ERGA Assembly Report

v24.10.15

Tags: ATLASea[INVALID TAG]

TxID	37643
ToLID	**xbCraRhiz1**
Species	Crassostrea rhizophorae
Class	Bivalvia
Order	Ostreida

Genome Traits	Expected	Observed
Haploid size (bp)	492,177,194	508,260,229
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.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

Curator notes

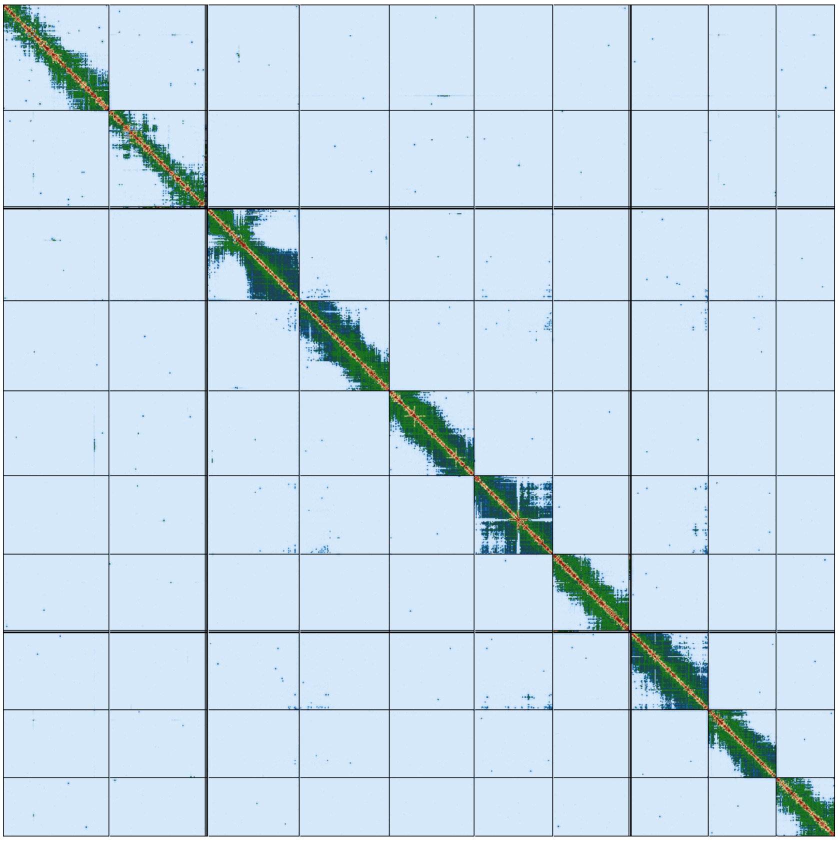
. Interventions/Gb: 70
. Contamination notes: ""
. Other observations: "The assembly of \'Crassostrea rhizophorae\' (xbCraRhiz1) is based on 69X PacBio data and 212X Arima Hi-C data generated as part of the ATLASea programme (https://www.atlasea.fr). The assembly process included the following steps: 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, 56 contigs were identified as contaminants (bacterial, archaeal, or viral), totaling 1.902 Mb (with the largest being 0.076 Mb). Additionally, 325 regions totaling 47.109 Mb (with the largest being 2.519 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, 22 haplotypic regions and 5 contaminant sequences were removed, totaling 10.077Mb and 0.092Mb, respectively (with the largest being 3.054Mb and 0.032Mb). 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	518,439,721	508,260,229
GC %	35.32	35.34
Gaps/Gbp	497.65	462.36
Total gap bp	29,500	29,000
Scaffolds	63	29
Scaffold N50	47,581,013	51,975,144
Scaffold L50	5	5
Scaffold L90	10	9
Contigs	304	264
Contig N50	7,488,295	8,143,000
Contig L50	20	19
Contig L90	92	86
QV	48.6935	48.8065
Kmer compl.	70.1154	69.2021
BUSCO sing.	96.8%	98.4%
BUSCO dupl.	1.0%	1.0%
BUSCO frag.	1.0%	0.1%
BUSCO miss.	1.2%	0.4%

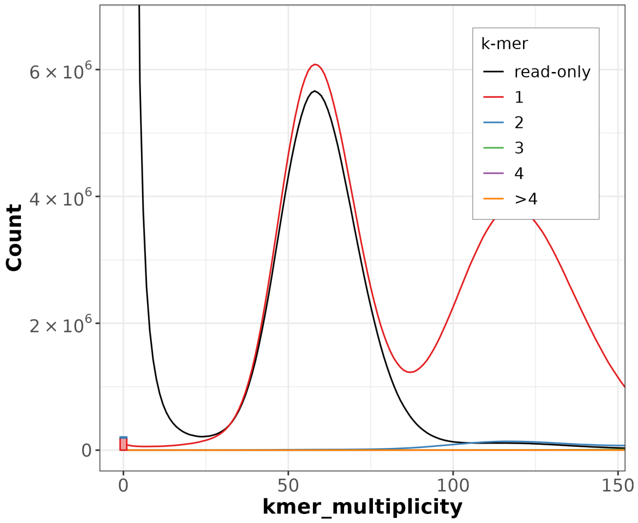
Warning! BUSCO versions or lineage datasets are not the same across results:
BUSCO: 5.8.2 (euk_genome_met, metaeuk) / Lineage: mollusca_odb12 (genomes:36, BUSCOs:4421)
BUSCO: 6.0.0 (euk_genome_min, miniprot) / Lineage: mollusca_odb12 (genomes:36, BUSCOs:4421)

HiC contact map of curated assembly

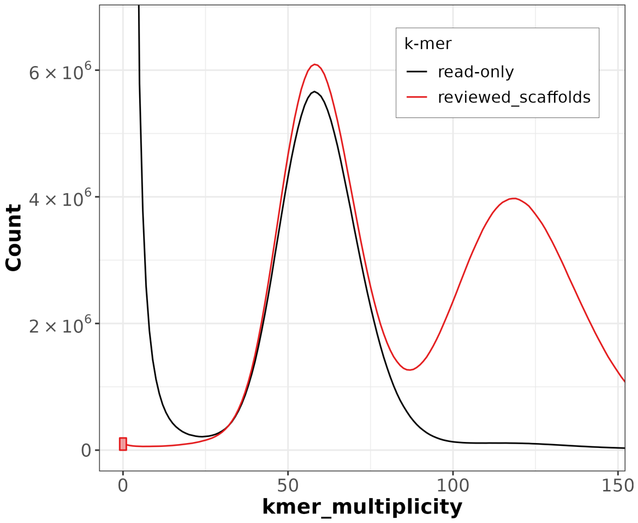


collapsed [LINK]

K-mer spectra of curated assembly

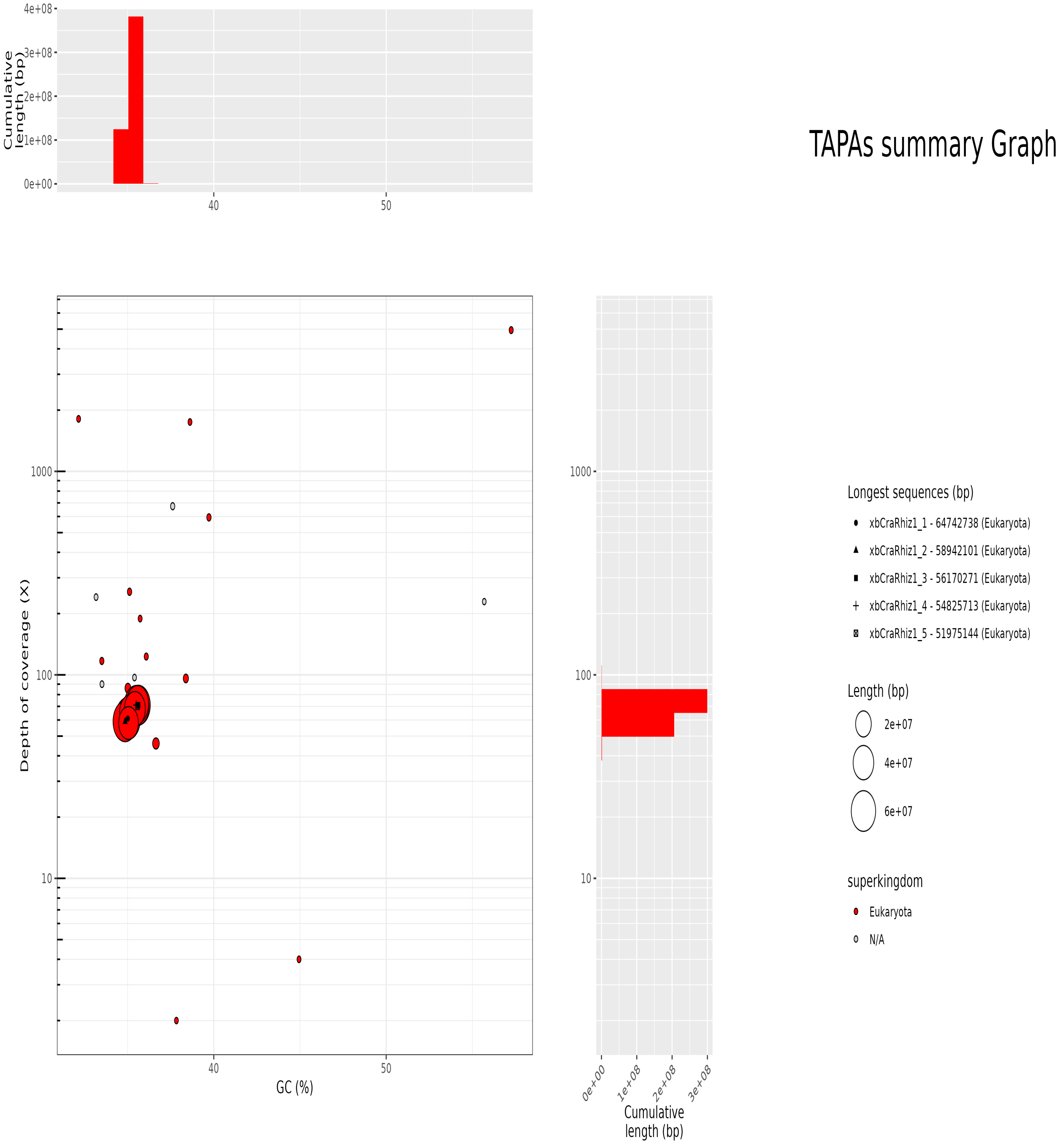


Distribution of k-mer counts per copy numbers found in asm



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	69	212

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: Simone Duprat
Affiliation: Genoscope

Date and time: 2025-10-28 15:29:45 CET