ERGA Assembly Report
v24.10.15

Tags: ATLASea[INVALID TAG]

TxID	39374
ToLID	**eeSphGran1**
Species	Sphaerechinus granularis
Class	Echinoidea
Order	Temnopleuroida

Genome Traits	Expected	Observed
Haploid size (bp)	748,755,816	992,419,902
Haploid Number	17 (source: direct)	20
Ploidy	2 (source: ancestor)	2
Sample Sex	Unknown	Unknown

EBP metrics summary and curation notes

Obtained EBP quality metric for collapsed: 6.7.Q55

The following metrics were automatically flagged as below EBP recommended standards or different from expected:

. Observed Haploid size (bp) has >20% difference with Expected
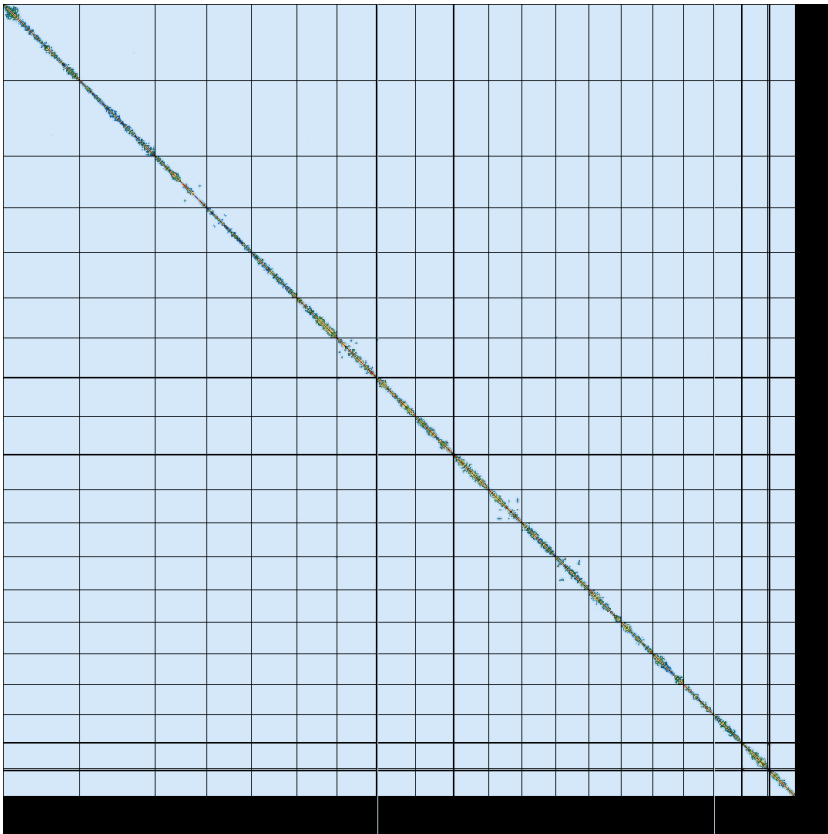. Observed Haploid Number is different from Expected

Curator notes

. Interventions/Gb: 111
. Contamination notes: "only 2 bacterial contigs were removed "
. Other observations: "18.8Mb were tagged as haplotigs and removed "

Quality metrics table

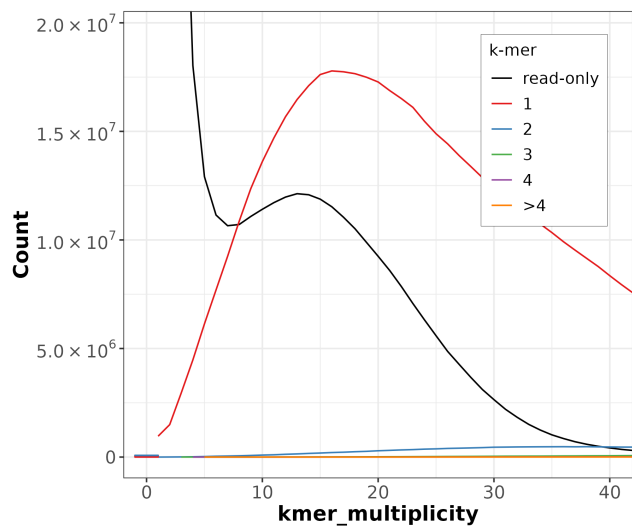| Metrics	Pre-curation collapsed	Curated collapsed
Total bp	1,011,272,610	992,419,902
GC %	36.42	36.41
Gaps/Gbp	870.19	895.79
Total gap bp	88,000	93,600
Scaffolds	352	306
Scaffold N50	42,375,516	44,956,304
Scaffold L50	9	9
Scaffold L90	21	19
Contigs	1,232	1,195
Contig N50	1,643,155	1,661,000
Contig L50	195	187
Contig L90	658	633
QV	43.3458	55.8817
Kmer compl.	72.837	73.349
BUSCO sing.	94.3%	
BUSCO dupl.	3.0%	
BUSCO frag.	1.3%	
BUSCO miss.	1.4%	

BUSCO: 5.4.3 (euk_genome_met, metaeuk) / Lineage: metazoa_odb10 (genomes:65, BUSCOs:954)
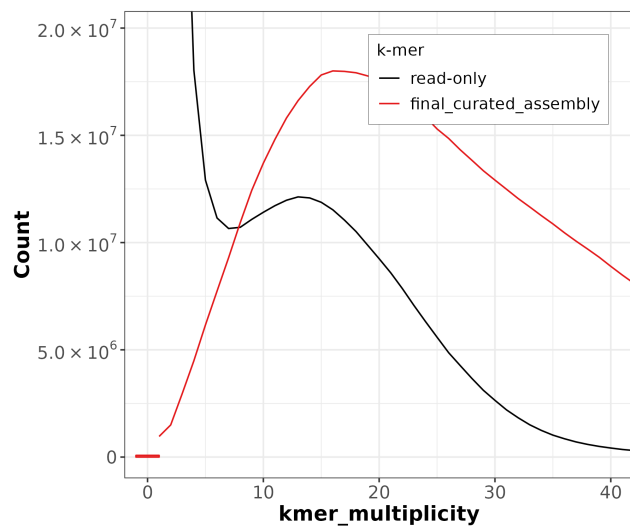
HiC contact map of curated assembly

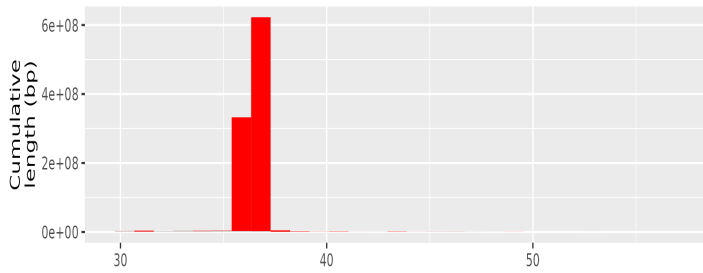**collapsed** [LINK]

K-mer spectra of curated assembly

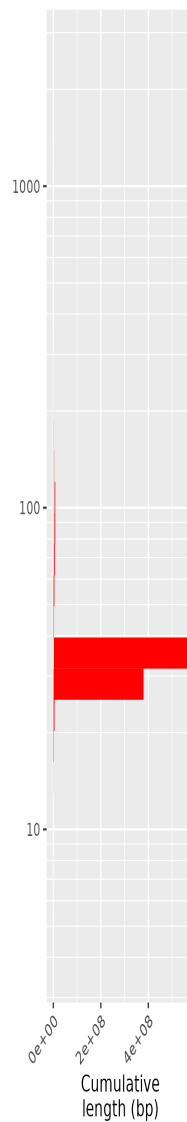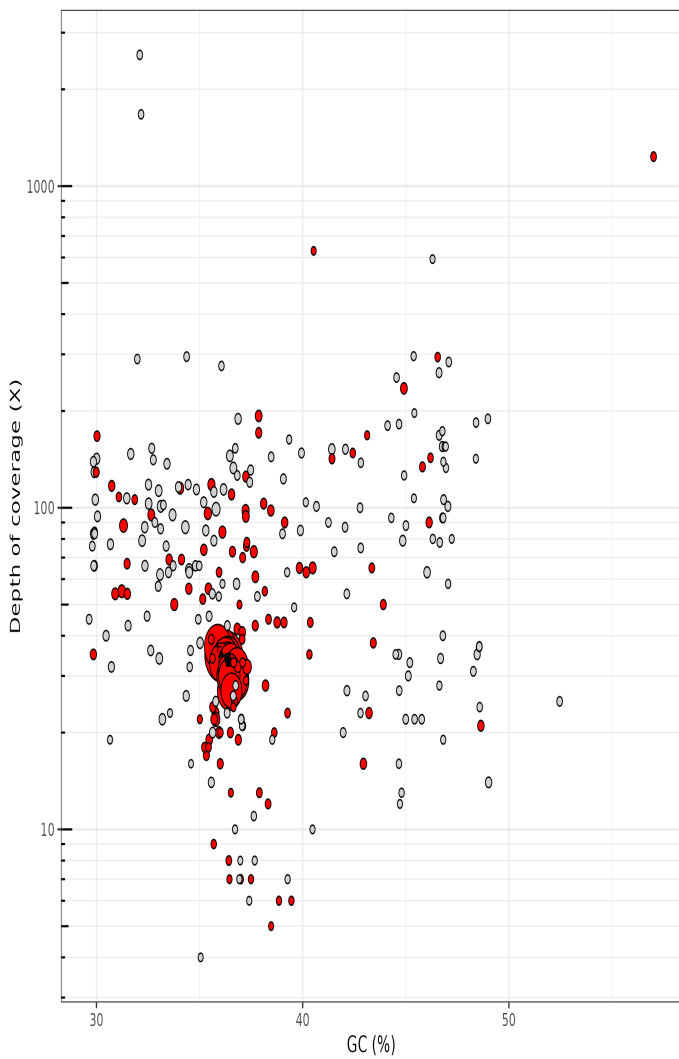Distribution of k-mer counts per copy
numbers found in asm

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

Post-curation contamination screening

TAPAs summary Graph

Longest sequences (bp)

- ● eeSphGran1_1_1 - 91140609 (Eukaryota)
- ▲ eeSphGran1_1_2 - 90287819 (Eukaryota)
- ■ eeSphGran1_1_3 - 61315862 (Eukaryota)
- + eeSphGran1_1_4 - 53993442 (Eukaryota)
- ⊠ eeSphGran1_1_5 - 53771520 (Eukaryota)

superkingdom

- ● Eukaryota
- ○ N/A

Length (bp)

- 2.5e+07
- 5.0e+07
- 7.5e+07

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	40	70

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 Belser
Affiliation: Genoscope

Date and time: 2024-12-13 20:54:25 CET