

ERGA Assembly Report

v24.10.15

Tags: ATLASEa[INVALID TAG]

TxID	537785
ToLID	xgNerPelo1
Species	Nerita peloronta
Class	Gastropoda
Order	Cycloneritida

Genome Traits	Expected	Observed
Haploid size (bp)	900,137,220	1,045,060,422
Haploid Number	9 (source: ancestor)	12
Ploidy	2 (source: ancestor)	2
Sample Sex	Unknown	Unknown

EBP metrics summary and curation notes

Obtained EBP quality metric for collapsed: 7.7.Q42

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

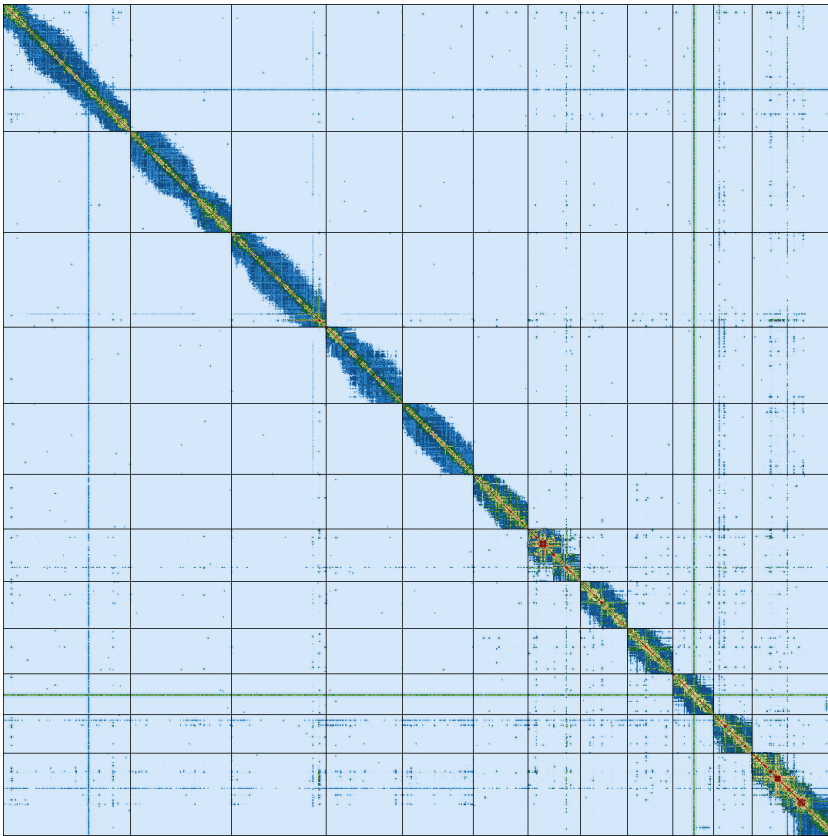
. Interventions/Gb: 6
. Contamination notes: ""
. Other observations: "The assembly of *Nerita peloronta* (xgNerPelo1) is based on 37.74X ONT data and 127.94X Arima Hi-C data generated as part of the ATLASEa programme (<https://www.atlasea.fr>). The assembly process included the following steps: initial 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, 1 contigs was identified as contaminants (bacterial, archaeal, or viral), totaling 0.09 Mb (with the largest being 0.09 Mb). Additionally, 45 regions totaling 18.748 Mb (with the largest being 4.419 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, 1 haplotypic region was removed, totaling 0.34Mb. 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	1,045,412,788	1,045,060,422
GC %	47.81	47.81
Gaps/Gbp	21.04	22.97
Total gap bp	2,200	2,700
Scaffolds	17	16
Scaffold N50	89,055,500	96,383,735
Scaffold L50	5	5
Scaffold L90	11	10
Contigs	39	40
Contig N50	67,937,159	67,937,159
Contig L50	6	6
Contig L90	18	18
QV	42.1045	42.1059
Kmer compl.	85.7558	85.752
BUSCO sing.	97.1%	97.1%
BUSCO dupl.	0.6%	0.6%
BUSCO frag.	0.3%	0.3%
BUSCO miss.	1.9%	1.9%

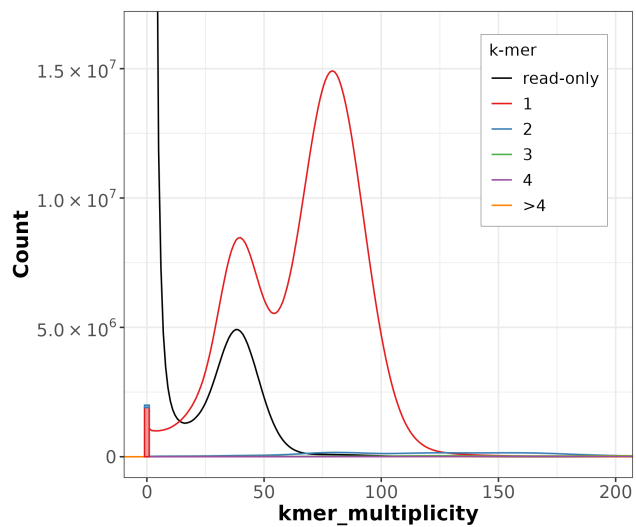
BUSCO: 6.1.0 (euk_genome_min, miniprot) / Lineage: mollusca_odb12.2 (genomes:31, BUSCOs:4236)

HiC contact map of curated assembly

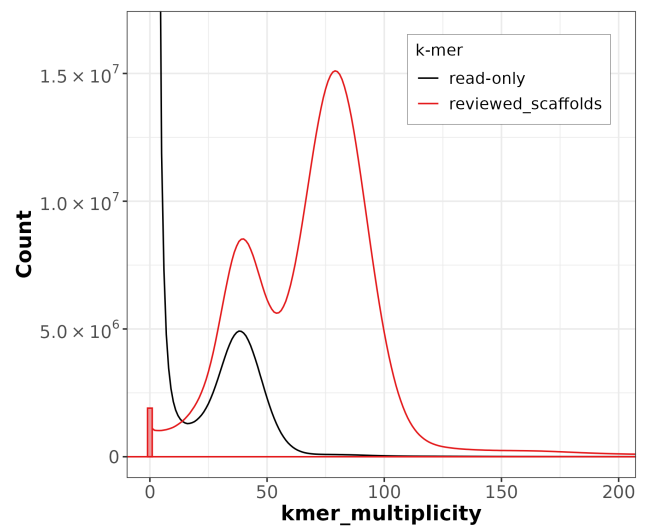


collapsed [\[LINK\]](#)

K-mer spectra of curated assembly

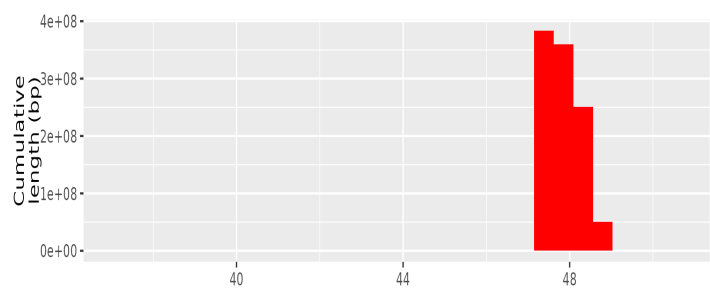


Distribution of k-mer counts per copy numbers found in asm

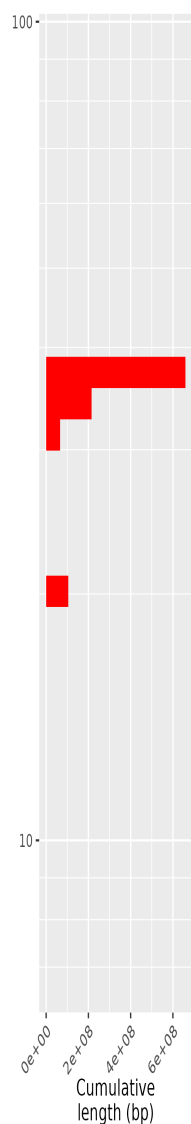
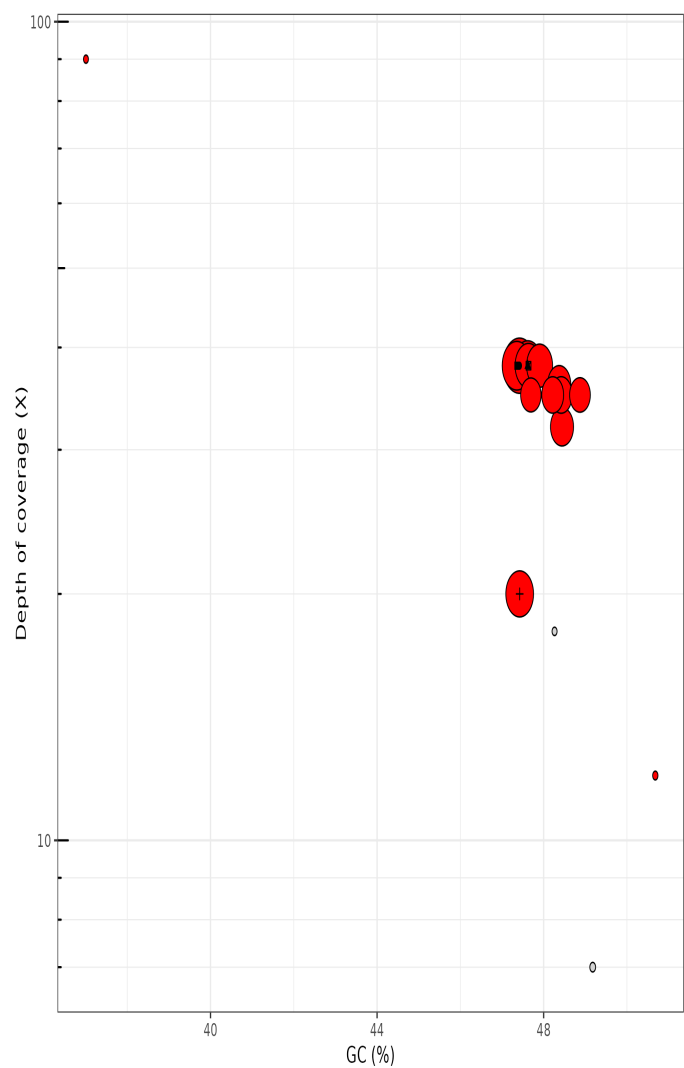


Distribution of k-mer counts coloured by their presence in reads/assemblies

Post-curation contamination screening



TAPAs summary Graph



Longest sequences (bp)

- xgNerPelo1_1 - 160461574 (Eukaryota)
- ▲ xgNerPelo1_2 - 126328411 (Eukaryota)
- xgNerPelo1_3 - 118952126 (Eukaryota)
- + xgNerPelo1_X - 104339180 (Eukaryota)
- ▣ xgNerPelo1_4 - 96383735 (Eukaryota)

Length (bp)

- 4.0e+07
- 8.0e+07
- 1.2e+08
- 1.6e+08

superkingdom

- Eukaryota
- N/A

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	37	127

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: Emilie Teodori

Affiliation: Genoscope

Date and time: 2026-08-28 23:09:05 CEST