ERGA Assembly Report

v24.10.15

Tags: ATLASea[INVALID TAG]

TxID	703305
ToLID	**xgStePenn6**
Species	Steromphala pennanti
Class	Gastropoda
Order	Trochida

Genome Traits	Expected	Observed
Haploid size (bp)	1,194,257,162	1,230,656,043
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

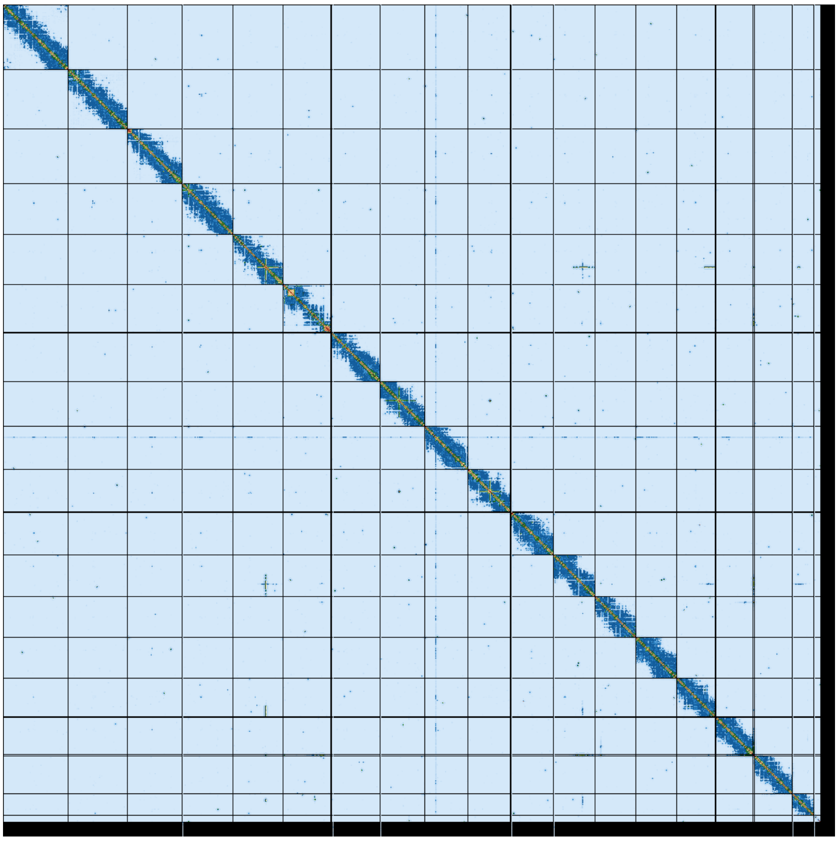
. Interventions/Gb: 47
. Contamination notes: ""
. Other observations: "The assembly of Steromphala pennanti (xgStePenn6) is based on 71X PacBio data and 162X Arima Hi-C data generated as part of the ATLASea programme (https://www.atlasea.fr). The assembly process included the following steps: initial 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, 65 contigs were identified as contaminants (bacterial), totaling 3.59 Mb (with the largest being 0.372 Mb). Additionally, 224 regions totaling 38.83 Mb (with the largest being 6.07 Mb) were identified as haplotypic duplications and removed. The mitochondrial genome was assembled using OATK. Finally, the primary assembly was analyzed and manually improved using Pretext. During manual curation, 18 haplotypic regions totaling 14.86 Mb were removed (with the largest being 3.3 Mb). 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	1,245,512,529	1,230,656,043
GC %	35.37	35.37
Gaps/Gbp	89.12	95.07
Total gap bp	11,100	14,500
Scaffolds	165	146
Scaffold N50	63,343,507	65,805,273
Scaffold L50	9	8
Scaffold L90	17	17
Contigs	276	263
Contig N50	19,544,000	19,544,000
Contig L50	22	22
Contig L90	72	70
QV	51.6735	51.6629
Kmer compl.	66.2206	65.8714
BUSCO sing.	97.6%	98.3%
BUSCO dupl.	1.3%	0.5%
BUSCO frag.	0.6%	0.6%
BUSCO miss.	0.5%	0.5%

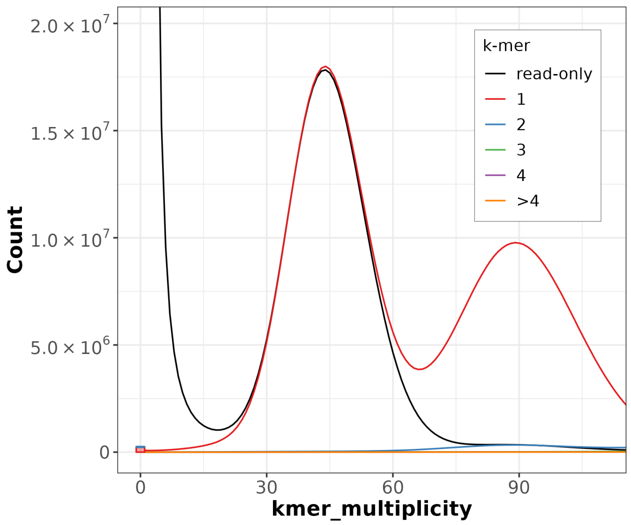
BUSCO: 6.0.0 (euk_genome_min, miniprot) / Lineage: mollusca_odb12 (genomes:36, BUSCOs:4421)

HiC contact map of curated assembly

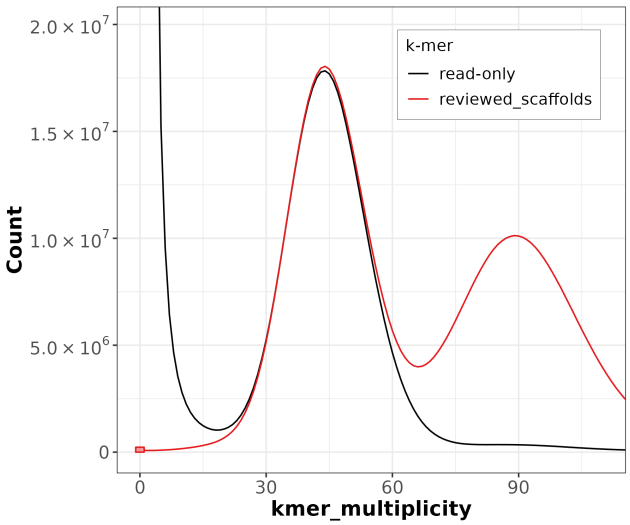


collapsed [LINK]

K-mer spectra of curated assembly

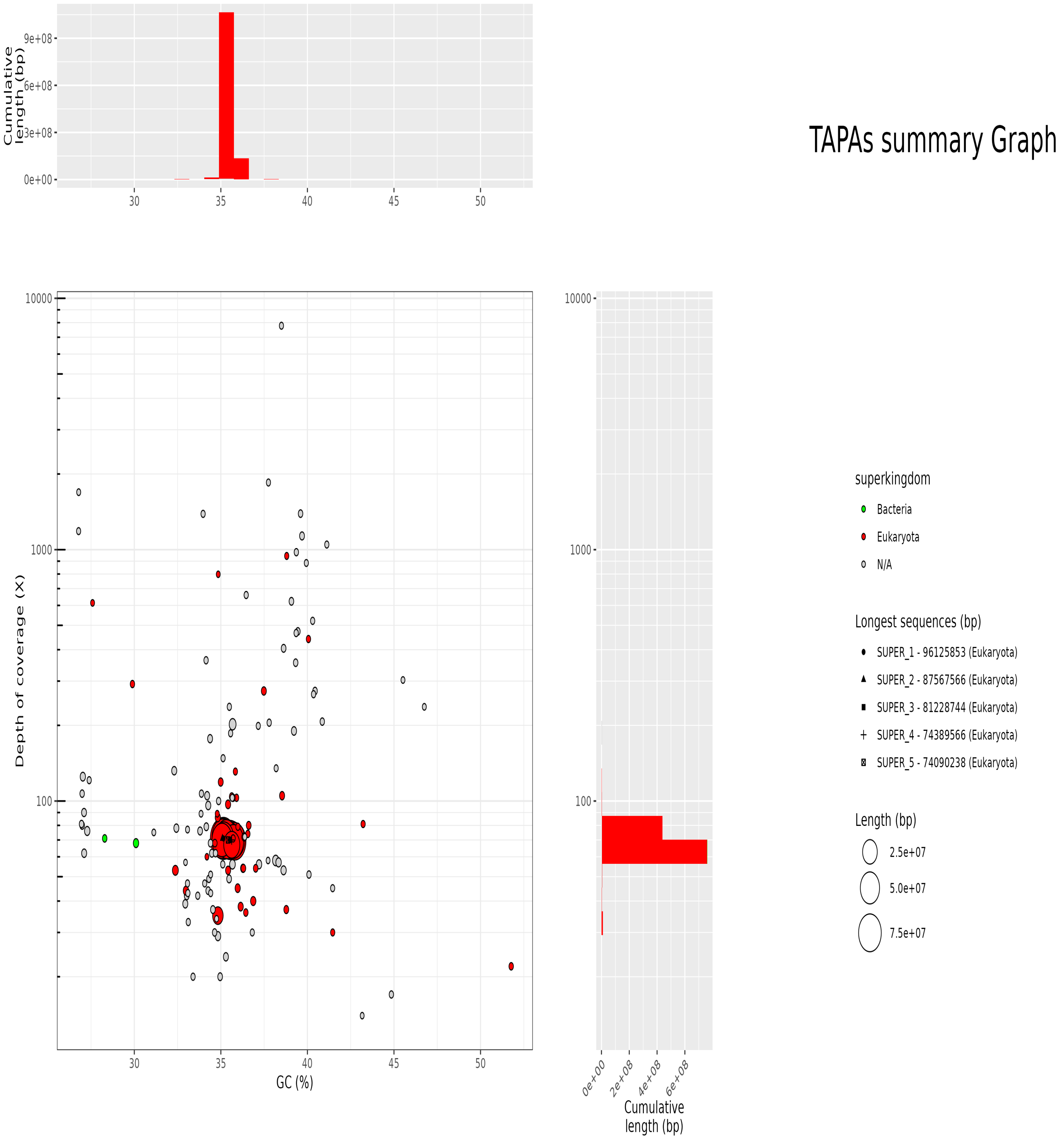


Distribution of k-mer counts per copy numbers found in asm



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

Post-curation contamination screening



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	71	162

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: Adama Ndar
Affiliation: Genoscope

Date and time: 2025-11-13 22:12:32 CET