

# ERGA Assembly Report

v24.10.15

Tags: ERGA-BGE

TxID	31175
ToLID	<b>eeDiaSeto1</b>
Species	Diadema setosum
Class	Echinoidea
Order	Diadematoida

Genome Traits	Expected	Observed
Haploid size (bp)	744,053,118	912,794,292
Haploid Number	21 (source: ancestor)	22
Ploidy	2 (source: ancestor)	2
Sample Sex	Unknown	Unknown

## EBP metrics summary and curation notes

Obtained EBP quality metric for collapsed: 6.7.Q52

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
- . Kmer completeness value is less than 90 for collapsed

### Curator notes

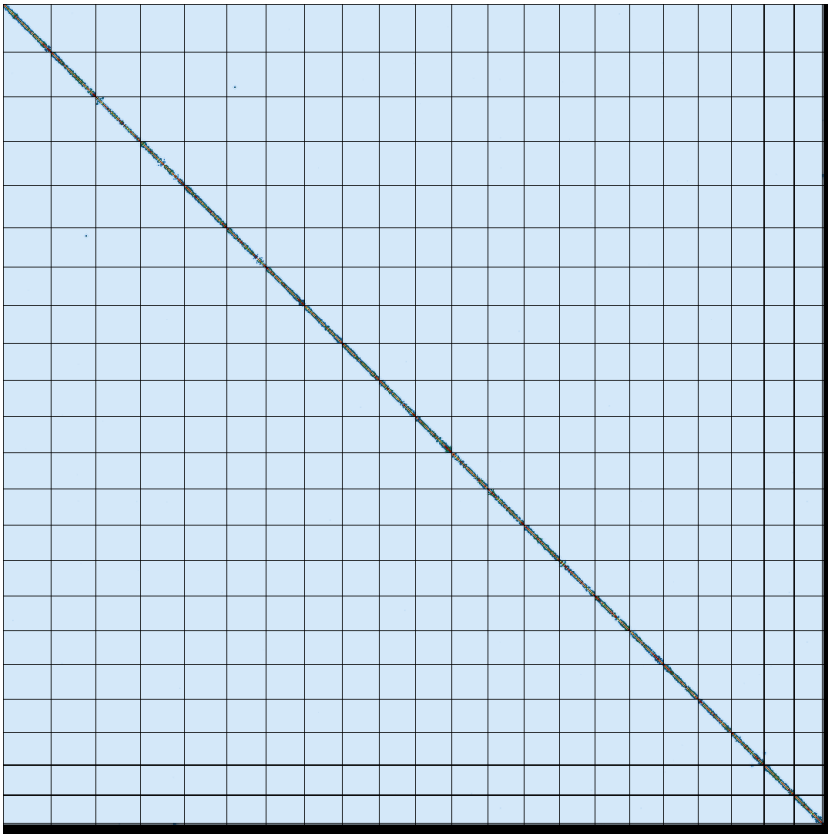
- . Interventions/Gb: 53
- . Contamination notes: "None "
- . Other observations: "Haplotigs confirmed by a decrease in coverage. The mitochondrial sequence has been added. "

## Quality metrics table

Metrics	Pre-curation collapsed	Curated collapsed
Total bp	917,618,006	912,794,292
GC %	38.35	38.36
Gaps/Gbp	0	705.53
Total gap bp	0	66,200
Scaffolds	701	102
Scaffold N50	2,320,055	39,828,611
Scaffold L50	128	11
Scaffold L90	394	20
Contigs	701	746
Contig N50	2,320,055	2,178,457
Contig L50	128	133
Contig L90	394	414
QV	8.26904	52.2797
Kmer compl.	0.0141886	72.1593
BUSCO sing.	95.5%	96.6%
BUSCO dupl.	1.0%	0.1%
BUSCO frag.	2.8%	2.7%
BUSCO miss.	0.6%	0.6%

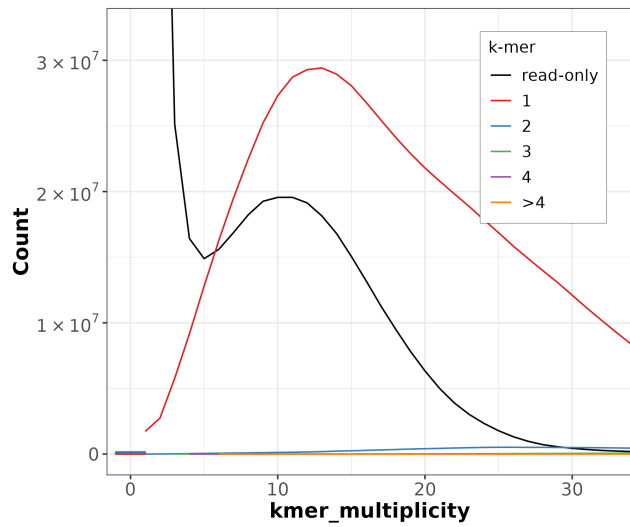
BUSCO: 5.8.2 (euk\_genome\_met, metaeuk) / Lineage: metazoa\_odb12 (genomes:206, BUSCOs:672)

# HiC contact map of curated assembly

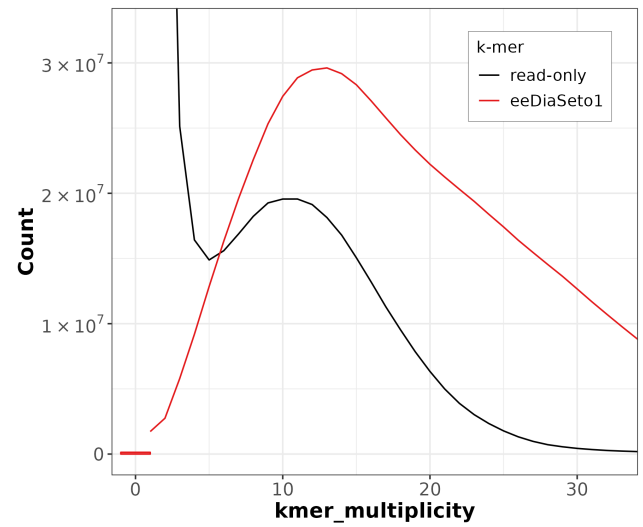


collapsed [\[LINK\]](#)

# K-mer spectra of curated assembly

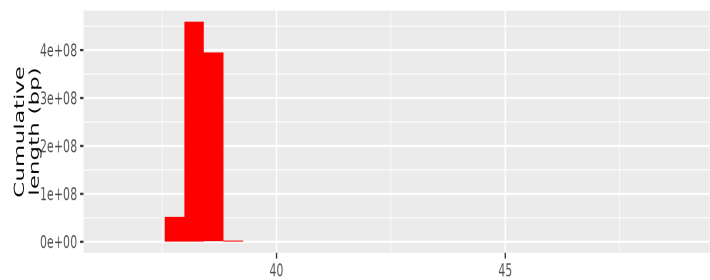


Distribution of k-mer counts per copy numbers found in asm



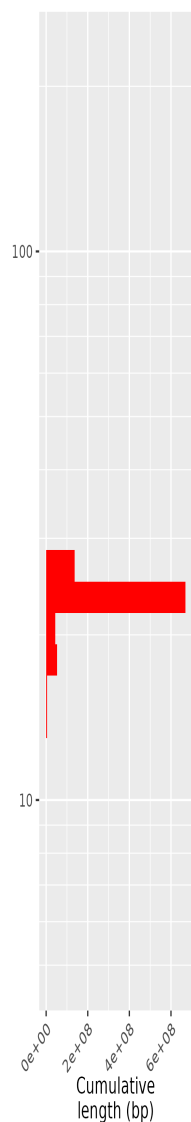
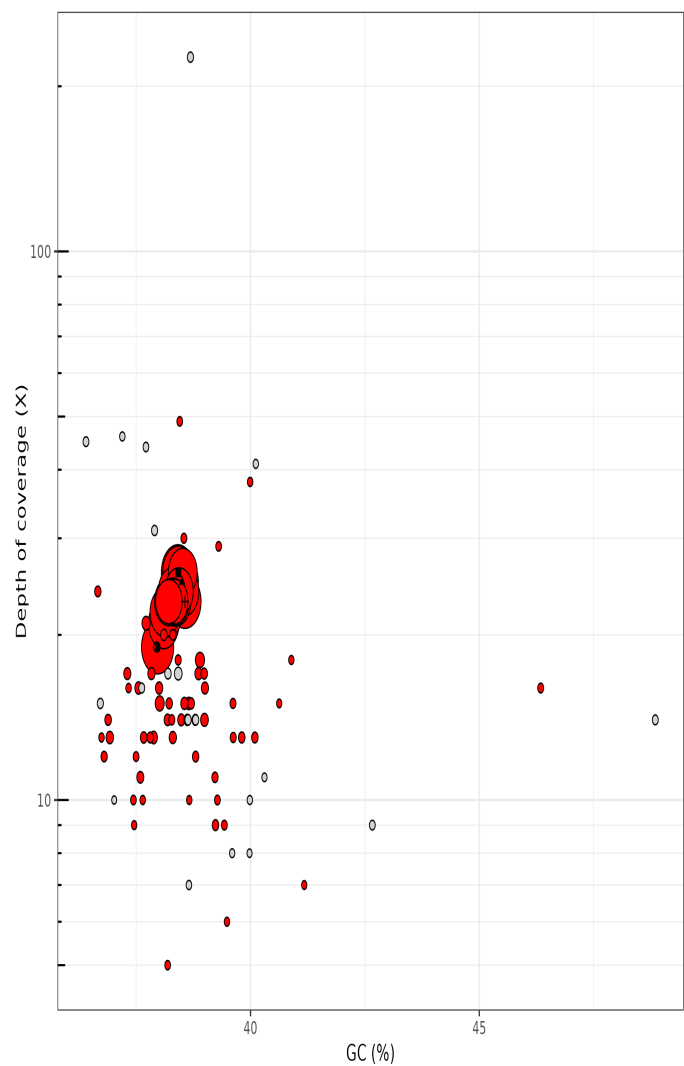
Distribution of k-mer counts coloured by their presence in reads/assemblies

# Post-curation contamination screening



## TAPAs summary Graph

(1 0X contig has been hidden)



### Longest sequences (bp)

- eeDiaSeto1\_1 - 52426318 (Eukaryota)
- ▲ eeDiaSeto1\_2 - 49285881 (Eukaryota)
- eeDiaSeto1\_3 - 48881309 (Eukaryota)
- + eeDiaSeto1\_4 - 48786125 (Eukaryota)
- ⊠ eeDiaSeto1\_5 - 46275308 (Eukaryota)

### Length (bp)

- 1e+07
- 2e+07
- 3e+07
- 4e+07
- 5e+07

### superkingdom

- Eukaryota
- N/A

**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	30	50

# 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: Arnaud Couloux

Affiliation: Genoscope

Date and time: 2025-05-22 01:16:08 CEST