ERGA Assembly Report

v24.10.15

Tags: ATLASea[INVALID TAG]

TxID	744580
ToLID	**xgEmaOcta1_**
Species	Emarginula octaviana
Class	Gastropoda
Order	Lepetellida

Genome Traits	Expected	Observed
Haploid size (bp)	1,516,848,270	1,554,981,467
Haploid Number	9 (source: ancestor)	11
Ploidy	2 (source: ancestor)	2
Sample Sex	Unknown	Unknown

EBP metrics summary and curation notes

Obtained EBP quality metric for collapsed: 7.8.Q48

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
. Assembly length loss > 3% for collapsed

Curator notes

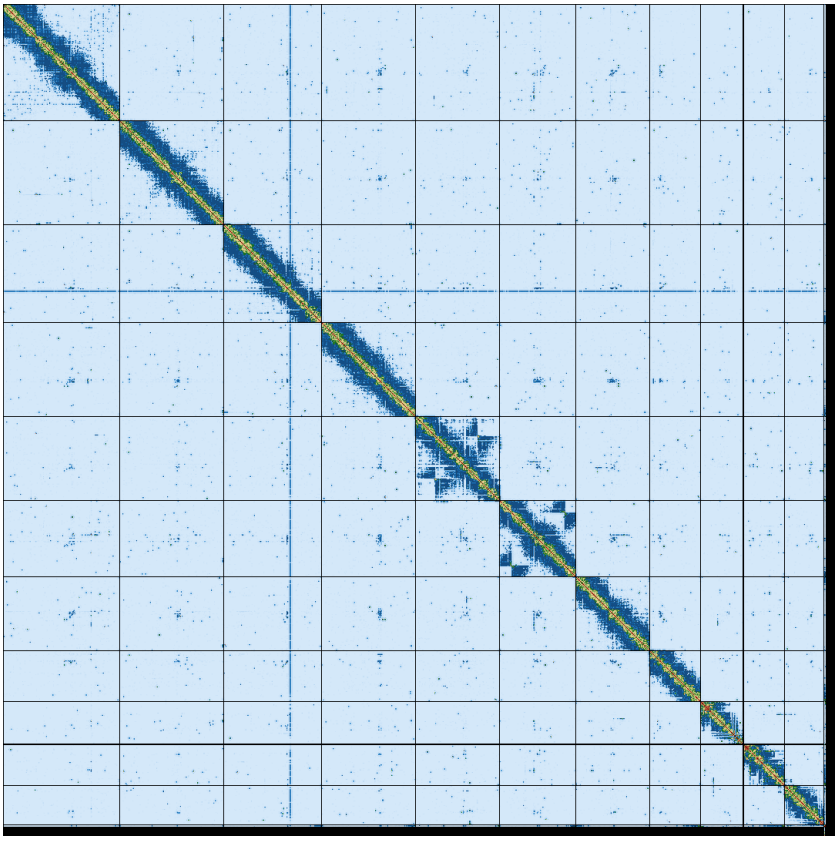
. Interventions/Gb: 48
. Contamination notes: ""
. Other observations: "The assembly of Emarginula octaviana (xgEmaOcta1_) is based on 53X PacBio data and 163X Arima Hi-C data generated as part of the ATLASea programme (https://www.atlasea.fr). The assembly process included the following steps: initial 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, 9 contigs were identified as contaminants (bacterial, archaeal, or viral), totaling 0.503 Mb (with the largest being 0.109 Mb). Additionally, 227 regions totaling 164.645 Mb (with the largest being 7.382 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, 32 haplotypic regions were removed, totaling 125.951Mb (with the largest being 13.766Mb). 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	1,680,930,437	1,554,981,467
GC %	37.95	37.94
Gaps/Gbp	79.12	75.89
Total gap bp	13,300	14,600
Scaffolds	109	101
Scaffold N50	174,479,137	158,217,951
Scaffold L50	5	5
Scaffold L90	10	10
Contigs	242	219
Contig N50	26,076,000	27,970,919
Contig L50	17	15
Contig L90	74	61
QV	48.231	48.2385
Kmer compl.	66.6061	63.2567
BUSCO sing.	91.0%	98.3%
BUSCO dupl.	8.1%	0.8%
BUSCO frag.	0.2%	0.2%
BUSCO miss.	0.7%	0.7%

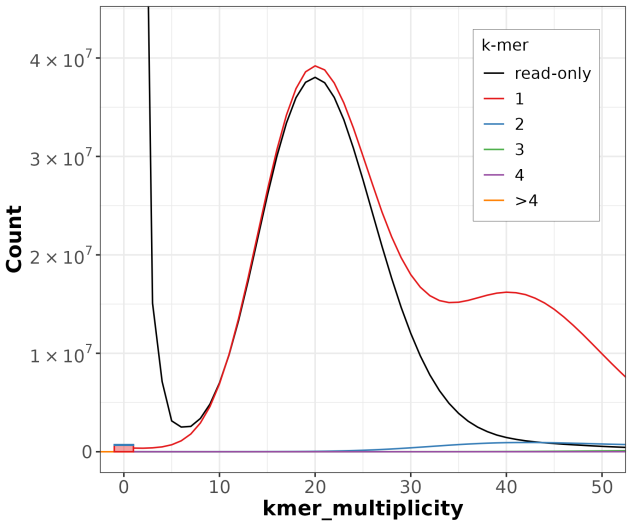
BUSCO: 6.0.0 (euk_genome_min, miniprot) / Lineage: mollusca_odb12 (genomes:36, BUSCOs:4421)

HiC contact map of curated assembly

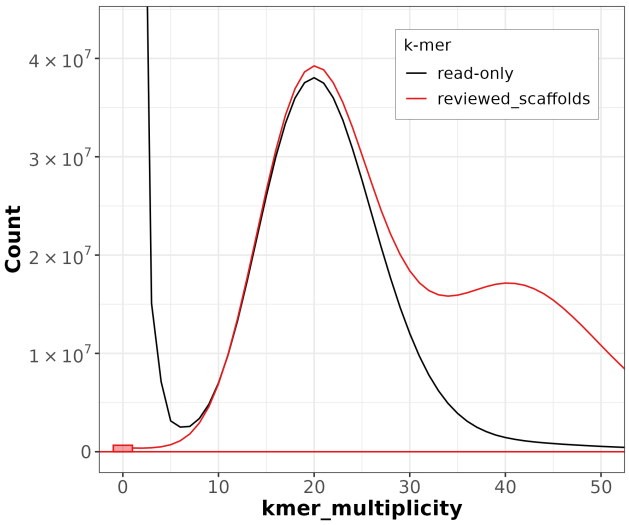


collapsed [LINK]

K-mer spectra of curated assembly

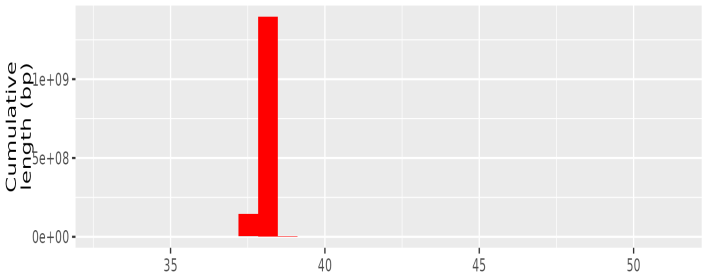


Distribution of k-mer counts per copy numbers found in asm

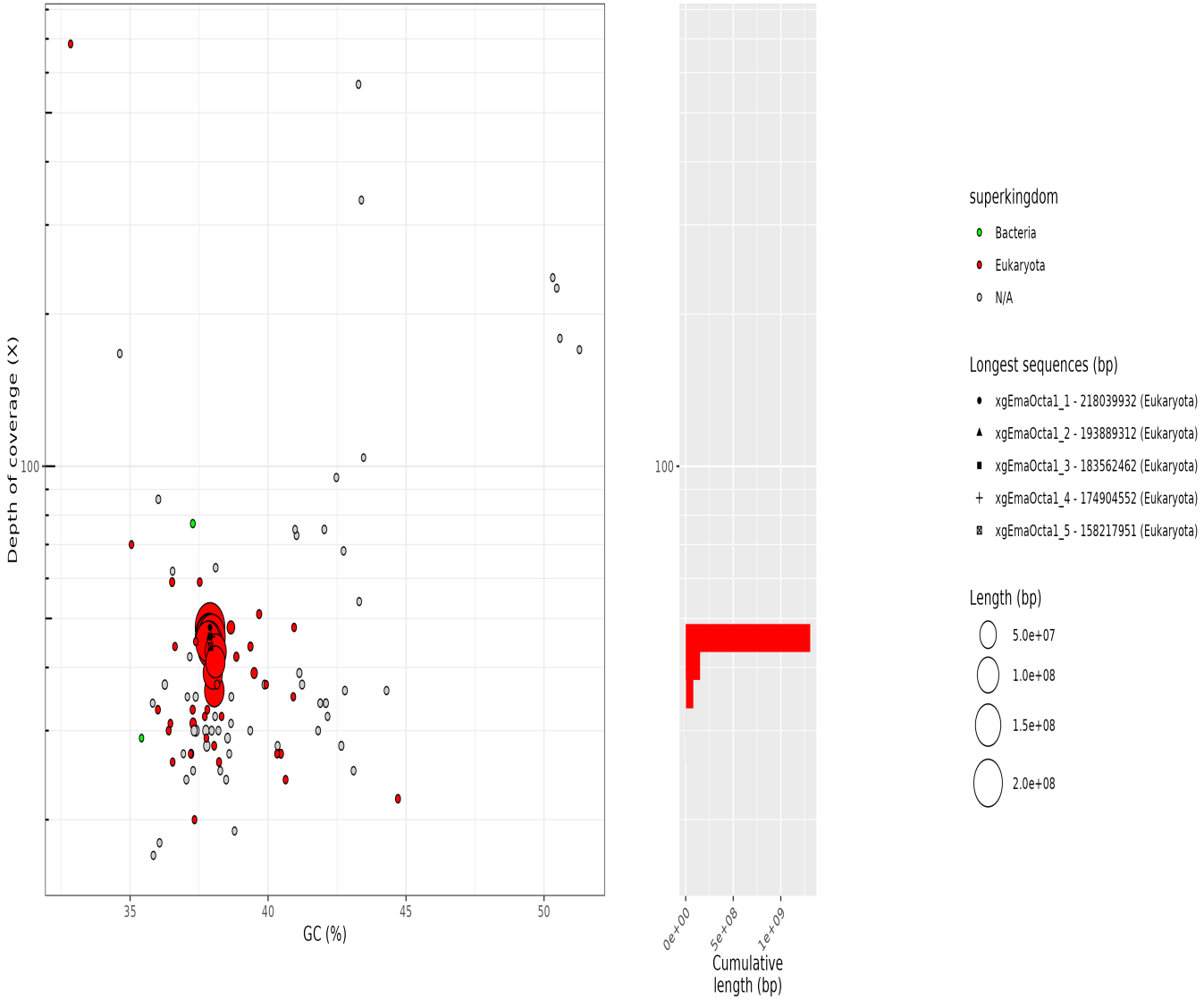


Distribution of k-mer counts coloured by their presence in reads/assemblies

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	57	176

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: 2026-06-08 16:54:15 CEST