ERGA Assembly Report

v24.10.15

Tags: ATLASea[INVALID TAG]

TxID	3261357
ToLID	**wtAphMult2**
Species	Aphelochaeta multibranchis
Class	Polychaeta
Order	Terebellida

Genome Traits	Expected	Observed
Haploid size (bp)	603,429,106	629,883,952
Haploid Number	3 (source: ancestor)	13
Ploidy	2 (source: ancestor)	2
Sample Sex	Unknown	Unknown

EBP metrics summary and curation notes

Obtained EBP quality metric for collapsed: 6.7.Q49

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

Curator notes

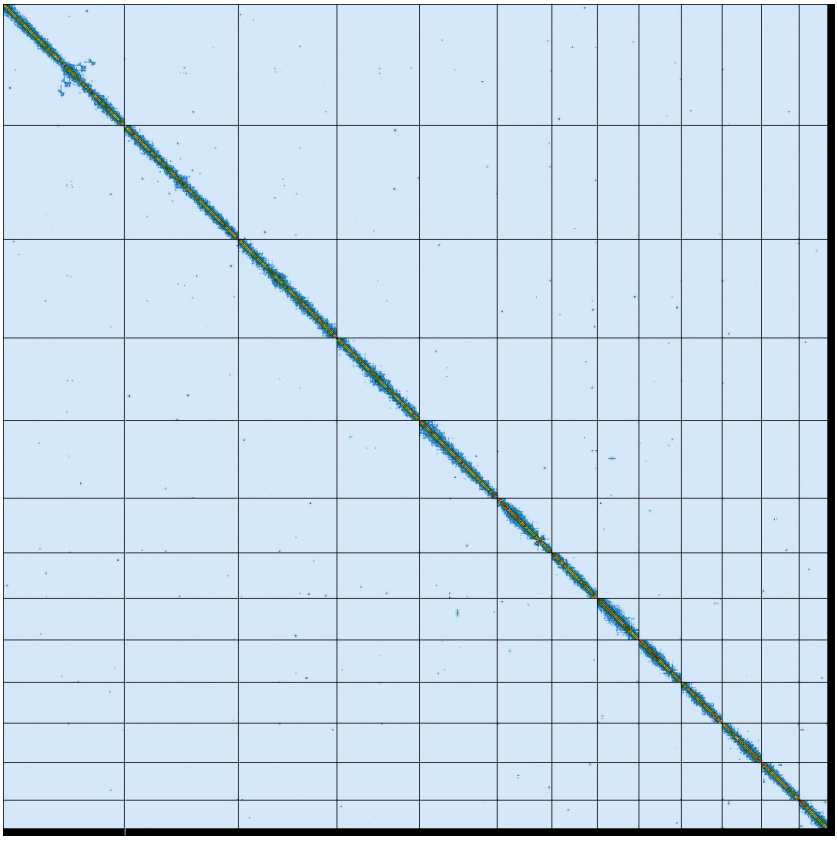
. Interventions/Gb: 260
. Contamination notes: ""
. Other observations: "The assembly of Aphelochaeta multibranchis (wtAphMult2) is based on 60X PacBio data and 369X Arima Hi-C data generated as part of the ATLASea programme (https://www.atlasea.fr). The assembly process included the following steps: 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, 6 contigs were identified as contaminants (bacterial, archaeal, or viral), totaling 0.1 Mb (with the largest being 0.026 Mb). Additionally, 578 regions totaling 67.779 Mb (with the largest being 1.297 Mb) were identified as haplotypic duplications and removed. Mitochondrial genome was assembled using OATK. Finally, the primary assembly was analyzed and manually improved using Pretext. During manual curation, 78 haplotypic regions and 3 contaminant sequences were removed, totaling 52.188Mb and 0.622Mb, respectively (with the largest being 2.636Mb and 0.268Mb). 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	682,705,281	629,883,952
GC %	43.84	43.85
Gaps/Gbp	295.88	304.82
Total gap bp	20,200	26,000
Scaffolds	131	103
Scaffold N50	44,791,636	59,385,535
Scaffold L50	5	5
Scaffold L90	12	11
Contigs	333	295
Contig N50	4,992,146	4,992,146
Contig L50	45	41
Contig L90	135	122
QV	49.7584	49.7799
Kmer compl.	70.7863	67.4562
BUSCO sing.	91.1%	97.6%
BUSCO dupl.	8.1%	1.4%
BUSCO frag.	0.5%	0.5%
BUSCO miss.	0.4%	0.5%

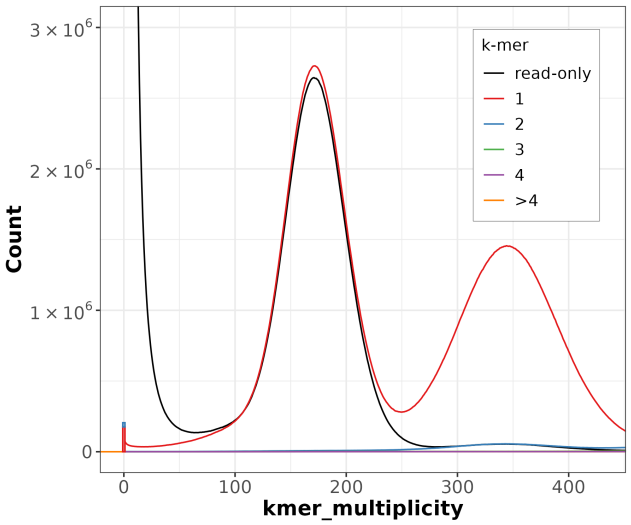
BUSCO: 6.0.0 (euk_genome_min, miniprot) / Lineage: lophotrochozoa_odb12 (genomes:75, BUSCOs:1252)

HiC contact map of curated assembly

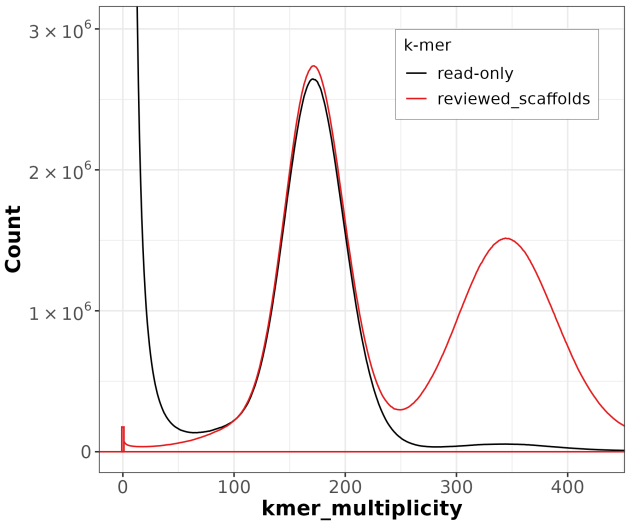


collapsed [LINK]

K-mer spectra of curated assembly

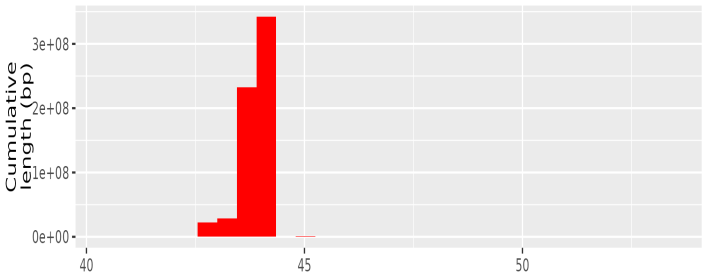


Distribution of k-mer counts per copy numbers found in asm

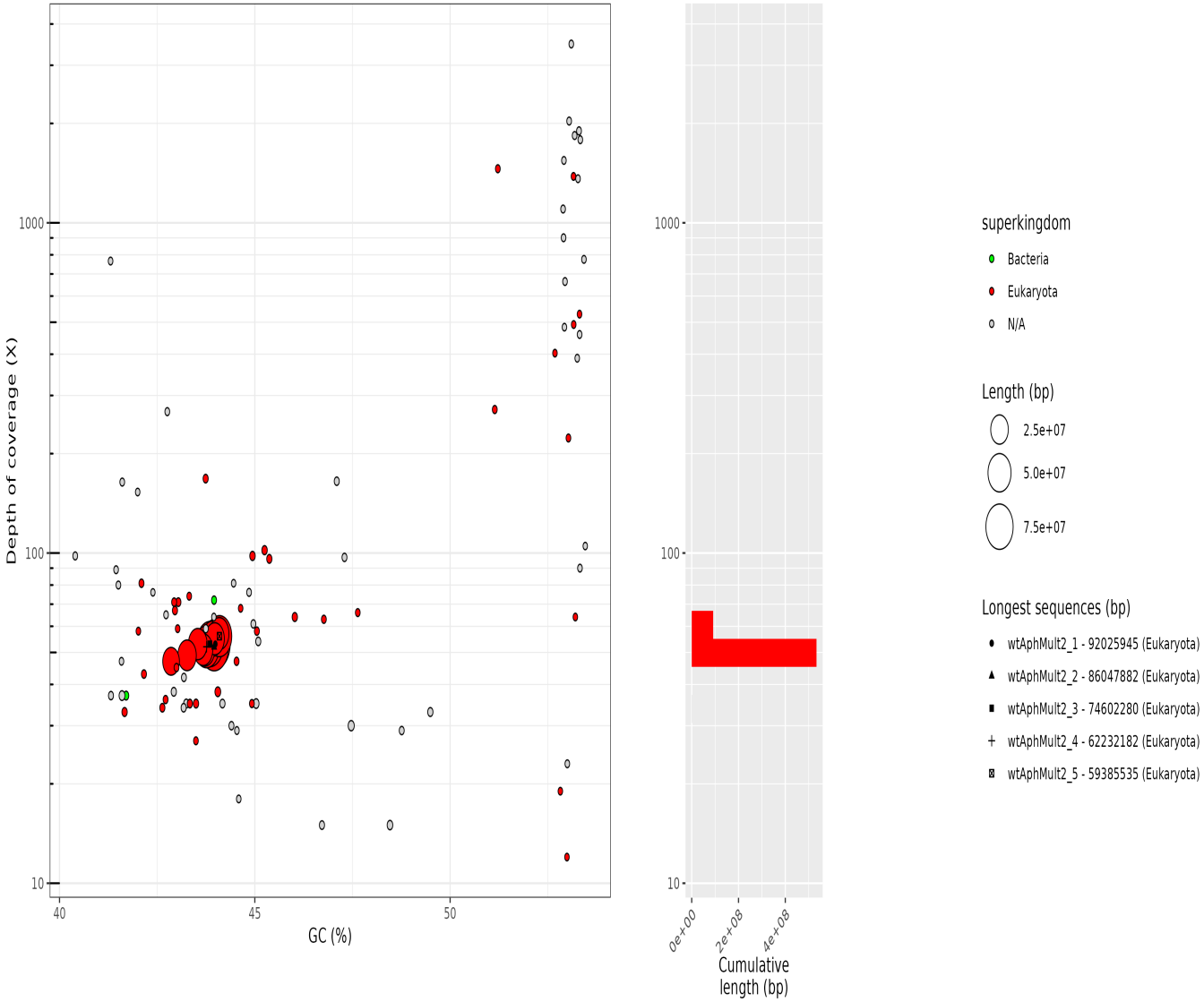


Distribution of k-mer counts coloured by their presence in reads/assemblies

Post-curation contamination screening



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	Long reads	Arima
Coverage	65	400

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: 2026-03-27 04:52:27 CET