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	Unknown

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

. Kmer completeness value is less than 90 for collapsed
. Not 90% of assembly in chromosomes for collapsed

Curator notes

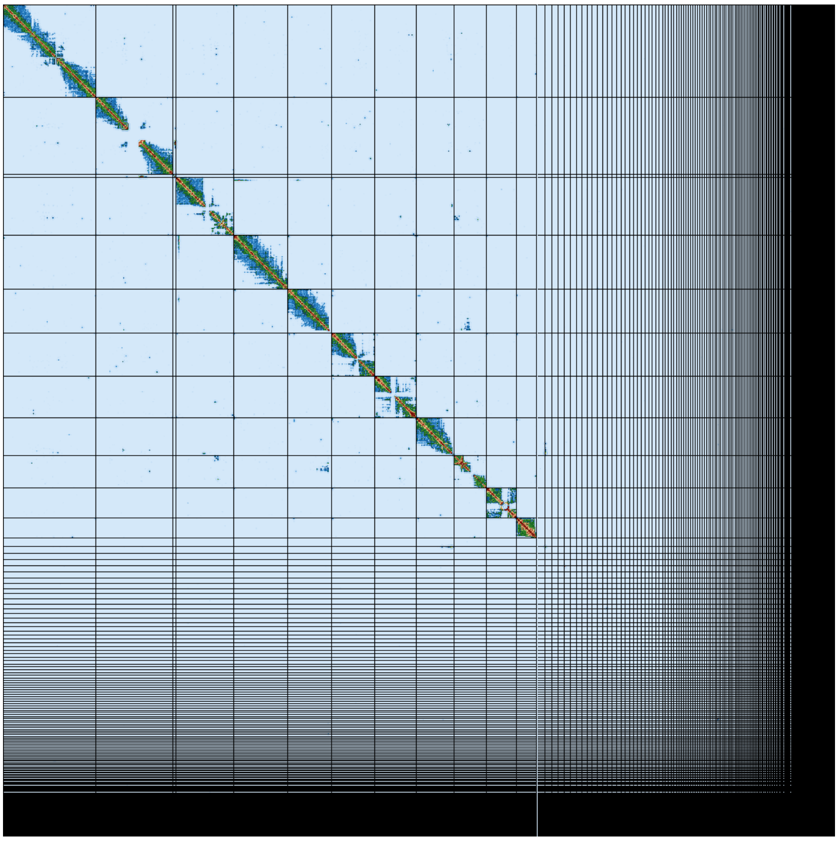
. Interventions/Gb: 782
. Contamination notes: ""
. Other observations: "The assembly of Dailognatha quadricollis (icDaiQuad1) is based on 42X PacBio data and 177X Arima Hi-C data generated as part of the European Reference Genome Atlas (ERGA, https://www.erga-biodiversity.eu/) via the Biodiversity Genomics 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. X chromosome was identified after doing an alignment the reference genome of Tenebrio molitor (icTenMoli1.1). "

Quality metrics table

Metrics	Pre-curation collapsed	Curated collapsed
Total bp	474,022,430	524,088,546
GC %	33.61	33.51
Gaps/Gbp	29.53	135.47
Total gap bp	1,400	13,800
Scaffolds	119	323
Scaffold N50	17,893,662	23,571,702
Scaffold L50	9	8
Scaffold L90	34	82
Contigs	133	394
Contig N50	12,558,000	4,327,000
Contig L50	12	27
Contig L90	47	136
QV	51.2232	65.1004
Kmer compl.	78.8857	75.709
BUSCO sing.	87.4%	96.6%
BUSCO dupl.	9.5%	0.7%
BUSCO frag.	1.8%	0.7%
BUSCO miss.	1.2%	2.1%

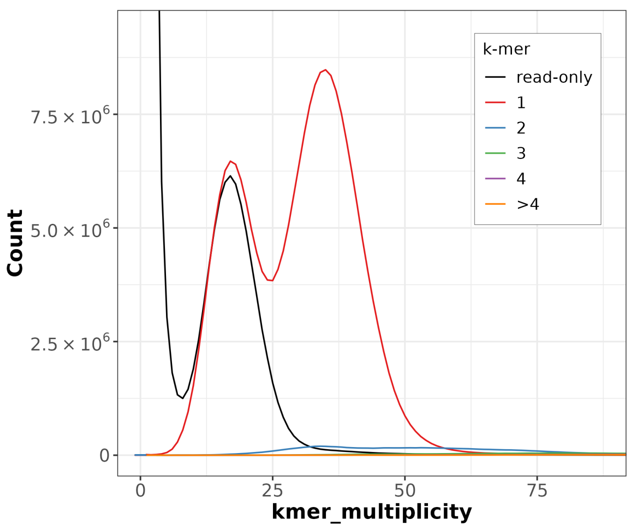
Warning! BUSCO versions or lineage datasets are not the same across results:
BUSCO: 5.8.2 (euk_genome_min, miniprot) / Lineage: polyphaga_odb12 (genomes:60, BUSCOs:4010)
BUSCO: 5.8.2 (euk_genome_met, metaeuk) / Lineage: polyphaga_odb12 (genomes:60, BUSCOs:4010)

HiC contact map of curated assembly

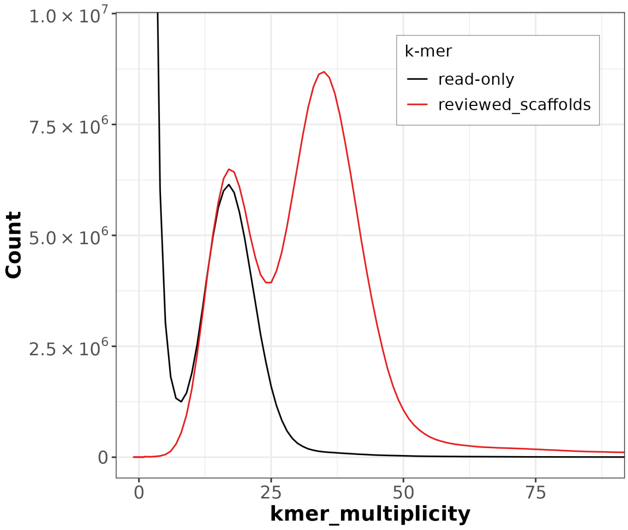


collapsed [LINK]

K-mer spectra of curated assembly

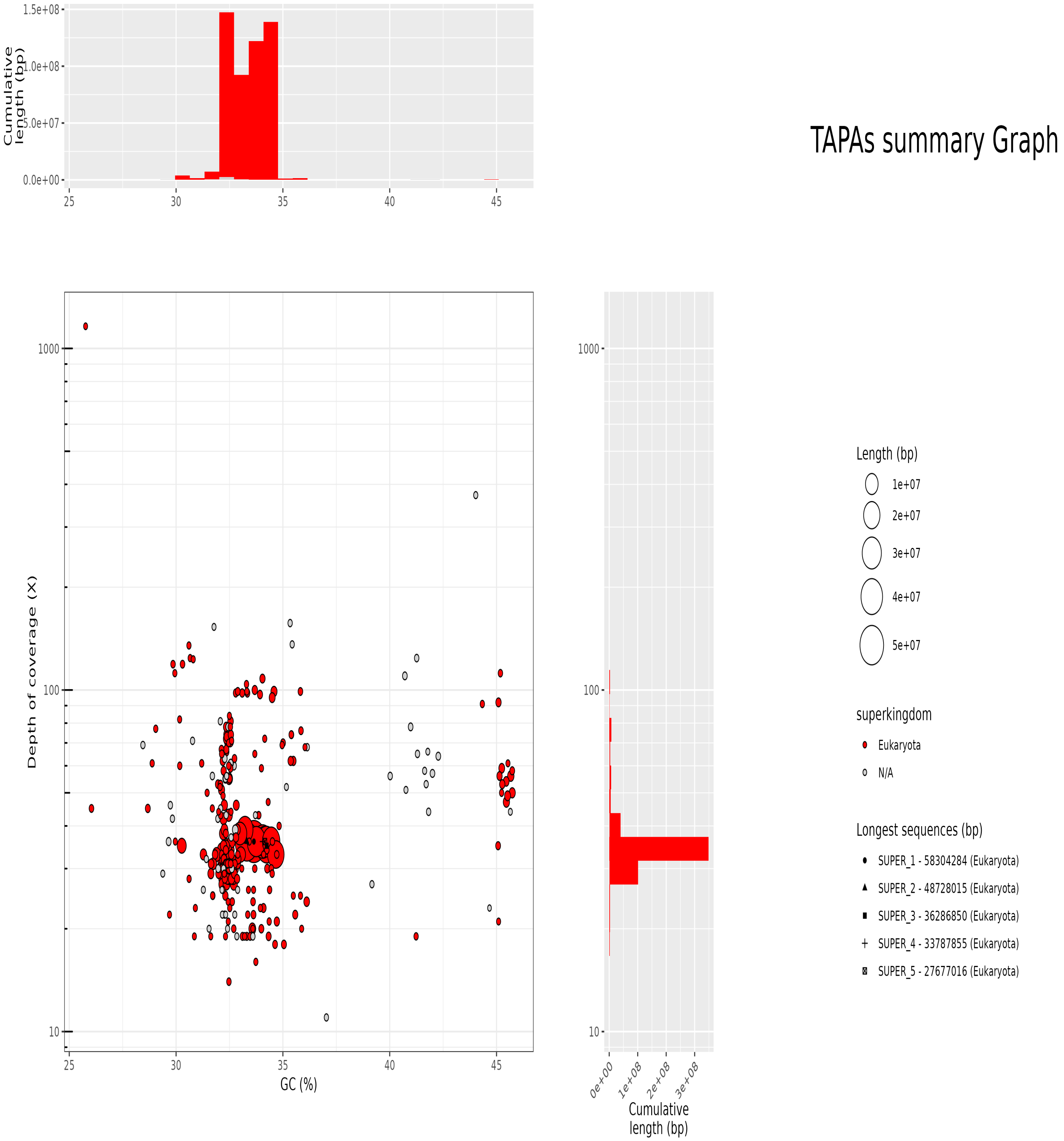


Distribution of k-mer counts per copy numbers found in asm



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

Post-curation contamination screening



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	41	206

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: 2025-04-10 14:18:14 CEST