ERGA Assembly Report

v24.04.03_beta

Tags: ERGA-BGE

TxID	112209
ToLID	**xbPinRadi1**
Species	Pinctada radiata
Class	Bivalvia
Order	Pterioida

Genome Traits	Expected	Observed
Haploid size (bp)	866,069,963	950,922,884
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.Q60

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: 94
. Contamination notes: "A few bacterial contigs were removed post-assembly"
. Other observations: "A large heterozygous rearrangement was found and left intact on SUPER_2"

Quality metrics table

| Metrics	Pre-curation collapsed	Curated collapsed
Total bp	1,005,790,220	950,922,884
GC %	35.44	35.43
Gaps/Gbp	158.08	171.41
Total gap bp	15,900	19,400
Scaffolds	73	70
Scaffold N50	63,972,673	63,751,225
Scaffold L50	7	6
Scaffold L90	14	13
Contigs	232	233
Contig N50	8,276,876	8,181,964
Contig L50	36	34
Contig L90	114	110
QV	60.8643	60.8162
Kmer compl.	67.0763	64.454
BUSCO sing.	94.5%	96.5%
BUSCO dupl.	3.5%	1.6%
BUSCO frag.	2.0%	2.0%
BUSCO miss.	0.0%	0.1%

BUSCO 5.4.3 Lineage: eukaryota_odb10 (genomes:70, BUSCOs:255)
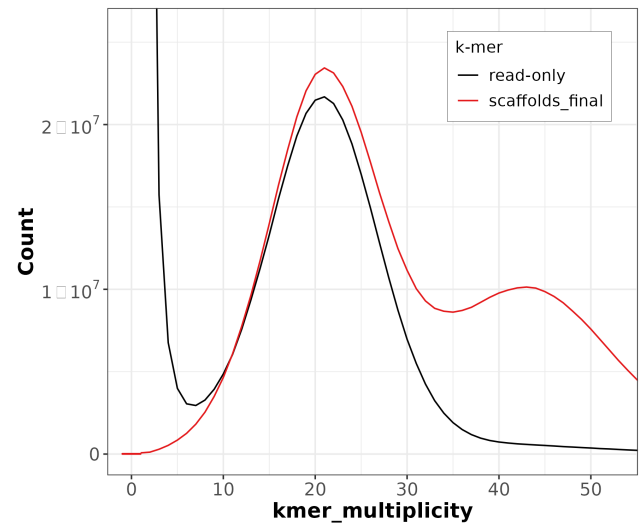
HiC contact map of curated assembly
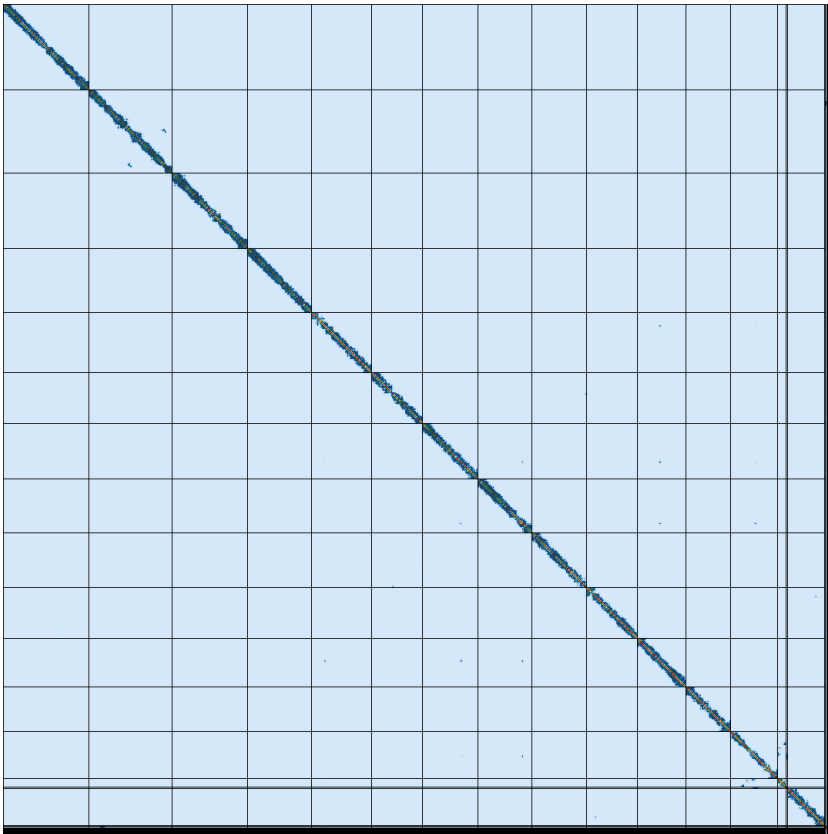
collapsed [LINK]
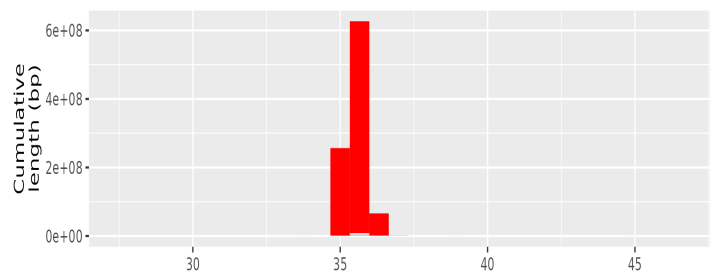
K-mer spectra of curated assembly
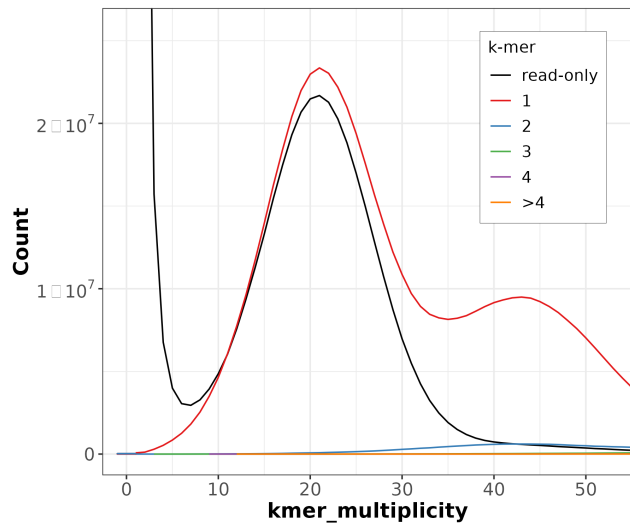
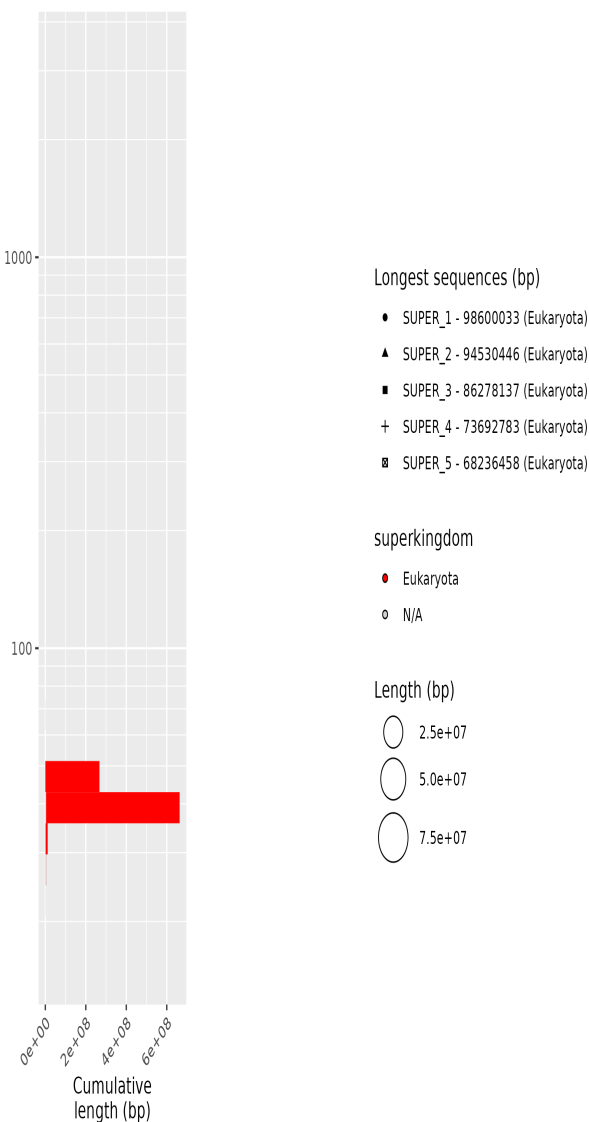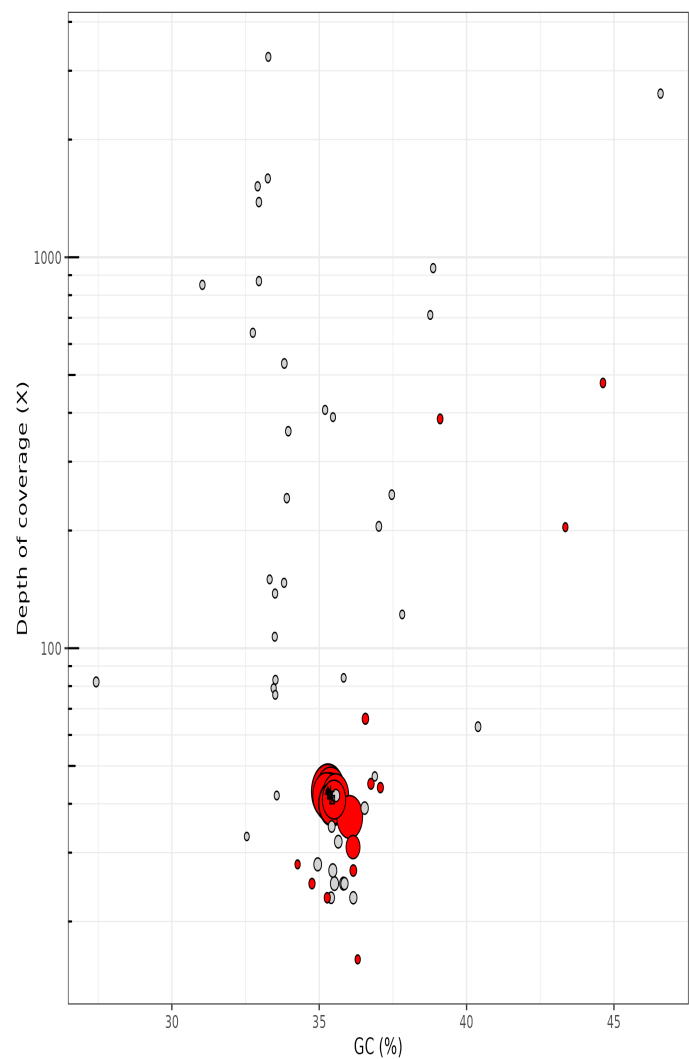Distribution of k-mer counts per copy numbers found in asm

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

Post-curation contamination screening

TAPAs summary Graph

Longest sequences (bp)

● SUPER_1 - 98600033 (Eukaryota)

▲ SUPER_2 - 94530446 (Eukaryota)

■ SUPER_3 - 86278137 (Eukaryota)

+ SUPER_4 - 73692783 (Eukaryota)

⊠ SUPER_5 - 68236458 (Eukaryota)

superkingdom

● Eukaryota

○ N/A

Length (bp)

○ 2.5e+07

○ 5.0e+07

○ 7.5e+07

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	PACBIO Hifi	HiC (Arima (4-enzymes))
Coverage	42	54

Assembly pipeline

- **Hifiasm**
    |_ *ver:* 0.19.5-r593
    |_ *key param:* NA
- **purge_dups**
    |_ *ver:* 1.2.5
    |_ *key param:* NA

Curation 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

Submitter: Benjamin Istace
Affiliation: Genoscope

Date and time: 2024-07-11 08:56:34 CEST