ERGA Assembly Report

v24.10.15

Tags: ERGA-BGE

TxID	3230685
ToLID	**icPseApti1**
Species	Pseudanchomenus aptinoides
Class	Insecta
Order	Coleoptera

Genome Traits	Expected	Observed
Haploid size (bp)	828,164,504	1,081,189,266
Haploid Number	19 (source: ancestor)	21
Ploidy	2 (source: ancestor)	2
Sample Sex	Unknown	Unknown

EBP metrics summary and curation notes

Obtained EBP quality metric for collapsed: 6.7.Q62

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: 116
. Contamination notes: ""
. Other observations: "The assembly of Pseudanchomenus aptinoides (icPseApti1) is based on 33X PacBio data and 181X 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, 35 contigs were identified as contaminants (bacterial, archaeal, or viral), totaling 9.446 Mb (with the largest being 4.384 Mb). Additionally, 1145 regions totaling 121.128 Mb (with the largest being 0.567 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. During manual curation, 11 haplotypic regions and and XX contaminant sequences were removed, totaling 8.298Mb and XXMb respectively, (with the largest being 1.993Mb and XX Mb).Chromosome-scale scaffolds

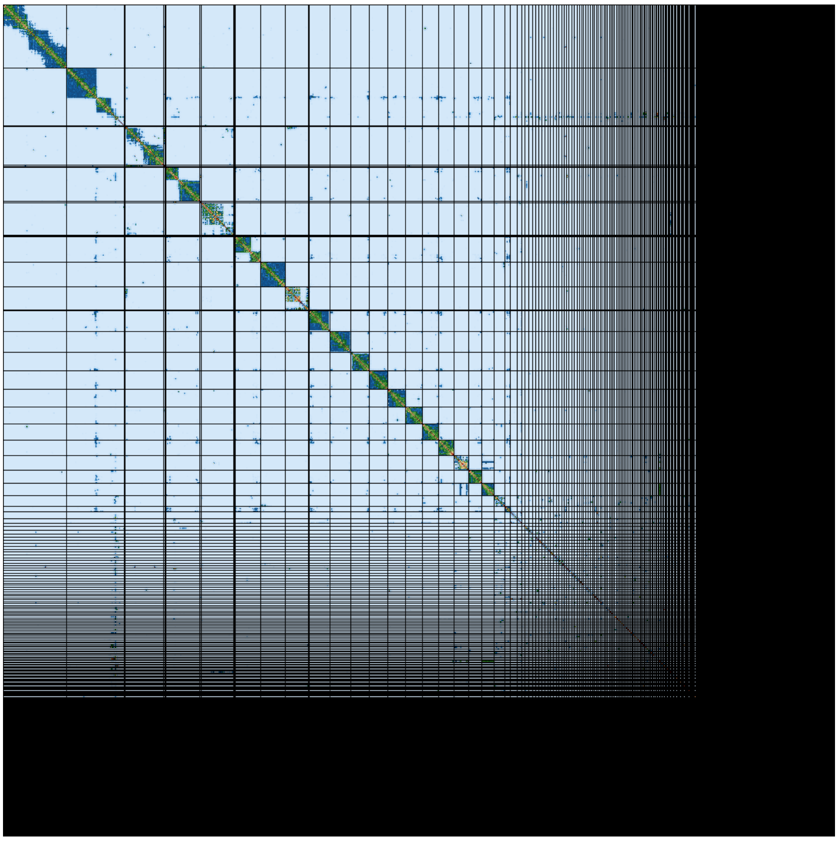
confirmed by Hi-C data were named in order of size. "

Quality metrics table

Metrics	Pre-curation collapsed	Curated collapsed
Total bp	1,089,488,111	1,081,189,266
GC %	31.53	31.53
Gaps/Gbp	405.7	359.79
Total gap bp	45,200	45,600
Scaffolds	1,079	686
Scaffold N50	6,363,035	20,986,798
Scaffold L50	27	15
Scaffold L90	375	222
Contigs	1,504	1,075
Contig N50	2,362,129	3,270,678
Contig L50	69	55
Contig L90	629	468
QV	62.3893	62.3906
Kmer compl.	85.4298	85.2809
BUSCO sing.	90.5%	93.0%
BUSCO dupl.	5.1%	4.9%
BUSCO frag.	1.5%	0.5%
BUSCO miss.	2.8%	1.7%

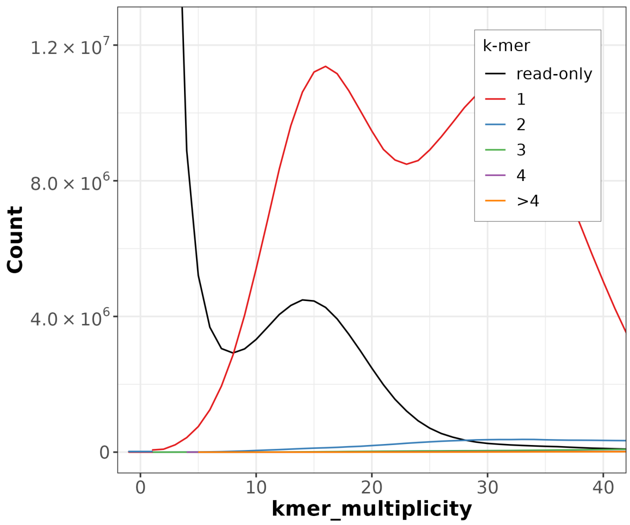
Warning! BUSCO versions or lineage datasets are not the same across results:
BUSCO: 5.8.2 (euk_genome_met, metaeuk) / Lineage: coleoptera_odb12 (genomes:64, BUSCOs:3729)
BUSCO: 6.0.0 (euk_genome_min, miniprot) / Lineage: coleoptera_odb12 (genomes:64, BUSCOs:3729)

HiC contact map of curated assembly

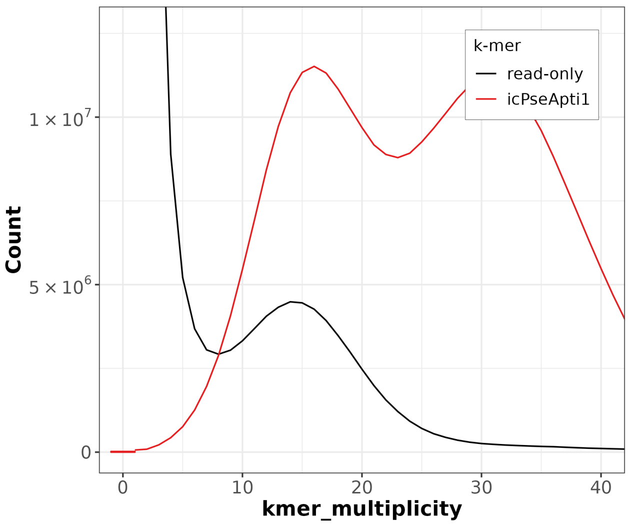


collapsed [LINK]

K-mer spectra of curated assembly

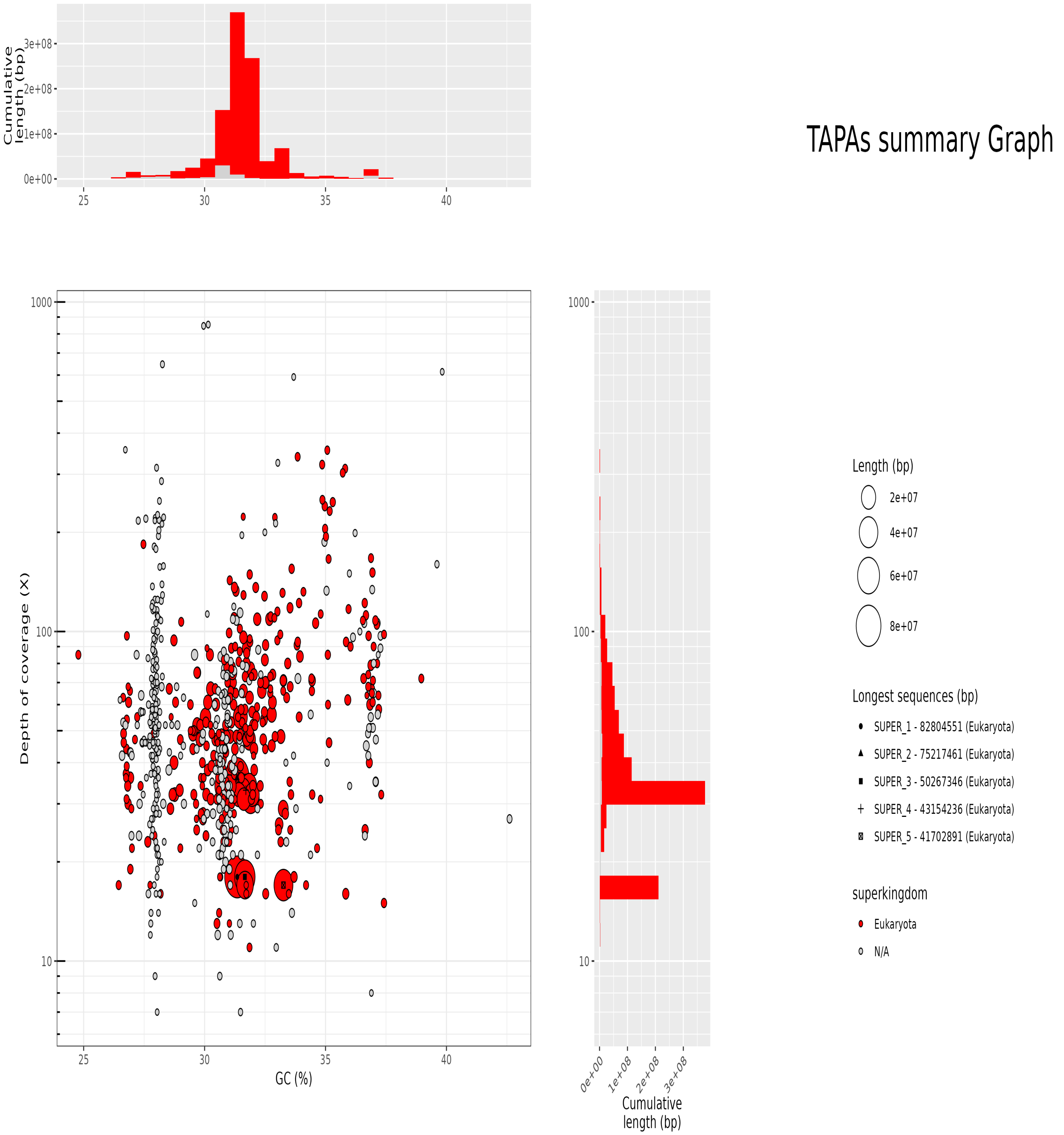


Distribution of k-mer counts per copy numbers found in asm



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	33	181

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: Sophie Mangenot
Affiliation: Genoscope

Date and time: 2025-12-11 09:23:31 CET