ERGA Assembly Report

v24.10.15

Tags: ATLASea[INVALID TAG]

TxID	13833
ToLID	**eoOphLong1**
Species	Ophioderma
Class	Ophiuroidea
Order	Amphilepidida

Genome Traits	Expected	Observed
Haploid size (bp)	2,754,979,138	3,243,334,633
Haploid Number	240 (source: descendant)	25
Ploidy	24 (source: descendant)	2
Sample Sex	Unknown	Unknown

EBP metrics summary and curation notes

Obtained EBP quality metric for collapsed: 6.8.Q45

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

. Observed Haploid Number is different from Expected
. Observed Ploidy is different from Expected
. Kmer completeness value is less than 90 for collapsed

Curator notes

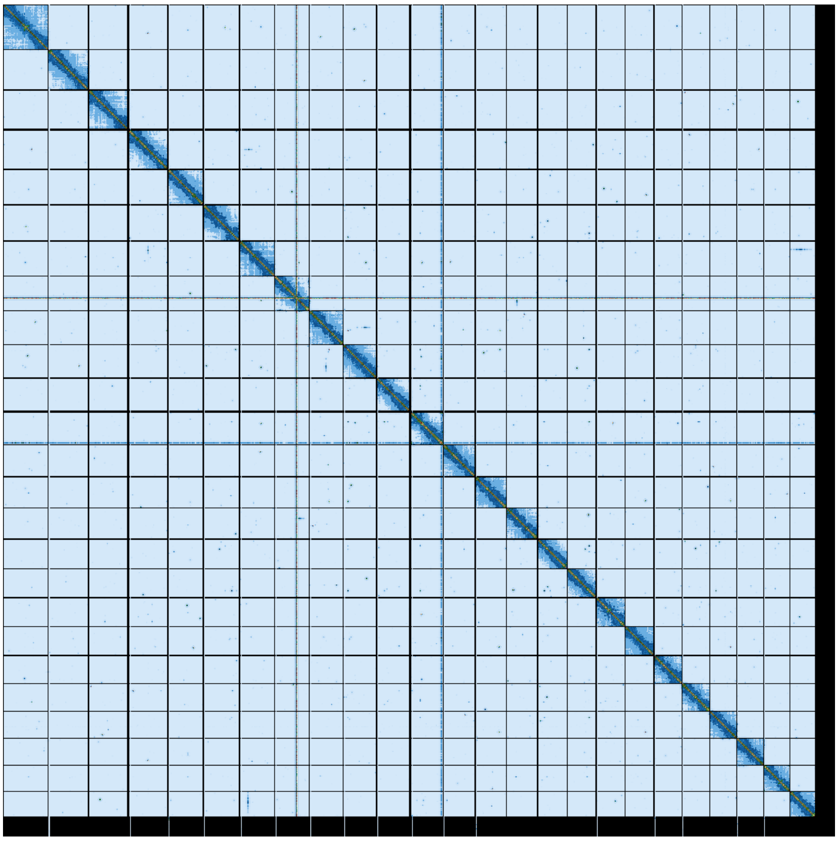
. Interventions/Gb: 68
. Contamination notes: ""
. Other observations: "The assembly of Ophioderma laugicaudum (WORMS accepted name of Ophioderma laugicauda eoOphLong1) is based on 38X PacBio data and Arima HighCoverage Hi-C data generated as part of the ATLASea programme (https://www.atlasea.fr).The 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, 9 contigs were identified as contaminants (bacterial, archaeal, or viral), totaling 325Kb (with the largest being 73 Kb). Additionally, 2,771 regions totaling 419 Mb (with the largest being 3.4 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, 71 haplotypic regions were removed,totaling 54 Mb (with the largest being 2.2 Mb) "

Quality metrics table

Metrics	Pre-curation collapsed	Curated collapsed
Total bp	3,298,763,186	3,243,334,633
GC %	37.78	37.78
Gaps/Gbp	724.82	736.9
Total gap bp	239,100	248,200
Scaffolds	1,401	933
Scaffold N50	119,052,542	124,730,492
Scaffold L50	13	12
Scaffold L90	25	23
Contigs	3,792	3,323
Contig N50	2,538,273	2,604,949
Contig L50	389	376
Contig L90	1,423	1,374
QV	57.4277	45.1717
Kmer compl.	72.0868	72.135
BUSCO sing.	94.9%	95.3%
BUSCO dupl.	3.1%	2.4%
BUSCO frag.	2.0%	2.4%
BUSCO miss.	0.0%	0.1%

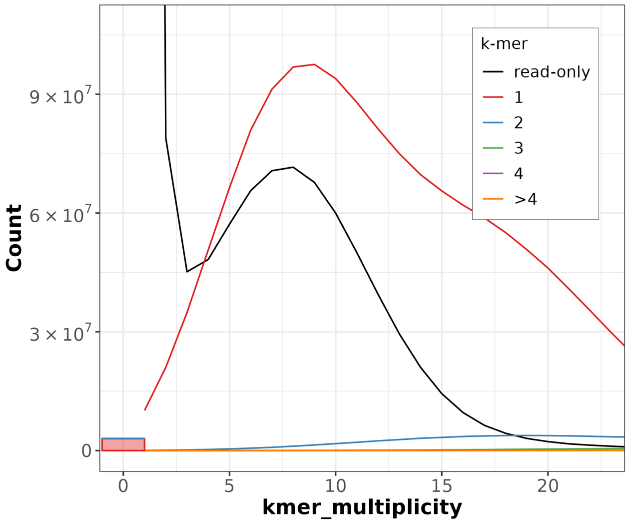
BUSCO: 5.4.3 (euk_genome_met, metaeuk) / Lineage: eukaryota_odb10 (genomes:70, BUSCOs:255)

HiC contact map of curated assembly

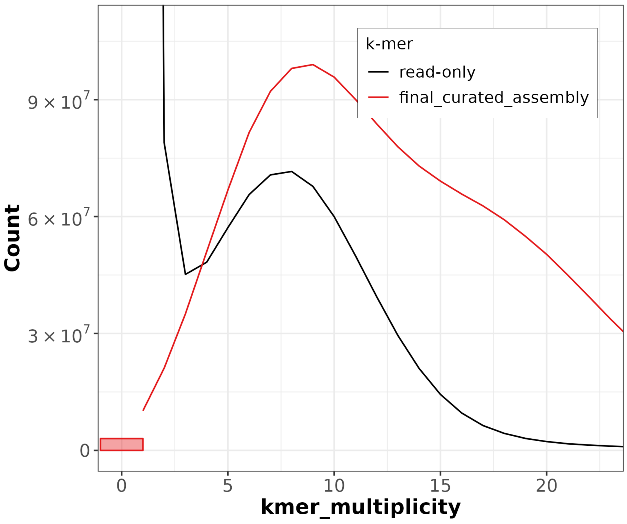


collapsed [LINK]

K-mer spectra of curated assembly

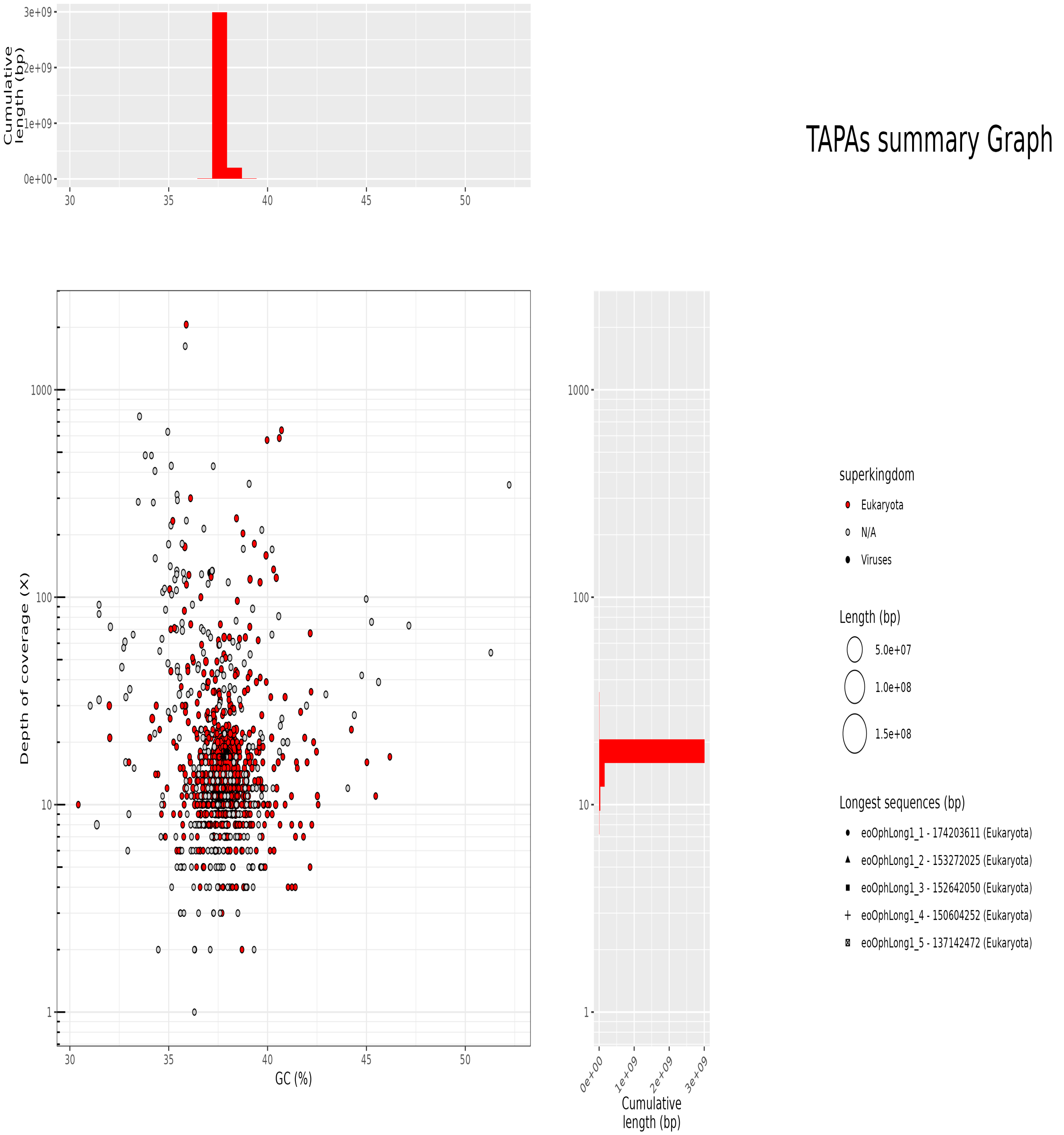


Distribution of k-mer counts per copy numbers found in asm



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	38	173

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: Caroline Belser
Affiliation: Genoscope

Date and time: 2025-02-01 18:29:17 CET