

ERGA Assembly Report

v24.09.10

Tags: ERGA-BGE

TxID	1980904
ToLID	xgMetNovel
Species	Metafruticicola noverca
Class	Gastropoda
Order	Stylommatophora

Genome Traits	Expected	Observed
Haploid size (bp)	1,632,108,444	1,762,538,353
Haploid Number	21 (source: ancestor)	26
Ploidy	2 (source: ancestor)	2
Sample Sex	Unknown	Unknown

EBP metrics summary and curation notes

Obtained EBP quality metric for collapsed: 6.7.Q51

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
- . Assembly length loss > 3% for collapsed
- . More than 1000 gaps/Gbp for collapsed

Curator notes

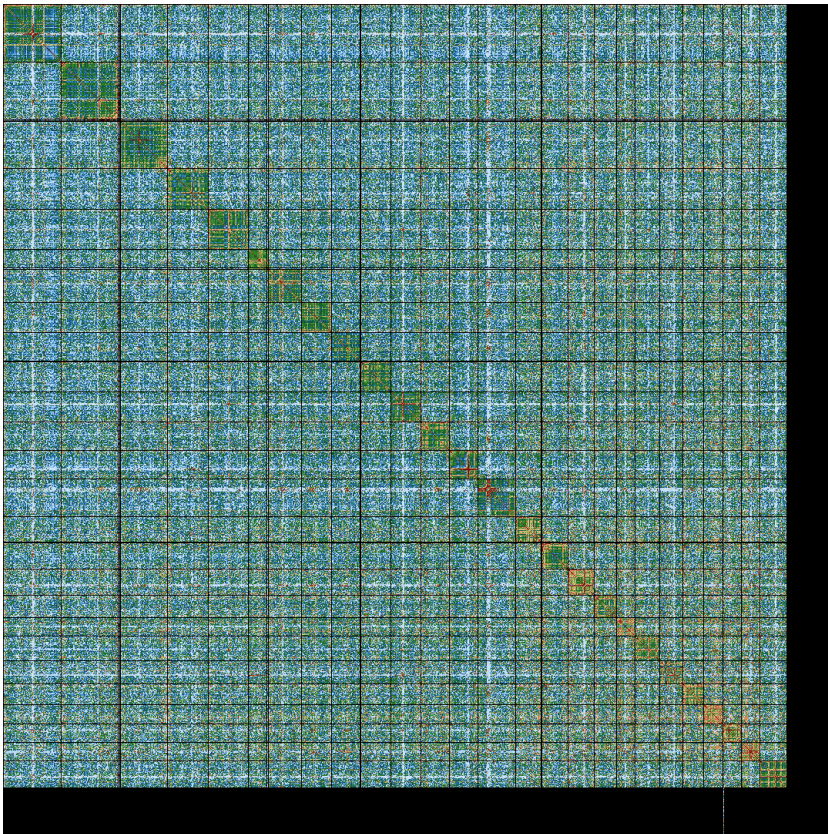
- . Interventions/Gb: 527
- . Contamination notes: "21 bacterial contigs have been removed before scaffolding "
- . Other observations: "lots of haplotypic duplications have been removed during the manual review. SUPER_8: end of the chromosome with an allelic duplication which has been left in place. "

Quality metrics table

Metrics	Pre-curation collapsed	Curated collapsed
Total bp	2,109,473,454	1,762,538,353
GC %	41.44	41.34
Gaps/Gbp	1,336.83	1,255.58
Total gap bp	282,000	223,900
Scaffolds	1,630	1,259
Scaffold N50	65,344,123	61,362,293
Scaffold L50	12	11
Scaffold L90	52	25
Contigs	4,450	3,472
Contig N50	1,029,000	1,270,000
Contig L50	620	405
Contig L90	2,068	1,502
QV	52.1751	51.8683
Kmer compl.	91.1952	85.6467
BUSCO sing.	81.2%	85.9%
BUSCO dupl.	16.1%	8.2%
BUSCO frag.	0.8%	0.8%
BUSCO miss.	1.9%	5.1%

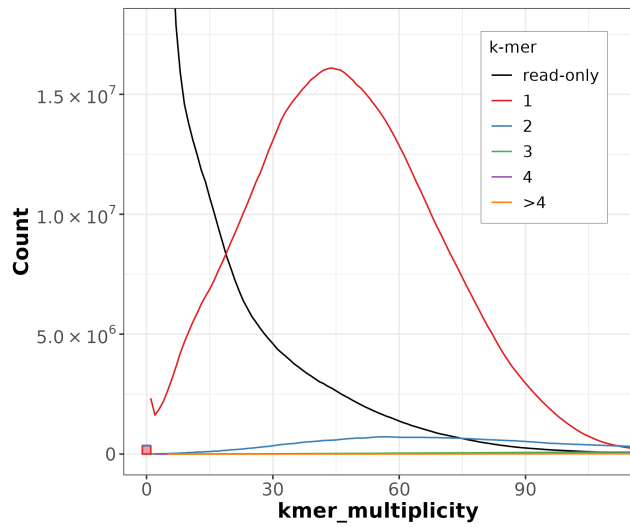
BUSCO 5.4.3 Lineage: eukaryota_odb10 (genomes:70, BUSCOs:255)

HiC contact map of curated assembly

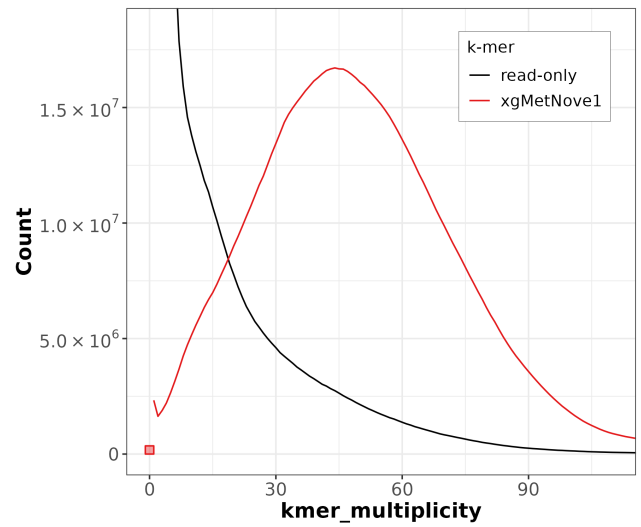


collapsed [\[LINK\]](#)

K-mer spectra of curated assembly

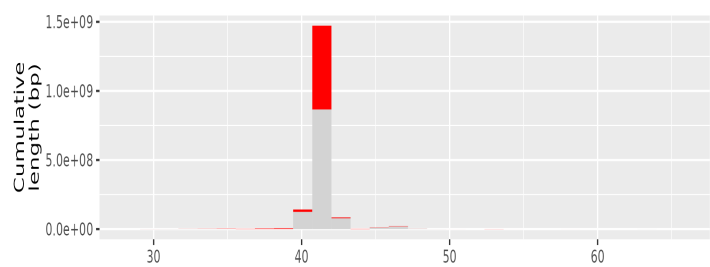


Distribution of k-mer counts per copy numbers found in asm

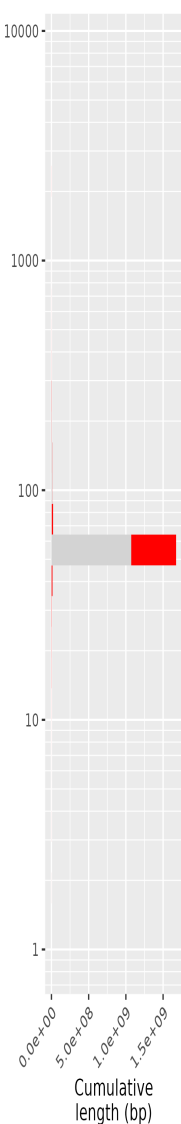
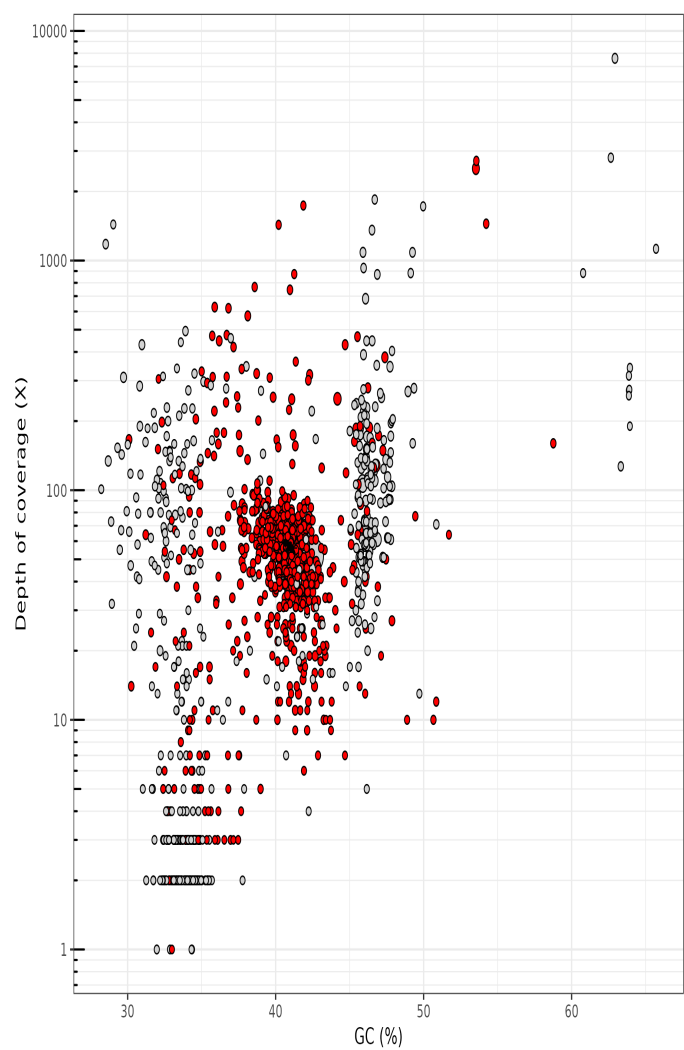


Distribution of k-mer counts coloured by their presence in reads/assemblies

Post-curation contamination screening



TAPAs summary Graph



- superkingdom
- Eukaryota
 - N/A
- Length (bp)
- 3.0e+07
 - 6.0e+07
 - 9.0e+07
 - 1.2e+08
- Longest sequences (bp)
- SUPER_1 - 123659894 (N/A)
 - ▲ SUPER_2 - 120448968 (N/A)
 - SUPER_3 - 98728621 (N/A)
 - + SUPER_4 - 87656732 (N/A)
 - SUPER_5 - 84022393 (Eukaryota)

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	63	29

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: S.Duprat

Affiliation: Genoscope

Date and time: 2024-10-18 02:25:33 CEST