Check for updates

DATA NOTE

REVISED The genome sequence of the 6-spot burnet, *Zygaena filipendulae* (Linnaeus, 1758) [version 2; peer review: 3 approved with reservations]

Douglas Boyes[1+], Liam M. Crowley[2], 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, Chelsea Skojec[3], David Plotkin[4], Akito Y. Kawahara[4], Darwin Tree of Life Consortium

[1]UK Centre for Ecology and Hydrology, Wallingford, Oxfordshire, UK
[2]Department of Zoology, University of Oxford, Oxford, UK
[3]Kawahara Lab, University of Florida, Gainesville, USA
[4]McGuire center for Lepidoptera and Biodiversity, Florida Museum of Natural History, Gainsville, USA

[+] Deceased author

v2 **First published:** 01 Aug 2022, **7**:197
https://doi.org/10.12688/wellcomeopenres.17924.1
Latest published: 18 Mar 2024, **7**:197
https://doi.org/10.12688/wellcomeopenres.17924.2

Abstract

We present a genome assembly from an individual female *Zygaena filipendulae* (6-spot burnet; Arthropoda; Insecta; Lepidoptera; Zygaenidae). The genome sequence is 365.9 megabases in span. The majority of the assembly (99.99%) is scaffolded into 31 chromosomal pseudomolecules, with the W and Z sex chromosomes assembled. The complete mitochondrial genome was also assembled and is 15.6 kilobases in length. Gene annotation of this assembly on Ensembl has identified 12,493 protein coding genes.

Keywords

Zygaena filipendulae, 6-spot burnet, genome sequence, chromosomal, Arthropoda

This article is included in the Tree of Life gateway.

Open Peer Review

Approval Status ? ? ?

	1	2	3
version 2 (revision) 18 Mar 2024			
version 1 01 Aug 2022	? view	? view	? view

1. **Wannes Dermauw**, Flanders Research Institute for Agriculture, Fisheries and Food, Merelbeke, Belgium
2. **Rachel A. Steward**, Lunds Universitet, Lund, Sweden
3. **Érika de Castro**, Cambridge University, Cambridge, 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: Boyes D: Investigation, Resources; **Crowley LM**: Investigation, Resources; **Skojec C**: Writing – Original Draft Preparation; **Plotkin D**: Writing – Original Draft Preparation; **Kawahara AY**: 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, 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.

Copyright: © 2024 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, Crowley LM, University of Oxford and Wytham Woods Genome Acquisition Lab *et al.* **The genome sequence of the 6-spot burnet, *Zygaena filipendulae* (Linnaeus, 1758) [version 2; peer review: 3 approved with reservations]** Wellcome Open Research 2024, **7**:197 https://doi.org/10.12688/wellcomeopenres.17924.2

First published: 01 Aug 2022, **7**:197 https://doi.org/10.12688/wellcomeopenres.17924.1

REVISED **Amendments from Version 1**

We have corrected the number of single-copy orthologues for lepidoptera_odb10.

Any further responses from the reviewers can be found at the end of the article

Species taxonomy

Eukaryota; Metazoa; Ecdysozoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Lepidoptera; Glossata; Ditrysia; Zygaenoidea; Zygaenidae; Zygaeninae; Zygaena; *Zygaena filipendulae* (Linnaeus, 1758) (NCBI:txid287375).

Background

The six-spot burnet moth, *Zygaena filipendulae* (Linnaeus, 1758) is an aposematic, chemically defended, day-flying moth in the family *Zygaenidae* with a distribution that ranges across Europe. There are 98 described species of burnet moths in *Zygaena* (Hofmann & Gerald Tremewan, 2005). Some *Zygaena* species have become model organisms to study the evolution of chemical defences (Zagrobelny *et al.*, 2019). Forewings of *Z. filipendulae* are black and distinctively marked with six red spots. This species can biosynthesize cyanogenic glucosides *de novo*, or obtain them from *Fabaceae* host plants, storing cyanoglucosides in cuticular cavities and hemolymph, for later use as a defensive secretion (Franzl *et al.*, 1986). The three enzymes involved in the evolution of biosynthesis in *Z. filipendulae* are two cytochrome P450s and a UDP-glycosyltransferase (Zagrobelny *et al.*, 2019). A genome of *Z. filipendulae* is much needed especially in order to understand the genetics of cyanogenic glucoside biosynthesis.

Genome sequence report

The genome was sequenced from a single female *Z. filipendulae* collected from Ant Hills region, Wytham, Berkshire, UK (Figure 1). A total of 58-fold coverage in Pacific Biosciences single-molecule HiFi long reads and 92-fold coverage in 10X Genomics read clouds were generated. Primary assembly contigs were scaffolded with chromosome conformation Hi-C data. Manual assembly curation corrected 3 missing/misjoins and removed 0 haplotypic duplications, reducing the assembly size by 0.004% and the scaffold number by 8.33% and the scaffold N50 remained the same.

The final assembly has a total length of 365.9 Mb in 55 sequence scaffolds with a scaffold N50 of 12.6 Mb (Table 1). The

Figure 1. Image of the *Zygaena filipendulae* specimen (ilZygFili1) taken prior to preservation and processing.

Table 1. Genome data for *Zygaena filipendulae*, ilZygFili1.2.

Project accession data	
Assembly identifier	ilZygFili1.2
Species	*Zygaena filipendulae*
Specimen	ilZygFili1 (genome assembly); ilZygFili2 (Hi-C); ilZygFili3 (RNA-Seq)
NCBI taxonomy ID	287375
BioProject	PRJEB44832
BioSample ID	SAMEA7519846
Isolate information	Female, whole organism (ilZygFili1); head/thorax tissue (ilZygFili2); abdomen tissue (ilZygFili3)
Raw data accessions	
PacificBiosciences SEQUEL II	ERR6436369
10X Genomics Illumina	ERR6054694-ERR6054701; ERR6054703-ERR6054706
Hi-C Illumina	ERR6054702
PolyA RNA-Seq Illumina	ERR9434973
Genome assembly	
Assembly accession	GCA_907165275.2
Accession of alternate haplotype	GCA_907165265.2
Span (Mb)	365.9
Number of contigs	68
Contig N50 length (Mb)	12.6
Number of scaffolds	55
Scaffold N50 length (Mb)	12.6
Longest scaffold (Mb)	16.1
BUSCO* genome score	C:97.8%[S:97.3%,D:0.5%], F:0.5%,M:1.7%,n:5286
Genome annotation	
Number of protein-coding genes	12,493
Average length of coding sequence (bp)	11,524.94
Average number of exons per transcript	6.70

*BUSCO scores based on the lepidoptera_odb10 BUSCO set using v5.2.2. 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/ilZygFili1.2/dataset/CAJRBF02/busco#Filters.

majority, 99.99%, of the assembly sequence was assigned to 31 chromosomal-level scaffolds, representing 29 autosomes (numbered by sequence length) and the W and Z sex chromosomes (Figure 2–Figure 5; Table 2).

The assembly has a BUSCO v5.2.2 (Manni *et al.*, 2021) completeness of 97.8% (single 97.3%, duplicated 0.5%) using the lepidoptera_odb10 reference set (n=5,286). While not fully phased, the assembly deposited is of one haplotype.

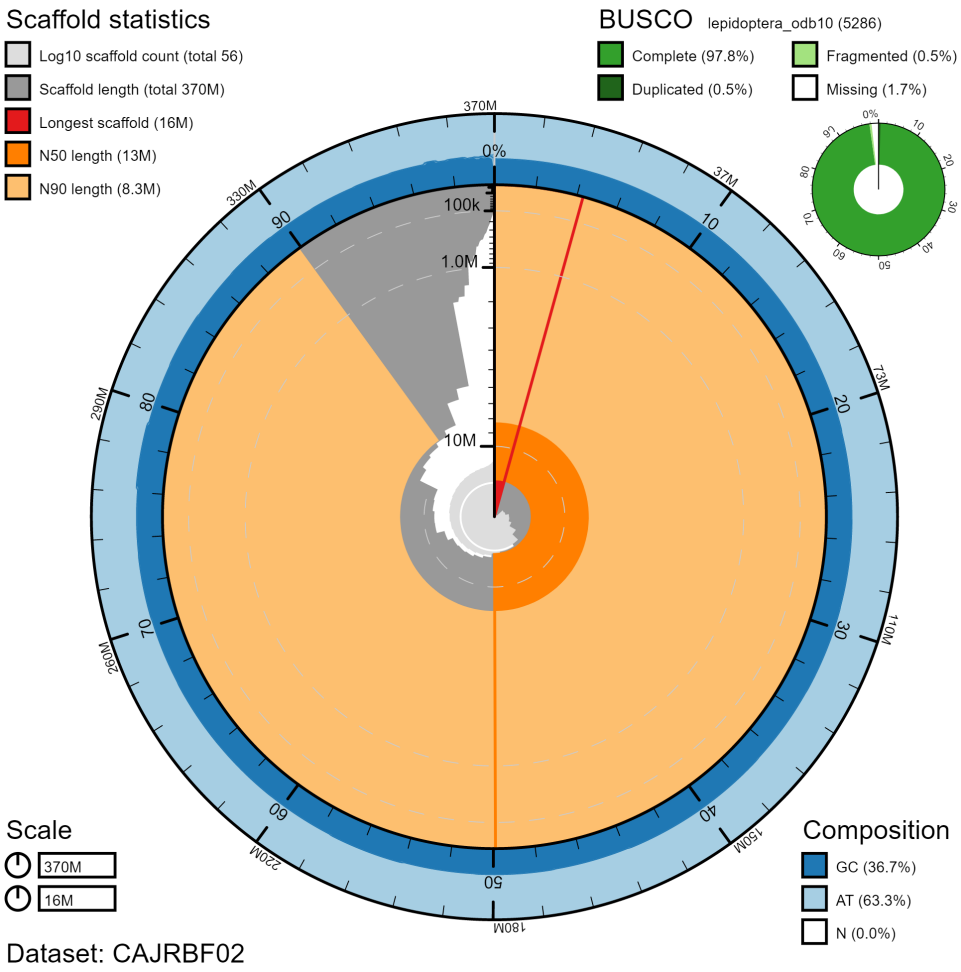

Figure 2. Genome assembly of *Zygaena filipendulae*, ilZygFili1.2: metrics. The BlobToolKit Snailplot shows N50 metrics and BUSCO gene completeness. The main plot is divided into 1,000 size-ordered bins around the circumference with each bin representing 0.1% of the 365,946,273 bp assembly. The distribution of chromosome lengths is shown in dark grey with the plot radius scaled to the longest chromosome present in the assembly (16,101,494 bp, shown in red). Orange and pale-orange arcs show the N50 and N90 chromosome lengths (12,640,274 and 8,250,661 bp), respectively. The pale grey spiral shows the cumulative chromosome 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/ilZygFili1.2/dataset/CAJRBF02/snail#Filters.

Contigs corresponding to the second haplotype have also been deposited.

Genome annotation report

The ilZygFili1.1 genome has been annotated using the Ensembl rapid annotation pipeline (Table 1; GCA_907165275.1 https://rapid.ensembl.org/Zygaena_filipendulae_GCA_907165275.1/Info/Index). The resulting annotation includes 20,201 transcribed mRNAs from 12,493 protein-coding and 1,770 non-coding genes.

Methods

Sample acquisition and nucleic acid extraction

Two *Z. filipendulae* specimens (ilZygFili1, genome assembly; and ilZygFili3, RNA-Seq) were collected using a net from Ant Hills region and Wytham woods, Wytham, Berkshire, UK (latitude 51.765, longitude -1.327) by Douglas Boyes (University of Oxford). The specimens were identified by Douglas Boyes and snap-frozen on dry ice. A further *Z. filipendulae* specimen (ilZygFili2, Hi-C) was collected using a net from Wytham woods, Berkshire, UK (latitude 51.771, longitude -1.338)

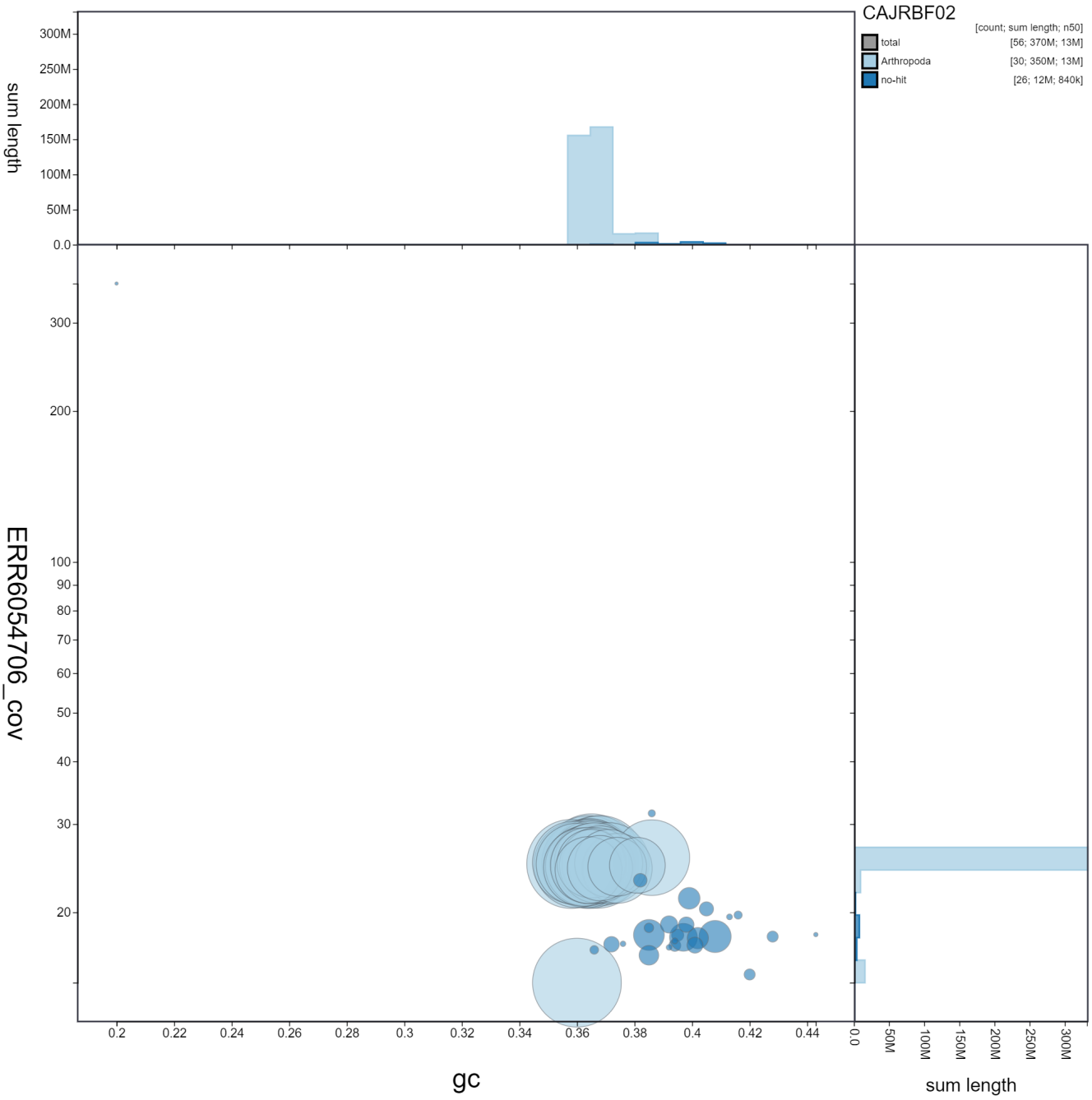

Figure 3. Genome assembly of *Zygaena filipendulae*, ilZygFili1.2: GC coverage. BlobToolKit GC-coverage plot. Scaffolds are coloured by phylum. Circles are sized in proportion to chromosome length Histograms show the distribution of chromosome length sum along each axis. An interactive version of this figure is available at https://blobtoolkit.genomehubs.org/view/ilZygFili1.2/dataset/CAJRBF02/blob#Filters.

by Liam Crowley (University of Oxford). The specimen was identified by Liam Crowley and snap-frozen on dry ice.

DNA was extracted at the Tree of Life laboratory, Wellcome Sanger Institute. The ilZygFili1 sample was weighed and dissected on dry ice. Whole organism tissue was cryogenically disrupted to a fine powder using a Covaris cryoPREP Automated Dry Pulveriser, receiving multiple impacts. Fragment size analysis of 0.01–0.5 ng of DNA was then performed using an Agilent FemtoPulse. High molecular weight (HMW) DNA was extracted using the Qiagen MagAttract HMW DNA extraction kit. Low molecular weight DNA was removed from

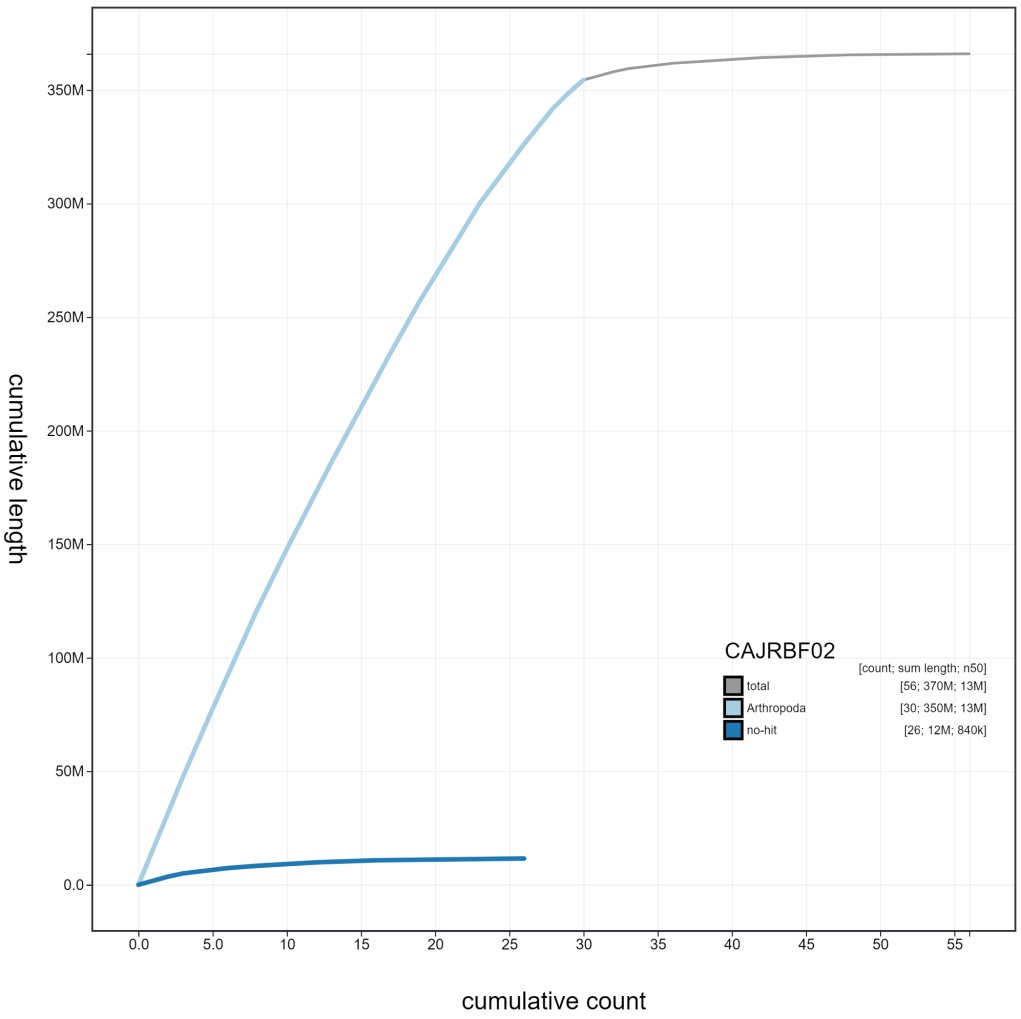

Figure 4. Genome assembly of *Zygaena filipendulae*, ilZygFili1.2: cumulative sequence. BlobToolKit cumulative sequence plot. The grey line shows cumulative length for all scaffolds. Coloured lines show cumulative lengths of scaffolds assigned to each phylum using the buscogenes taxrule. An interactive version of this figure is available at https://blobtoolkit.genomehubs.org/view/ilZygFili1.2/dataset/CAJRBF02/cumulative#Filters.

a 200-ng aliquot of extracted DNA using 0.8X AMpure XP purification kit prior to 10X Chromium sequencing; a minimum of 50 ng DNA was submitted for 10X sequencing. HMW DNA was sheared into an average fragment size between 12–20 kb in a Megaruptor 3 system with speed setting 30. Sheared DNA was purified by solid-phase reversible immobilisation using AMPure PB beads with a 1.8X ratio of beads to sample to remove the shorter fragments and concentrate the DNA sample. The concentration of the sheared and purified DNA was assessed using a Nanodrop spectrophotometer and Qubit Fluorometer and Qubit dsDNA High Sensitivity Assay kit. Fragment size distribution was evaluated by running the sample on the FemtoPulse system.

RNA was extracted from other abdomen tissue of ilZygFili3 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 RNA assessed using a Nanodrop spectrophotometer and Qubit Fluorometer using 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 Chromium read cloud sequencing libraries were constructed according to the manufacturers' instructions. Sequencing was performed by the Scientific Operations core at the Wellcome Sanger Institute on Pacific Biosciences SEQUEL II (HiFi), Illumina NovaSeq 6000 (10X) and Illumina HiSeq 4000 (RNA-Seq) instruments. Hi-C data were generated in the Tree of Life laboratory from head/thorax tissue of ilZygFili2

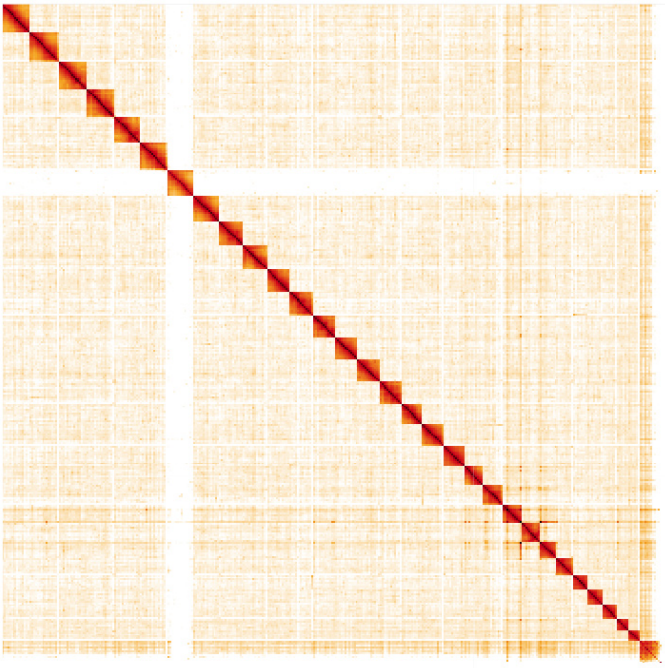

Figure 5. Genome assembly of *Zygaena filipendulae*, ilZygFili1.2: Hi-C contact map. Hi-C contact map of the ilZygFili1.2 assembly, visualised in HiGlass. Chromosomes are arranged in size order from left to right and top to bottom. The interactive Hi-C map can be viewed at https://genome-note-higlass.tol.sanger.ac.uk/l/?d=Aqyc_jJbQjuSzW9eMHqPQg.

Table 2. Chromosomal pseudomolecules in the genome assembly of *Zygaena filipendulae*, ilZygFili1.2.

INSDC accession	Chromosome	Size (Mb)	GC%
OU015649.1	1	16.1	36.5
OU015650.1	2	15.74	36.4
OU015651.1	3	15.74	36.7
OU015652.1	4	15.03	36.8
OU015653.1	5	14.95	36.4
OU015654.1	6	14.63	35.8
OU015656.1	7	14.34	36.3
OU015657.1	8	13.67	36
OU015658.1	9	12.99	36.5
OU015659.1	10	12.78	36.6
OU015660.1	11	12.77	35.9
OU015661.1	12	12.64	36.5
OU015662.1	13	12.32	36
OU015663.1	14	12.14	37.1
OU015664.1	15	11.96	36.3
OU015665.1	16	11.88	36.5
OU015666.1	17	11.8	36.6
OU015667.1	18	11.37	36.2
OU015668.1	19	10.74	36.8
OU015669.1	20	10.72	36.4
OU015670.1	21	10.57	38.6
OU015671.1	22	10.5	36.7
OU015672.1	23	9.02	37.4
OU015673.1	24	8.7	36.5
OU015674.1	25	8.53	37.1
OU015675.1	26	8.25	36.4
OU015676.1	27	7.85	36.8
OU015677.1	28	6.36	37.4
OU015678.1	29	5.72	38.1
OU015679.1	W	1.87	40.8
OU015655.1	Z	14.6	36
OU015680.1	MT	0.02	19.9
-	Unplaced	9.65	39.4

using the Arima v2 kit and sequenced on a NovaSeq 6000 instrument.

Genome assembly

Assembly was carried out with Hifiasm (Cheng *et al.*, 2021); haplotypic duplication was identified and removed with purge_dups (Guan *et al.*, 2020). One round of polishing was performed by aligning 10X Genomics read data to the assembly with longranger align, calling variants with freebayes (Garrison & Marth, 2012). The assembly was then scaffolded with Hi-C data (Rao *et al.*, 2014) using SALSA2 (Ghurye *et al.*, 2019). The assembly was checked for contamination and corrected using the gEVAL system (Chow *et al.*, 2016) as described previously (Howe *et al.*, 2021). Manual curation (Howe *et al.*, 2021) was performed using gEVAL, HiGlass (Kerpedjiev *et al.*, 2018) and Pretext. The mitochondrial genome was assembled using MitoHiFi (Uliano-Silva *et al.*, 2021), which performs annotation using MitoFinder (Allio *et al.*, 2020). The genome was analysed and BUSCO scores generated within the BlobToolKit environment (Challis *et al.*, 2020). Table 3 contains a list of all software tool versions used, where appropriate.

Table 3. Software tools used.

Software tool	Version	Source
Hifiasm	0.14-r312	Cheng *et al.*, 2021
purge_dups	1.2.3	Guan *et al.*, 2020
SALSA2	2.2	Ghurye *et al.*, 2019
longranger align	2.2.2	https://support.10xgenomics.com/genome-exome/software/pipelines/latest/advanced/other-pipelines
freebayes	1.3.1-17-gaa2ace8	Garrison & Marth, 2012
MitoHiFi	2.11	Uliano-Silva *et al.*, 2021
HiGlass	1.11.6	Kerpedjiev *et al.*, 2018
PretextView	0.2.x	https://github.com/wtsi-hpag/PretextView
BlobToolKit	3.0.5	Challis *et al.*, 2020

Genome annotation

The Ensembl gene annotation system (Aken *et al.*, 2016) was used to generate annotation for the *Z. filipendulae* assembly (GCA_907165275.1). 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).

Ethics/compliance issues

The materials that have contributed to this genome note have been supplied by a Darwin Tree of Life Partner. The submission of materials by a Darwin Tree of Life Partner is subject to the Darwin Tree of Life Project Sampling Code of Practice. By agreeing with and signing up to the Sampling Code of Practice, the Darwin Tree of Life Partner agrees they will meet the legal and ethical requirements and standards set out within this document in respect of all samples acquired for, and supplied to, the Darwin Tree of Life Project. Each transfer of samples is further undertaken according to a Research Collaboration Agreement or Material Transfer Agreement entered into by the Darwin Tree of Life Partner, Genome Research Limited (operating as the Wellcome Sanger Institute), and in some circumstances other Darwin Tree of Life collaborators.

Data availability

European Nucleotide Archive: Zygaena filipendulae (6-spot burnet). Accession number PRJEB44832; https://identifiers.org/ena.embl/PRJEB44832.

The genome sequence is released openly for reuse. The *Z. filipendulae* 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.6418202.

Members of the Darwin Tree of Life Barcoding collective are listed here: https://doi.org/10.5281/zenodo.6418156.

Members of the Wellcome Sanger Institute Tree of Life programme are listed here: https://doi.org/10.5281/zenodo.6418327.

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

Members of the Tree of Life Core Informatics collective are listed here: https://doi.org/10.5281/zenodo.6125046.

Members of the Darwin Tree of Life Consortium are listed here: https://doi.org/10.5281/zenodo.6418363.

References

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

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

Cheng H, Concepcion GT, Feng X, *et al.*: **Haplotype-Resolved *de Novo* Assembly Using Phased Assembly Graphs with Hifiasm.** *Nat Methods.* 2021; **18**(2): 170–75.
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

Franzl S, Nahrstedt A, Naumann CM: **Evidence for Site of Biosynthesis and Transport of the Cyanoglucosides Linamarin and Lotaustralin in Larvae of *Zygaena Trifolii* (Insecta: Lepidoptera).** *J Insect Physiol.* 1986; **32**(8): 705–9.
Publisher Full Text

Garrison E, Marth G: **Haplotype-Based Variant Detection from Short-Read Sequencing**. 2012; arXiv:1207.3907.
Publisher Full Text

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): 2896–98.
PubMed Abstract | Publisher Full Text | Free Full Text

Hofmann A, Gerald Tremewan W: **Contribution to the Knowledge of the Genus Zygaena Fabricius, 1775 in Iran (Lepidoptera, Zygaenidae). Part VII: Records from the Zarde Kuh Region with the Description of a New Species. - Linneana Belgica XX (1): 15-21.** *Linneana Belgica.* 2005.
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

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): 4647–54.
PubMed Abstract | Publisher Full Text | Free Full Text

Rao SS, 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

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

UniProt Consortium: **UniProt: A Worldwide Hub of Protein Knowledge.** *Nucleic Acids Res.* 2019; **47**(D1): D506–15.
PubMed Abstract | Publisher Full Text | Free Full Text

Zagrobelny M, Dalsten L, Hille A: **Colonization of Northern Europe by *Zygaena Filipendulae* (Lepidoptera).** *Ecol Evol.* 2019; **9**(8): 4796–4804.
PubMed Abstract | Publisher Full Text | Free Full Text

Open Peer Review

Current Peer Review Status: ? ? ?

Version 1

Reviewer Report 06 October 2023

https://doi.org/10.21956/wellcomeopenres.19862.r56023

© **2023 de Castro É.** 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.

?

Érika de Castro
[1] Zoology, Cambridge University, Cambridge, Cambridgeshire, UK
[2] Zoology, Cambridge University, Cambridge, Cambridgeshire, UK

This article is a report of genome sequencing and assembly of *Zygaena filipendulae* (Assembly identifier: ilZygFili1.2). Three specimens were used in this study, each for a different type of sequencing: ilZygFili1 for Hi-Fi, ilZygFili2 for Hi-C and ilZygFili3 for RNA-seq. HiFi produces highly accurate long reads which are ideal for a good quality assembly and it was combined with Hi-C to improve assembly at chromosomic level. The RNA-seq data generated in this study was used to make a gene annotation for the new assembly. All software/packages used in this project are listed in the methods and all the generated data is available on ENA (PRJEB44832). The assembly has 31 pseudo chromosomes which is the chromosome number of many related moths. As a female was used for the genome assembly, its sex chromosomes were heterozygous and the W and Z chromosome were also assembled in this genome. This genome assembly will be very important not only for lepidopterists, but also conservation biologists and chemical ecologists: *Z. filipendulae* is one of the most representative species of the *Zygaena* genus and these moths are in decline, with some species already in the red list. Moreover, this species has become a model to study the evolution of cyanogenesis in insects. I think this two topics could be better explained in the rationale for creating the dataset. I, hereby, recommend the article, but do encourage the authors to improve the “Background“ session and also have some minor questions and suggestions:

Format
Line numbers in the pdf would facilitated for reviewers to comment on specific sentences.

Keywords
You could add 'moth' or 'Lepidoptera'

Background

- In “*Some Zygaena species have become model organisms to study the evolution of chemical defence compounds*”, the word “compound” is pleonasm, I would say ‘chemical defences’

instead.

- Missing reference for the sentence "*Forewings of Z. filipendulae are black and distinctively marked with six red spots.*"

- Please rephrase as ' ...are two cytochrome P450s and a UDP-glycosyltransferase

Why is it important to "understand the evolution of cyanogenic glucoside biosynthesis"? As a chemical ecologist, I know this answer and I am very interested on this subject, but I think this would not be appealing for a broad audience. I would rephrase this session focusing on: (1) Populations of Zygaena moths are in decline, with some species even in the red list (Bergman et al., 2020), (2) Cyanogenic glucosides played an important role not only in the diversification Zygaena moths (aposematism and mimicry), but also in its life-history (Females prefer to mate with the most toxic males (Zagrobelny, Bak, Olsen, et al., 2007); They use these compounds as nuptial gifts which are then transferred to their eggs to protect their offspring (Zagrobelny et al., 2014); Larvae balance between biosynthesis and sequestration of cyanogenic glucosides depending on how much of these compounds they can get from their hostplant, but biosynthesis has higher fitness cost (Zagrobelny, Bak, Thorn Ekstrøm, et al., 2007)), and (3) There is a lack of knowledge on the genetic basis of their cyanogenic metabolism and mechanism that allows these compounds to play multiple roles in the biology of *Zygaena filipendulae*. In my opinion, it is not only about how they evolve to de novo biosynthesize their cyanogenic glucosides (CGs), but to sequester CGs (unknown transporter), to balance between biosynthesis and sequestration (unknown), to activate CGs to release cyanide (B-glu), to detoxify cyanide (B-Cas and Rhodanase), to turn-over the CGs when they need nitrogen (unknown), to perceive the CGs in their hostplant and their mate-partners (unknown), etc. The *Z. filipendulae* genome can aid studies in this broad aspect of the evolution of cyanogenic metabolism.

Methods

Sample acquisition and nucleic acid extraction:

- You don't need to specify which author collected and identified which samples.

- Add the manufacturer of TRIzol.

- Currently, there is no explanation in this session about the method used for the DNA extraction of ilZygFili2 for Hi-C. Please add that.

- I checked ENA and I could not find the information on sex and tissue of the sample used for transcriptomics. Please added this metadata info in ENA, if this is not there already, so the data to facilitated reuse.

Questions: Only the abdomen of the specimen ilZygFili3 were used for the RNA extraction and analyses and I would like to know the rationale for this decision. Moreover, there are other transcriptomic datasets of *Zygaena filipendulae* that are available at ENA and these could have being used to improve even more the genome annotation – why were these not used for the annotation?

Figures

Figure 1. Please keep just the first photo of moth where it is possible to see the six red dots and remove the second one and the tube photo. The tube photo could go to supplementary material, if there is tissue left and you think this info could be used in case other researchers want to re-

sequence the same sample – otherwise it is irrelevant for the content of the paper and even confusing, as the tube has a label (FF03455026) that it is different from the samples code (ilZygFili). The legend also does not say to which specimen the photo correspond to (ilZygFili 1, 2 or 3?)

References:

- Bergman, K. O., Burman, J., Jonason, D., Larsson, M. C., Ryrholm, N., Westerberg, L., & Milberg, P. (2020). Clear-cuts are temporary habitats, not matrix, for endangered grassland burnet moths (Zygaena spp.). *Journal of Insect Conservation*, *24*(2), 269–277. https://doi.org/10.1007/S10841-019-00193-3/FIGURES/2

- Zagrobelny, M., Bak, S., Olsen, C. E., & Møller, B. L. (2007). Intimate roles for cyanogenic glucosides in the life cycle of Zygaena filipendulae (Lepidoptera, Zygaenidae). *Insect Biochemistry and Molecular Biology*, *37*(11), 1189–1197. https://doi.org/10.1016/J.IBMB.2007.07.008

- Zagrobelny, M., Bak, S., Thorn Ekstrøm, C., Erik Olsen, C., & Lindberg Møller, B. (2007). The cyanogenic glucoside composition of Zygaena filipendulae (Lepidoptera: Zygaenidae) as effected by feeding on wild-type and transgenic lotus populations with variable cyanogenic glucoside profiles. *Insect Biochemistry and Molecular Biology*, *37*(1), 10–18. https://doi.org/10.1016/J.IBMB.2006.09.008

- Zagrobelny, M., Olsen, C. E., Pentzold, S., Fürstenberg-Hägg, J., Jørgensen, K., Bak, S., Møller, B. L., & Motawia, M. S. (2014). Sequestration, tissue distribution and developmental transmission of cyanogenic glucosides in a specialist insect herbivore. *Insect Biochemistry and Molecular Biology*, *44*(1), 44–53. https://doi.org/10.1016/J.IBMB.2013.11.003

References

1. Bergman K, Burman J, Jonason D, Larsson M, et al.: Clear-cuts are temporary habitats, not matrix, for endangered grassland burnet moths (Zygaena spp.). *Journal of Insect Conservation*. 2020; **24** (2): 269-277 Publisher Full Text
2. Zagrobelny M, Bak S, Olsen CE, Møller BL: Intimate roles for cyanogenic glucosides in the life cycle of Zygaena filipendulae (Lepidoptera, Zygaenidae).*Insect Biochem Mol Biol*. 2007; **37** (11): 1189-97 PubMed Abstract | Publisher Full Text
3. Zagrobelny M, Bak S, Ekstrøm CT, Olsen CE, et al.: The cyanogenic glucoside composition of Zygaena filipendulae (Lepidoptera: Zygaenidae) as effected by feeding on wild-type and transgenic lotus populations with variable cyanogenic glucoside profiles.*Insect Biochem Mol Biol*. 2007; **37** (1): 10-8 PubMed Abstract | Publisher Full Text
4. Zagrobelny M, Olsen CE, Pentzold S, Fürstenberg-Hägg J, et al.: Sequestration, tissue distribution and developmental transmission of cyanogenic glucosides in a specialist insect herbivore.*Insect Biochem Mol Biol*. 2014; **44**: 44-53 PubMed Abstract | Publisher Full Text

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

Partly

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: Chemical ecologist

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 05 October 2023

https://doi.org/10.21956/wellcomeopenres.19862.r66242

© 2023 Steward 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.

?

Rachel A. Steward
[1] Biology, Lunds Universitet, Lund, Skåne County, Sweden
[2] Biology, Lunds Universitet, Lund, Skåne County, Sweden

Boyes & Crowley present a chromosome level, highly complete genome and preliminary gene annotation for *Zygaena filipendulae*, well-studied aposematic moth. This is the only available reference genome for the genus, and represents an important resource for understanding the genomic basis of cyanogenic glucosides sequestration and biosynthesis. Assembly, annotation and data are available from ENA and Darwin Tree of Life, and this genome note provides adequate information about sequencing and assembly methods and final assembly quality. I especially appreciated the table of tools and versions used, with citations. I have several minor comments, but overall this is a useful and concise contribution.

Minor comments:

1. The annotation deserves some kind of quality assessment, e.g., BUSCO on the protein coding transcripts. The authors also failed to mention that the sex of the sample for RNAseq was unknown, and thus it is unclear whether annotation of the W chromosome was also informed by gene expression data.

2. More detail on the gene annotation methods would be useful, including RNA source tissue (thorax, available on ToLQC) and RNA library prep, a brief synopsis of the Ensembl gene annotation system (which aligner was used, which gene prediction software, which protein databases, etc.), whether this annotation includes UTRs, etc.

3. The number of Lepidoptera_odb10 SCOs is 5286 (Manni et al. 2021. Mol Biol Evol). Boyes & Crowley cite it as 954. Either the information on which database was used is incorrect, or the number of SCOs used is incorrect. Please revise.

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?
Partly

Are the datasets clearly presented in a useable and accessible format?
Yes

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Gene expression, alternative splicing, plant-insect interactions, population genetics, ecological 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, however I have significant reservations, as outlined above.

Reviewer Report 03 May 2023

https://doi.org/10.21956/wellcomeopenres.19862.r56020

© 2023 Dermauw W. 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.

?

Wannes Dermauw
[1] Plant Sciences, Flanders Research Institute for Agriculture, Fisheries and Food, Merelbeke, Oost-Vlaanderen, Belgium
[2] Plant Sciences, Flanders Research Institute for Agriculture, Fisheries and Food, Merelbeke, Oost-Vlaanderen, Belgium

Boyes et al. report the genome sequence of a moth species belonging to the Zygaenidae, a family that has been well studied for the evolution of chemical defense compounds. The genome was sequenced from a single female, with a genome assembly size of 365.9 MB. RNAseq and Hi-C data were generated from other specimens and used for annotation and chromosome conformation, respectively. This is a clear and concise report that merits indexing as a data note. I only have a few minor comments:

- Authors provided a BUSCO genome score, but ideally also calculate a BUSCO score for the

predicted gene set (12,493 genes). A brief comparison with other sequenced lepidopteran genomes would also be recommended.

- Page 4: it is not clear to me how a reader can obtain access to the predicted gene set; authors should upload a gff file and the predicted CDS to NCBI or include this data as a supplementary file(s)

- Page 3, 1st paragraph: add the total number of reads that were generated using each sequencing technology

- Page 3: authors should briefly explain how they determined W and Z sex chromosomes

- Methods/Genome Assembly: This section is very general, and I would suggest the authors to include scripts or more details re software settings that were used.

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?
Partly

Are the datasets clearly presented in a useable and accessible format?
Yes

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: molecular entomology, genomics, horizontal gene transfer, pesticide resistance

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.