

ERGA Assembly Report

v24.10.15

Tags: ERGA-BGE

TxID	268458
ToLID	icMylQuad2
Species	<i>Mylabris quadripunctata</i>
Class	Insecta
Order	Coleoptera

Genome Traits	Expected	Observed
Haploid size (bp)	264,679,698	190,378,444
Haploid Number	11 (source: ancestor)	10
Ploidy	2 (source: ancestor)	2
Sample Sex	Unknown	Unknown

EBP metrics summary and curation notes

Obtained EBP quality metric for collapsed: 7.7.Q40

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

- Observed Haploid size (bp) has >20% difference with Expected
- Observed Haploid Number is different from Expected
- Kmer completeness value is less than 90 for collapsed

Curator notes

- Interventions/Gb: 121
- Contamination notes: ""
- Other observations: "The assembly of *Mylabris quadripunctata* (icMylQuad2) is based on 43,88X PacBio data and 242,75X 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 Nextdenovo, removal of contaminant sequences using Context, removal of haplotypic duplications using purge_dups, and Hi-C-based scaffolding with YaHS. No contigs were found to be contaminants (bacterial, archaeal, or viral). Additionally, 174 regions totaling 16.062 Mb (with the largest being 0.338 Mb) were identified as haplotypic duplications and removed. The mitochondrial genome was assembled using OATK. Finally, the primary assembly was analyzed and manually improved using Pretext. Chromosome-scale scaffolds confirmed by Hi-C data were named in order of size. Using comparative genomics analysis with *Mylabris sibirica*, *Hycleus marcipoli*, *Tenebrio molitor*, and *Harmonia axyridis*; we have been able to identified chromosome X and tag it accordingly. "

Quality metrics table

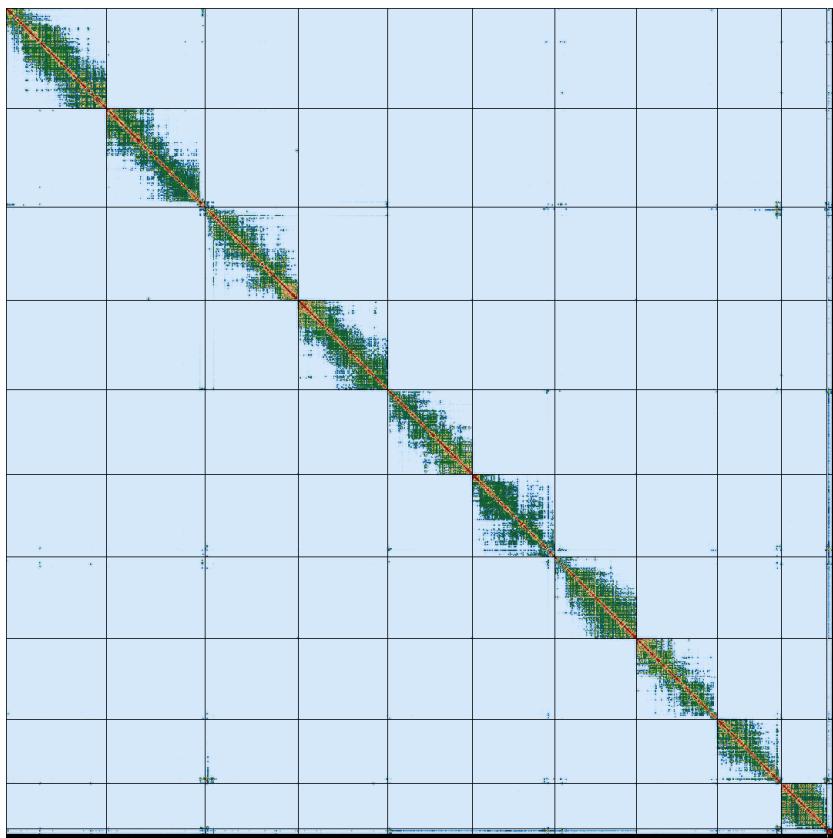
Metrics	Pre-curation collapsed	Curated collapsed
Total bp	190,361,040	190,378,444
GC %	32	32
Gaps/Gbp	378.23	351.93
Total gap bp	8,000	8,400
Scaffolds	32	34
Scaffold N50	19,246,236	19,246,236
Scaffold L50	5	5
Scaffold L90	10	9
Contigs	95	101
Contig N50	10,358,000	10,173,611
Contig L50	7	7
Contig L90	19	20
QV	40.7318	40.7322
Kmer compl.	75.058	75.058
BUSCO sing.	96.7%	98.4%
BUSCO dupl.	0.2%	0.4%
BUSCO frag.	0.8%	0.1%
BUSCO miss.	2.3%	1.1%

Warning! BUSCO versions or lineage datasets are not the same across results:

BUSCO: 5.8.2 (euk_genome_met, metaeuk) / Lineage: polyphaga_odb12 (genomes:60, BUSCOs:4010)

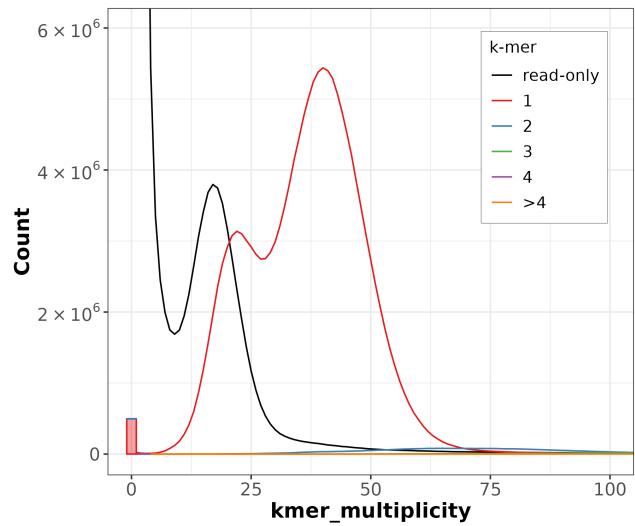
BUSCO: 6.0.0 (euk_genome_min, miniprot) / Lineage: polyphaga_odb12 (genomes:60, BUSCOs:4010)

HiC contact map of curated assembly

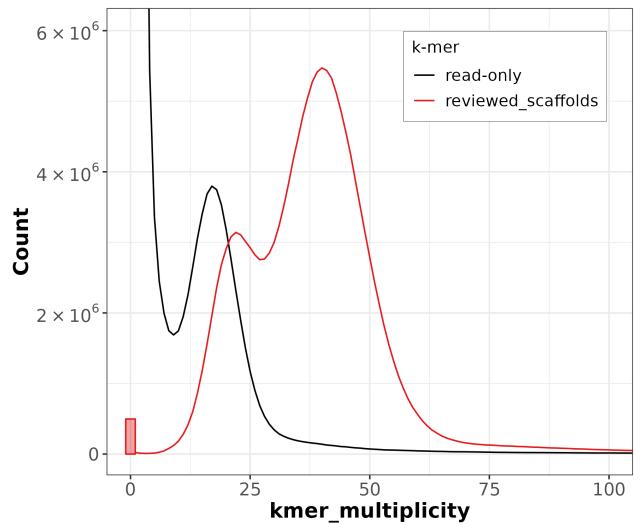


collapsed [\[LINK\]](#)

K-mer spectra of curated assembly

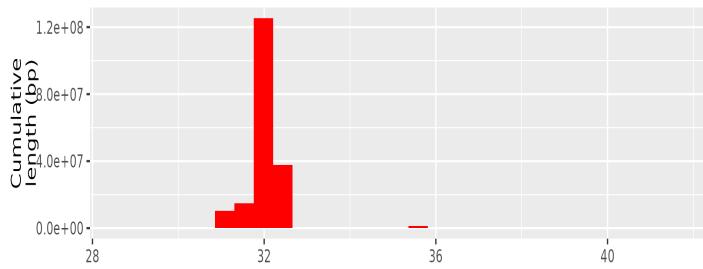


Distribution of k-mer counts per copy numbers found in asm

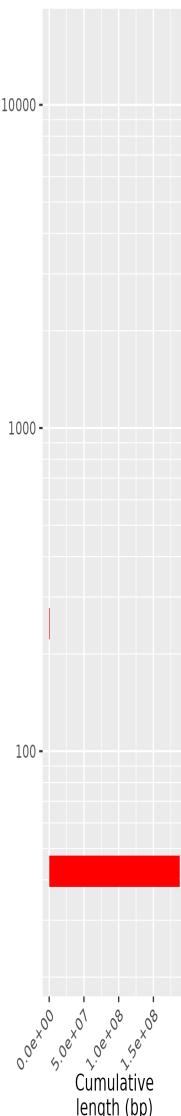
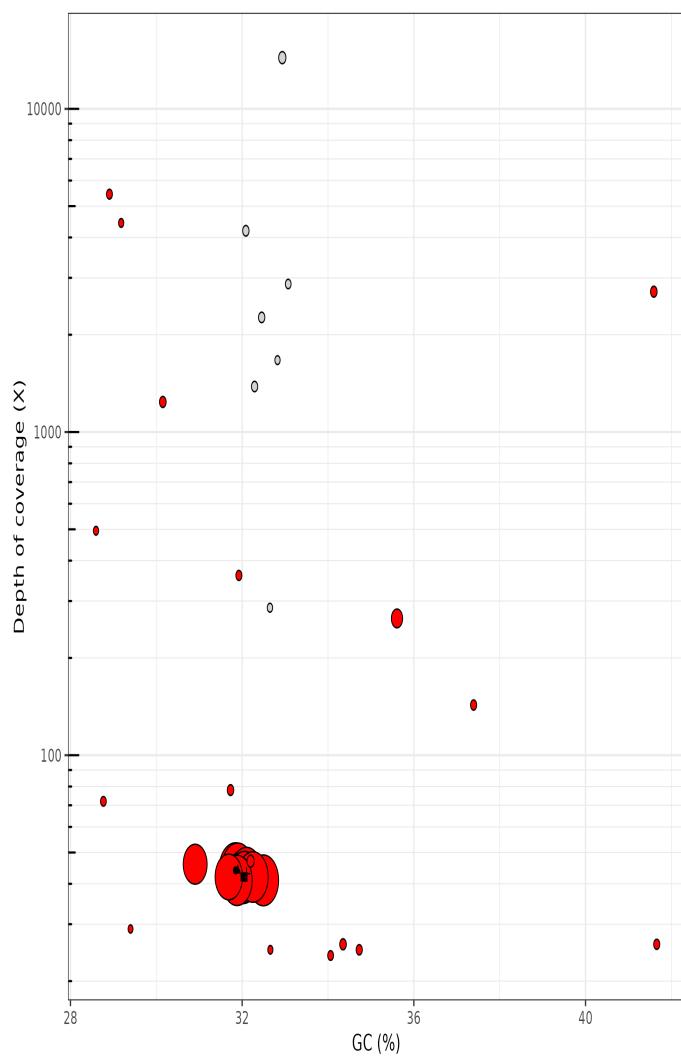


Distribution of k-mer counts coloured by their presence in reads/assemblies

Post-curation contamination screening



TAPAs summary Graph



Longest sequences (bp)

- icMyQuad2_1 - 23016639 (Eukaryota)
- ▲ icMyQuad2_2 - 22630603 (Eukaryota)
- icMyQuad2_3 - 21170703 (Eukaryota)
- + icMyQuad2_4 - 20608144 (Eukaryota)
- ▣ icMyQuad2_5 - 19246236 (Eukaryota)

superkingdom

- Eukaryota
- N/A

Length (bp)

- 5.0e+06
- 1.0e+07
- 1.5e+07
- 2.0e+07

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	43	242

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: Phuong Doan
Affiliation: Genoscope

Date and time: 2025-10-25 13:49:33 CEST