ERGA Assembly Report

v24.10.15

Tags: ATLASea[INVALID TAG]

TxID	399303
ToLID	**xbAbrAlba5**
Species	Abra alba
Class	Bivalvia
Order	Cardiida

Genome Traits	Expected	Observed
Haploid size (bp)	1,343,012,087	1,376,633,745
Haploid Number	26 (source: ancestor)	19
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
. Assembly length loss > 3% for collapsed

Curator notes

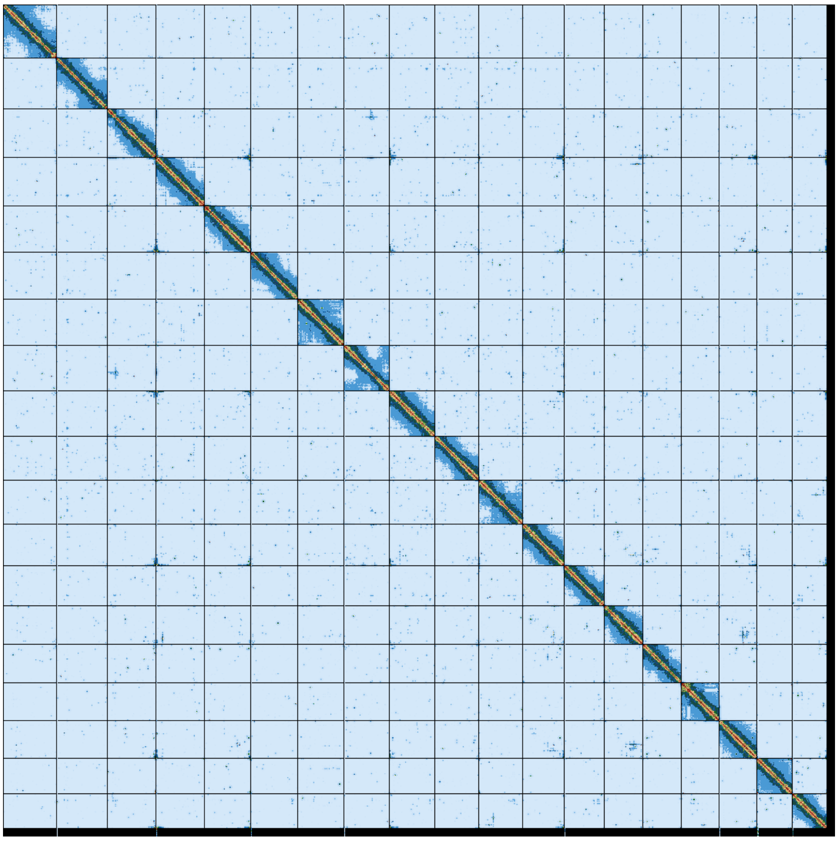
. Interventions/Gb: 268
. Contamination notes: ""
. Other observations: "The assembly of Abra alba (xbAbrAlba5.1) is based on 60X PacBio data and 300X Arima Hi-C data generated as part of the ATLASea programme (https://www.atlasea.fr). The assembly process included the following steps: initial PacBio+Hi-C assembly generation with Hifiasm, removal of contaminant sequences using Context, and Hi-C-based scaffolding with YaHS of each haplotype. In total, 54 contigs were identified as contaminants (bacterial), totaling 5.34 Mb (with the largest being 2.132 Mb). The mitochondrial genome was assembled using OATK. Both haplotypes were analyzed and manually improved using Pretext. During manual curation, 20 haplotypic regions were removed, totaling 38.55 Mb (with the largest being 6.793 Mb). 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	1,963,317,447	1,376,633,745
GC %	38	38.06
Gaps/Gbp	412.06	453.28
Total gap bp	93,700	73,600
Scaffolds	136	279
Scaffold N50	82,551,375	75,392,811
Scaffold L50	11	9
Scaffold L90	25	17
Contigs	898	903
Contig N50	4,229,873	3,868,558
Contig L50	130	109
Contig L90	433	367
QV	48.3719	48.0326
Kmer compl.	78.6339	59.2856
BUSCO sing.	59.8%	97.7%
BUSCO dupl.	39.0%	1.1%
BUSCO frag.	0.6%	0.6%
BUSCO miss.	0.6%	0.6%

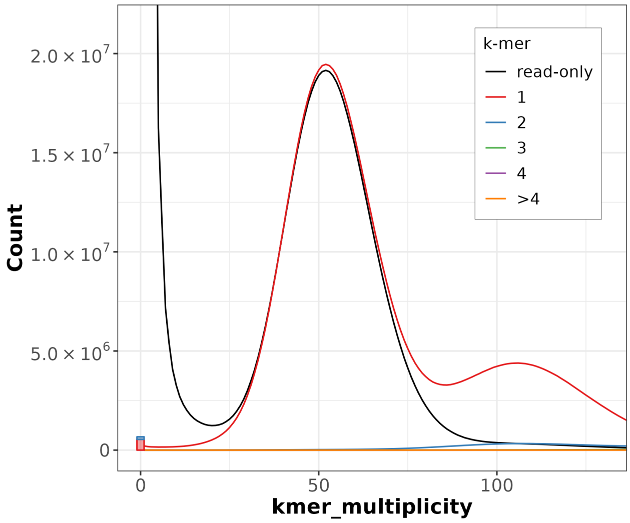
BUSCO: 6.0.0 (euk_genome_min, miniprot) / Lineage: mollusca_odb12 (genomes:36, BUSCOs:4421)

HiC contact map of curated assembly

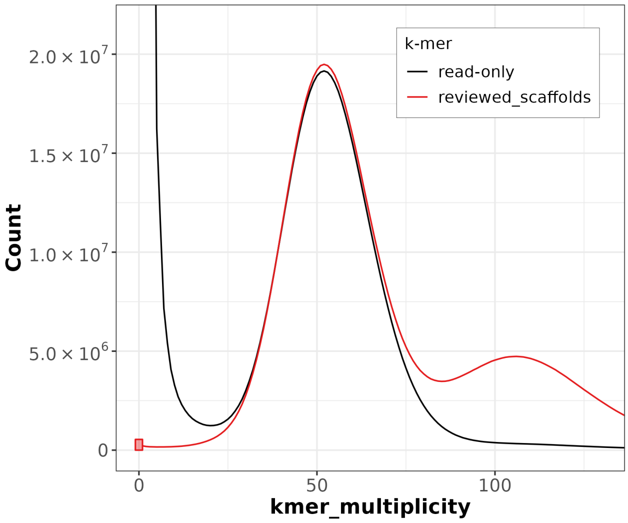


collapsed [LINK]

K-mer spectra of curated assembly

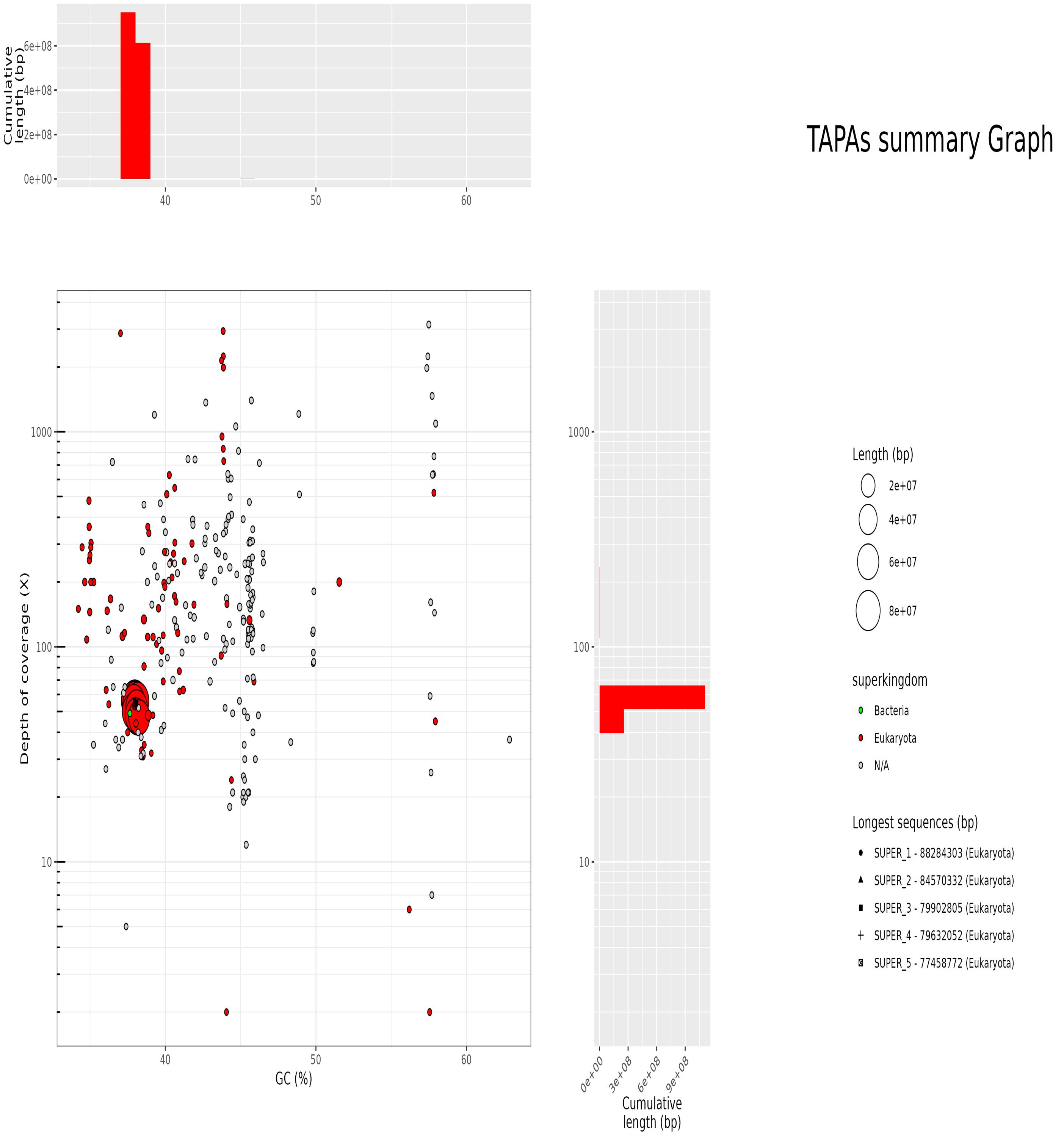


Distribution of k-mer counts per copy numbers found in asm



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	60	301

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: Adama Ndar
Affiliation: Genoscope

Date and time: 2025-12-21 21:39:16 CET