ERGA Assembly Report

v24.10.15

Tags: ATLASea[INVALID TAG]

TxID	164121
ToLID	**xgCitPica1**
Species	Cittarium pica
Class	Gastropoda
Order	Trochida

Genome Traits	Expected	Observed
Haploid size (bp)	675,842,208	687,752,580
Haploid Number	9 (source: ancestor)	18
Ploidy	2 (source: ancestor)	2
Sample Sex	Unknown	Unknown

EBP metrics summary and curation notes

Obtained EBP quality metric for collapsed: 7.7.Q51

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

Curator notes

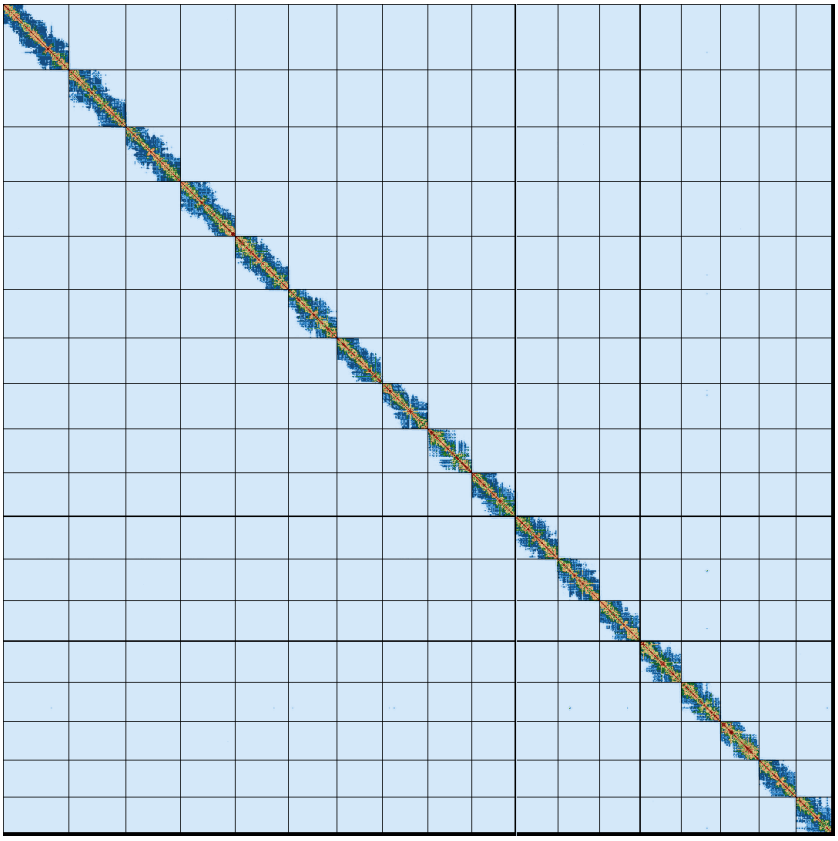
. Interventions/Gb: 4
. Contamination notes: ""
. Other observations: "The assembly of Cittarium pica (xgCitPica1) is based on 36X PacBio data and 68X Arima Hi-C data generated as part of the ATLASea programme (https://www.atlasea.fr). The assembly process included the following steps: initial PacBio 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, 152 contigs were identified as contaminants (bacterial, archaeal, or viral), totaling 7.508 Mb (with the largest being 0.176 Mb). Additionally, 26 regions totaling 2.609 Mb (with the largest being 0.721 Mb) were identified as haplotypic duplications and removed. Finally, the primary assembly was analyzed and manually improved using Pretext. During manual curation, 2 haplotypic regions and 2 contaminant sequences were removed, totaling 2.59Mb and 0.108Mb, respectively (with the largest being 2.355Mb and 0.069Mb). 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	690,457,140	687,752,580
GC %	36.7	36.7
Gaps/Gbp	15.93	15.99
Total gap bp	1,100	1,200
Scaffolds	45	42
Scaffold N50	37,531,372	37,478,293
Scaffold L50	8	8
Scaffold L90	16	16
Contigs	56	53
Contig N50	34,684,446	34,684,446
Contig L50	9	9
Contig L90	20	20
QV	51.6448	51.6494
Kmer compl.	83.0778	82.9396
BUSCO sing.	99.2%	99.2%
BUSCO dupl.	0.5%	0.4%
BUSCO frag.	0.1%	0.1%
BUSCO miss.	0.2%	0.2%

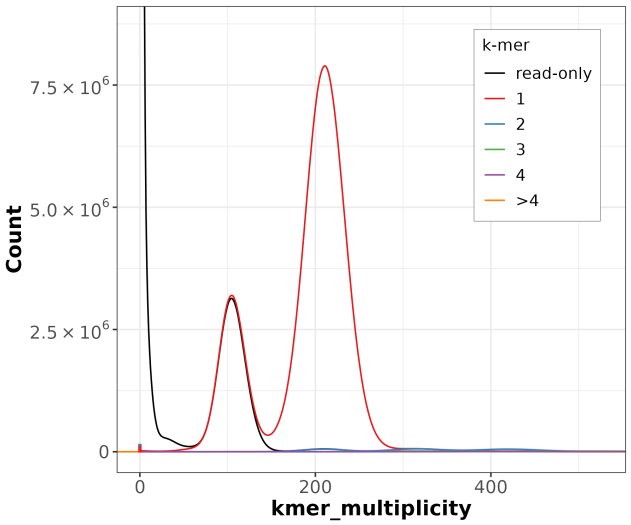
BUSCO: 6.0.0 (euk_genome_min, miniprot) / Lineage: mollusca_odb12 (genomes:36, BUSCOs:4421)

HiC contact map of curated assembly

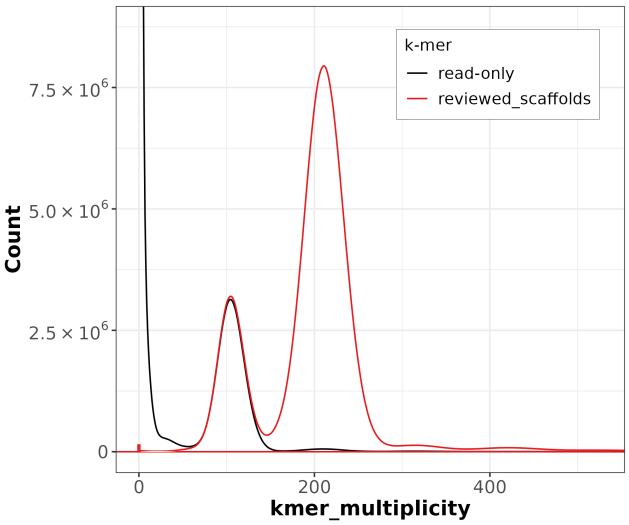


collapsed [LINK]

K-mer spectra of curated assembly

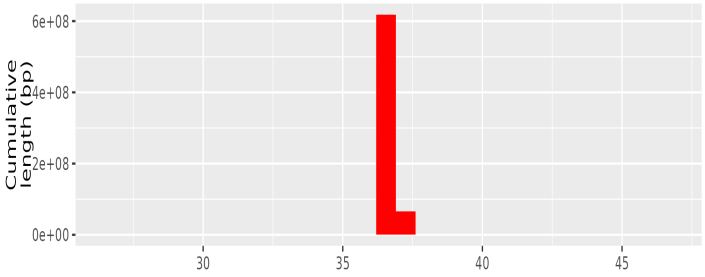


Distribution of k-mer counts per copy numbers found in asm

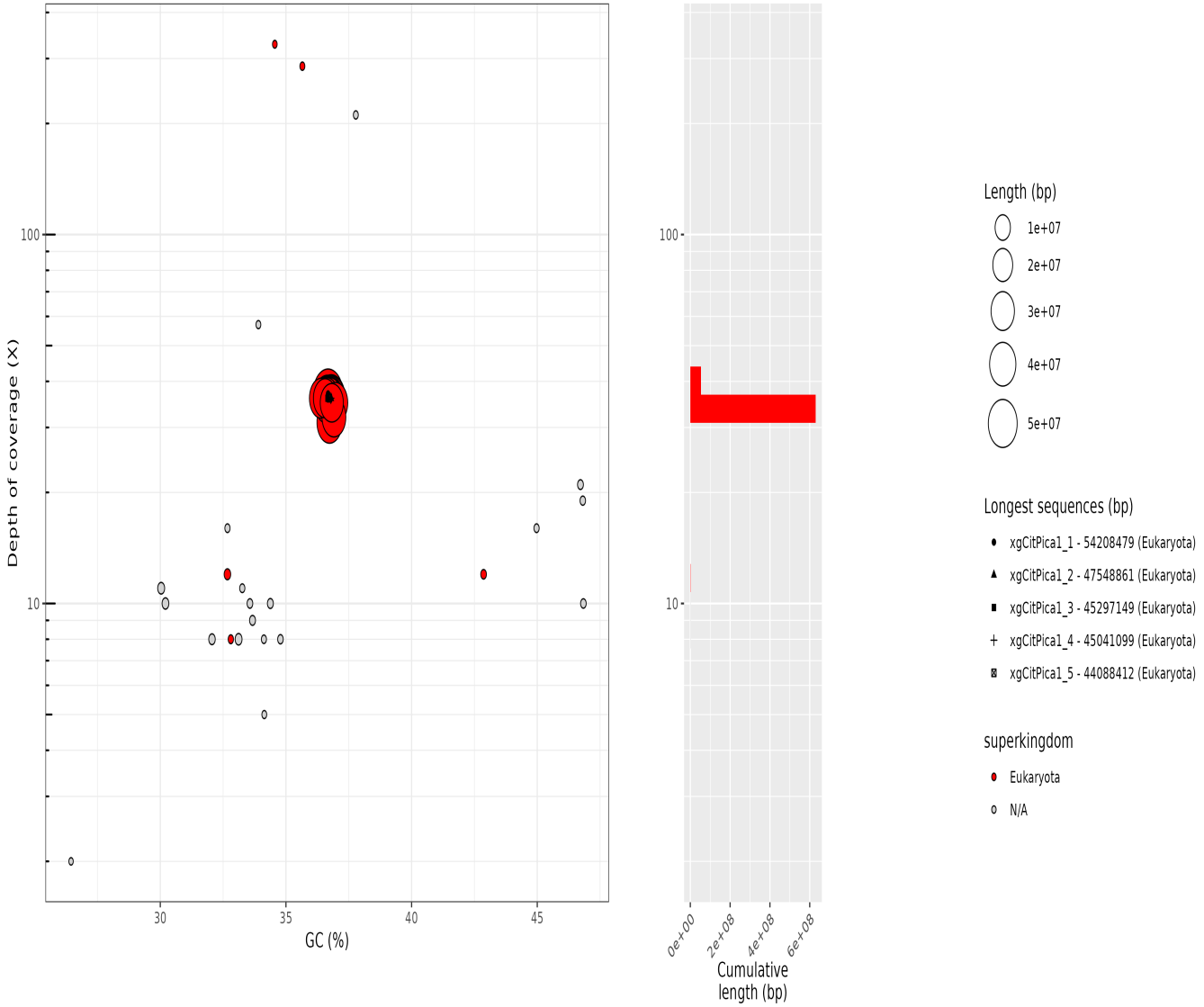


Distribution of k-mer counts coloured by their presence in reads/assemblies

Post-curation contamination screening



TAPAs summary Graph



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	36	68

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: 2026-04-27 15:53:00 CEST