ERGA Assembly Report

v24.10.15

Tags: ERGA-BGE

TxID	287327
ToLID	**ilZygLaet5**
Species	Zygaena laeta
Class	Insecta
Order	Lepidoptera

Genome Traits	Expected	Observed
Haploid size (bp)	311,592,246	362,582,388
Haploid Number	30 (source: ancestor)	30
Ploidy	2 (source: ancestor)	2
Sample Sex	Unknown	Unknown

EBP metrics summary and curation notes

Obtained EBP quality metric for collapsed: 6.7.Q55

The following metrics were automatically flagged as below EBP recommended standards or different from expected:

. Kmer completeness value is less than 90 for collapsed

Curator notes

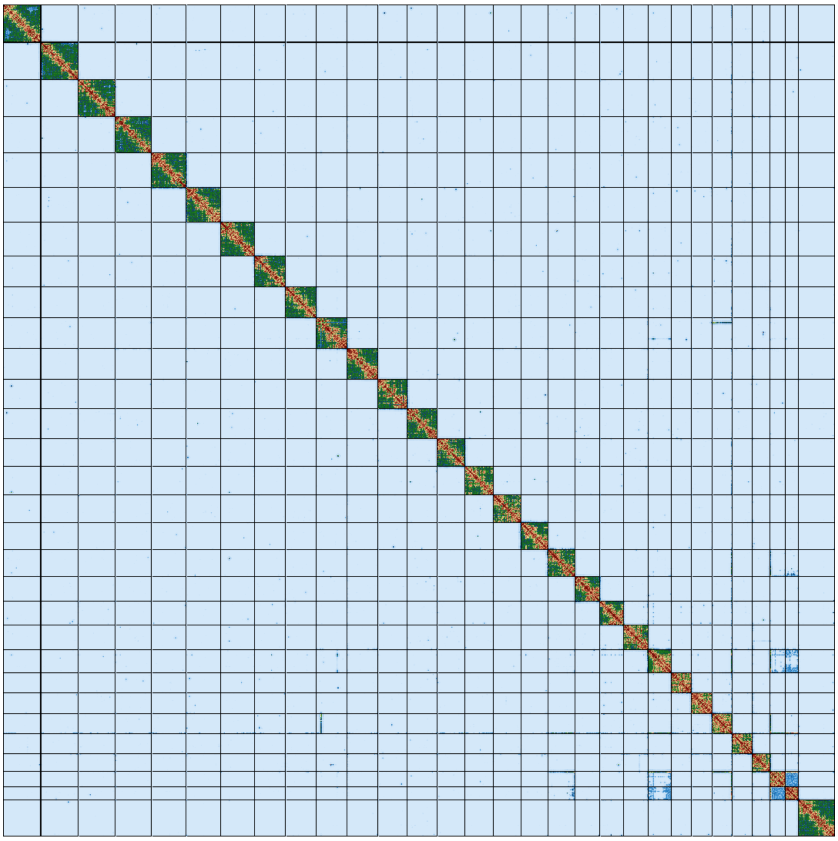
. Interventions/Gb: 52
. Contamination notes: ""
. Other observations: "The assembly of Zygaena laeta (ilZygLaet5.1) is based on 44X PacBio data and 354X of Arima Hi-C data generated as part of the European Reference Genome Atlas (ERGA, https://www.erga-biodiversity.eu/) via the Biodiversity Genomics Europe project (BGE, https://biodiversitygenomics.eu/). The assembly process included the following steps: initial PacBio assembly generation with Hifiasm, removal of contaminant sequences using Context, removal of haplotypic duplications using purge_dups, and Hi-C-based scaffolding with YaHS. In total, 1 contig of 0.016 Mb was identified as contaminant (bacterial). Additionally, 503 regions totaling 37.471 Mb (with the largest being 0.451 Mb) were identified as haplotypic duplications and removed. Finally, the primary assembly was analyzed and manually improved using Pretext. During manual curation, 7 haplotypic regions and 1 contaminant sequence were removed, totaling 3.478 Mb and 0.027 Mb (with the largest being 0.611 Mb and 0.027 Mb). Chromosome-scale scaffolds confirmed by Hi-C data were named in order of size. "

Quality metrics table

Metrics	Pre-curation collapsed	Curated collapsed
Total bp	366,086,161	362,582,388
GC %	36.82	36.83
Gaps/Gbp	789.43	728.11
Total gap bp	41,200	40,600
Scaffolds	71	54
Scaffold N50	13,050,123	13,050,223
Scaffold L50	13	13
Scaffold L90	26	26
Contigs	330	318
Contig N50	1,862,462	1,876,880
Contig L50	58	57
Contig L90	181	176
QV	55.6572	55.6629
Kmer compl.	86.7695	86.6051
BUSCO sing.	95.4%	96.1%
BUSCO dupl.	1.0%	0.4%
BUSCO frag.	1.8%	1.8%
BUSCO miss.	1.7%	1.8%

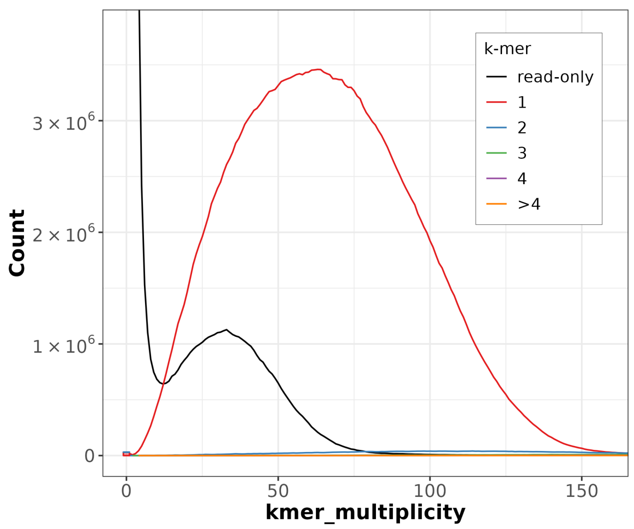
BUSCO: 5.8.2 (euk_genome_met, metaeuk) / Lineage: lepidoptera_odb12 (genomes:79, BUSCOs:5760)

HiC contact map of curated assembly

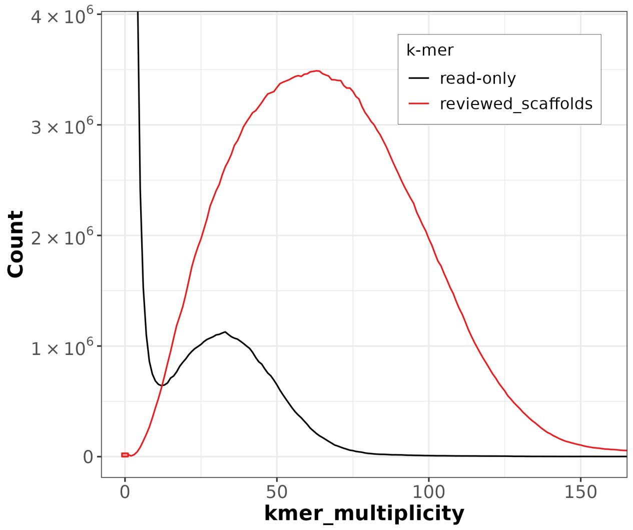


collapsed [LINK]

K-mer spectra of curated assembly

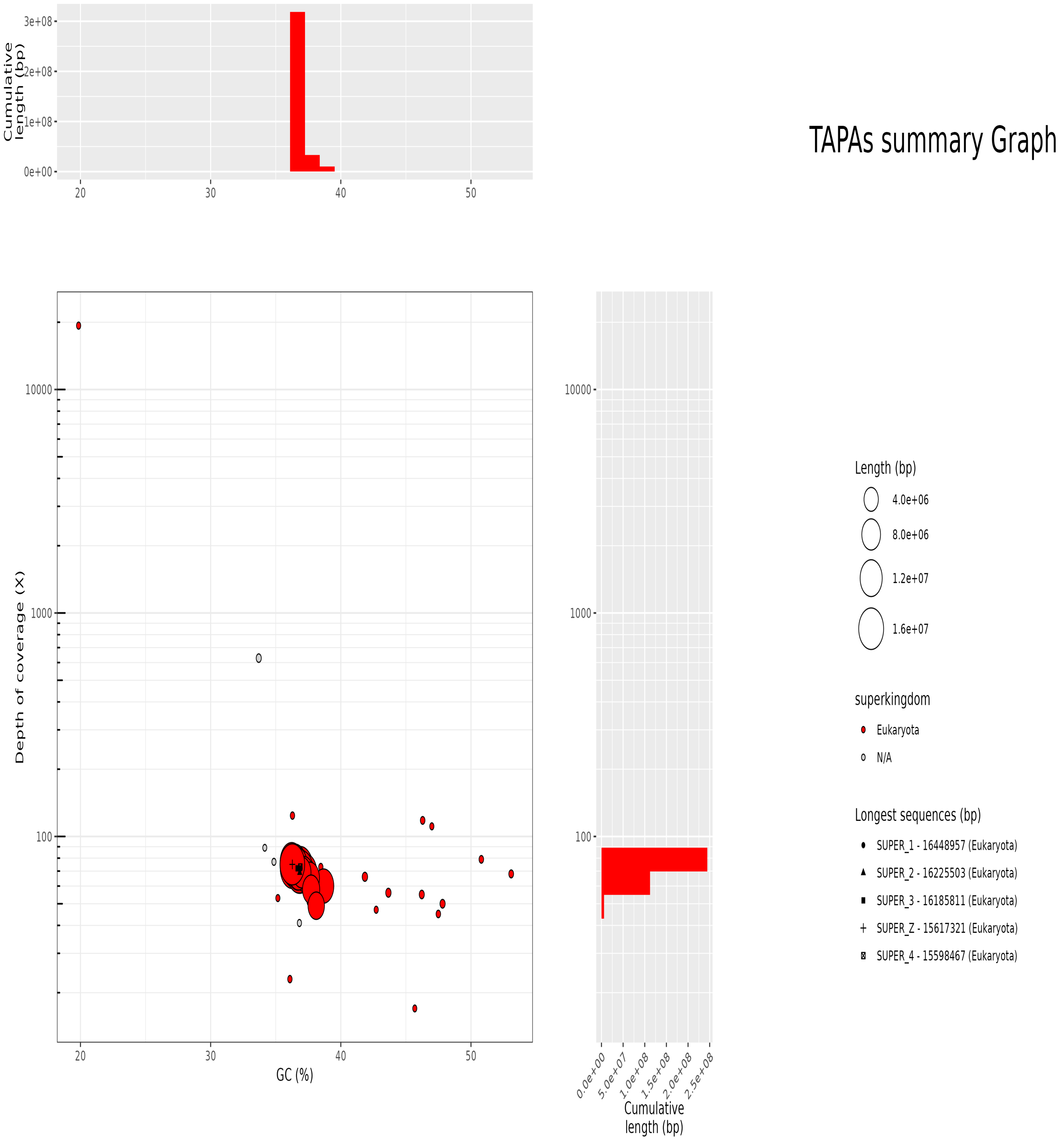


Distribution of k-mer counts per copy numbers found in asm



Distribution of k-mer counts coloured by their presence in reads/assemblies

Post-curation contamination screening



collapsed. Bubble plot circles are scaled by sequence length, positioned by coverage and GC proportion, and coloured by taxonomy. Histograms show total assembly length distribution on each axis.

Data profile

Data	Long reads	Arima
Coverage	86	354

Assembly pipeline

```
 - Hifiasm
     |_ ver: 0.19.5-r593
     |_ key param: NA
 - purge_dups
     |_ ver: 1.2.5
     |_ key param: NA
 - YaHS
     |_ ver: 1.2
     |_ key param: NA
```

Curation pipeline

```
 - PretextMap
     |_ ver: 0.1.9
     |_ key param: NA
 - PretextView
     |_ ver: 0.2.5
     |_ key param: NA
```

Submitter: Adama Ndar
Affiliation: Genoscope

Date and time: 2025-09-19 21:12:09 CEST