ERGA Assembly Report
v24.10.15

Tags: ERGA-BGE

TxID	31175
ToLID	**eeDiaSeto1**
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
. Assembly length loss > 3% for collapsed

Curator notes

. Interventions/Gb: 53
. Contamination notes: "None "
. Other observations: "Haplotigs confirmed by a decrease in coverage "

Quality metrics table

| Metrics	Pre-curation collapsed	Curated collapsed
Total bp	1,113,293,332	912,794,292
GC %	38.36	38.36
Gaps/Gbp	0	705.53
Total gap bp	0	66,200
Scaffolds	1,185	102
Scaffold N50	2,005,209	39,828,611
Scaffold L50	171	11
Scaffold L90	587	20
Contigs	1,185	746
Contig N50	2,005,209	2,178,457
Contig L50	171	133
Contig L90	587	414
QV	54.4934	52.2797
Kmer compl.	78.9364	72.1593
BUSCO sing.	76.0%	96.6%
BUSCO dupl.	21.0%	0.1%
BUSCO frag.	2.4%	2.7%
BUSCO miss.	0.6%	0.6%

BUSCO: 5.8.2 (euk_genome_met, metaeuk) / Lineage: metazoa_odb12 (genomes:206, BUSCOs:672)
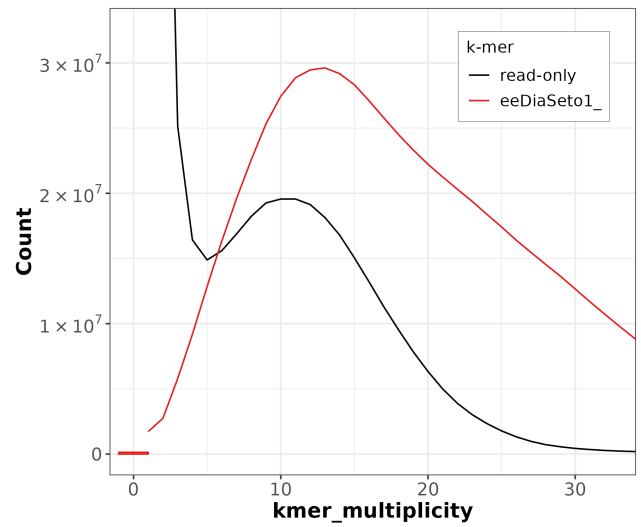
HiC contact map of curated assembly
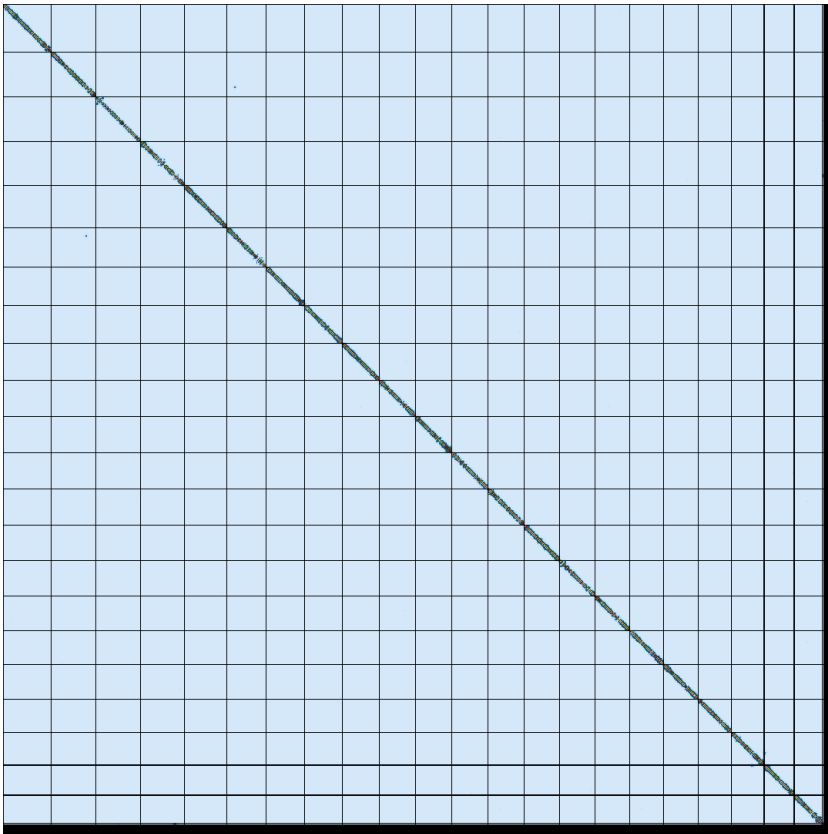
collapsed [LINK]
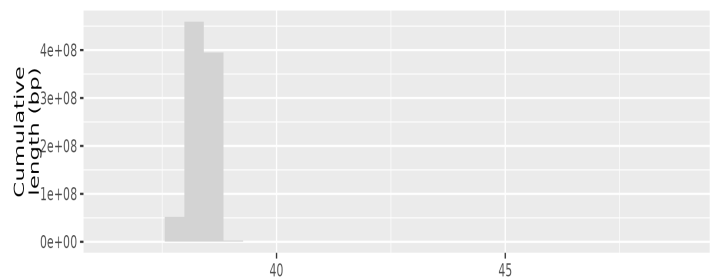
K-mer spectra of curated assembly
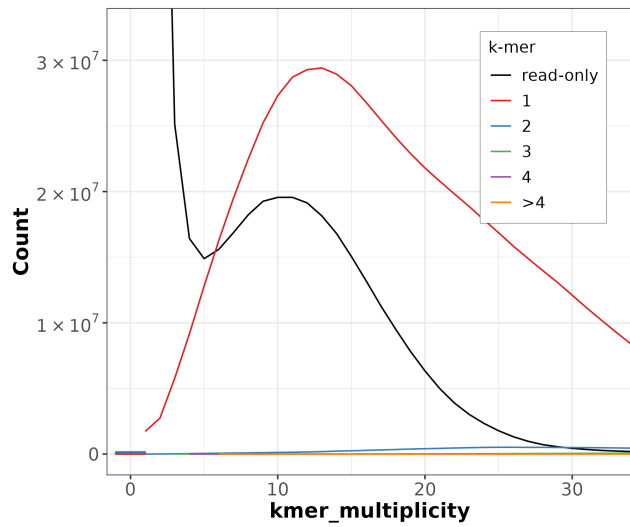
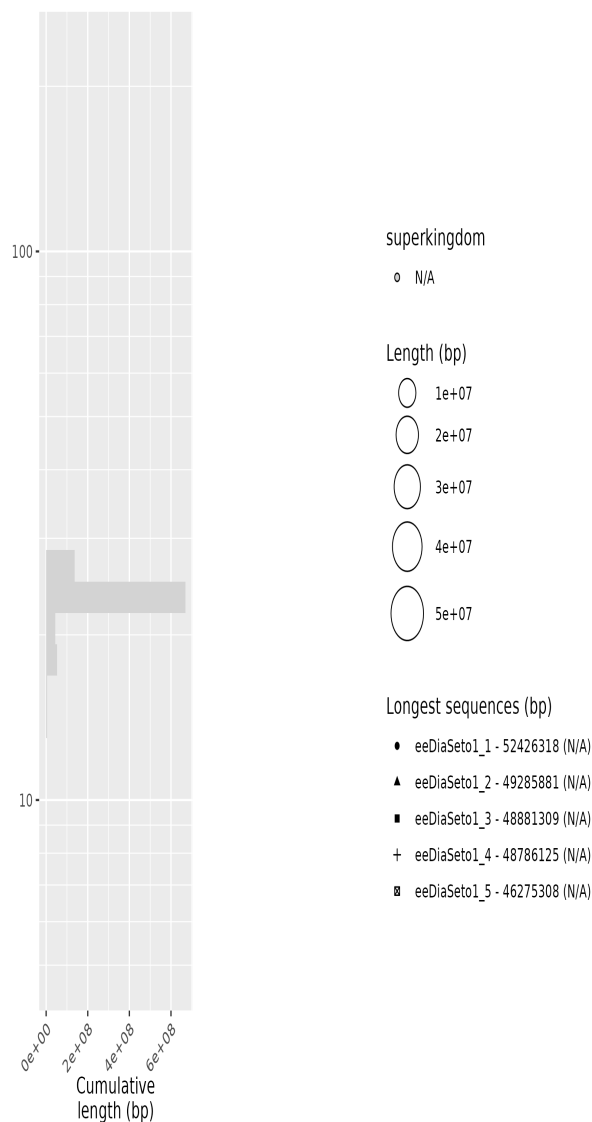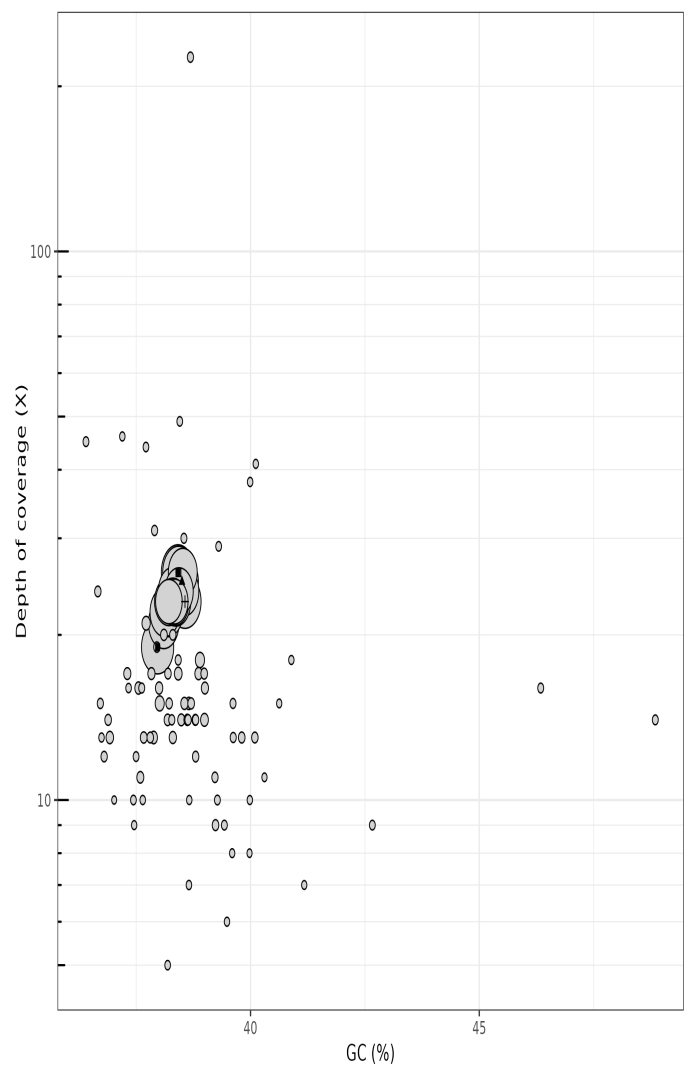Distribution of k-mer counts per copy
numbers found in asm

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)

superkingdom

○ N/A

Length (bp)

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

Longest sequences (bp)

• eeDiaSeto1_1 - 52426318 (N/A)
▲ eeDiaSeto1_2 - 49285881 (N/A)
■ eeDiaSeto1_3 - 48881309 (N/A)
+ eeDiaSeto1_4 - 48786125 (N/A)
⊠ eeDiaSeto1_5 - 46275308 (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: 2025-03-27 15:40:26 CET