

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 | 1,732,642,223 |
| 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

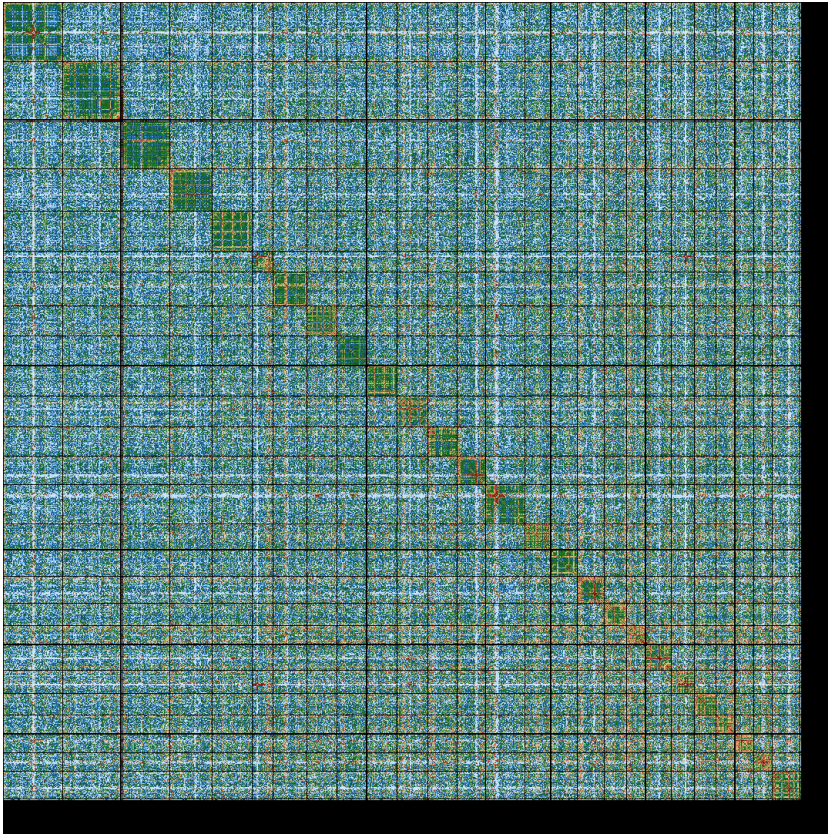
- . Interventions/Gb: 537
- . Contamination notes: "21 bacterial contigs have been removed before scaffolding "
- . Other observations: "lots of haplotypic duplications have been removed during the manual review. "

Quality metrics table

| Metrics | Pre-curation collapsed | Curated collapsed |
|--------------|------------------------|-------------------|
| Total bp | 2,109,473,454 | 1,732,642,223 |
| GC % | 41.44 | 41.35 |
| Gaps/Gbp | 1,336.83 | 1,275.51 |
| Total gap bp | 282,000 | 224,100 |
| Scaffolds | 1,630 | 956 |
| Scaffold N50 | 65,344,123 | 61,362,293 |
| Scaffold L50 | 12 | 11 |
| Scaffold L90 | 52 | 24 |
| Contigs | 4,450 | 3,166 |
| Contig N50 | 1,029,000 | 1,295,557 |
| Contig L50 | 620 | 394 |
| Contig L90 | 2,068 | 1,412 |
| QV | 52.1751 | 51.8527 |
| Kmer compl. | 91.1952 | 85.4274 |
| BUSCO sing. | 81.2% | 88.6% |
| BUSCO dupl. | 16.1% | 5.5% |
| BUSCO frag. | 0.8% | 0.8% |
| BUSCO miss. | 1.9% | 5.1% |

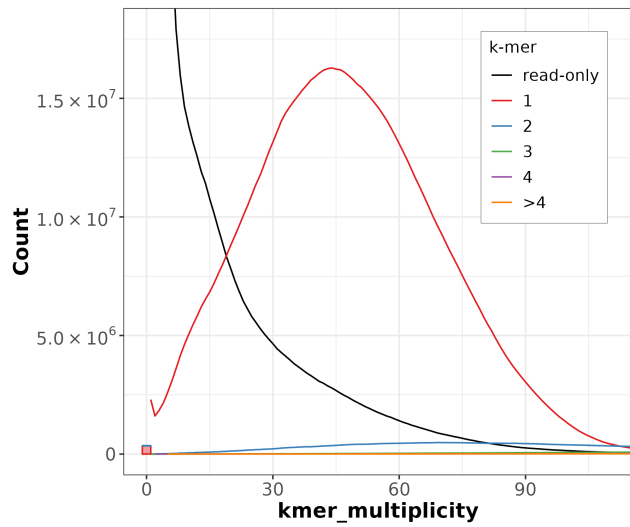
BUSCO 5.4.3 Lineage: eukaryota_odb10 (genomes:70, BUSCOs:255)

HiC contact map of curated assembly

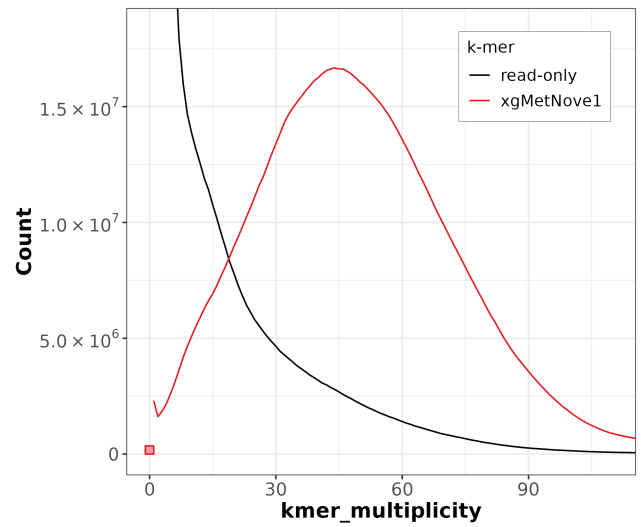


collapsed [\[LINK\]](#)

K-mer spectra of curated assembly

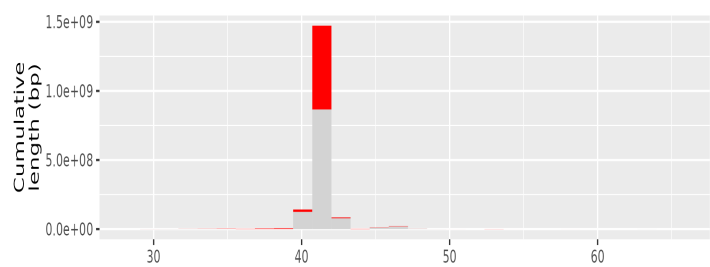


Distribution of k-mer counts per copy numbers found in asm

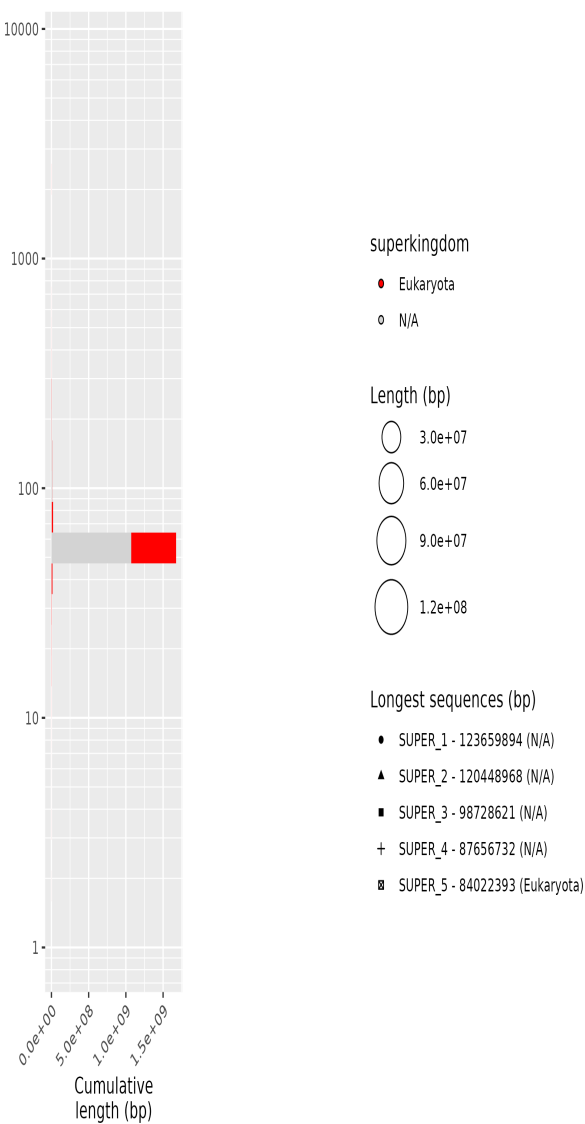
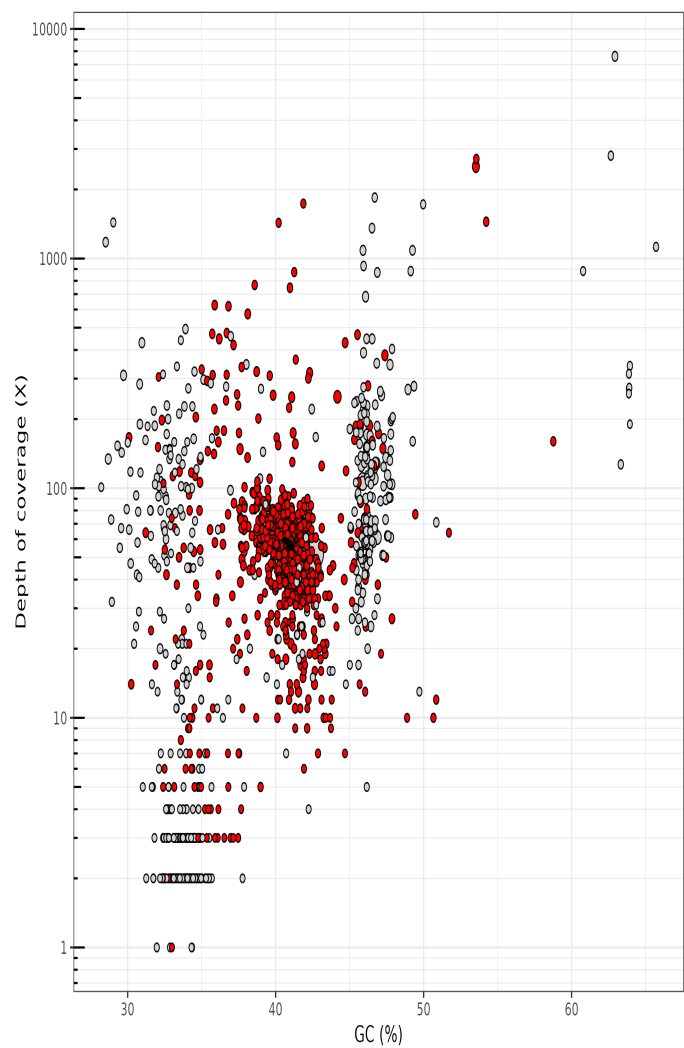


Distribution of k-mer counts coloured by their presence in reads/assemblies

Post-curation contamination screening



TAPAs summary Graph



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: 2024-11-08 11:41:36 CET