ERGA Assembly Report

v24.10.15

Tags: ATLASea[INVALID TAG]

TxID	120367
ToLID	**xbBarBarb1**
Species	Barbatia barbata
Class	Bivalvia
Order	Arcoida

Genome Traits	Expected	Observed
Haploid size (bp)	1,446,271,143	1,465,975,088
Haploid Number	14 (source: ancestor)	20
Ploidy	2 (source: ancestor)	2
Sample Sex	Unknown	Unknown

EBP metrics summary and curation notes

Obtained EBP quality metric for collapsed: 7.7.Q47

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

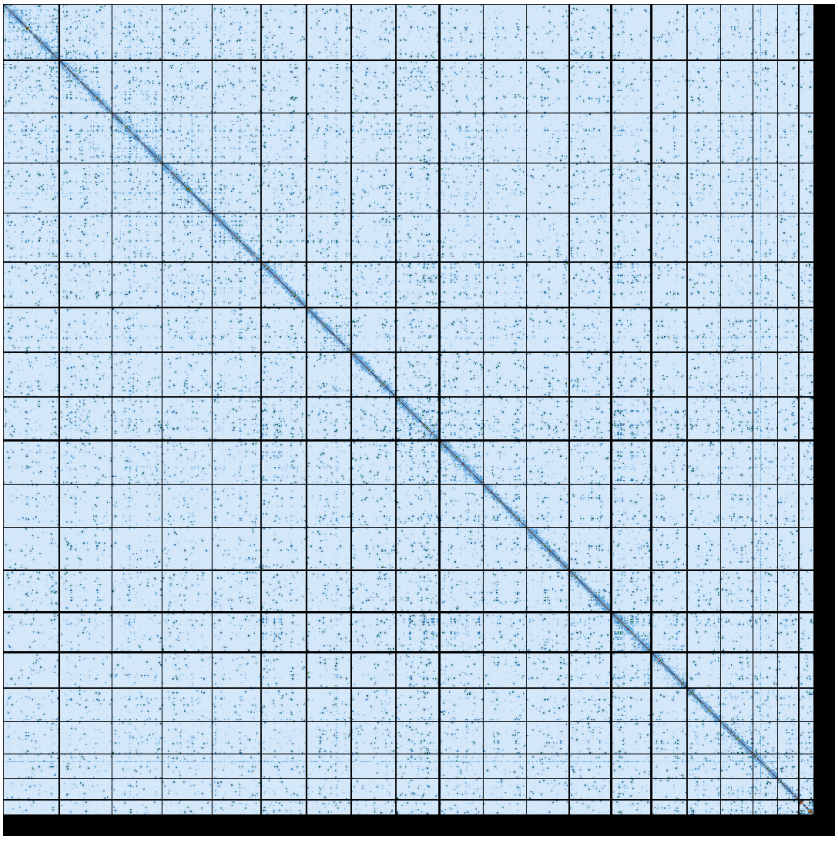
. Interventions/Gb: 3
. Contamination notes: ""
. Other observations: "The assembly of Barbatia barbata (xbBarBarb1) is based on 54.17X PacBio data and 139.72X Arima Hi-C data generated as part of the ATLASea programme (https://www.atlasea.fr). The assembly process included the following steps: initial 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, 8 contigs were identified as contaminants (bacterial, archaeal, or viral), totaling 0.227 Mb (with the largest being 0.053 Mb). Additionally, 364 regions totaling 40.163 Mb (with the largest being 6.23 Mb) were identified as haplotypic duplications and removed. Finally, the primary assembly was analyzed and manually improved using Pretext. During manual curation, 1 haplotypic region was removed, totaling 1.56Mb. The telomeric pattern AACCCT 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	1,491,396,325	1,465,975,088
GC %	33.6	33.6
Gaps/Gbp	82.47	82.54
Total gap bp	12,300	14,100
Scaffolds	352	342
Scaffold N50	78,840,100	76,914,500
Scaffold L50	8	9
Scaffold L90	17	18
Contigs	475	463
Contig N50	22,005,000	22,005,000
Contig L50	22	22
Contig L90	70	68
QV	47.3963	47.4135
Kmer compl.	69.991	69.4328
BUSCO sing.	95.7%	97.4%
BUSCO dupl.	2.3%	0.7%
BUSCO frag.	1.0%	1.0%
BUSCO miss.	0.9%	0.9%

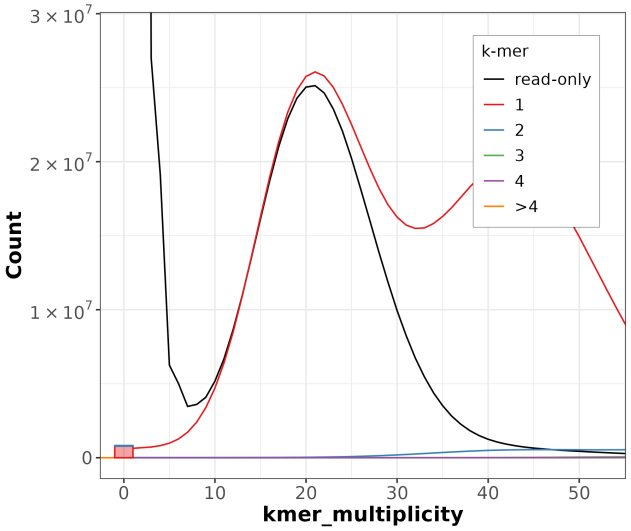
Warning! BUSCO versions or lineage datasets are not the same across results:
BUSCO: 6.0.0 (euk_genome_min, miniprot) / Lineage: mollusca_odb12 (genomes:36, BUSCOs:4421)
BUSCO: 6.1.0 (euk_genome_min, miniprot) / Lineage: mollusca_odb12 (genomes:36, BUSCOs:4421)

HiC contact map of curated assembly

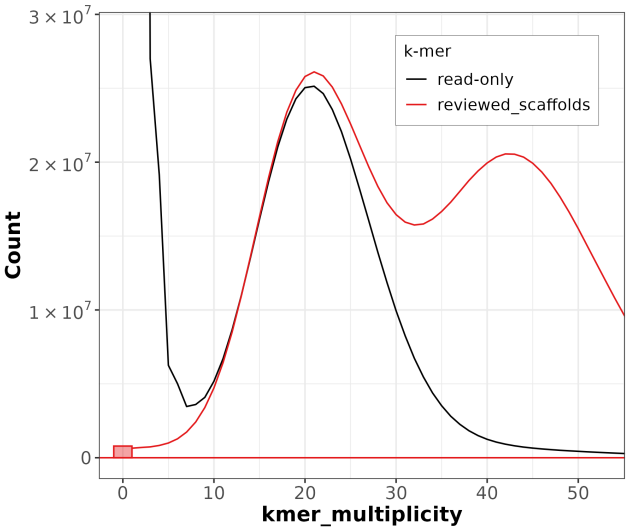


collapsed [LINK]

K-mer spectra of curated assembly

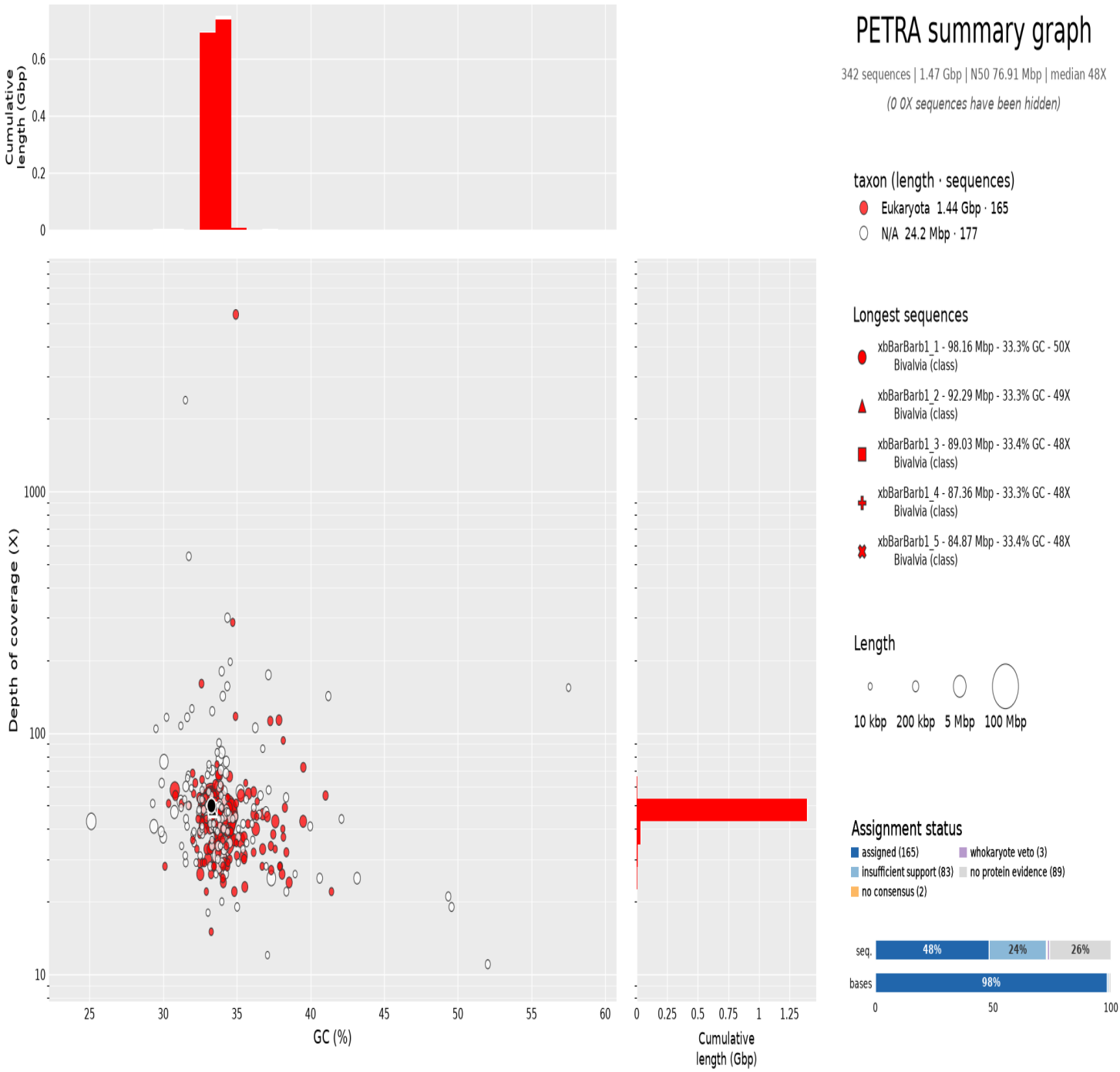


Distribution of k-mer counts per copy numbers found in asm



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	55	142

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: Eleanore Lacoste
Affiliation: Genoscope

Date and time: 2026-10-07 11:08:08 CEST