ERGA Assembly Report

v24.10.15

Tags: ATLASea[INVALID TAG]

TxID	444096
ToLID	**xgNasNiti1**
Species	Nassarius nitidus
Class	Gastropoda
Order	Neogastropoda

Genome Traits	Expected	Observed
Haploid size (bp)	2,200,749,236	2,458,602,159
Haploid Number	14 (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.Q41

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

Curator notes

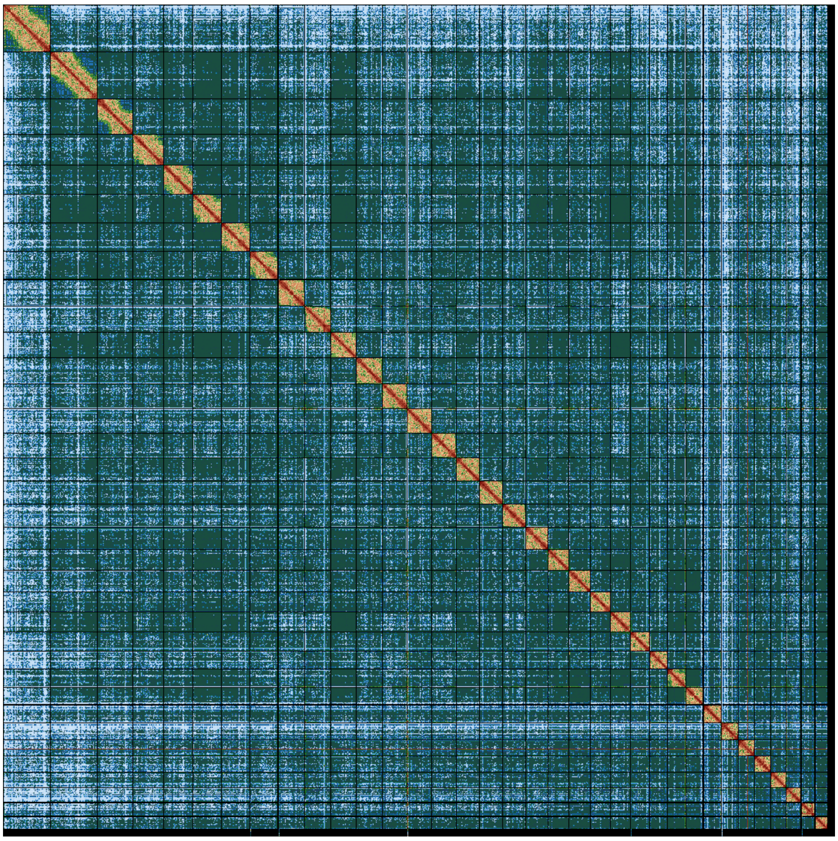
. Interventions/Gb: 70
. Contamination notes: ""
. Other observations: "The assembly of Nassarius nitidus (xgNasNiti1) is based on 47X ONT data and 135X Arima Hi-C data generated as part of the ATLASea programme (https://www.atlasea.fr). The assembly process included the following steps: initial ONT 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, 788 contigs were identified as contaminants (bacterial, archaeal, or viral), totaling 163.192 Mb (with the largest being 4.942 Mb). Additionally, 2156 regions totaling 209.183 Mb (with the largest being 0.984 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, 17 haplotypic regions were removed, totaling 8.40Mb, (with the largest being 1.241Mb). 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	2,467,086,632	2,458,602,159
GC %	41.46	41.46
Gaps/Gbp	759.2	745.14
Total gap bp	257,400	263,900
Scaffolds	477	360
Scaffold N50	69,470,540	72,607,271
Scaffold L50	14	14
Scaffold L90	31	30
Contigs	2,243	2,192
Contig N50	2,358,760	2,362,126
Contig L50	320	319
Contig L90	1,072	1,067
QV	41.6485	41.6482
Kmer compl.	72.0413	71.9088
BUSCO sing.	71.8%	71.8%
BUSCO dupl.	12.1%	12.1%
BUSCO frag.	11.2%	11.2%
BUSCO miss.	5.0%	5.0%

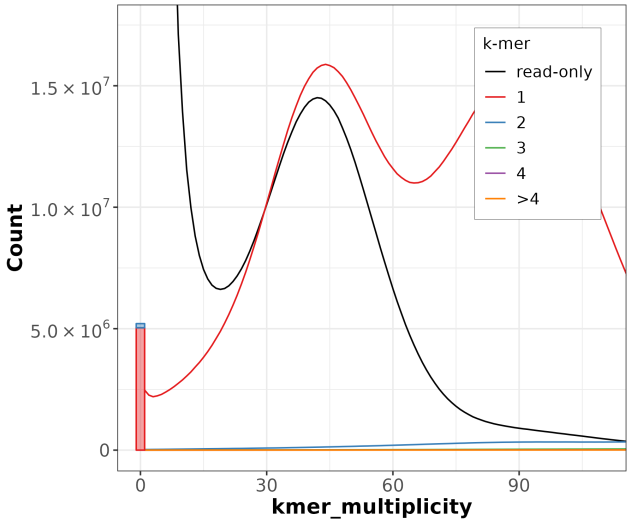
BUSCO: 5.8.2 (euk_genome_met, metaeuk) / Lineage: mollusca_odb12 (genomes:36, BUSCOs:4421)

HiC contact map of curated assembly

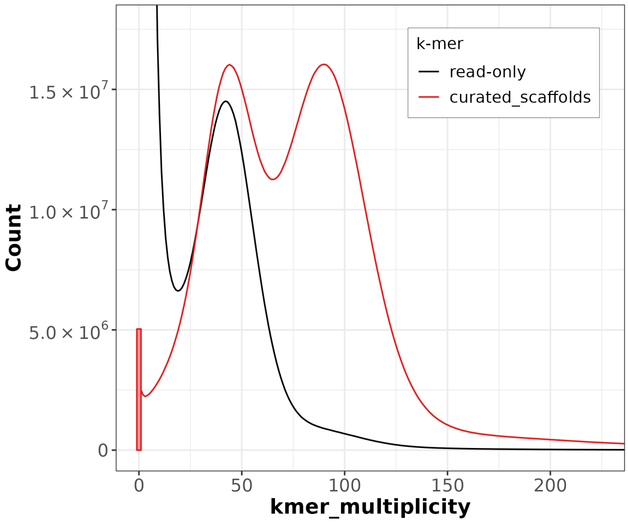


collapsed [LINK]

K-mer spectra of curated assembly

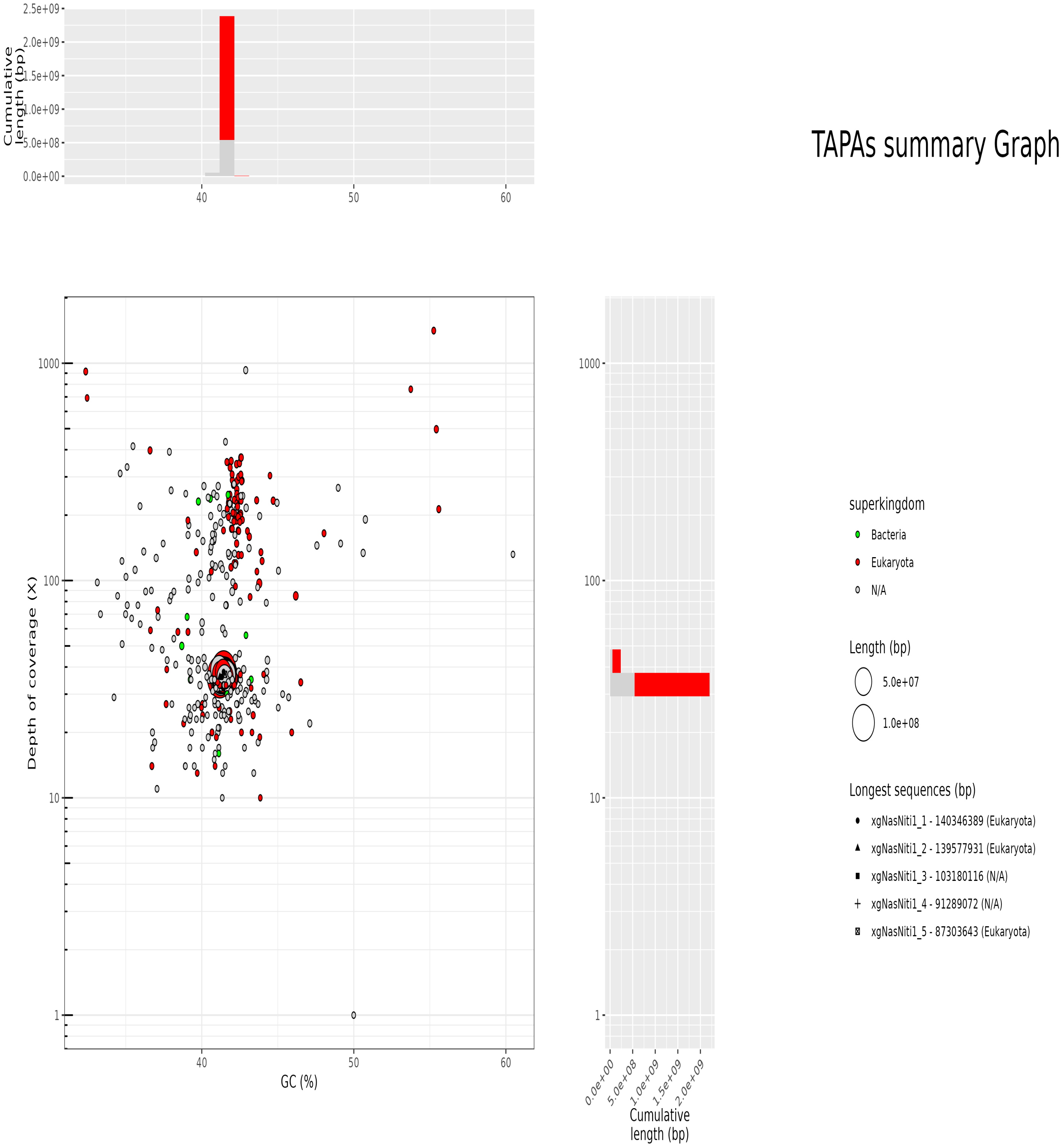


Distribution of k-mer counts per copy numbers found in asm



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

Post-curation contamination screening



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	47	134

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: 2025-08-31 04:12:10 CEST