ERGA Assembly Report

v24.10.15

Tags: ATLASea[INVALID TAG]

TxID	1981583
ToLID	**wpFimZetl1**
Species	Fimbriosthenelais zetlandica
Class	Polychaeta
Order	Phyllodocida

Genome Traits	Expected	Observed
Haploid size (bp)	433,233,805	447,105,712
Haploid Number	10 (source: ancestor)	11
Ploidy	2 (source: ancestor)	2
Sample Sex	Unknown	Unknown

EBP metrics summary and curation notes

Obtained EBP quality metric for collapsed: 7.7.Q47

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

Curator notes

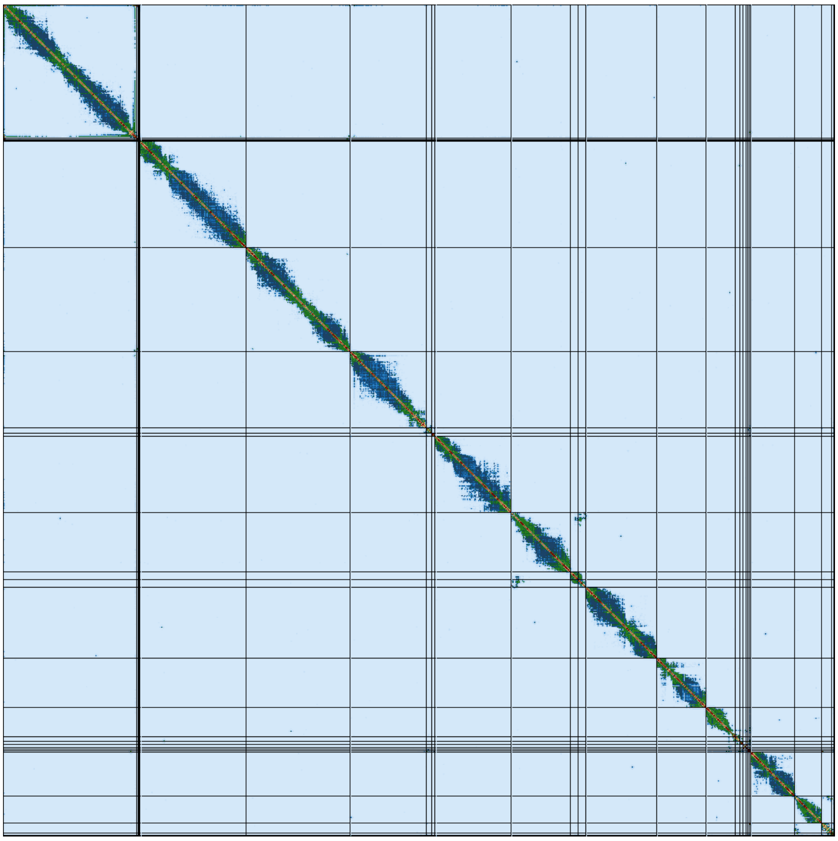
. Interventions/Gb: 87
. Contamination notes: ""
. Other observations: "The assembly of Fimbriosthenelais zetlandica (wpFimZetl1) is based on 55X PacBio data and 125X Arima Hi-C data generated as part of the ATLASea programme (https://www.atlasea.fr). The assembly process included the following steps: 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. In total, 3 contigs were identified as contaminants (bacterial, archaeal, or viral), totaling 0.04 Mb (with the largest being 0.016 Mb). Additionally, 339 regions totaling 30.876 Mb (with the largest being 2.58 Mb) were identified as haplotypic duplications and removed. The mitochondrial genome was assembled using OATK. During manual curation, 3 haplotypic regions were removed, totaling 3,107Mb (with the largest being 1,648Mb). 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	450,227,675	447,105,712
GC %	40.2	40.2
Gaps/Gbp	399.8	402.59
Total gap bp	21,300	23,300
Scaffolds	107	79
Scaffold N50	42,529,737	41,239,164
Scaffold L50	4	4
Scaffold L90	11	10
Contigs	273	259
Contig N50	11,391,945	11,391,945
Contig L50	11	11
Contig L90	55	54
QV	47.6019	47.6395
Kmer compl.	77.695	77.3259
BUSCO sing.	88.6%	97.9%
BUSCO dupl.	1.4%	0.9%
BUSCO frag.	8.5%	0.3%
BUSCO miss.	1.6%	0.9%

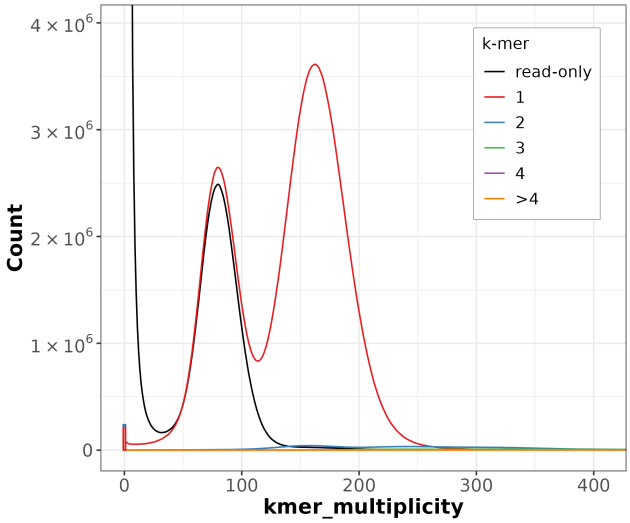
Warning! BUSCO versions or lineage datasets are not the same across results:
BUSCO: 5.8.2 (euk_genome_met, metaeuk) / Lineage: lophotrochozoa_odb12 (genomes:75, BUSCOs:1252)
BUSCO: 6.0.0 (euk_genome_min, miniprot) / Lineage: lophotrochozoa_odb12 (genomes:75, BUSCOs:1252)

HiC contact map of curated assembly

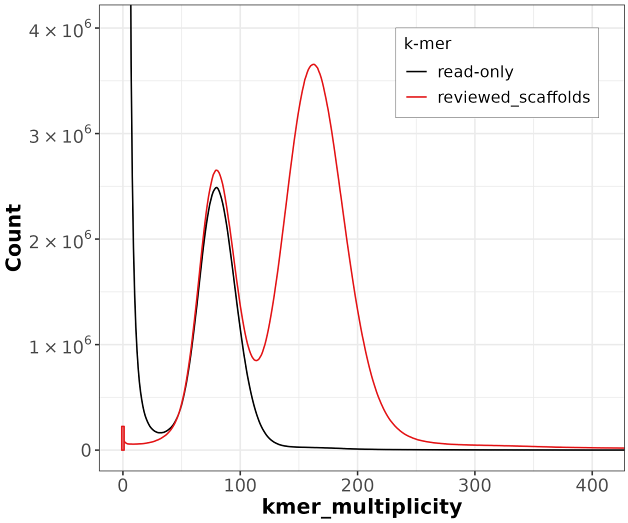


collapsed [LINK]

K-mer spectra of curated assembly

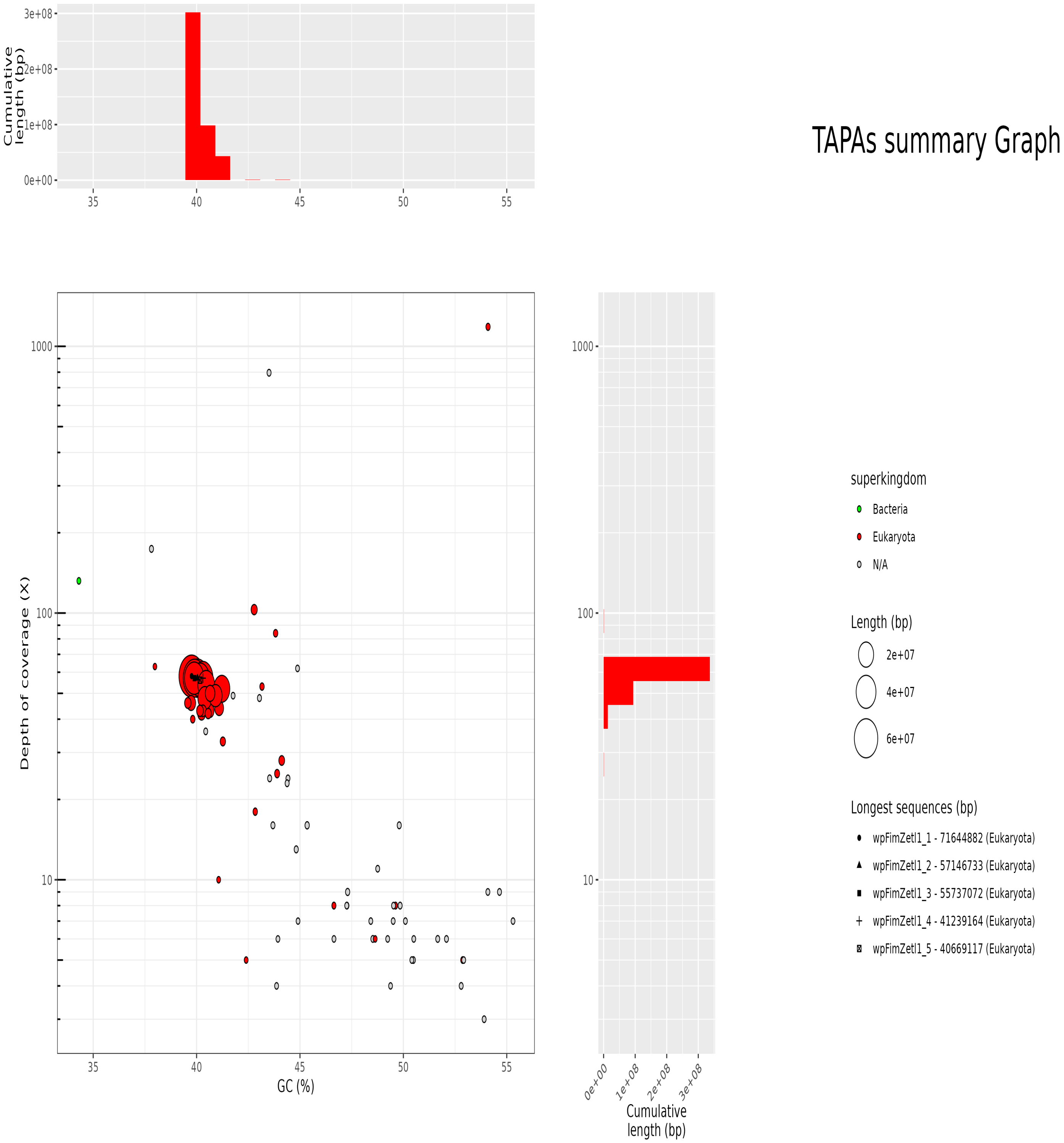


Distribution of k-mer counts per copy numbers found in asm



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	55	125

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: Arnaud Couloux
Affiliation: Genoscope

Date and time: 2025-10-28 06:45:58 CET