

ERGA Assembly Report

v24.10.15

Tags: ATLASea[INVALID TAG]

TxID	335776
ToLID	xgBolRugol
Species	Bolma rugosa
Class	Gastropoda
Order	Trochida

Genome Traits	Expected	Observed
Haploid size (bp)	2,966,988,005	3,034,741,950
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: 7.8.Q63

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

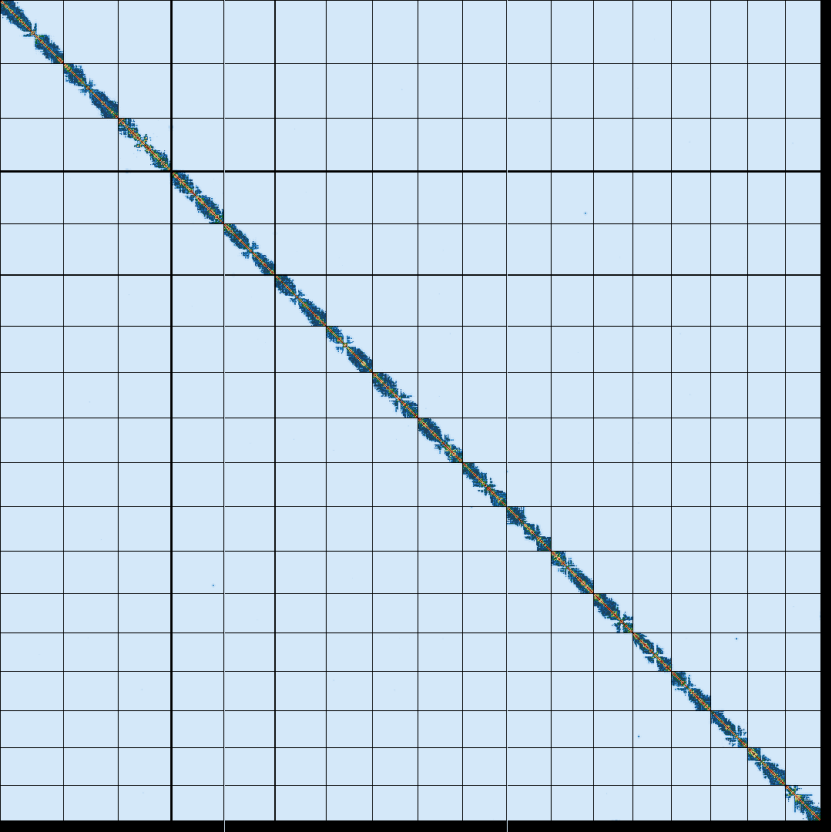
- . Interventions/Gb: 18
- . Contamination notes: "96 bacterial sequences of 8.9 Mb were detected and removed before scaffolding. The post-curation contamination screening plot corresponds to the assembly before bacterial sequences are removed. "
- . Other observations: "15 contigs of 10.8 Mb were tagged as haplotigs. "

Quality metrics table

Metrics	Pre-curation collapsed	Curated collapsed
Total bp	3,045,660,594	3,034,741,950
GC %	39.61	39.61
Gaps/Gbp	94.56	97.21
Total gap bp	28,800	31,800
Scaffolds	167	161
Scaffold N50	162,738,200	166,155,781
Scaffold L50	9	8
Scaffold L90	17	17
Contigs	455	456
Contig N50	18,070,829	18,070,829
Contig L50	51	51
Contig L90	176	178
QV	46.825	63.0868
Kmer compl.	75.3821	75.4297
BUSCO sing.	84.9%	85.0%
BUSCO dupl.	1.5%	1.4%
BUSCO frag.	4.1%	4.1%
BUSCO miss.	9.5%	9.5%

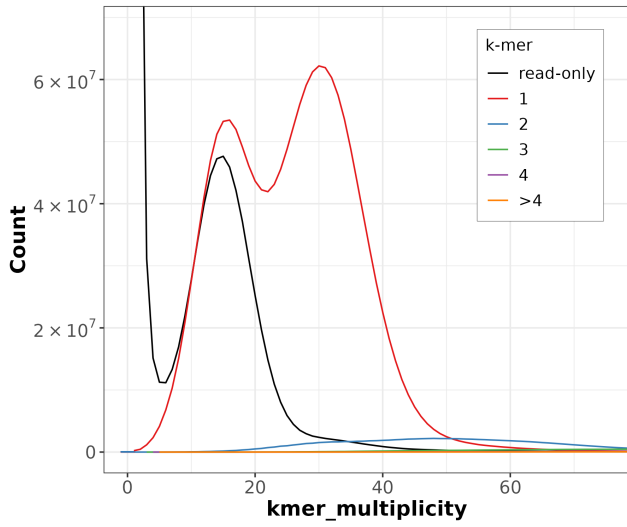
BUSCO: 5.4.3 (euk_genome_met, metaeuk) / Lineage: mollusca_odb10 (genomes:7, BUSCOs:5295)

HiC contact map of curated assembly

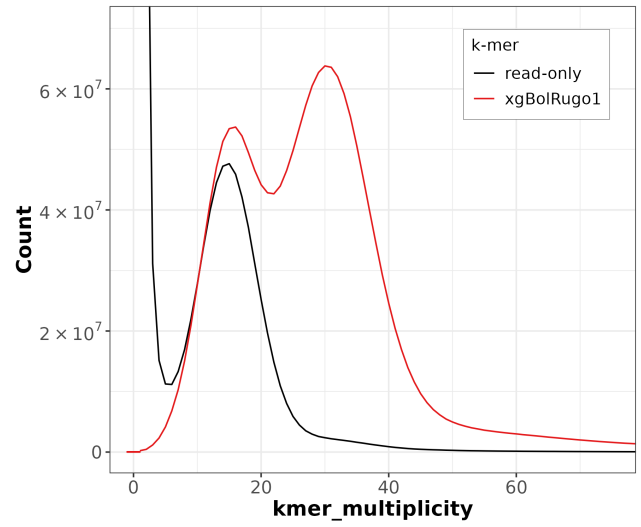


collapsed [\[LINK\]](#)

K-mer spectra of curated assembly

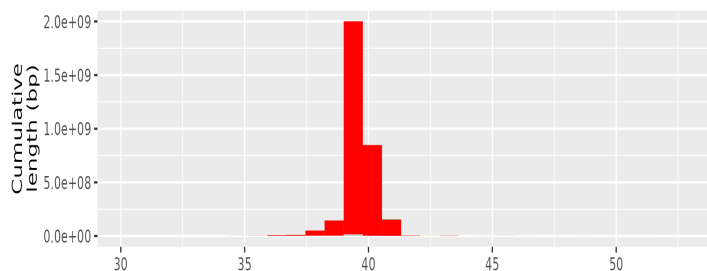


Distribution of k-mer counts per copy numbers found in asm

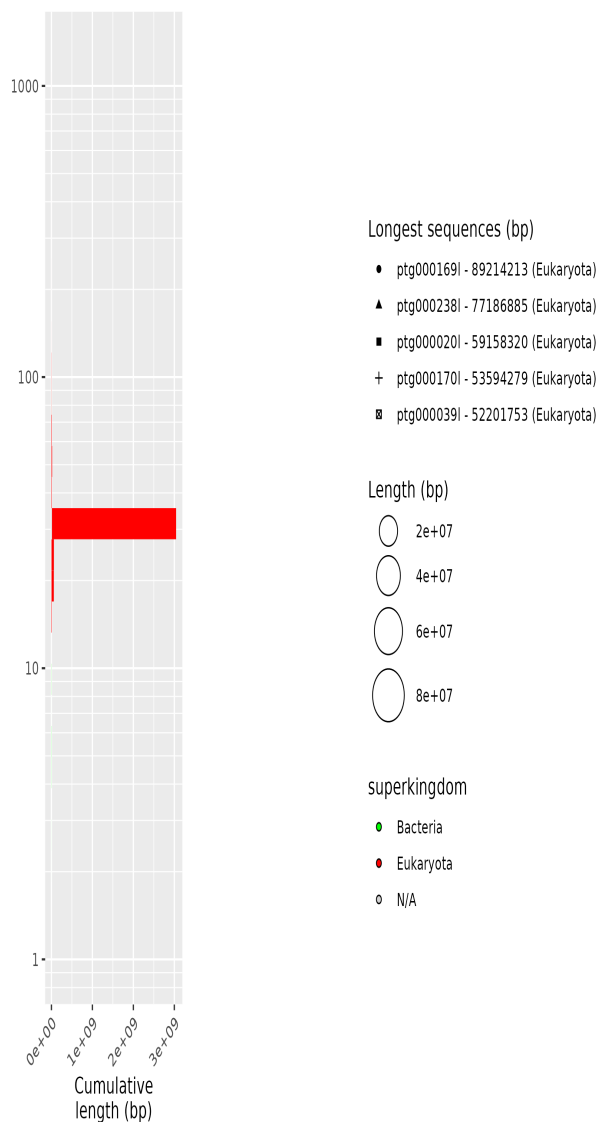
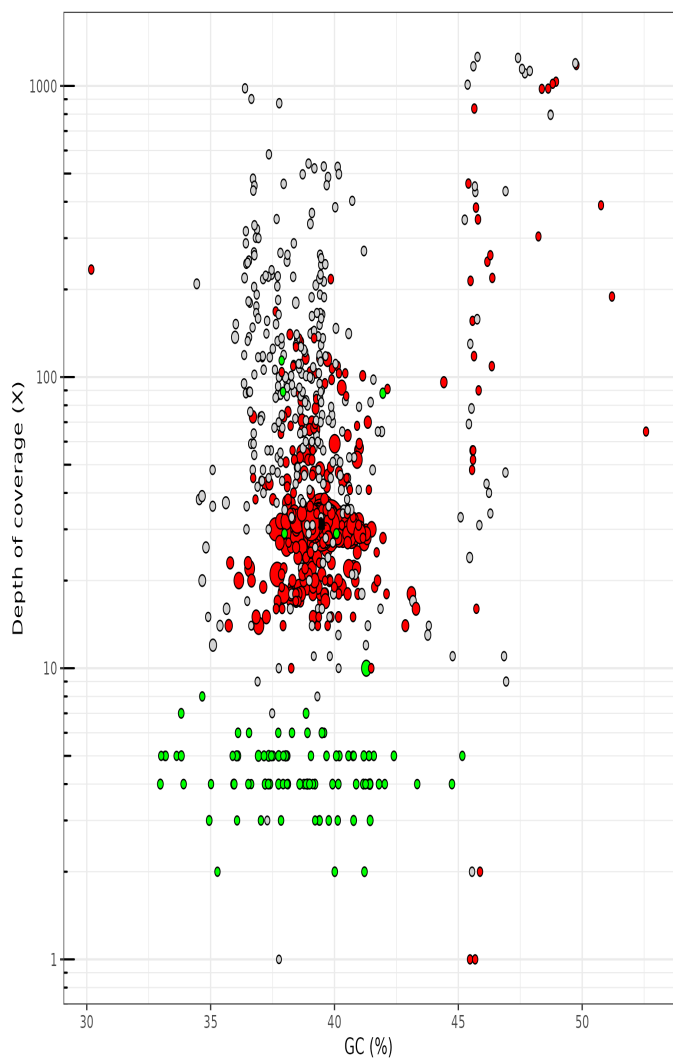


Distribution of k-mer counts coloured by their presence in reads/assemblies

Post-curation contamination screening



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	PACBIO Hifi	Arima
Coverage	32	30

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: 2024-12-06 01:48:05 CET