




DATA NOTE

# The genome sequence of the Beautiful Golden Y, *Autographa pulchrina* (Haworth, 1809)

[version 1; peer review: 2 approved]

Douglas Boyes<sup>1+</sup>, Kaouthar Eljounaidi <sup>2</sup>,  
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

<sup>1</sup>UK Centre for Ecology & Hydrology, Wallingford, England, UK<sup>2</sup>Department of Biology, University of York, York, England, UK

+ Deceased author

---

**V1** First published: 30 Aug 2023, 8:375  
<https://doi.org/10.12688/wellcomeopenres.19840.1>  
Latest published: 30 Aug 2023, 8:375  
<https://doi.org/10.12688/wellcomeopenres.19840.1>

---

## Abstract

We present a genome assembly from an individual female *Autographa pulchrina* (the Beautiful Golden Y; Arthropoda; Insecta; Lepidoptera; Noctuidae). The genome sequence is 426.2 megabases in span. Most of the assembly is scaffolded into 32 chromosomal pseudomolecules, including the Z and W sex chromosomes. The mitochondrial genome has also been assembled and is 15.25 kilobases in length. Gene annotation of this assembly on Ensembl identified 12,916 protein coding genes.

## Keywords




*Autographa pulchrina*, Beautiful Golden Y, genome sequence, chromosomal, Lepidoptera



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

## Open Peer Review

**Approval Status**  

	1	2
<b>version 1</b> 30 Aug 2023	 <a href="#">view</a>	 <a href="#">view</a>
1. <b>Reza Zahiri</b> , State Museum of Natural History Karlsruhe, Karlsruhe, Germany		
2. <b>Marko Mutanen</b>  , University of Oulu, Oulu, Finland		

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:** **Boyes D:** Investigation, Resources; **Eljounaidi K:** 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:** © 2023 Boyes D *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:** Boyes D, Eljounaidi K, University of Oxford and Wytham Woods Genome Acquisition Lab *et al.* **The genome sequence of the Beautiful Golden Y, *Autographa pulchrina* (Haworth, 1809) [version 1; peer review: 2 approved]** Wellcome Open Research 2023, 8:375 <https://doi.org/10.12688/wellcomeopenres.19840.1>

**First published:** 30 Aug 2023, 8:375 <https://doi.org/10.12688/wellcomeopenres.19840.1>

## Species taxonomy

Eukaryota; Metazoa; Eumetazoa; Bilateria; Protostomia; Ecdysozoa; Panarthropoda; Arthropoda; Mandibulata; Pancrustacea; Hexapoda; Insecta; Dicondylia; Pterygota; Neoptera; Endopterygota; Amphiesmenoptera; Lepidoptera; Glossata; Neolepidoptera; Heteroneura; Ditrysia; Obtectomera; Noctuoidea; Noctuidae; Plusiinae; *Autographa*; *Autographa pulchrina* (Haworth, 1809) (NCBI:txid987893).

## Background

*Autographa pulchrina*, also referred to as the Beautiful Golden Y, is a member of the Noctuidae family. This family significantly contributes to the overall diversity of the Lepidoptera order, accounting for more than 12,000 described species and thousands more to be described (Keegan *et al.*, 2021). The Beautiful Golden Y has a striking appearance, with a forewing which is predominantly dark reddish-brown hue, extensively marbled with light purplish brown with a characteristic kidney-shaped mark, sharply pinched and finely edged in golden colour (Waring *et al.*, 2017). This mark gives the common name 'Golden Y' to Beautiful Golden Y and several other 'Y' species. *A. pulchrina* is similar in appearance to its close relative, the Plain Golden Y (*Autographa jota*), as both are reddish with a similar Y mark that is usually separated into a dot and a V, and they also have a similar flight season. However, they may be distinguished by close comparison of the forewings (British Lepidoptera, 2018; Wagner, 2023).

*A. pulchrina* is commonly found across Europe and has also been recorded in Russia and Kazakhstan (GBIF Secretariat, 2023). In Britain, it is widely distributed and commonly sighted, inhabiting diverse environments such as woodlands, gardens, moors, hedgerows, and heathlands (Waring *et al.*, 2017). This species has one generation, and is on the wing in June and July, slightly earlier than *A. jota*. The larvae, which feed on low plants, overwinter in this stage, and pupate in May or June (Kimber, 2023). The larvae are polyphagous and have been recorded feeding on plants such as nettle (*Urtica*) and honeysuckle (*Lonicera*) in Britain (Robinson *et al.*, 2023).

The genome of *A. pulchrina* was sequenced as part of the Darwin Tree of Life Project, a collaborative effort to sequence all named eukaryotic species in the Atlantic Archipelago of Britain and Ireland. The genome sequence will enrich the DNA repertoire for phylogenetic and evolutionary investigations and will lead to better understanding of the diversification dynamics of the Lepidoptera order.

## Genome sequence report

The genome was sequenced from one female *Autographa pulchrina* (Figure 1) collected from Wytham Woods, Oxfordshire, UK (51.77, -1.34). A total of 47-fold coverage in Pacific Biosciences single-molecule HiFi long reads and 80-fold coverage in 10X Genomics read clouds was generated. Primary assembly contigs were scaffolded with chromosome conformation Hi-C data. Manual assembly curation corrected 93 missing joins or misjoins and removed 31 haplotypic



**Figure 1.** Photograph of the *Autographa pulchrina* (ilAutPulc1) specimen used for genome sequencing.

duplications, reducing the assembly length by 1.26% and the scaffold number by 33.82%, and increasing the scaffold N50 by 4.79%.

The final assembly has a total length of 426.2 Mb in 135 sequence scaffolds with a scaffold N50 of 13.5 Mb (Table 1). Most (99.03%) of the assembly sequence was assigned to 32 chromosomal-level scaffolds, representing 30 autosomes and the Z and W sex chromosomes. 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 56.1 with *k*-mer completeness of 99.99%, and the assembly has a BUSCO v5.3.2 completeness of 98.9% (single = 98.6%, duplicated = 0.3%), using the lepidoptera\_odb10 reference set ( $n = 5,286$ ).

Metadata for specimens, spectral estimates, sequencing runs, contaminants and pre-curation assembly statistics can be found at <https://links.tol.sanger.ac.uk/species/987893>.

## Genome annotation report

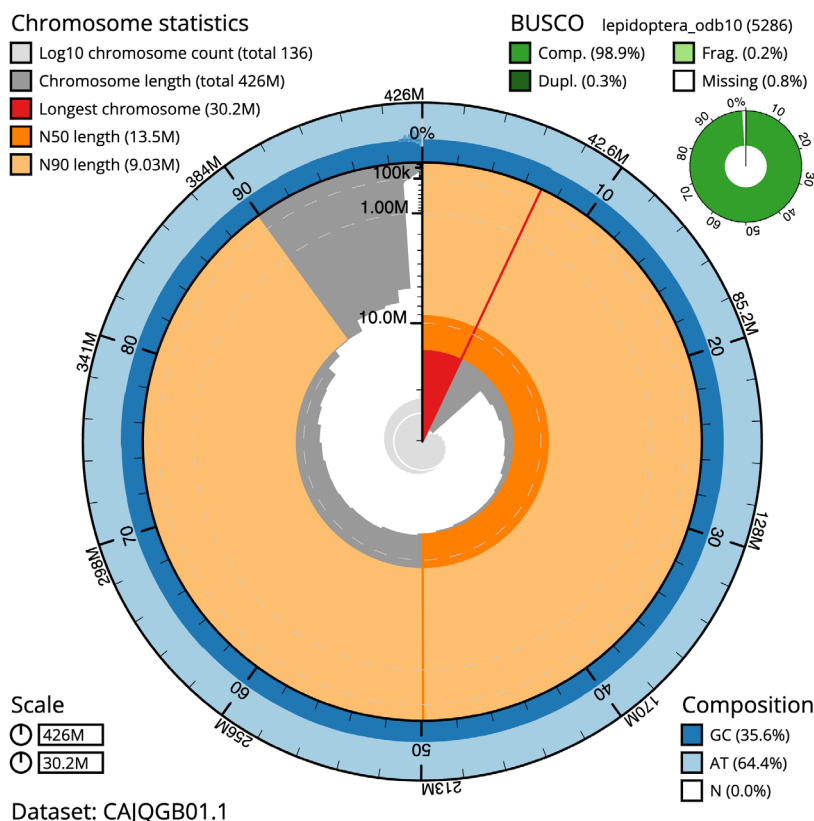
The *Autographa pulchrina* genome assembly (GCA\_905475315.1) was annotated using the Ensembl rapid annotation pipeline (Table 1; [https://rapid.ensembl.org/Autographa\\_pulchrina\\_GCA\\_905475315.1/Info/Index](https://rapid.ensembl.org/Autographa_pulchrina_GCA_905475315.1/Info/Index)). The resulting annotation includes 24,965 transcribed mRNAs from 12,916 protein-coding and 3,056 non-coding genes.

**Table 1. Genome data for *Autographa pulchrina*, ilAutPulc1.1.**

Project accession data		
Assembly identifier	ilAutPulc1.1	
Species	<i>Autographa pulchrina</i>	
Specimen	ilAutPulc1	
NCBI taxonomy ID	987893	
BioProject	PRJEB43793	
BioSample ID	SAMEA7520527	
Isolate information	ilAutPulc1, female: head and thorax (DNA sequencing and Hi-C), abdomen (RNA sequencing)	
Assembly metrics*		Benchmark
Consensus quality (QV)	56.1	$\geq 50$
k-mer completeness	99.99%	$\geq 95\%$
BUSCO**	C:98.9%[S:98.6%,D:0.3%], F:0.2%,M:0.8%,n:5,286	$C \geq 95\%$
Percentage of assembly mapped to chromosomes	99.03%	$\geq 95\%$
Sex chromosomes	Z and W chromosomes	<i>localised homologous pairs</i>
Organelles	Mitochondrial genome assembled	<i>complete single alleles</i>
Raw data accessions		
PacificBiosciences SEQUEL II	ERR6412366	
10X Genomics Illumina	ERR6054588, ERR6054586, ERR6054587, ERR6054589	
Hi-C Illumina	ERR6054585	
PolyA RNA-Seq Illumina	ERR6054590	
Genome assembly		
Assembly accession	GCA_905475315.1	
Accession of alternate haplotype	GCA_905475435.1	
Span (Mb)	426.2	
Number of contigs	213	
Contig N50 length (Mb)	9.6	
Number of scaffolds	135	
Scaffold N50 length (Mb)	13.5	
Longest scaffold (Mb)	30.2	
Genome annotation		
Number of protein-coding genes	12,916	
Number of non-coding genes	3,056	
Number of gene transcripts	24,965	

\* 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 lepidoptera\_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/ilAutPulc1.1/dataset/CAJQGB01.1/busco>.



**Figure 2. Genome assembly of *Autographa pulchrina*, ilAutPulc1.1: 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 426,240,943 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 (30,169,485 bp, shown in red). Orange and pale-orange arcs show the N50 and N90 scaffold lengths (13,453,062 and 9,030,163 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 lepidoptera\_odb10 set is shown in the top right. An interactive version of this figure is available at <https://blobtoolkit.genomehubs.org/view/ilAutPulc1.1/dataset/CAJQGB01.1/snail>.

## Methods

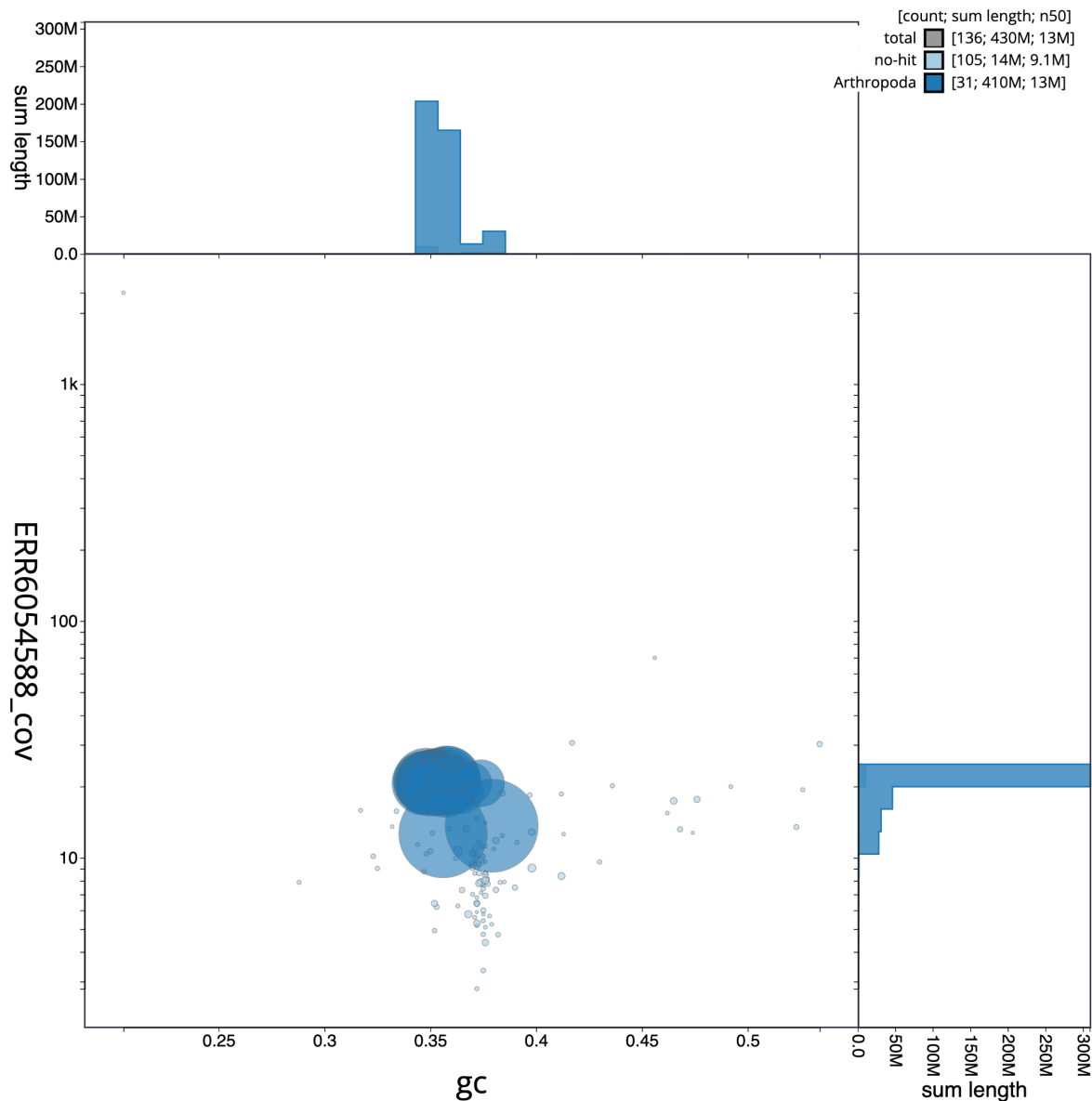
### Sample acquisition and nucleic acid extraction

A female *Autographa pulchrina* (specimen ID Ox000405, individual ilAutPulc1) was collected from Wytham Woods, Oxfordshire (biological vice-county Berkshire), UK (latitude 51.77, longitude -1.34) on 2020-05-22 using a light trap. The specimen was collected and identified by Douglas Boyes (University of Oxford) and preserved on dry ice in a CoolRack.

DNA was extracted at the Tree of Life laboratory, Wellcome Sanger Institute (WSI). The ilAutPulc1 sample was weighed and dissected on dry ice with tissue set aside for Hi-C sequencing. Head and thorax tissue was disrupted using a Nippi Power-masher fitted with a BioMasher pestle. High molecular weight (HMW) DNA was extracted using the Qiagen MagAttract HMW DNA extraction kit. Low molecular weight DNA was removed from a 20 ng aliquot of extracted DNA using the

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 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.

RNA was extracted from abdomen tissue of ilAutPulc1 in the Tree of Life Laboratory at the WSI using TRIzol, according to the manufacturer's instructions. RNA was then eluted in 50 µl RNase-free water and its concentration assessed using a Nanodrop spectrophotometer and Qubit Fluorometer using



**Figure 3. Genome assembly of *Autographa pulchrina*, ilAutPulc1.1: 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/ilAutPulc1.1/dataset/CAJQGB01.1/blob>.

the Qubit RNA Broad-Range (BR) Assay kit. Analysis of the integrity of the RNA was done using Agilent RNA 6000 Pico Kit and Eukaryotic Total RNA assay.

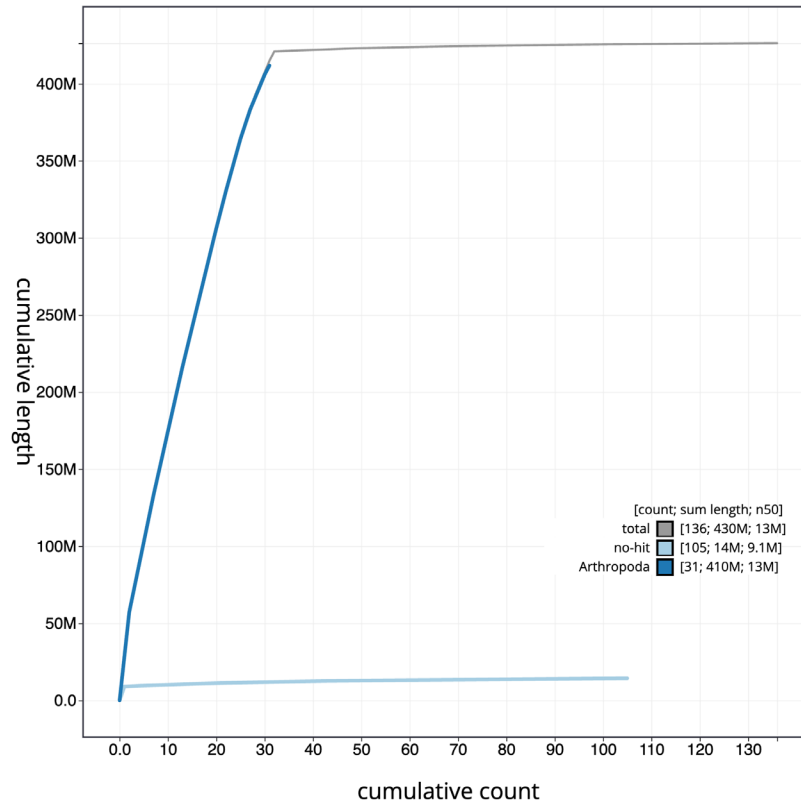
### Sequencing

Pacific Biosciences HiFi circular consensus and 10X Genomics read cloud 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 II (HiFi), Illumina HiSeq 4000 (RNA-Seq) and HiSeq

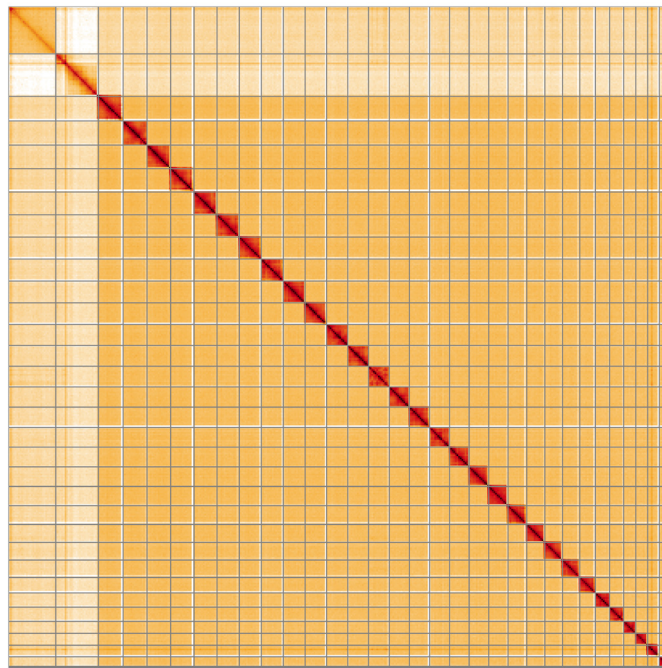
X Ten (10X) instruments. Hi-C data were also generated from remaining head and thorax tissue of ilAutPulc1 using the Arima2 kit and sequenced on the HiSeq X Ten instrument.

### Genome assembly, curation and evaluation

Assembly was carried out with HiCanu (Nurk *et al.*, 2020) and haplotypic duplication was identified and removed with purge\_dups (Guan *et al.*, 2020). One round of polishing was performed by aligning 10X Genomics read data to the assembly with Long Ranger ALIGN, calling variants with FreeBayes (Garrison & Marth, 2012). The assembly was then scaffolded with Hi-C data (Rao *et al.*, 2014) using SALSA2



**Figure 4. Genome assembly of *Autographa pulchrina*, ilAutPulc1.1: 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 buscodegenes taxrule. An interactive version of this figure is available at <https://blobtoolkit.genomehubs.org/view/ilAutPulc1.1/dataset/CAJQGB01.1/cumulative>.



**Figure 5. Genome assembly of *Autographa pulchrina*, ilAutPulc1.1: Hi-C contact map of the ilAutPulc1.1 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=dILrDprpQr-4Kqk1XxixkA>.

**Table 2. Chromosomal pseudomolecules in the genome assembly of *Autographa pulchrina*, iAutPulc1.**

Accession	Chromosome	Length (MB)	GC Percent
FR997763.1	1	15.71	35
FR997764.1	2	15.34	35.5
FR997765.1	3	15.02	35.5
FR997766.1	4	14.88	35.5
FR997767.1	5	14.63	34.5
FR997769.1	6	14.02	35
FR997768.1	7	14.21	34.5
FR997770.1	8	13.99	34.5
FR997771.1	9	13.93	35
FR997772.1	10	13.73	34.5
FR997773.1	11	13.45	34.5
FR997774.1	12	13.24	35
FR997775.1	13	13.22	35.5
FR997776.1	14	12.88	35
FR997777.1	15	12.86	34.5
FR997778.1	16	12.66	35.5
FR997779.1	17	12.54	35
FR997780.1	18	12.51	35
FR997781.1	19	12.27	34.5
FR997782.1	20	11.98	35.5
FR997783.1	21	11.51	34.5
FR997784.1	22	11.23	35
FR997785.1	23	11.07	35.5
FR997786.1	24	9.87	34.5
FR997787.1	25	9.08	35
FR997788.1	26	9.03	34.5
FR997789.1	27	7.7	35.5
FR997790.1	28	7.6	36
FR997791.1	29	7.37	37
FR997792.1	30	6.04	36.5
FR997761.1	W	30.17	37.5
FR997762.1	Z	27.17	35.5

(Ghurye *et al.*, 2019). The assembly was checked for contamination and corrected using the gEVAL system (Chow *et al.*, 2016) as described previously (Howe *et al.*, 2021). Manual curation was performed using gEVAL, HiGlass (Kerpedjiev *et al.*, 2018) and Pretext (Harry, 2022). The mitochondrial genome was assembled using MitoHiFi (Uliano-Silva *et al.*, 2023), 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 (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 3 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 *Autographa pulchrina* assembly (GCA\_905475315.1). Annotation was created primarily through alignment of transcriptomic data to the genome, with gap filling via 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



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

Software tool	Version	Source
BlobToolKit	4.1.7	<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>
FreeBayes	1.3.1-17-gaa2ace8	<a href="https://github.com/freebayes/freebayes">https://github.com/freebayes/freebayes</a>
gEVAL	N/A	<a href="https://geval.org.uk/">https://geval.org.uk/</a>
HiCanu	2.1	<a href="https://github.com/marbl/canu">https://github.com/marbl/canu</a>
HiGlass	1.11.6	<a href="https://github.com/higlass/higlass">https://github.com/higlass/higlass</a>
Long Ranger ALIGN	2.2.2	<a href="https://support.10xgenomics.com/genome-exome/software/pipelines/latest/advanced/other-pipelines">https://support.10xgenomics.com/genome-exome/software/pipelines/latest/advanced/other-pipelines</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>
SALSA	2.2	<a href="https://github.com/salsa-rs/salsa">https://github.com/salsa-rs/salsa</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>

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: *Autographa pulchrina* (beautiful golden Y). Accession number PRJEB43793; <https://identifiers.org/ena.embl/PRJEB43793>. (Wellcome Sanger Institute, 2021) The genome sequence is released openly for reuse. The *Autographa pulchrina* 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.4789928>.

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](#)
- 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](#)
- Bernt M, Donath A, Jühling F, et al.: **MITOS: Improved *de novo* metazoan mitochondrial genome annotation.** *Mol Phylogenet Evol.* 2013; **69**(2): 313–319.  
[PubMed Abstract](#) | [Publisher Full Text](#)
- British Lepidoptera: ***Autographa pulchrina* vs *jota* (Beautiful vs Plain Golden Y).** 2018; [Accessed 25 July 2023].  
[Reference Source](#)
- 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](#)
- 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](#)
- 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](#)
- Garrison E, Marth G: **Haplotype-based variant detection from short-read sequencing.** 2012.  
[Publisher Full Text](#)
- GBIF Secretariat: ***Autographa pulchrina* (Haworth, 1809).** *GBIF Backbone Taxonomy.* 2023; [Accessed 25 July 2023].  
[Reference Source](#)
- Ghurye J, Rhie 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](#)
- 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): g1aa153.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Keegan KL, Rota J, Zahiri R, et al.: **Toward a Stable Global Noctuidae (Lepidoptera) Taxonomy.** *Insect Systematics and Diversity.* 2021; **5**(3).  
[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](#)
- Kimber I: **Beautiful Golden Y *Autographa pulchrina* (Haworth, 1809).** UKMoths. 2023; [Accessed 25 July 2023].  
[Reference Source](#)
- Manni M, Berkeley MR, Seppely 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](#)
- Nurk S, Walenz BP, Rhie A, et al.: **HiCanu: Accurate assembly of segmental duplications, satellites, and allelic variants from high-fidelity long reads.** *Genome Res.* 2020; **30**(9): 1291–1305.  
[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.: **Mercury: Reference-free quality, completeness, and phasing assessment for genome assemblies.** *Genome Biol.* 2020; **21**(1): 245.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Robinson GS, Ackery PR, Kitching I, et al.: **HOSTS - a Database of the World's Lepidopteran Hostplants.** [Data set]. Natural History Museum, 2023; [Accessed 21 April 2023].  
[Publisher 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](#)
- Surana P, Muffato M, Qi G: **sanger-tol/readmapping: sanger-tol/readmapping v1.1.0 - Hebridean Black (1.1.0).** Zenodo. 2023a; [Accessed 21 July 2023].  
[Publisher Full Text](#)
- Surana P, Muffato M, Sadasivan Baby C: **sanger-tol/genomenote (v1.0.dev).** Zenodo. 2023b; [Accessed 21 July 2023].  
[Publisher Full Text](#)
- Uliano-Silva M, Ferreira GJRN, Krashenninnikova 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](#)
- Wagner W: ***Autographa pulchrina* (Haworth, 1809).** *Lepidoptera and their ecology.* 2023; [Accessed 25 July 2023].  
[Reference Source](#)
- Waring P, Townsend M, Lewington R: **Field Guide to the Moths of Great Britain and Ireland: Third Edition.** Bloomsbury Wildlife Guides, 2017.  
[Reference Source](#)
- Wellcome Sanger Institute: **The genome sequence of the Beautiful Golden Y, *Autographa pulchrina* (Haworth, 1809).** European Nucleotide Archive, [dataset], accession number PRJEB43793, 2021.

# Open Peer Review

Current Peer Review Status:  

---

## Version 1

Reviewer Report 16 January 2025

<https://doi.org/10.21956/wellcomeopenres.21972.r88454>

© 2025 Mutanen M. 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.



**Marko Mutanen** 

University of Oulu, Oulu, Finland

This is a well-prepared paper of a widely distributed and well-known species of *Plusiinae*. The genome is very useful for example to address the taxonomic problems in *Plusiinae* and the genus *Autographa*. I find it unfortunate that the authors compare this species with *A. jota*, which is a clearly distinct species, rather than its true sister *A. buraetica* that is present in the mainland Europe and could be easily overlooked. It is also so similar to *pulchrina* that their taxonomic status as separate species have sometimes been questioned. Also, because of this similarity, and possible overlook of *buraetica*, it would have been important to confirm that the studied specimen truly represents *pulchrina*. Based on the voucher photograph, this appears likely, but that this aspect is completely ignored is an issue. That *buraetica* perhaps is not known to occur in British Isles is not an excuse in cases like this. The DNA barcodes appear slightly, but constantly different.

- In the Introduction, its last paragraph, the species name is misspelled once: *A. pulichrina* > *A. pulchrina*.
- That the larvae feed on "low plants" is a bit misleading as I find its larvae typically in woody bushes such as willows. They themselves mention honeysuckle. I do not find its typical food plants "low plants" while they probably are used too.
- As a general comment that concerns many genome reports is that the species have been introduced like if they were occurring in British Isles only. I would appreciate a bit wider geographic focus with these reports.
- The genome part appears very strong as all the statistics of genome quality suggest so. For example, the BUSCO percentage is very high and well over the benchmark. K-mer completeness value is also very high.

**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:** Molecular taxonomy, phylogenetics, Lepidoptera, Sawflies, DNA barcoding, biodiversity genomics, dark taxa

**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 20 October 2023

<https://doi.org/10.21956/wellcomeopenres.21972.r66279>

© 2023 Zahiri R. 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.



**Reza Zahiri**

Entomology, State Museum of Natural History Karlsruhe, Karlsruhe, Baden-Württemberg, Germany

Review of "The genome sequence of the Beautiful Golden Y, *Autographa pulchrina*" by Douglas Boyes and Kaouthar Eljounaidi

This is a well-written paper, very straightforward and transparent from the Wellcome Centre and Darwin Tree of Life project. There's nothing much to report except a few minor comments can be found below:

**Abstract**

The species name should come with the author's name as appears for the first time in the paper.

How if you modify it this way:

... an individual female the Beautiful Golden Y (*Autographa pulchrina* (Haworth, 1809): Arthropoda; Insecta; ...

**Keywords**

Usually, keywords are meant to be different from the title, e.g., use Plusiinae, Noctuoidea, genome assembly

### Background

Line 1. ... member of the family Noctuidae.

Line 2. of the order Lepidoptera, ...

Line 3. this paper should be cited as well:

<https://www.mapress.com/zt/article/view/zootaxa.3148.1.41>

Line 7: Plain Golden Y (*Autographa jota*) needs the author's name

Line 8: by close comparison of the forewings - and probably genitalic characteristics?

Last paragraph. The genome of *A. pulichrina* was sequenced as part -> The genome of *A. pulichrina* is sequenced as part

Last line. dynamics of the Lepidoptera order -> dynamics of the order Lepidoptera

### Genome sequence report

Line 1. how is this specimen collected?

### Methods

Para. 1. ... and preserved on dry ice in a CoolRack. -> without being killed by a killing agent, if so please indicate.

### References

1. VAN NIEUKERKEN E, KAILA L, KITCHING I, KRISTENSEN N, et al.: Order Lepidoptera Linnaeus, 1758. In: Zhang, Z.-Q. (Ed.) Animal biodiversity: An outline of higher-level classification and survey of taxonomic richness. *Zootaxa*. 2011; **3148** (1). [Publisher Full Text](#)

### 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 biologist

**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.**

Reader Comment 20 Jan 2025

**jim mallet**

I believe Reza means <https://doi.org/10.11646/zootaxa.3148.1.41> rather than the link given in the review

***Competing Interests:*** No competing interests were disclosed.

