ERGA Assembly Report

v24.10.15

Tags: ATLASea[INVALID TAG]

TxID	3190879
ToLID	**xbGloRigi1**
Species	Globivenus rigida
Class	Bivalvia
Order	Venerida

Genome Traits	Expected	Observed
Haploid size (bp)	1,774,444,677	1,800,706,013
Haploid Number	19 (source: ancestor)	19
Ploidy	2 (source: ancestor)	2
Sample Sex	Unknown	Unknown

EBP metrics summary and curation notes

Obtained EBP quality metric for collapsed: 7.7.Q49

The following metrics were automatically flagged as below EBP recommended standards or different from expected:

. Kmer completeness value is less than 90 for collapsed

Curator notes

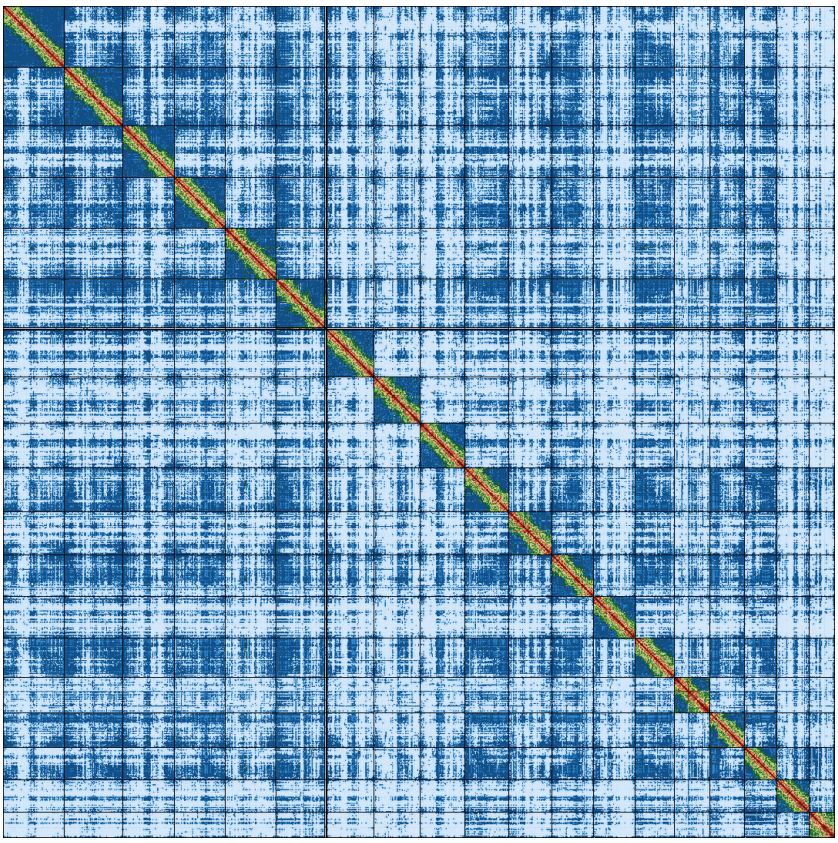
. Interventions/Gb: 18
. Contamination notes: ""
. Other observations: "The assembly of Globivenus rigida (xbGloRigi1) is based on 43X PacBio data and 213X Arima Hi-C data generated as part of the ATLASea programme (https://www.atlasea.fr). The assembly process included the following steps: initial PacBio 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, 68 contigs were identified as contaminants (bacterial, archaeal, or viral), totaling 4.784 Mb (with the largest being 0.31 Mb). Additionally, 380 regions totaling 29.561 Mb (with the largest being 3.854 Mb) were identified as haplotypic duplications and removed. Mitochondrial genome was assembled using OATK and, surprisingly, has a size of 40 kb, which is larger than the ~20 kb typically observed. Finally, the primary assembly was analyzed and manually improved using Pretext. During manual curation, 16 haplotypic regions and 1 contaminant sequence were removed, totaling 36.749Mb and 0.06Mb, respectively (with the largest being 10.283Mb and 0.06Mb). 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	1,837,472,160	1,800,706,013
GC %	34.54	34.53
Gaps/Gbp	60.41	62.2
Total gap bp	11,100	12,400
Scaffolds	67	60
Scaffold N50	98,313,306	96,291,983
Scaffold L50	9	9
Scaffold L90	17	17
Contigs	178	172
Contig N50	26,422,797	26,367,000
Contig L50	22	22
Contig L90	70	68
QV	49.1032	49.129
Kmer compl.	71.3224	70.646
BUSCO sing.	96.0%	97.9%
BUSCO dupl.	3.1%	1.1%
BUSCO frag.	0.4%	0.4%
BUSCO miss.	0.5%	0.6%

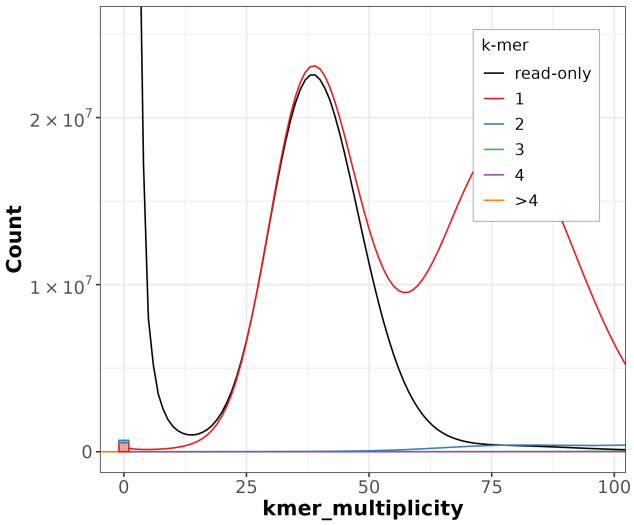
BUSCO: 6.0.0 (euk_genome_min, miniprot) / Lineage: mollusca_odb12 (genomes:36, BUSCOs:4421)

HiC contact map of curated assembly

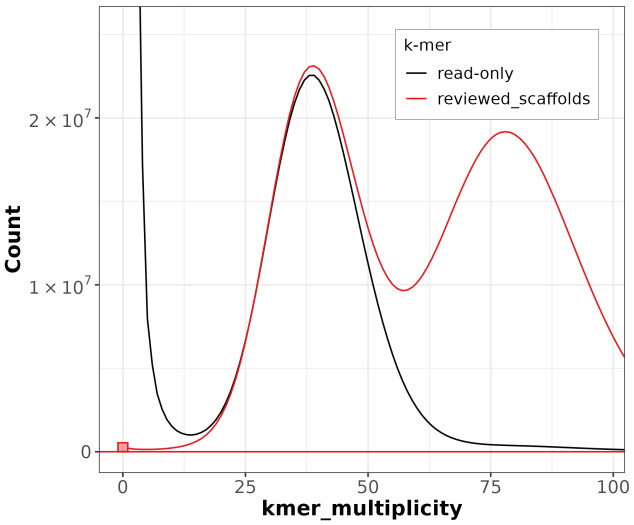


collapsed [LINK]

K-mer spectra of curated assembly

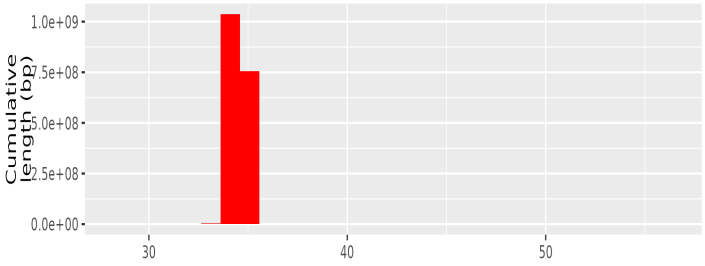


Distribution of k-mer counts per copy numbers found in asm

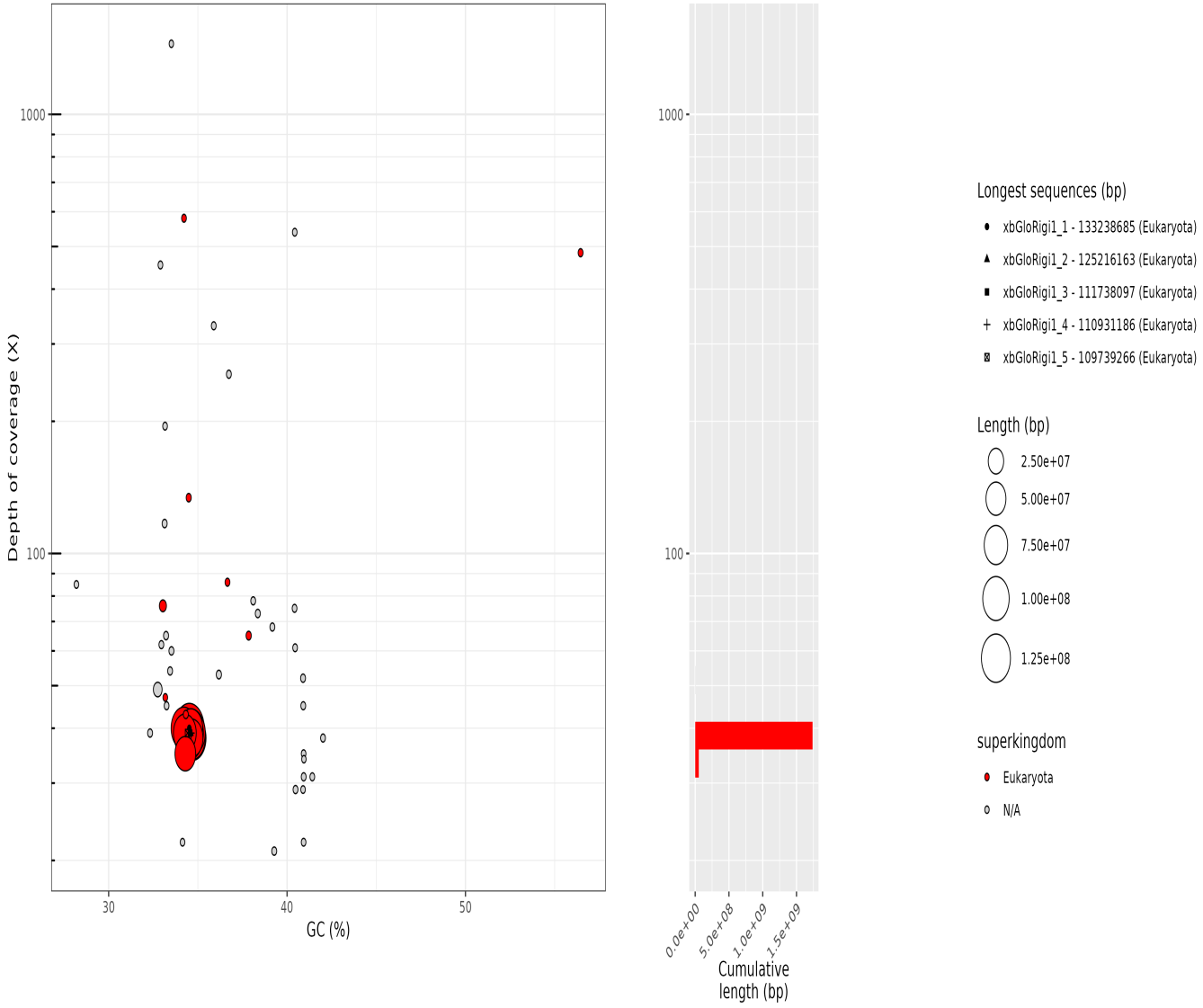


Distribution of k-mer counts coloured by their presence in reads/assemblies

Post-curation contamination screening



TAPAs summary Graph



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	44	217

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: Jean-Marc Aury
Affiliation: Genoscope

Date and time: 2026-04-23 09:18:48 CEST