ERGA Assembly Report

v24.10.15

Tags: ATLASea[INVALID TAG]

TxID	154632
ToLID	**xgTylPerv1**
Species	Tylodina perversa
Class	Gastropoda
Order	Umbraculida

Genome Traits	Expected	Observed
Haploid size (bp)	1,360,279,174	1,541,985,296
Haploid Number	5 (source: ancestor)	18
Ploidy	2 (source: ancestor)	2
Sample Sex	Unknown	Unknown

EBP metrics summary and curation notes

Obtained EBP quality metric for collapsed: 6.7.Q41

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
. Assembly length loss > 3% for collapsed
. Not 90% of assembly in chromosomes for collapsed

Curator notes

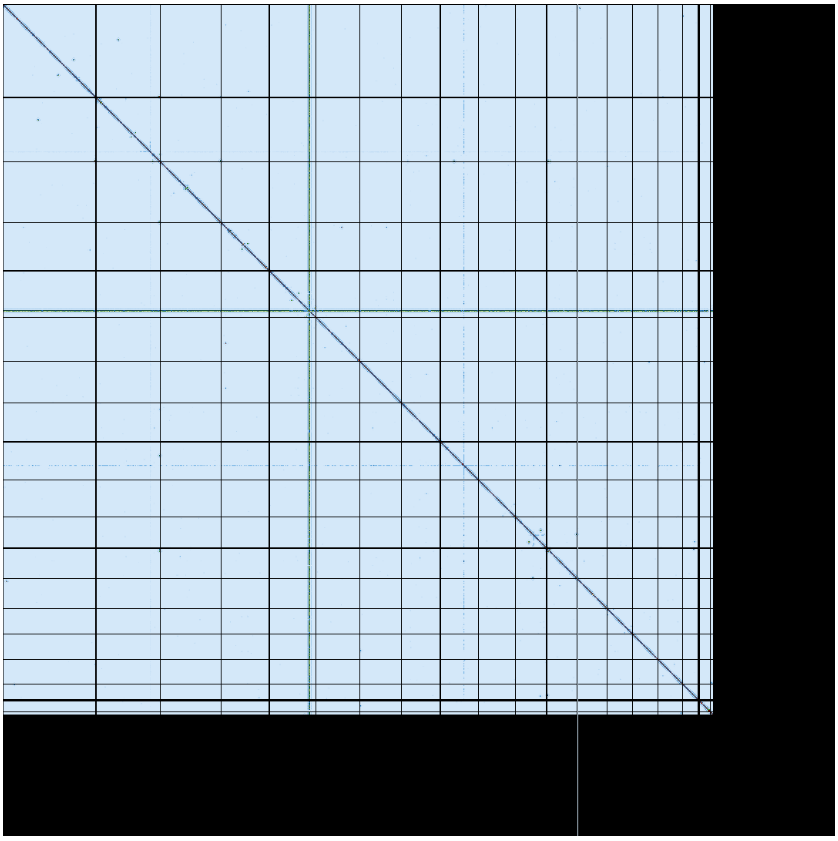
. Interventions/Gb: 152
. Contamination notes: ""
. Other observations: "The assembly of Tylodina perversa (xgTylPerv1) is based on 79X PacBio data and 204X Arima Hi-C data generated as part of the ATLASea programme (https://www.atlasea.fr). The assembly process included the following steps: initial 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, 257 contigs were identified as contaminants (bacterial and archaeal) , totaling 1.7 Mb (with the largest being 0.6 Mb) . Additionally, 3105 regions totaling 213 Mb (with the largest being 2.8 Mb) were identified as haplotypic duplications and removed. The mitochondrial genome was assembled using ptGAUL. Finally, the primary assembly was analyzed and manually improved using Pretext. During manual curation, 93 haplotypic regions and 6 contaminant sequences were removed, totaling 133.625 Mb and 0.649 Mb, respectively (with the largest being 6.7 Mb and 0.538Mb). 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	1,680,739,914	1,541,985,296
GC %	41.07	41.04
Gaps/Gbp	404.58	427.37
Total gap bp	68,000	75,000
Scaffolds	9,428	5,142
Scaffold N50	27,464,092	71,060,464
Scaffold L50	18	8
Scaffold L90	371	377
Contigs	10,108	5,801
Contig N50	3,228,000	3,565,190
Contig L50	130	108
Contig L90	923	933
QV	40.9031	41.2451
Kmer compl.	69.6949	65.9292
BUSCO sing.	81.7%	93.0%
BUSCO dupl.	9.5%	2.7%
BUSCO frag.	6.7%	3.0%
BUSCO miss.	2.1%	1.3%

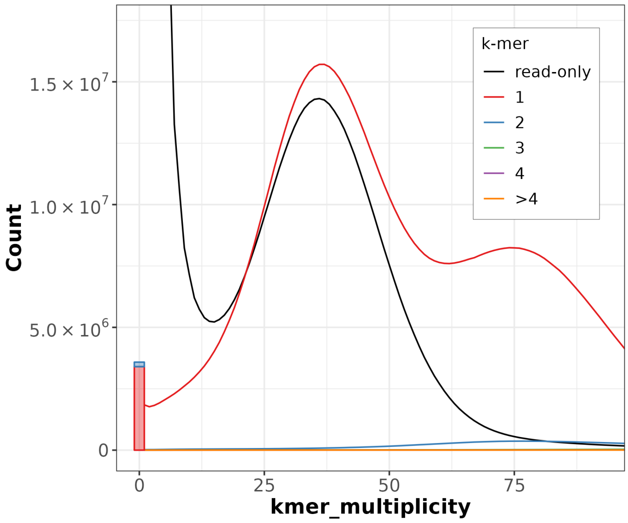
Warning! BUSCO versions or lineage datasets are not the same across results:
BUSCO: 5.8.2 (euk_genome_met, metaeuk) / Lineage: mollusca_odb12 (genomes:36, BUSCOs:4421)
BUSCO: 6.0.0 (euk_genome_min, miniprot) / Lineage: mollusca_odb12 (genomes:36, BUSCOs:4421)

HiC contact map of curated assembly

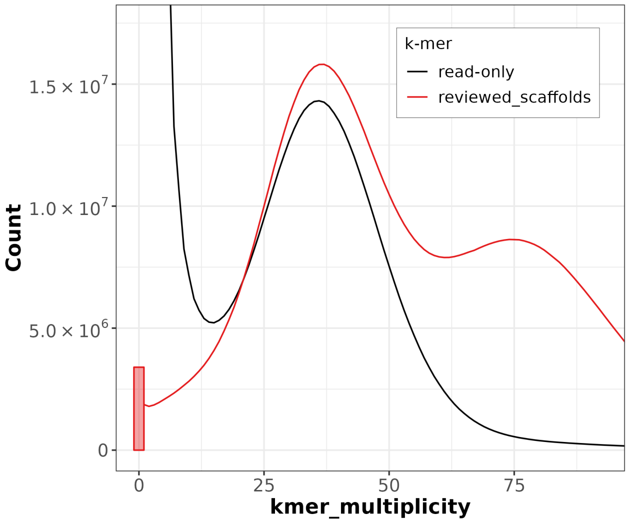


collapsed [LINK]

K-mer spectra of curated assembly

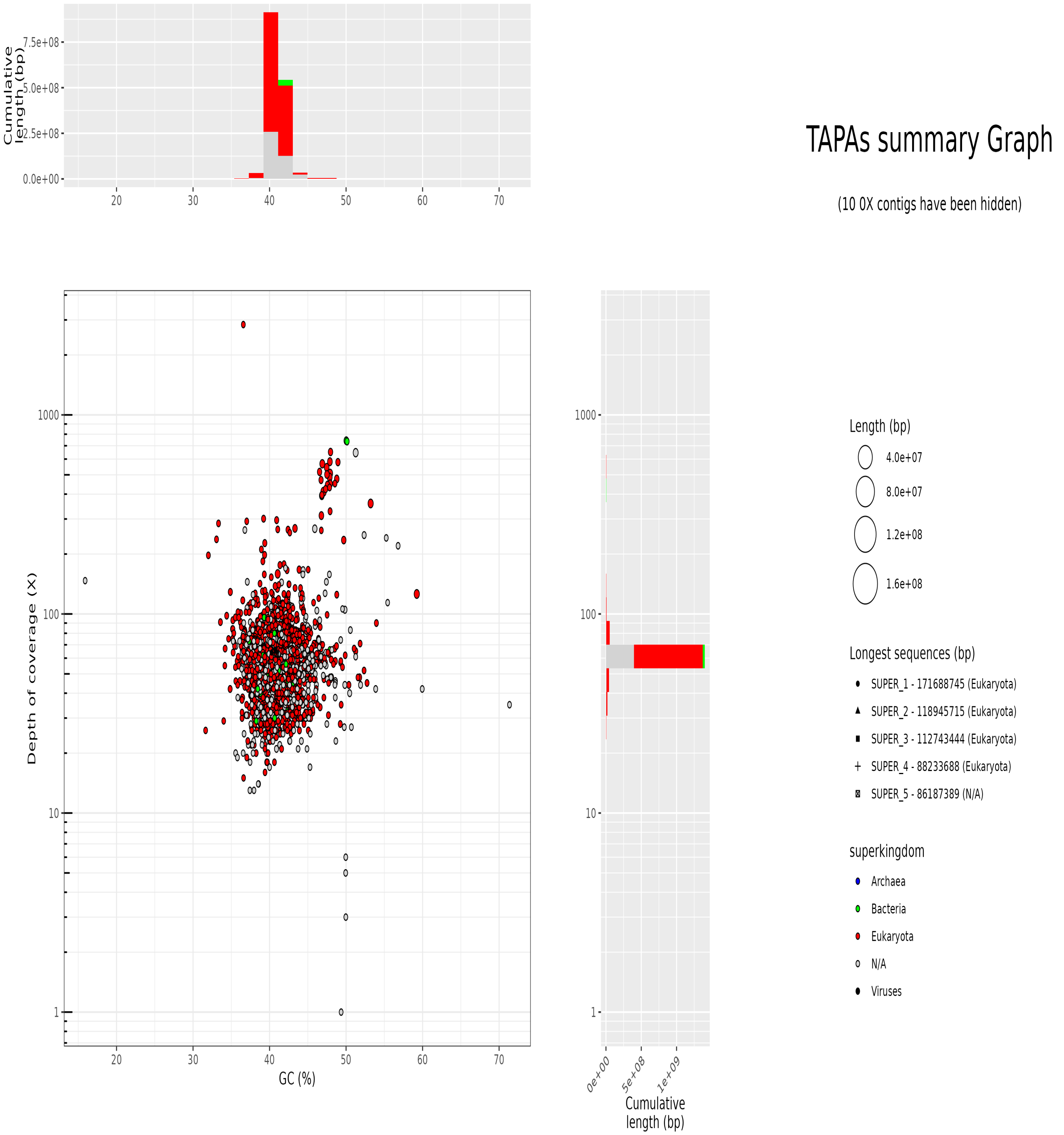


Distribution of k-mer counts per copy numbers found in asm



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	79	204

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: Sophie Layac
Affiliation: Genoscope

Date and time: 2025-10-25 03:03:23 CEST