

# ERGA Assembly Report

v24.10.15

Tags: ERGA-BGE

TxID	575809
ToLID	<b>icDaiQuad1</b>
Species	<i>Dailognatha quadricollis</i>
Class	Insecta
Order	Coleoptera

Genome Traits	Expected	Observed
Haploid size (bp)	415,868,280	844,488,098
Haploid Number	10 (source: ancestor)	22
Ploidy	2 (source: ancestor)	2
Sample Sex	Unknown	Unknown

## EBP metrics summary and curation notes

Obtained EBP quality metric for collapsed: 6.7.Q66

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
- . BUSCO single copy value is less than 90% for collapsed
- . BUSCO duplicated value is more than 5% for collapsed
- . Not 90% of assembly in chromosomes for collapsed

### Curator notes

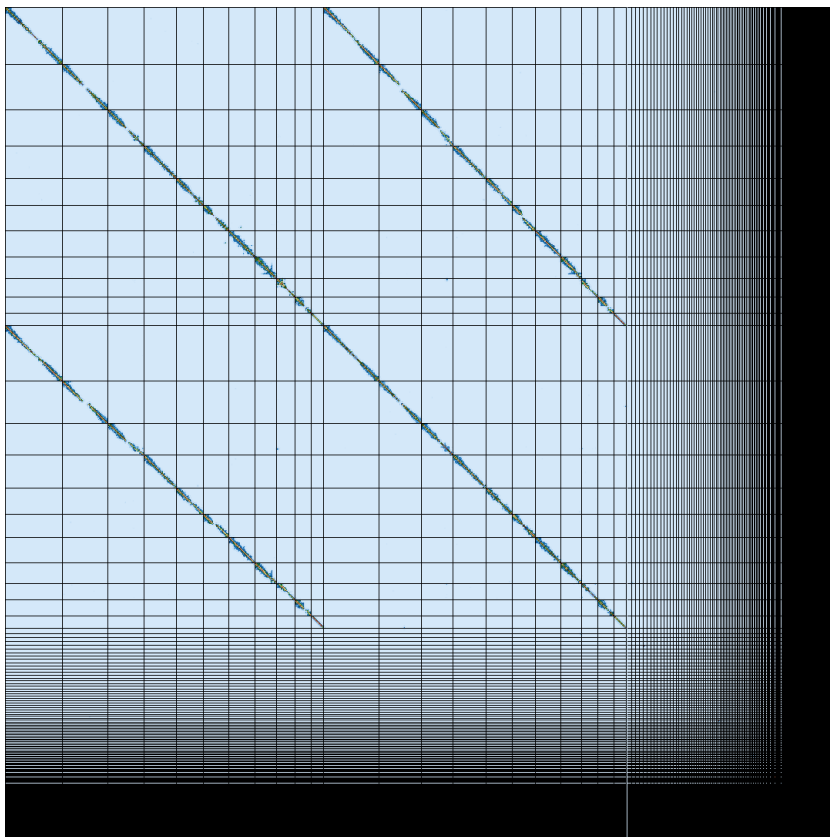
- . Interventions/Gb: 156
- . Contamination notes: ""
- . Other observations: "The assembly of *Dailognatha quadricollis* (icDaiQuad1) is based on 42X PacBio data and 177X Arima Hi-C data generated as part of the European Reference Genome Atlas (ERGA, <https://www.erga-biodiversity.eu/>) via the Biodiversity Genomics Europe project (BGE, <https://biodiversitygenomics.eu/>). The assembly process included the following steps: initial phased assembly with PacBio and Hi-C data using hifiasm, removal of contaminant sequences using Context, and Hi-C-based scaffolding with YaHS. In total, 0 contig were identified as contaminants (bacterial, archaeal, or viral). The mitochondrial genome was assembled using OATK. Finally, the assembly was analyzed and manually improved using Pretext. Chromosome-scale scaffolds confirmed by Hi-C data were named in order of size. The larger chromosomes were tagged as HAP1. X chromosome was identified after doing an alignment the reference genome of *Tenebrio molitor* (cTenMol1.1). "

# Quality metrics table

Metrics	Pre-curation collapsed	Curated collapsed
Total bp	844,662,313	844,488,098
GC %	33.67	33.67
Gaps/Gbp	58.01	118.41
Total gap bp	4,900	19,300
Scaffolds	441	387
Scaffold N50	31,477,100	25,988,060
Scaffold L50	10	12
Scaffold L90	106	79
Contigs	490	487
Contig N50	7,165,168	7,165,168
Contig L50	30	30
Contig L90	155	157
QV	50.5218	66.4102
Kmer compl.	98.992	98.444
BUSCO sing.	3.5%	1.6%
BUSCO dupl.	96.3%	98.3%
BUSCO frag.	0.1%	0.1%
BUSCO miss.	0.1%	0.0%

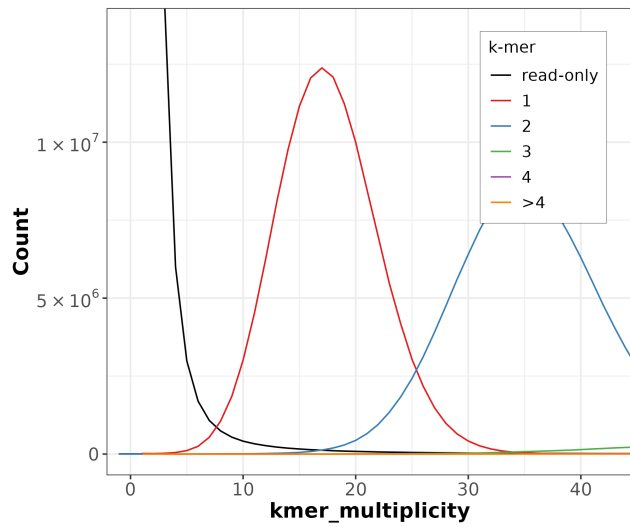
BUSCO: 5.4.3 (euk\_genome\_met, metaeuk) / Lineage: endopterygota\_odb10 (genomes:56, BUSCOs:2124)

# HiC contact map of curated assembly

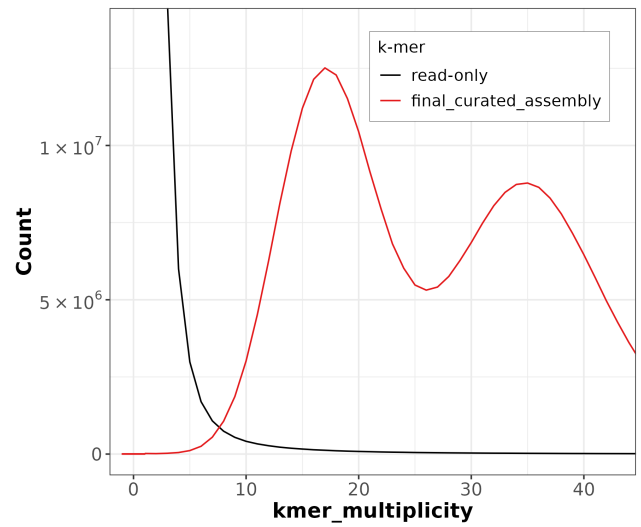


collapsed [\[LINK\]](#)

# K-mer spectra of curated assembly

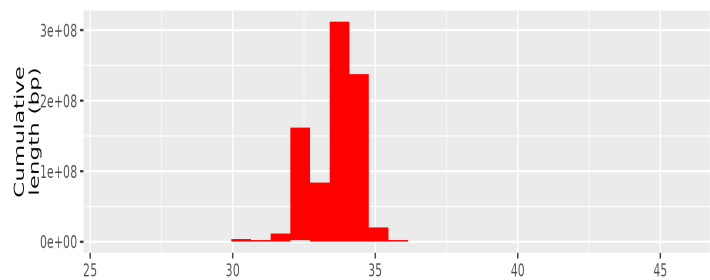


Distribution of k-mer counts per copy numbers found in asm

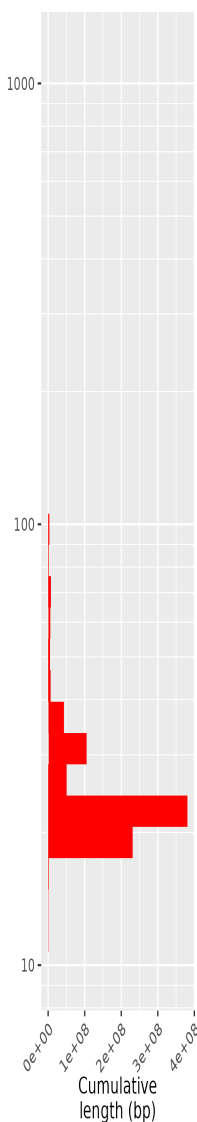
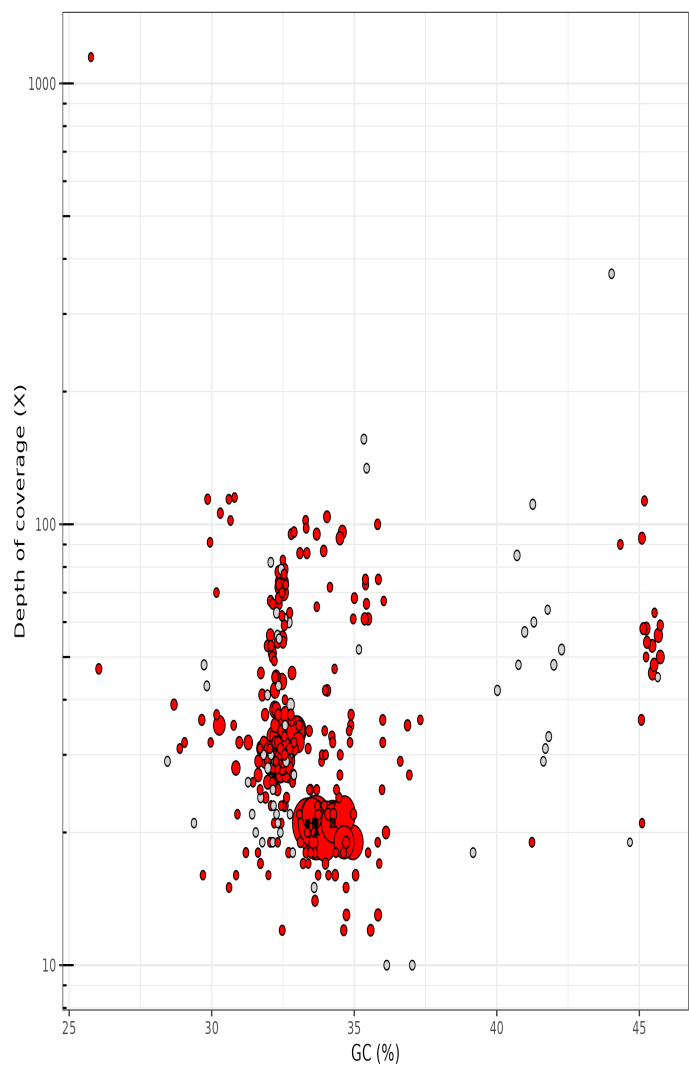


Distribution of k-mer counts coloured by their presence in reads/assemblies

# Post-curation contamination screening



TAPAs summary Graph



Longest sequences (bp)

- Hap1\_SUPER\_1 - 58542736 (Eukaryota)
- ▲ Hap2\_SUPER\_1 - 56523129 (Eukaryota)
- Hap1\_SUPER\_2 - 46214615 (Eukaryota)
- + Hap2\_SUPER\_2 - 42955530 (Eukaryota)
- Hap1\_SUPER\_3 - 36286850 (Eukaryota)

Length (bp)

- 1e+07
- 2e+07
- 3e+07
- 4e+07
- 5e+07

superkingdom

- Eukaryota
- N/A

**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	PACBIO Hifi	Arima
Coverage	42	177

# 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: Emilie Teodori

Affiliation: Genoscope

Date and time: 2025-02-11 09:09:15 CET