

ERGA Assembly Report

v24.10.15

Tags: ERGA-BGE

TxID	575809
ToLID	icDaiQuad1
Species	Dailognatha quadricollis
Class	Insecta
Order	Coleoptera

Genome Traits	Expected	Observed
Haploid size (bp)	415,868,280	524,088,546
Haploid Number	10 (source: ancestor)	11
Ploidy	2 (source: ancestor)	2
Sample Sex	Unknown	XX

EBP metrics summary and curation notes

Obtained EBP quality metric for collapsed: 6.7.Q65

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
- . Observed sex is different from Sample sex
- . Kmer completeness value is less than 90 for collapsed
- . Not 90% of assembly in chromosomes for collapsed

Curator notes

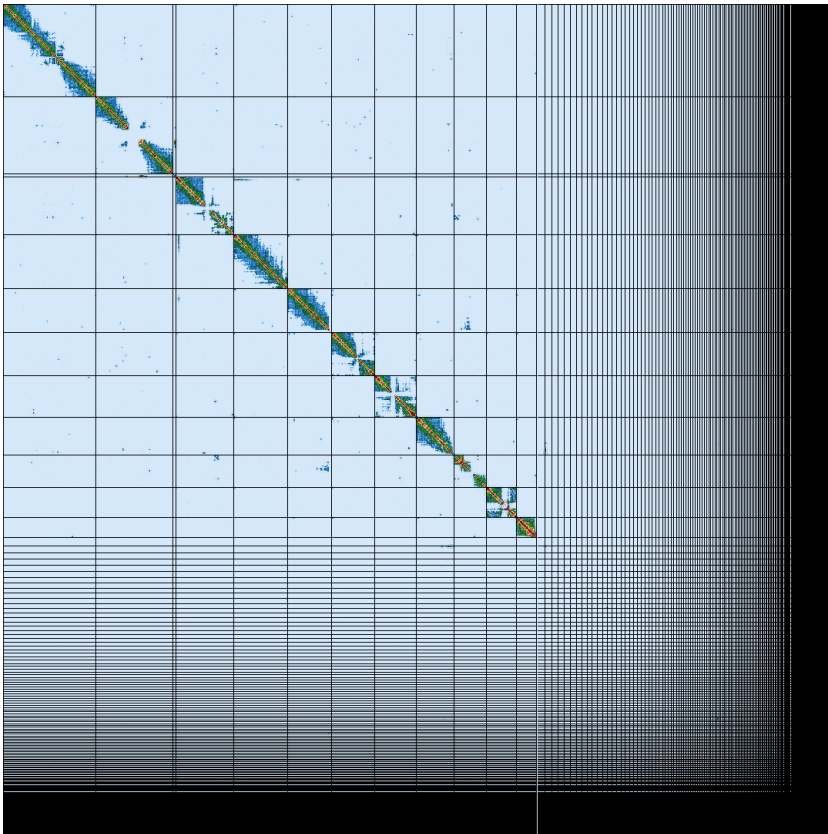
- . Interventions/Gb: 782
- . Contamination notes: ""
- . Other observations: "The assembly of Dailognatha quadricollis (icDaiQuad1) is based on 42X PacBio data and 177X Arima Hi-C data generated as part of the European Reference Genome Atlas (ERGA, <https://www.erga-biodiversity.eu/>) via the Biodiversity 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 (icTenMol1.1). "

Quality metrics table

Metrics	Pre-curation collapsed	Curated collapsed
Total bp	392,707,419	524,088,546
GC %	33.65	33.51
Gaps/Gbp	0	135.47
Total gap bp	0	13,800
Scaffolds	197	323
Scaffold N50	12,785,753	23,571,702
Scaffold L50	10	8
Scaffold L90	50	82
Contigs	197	394
Contig N50	12,785,753	4,327,000
Contig L50	10	27
Contig L90	50	136
QV	50.3522	65.1004
Kmer compl.	69.3685	75.709
BUSCO sing.	91.6%	98.7%
BUSCO dupl.	0.7%	1.1%
BUSCO frag.	0.1%	0.1%
BUSCO miss.	7.6%	0.1%

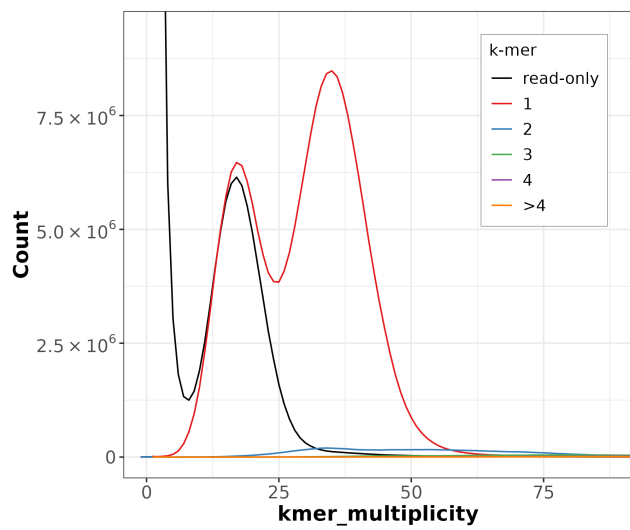
BUSCO: 5.4.3 (euk_genome_met, metaeuk) / Lineage: endopterygota_odb10 (genomes:56, BUSCOs:2124)

HiC contact map of curated assembly

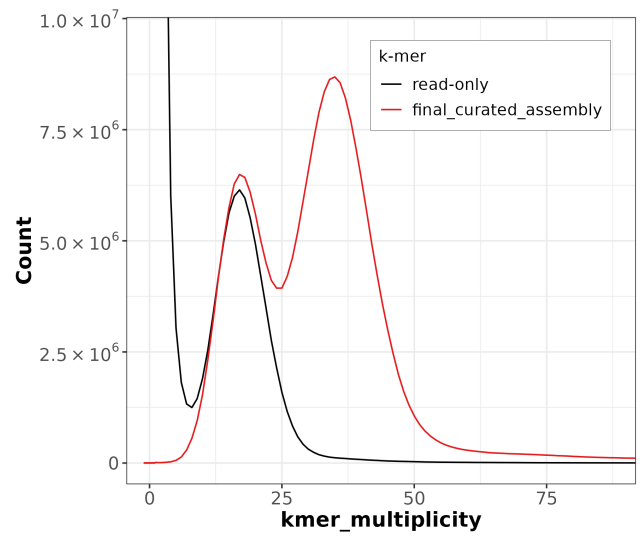


collapsed [\[LINK\]](#)

K-mer spectra of curated assembly

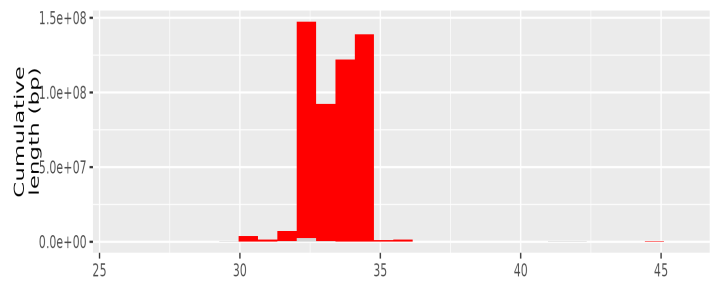


Distribution of k-mer counts per copy numbers found in asm

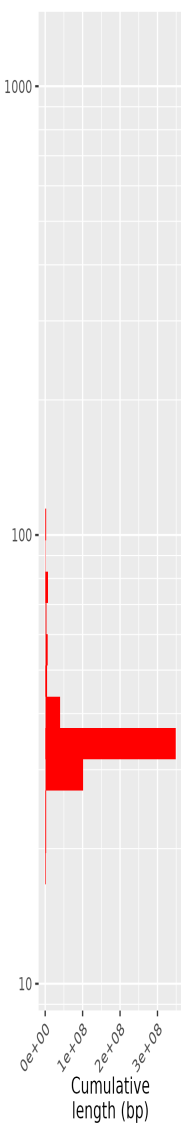
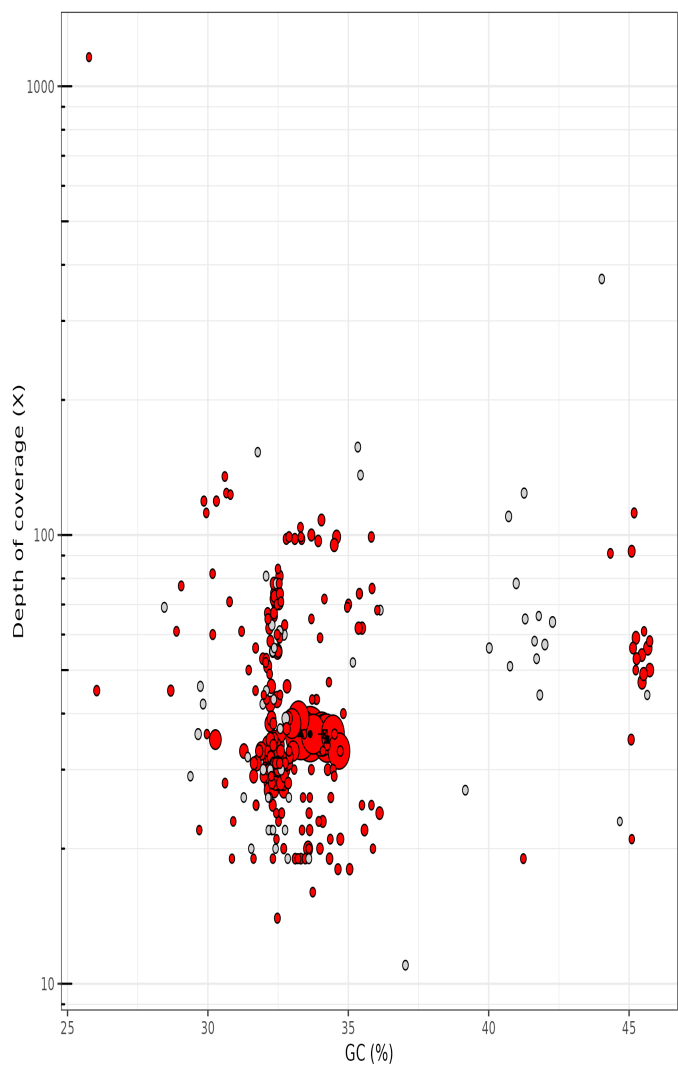


Distribution of k-mer counts coloured by their presence in reads/assemblies

Post-curation contamination screening



TAPAs summary Graph



- Length (bp)
- 1e+07
 - 2e+07
 - 3e+07
 - 4e+07
 - 5e+07
- superkingdom
- Eukaryota
 - N/A
- Longest sequences (bp)
- SUPER_1 - 58304284 (Eukaryota)
 - ▲ SUPER_2 - 48728015 (Eukaryota)
 - SUPER_3 - 36286850 (Eukaryota)
 - + SUPER_4 - 33787855 (Eukaryota)
 - SUPER_5 - 27677016 (Eukaryota)

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: 2025-02-28 08:38:59 CET