ERGA Assembly Report

v24.10.15

Tags: ATLASea[INVALID TAG]

TxID	195260
ToLID	**wuOphNegl1**
Species	Ophelia neglecta
Class	Polychaeta
Order	NA

Genome Traits	Expected	Observed
Haploid size (bp)	178,633,194	762,351,503
Haploid Number	9 (source: ancestor)	8
Ploidy	2 (source: ancestor)	2
Sample Sex	Unknown	Unknown

EBP metrics summary and curation notes

Obtained EBP quality metric for collapsed: 6.8.Q35

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

. Observed Haploid size (bp) has >20% difference with Expected
. Observed Haploid Number is different from Expected

. QV value is less than 40 for collapsed
. Kmer completeness value is less than 90 for collapsed
. More than 1000 gaps/Gbp for collapsed

Curator notes

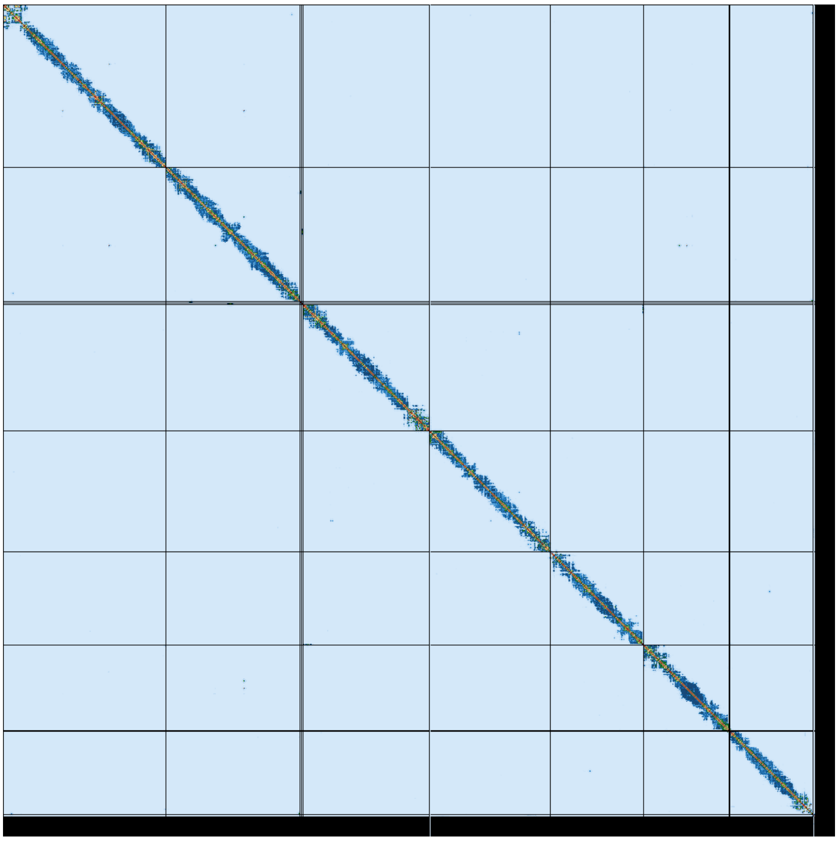
. Interventions/Gb: 295
. Contamination notes: ""
. Other observations: "The assembly of Ophelia neglecta (wuOphNegl1) is based on 35X PacBio data and 25X Arima Hi-C data generated as part of the ATLASea programme (https://www.atlasea.fr). The assembly process included the following steps: initial PacBio 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, 148 contigs were identified as contaminants (bacterial), totaling 7 Mb (with the largest being 1.7 Mb). Additionally, 1041 regions totaling 63 Mb (with the largest being 0.4 Mb) were identified as haplotypic duplications and removed. The mitochondrial genome was assembled using OATK. Finally, the primary assembly was analyzed and manually improved using Pretext. During manual curation, 44 haplotypic regions and 4 contaminant sequences were removed, totaling 11.3 Mb and 166 kb, respectively (with the largest being 0.9 Mb and 102 kb). 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	775,104,726	762,351,503
GC %	39.23	39.19
Gaps/Gbp	961.16	1,015.28
Total gap bp	74,500	86,600
Scaffolds	370	223
Scaffold N50	111,741,015	115,945,607
Scaffold L50	4	3
Scaffold L90	7	7
Contigs	1,115	997
Contig N50	3,201,000	3,382,390
Contig L50	50	46
Contig L90	370	341
QV	35.1032	35.1955
Kmer compl.	72.9299	72.2795
BUSCO sing.	89.2%	90.2%
BUSCO dupl.	3.0%	1.8%
BUSCO frag.	6.5%	6.6%
BUSCO miss.	1.4%	1.4%

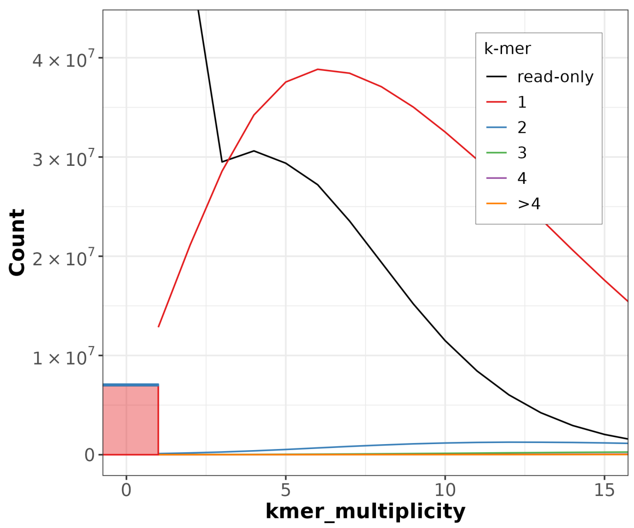
BUSCO: 5.8.2 (euk_genome_met, metaeuk) / Lineage: lophotrochozoa_odb12 (genomes:75, BUSCOs:1252)

HiC contact map of curated assembly

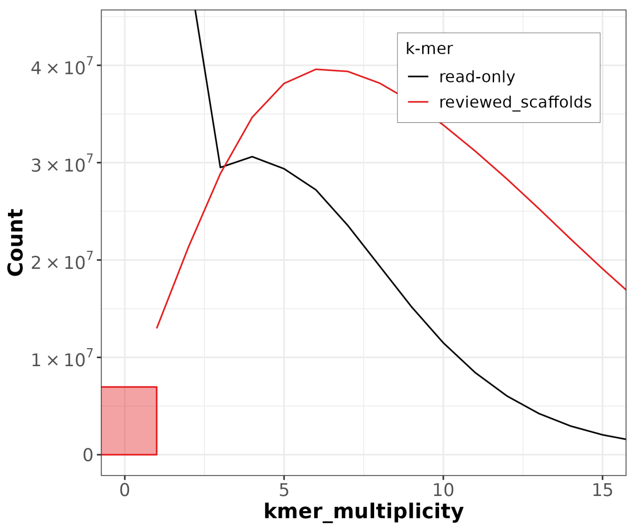


collapsed [LINK]

K-mer spectra of curated assembly

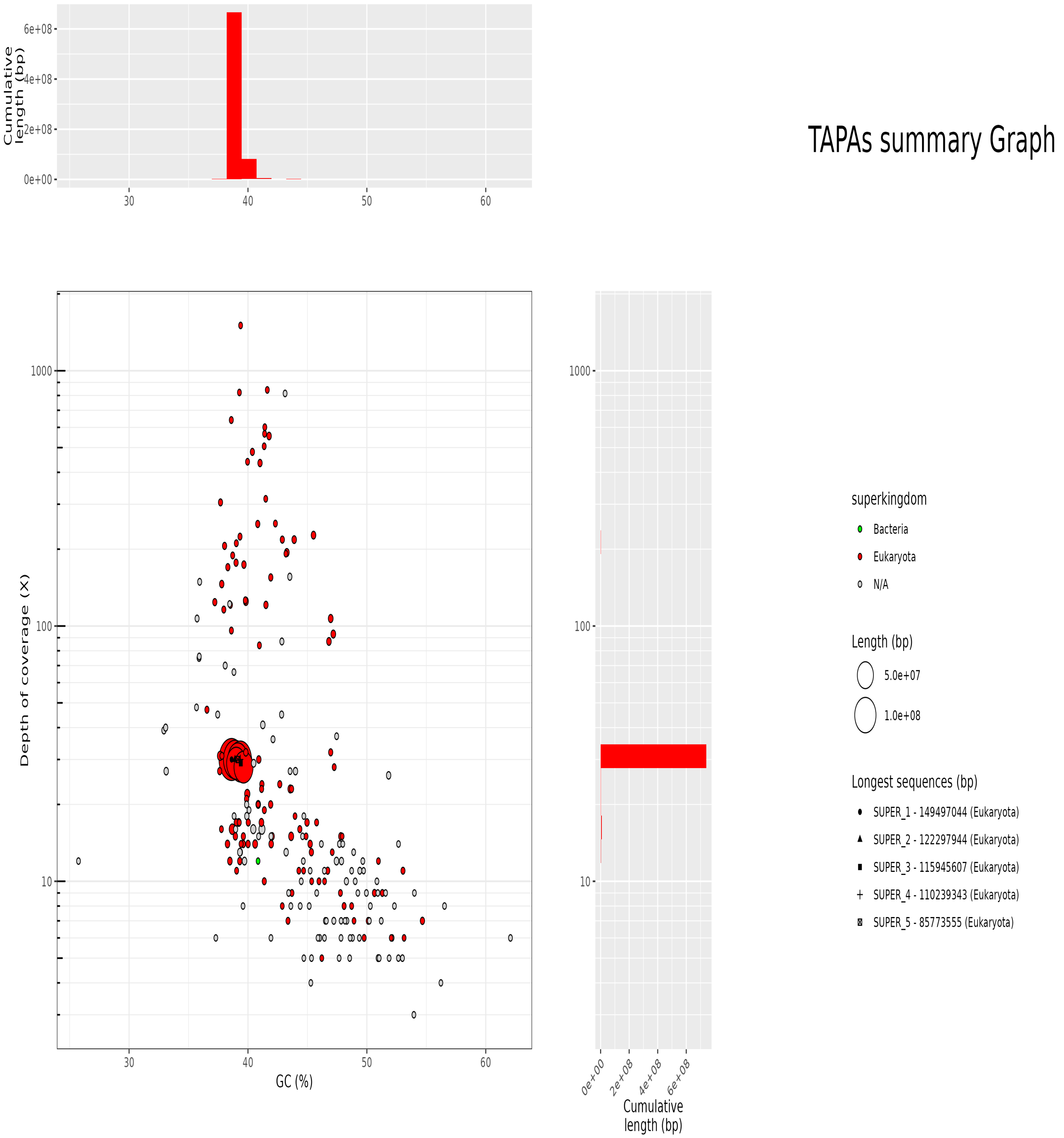


Distribution of k-mer counts per copy numbers found in asm



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	PACBIO Hifi	Arima
Coverage	116	152

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: 2025-05-17 01:56:00 CEST