ERGA Assembly Report

v24.10.15

Tags: ATLASea[INVALID TAG]

TxID	658864
ToLID	**xgCerLitt1**
Species	Cerithium litteratum
Class	Gastropoda
Order	NA

Genome Traits	Expected	Observed
Haploid size (bp)	636,086,152	695,434,183
Haploid Number	15 (source: ancestor)	18
Ploidy	2 (source: ancestor)	2
Sample Sex	Unknown	Unknown

EBP metrics summary and curation notes

Obtained EBP quality metric for collapsed: 6.7.Q43

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

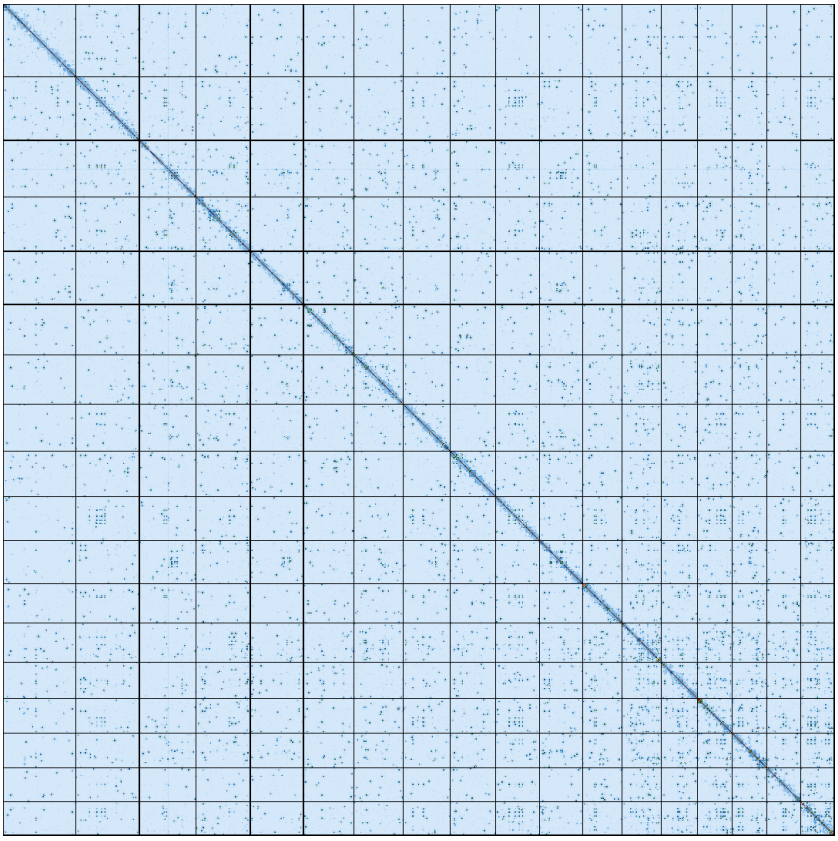
. Interventions/Gb: 57
. Contamination notes: ""
. Other observations: "The assembly of Cerithium litteratum (xgCerLitt1) is based on 34X ONT data and 184X Arima Hi-C data generated as part of the ATLASea programme (https://www.atlasea.fr). The assembly process included the following steps: initial ONT 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, 6 contigs were identified as contaminants (bacterial, archaeal, or viral), totaling 0.62 Mb (with the largest being 0.253 Mb). Additionally, 356 regions totaling 185.496 Mb (with the largest being 3.218 Mb) were identified as haplotypic duplications and removed. Mitochondrial genome was assembled using OATK. Finally, the primary assembly was analyzed and manually improved using Pretext. During manual curation, 13 haplotypic regions were removed, totaling 7.846Mb (with the largest being 1.551Mb). 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	703,330,184	695,434,183
GC %	45.03	45.03
Gaps/Gbp	365.4	372.43
Total gap bp	25,700	27,500
Scaffolds	40	37
Scaffold N50	39,982,835	39,443,376
Scaffold L50	8	8
Scaffold L90	16	16
Contigs	297	296
Contig N50	4,214,870	4,214,870
Contig L50	54	54
Contig L90	160	159
QV	43.3918	43.4365
Kmer compl.	61.6332	61.2099
BUSCO sing.	96.5%	94.0%
BUSCO dupl.	1.5%	0.7%
BUSCO frag.	0.9%	3.9%
BUSCO miss.	1.1%	1.5%

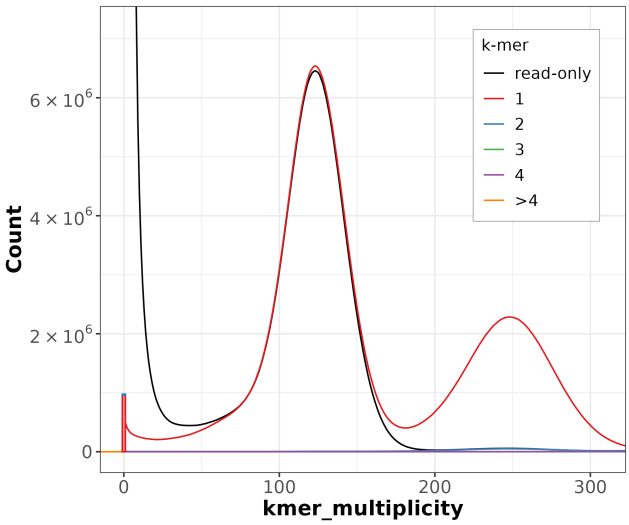
Warning! BUSCO versions or lineage datasets are not the same across results:
BUSCO: 6.0.0 (euk_genome_min, miniprot) / Lineage: mollusca_odb12 (genomes:36, BUSCOs:4421)
BUSCO: 5.8.2 (euk_genome_min, miniprot) / Lineage: mollusca_odb12 (genomes:36, BUSCOs:4421)

HiC contact map of curated assembly

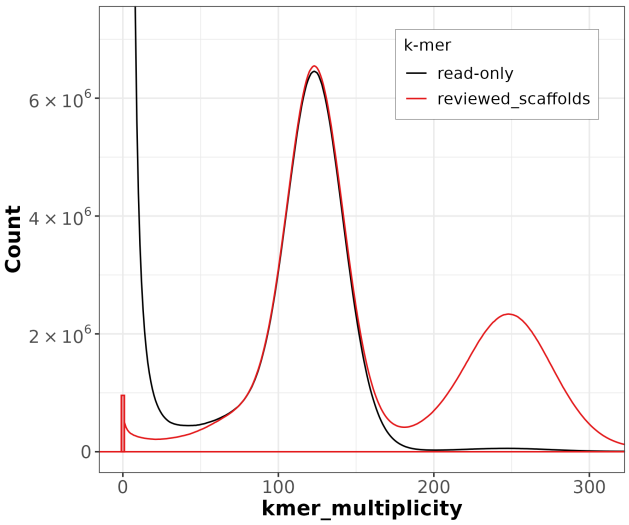


collapsed [LINK]

K-mer spectra of curated assembly

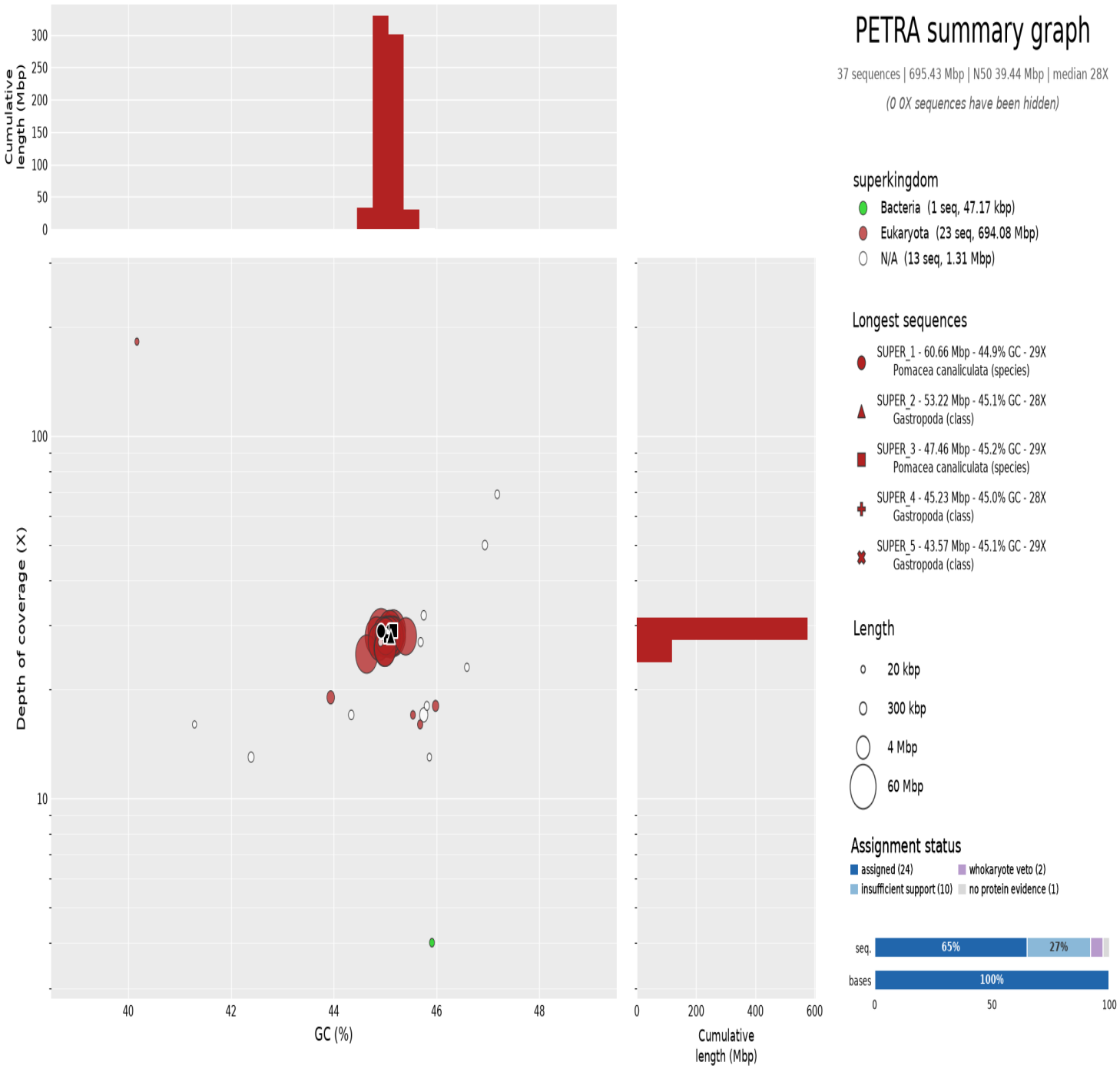


Distribution of k-mer counts per copy numbers found in asm



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	34	185

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:* 0.2.4
  |_ *key param:* NA
- **PretextView**
  |_ *ver:* 1.0.7
  |_ *key param:* NA

Submitter: Mohamed Meftah Saoues
Affiliation: Genoscope

Date and time: 2026-09-29 07:34:07 CEST