ERGA Assembly Report

v24.10.15

Tags: ATLASea[INVALID TAG]

TxID	31178
ToLID	**eeEchMath1**
Species	Echinometra mathaei
Class	Echinoidea
Order	Camarodonta

Genome Traits	Expected	Observed
Haploid size (bp)	849,235,818	875,599,765
Haploid Number	16 (source: ancestor)	21
Ploidy	2 (source: ancestor)	2
Sample Sex	Unknown	Unknown

EBP metrics summary and curation notes

Obtained EBP quality metric for collapsed: 7.7.Q46

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

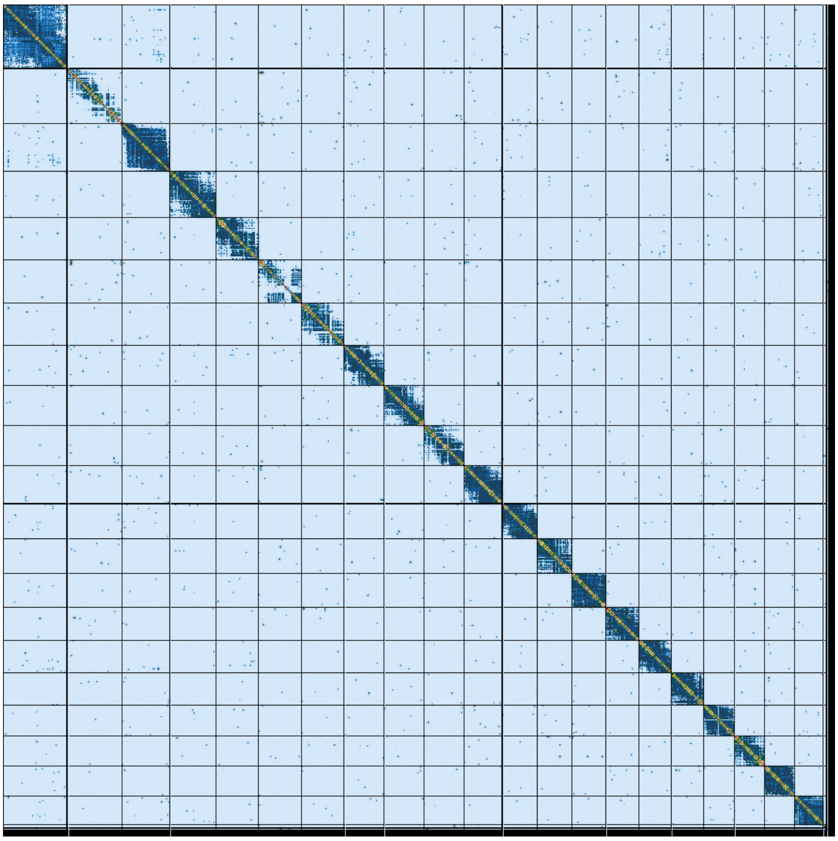
. Interventions/Gb: 169
. Contamination notes: ""
. Other observations: "The assembly of Echinometra mathaei (eeEchMath1) is based on 56X PacBio data and 294X Arima Hi-C data generated as part of the ATLASea programme (https://www.atlasea.fr).The assembly process included the following steps: initial PacBio assembly generation with Hifiasm in Hi-C integrated assembly mode, removal of contaminant sequences using Context and Hi-C based scaffolding with YaHS. In total, 921 contigs were identified as contaminants (bacterial, archaeal, or viral), totaling 108 Mb (with the largest being 0.919 Mb). The mitochondrial genome was assembled using OATK. Curation was performed first on both haplotypes simultaneously using Pretext, and then in a second round individually. Chromosome-scale scaffolds confirmed by Hi-C data were named in order of size. 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	1,104,372,934	875,599,765
GC %	37.12	37.12
Gaps/Gbp	248.1	181.59
Total gap bp	28,500	22,400
Scaffolds	120	170
Scaffold N50	51,015,210	42,099,637
Scaffold L50	8	9
Scaffold L90	19	19
Contigs	389	329
Contig N50	10,640,000	10,216,854
Contig L50	30	26
Contig L90	111	86
QV	46.1168	46.1312
Kmer compl.	64.91	54.3028
BUSCO sing.	75.9%	98.4%
BUSCO dupl.	23.5%	1.0%
BUSCO frag.	0.1%	0.3%
BUSCO miss.	0.4%	0.3%

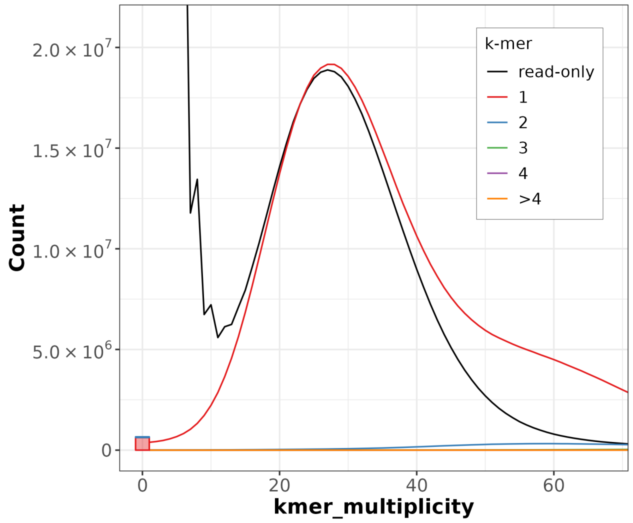
BUSCO: 6.0.0 (euk_genome_min, miniprot) / Lineage: metazoa_odb12 (genomes:206, BUSCOs:672)

HiC contact map of curated assembly

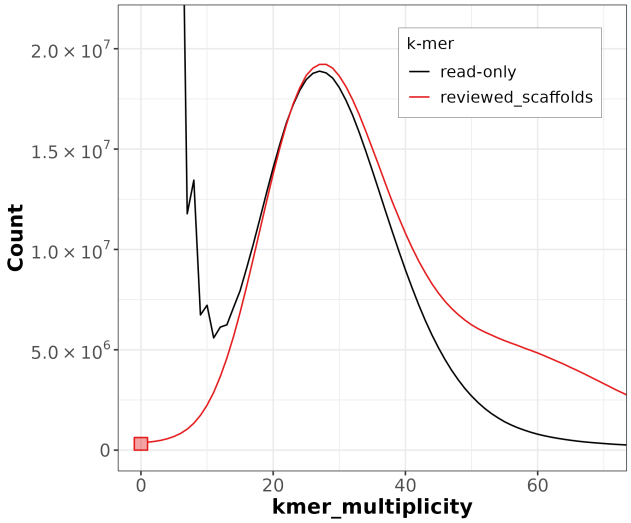


collapsed [LINK]

K-mer spectra of curated assembly

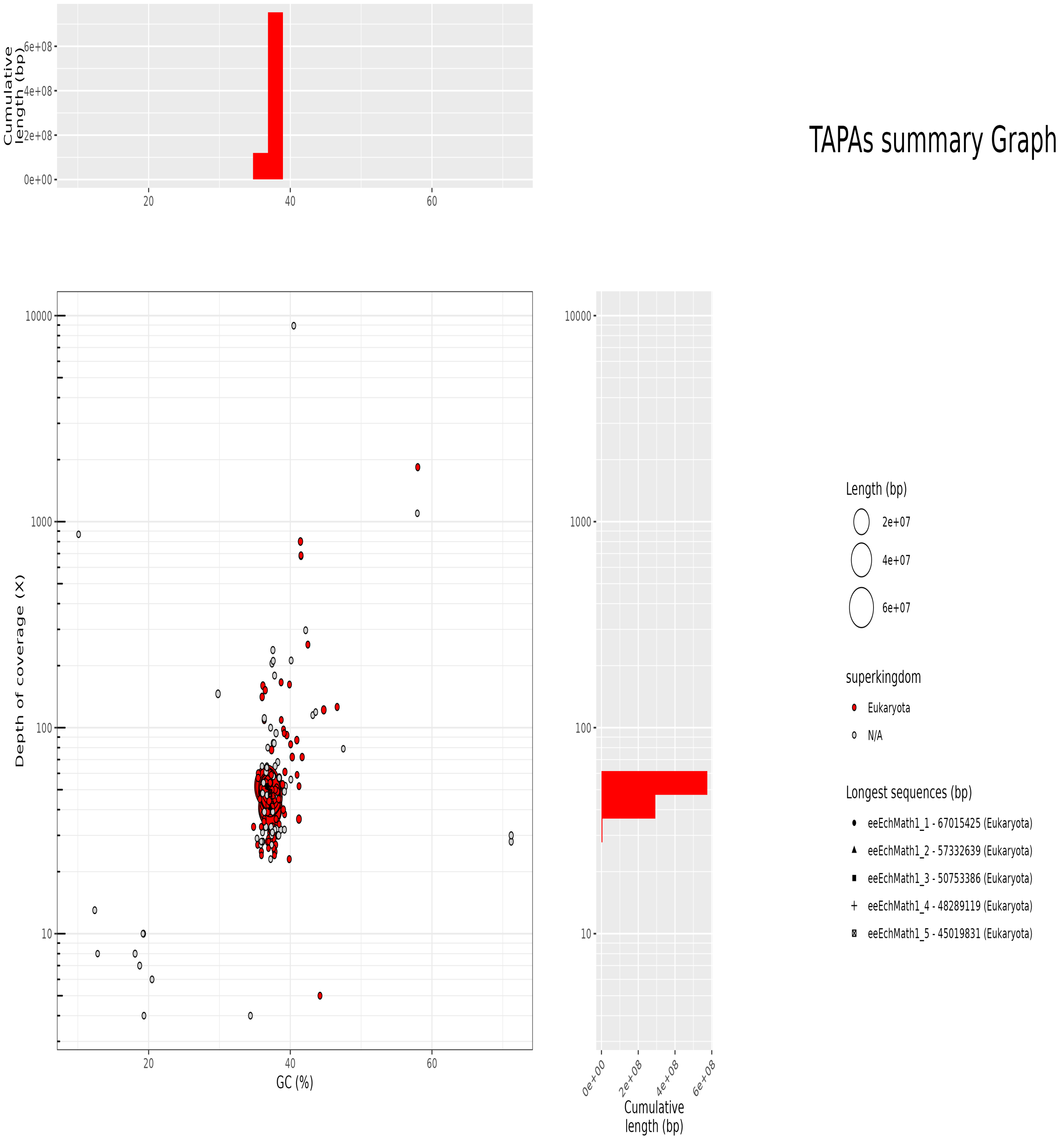


Distribution of k-mer counts per copy numbers found in asm



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	56	294

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 Menguy
Affiliation: Genoscope

Date and time: 2026-01-17 09:40:42 CET