ERGA Assembly Report

v24.10.15

Tags: ATLASea[INVALID TAG]

TxID	75124
ToLID	**xgCypTigr1**
Species	Cypraea tigris
Class	Gastropoda
Order	Littorinimorpha

Genome Traits	Expected	Observed
Haploid size (bp)	1,438,783,631	1,545,381,438
Haploid Number	17 (source: ancestor)	37
Ploidy	2 (source: ancestor)	2
Sample Sex	Unknown	Unknown

EBP metrics summary and curation notes

Obtained EBP quality metric for collapsed: 6.7.Q44

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

. Observed Haploid Number is different from Expected

. Kmer completeness value is less than 90 for collapsed
. BUSCO single copy value is less than 90% for collapsed
. BUSCO duplicated value is more than 5% for collapsed

Curator notes

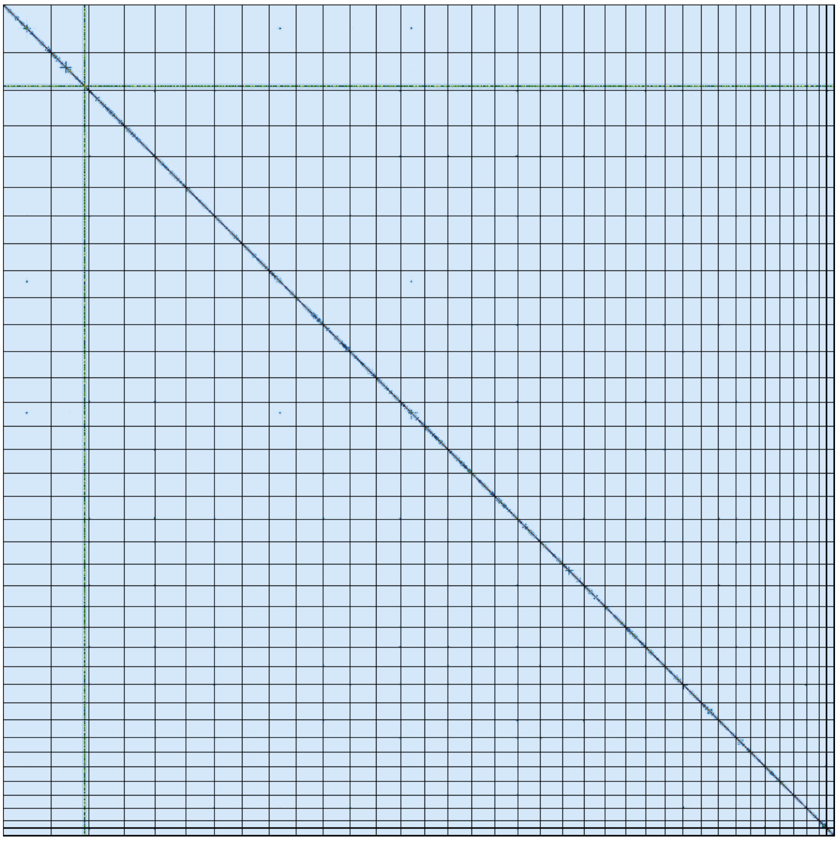
. Interventions/Gb: 58
. Contamination notes: ""
. Other observations: "The assembly of Cypraea tigris (xgCypTigr1) is based on 43X ONT data and 178X Arima Hi-C data generated as part of the ATLASea programme (https://www.atlasea.fr). The assembly process included the following steps: initial ONT assembly 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, 20 contigs were identified as contaminants (bacterial, archaeal, or viral), totaling 2.725 Mb (with the largest being 0.728 Mb). Additionally, 393 regions totaling 52.998 Mb (with the largest being 1.749 Mb) were identified as haplotypic duplications and removed. Finally, the primary assembly was analyzed and manually improved using Pretext.The mitochondrial genome was assembled using ptGAUL. During manual curation, 22 haplotypic regions and 8 contaminant sequences were removed, totaling 8.679Mb and 1.088Mb, respectively (with the largest being 1.156Mb and 0.391Mb). 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	1,555,147,505	1,545,381,438
GC %	42.29	42.28
Gaps/Gbp	177.48	177.3
Total gap bp	27,600	30,700
Scaffolds	92	73
Scaffold N50	49,615,095	44,261,810
Scaffold L50	13	14
Scaffold L90	28	31
Contigs	368	347
Contig N50	8,957,251	8,957,251
Contig L50	49	49
Contig L90	166	164
QV	44.1395	44.141
Kmer compl.	81.2502	80.9435
BUSCO sing.	76.0%	76.1%
BUSCO dupl.	21.8%	21.6%
BUSCO frag.	0.9%	0.9%
BUSCO miss.	1.2%	1.4%

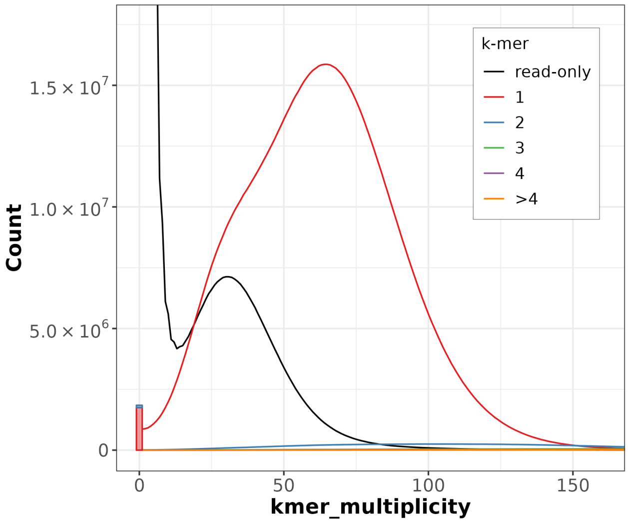
BUSCO: 6.0.0 (euk_genome_min, miniprot) / Lineage: mollusca_odb12 (genomes:36, BUSCOs:4421)

HiC contact map of curated assembly

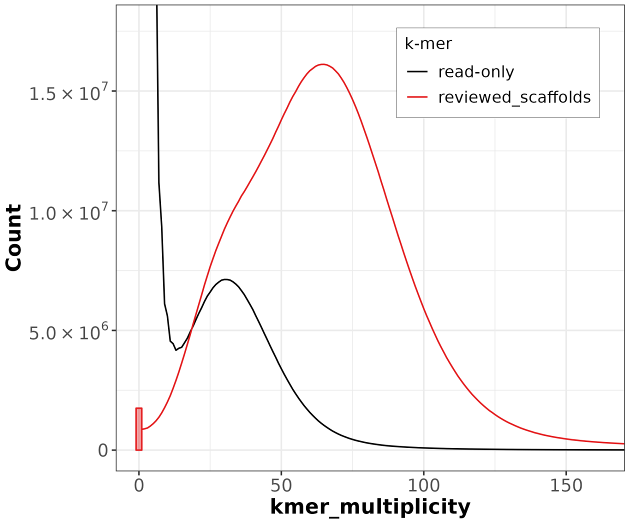


collapsed [LINK]

K-mer spectra of curated assembly

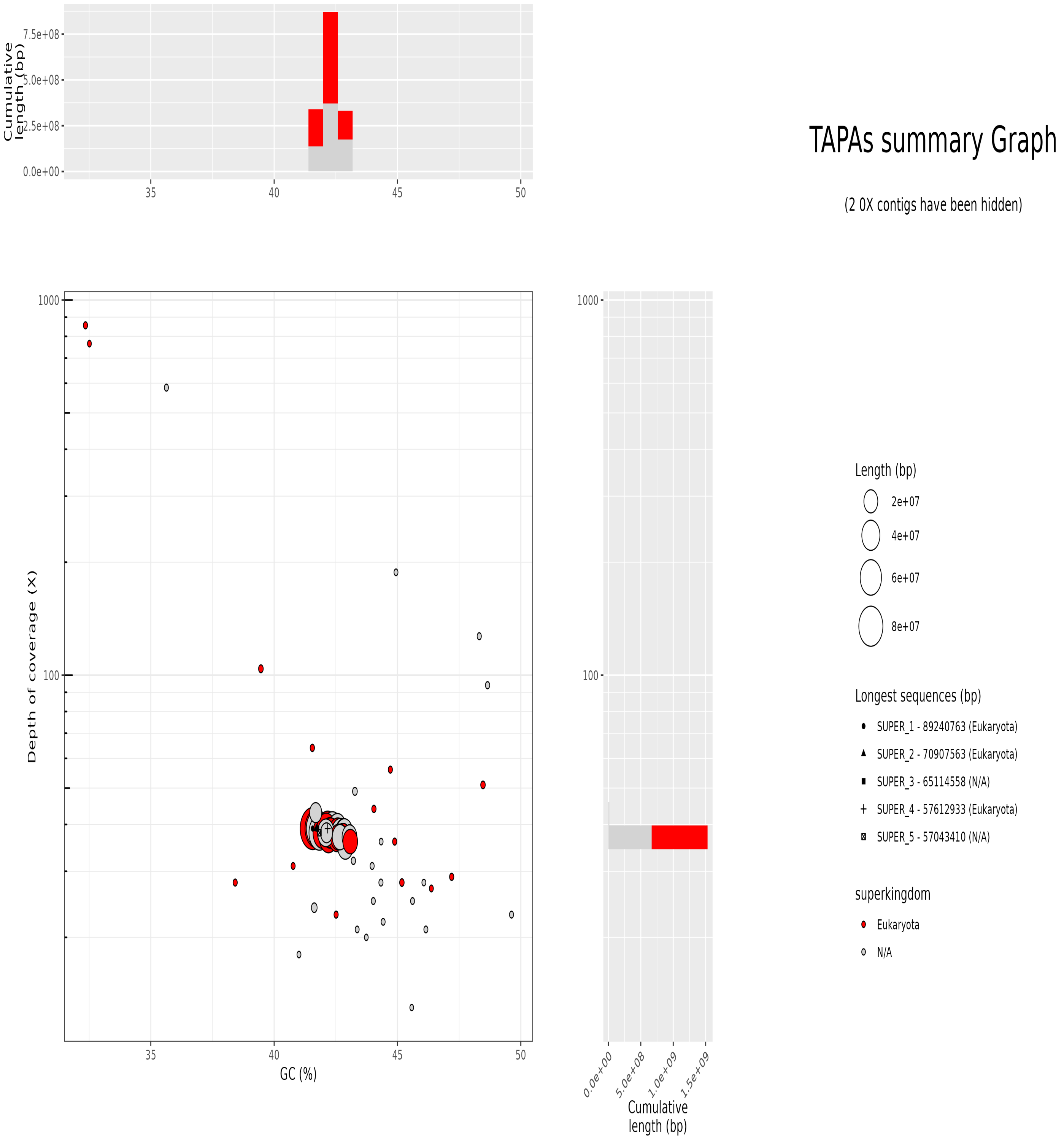


Distribution of k-mer counts per copy numbers found in asm



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	43	178

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 Layac
Affiliation: Genoscope

Date and time: 2025-12-13 19:57:10 CET