

ERGA Assembly Report

v24.09.10

Tags: ERGA-BGE

TxID	31175
ToLID	eeDiaSeto1
Species	Diadema setosum
Class	Echinoidea
Order	Diadematoida

Genome Traits	Expected	Observed
Haploid size (bp)	744,053,118	912,952,438
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.Q54

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

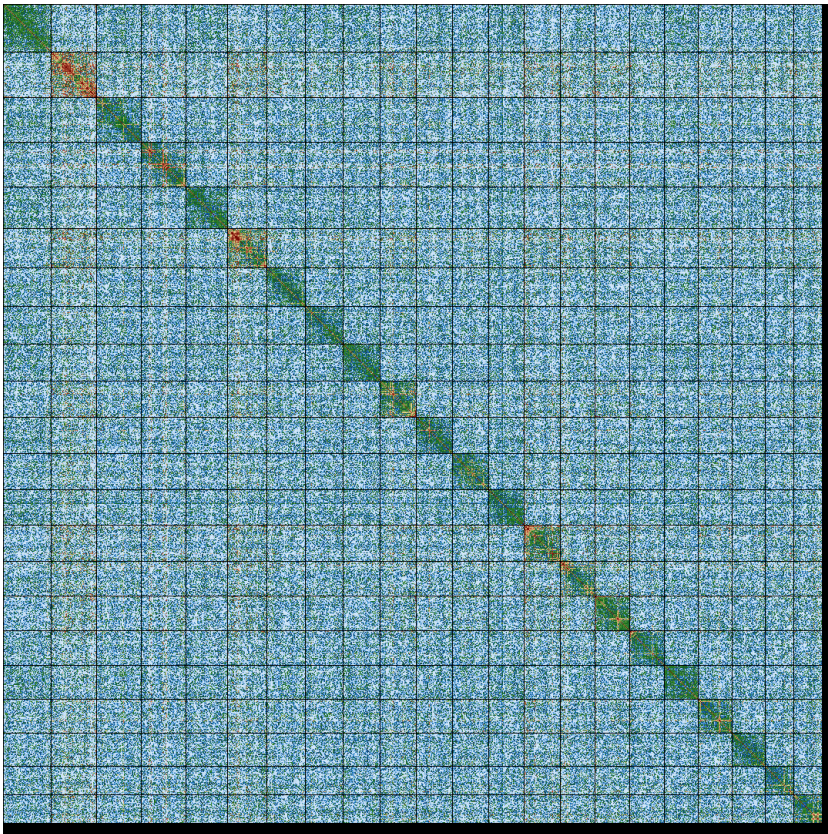
- . Interventions/Gb: 53
- . Contamination notes: "None "
- . Other observations: "Haplotigs confirmed by a decrease in coverage "

Quality metrics table

Metrics	Pre-curation collapsed	Curated collapsed
Total bp	917,618,006	912,952,438
GC %	38.35	38.35
Gaps/Gbp	0	727.31
Total gap bp	0	68,300
Scaffolds	701	98
Scaffold N50	2,320,055	39,828,711
Scaffold L50	128	11
Scaffold L90	394	20
Contigs	701	762
Contig N50	2,320,055	2,171,880
Contig L50	128	133
Contig L90	394	416
QV		54.4392
Kmer compl.		72.1726
BUSCO sing.	96.1%	97.6%
BUSCO dupl.	2.0%	0.4%
BUSCO frag.	1.2%	1.2%
BUSCO miss.	0.7%	0.8%

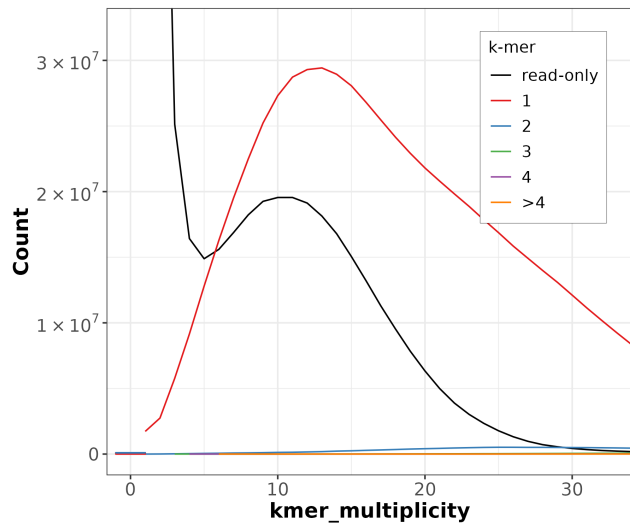
BUSCO 5.4.3 Lineage: eukaryota_odb10 (genomes:70, BUSCOs:255)

HiC contact map of curated assembly

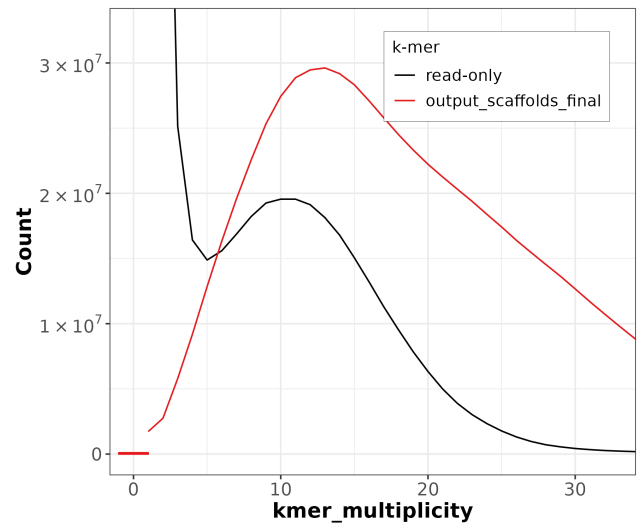


collapsed [\[LINK\]](#)

K-mer spectra of curated assembly

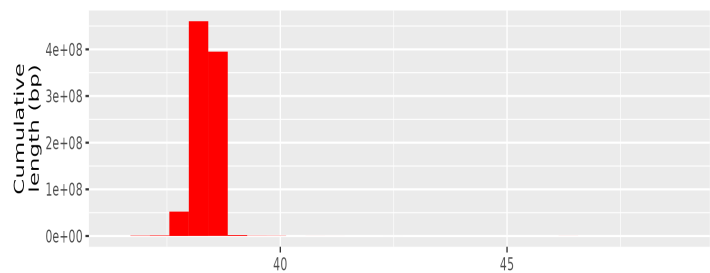


Distribution of k-mer counts per copy numbers found in asm

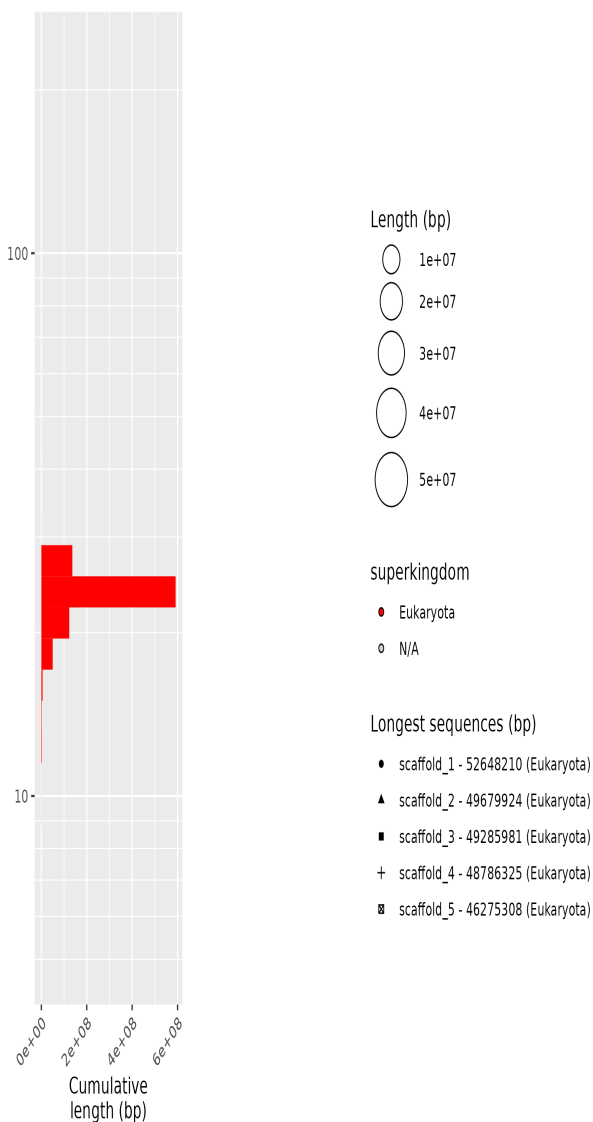
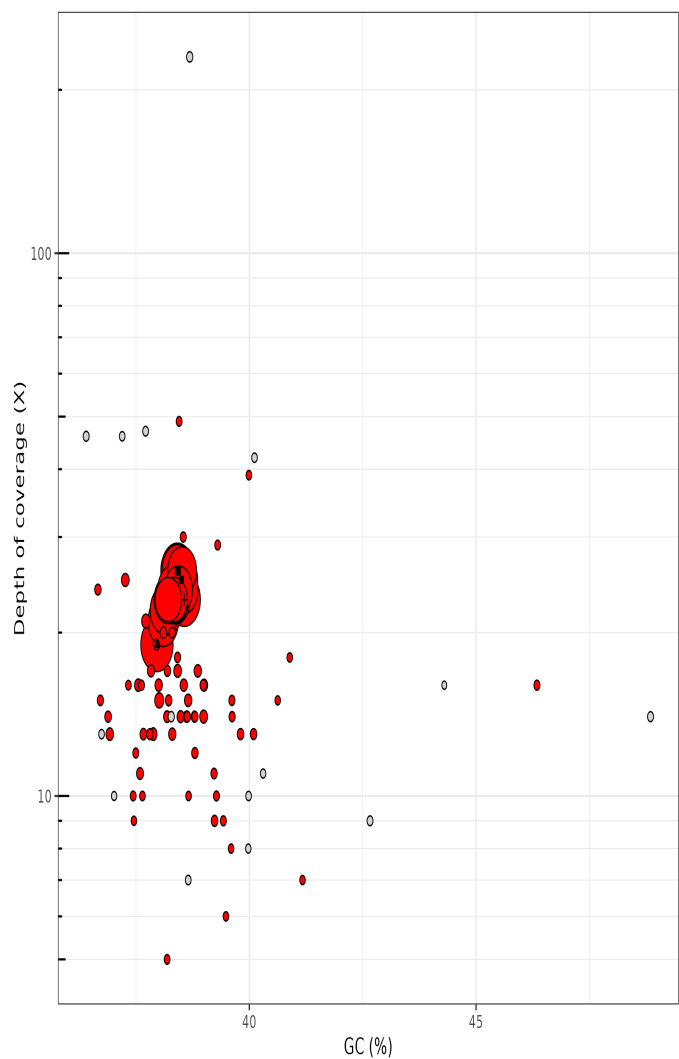


Distribution of k-mer counts coloured by their presence in reads/assemblies

Post-curation contamination screening



TAPAs summary Graph



- Length (bp)
- 1e+07
 - 2e+07
 - 3e+07
 - 4e+07
 - 5e+07
- superkingdom
- Eukaryota
 - N/A
- Longest sequences (bp)
- scaffold_1 - 52648210 (Eukaryota)
 - scaffold_2 - 49679924 (Eukaryota)
 - scaffold_3 - 49285981 (Eukaryota)
 - scaffold_4 - 48786325 (Eukaryota)
 - scaffold_5 - 46275308 (Eukaryota)

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 (4-enz)
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: 2024-09-17 15:56:51 CEST