ERGA Assembly Report

v24.10.15

Tags: ATLASea[INVALID TAG]

TxID	1048598
ToLID	**xbCteMiti1**
Species	Ctenoides mitis
Class	Bivalvia
Order	Limoida

Genome Traits	Expected	Observed
Haploid size (bp)	1,392,391,047	1,422,448,557
Haploid Number	14 (source: ancestor)	18
Ploidy	2 (source: ancestor)	2
Sample Sex	Unknown	Unknown

EBP metrics summary and curation notes

Obtained EBP quality metric for collapsed: 7.7.Q50

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

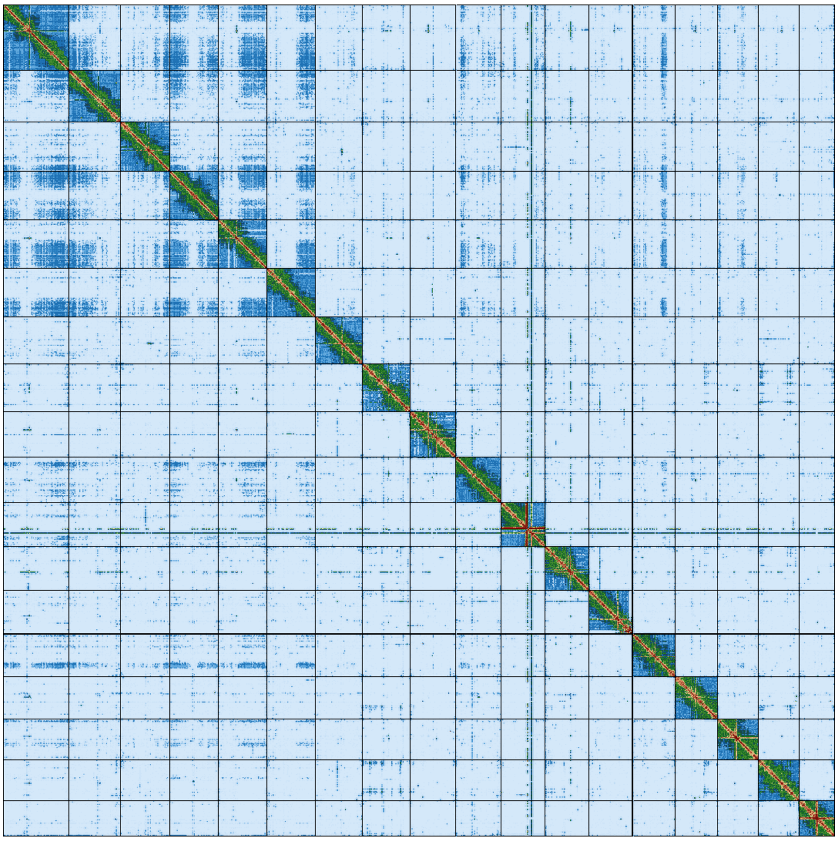
. Interventions/Gb: 6
. Contamination notes: ""
. Other observations: "The assembly of Ctenoides mitis (xbCteMiti1) is based on 48X PacBio data and 167X Arima Hi-C data generated as part of the ATLASea programme (https://www.atlasea.fr). The assembly process included the following steps: initial PacBio 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, 2 contigs were identified as contaminants (bacterial, archaeal, or viral), totaling 0.092 Mb (with the largest being 0.048 Mb). Additionally, 125 regions totaling 51.533 Mb (with the largest being 14.644 Mb) were identified as haplotypic duplications and removed. During manual curation, 3 haplotypic regions were removed, totaling 7.94Mb (with the largest being 3.97Mb). 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	1,430,353,598	1,422,448,557
GC %	35.81	35.81
Gaps/Gbp	15.38	18.28
Total gap bp	2,200	3,100
Scaffolds	37	33
Scaffold N50	77,752,964	77,752,964
Scaffold L50	9	9
Scaffold L90	17	16
Contigs	59	59
Contig N50	49,475,702	49,475,702
Contig L50	10	10
Contig L90	27	27
QV	50.0442	50.0428
Kmer compl.	68.4517	68.2712
BUSCO sing.	97.1%	97.6%
BUSCO dupl.	1.3%	0.7%
BUSCO frag.	0.5%	0.5%
BUSCO miss.	1.1%	1.1%

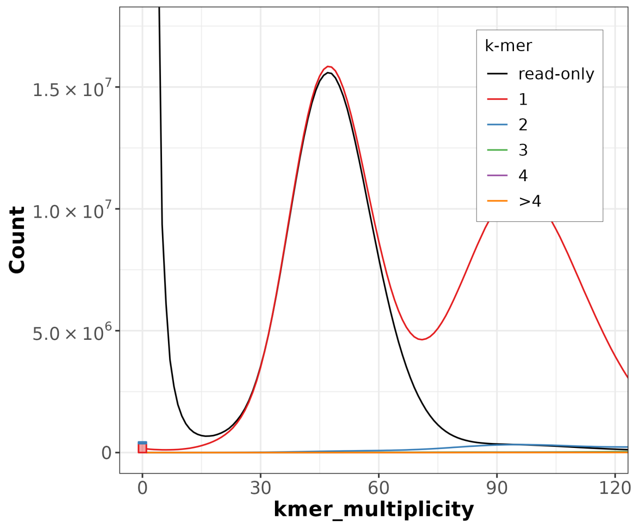
BUSCO: 6.0.0 (euk_genome_min, miniprot) / Lineage: mollusca_odb12 (genomes:36, BUSCOs:4421)

HiC contact map of curated assembly

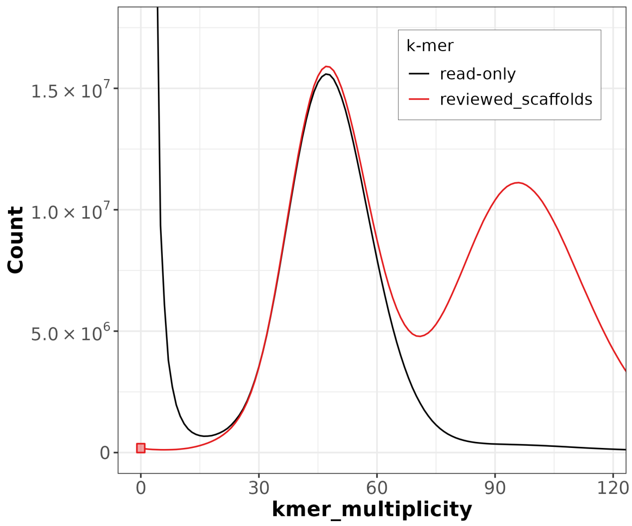


collapsed [LINK]

K-mer spectra of curated assembly

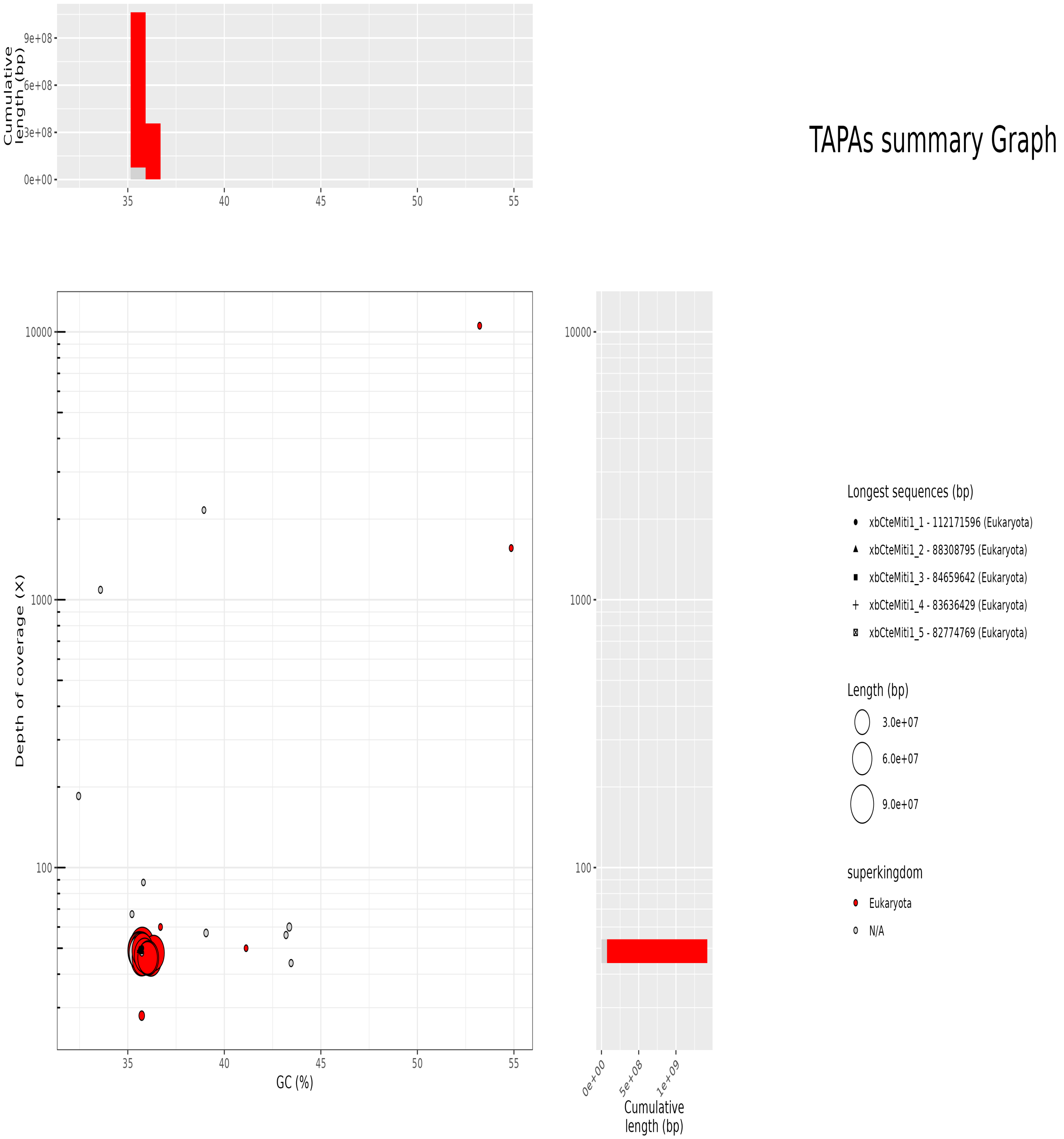


Distribution of k-mer counts per copy numbers found in asm



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	48	167

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: Emilie Teodori
Affiliation: Genoscope

Date and time: 2025-12-14 14:34:33 CET