# ERGA Assembly Report

v24.10.15

Tags: ATLASea[INVALID TAG]

| | |
|---|---|
| TxID | 72721 |
| ToLID | **eaEchSepo1.1** |
| Species | Echinaster sepositus |
| Class | Asteroidea |
| Order | Spinulosida |

| Genome Traits | Expected | Observed |
|---|---|---|
| Haploid size (bp) | 613,091,443 | 627,020,772 |
| Haploid Number | 18 (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.Q59

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

### Curator notes

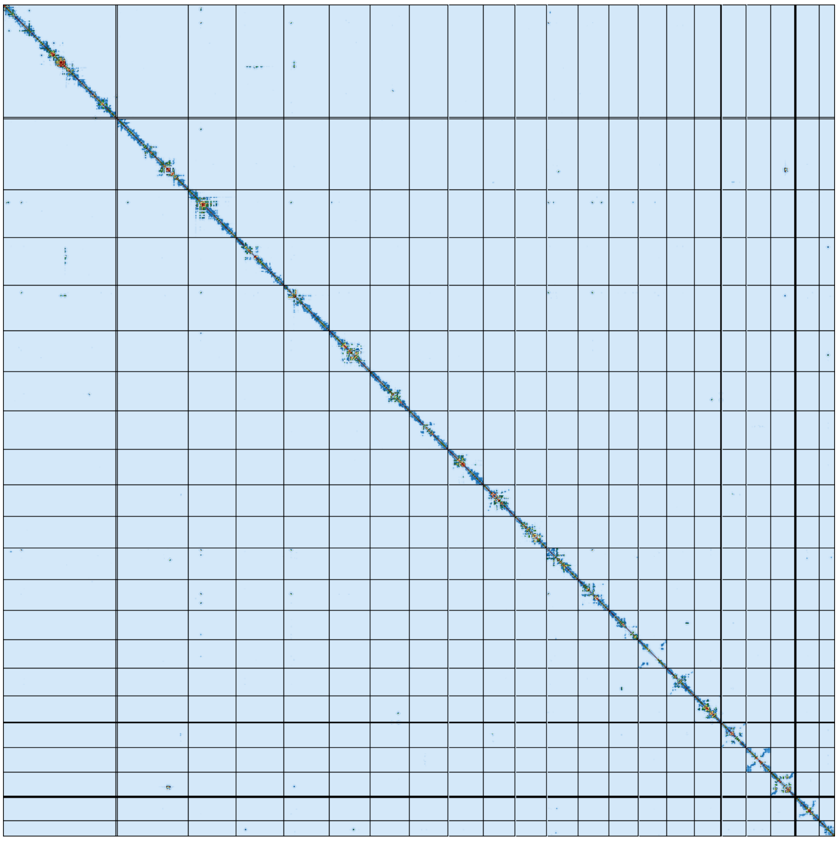
. Interventions/Gb: 604
. Contamination notes: "We identified and removed 183 contigs as bacterial or viral sequences, totaling 13.3Mb, with the largest being 479Kb. "
. Other observations: "The genome contains large centromeric regions with uncertain organization. We removed 434 haplotypic regions, totaling 142Mb, with the largest being 9.4Mb. "

# Quality metrics table

| Metrics | Pre-curation collapsed | Curated collapsed |
|---|---|---|
| Total bp | 633,618,056 | 627,020,772 |
| GC % | 41.93 | 41.92 |
| Gaps/Gbp | 260.41 | 524.7 |
| Total gap bp | 16,500 | 54,600 |
| Scaffolds | 184 | 40 |
| Scaffold N50 | 26,037,565 | 28,592,635 |
| Scaffold L50 | 9 | 8 |
| Scaffold L90 | 25 | 19 |
| Contigs | 349 | 369 |
| Contig N50 | 4,170,447 | 4,092,000 |
| Contig L50 | 49 | 51 |
| Contig L90 | 151 | 160 |
| QV | 46.4269 | 59.7099 |
| Kmer compl. | 79.3067 | 80.1215 |
| BUSCO sing. | 95.8% | 96.6% |
| BUSCO dupl. | 1.2% | 0.3% |
| BUSCO frag. | 1.0% | 1.0% |
| BUSCO miss. | 2.0% | 2.1% |

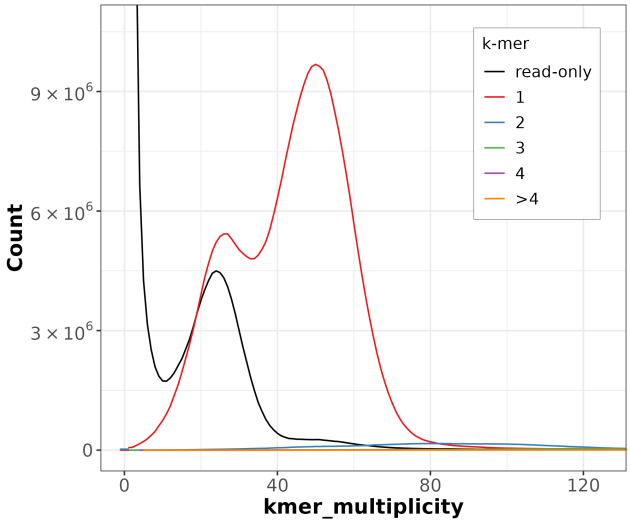
BUSCO: 5.4.3 (euk_genome_met, metaeuk) / Lineage: metazoa_odb10 (genomes:65, BUSCOs:954)

# HiC contact map of curated assembly

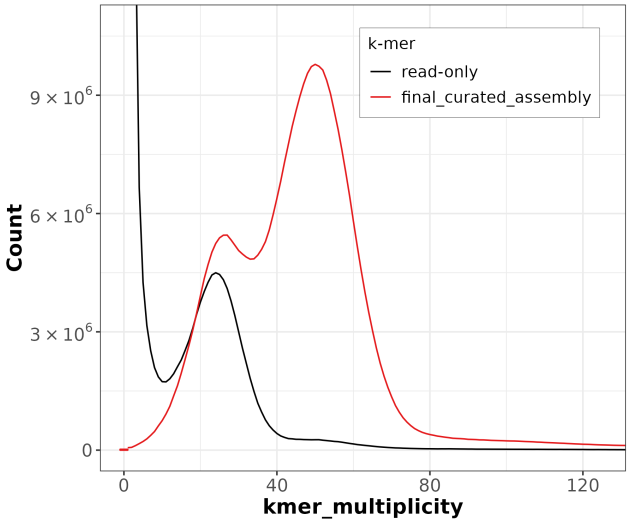


**collapsed** [LINK]

# K-mer spectra of curated assembly

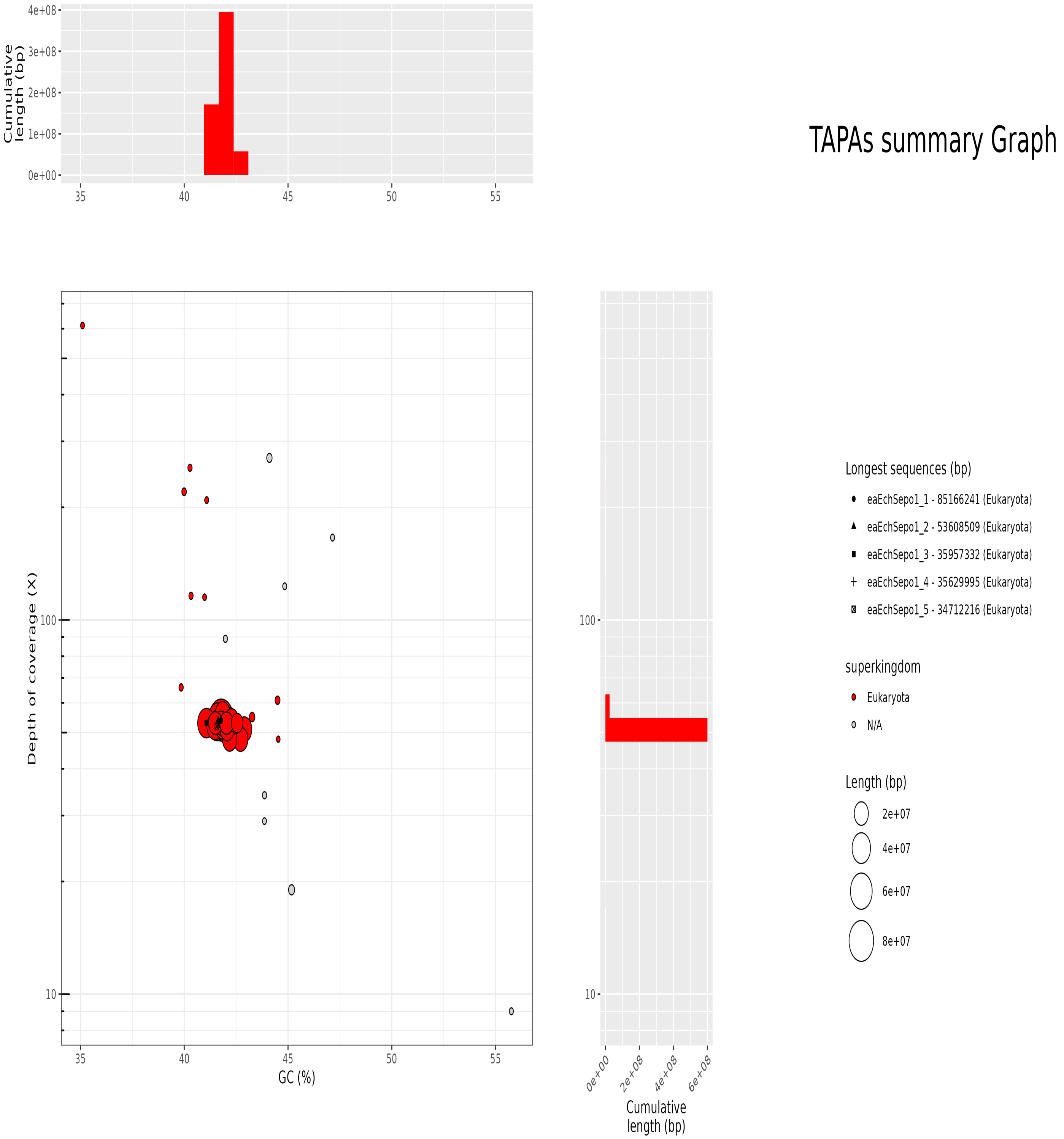


Distribution of k-mer counts per copy numbers found in asm



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 | 41 | 115 |

# 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: Jean-Marc Aury
Affiliation: Genoscope

Date and time: 2024-12-09 17:24:45 CET