ERGA Assembly Report

v24.10.15

Tags: ATLASea[INVALID TAG]

TxID	2709805
ToLID	**xgRocNilo1**
Species	Rochia nilotica
Class	Gastropoda
Order	Trochida

Genome Traits	Expected	Observed
Haploid size (bp)	1,351,252,191	1,394,759,703
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: 7.7.Q49

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

. Observed Haploid Number is different from Expected

Curator notes

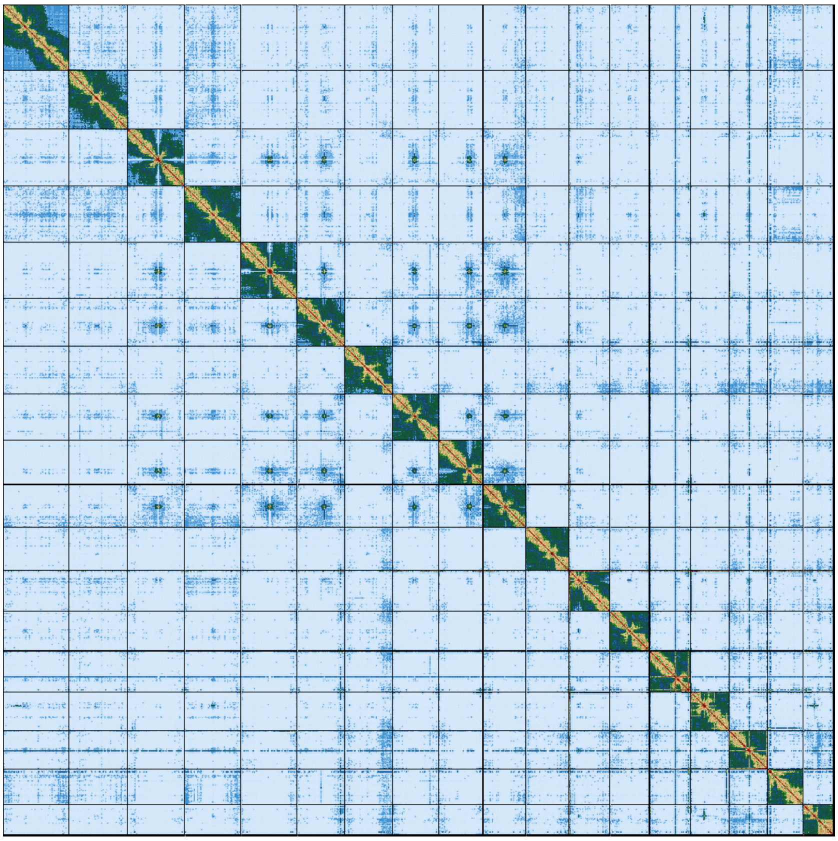
. Interventions/Gb: 70
. Contamination notes: ""
. Other observations: "The assembly of Rochia nilotica (xgRocNilo1) is based on 41X PacBio data and 201X Arima Hi-C data generated as part of the ATLASea programme (https://www.atlasea.fr). The assembly process included the following steps: initial PacBio assembly generation with Hifiasm, removal of contaminant sequences using Context, removal of haplotypic duplications using purge_dups, and Hi-C-based scaffolding with YaHS. No contigs were found to be contaminants (bacterial, archaeal, or viral). Additionally, 179 regions totaling 26.247 Mb (with the largest being 2.144 Mb) were identified as haplotypic duplications and removed. Mitochondrial genome was assembled using OATK. During manual curation, 6 haplotypic regions were removed, totaling 3,384,646 pb (with the largest being 874,703 pb). 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	1,398,160,267	1,394,759,703
GC %	37.55	37.55
Gaps/Gbp	85.83	93.21
Total gap bp	12,000	14,900
Scaffolds	67	60
Scaffold N50	77,765,722	77,910,744
Scaffold L50	8	8
Scaffold L90	16	16
Contigs	187	190
Contig N50	15,596,904	15,596,904
Contig L50	29	29
Contig L90	87	88
QV	49.5508	49.5478
Kmer compl.	92.279	92.225
BUSCO sing.	98.7%	98.8%
BUSCO dupl.	0.8%	0.6%
BUSCO frag.	0.2%	0.2%
BUSCO miss.	0.4%	0.5%

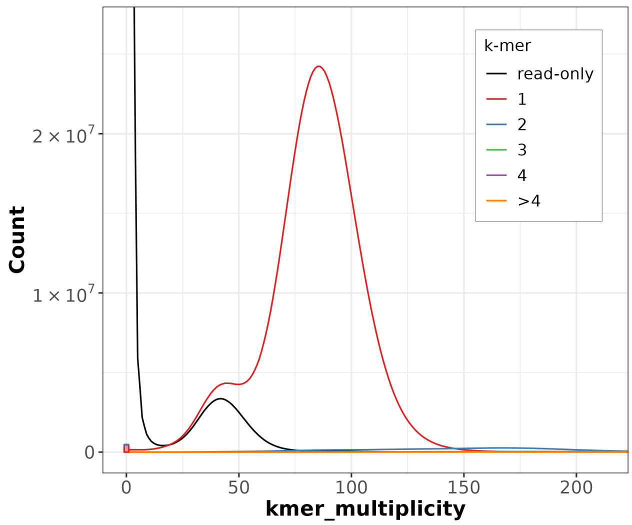
BUSCO: 6.0.0 (euk_genome_min, miniprot) / Lineage: mollusca_odb12 (genomes:36, BUSCOs:4421)

HiC contact map of curated assembly

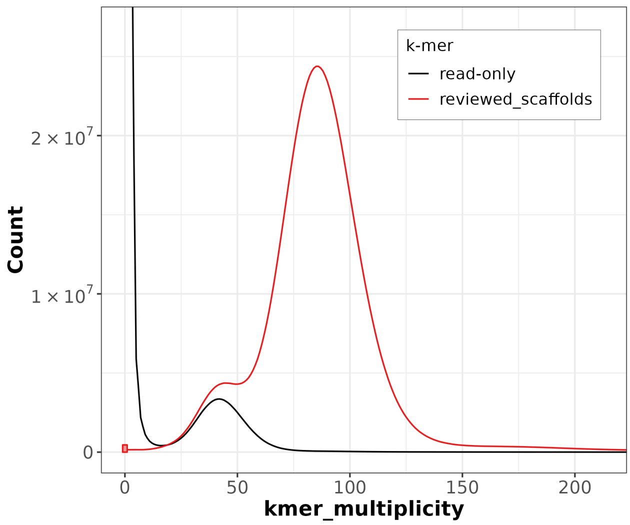


collapsed [LINK]

K-mer spectra of curated assembly

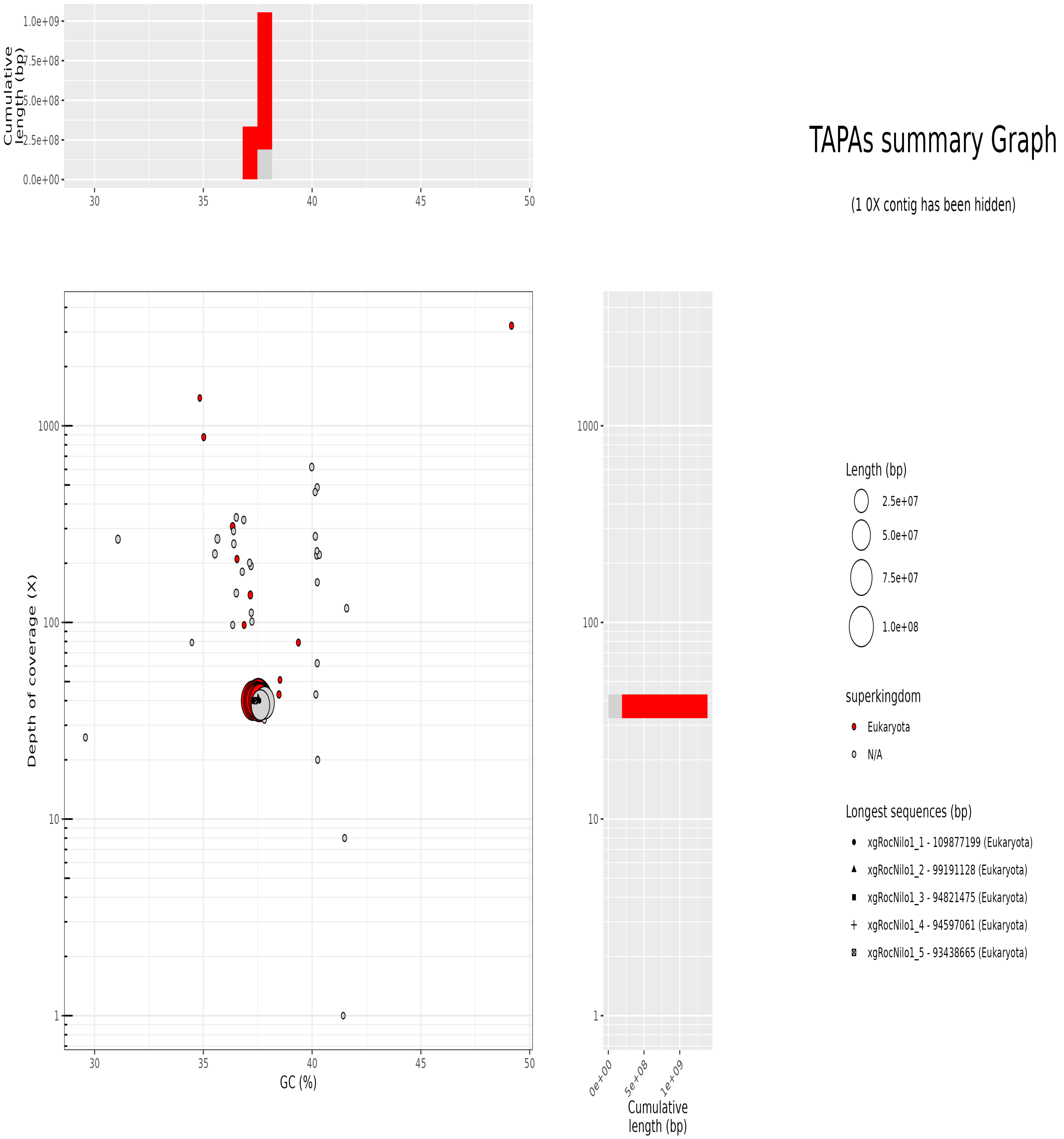


Distribution of k-mer counts per copy numbers found in asm



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

Post-curation contamination screening



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	201

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: Lola Demirdjian
Affiliation: Genoscope

Date and time: 2025-12-21 02:57:55 CET