ERGA Assembly Report

v24.09.10

Tags: ATLASea[INVALID TAG]

TxID	87958
ToLID	**xgPatCaer1.1**
Species	Patella caerulea
Class	Gastropoda
Order	NA

Genome Traits	Expected	Observed
Haploid size (bp)	700,902,420	725,889,058
Haploid Number	9 (source: ancestor)	9
Ploidy	2 (source: ancestor)	2
Sample Sex	Unknown	Unknown

EBP metrics summary and curation notes

Obtained EBP quality metric for collapsed: 7.7.Q68

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

. Kmer completeness value is less than 90 for collapsed
. BUSCO single copy value is less than 90% for collapsed
. Assembly length loss > 3% for collapsed

Curator notes

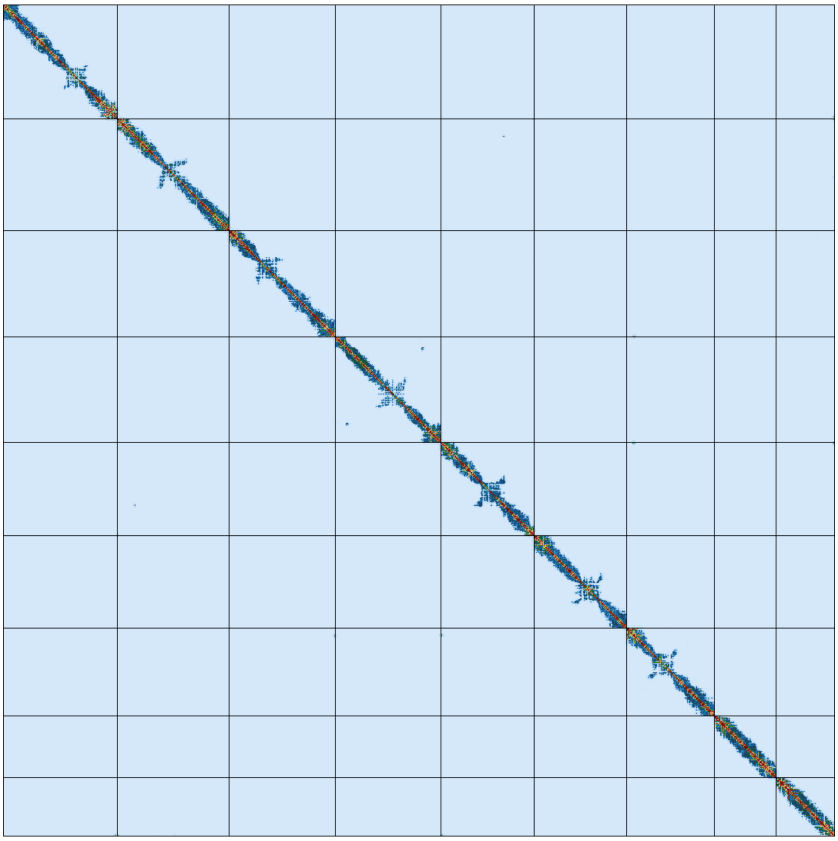
. Interventions/Gb: 30
. Contamination notes: "81 contigs were identified as contaminant sequences and removed, totaling 2.1 Mb, with the largest being 79 Kb. "
. Other observations: "Few modifications, contigs were already mostly chromosome scale. However, eleven haplotypic regions were removed, totaling 36Mb, with the largest being 19Mb (entire contig). Contigs were already mostly chromosome scale. "

Quality metrics table

Metrics	Pre-curation collapsed	Curated collapsed
Total bp	763,088,660	725,889,058
GC %	36.05	36.05
Gaps/Gbp	31.45	35.82
Total gap bp	2,400	3,400
Scaffolds	60	19
Scaffold N50	93,002,498	92,239,536
Scaffold L50	4	4
Scaffold L90	8	8
Contigs	84	45
Contig N50	30,243,000	30,569,455
Contig L50	10	9
Contig L90	24	22
QV	45.6338	68.5752
Kmer compl.	72.9186	71.4784
BUSCO sing.	83.2%	87.5%
BUSCO dupl.	5.3%	1.0%
BUSCO frag.	4.8%	4.8%
BUSCO miss.	6.7%	6.7%

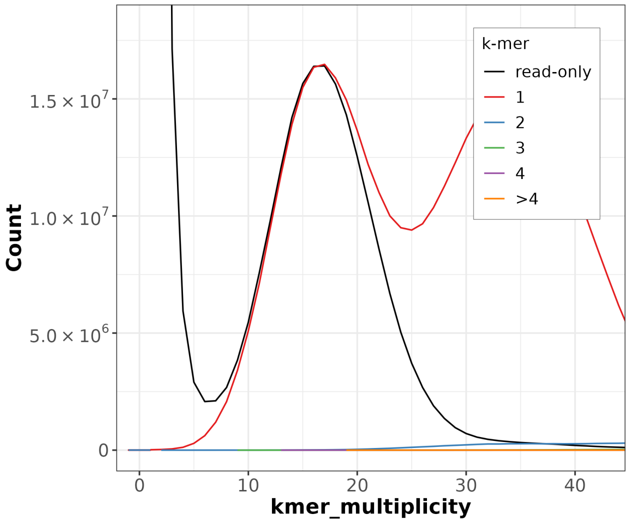
BUSCO 5.4.3 Lineage: mollusca_odb10 (genomes:7, BUSCOs:5295)

HiC contact map of curated assembly

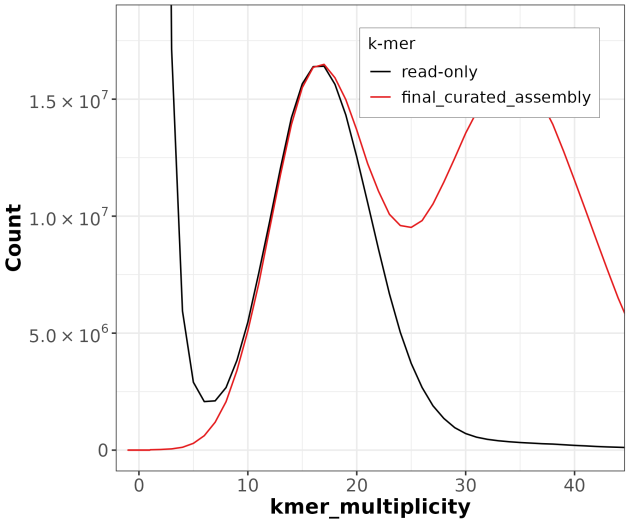


collapsed [LINK]

K-mer spectra of curated assembly

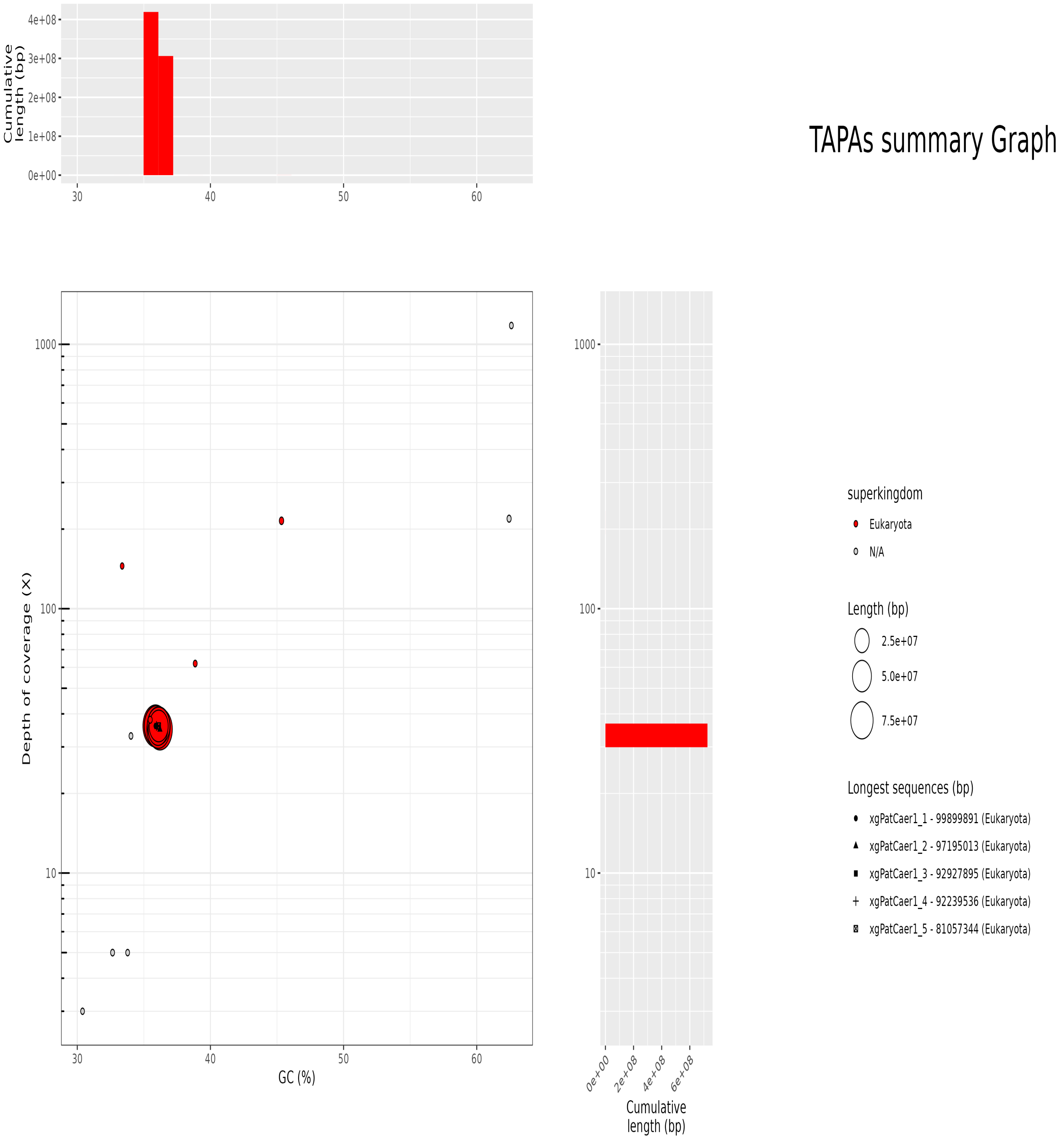


Distribution of k-mer counts per copy numbers found in asm



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	PACBIO Hifi	Arima (4-enz)
Coverage	36	50

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: Jean-Marc Aury
Affiliation: Genoscope

Date and time: 2024-11-15 08:13:54 CET