ERGA Assembly Report
v24.10.15

Tags: ATLASea[INVALID TAG]

TxID	36103
ToLID	**xgHalTube1**
Species	Haliotis tuberculata
Class	Gastropoda
Order	Lepetellida

Genome Traits	Expected	Observed
Haploid size (bp)	1,243,856,902	1,357,172,450
Haploid Number	9 (source: ancestor)	14
Ploidy	2 (source: ancestor)	2
Sample Sex	Unknown	Unknown

EBP metrics summary and curation notes

Obtained EBP quality metric for collapsed: 7.7.Q64

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
. BUSCO single copy value is less than 90% for collapsed

Curator notes

. Interventions/Gb: 79
. Contamination notes: ""
. Other observations: "The assembly of Haliotis tuberculata (xgHalTube1.1) is based on 59X of PacBio data and 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, 43 contigs were identified as contaminants (bacterial, archaeal, or viral), totaling 7.7 Mb (with the largest being 1.2 Mb). Additionally, 245 regions totaling 125 Mb (with the largest being 7.5 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, 23 haplotypic regions and 2 contaminant sequence were removed, totaling 16 Mb and 0.2 Mb, respectively (with the largest being 4.2 Mb and 0.1 Mb). Alignment of the Haliotis tuberculata assembly on a reference (Haliotis asinina) shows that scaffold_15 and SUPER_4 align on the same chromosome of the reference. On the HiC contact map, the beginning of scaffold_15

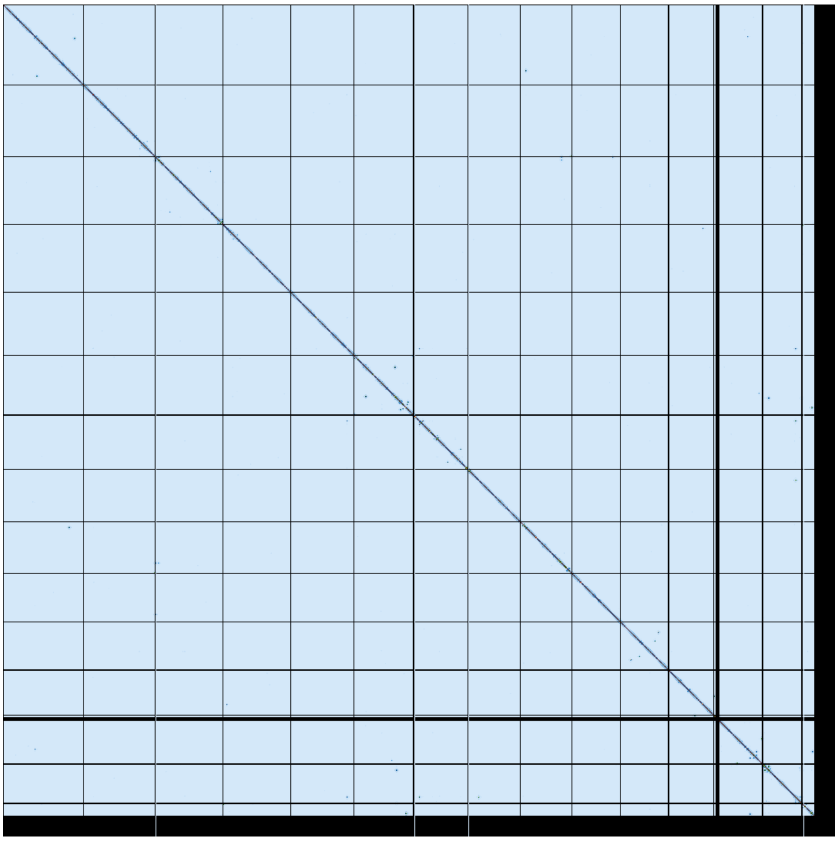
appears to have contacts with SUPER_4, but there is no gap, making it difficult to make a decision. Chromosome-scale scaffolds confirmed by Hi-C data were named in order of size. "

Quality metrics table

Metrics	Pre-curation collapsed	Curated collapsed
Total bp	1,366,964,971	1,357,172,450
GC %	41.45	41.45
Gaps/Gbp	155.82	162.84
Total gap bp	21,300	24,000
Scaffolds	383	355
Scaffold N50	84,607,011	88,135,511
Scaffold L50	7	7
Scaffold L90	14	13
Contigs	596	576
Contig N50	23,232,335	23,232,335
Contig L50	21	21
Contig L90	68	66
QV	45.8477	64.2684
Kmer compl.	67.8464	67.6197
BUSCO sing.	85.0%	85.3%
BUSCO dupl.	1.6%	1.3%
BUSCO frag.	4.3%	4.3%
BUSCO miss.	9.1%	9.1%

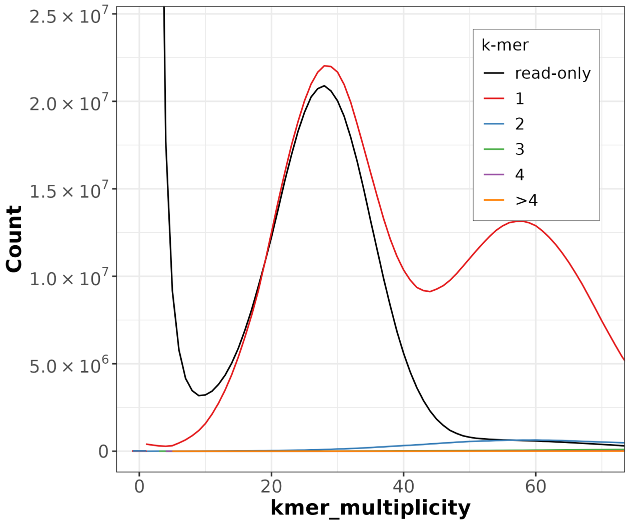
BUSCO: 5.4.3 (euk_genome_met, metaeuk) / Lineage: mollusca_odb10 (genomes:7, BUSCOs:5295)

HiC contact map of curated assembly

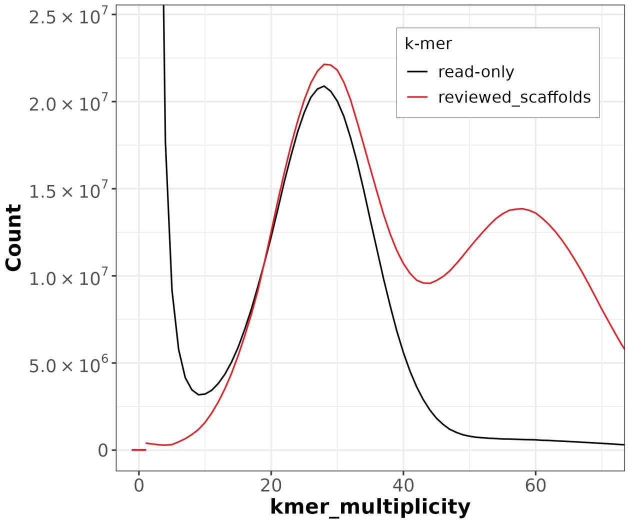


collapsed [LINK]

K-mer spectra of curated assembly

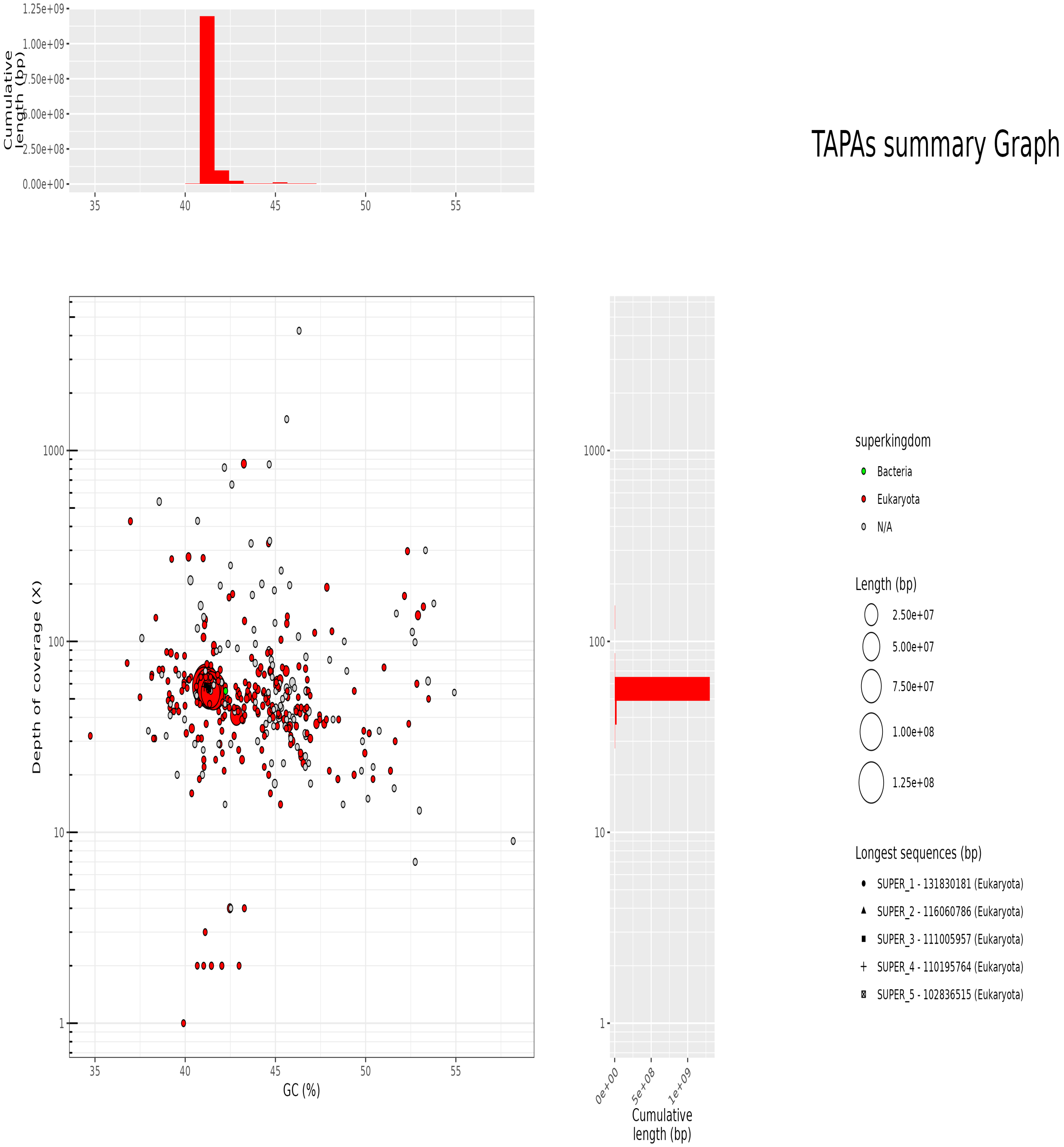


Distribution of k-mer counts per copy numbers found in asm



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	PACBIO Hifi	Arima
Coverage	58	68

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: 2025-02-27 14:41:49 CET