

ERGA Assembly Report

v24.10.15

Tags: ATLASea[INVALID TAG]

TxID	2006579
ToLID	wpEutOcul1.1
Species	Euthalenessa
Class	Polychaeta
Order	Phyllodocida

Genome Traits	Expected	Observed
Haploid size (bp)	1,167,406,580	1,206,749,674
Haploid Number	10 (source: ancestor)	12
Ploidy	2 (source: ancestor)	2
Sample Sex	Unknown	Unknown

EBP metrics summary and curation notes

Obtained EBP quality metric for collapsed: 7.7.Q61

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

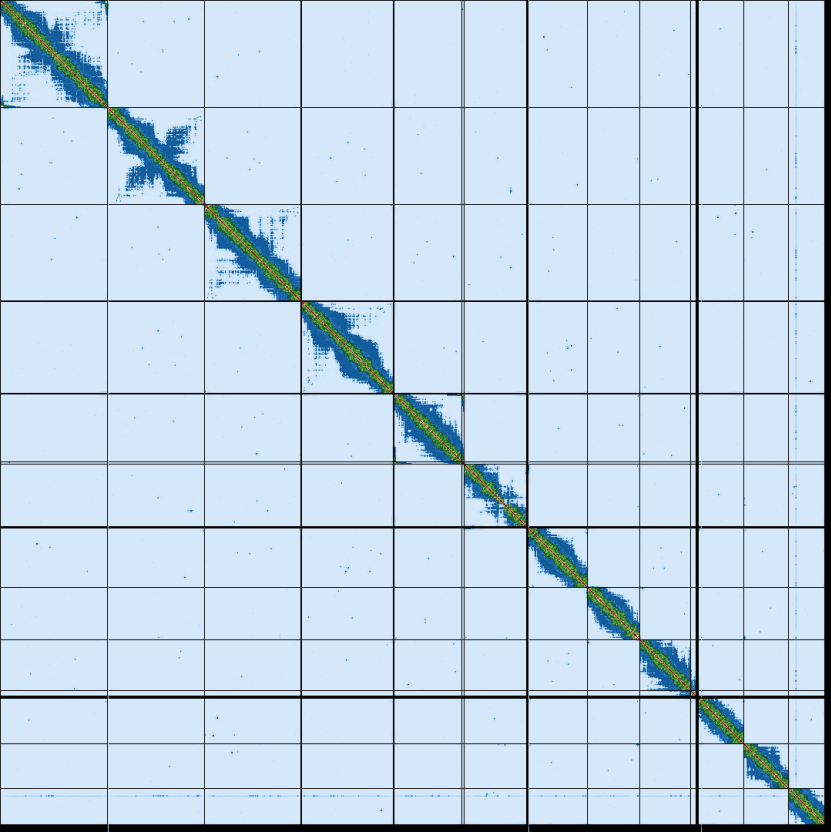
- . Interventions/Gb: 110
- . Contamination notes: ""
- . Other observations: "The assembly of *Euthalenessa oculata* (wpEutOcul1.1) is based on 38X PacBio data and Arima Hi-C data generated as part of the ATLASea programme (<https://www.atlasea.fr>). The assembly process included the following steps: 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, 51 contigs were identified as contaminants (bacterial, archaeal, or viral), totaling 1.9 Mb (with the largest being 0.25 Mb). Additionally, 823 regions totaling 192 Mb (with the largest being 6.69 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, 45 haplotypic regions were removed, totaling 42.3 Mb (with the largest being 3.75 Mb). Chromosome-scale scaffolds confirmed by Hi-C data were named in order of size. Chromosome 9 contains several unlocalized contigs, with the largest being 7.7 Mb. It may correspond to the end of the chromosome, but contigs are hard to organize in this region "

Quality metrics table

Metrics	Pre-curation collapsed	Curated collapsed
Total bp	1,249,489,757	1,206,749,674
GC %	41.06	41.06
Gaps/Gbp	351.34	343.9
Total gap bp	43,900	44,700
Scaffolds	392	285
Scaffold N50	103,062,048	98,633,779
Scaffold L50	5	5
Scaffold L90	11	11
Contigs	831	700
Contig N50	11,266,701	13,944,052
Contig L50	30	25
Contig L90	137	109
QV	46.9845	61.6959
Kmer compl.	72.6226	71.9797
BUSCO sing.	92.0%	93.8%
BUSCO dupl.	2.8%	1.0%
BUSCO frag.	1.8%	1.8%
BUSCO miss.	3.4%	3.4%

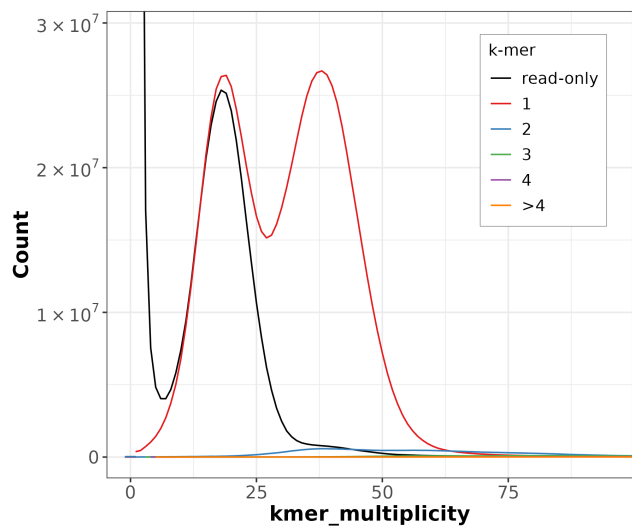
BUSCO: 5.4.3 (euk_genome_met, metaeuk) / Lineage: metazoa_odb10 (genomes:65, BUSCOs:954)

HiC contact map of curated assembly

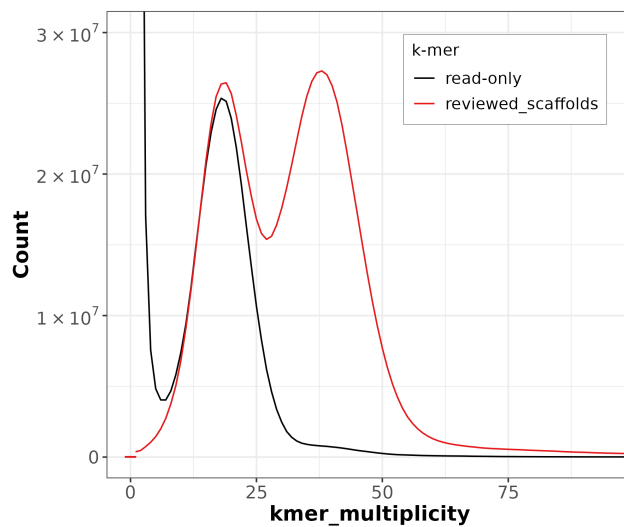


collapsed [\[LINK\]](#)

K-mer spectra of curated assembly

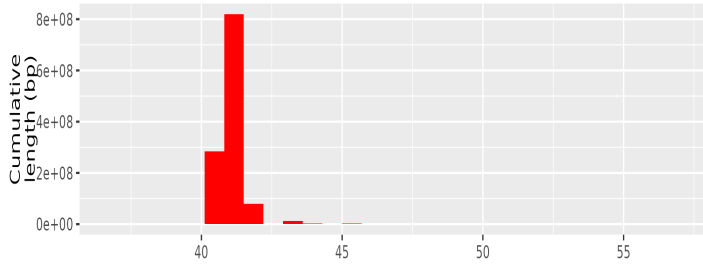


Distribution of k-mer counts per copy numbers found in asm

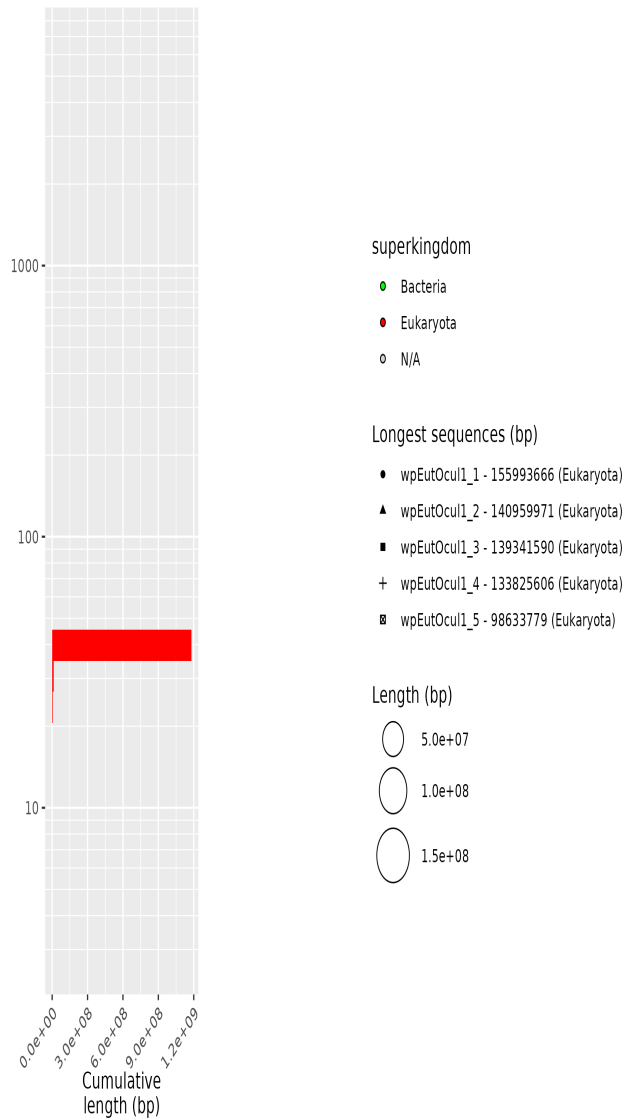
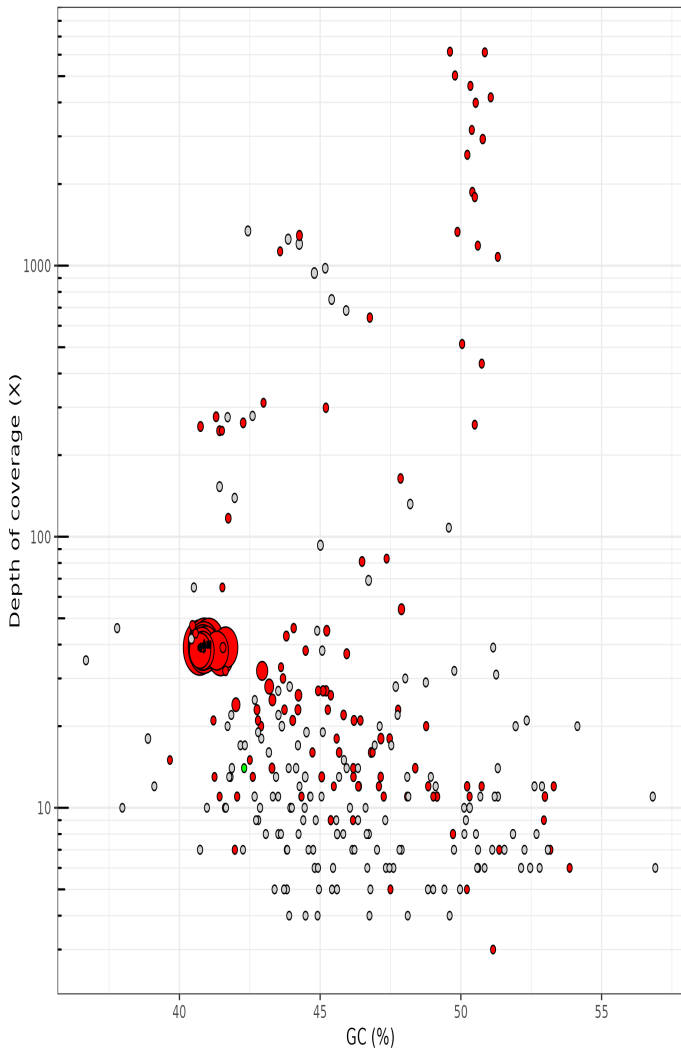


Distribution of k-mer counts coloured by their presence in reads/assemblies

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	PACBIO Hifi	Arima
Coverage	38	133

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: Jean-Marc Aury

Affiliation: Genoscope

Date and time: 2025-02-04 14:20:47 CET