

ERGA Assembly Report

v24.09.10

Tags: ATLASea[INVALID TAG]

TxID	488666
ToLID	xgNevJose1.1
Species	Neverita josephinia
Class	Gastropoda
Order	Littorinimorpha

Genome Traits	Expected	Observed
Haploid size (bp)	1,607,754,785	1,731,099,993
Haploid Number	17 (source: ancestor)	17
Ploidy	2 (source: ancestor)	2
Sample Sex	Unknown	Unknown

EBP metrics summary and curation notes

Obtained EBP quality metric for collapsed: 6.8.Q62

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

- . BUSCO single copy value is less than 90% for collapsed

Curator notes

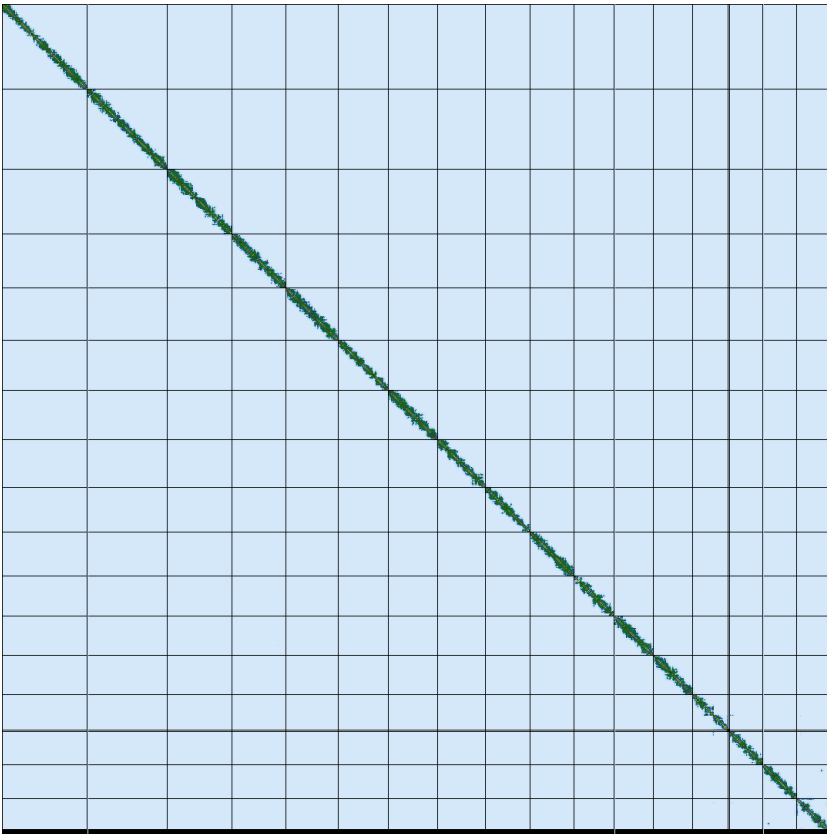
- . Interventions/Gb: 1
- . Contamination notes: "16 bacterial sequences of 4.6 Mb were removed before scaffolding. "
- . Other observations: "Manual curation was minimal because we already had the chromosomes after scaffolding with yahs. Four sequences of 2.6Mb have been assigned as haplotig and removed. The post-curation contamination screening plot corresponds to the assembly before bacterial sequences are removed. "

Quality metrics table

Metrics	Pre-curation collapsed	Curated collapsed
Total bp	1,733,806,305	1,731,099,993
GC %	42.34	42.35
Gaps/Gbp	419.31	419.96
Total gap bp	72,700	72,900
Scaffolds	154	109
Scaffold N50	101,996,338	101,996,338
Scaffold L50	7	7
Scaffold L90	15	15
Contigs	881	836
Contig N50	5,572,203	5,574,000
Contig L50	87	86
Contig L90	360	358
QV	44.2403	62.3093
Kmer compl.	98.3623	98.2109
BUSCO sing.	78.7%	78.7%
BUSCO dupl.	0.5%	0.6%
BUSCO frag.	4.8%	4.8%
BUSCO miss.	16.0%	15.9%

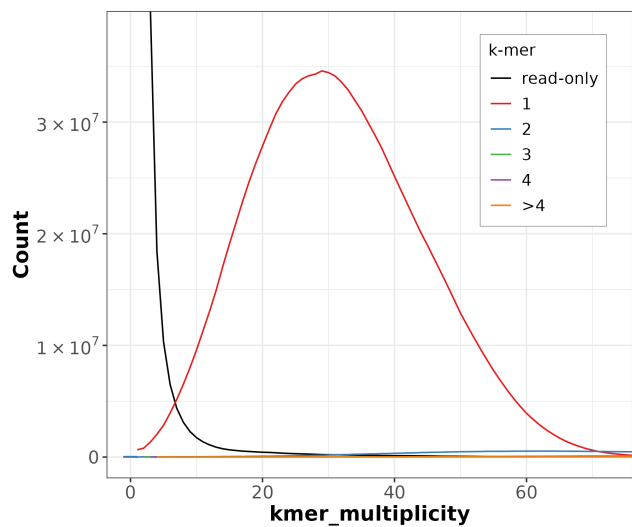
BUSCO 5.4.3 Lineage: mollusca_odb10 (genomes:7, BUSCOs:5295)

HiC contact map of curated assembly

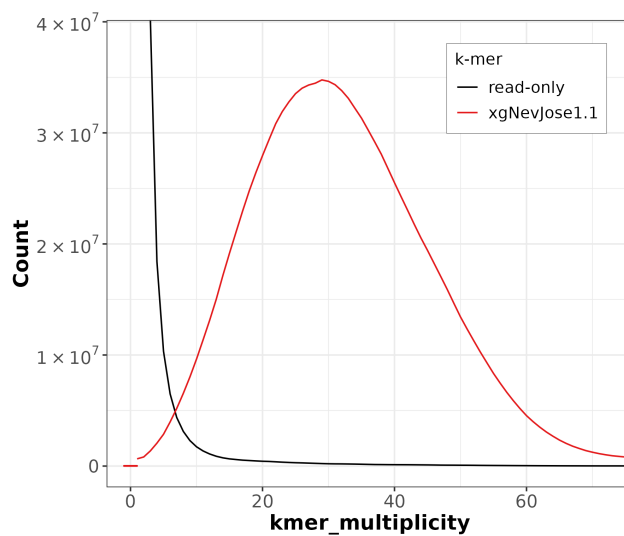


collapsed [\[LINK\]](#)

K-mer spectra of curated assembly

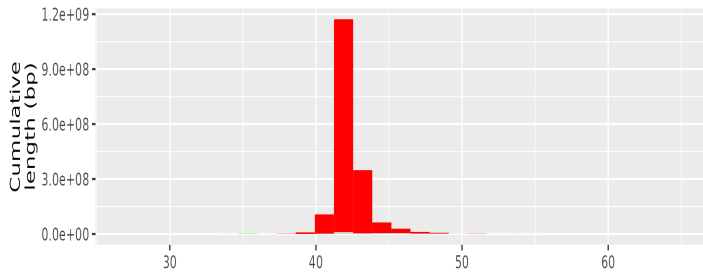


Distribution of k-mer counts per copy numbers found in asm

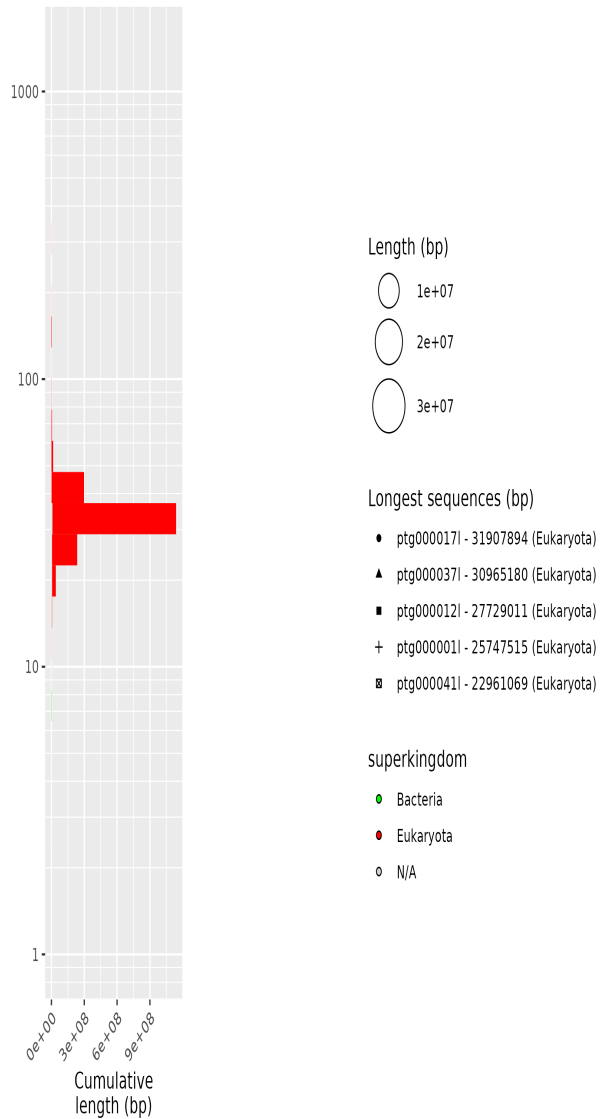
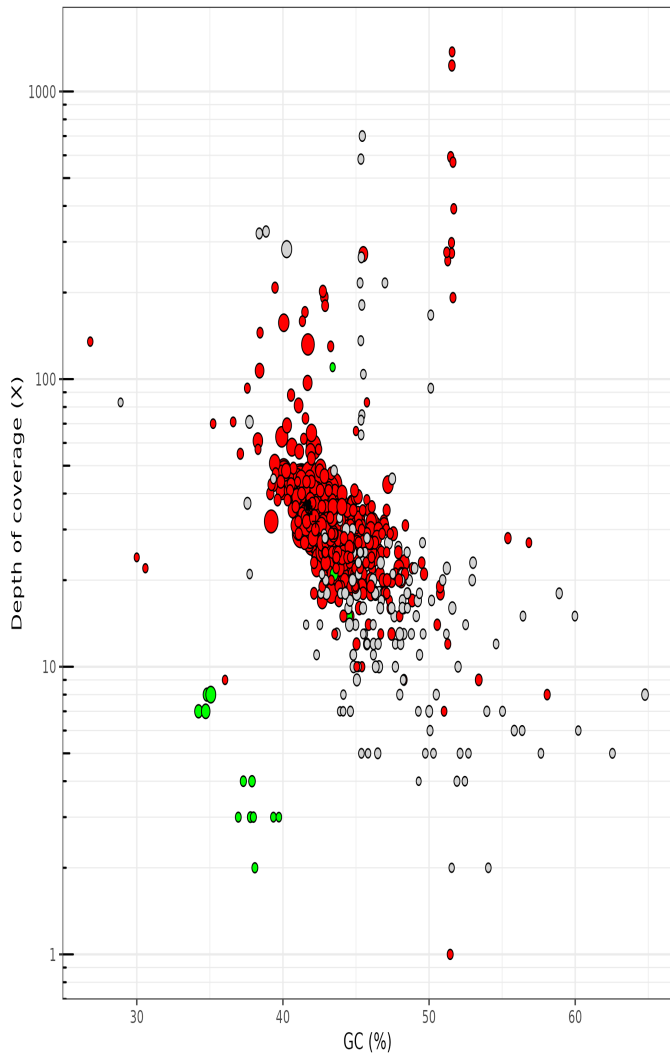


Distribution of k-mer counts coloured by their presence in reads/assemblies

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 (4-enz)
Coverage	35	23

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: Adama Ndar
Affiliation: Genoscope

Date and time: 2024-11-06 13:22:30 CET