

ERGA Assembly Report

v24.04.03_beta

Tags: ERGA-BGE

TxID	44597
ToLID	xbArcNoae1
Species	Arca noae
Class	Bivalvia
Order	Arcoida

Genome Traits	Expected	Observed
Haploid size (bp)	1,387,861,013	1,496,024,100
Haploid Number	14 (source: ancestor)	19
Ploidy	2 (source: ancestor)	2
Sample Sex	Unknown	unknown

EBP metrics summary and curation notes

Obtained EBP quality metric for collapsed: 7.7.Q59

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

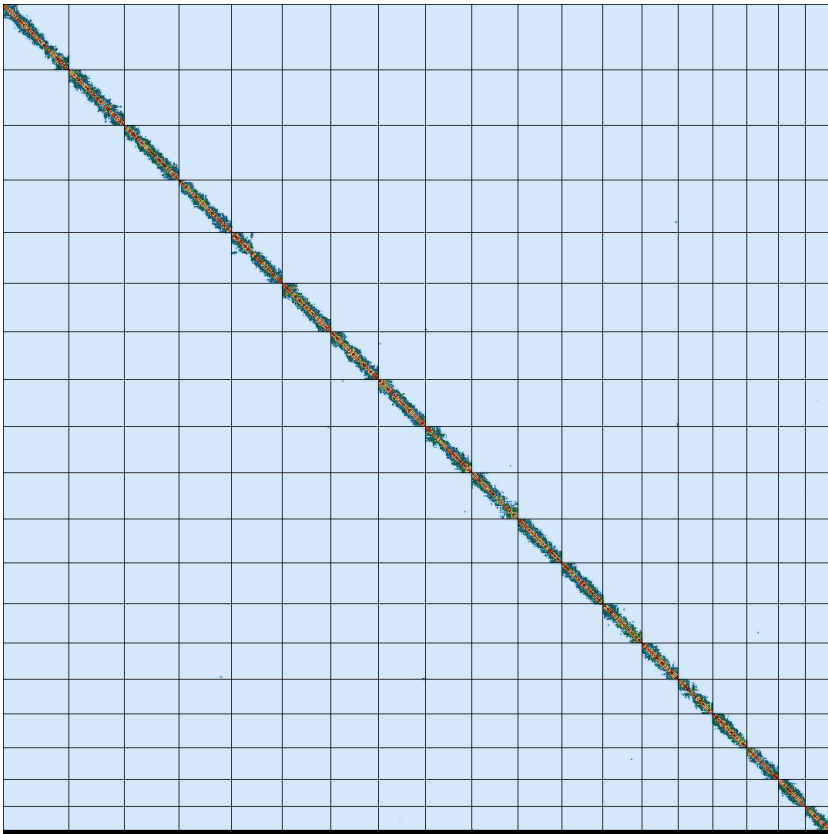
- . Interventions/Gb: 101
- . Contamination notes: "30 bacterian contigs removed after assembly "
- . Other observations: "Numerous haplotypic duplications removed during manual curation. Inversions between haplotypes are seen in the following locations: SUPER_2 from 63.89 to 75.1 Mbp, SUPER_6 from 6.87 to 15.0 Mbp, SUPER_9 from 12.28 to 26.61 Mbp, SUPER_17 from 5.52 to 13.81 Mbp, SUPER_19 from beginning to 13.30 Mbp "

Quality metrics table

Metrics	Pre-curation collapsed	Curated collapsed
Total bp	1,540,111,515	1,496,024,100
GC %	33.57	33.57
Gaps/Gbp	101.94	92.24
Total gap bp	15,700	16,000
Scaffolds	135	119
Scaffold N50	85,310,971	84,746,044
Scaffold L50	9	8
Scaffold L90	17	17
Contigs	292	257
Contig N50	17,667,724	20,465,118
Contig L50	24	22
Contig L90	87	74
QV	59.4982	59.5106
Kmer compl.	63.2552	61.9488
BUSCO sing.	96.9%	98.4%
BUSCO dupl.	2.4%	0.8%
BUSCO frag.	0.8%	0.8%
BUSCO miss.	0.1%	0.0%

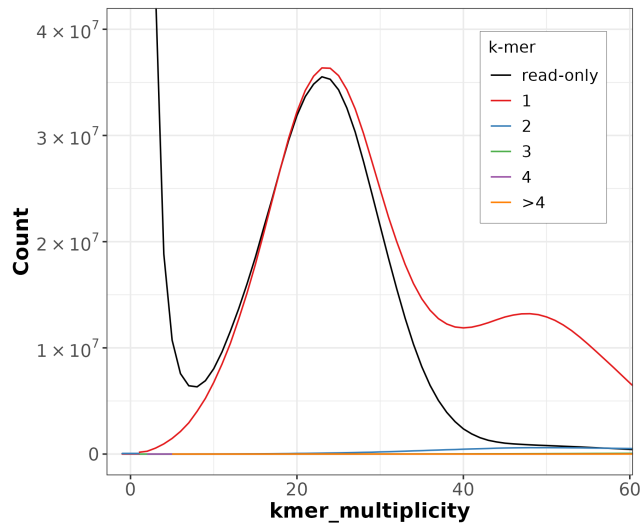
BUSCO 5.4.3 Lineage: eukaryota_odb10 (genomes:70, BUSCOs:255)

HiC contact map of curated assembly

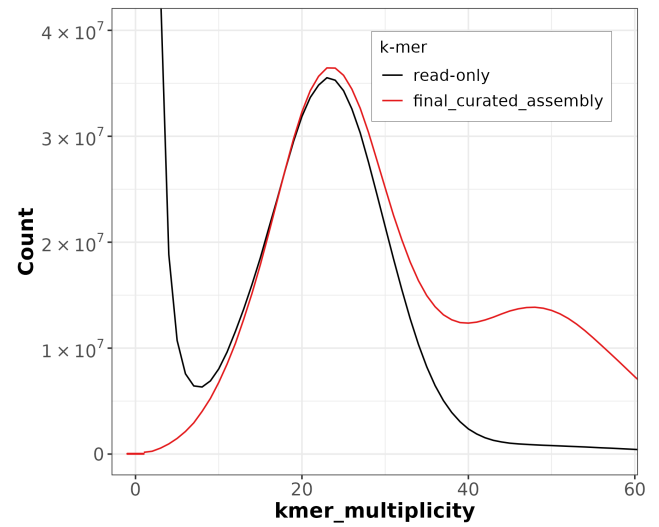


collapsed [\[LINK\]](#)

K-mer spectra of curated assembly

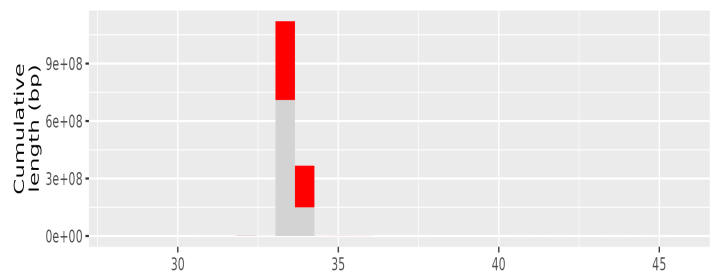


Distribution of k-mer counts per copy numbers found in asm

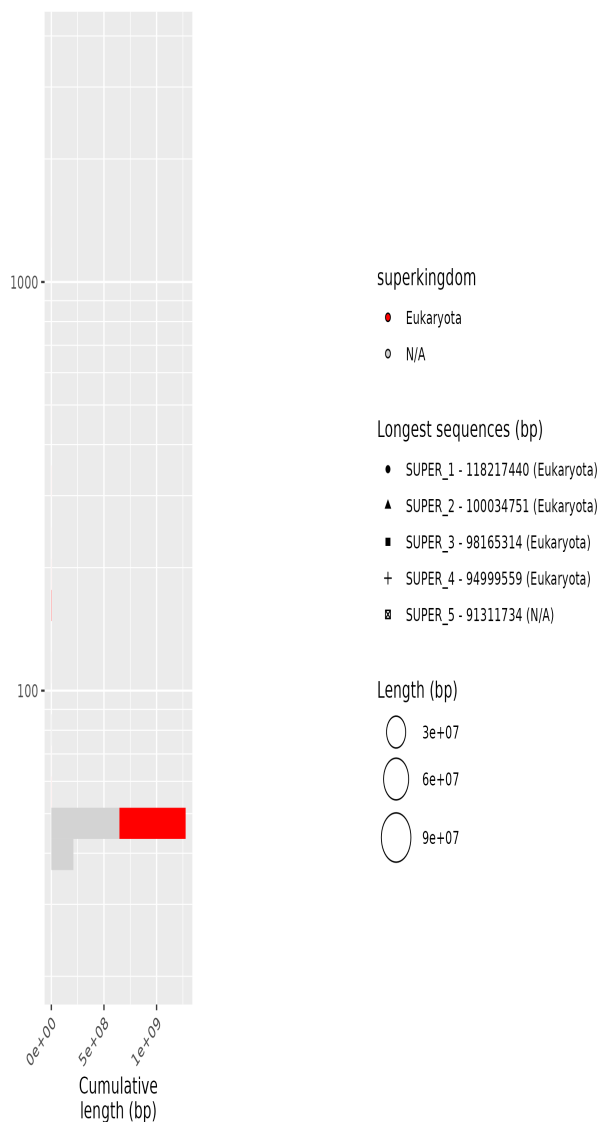
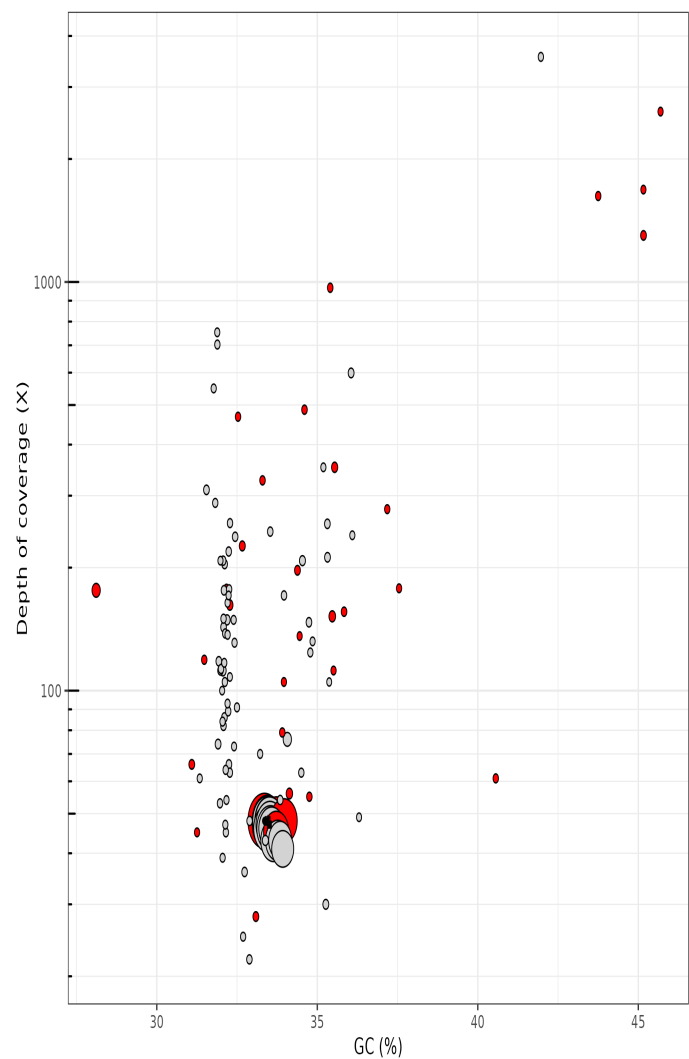


Distribution of k-mer counts coloured by their presence in reads/assemblies

Post-curation contamination screening



TAPAs summary Graph



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	Omnic
Coverage	48	24

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: Emilie Teodori

Affiliation: Genoscope

Date and time: 2024-09-02 09:04:39 CEST