ERGA Assembly Report

v24.10.15

Tags: ATLASea[INVALID TAG]

TxID	2010383
ToLID	**xgBolHeni1**
Species	Bolma henica
Class	Gastropoda
Order	Trochida

Genome Traits	Expected	Observed
Haploid size (bp)	3,180,826,705	3,342,504,185
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: 6.8.Q48

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

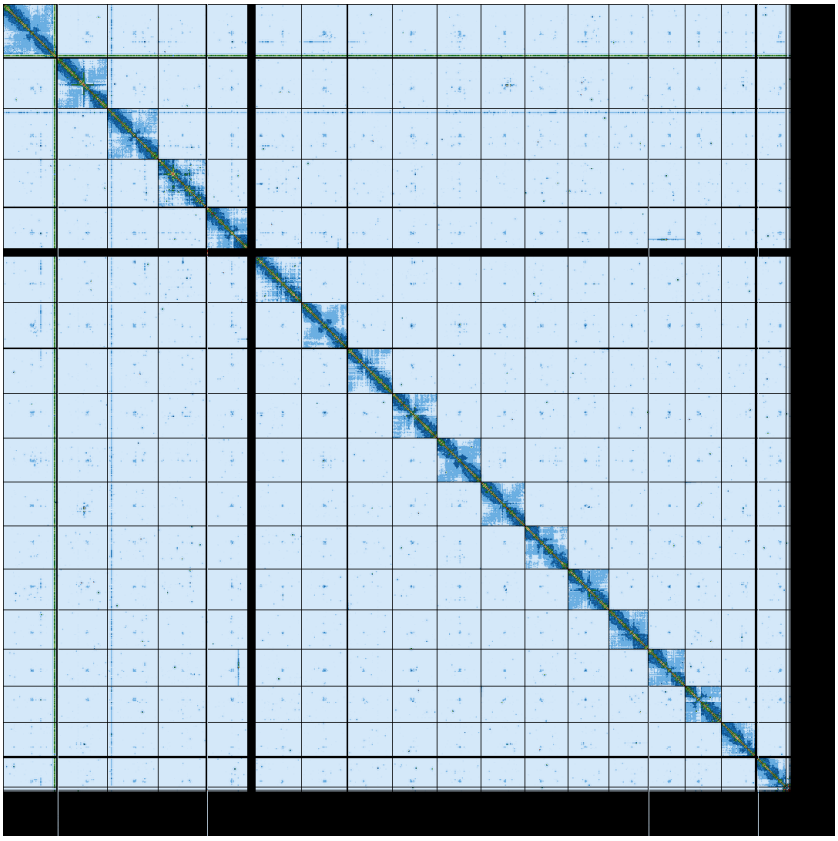
. Interventions/Gb: 37
. Contamination notes: ""
. Other observations: "The assembly of Bolma henica (xgBolHeni1) is based on 41X PacBio data and 160X Arima Hi-C data generated as part of the ATLASea programme (https://www.atlasea.fr). The assembly process included the following steps: initial PacBio 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, 88 contigs were identified as contaminants (bacterial, archaeal, or viral), totaling 9.251 Mb (with the largest being 1.814 Mb). Additionally, 1716 regions totaling 215.511 Mb (with the largest being 1.175 Mb) were identified as haplotypic duplications and removed. Finally, the primary assembly was analyzed and manually improved using Pretext. During manual curation, 4 haplotypic regions were removed, totaling 5.5Mb (with the largest being 2.781Mb). 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	3,355,145,426	3,342,504,185
GC %	38.66	38.66
Gaps/Gbp	550.5	558.86
Total gap bp	184,700	194,000
Scaffolds	979	932
Scaffold N50	174,323,871	177,261,904
Scaffold L50	9	9
Scaffold L90	18	18
Contigs	2,826	2,800
Contig N50	3,382,450	3,381,697
Contig L50	292	291
Contig L90	1,068	1,059
QV	48.1036	48.104
Kmer compl.	77.0919	76.9977
BUSCO sing.	95.8%	96.2%
BUSCO dupl.	3.7%	3.3%
BUSCO frag.	0.2%	0.2%
BUSCO miss.	0.3%	0.3%

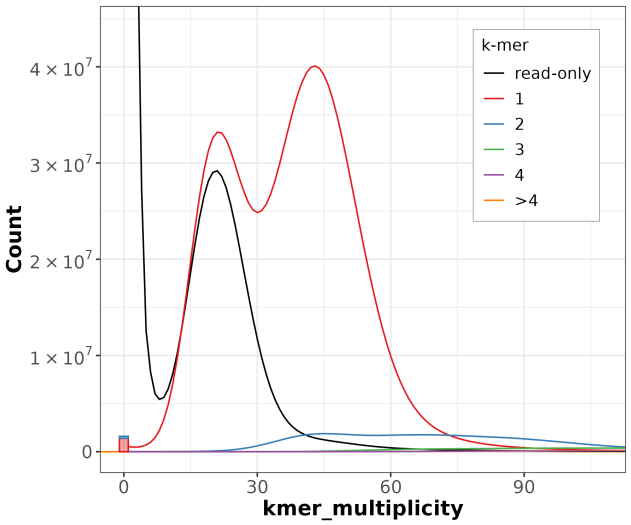
BUSCO: 6.0.0 (euk_genome_min, miniprot) / Lineage: mollusca_odb12 (genomes:36, BUSCOs:4421)

HiC contact map of curated assembly

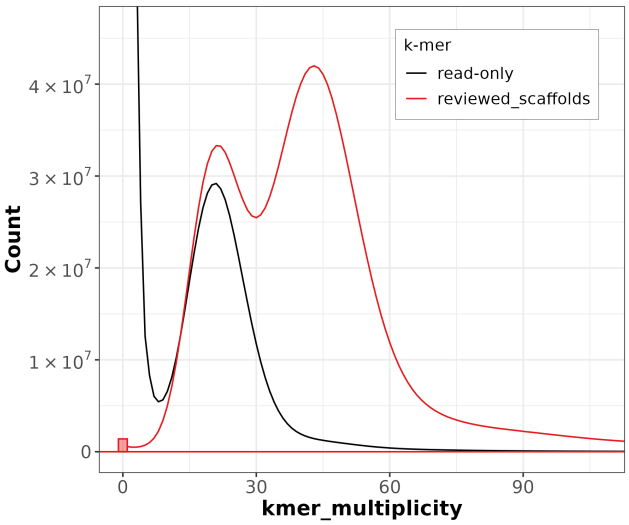


collapsed [LINK]

K-mer spectra of curated assembly

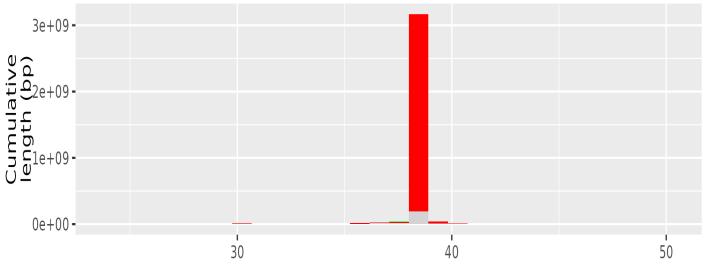


Distribution of k-mer counts per copy numbers found in asm

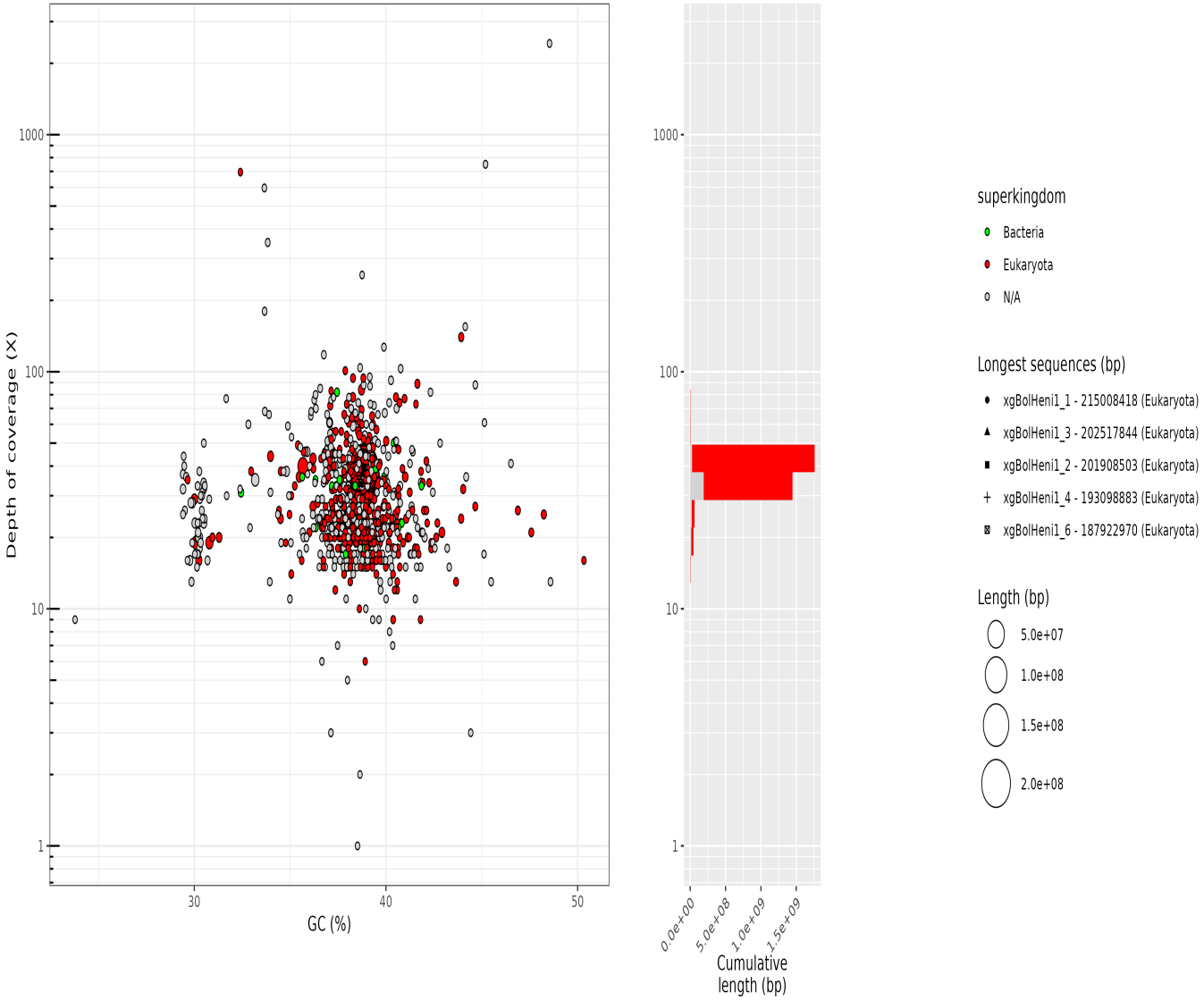


Distribution of k-mer counts coloured by their presence in reads/assemblies

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	41	161

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: Mohamed Meftah Saoues
Affiliation: Genoscope

Date and time: 2026-04-20 08:40:44 CEST