ERGA Assembly Report

v24.10.15

Tags: ATLASea[INVALID TAG]

TxID	496813
ToLID	**xgCerNodu1**
Species	Cerithium nodulosum
Class	Gastropoda
Order	NA

Genome Traits	Expected	Observed
Haploid size (bp)	733,733,189	794,547,244
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: 7.7.Q42

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

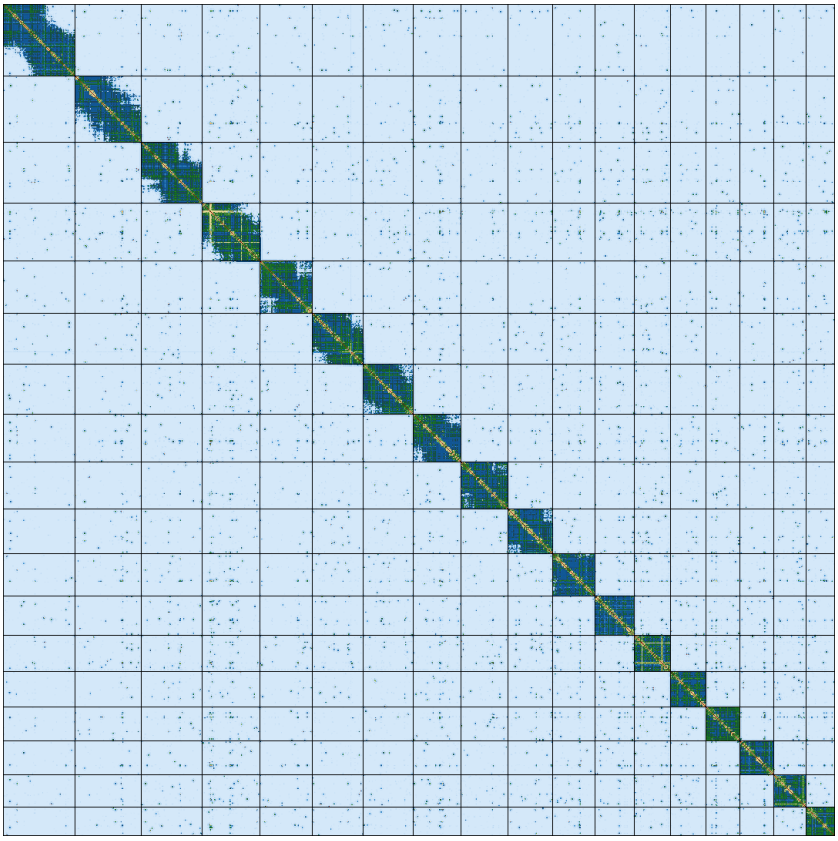
. Interventions/Gb: 41
. Contamination notes: ""
. Other observations: "The assembly of Cerithium nodulosum (xgCerNodu1) is based on 61X ONT data and 138X Arima Hi-C data generated as part of the ATLASea programme (https://www.atlasea.fr). The assembly process included the following steps: initial ONT 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, 17 contigs were identified as contaminants (bacterial, archaeal, or viral), totaling 1.892 Mb (with the largest being 0.352 Mb). Additionally, 22 regions totaling 9.217 Mb (with the largest being 3.492 Mb) were identified as haplotypic duplications and removed. Finally, the primary assembly was analyzed and manually improved using Pretext. During manual curation, 10 haplotypic regions and 1 contaminant sequence were removed, totaling 8.246Mb and 0.061Mb, respectively (with the largest being 2.214Mb and 0.061Mb). The telomeric pattern TTAGGG was identified with TelFinder and used to generate the Pretext Telomeres tracks. 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	802,839,266	794,547,244
GC %	44.92	44.92
Gaps/Gbp	41.1	42.79
Total gap bp	3,300	4,400
Scaffolds	24	22
Scaffold N50	46,861,492	45,738,299
Scaffold L50	8	8
Scaffold L90	16	16
Contigs	57	56
Contig N50	25,091,851	26,231,284
Contig L50	11	10
Contig L90	30	29
QV	42.8175	42.8211
Kmer compl.	85.6245	85.462
BUSCO sing.	96.8%	94.3%
BUSCO dupl.	1.6%	0.7%
BUSCO frag.	0.8%	4.0%
BUSCO miss.	0.7%	1.0%

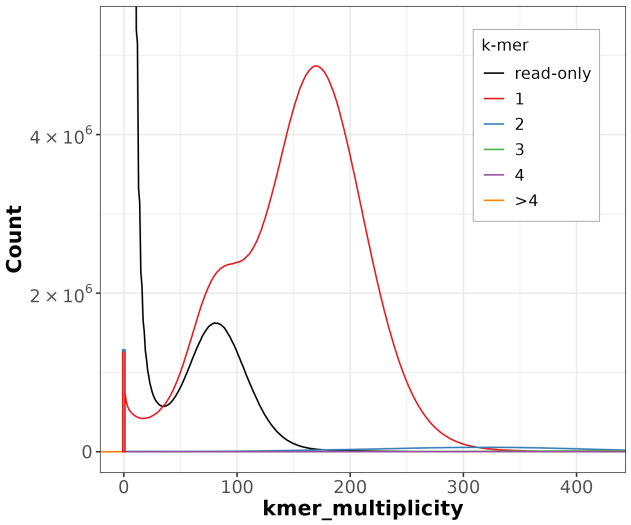
Warning! BUSCO versions or lineage datasets are not the same across results:
BUSCO: 6.0.0 (euk_genome_min, miniprot) / Lineage: mollusca_odb12 (genomes:36, BUSCOs:4421)
BUSCO: 5.8.2 (euk_genome_min, miniprot) / Lineage: mollusca_odb12 (genomes:36, BUSCOs:4421)

HiC contact map of curated assembly

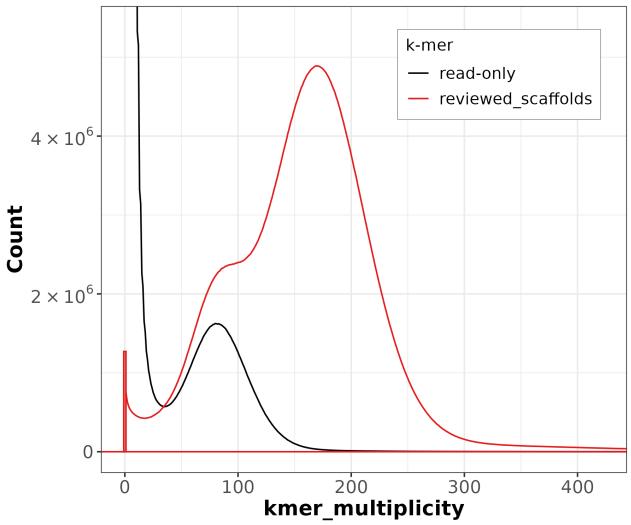


collapsed [LINK]

K-mer spectra of curated assembly

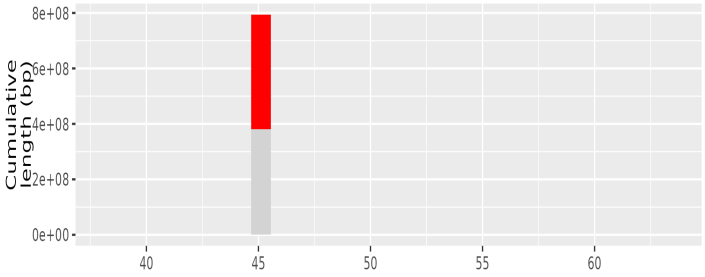


Distribution of k-mer counts per copy numbers found in asm

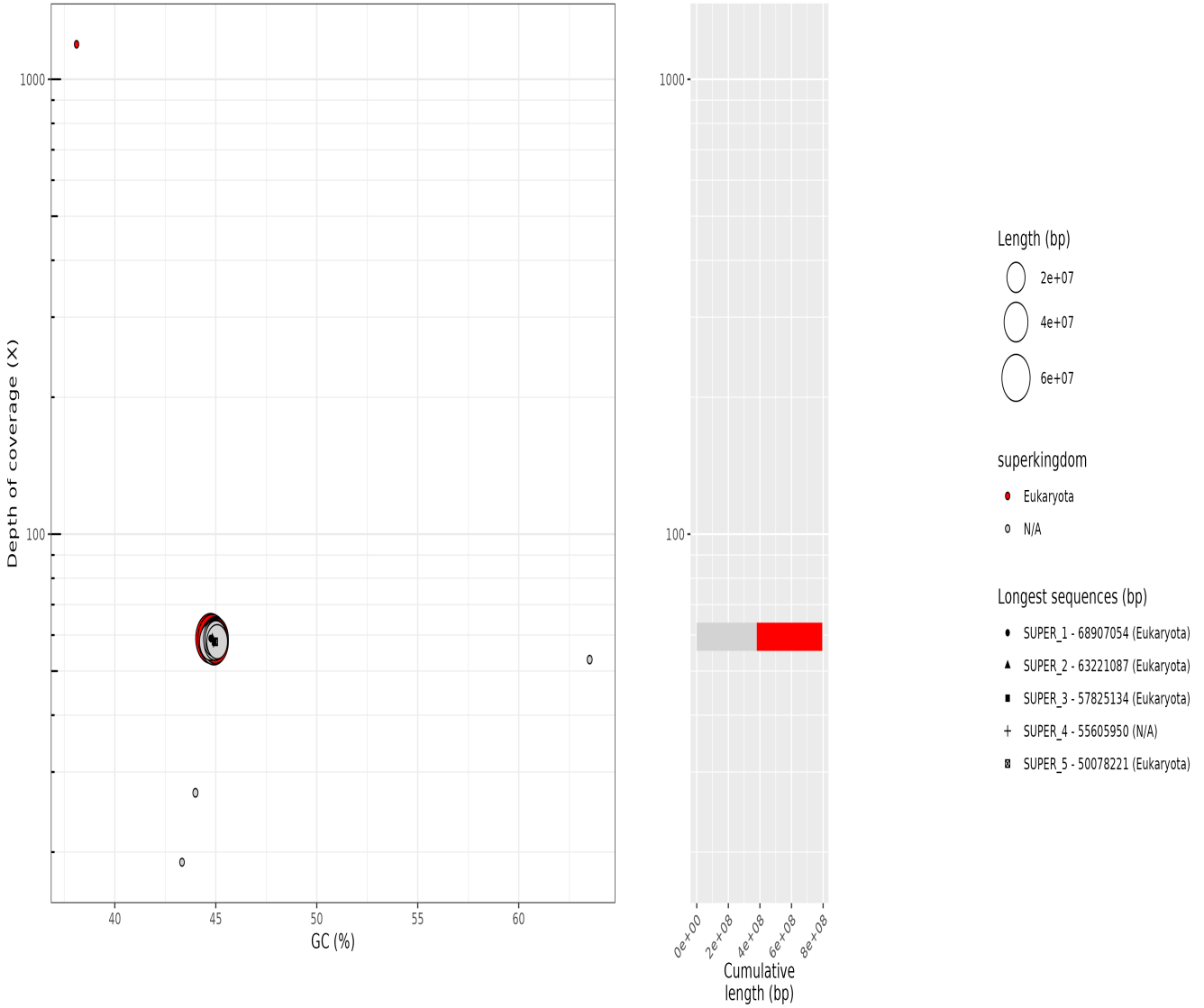


Distribution of k-mer counts coloured by their presence in reads/assemblies

Post-curation contamination screening



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	62	139

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: 2026-08-01 15:36:29 CEST