ERGA Assembly Report

v24.10.15

Tags: ATLASea[INVALID TAG]

TxID	2733422
ToLID	**wsAnaOrst1**
Species	Anamobaea orstedii
Class	Polychaeta
Order	Sabellida

Genome Traits	Expected	Observed
Haploid size (bp)	1,402,528,639	4,988,074,082
Haploid Number	4 (source: ancestor)	14
Ploidy	2 (source: ancestor)	2
Sample Sex	Unknown	Unknown

EBP metrics summary and curation notes

Obtained EBP quality metric for collapsed: 7.8.Q38

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
. Assembly length loss > 3% for collapsed

Curator notes

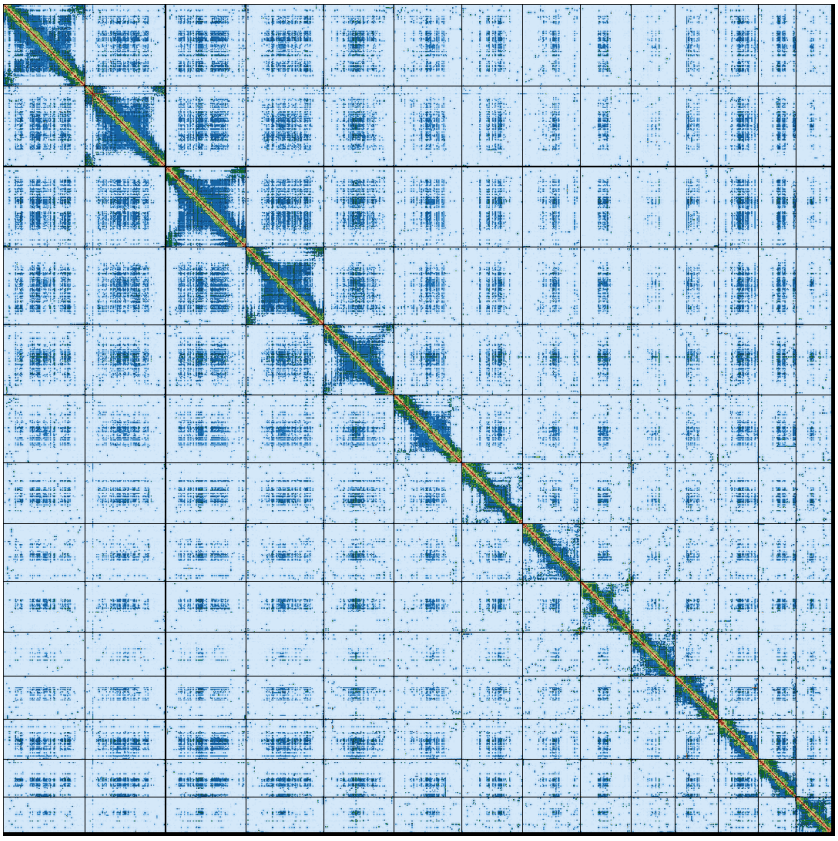
. Interventions/Gb: 51
. Contamination notes: ""
. Other observations: "The assembly of Anamobaea orstedii (wsAnaOrst1) is based on 42X PacBio data and 39X Arima Hi-C data generated as part of the ATLASea programme (https://www.atlasea.fr). The assembly process included the following steps: initial PacBio 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, 15 contigs were identified as contaminants (bacterial, archaeal, or viral), totaling 0.33 Mb (with the largest being 0.049 Mb). Additionally, 1336 regions totaling 247.485 Mb (with the largest being 8.315 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, 120 haplotypic regions were removed, totaling 343.757Mb (with the largest being 26.937Mb). 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	5,331,822,261	4,988,074,082
GC %	34.72	34.73
Gaps/Gbp	58.7	60.74
Total gap bp	31,300	40,700
Scaffolds	329	285
Scaffold N50	429,301,325	409,195,686
Scaffold L50	6	6
Scaffold L90	13	12
Contigs	642	588
Contig N50	27,481,359	28,023,542
Contig L50	61	56
Contig L90	194	178
QV	38.5354	38.5976
Kmer compl.	71.2497	68.3315
BUSCO sing.	94.2%	96.7%
BUSCO dupl.	3.8%	1.0%
BUSCO frag.	1.0%	1.0%
BUSCO miss.	1.0%	1.3%

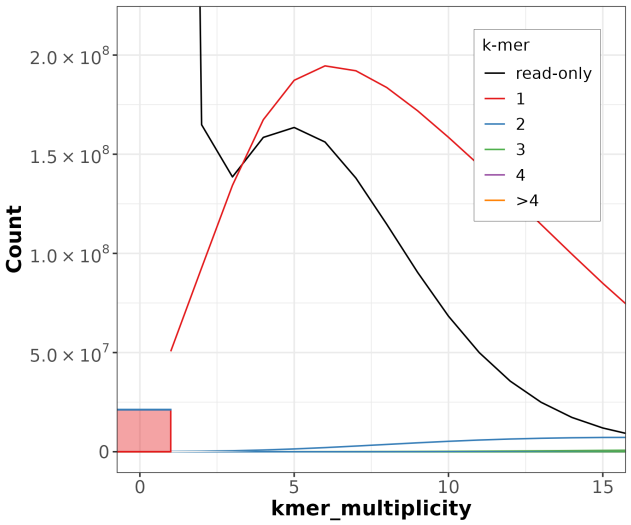
BUSCO: 6.1.0 (euk_genome_min, miniprot) / Lineage: lophotrochozoa_odb12 (genomes:75, BUSCOs:1252)

HiC contact map of curated assembly

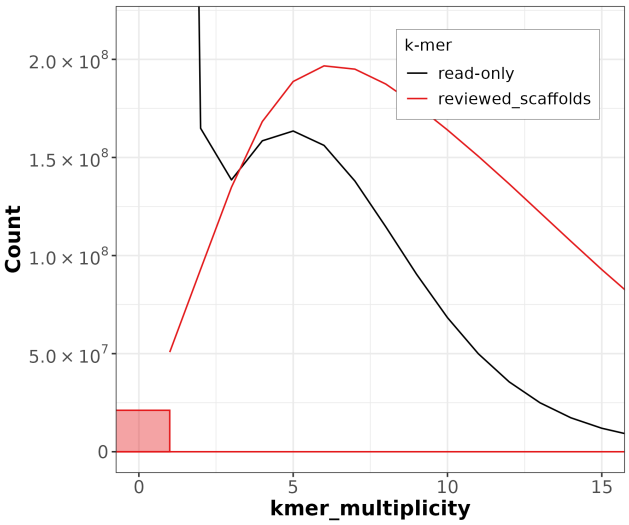


collapsed [LINK]

K-mer spectra of curated assembly

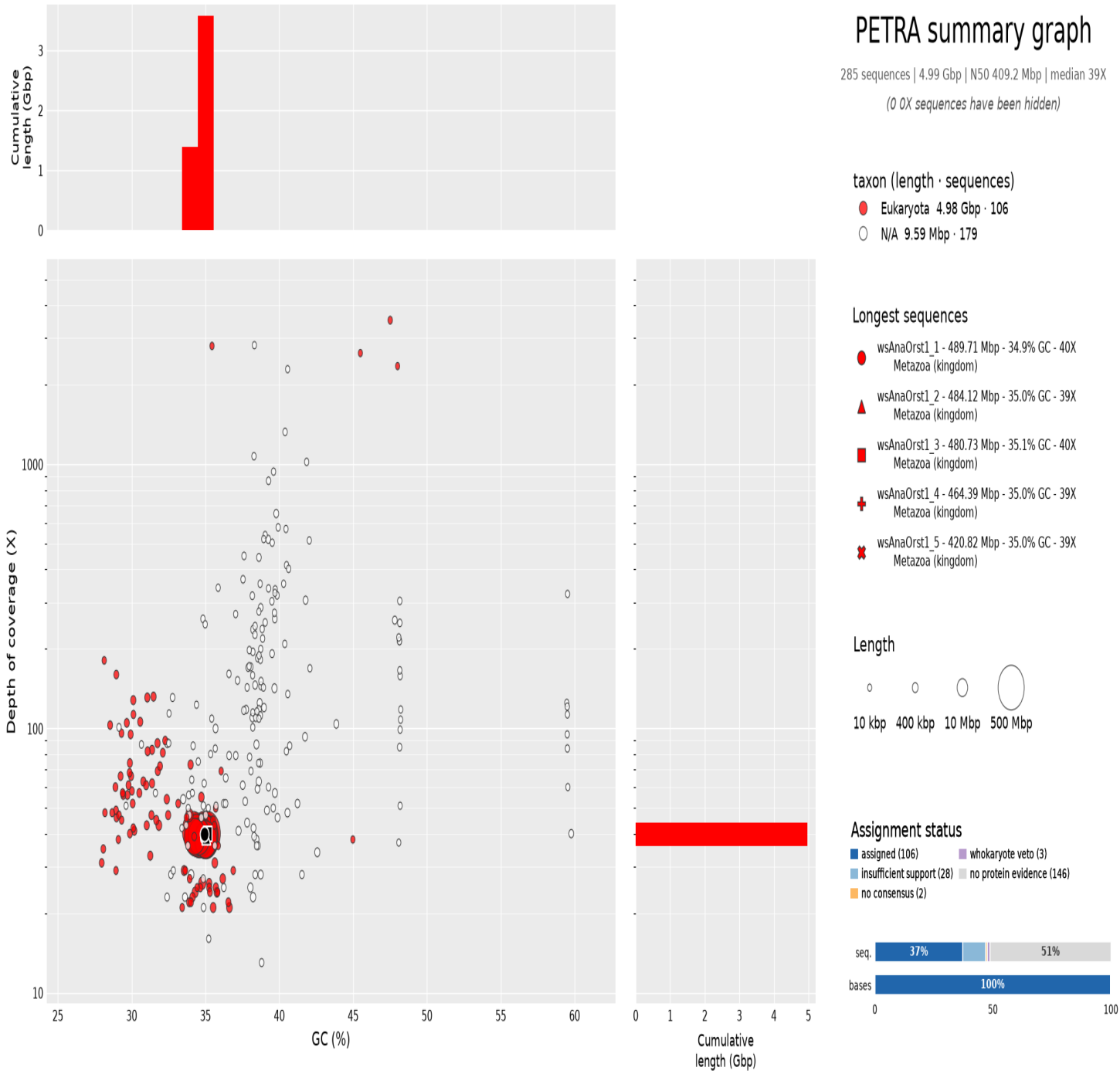


Distribution of k-mer counts per copy numbers found in asm



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	45	41

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:* 0.2.4
  - *key param:* NA
- **PretextView**
  - *ver:* 1.0.7
  - *key param:* NA

Submitter: Jean-Marc Aury
Affiliation: Genoscope

Date and time: 2026-10-07 11:33:04 CEST