

ERGA Assembly Report

v24.10.15

Tags: ATLASea[INVALID TAG]

TxID	2724334
ToLID	xbSacMor01
Species	Saccostrea mordoides
Class	Bivalvia
Order	Ostreida

Genome Traits	Expected	Observed
Haploid size (bp)	676,011,082	710,418,182
Haploid Number	14 (source: ancestor)	10
Ploidy	2 (source: ancestor)	2
Sample Sex	Unknown	Unknown

EBP metrics summary and curation notes

Obtained EBP quality metric for collapsed: 6.7.Q45

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

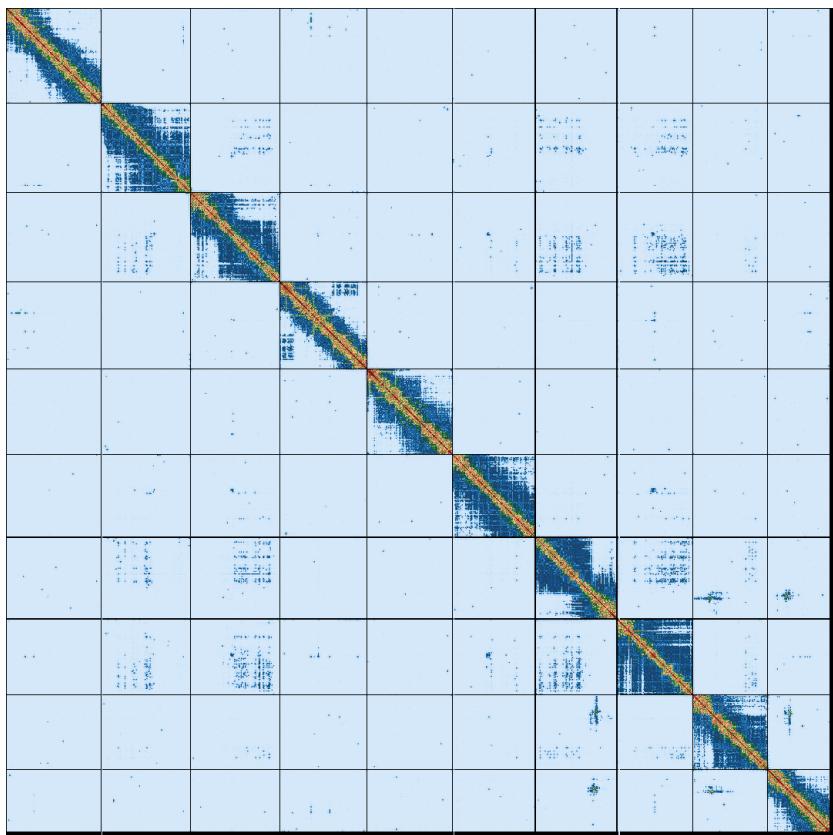
- . Interventions/Gb: 116
- . Contamination notes: ""
- . Other observations: "The assembly of Saccostrea mordoides (xbSacMor01) is based on 63X PacBio data and 230X Arima Hi-C data generated as part of the ATLASea programme (<https://www.atlasea.fr>). The assembly process included the following steps: initial PacBio phased assembly generation with Hifiasm using Hi-C data, removal of contaminant sequences using Context, and Hi-C-based scaffolding with YaHS. In total, 3 contigs were identified as contaminants (bacterial, archaeal, or viral), totaling 0.091 Mb (with the largest being 0.037 Mb). Additionally, 161 regions totaling 63.251 Mb (with the largest being 3.333 Mb) were identified as haplotypic duplications and removed. Mitochondrial genome was assembled using OATK. During manual curation, 2 haplotypic regions were removed, totaling 1.07Mb (with the largest being 0.88Mb). A first round of curation was made on the combined map using Pretext followed by a curation made on each haplotype. 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	812,423,620	710,418,182
GC %	33.84	33.9
Gaps/Gbp	115.7	388.5
Total gap bp	9,400	32,800
Scaffolds	38	213
Scaffold N50	69,131,551	72,628,033
Scaffold L50	5	5
Scaffold L90	10	9
Contigs	132	489
Contig N50	14,063,468	4,336,661
Contig L50	20	51
Contig L90	66	166
QV	45.3231	45.198
Kmer compl.	69.2848	62.1851
BUSCO sing.	90.6%	98.5%
BUSCO dupl.	9.1%	1.1%
BUSCO frag.	0.0%	0.1%
BUSCO miss.	0.3%	0.3%

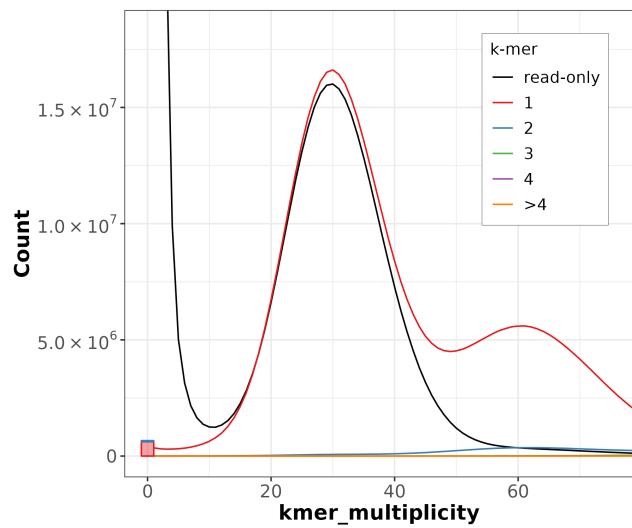
BUSCO: 6.0.0 (euk_genome_min, miniprot) / Lineage: mollusca_odb12 (genomes:36, BUSCOs:4421)

HiC contact map of curated assembly

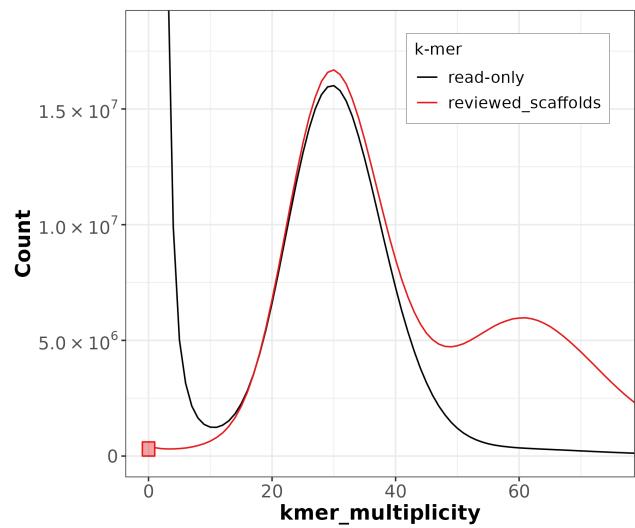


collapsed [\[LINK\]](#)

K-mer spectra of curated assembly

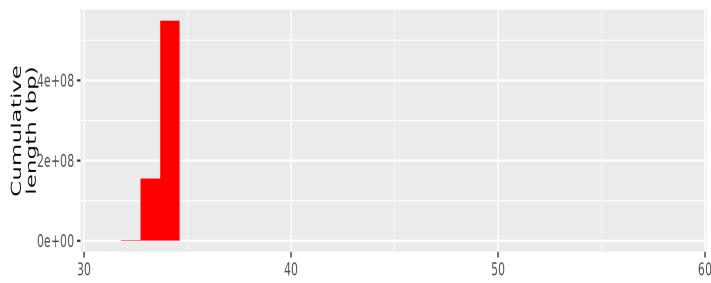


Distribution of k-mer counts per copy numbers found in asm

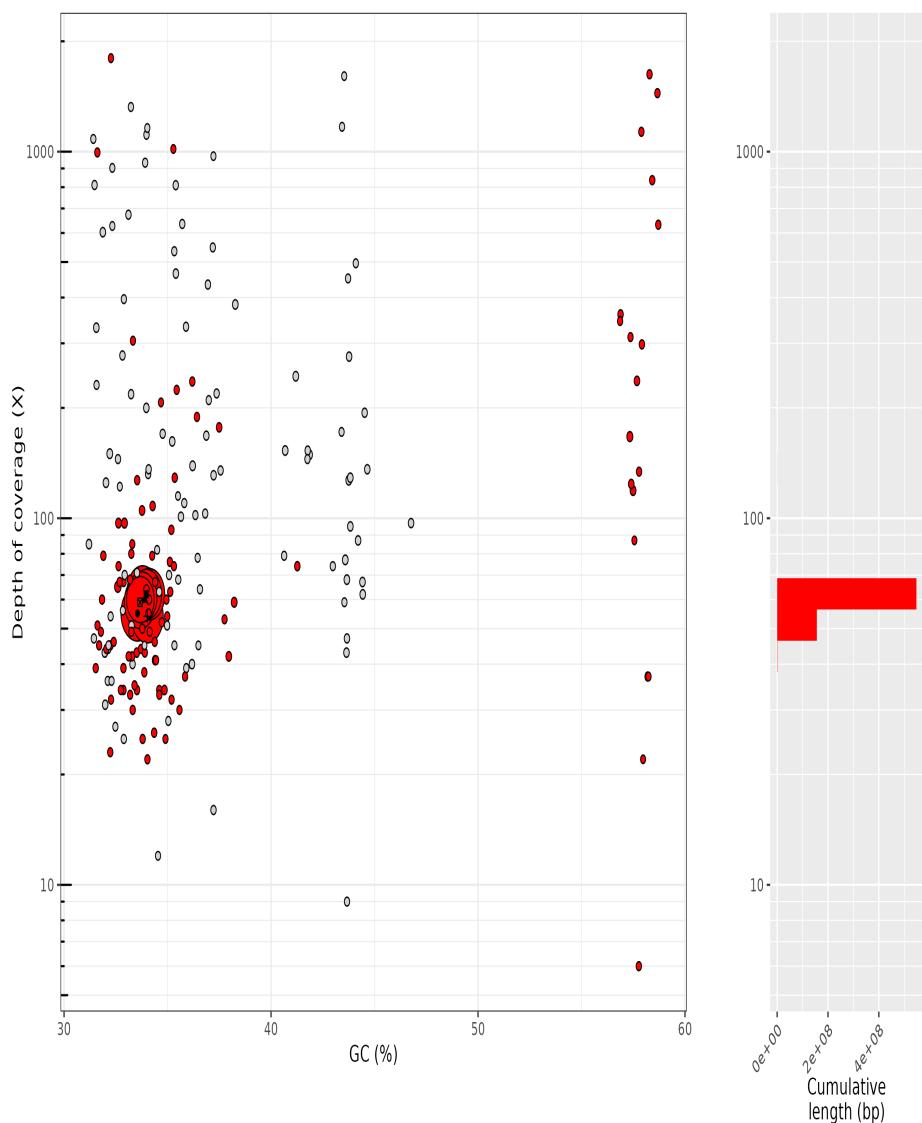


Distribution of k-mer counts coloured by their presence in reads/assemblies

Post-curation contamination screening



TAPAs summary Graph



Length (bp)

- 2e+07
- 4e+07
- 6e+07
- 8e+07

superkingdom

- Eukaryota
- N/A

Longest sequences (bp)

- xbSacMoro1_1 - 81369808 (Eukaryota)
- ▲ xbSacMoro1_2 - 76319830 (Eukaryota)
- xbSacMoro1_3 - 76055168 (Eukaryota)
- + xbSacMoro1_4 - 74416220 (Eukaryota)
- ☒ xbSacMoro1_5 - 72628033 (Eukaryota)

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	63	230

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: 2025-11-28 04:29:30 CET