, with a particularly long pronotum, metallic green hues and, in females, two large dark patches on each fore wing. Like most parasitoid wasps, *G. pulcherrimus* is poorly known but it is fairly easily found in deciduous woodland and has a wide range across Europe (Noyes, 2019), recorded from the Republic of Ireland, Northern Ireland (Thuróczy & O'Connor, 2009) and from England, but not yet from Wales or Scotland (Dale-Skey *et al.*, 2016).

Despite its relative conspicuousness, knowledge about the biology of *G. pulcherrimus* is not extensive. It belongs to the subfamily Trigonoderinae, which all seem to be parasitoids of Coleoptera (Graham, 1969). *Gastracanthus pulcherrimus* is associated with wood and has been reported as a parasitoid of a *Sphenoptera* species (Coleoptera: Buprestidae (Ghahari & Huang, 2012). Graham (1969) expresses doubt about a supposed rearing from an adult byrrhid beetle. By analogy with related Pteromalidae, *G. pulcherrimus* is presumably an idiobiont ectoparasitoid, i.e., hosts will be permanently paralysed, and the larva develops externally on the host.

The family Pteromalidae, for a long time an unwieldy assemblage of disparate chalcid lineages, has recently been split into many different, hopefully monophyletic families (Burks *et al.*, 2022). *Gastracanthus* is among ten genera comprising the subfamily Trigonoderinae, itself one of eight subfamilies within the diverse and species rich Pteromalidae family. All previously published genomes belong to species in the subfamily Pteromalinae (Martinson *et al.*, 2017; Werren *et al.*, 2010).

## Genome sequence report

The genome was sequenced from one female *Gastracanthus pulcherrimus* (Figure 1) collected from Wytham Woods, UK (51.77, -1.31). A total of 27-fold coverage in Pacific Biosciences single-molecule HiFi long reads was generated. Primary assembly contigs were scaffolded with chromosome conformation Hi-C data. Manual assembly curation corrected 48 missing joins or mis-joins and removed one haplotypic duplication, reducing the scaffold number by 6.49%, and increasing the scaffold N50 by 0.45%.

The final assembly has a total length of 1,010.0 Mb in 143 sequence scaffolds with a scaffold N50 of 186.8 Mb (Table 1).



**Figure 1. Photograph of the *Gastracanthus pulcherrimus* (iyGasPulc2) specimen used for genome sequencing.**

Most (95.67%) of the assembly sequence was assigned to 5 chromosomal-level scaffolds. Chromosome-scale scaffolds confirmed by the Hi-C data are named in order of size (Figure 2–Figure 5; Table 2). While not fully phased, the assembly deposited is of one haplotype. Contigs corresponding to the second haplotype have also been deposited. The mitochondrial genome was also assembled and can be found as a contig within the multifasta file of the genome submission.

The estimated Quality Value (QV) of the final assembly is 54.3 with *k*-mer completeness of 99.98%, and the assembly has a BUSCO v5.3.2 completeness of 92.6% (single = 91.7%, duplicated = 0.9%), using the hymenoptera\_odb10 reference set (*n* = 5,991).

Metadata for specimens, spectral estimates, sequencing runs, contaminants and pre-curation assembly statistics can be found at <https://links.tol.sanger.ac.uk/species/2922068>.

## Methods

### Sample acquisition and nucleic acid extraction

Two adult *Gastracanthus pulcherrimus* were collected from Bert's Pheasant Pen, Wytham Woods, Oxfordshire (biological vice-county Berkshire), UK (latitude 51.77, longitude -1.31) on 2021-09-02, using a light trap. The collectors were Gavin Broad, Chris Fletcher and Inez Januszczak (Natural History Museum). The specimens were identified by Gavin Broad and then dry-frozen at -80°C. The specimen used for DNA sequencing was specimen ID NHMUK014451697, ToLID iyGasPulc2, while specimen ID NHMUK014451655, ToLID iyGasPulc1 was used for Hi-C scaffolding.

DNA was extracted at the Tree of Life laboratory, Wellcome Sanger Institute (WSI). The iyGasPulc2 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. High molecular weight (HMW) DNA was extracted using the Qiagen

**Table 1. Genome data for *Gastracanthus pulcherrimus*, iyGasPulc2.1.**

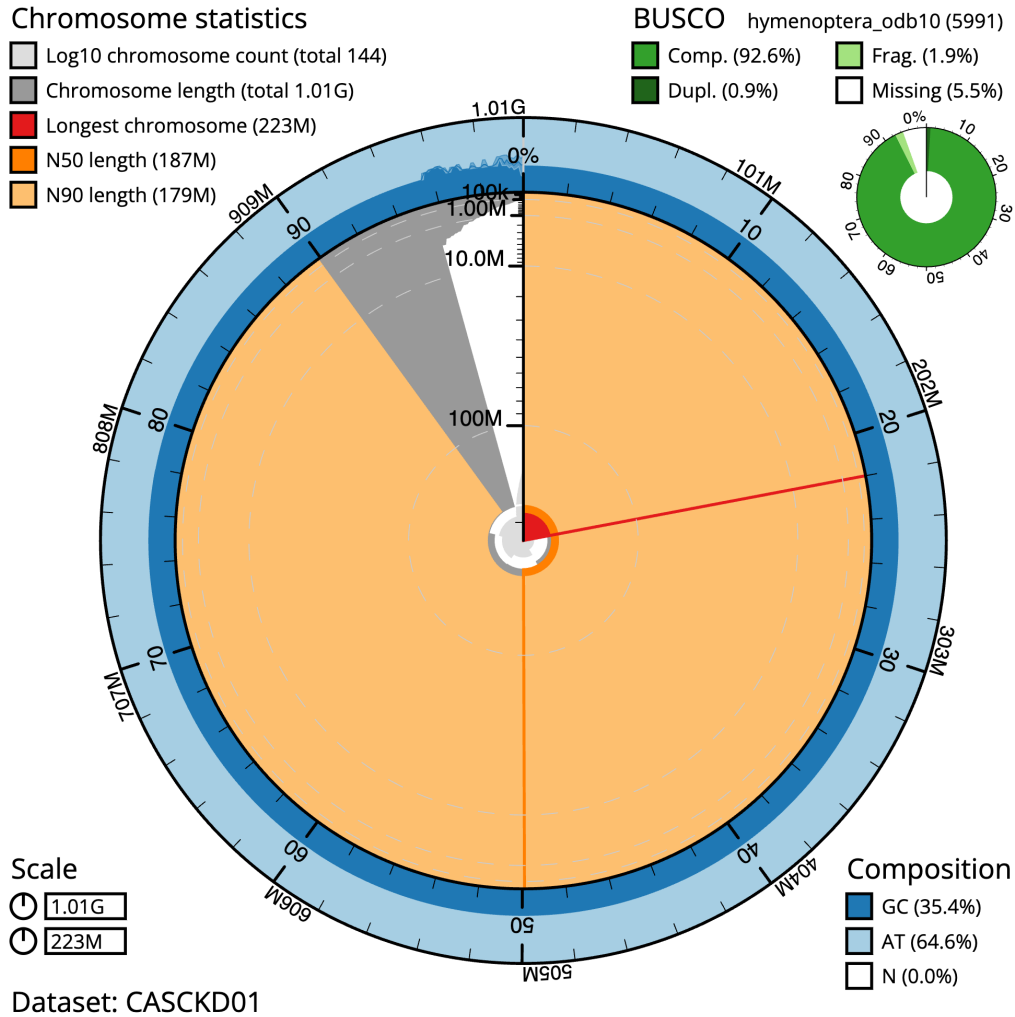
Project accession data		
Assembly identifier	iyGasPulc2.1	
Species	<i>Gastracanthus pulcherrimus</i>	
Specimen	iyGasPulc2	
NCBI taxonomy ID	2922068	
BioProject	PRJEB59075	
BioSample ID	SAMEA110043171	
Isolate information	iyGasPulc2, female: whole organism (DNA sequencing) iyGasPulc1: whole organisms (Hi-C scaffolding)	
Assembly metrics*		Benchmark
Consensus quality (QV)	54.3	≥ 50
k-mer completeness	99.98%	≥ 95%
BUSCO**	C:92.6%[S:91.7%,D:0.9%], F:1.9%,M:5.5%,n:5,991	C ≥ 95%
Percentage of assembly mapped to chromosomes	95.67%	≥ 95%
Sex chromosomes	-	localised homologous pairs
Organelles	Mitochondrial genome assembled	complete single alleles
Raw data accessions		
PacificBiosciences SEQUEL II	ERR10798431	
Hi-C Illumina	ERR10802452	
Genome assembly		
Assembly accession	GCA_949152435.1	
Accession of alternate haplotype	GCA_949152415.1	
Span (Mb)	1,010.0	
Number of contigs	954	
Contig N50 length (Mb)	1.9	
Number of scaffolds	143	
Scaffold N50 length (Mb)	186.8	
Longest scaffold (Mb)	222.5	

\* Assembly metric benchmarks are adapted from column VGP-2020 of “Table 1: Proposed standards and metrics for defining genome assembly quality” from (Rhie *et al.*, 2021).

\*\* BUSCO scores based on the hymenoptera\_odb10 BUSCO set using v5.3.2. 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/iyGasPulc2.1/dataset/CASCKD01/busco>.

MagAttract HMW DNA extraction kit. HMW DNA was sheared into an average fragment size of 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.



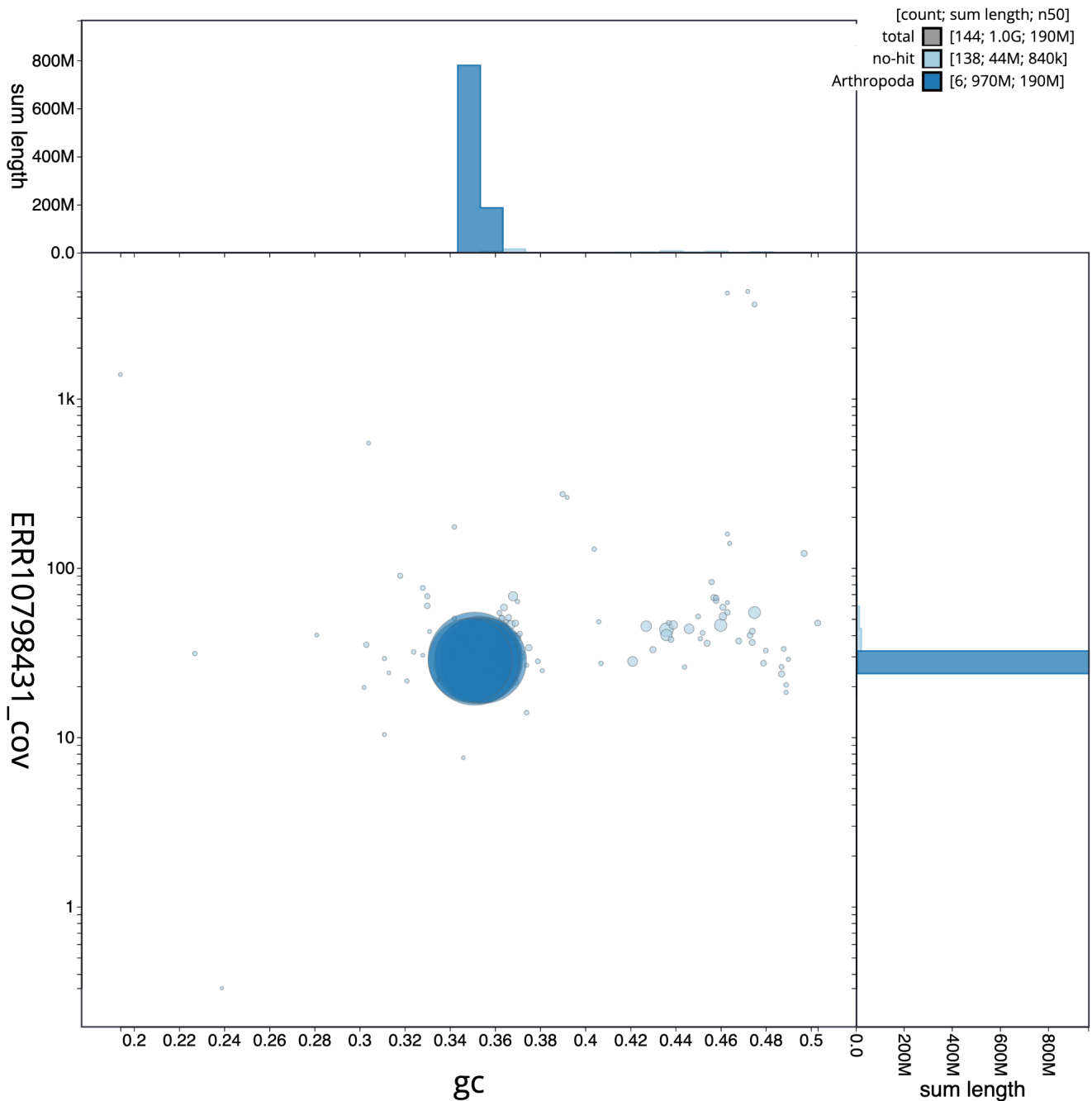
**Figure 2. Genome assembly of *Gastracanthus pulcherrimus*, iyGasPulc2.1: metrics.** The BlobToolKit Snailplot shows N50 metrics and BUSCO gene completeness. The main plot is divided into 1,000 size-ordered bins around the circumference with each bin representing 0.1% of the 1,010,024,266 bp assembly. The distribution of scaffold lengths is shown in dark grey with the plot radius scaled to the longest scaffold present in the assembly (222,508,353 bp, shown in red). Orange and pale-orange arcs show the N50 and N90 scaffold lengths (186,761,157 and 179,352,141 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/iyGasPulc2.1/dataset/CASCKD01/snail>.

### Sequencing

Pacific Biosciences HiFi circular consensus DNA sequencing libraries were constructed according to the manufacturers' instructions. DNA sequencing was performed by the Scientific Operations core at the WSI on a Pacific Biosciences SEQUEL IIe (HiFi) instrument. Hi-C data were also generated from whole organism tissue of iyGasPulc1 using the Arima2 kit and sequenced on the Illumina NovaSeq 6000 instrument.

### Genome assembly, curation and evaluation

Assembly was carried out with Hifiasm (Cheng *et al.*, 2021) and haplotypic duplication was identified and removed with purge\_dups (Guan *et al.*, 2020). The assembly was then scaffolded with Hi-C data (Rao *et al.*, 2014) using YaHS (Zhou *et al.*, 2023). The assembly was checked for contamination and corrected using the gEVAL system (Chow *et al.*, 2016) as described previously (Howe *et al.*, 2021). Manual curation was



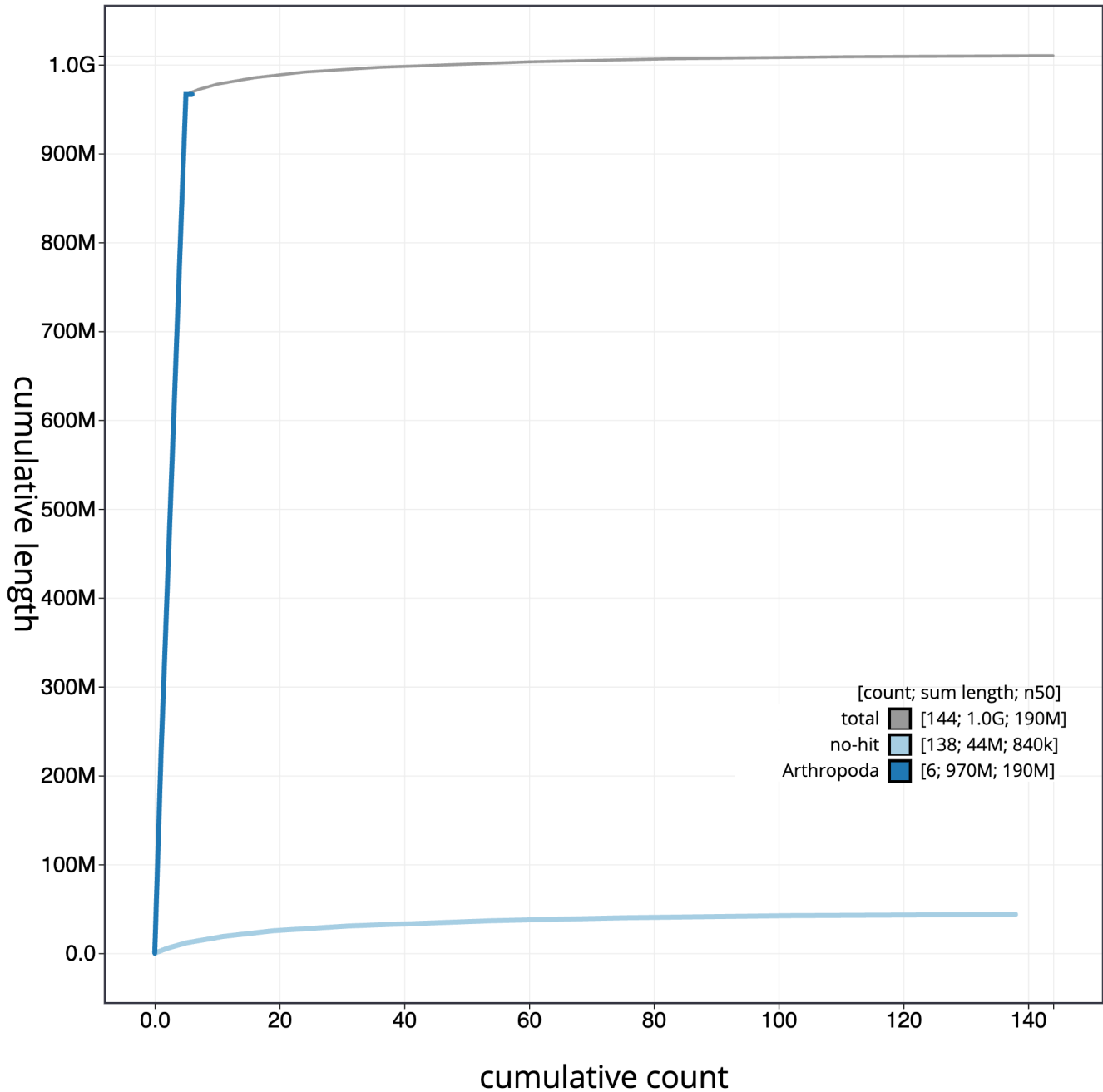
**Figure 3. Genome assembly of *Gastracanthus pulcherrimus*, iyGasPulc2.1: BlobToolKit GC-coverage plot.** Scaffolds are coloured by phylum. Circles are sized in proportion to scaffold length. Histograms show the distribution of scaffold length sum along each axis. An interactive version of this figure is available at <https://blobtoolkit.genomehubs.org/view/iyGasPulc2.1/dataset/CASCKD01/blob>.

performed using gEVAL, HiGlass (Kerpedjiev *et al.*, 2018) and Pretext (Harry, 2022). The mitochondrial genome was assembled using MitoHiFi (Uliano-Silva *et al.*, 2022), which runs MitoFinder (Allio *et al.*, 2020) or MITOS (Bernt *et al.*, 2013) and uses these annotations to select the final mitochondrial

contig and to ensure the general quality of the sequence.

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

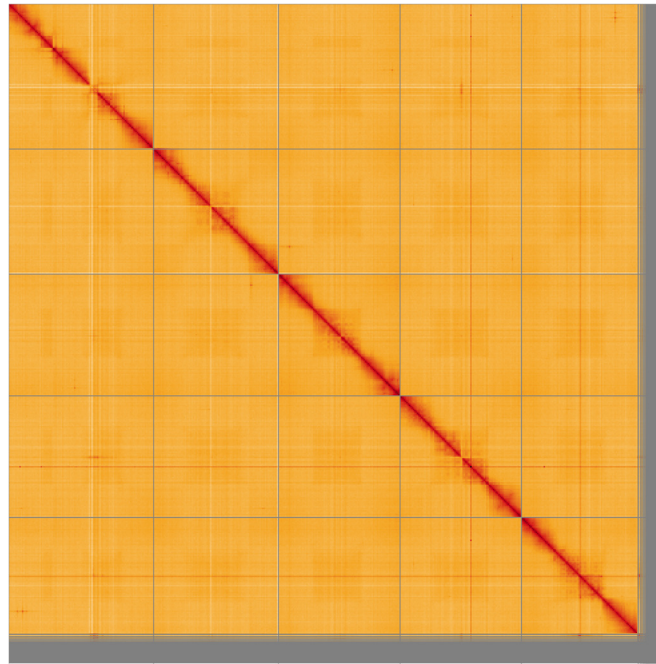




**Figure 4. Genome assembly of *Gastracanthus pulcherrimus*, iyGasPulc2.1: BlobToolKit cumulative sequence plot.** The grey line shows cumulative length for all scaffolds. Coloured lines show cumulative lengths of scaffolds assigned to each phylum using the buscogenes taxrule. An interactive version of this figure is available at <https://blobtoolkit.genomehubs.org/view/iyGasPulc2.1/dataset/CASCKD01/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



**Figure 5. Genome assembly of *Gastracanthus pulcherrimus*, iyGasPulc2.1: Hi-C contact map of the iyGasPulc2.1 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/I/?d=Yrs3ERn6RL-maDBtEFqvcA>.

**Table 2. Chromosomal pseudomolecules in the genome assembly of *Gastracanthus pulcherrimus*, iyGasPulc2.**

INSDC accession	Chromosome	Length (Mb)	GC%
OX424572.1	1	222.51	35.0
OX424573.1	2	191.63	35.5
OX424574.1	3	186.76	35.5
OX424575.1	4	186.09	35.0
OX424576.1	5	179.35	35.0
OX424577.1	MT	0.02	19.5

was analysed within the BlobToolKit environment (Challis *et al.*, 2020) and BUSCO scores (Manni *et al.*, 2021; Simão *et al.*, 2015) were calculated.

Table 3 contains a list of relevant software tool versions and sources.

#### Wellcome Sanger Institute – Legal and Governance

The materials that have contributed to this genome note have been supplied by a Darwin Tree of Life Partner. The submission of materials by a Darwin Tree of Life Partner is subject to the ‘Darwin Tree of Life Project Sampling Code of

Practice’, which can be found in full on the Darwin Tree of Life website [here](#). By agreeing with and signing up to the Sampling Code of Practice, the Darwin Tree of Life Partner agrees they will meet the legal and ethical requirements and standards set out within this document in respect of all samples acquired for, and supplied to, the Darwin Tree of Life Project.

Further, the Wellcome Sanger Institute employs a process whereby due diligence is carried out proportionate to the nature of the materials themselves, and the circumstances under which they have been/are to be collected and provided for use. The purpose of this is to address and mitigate any potential legal and/or ethical implications of receipt and use of the materials as part of the research project, and to ensure that in 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.

**Table 3. Software tools: versions and sources.**

Software tool	Version	Source
BlobToolKit	4.1.5	<a href="https://github.com/blobtoolkit/blobtoolkit">https://github.com/blobtoolkit/blobtoolkit</a>
BUSCO	5.3.2	<a href="https://gitlab.com/ezlab/busco">https://gitlab.com/ezlab/busco</a>
gEVAL	N/A	<a href="https://geval.org.uk/">https://geval.org.uk/</a>
Hifiasm	0.16.1-r375	<a href="https://github.com/chhylp123/hifiasm">https://github.com/chhylp123/hifiasm</a>
HiGlass	1.11.6	<a href="https://github.com/higlass/higlass">https://github.com/higlass/higlass</a>
Merqury	MerquryFK	<a href="https://github.com/thegenemyers/MERQURY.FK">https://github.com/thegenemyers/MERQURY.FK</a>
MitoHiFi	2	<a href="https://github.com/marcelauliano/MitoHiFi">https://github.com/marcelauliano/MitoHiFi</a>
PretextView	0.2	<a href="https://github.com/wtsi-hpag/PretextView">https://github.com/wtsi-hpag/PretextView</a>
purge_dups	1.2.3	<a href="https://github.com/dfguan/purge_dups">https://github.com/dfguan/purge_dups</a>
sanger-tol/genomenote	v1.0	<a href="https://github.com/sanger-tol/genomenote">https://github.com/sanger-tol/genomenote</a>
sanger-tol/readmapping	1.1.0	<a href="https://github.com/sanger-tol/readmapping/tree/1.1.0">https://github.com/sanger-tol/readmapping/tree/1.1.0</a>
YaHS	1.2a	<a href="https://github.com/c-zhou/yahs">https://github.com/c-zhou/yahs</a>

### Data availability

European Nucleotide Archive: *Gastracanthus pulcherrimus*. Accession number PRJEB59075; <https://identifiers.org/ena.embl/PRJEB59075>. (Wellcome Sanger Institute, 2023)

The genome sequence is released openly for reuse. The *Gastracanthus pulcherrimus* genome sequencing initiative is part of the Darwin Tree of Life (DTOL) project. All raw sequence data and the assembly have been deposited in INSDC databases. The genome will be annotated using available RNA-Seq data and presented through the [Ensembl](#) pipeline at the European Bioinformatics Institute. Raw data and assembly accession identifiers are reported in [Table 1](#).

### Author information

Members of the Natural History Museum Genome Acquisition Lab are listed here: <https://doi.org/10.5281/zenodo.4790042>.

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 programme are listed here: <https://doi.org/10.5281/zenodo.4783585>.

Members of Wellcome Sanger Institute Scientific Operations: DNA Pipelines collective are listed here: <https://doi.org/10.5281/zenodo.4790455>.

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](#)

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](#)

Bernt M, Donath A, Jühling F, et al.: **MITOS: Improved *de novo* metazoan mitochondrial genome annotation.** *Mol Phylogenet Evol*. 2013; **69**(2): 313–9.

[PubMed Abstract](#) | [Publisher Full Text](#)

Burks R, Mitroiu MD, Fusu L, et al.: **From hell's heart I stab at thee! A**

**determined approach towards a monophyletic Pteromalidae and reclassification of Chalcidoidea (Hymenoptera).** *J Hymenopt Res*. 2022; **94**(2): 13–88.

[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): 2508–10.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Dale-Skey N, Askew RR, Noyes JS, *et al.*: **Checklist of British and Irish Hymenoptera - Chalcidoidea and Mymarommatoidea.** *Biodivers Data J.* 2016; **4**: e8013.  
[PubMed Abstract](#) | [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](#)
- Ghahari H, Huang J: **A study of the Pteromalidae (Hymenoptera: Chalcidoidea) from western and northwestern Iran.** *Archives of Biological Sciences.* 2012; **64**(1): 353–357.  
[Publisher Full Text](#)
- Graham MWRdeV: **The Pteromalidae of north-western Europe (Hymenoptera: Chalcidoidea).** *Bull Br Mus Entomology Supplement.* 1969; **16**: 1–908.  
[Publisher Full Text](#)
- Guan D, McCarthy SA, Wood J, *et al.*: **Identifying and removing haplotypic duplication in primary genome assemblies.** *Bioinformatics.* 2020; **36**(9): 2896–2898.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Harry E: **PretextView (Paired REad TEXTure Viewer): A desktop application for viewing pretext contact maps.** 2022; [Accessed 19 October 2022].  
[Reference Source](#)
- Howe K, Chow W, Collins J, *et al.*: **Significantly improving the quality of genome assemblies through curation.** *GigaScience.* Oxford University Press, 2021; **10**(1): gaa153.  
[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, prokaryotic, and viral genomes.** *Mol Biol Evol.* 2021; **38**(10): 4647–4654.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Martinson EO, Mrinalini, Kelkar YD, *et al.*: **The Evolution of Venom by Co-option of Single-Copy Genes.** *Curr Biol.* 2017; **27**(13): 2007–2013.e8.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Noyes JS: **Universal Chalcidoidea Database.** 2019; [Accessed 7 July 2023].  
[Reference Source](#)
- Rao SSP, Huntley MH, Durand NC, *et al.*: **A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping.** *Cell.* 2014; **159**(7): 1665–80.  
[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): 3210–2.  
[PubMed Abstract](#) | [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](#)
- Thuróczy C, O'Connor JP: **Records of Irish Chalcidoidea (Hymenoptera) including thirteen species new to Ireland.** *Irish Naturalist's Journal.* 2009; **30**: 126.  
[Reference Source](#)
- Uliano-Silva M, Ferreira JGRN, Krashenninnikova K, *et al.*: **MitoHiFi: a python pipeline for mitochondrial genome assembly from PacBio High Fidelity reads.** *bioRxiv.* 2022.  
[Publisher 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](#)
- Wellcome Sanger Institute: **The genome sequence of the chalcid wasp, *Gastrancistrus pulcherrimus* (Westwood, 1833).** European Nucleotide Archive. [dataset], accession number PRJEB59075, 2023.
- Werren JH, Richards S, Desjardins CA, *et al.*: **Functional and Evolutionary Insights from the Genomes of Three Parasitoid *Nasonia* Species.** *Science.* 2010; **327**(5963): 343–8.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free 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 16 September 2024

<https://doi.org/10.21956/wellcomeopenres.21890.r95768>

© 2024 Garcia-Souto D. This is an open access peer review report 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.



**Daniel Garcia-Souto** 

Bioquímica, xenética e inmunoloxía, Universidade de Vigo (Ringgold ID: 16784), Vigo, Galicia, Spain

This manuscript presents a chromosome-level genome assembly of *Gastracanthus pulcherrimus*, a species of chalcid wasp in the family Pteromalidae. The study provides significant insights into the genomic structure of this relatively understudied species, with a genome span of 1,010.0 megabases. The assembly was primarily scaffolded into five chromosomal pseudomolecules, covering 95.67% of the genome, which highlights the high degree of completeness in the assembly. The authors also sequenced and assembled the mitochondrial genome, which is 24.4 kilobases in length, potentially indicating expansions or repeats in the control region that would, in my honest opinion, require some validation.

The manuscript makes a notable contribution to the genomic data of the genus, as this is one of the few genomes from the Pteromalidae family outside the Pteromalinae subfamily. Although the final assembly's BUSCO score is slightly lower than the desired threshold (92.6% completeness), this is still a strong result for a genome of such complexity.

**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:** Genomics, Molecular Cytogenetics, Epigenetics, Karyotyping.

**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 04 September 2024

<https://doi.org/10.21956/wellcomeopenres.21890.r95766>

© 2024 Koch J. This is an open access peer review report 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.



**Jonathan Berenguer Uhuad Koch** 

USDA-ARS Pollinating Insect Research Unit, Maryland, USA

The article describes a genome assembly of *Gastracanthus pulcherrimus* from a female specimen. The authors clearly articulate the generation of the genome assembly using PacBio HiFi and HiC. The genome assembly still needs to be annotated, and is currently in the queue.

**One minor point**

- Rephrase "The family Pteromalidae, for a long time an unwieldy assemblage of disparate chalcid lineages, has recently been split into many different, hopefully monophyletic families" to "The family Pteromalidae, for a long time an unwieldy assemblage of disparate chalcid lineages, has recently been split into many different, likely monophyletic families"

**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, 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.**

---