

ERGA Assembly Report

v24.10.15

Tags: ATLASEa[INVALID TAG]

TxID	676947
ToLID	xgBerEdwa1
Species	Berthellina edwardsii
Class	Gastropoda
Order	Pleurobranchida

Genome Traits	Expected	Observed
Haploid size (bp)	892,451,113	932,957,193
Haploid Number	11 (source: ancestor)	12
Ploidy	2 (source: ancestor)	2
Sample Sex	Unknown	Unknown

EBP metrics summary and curation notes

Obtained EBP quality metric for collapsed: 6.7.Q42

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

- . Observed Haploid Number is different from Expected

Curator notes

- . Interventions/Gb: 96
- . Contamination notes: ""
- . Other observations: "The assembly of *Berthellina edwardsii* (xgBerEdwa1) is based on 29X ONT data and 129X Arima Hi-C data generated as part of the ATLASEa programme (<https://www.atlasea.fr>). The assembly process included the following steps: initial ONT 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, 31 contigs were identified as contaminants (bacterial, archaeal, or viral), totaling 1.677 Mb (with the largest being 0.187 Mb). Additionally, 480 regions totaling 51.077 Mb (with the largest being 0.782 Mb) were identified as haplotypic duplications and removed. Mitochondrial genome was assembled using ptGAUL. Finally, the primary assembly was analyzed and manually improved using Pretext. During manual curation, 24 haplotypic regions were removed, totaling 11.261Mb (with the largest being 6.307Mb). Chromosome-scale scaffolds confirmed by Hi-C data were named in order of size. "

Quality metrics table

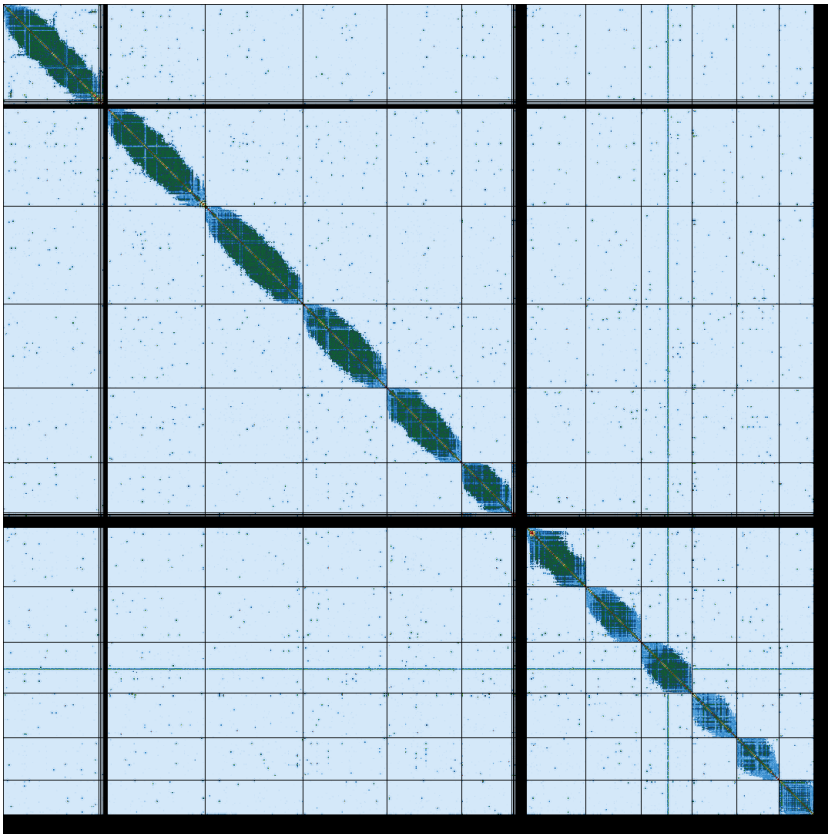
Metrics	Pre-curation collapsed	Curated collapsed
Total bp	944,224,978	932,957,193
GC %	37.34	37.34
Gaps/Gbp	258.41	288.33
Total gap bp	24,400	31,800
Scaffolds	322	293
Scaffold N50	82,913,772	83,783,343
Scaffold L50	5	5
Scaffold L90	12	11
Contigs	566	562
Contig N50	6,589,644	6,446,617
Contig L50	44	44
Contig L90	172	169
QV	42.9281	42.9279
Kmer compl.	93.8454	93.3217
BUSCO sing.	96.9%	95.4%
BUSCO dupl.	1.5%	1.1%
BUSCO frag.	0.6%	2.3%
BUSCO miss.	0.9%	1.2%

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)

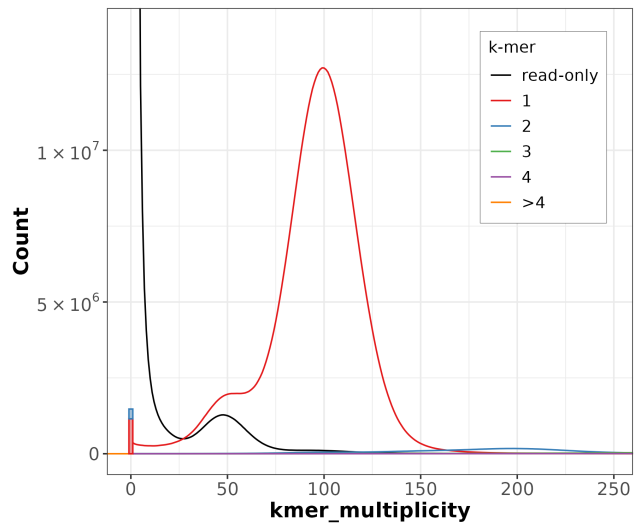
BUSCO: 5.8.2 (euk_genome_min, miniprot) / Lineage: mollusca_odb12 (genomes:36, BUSCOs:4421)

HiC contact map of curated assembly

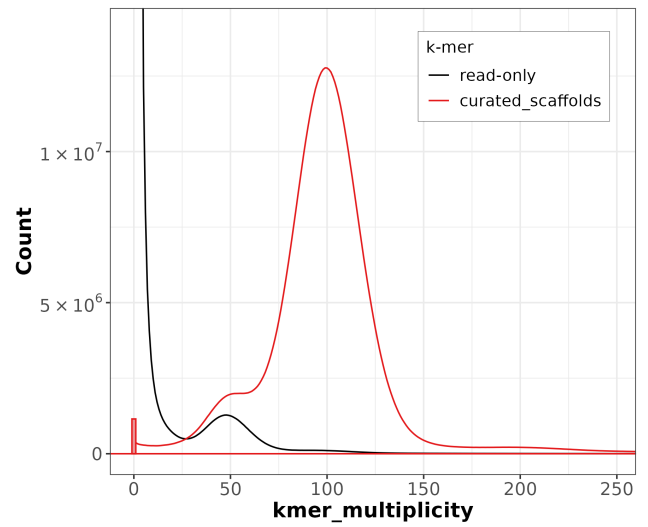


collapsed [\[LINK\]](#)

K-mer spectra of curated assembly

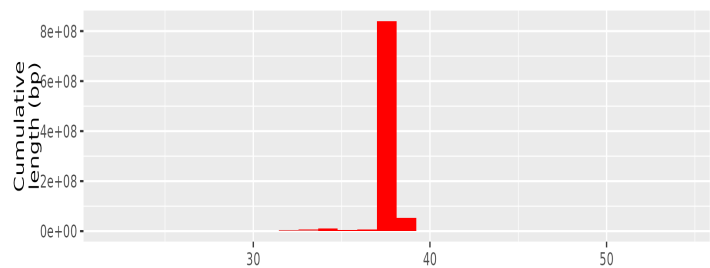


Distribution of k-mer counts per copy numbers found in asm



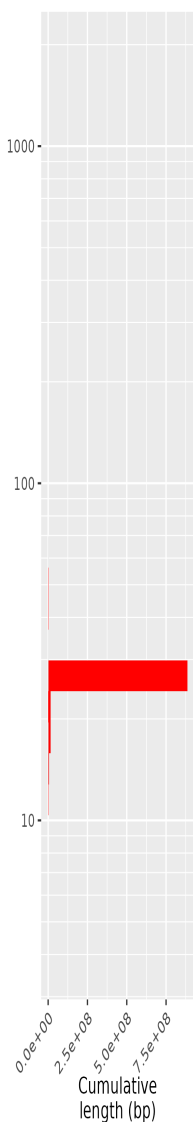
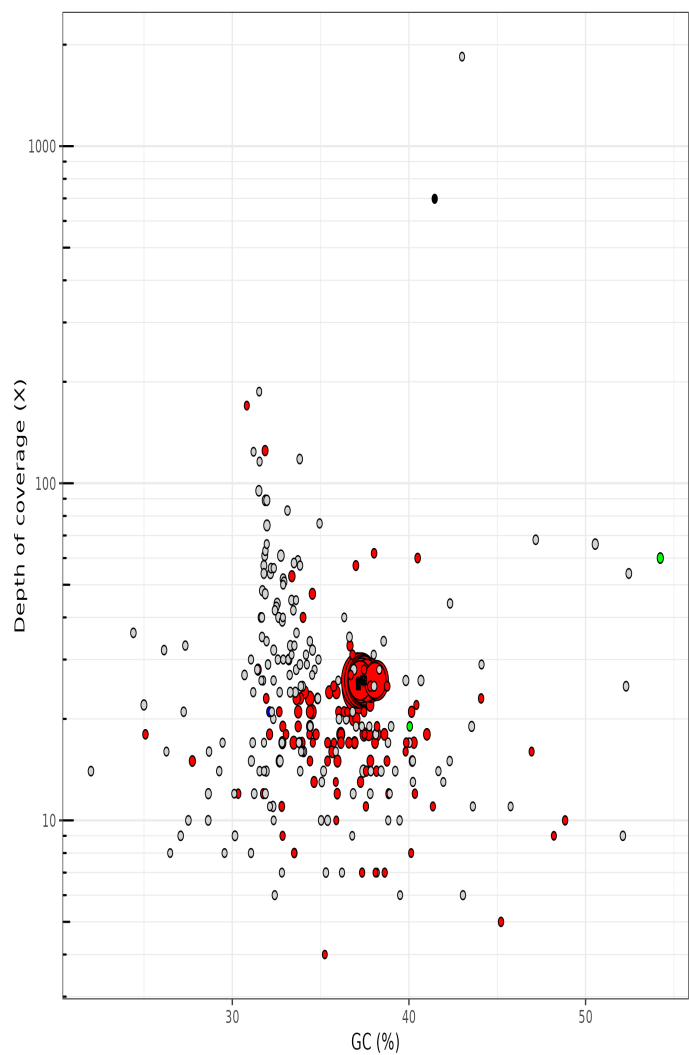
Distribution of k-mer counts coloured by their presence in reads/assemblies

Post-curation contamination screening



TAPAs summary Graph

(1 0X contig has been hidden)



Longest sequences (bp)

- xgBerEdwa1_2 - 110376511 (Eukaryota)
- ▲ xgBerEdwa1_3 - 109426926 (Eukaryota)
- xgBerEdwa1_1 - 107385469 (Eukaryota)
- + xgBerEdwa1_4 - 93969286 (Eukaryota)
- xgBerEdwa1_5 - 83783343 (Eukaryota)

Length (bp)

- 2.5e+07
- 5.0e+07
- 7.5e+07
- 1.0e+08

superkingdom

- Archaea
- Bacteria
- Eukaryota
- N/A
- Viruses

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	29	130

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: 2026-08-06 19:04:54 CEST