

ERGA Assembly Report

v24.10.15

Tags: ATLASea[INVALID TAG]

TxID	1548796
ToLID	jaParCine1
Species	Paranemonia cinerea
Class	Anthozoa
Order	Actiniaria

Genome Traits	Expected	Observed
Haploid size (bp)	308,559,283	326,819,555
Haploid Number	6 (source: ancestor)	19
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

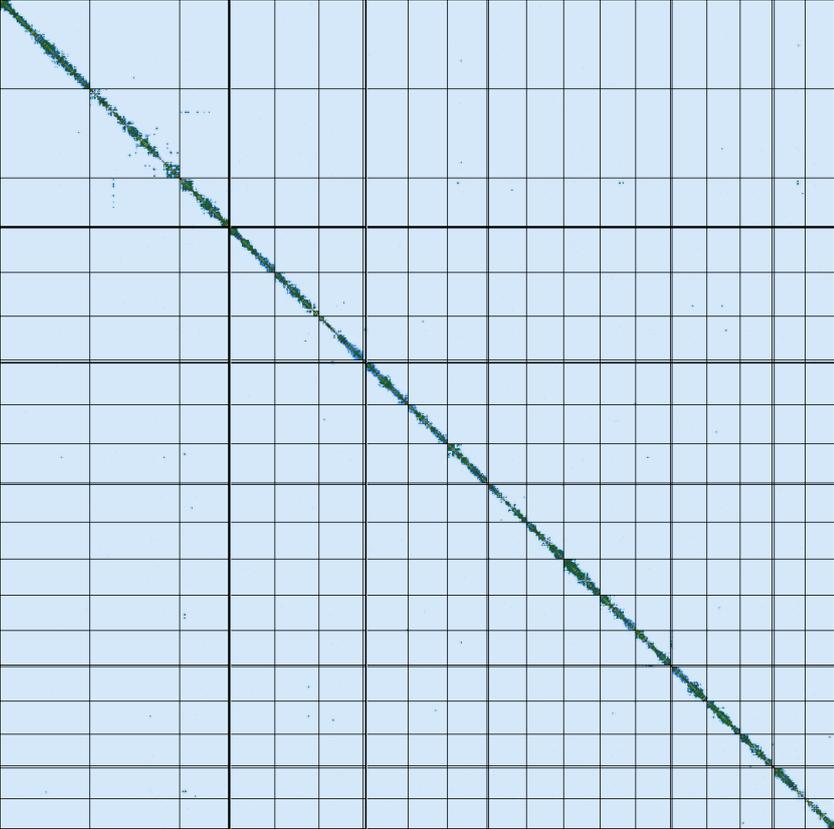
. Interventions/Gb: 329
. Contamination notes: ""
. Other observations: "The assembly of *Paranemonia cinerea* (jaParCine1) is based on 59X PacBio data and 162 X Arima Hi-C data generated as part of the ATLASea programme (<https://www.atlasea.fr>). The assembly process included the following steps: initial 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, 1 contig was identified as contaminants (bacterial), totaling 2 321 pb . Additionally, 110 regions regions totaling 8 Mb (with the largest being 567 kb) 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 57 haplotypic regions and 18 contaminant sequences were removed, totaling 20 Mb and 632 kb , respectively (with the largest being 5 Mb and 101 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	347,186,041	326,819,555
GC %	37.75	37.74
Gaps/Gbp	69.13	70.38
Total gap bp	2,400	3,700
Scaffolds	80	38
Scaffold N50	17,267,776	15,678,110
Scaffold L50	7	8
Scaffold L90	17	17
Contigs	104	61
Contig N50	13,074,213	13,074,213
Contig L50	11	11
Contig L90	28	28
QV	46.3726	61.6524
Kmer compl.	79.4256	77.4761
BUSCO sing.	92.6%	95.2%
BUSCO dupl.	3.7%	0.8%
BUSCO frag.	2.0%	2.0%
BUSCO miss.	1.7%	2.0%

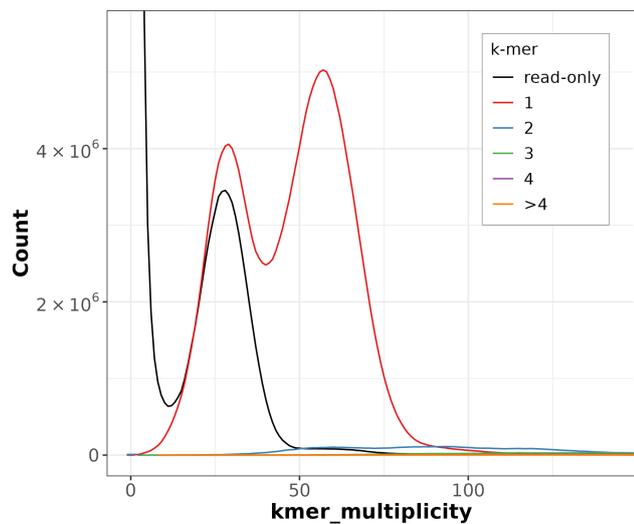
BUSCO: 5.4.3 (euk_genome_met, metaeuk) / Lineage: metazoa_odb10 (genomes:65, BUSCOs:954)

HiC contact map of curated assembly

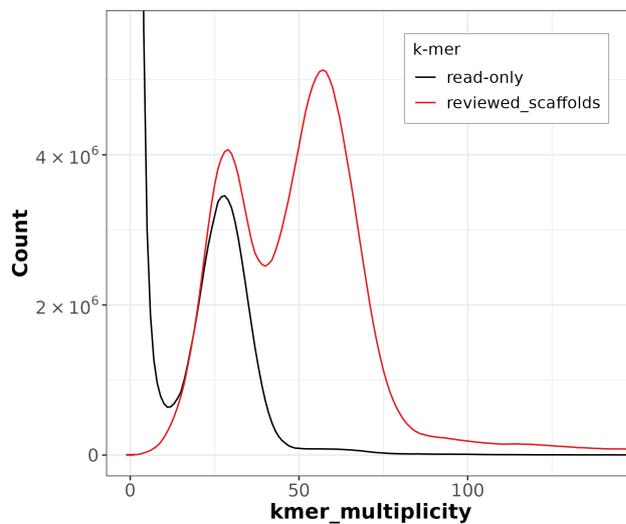


collapsed [\[LINK\]](#)

K-mer spectra of curated assembly

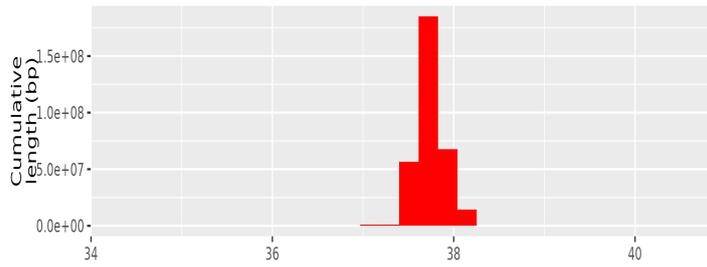


Distribution of k-mer counts per copy numbers found in asm

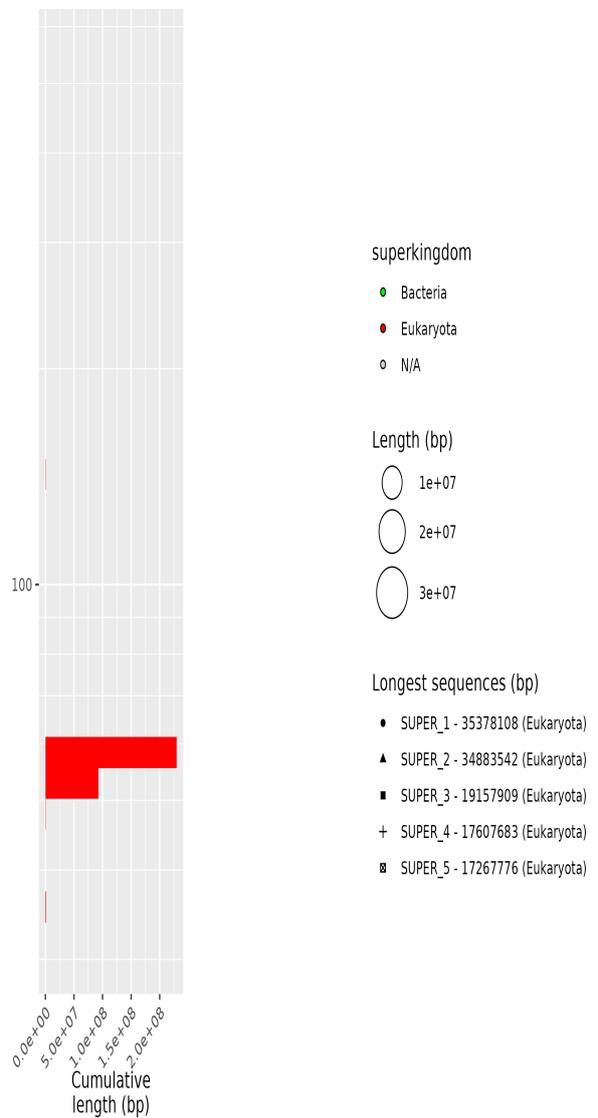
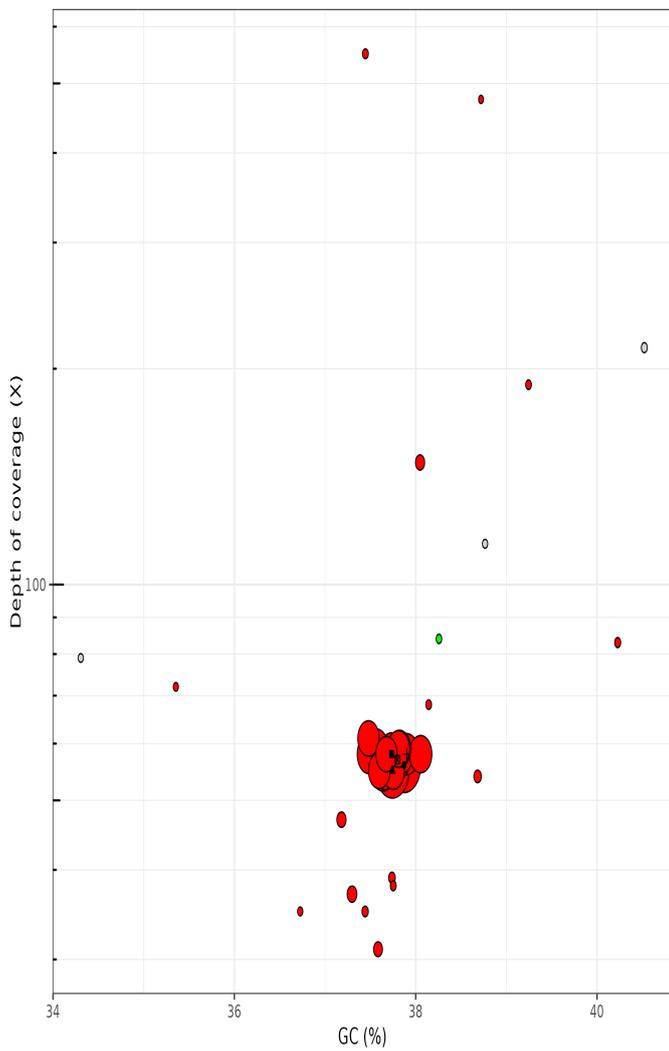


Distribution of k-mer counts coloured by their presence in reads/assemblies

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	59	162

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: 2025-04-08 22:58:20 CEST