ERGA Assembly Report
v24.10.15

Tags: ATLASea[INVALID TAG]

TxID	179646
ToLID	**xgBolBran1**
Species	Bolinus brandaris
Class	Gastropoda
Order	Neogastropoda

Genome Traits	Expected	Observed
Haploid size (bp)	2,221,704,037	2,490,557,693
Haploid Number	28 (source: direct)	35
Ploidy	2 (source: ancestor)	2
Sample Sex	Unknown	Unknown

EBP metrics summary and curation notes

Obtained EBP quality metric for collapsed: 6.7.Q51

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
. BUSCO duplicated value is more than 5% for collapsed
. More than 1000 gaps/Gbp for collapsed

Curator notes

. Interventions/Gb: 70
. Contamination notes: ""
. Other observations: "The assembly of Bolinus brandaris(xgBolBran) isbased on 100x PacBio data and 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, 535 contigs were identified as contaminants (bacterial, archaeal, or viral), totaling 66 Mb (with the largest being 1.4 Mb). Additionally, 2,784 region of 356 Mb were identified as haplotypic duplications and removed. The mitochondrial genome was assembled using OATK. Finally, the primary assembly was analyzed and manually improved using Pretext. During manual curation, 1 haplotypic region of 356 kb was removed. 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	2,491,975,237	2,490,557,693
GC %	41.39	41.39
Gaps/Gbp	1,859.57	1,868.66
Total gap bp	463,400	472,600
Scaffolds	673	548
Scaffold N50	74,274,832	73,600,153
Scaffold L50	13	14
Scaffold L90	29	31
Contigs	5,307	5,202
Contig N50	1,092,384	1,096,456
Contig L50	663	662
Contig L90	2,463	2,463
QV	41.8373	51.445
Kmer compl.	85.502	87.5974
BUSCO sing.	65.1%	65.1%
BUSCO dupl.	13.7%	13.6%
BUSCO frag.	4.4%	4.4%
BUSCO miss.	16.8%	16.9%

BUSCO: 5.4.3 (euk_genome_met, metaeuk) / Lineage: mollusca_odb10 (genomes:7, BUSCOs:5295)
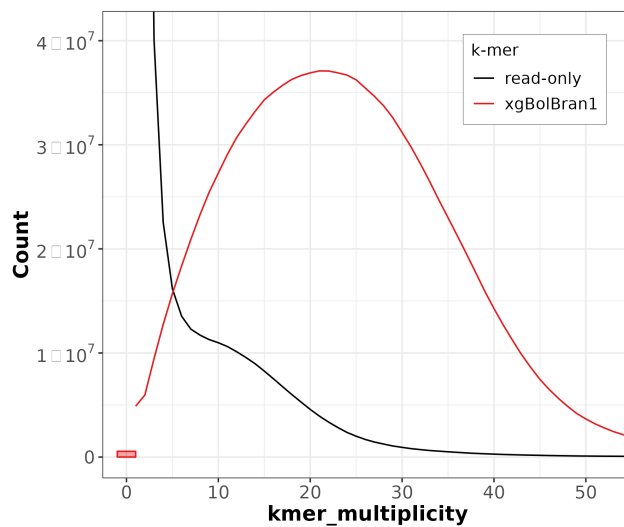
HiC contact map of curated assembly
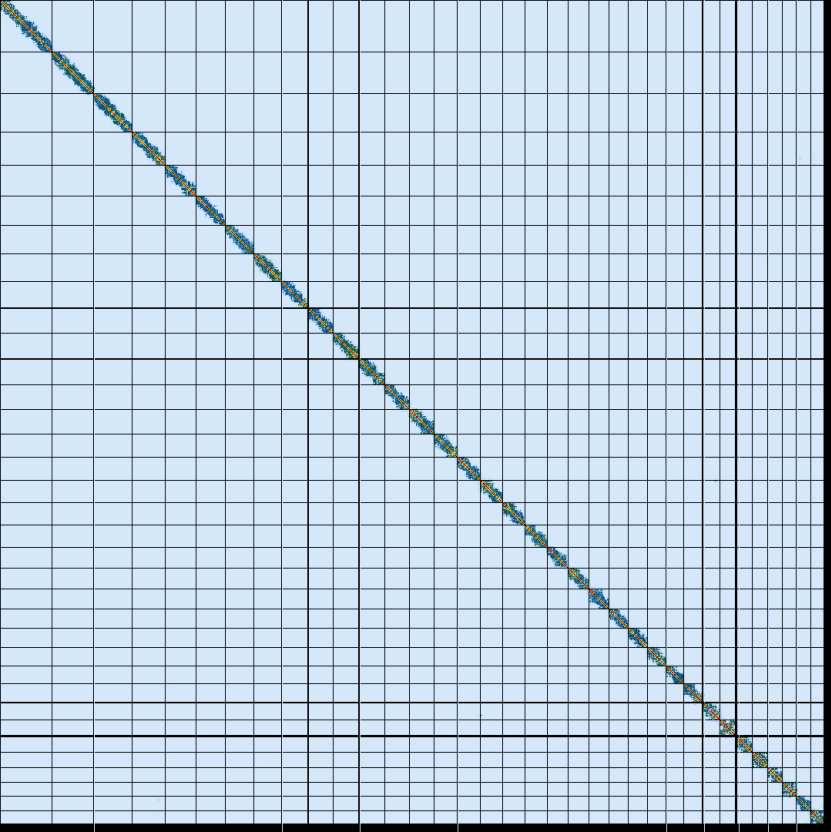
collapsed [LINK]

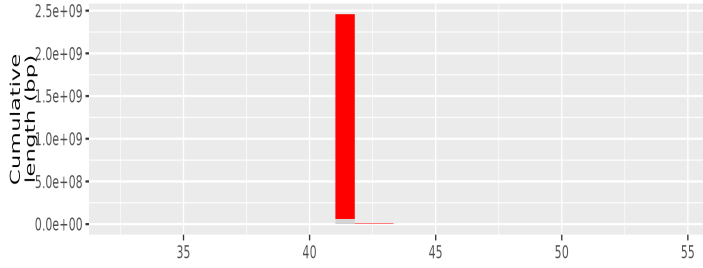# K-mer spectra of curated assembly
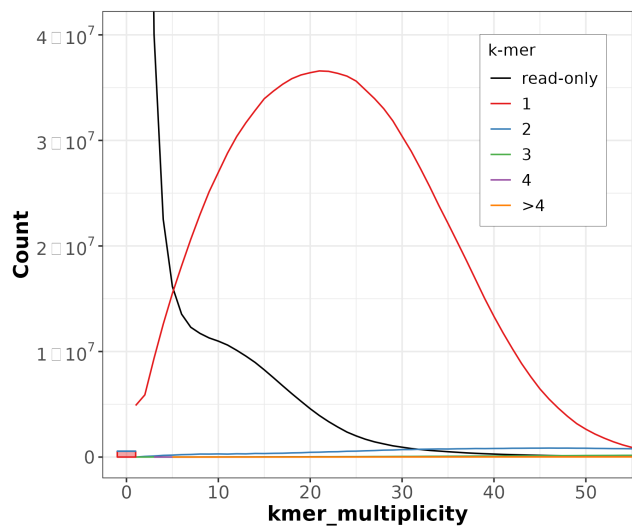
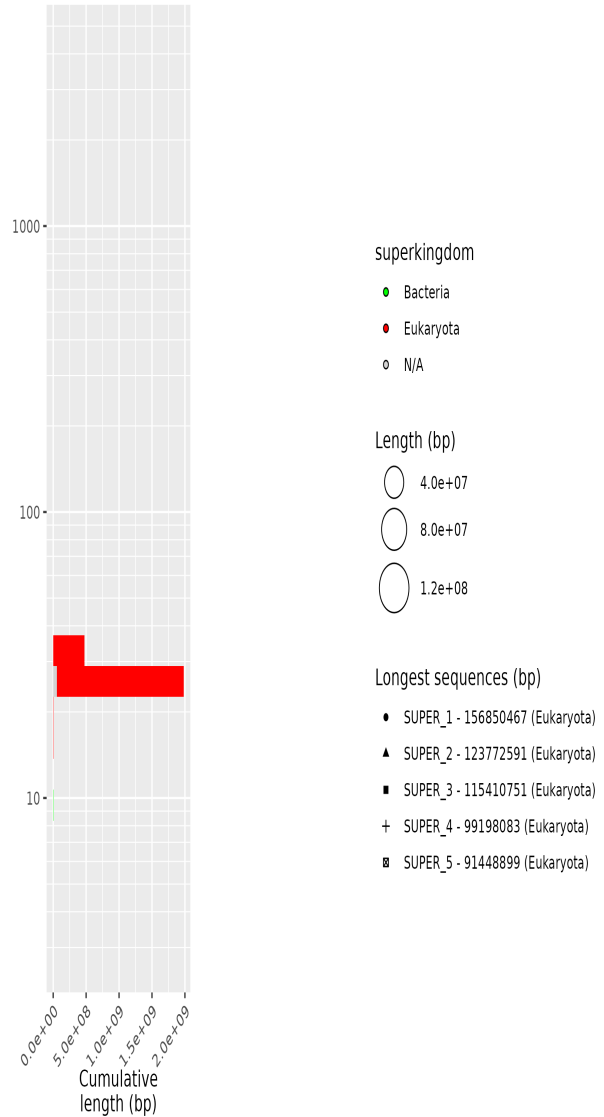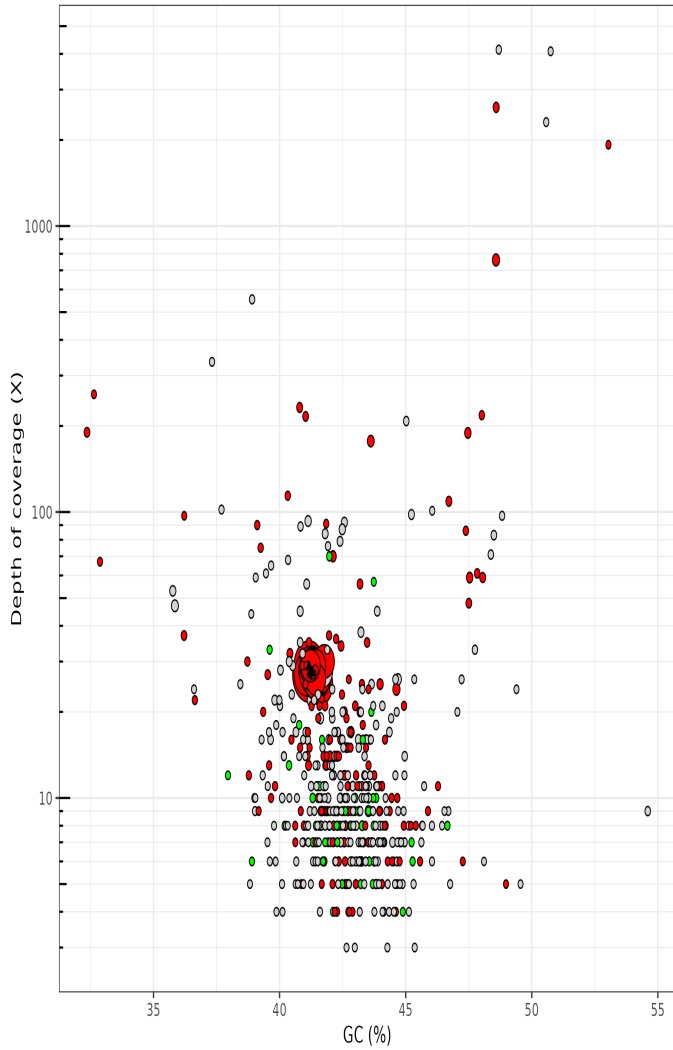Distribution of k-mer counts per copy
numbers found in asm

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	100	100

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: Benjamin Istace
Affiliation: Genoscope

Date and time: 2025-01-27 08:29:36 CET