ERGA Assembly Report

v24.10.15

Tags: ATLASea[INVALID TAG]

TxID	529747
ToLID	**xgLitTube1**
Species	Lithopoma tuber
Class	Gastropoda
Order	Trochida

Genome Traits	Expected	Observed
Haploid size (bp)	1,281,585,116	1,282,231,546
Haploid Number	9 (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

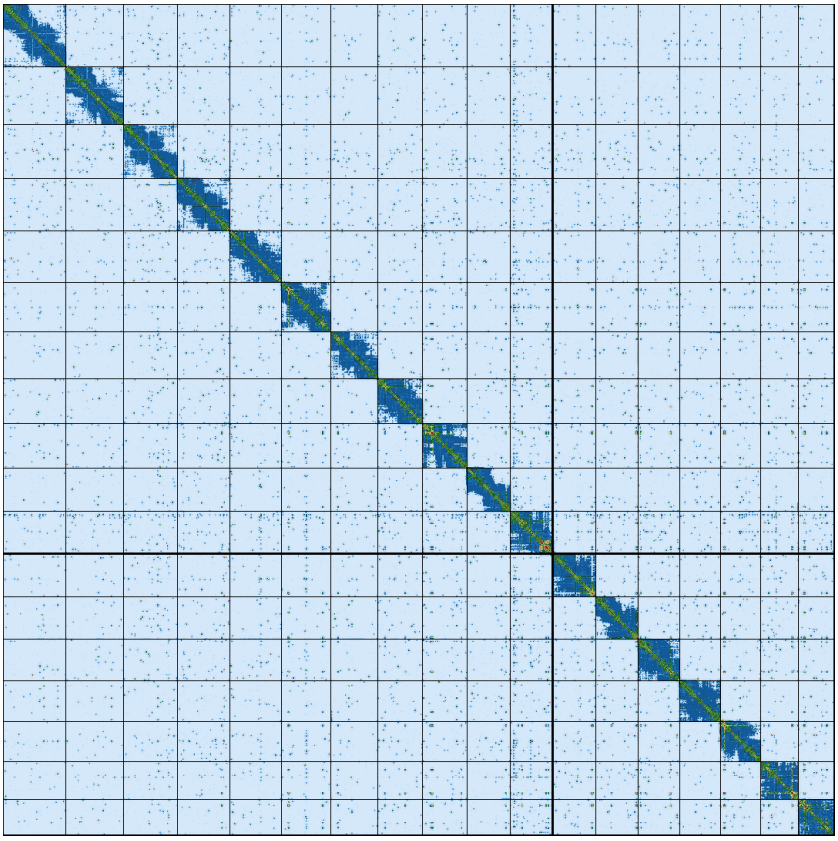
. Interventions/Gb: 37
. Contamination notes: ""
. Other observations: "The assembly of Lithopoma tuber (xgLitTube1) is based on 44X PacBio data and 186X Arima Hi-C data generated as part of the ATLASea programme (https://www.atlasea.fr). The assembly process included the following steps: initial PacBio 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 3.802 Mb (with the largest being 0.824 Mb). Additionally, 159 regions totaling 20.323 Mb (with the largest being 2.996 Mb) were identified as haplotypic duplications and removed. Mitochondrial genome was assembled using OATK. Finally, the primary assembly was analyzed and manually improved using Pretext. During manual curation, 4 haplotypic regions and 4 contaminant sequences were removed, totaling 6.878Mb and 0.151Mb, respectively (with the largest being 2.498Mb and 0.096Mb). The telomeric pattern TTAGGG was identified with TelFinder and used to generate the Pretext Telomeres tracks. 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	1,289,293,219	1,282,231,546
GC %	38.18	38.19
Gaps/Gbp	62.05	76.43
Total gap bp	8,000	12,800
Scaffolds	59	40
Scaffold N50	69,675,821	69,773,121
Scaffold L50	8	8
Scaffold L90	17	16
Contigs	139	138
Contig N50	24,100,166	24,278,289
Contig L50	17	16
Contig L90	58	57
QV	50.656	50.7001
Kmer compl.	77.0821	76.9769
BUSCO sing.	98.6%	99.2%
BUSCO dupl.	0.9%	0.4%
BUSCO frag.	0.0%	0.0%
BUSCO miss.	0.4%	0.4%

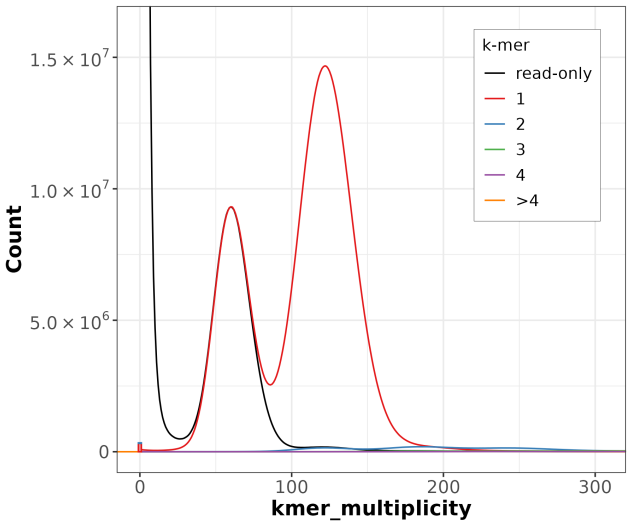
BUSCO: 6.1.0 (euk_genome_min, miniprot) / Lineage: mollusca_odb12.2 (genomes:31, BUSCOs:4236)

HiC contact map of curated assembly

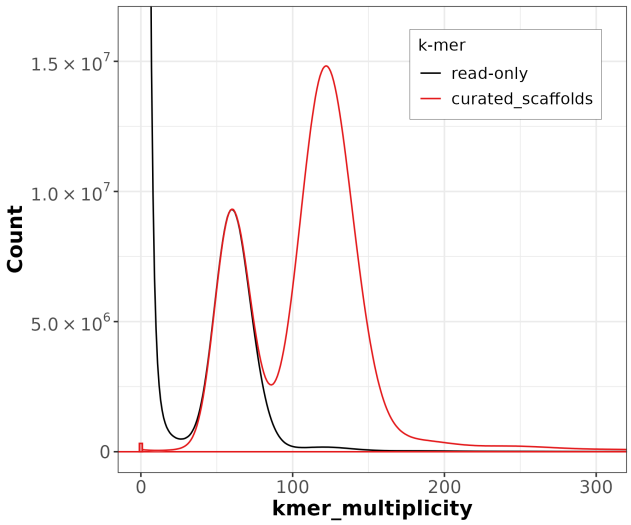


collapsed [LINK]

K-mer spectra of curated assembly

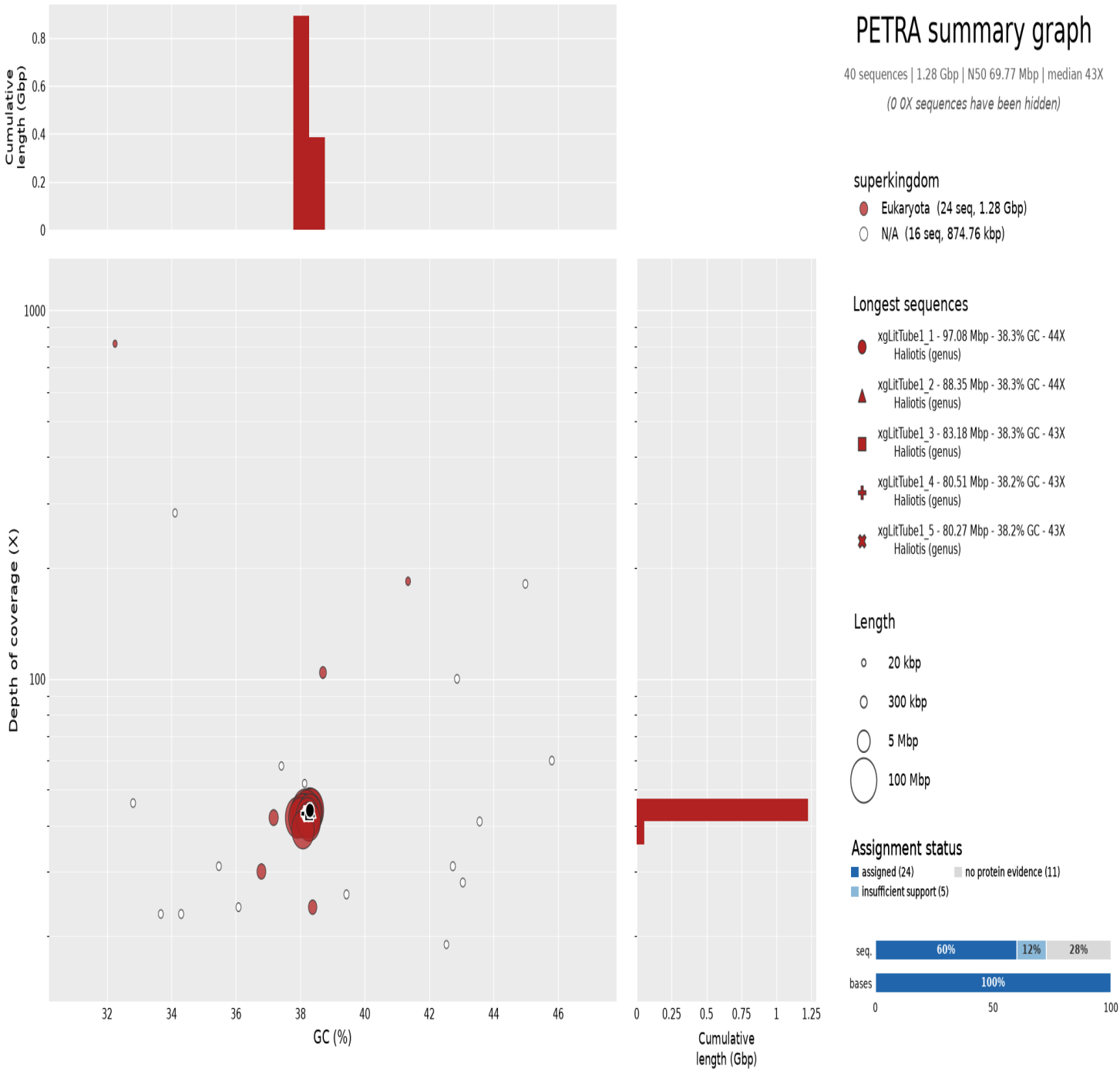


Distribution of k-mer counts per copy numbers found in asm



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	45	187

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:* 0.2.4
  |_ *key param:* NA
- **PretextView**
  |_ *ver:* 1.0.7
  |_ *key param:* NA

Submitter: Sophie Layac
Affiliation: Genoscope

Date and time: 2026-09-27 21:31:47 CEST