ERGA Assembly Report

v24.10.15

Tags: ATLASea[INVALID TAG]

TxID	102329
ToLID	**xbPinMarg2**
Species	Pinctada margaritifera
Class	Bivalvia
Order	Pterioida

Genome Traits	Expected	Observed
Haploid size (bp)	1,093,903,138	1,146,510,992
Haploid Number	14 (source: ancestor)	14
Ploidy	2 (source: ancestor)	2
Sample Sex	Unknown	Unknown

EBP metrics summary and curation notes

Obtained EBP quality metric for collapsed: 6.7.Q44

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

. Kmer completeness value is less than 90 for collapsed
. More than 1000 gaps/Gbp for collapsed

Curator notes

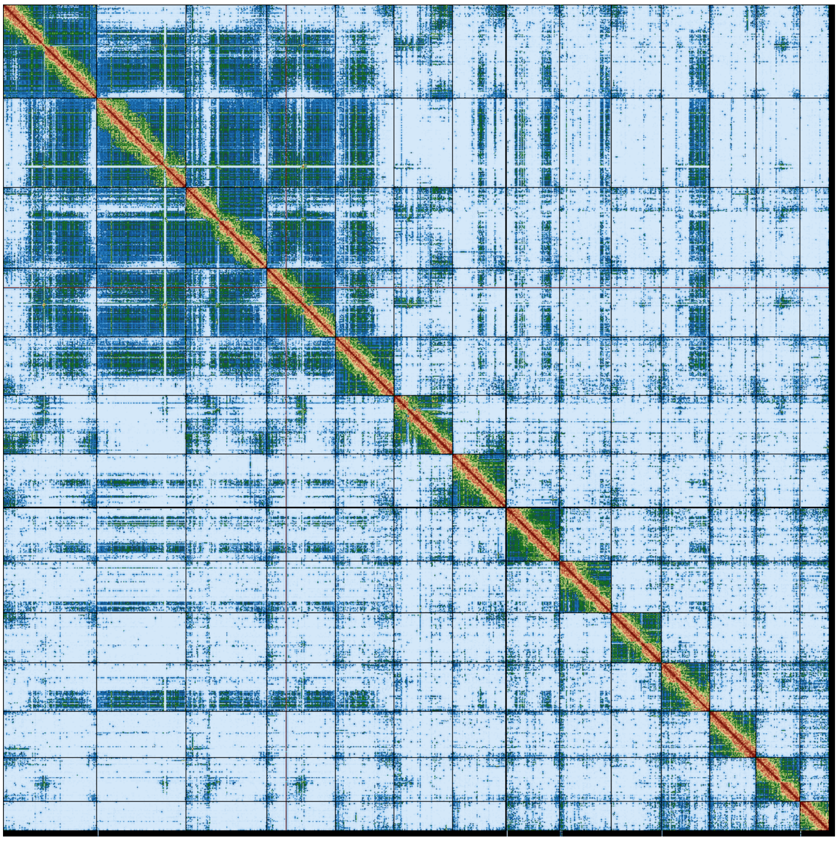
. Interventions/Gb: 60
. Contamination notes: ""
. Other observations: "The assembly of Pinctada margaritifera (xbPinMarg2) is based on 46X PacBio data and 215X Arima Hi-C data generated as part of the ATLASea programme (https://www.atlasea.fr). The assembly process included the following steps: initial PacBio+Hi-C assembly generation with Hifiasm, removal of contaminant sequences using Context, and Hi-C-based scaffolding with YaHS of each haplotype. In total, 6 contigs were identified as contaminants (bacterial, archaeal, or viral), totaling 0.412 Mb (with the largest being 0.139 Mb). The mitochondrial genome was assembled using OATK. Both haplotypes were analyzed and manually improved using Pretext. 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	1,137,223,033	1,146,510,992
GC %	35.99	36.03
Gaps/Gbp	522.32	1,018.74
Total gap bp	72,800	124,500
Scaffolds	81	206
Scaffold N50	76,687,450	80,408,348
Scaffold L50	6	6
Scaffold L90	13	12
Contigs	641	1,374
Contig N50	3,679,268	1,605,988
Contig L50	95	224
Contig L90	312	713
QV	44.9997	44.4496
Kmer compl.	65.2566	65.6129
BUSCO sing.	90.5%	97.5%
BUSCO dupl.	0.5%	0.6%
BUSCO frag.	6.3%	0.6%
BUSCO miss.	2.7%	1.3%

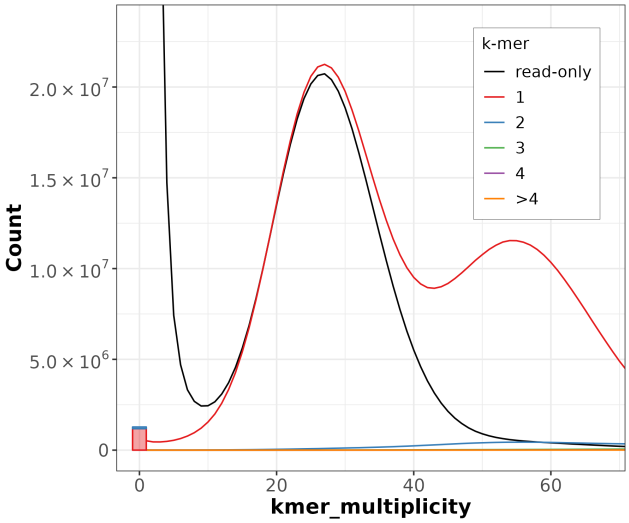
Warning! BUSCO versions or lineage datasets are not the same across results:
BUSCO: 5.8.2 (euk_genome_met, metaeuk) / Lineage: mollusca_odb12 (genomes:36, BUSCOs:4421)
BUSCO: 6.0.0 (euk_genome_min, miniprot) / Lineage: mollusca_odb12 (genomes:36, BUSCOs:4421)

HiC contact map of curated assembly

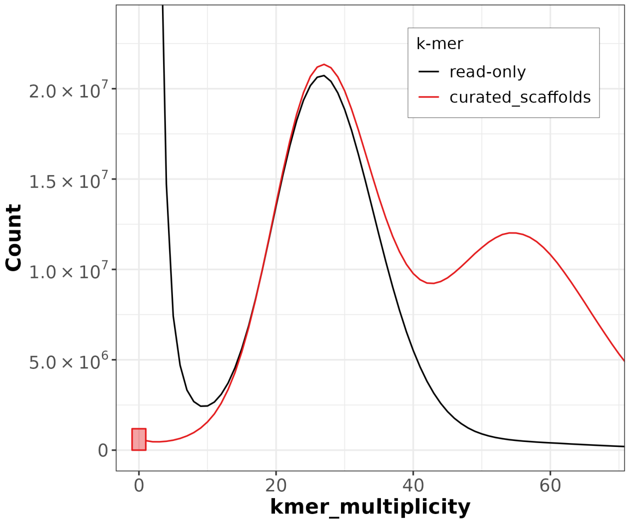


collapsed [LINK]

K-mer spectra of curated assembly

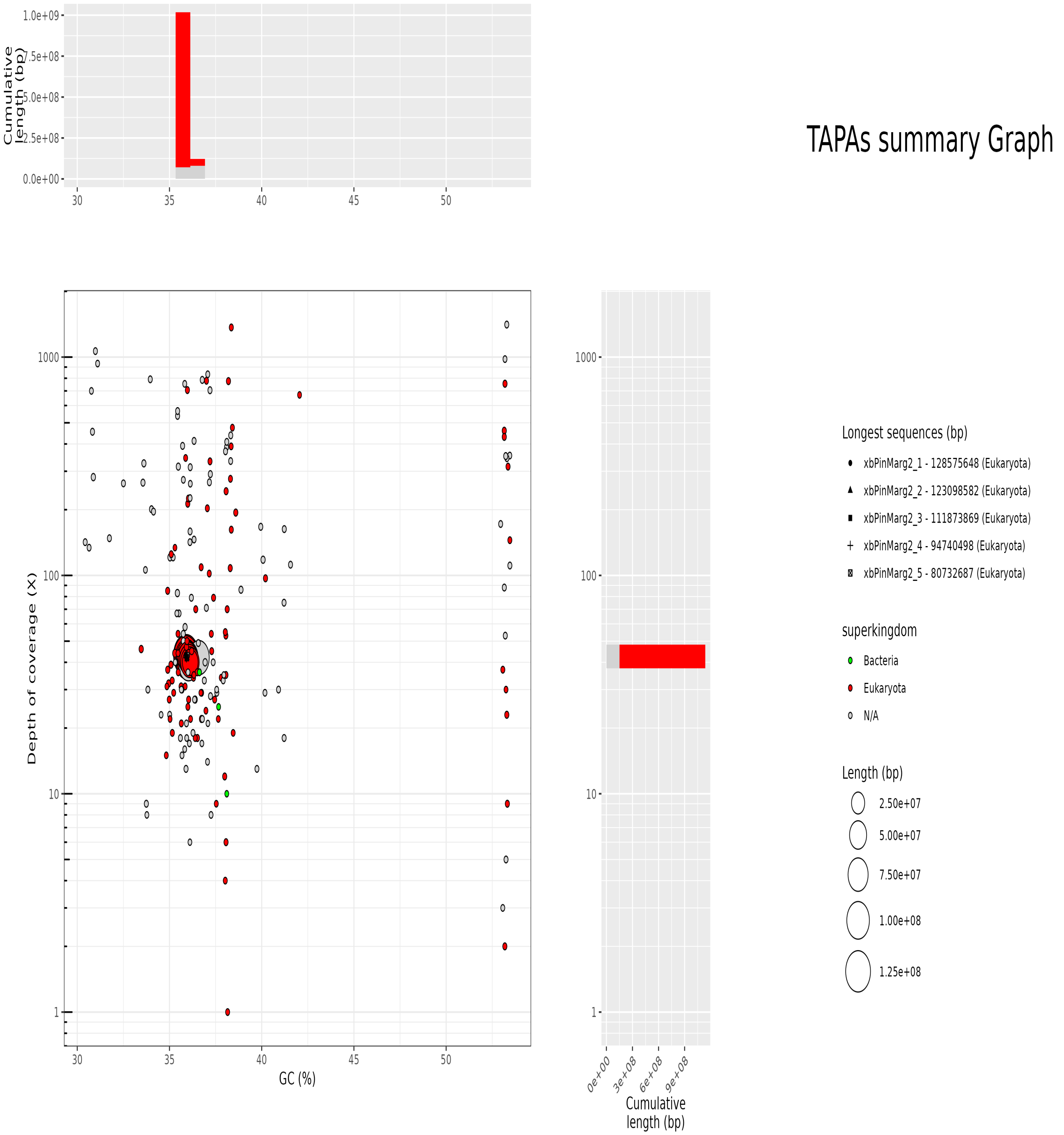


Distribution of k-mer counts per copy numbers found in asm



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	45	214

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: 2025-11-20 02:16:37 CET