ERGA Assembly Report

v24.10.15

Tags: ATLASea[INVALID TAG]

TxID	509997
ToLID	**xbTalMult1**
Species	Talochlamys multistriata
Class	Bivalvia
Order	Pectinida

Genome Traits	Expected	Observed
Haploid size (bp)	1,166,818,007	1,186,413,610
Haploid Number	19 (source: ancestor)	19
Ploidy	2 (source: ancestor)	2
Sample Sex	Unknown	Unknown

EBP metrics summary and curation notes

Obtained EBP quality metric for collapsed: 6.7.Q48

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

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

Curator notes

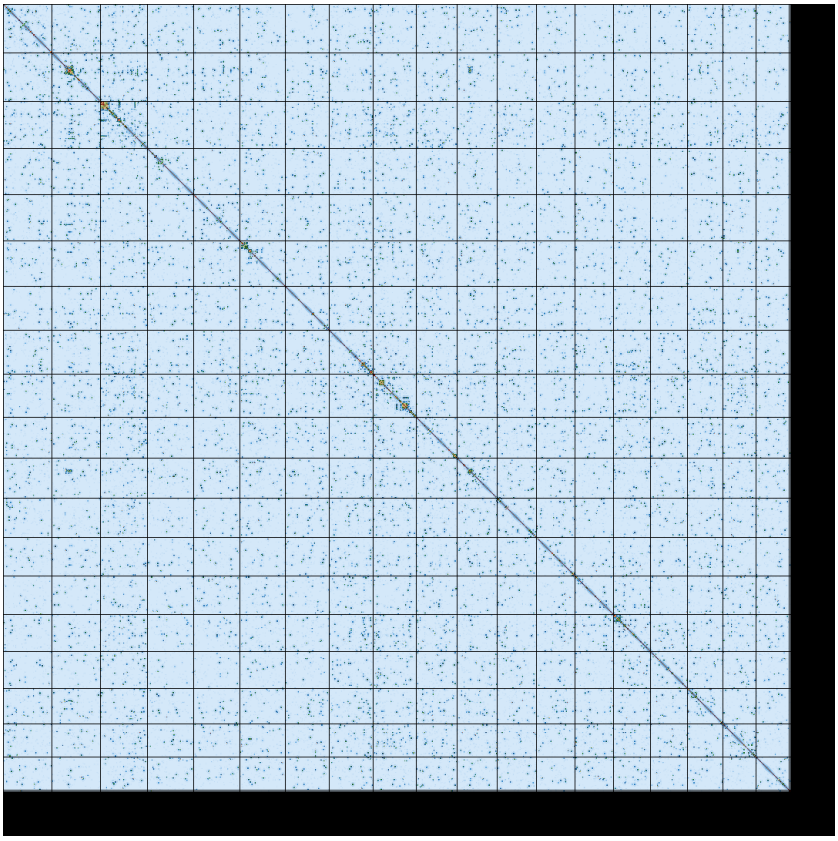
. Interventions/Gb: 107
. Contamination notes: ""
. Other observations: "The assembly of Talochlamys multistriata (xbTalMult1) is based on 34.42X PacBio data and 108.71X Arima Hi-C data generated as part of the ATLASea programme (https://www.atlasea.fr). The assembly process included the following steps: initial PacBio+HiC assembly generation with Hifiasm, removal of contaminant sequences using Context, and Hi-C-based scaffolding with YaHS of each haplotype. In total, 73 contigs were identified as contaminants (bacterial, archaeal, or viral), totaling 2.552 Mb (with the largest being 0.246 Mb). The mitochondrial genome was assembled using Oatk. Finally, Both haplotypes were analyzed and manually improved using Pretext. During manual curation, 8 haplotypic regions were removed, totaling 2.427Mb (with the largest being 1.487Mb). 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	1,753,839,834	1,186,413,610
GC %	36.94	36.95
Gaps/Gbp	181.89	261.29
Total gap bp	31,900	37,700
Scaffolds	2,573	1,003
Scaffold N50	73,247,304	57,842,443
Scaffold L50	10	10
Scaffold L90	36	18
Contigs	2,892	1,313
Contig N50	8,303,000	6,368,809
Contig L50	59	51
Contig L90	236	199
QV	49.0346	48.7123
Kmer compl.	80.3968	59.4541
BUSCO sing.	48.9%	93.3%
BUSCO dupl.	50.1%	2.0%
BUSCO frag.	0.8%	0.9%
BUSCO miss.	0.3%	3.8%

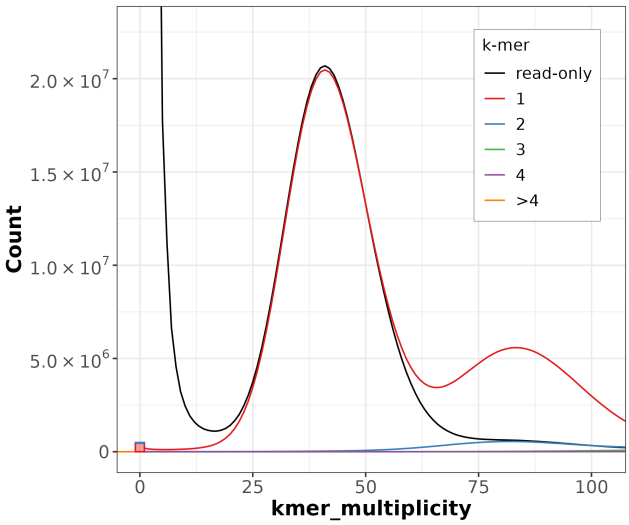
Warning! BUSCO versions or lineage datasets are not the same across results:
BUSCO: 6.0.0 (euk_genome_min, miniprot) / Lineage: mollusca_odb12 (genomes:36, BUSCOs:4421)
BUSCO: 6.1.0 (euk_genome_min, miniprot) / Lineage: mollusca_odb12 (genomes:36, BUSCOs:4421)

HiC contact map of curated assembly

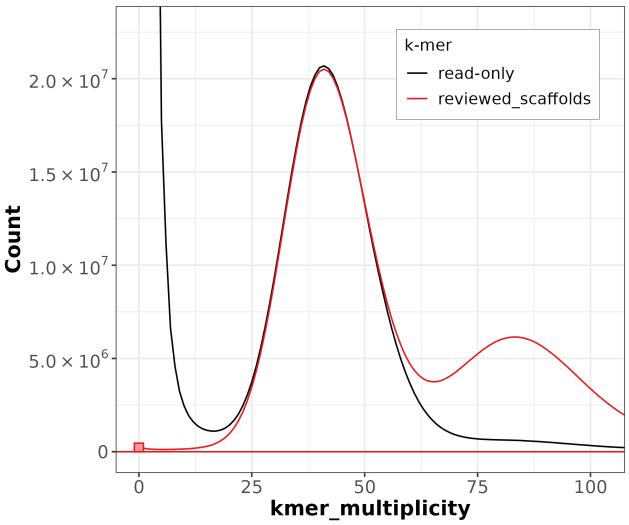


collapsed [LINK]

K-mer spectra of curated assembly

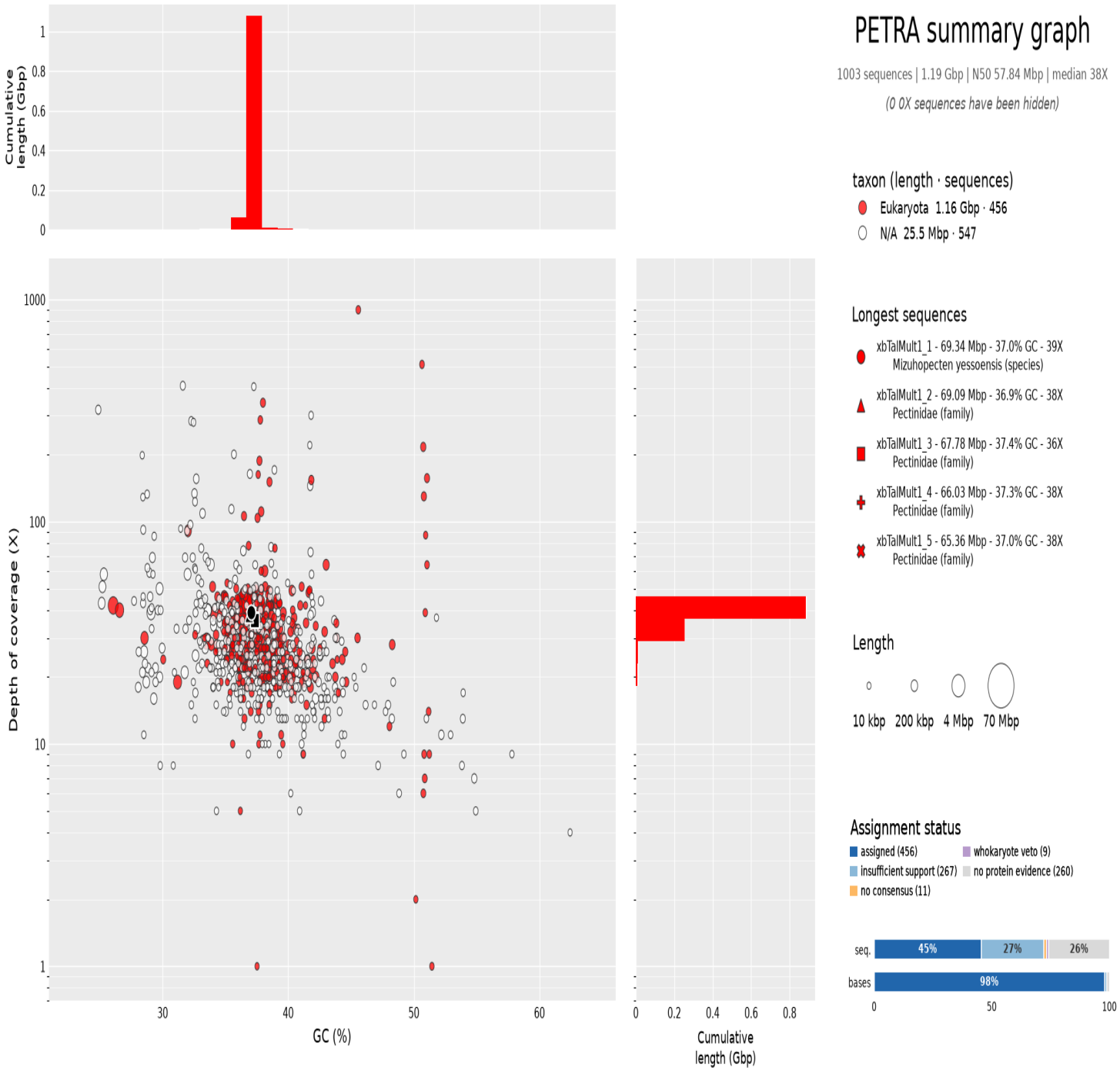


Distribution of k-mer counts per copy numbers found in asm



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	50	160

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: Emilie Teodori
Affiliation: Genoscope

Date and time: 2026-10-07 02:59:31 CEST