

ERGA Assembly Report

v24.09.10

Tags: ERGA-BGE

TxID	491806
ToLID	ehHolSanc2
Species	Holothuria sanctori
Class	Holothuroidea
Order	Aspidochirotida

Genome Traits	Expected	Observed
Haploid size (bp)	1,028,481,939	1,171,879,570
Haploid Number	21 (source: ancestor)	23
Ploidy	2 (source: ancestor)	2
Sample Sex	Unknown	Unknown

EBP metrics summary and curation notes

Obtained EBP quality metric for collapsed: 7.7.Q61

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

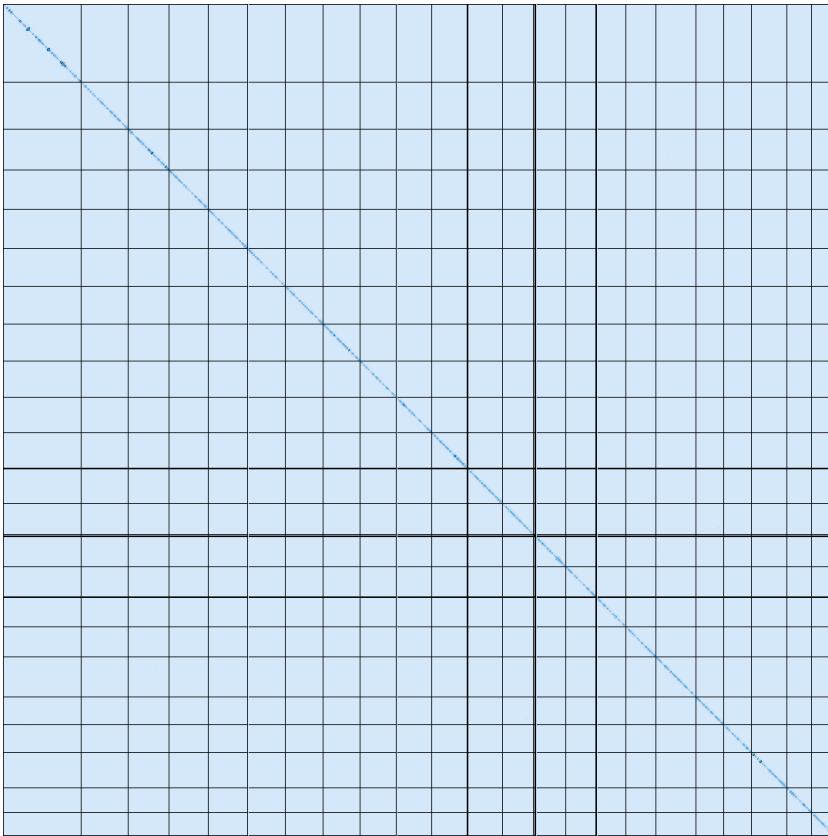
- . Interventions/Gb: 35
- . Contamination notes: ""
- . Other observations: "A large number of duplications have been removed and many sequences at the end of the scaffold have been replaced in the chromosomes "

Quality metrics table

Metrics	Pre-curation collapsed	Curated collapsed
Total bp	1,194,060,647	1,171,879,570
GC %	39.37	39.37
Gaps/Gbp	35.17	87.04
Total gap bp	4,200	17,400
Scaffolds	59	43
Scaffold N50	49,855,781	50,699,391
Scaffold L50	10	10
Scaffold L90	21	20
Contigs	101	145
Contig N50	31,981,105	17,819,081
Contig L50	16	22
Contig L90	42	65
QV	61.1747	61.1753
Kmer compl.	70.1165	69.3228
BUSCO sing.	97.6%	97.6%
BUSCO dupl.	0.8%	0.8%
BUSCO frag.	1.2%	1.2%
BUSCO miss.	0.4%	0.4%

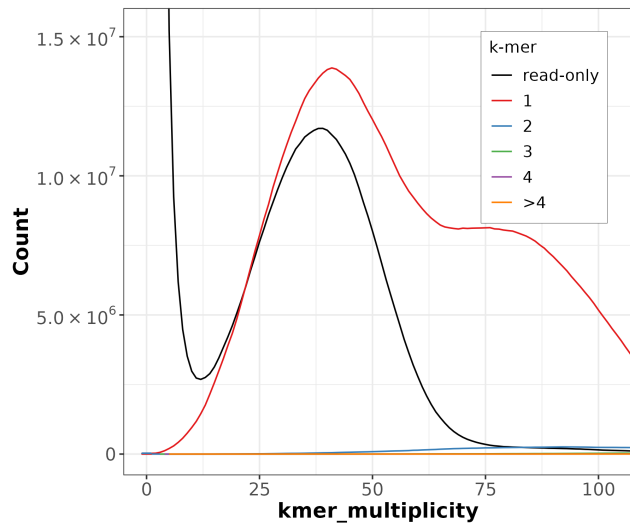
BUSCO 5.4.3 Lineage: eukaryota_odb10 (genomes:70, BUSCOs:255)

HiC contact map of curated assembly

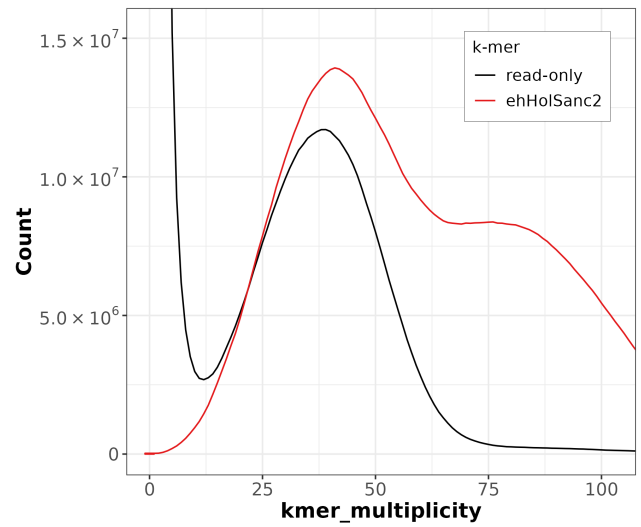


collapsed [\[LINK\]](#)

K-mer spectra of curated assembly

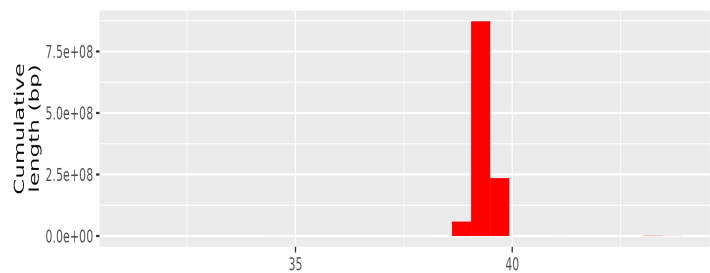


Distribution of k-mer counts per copy numbers found in asm

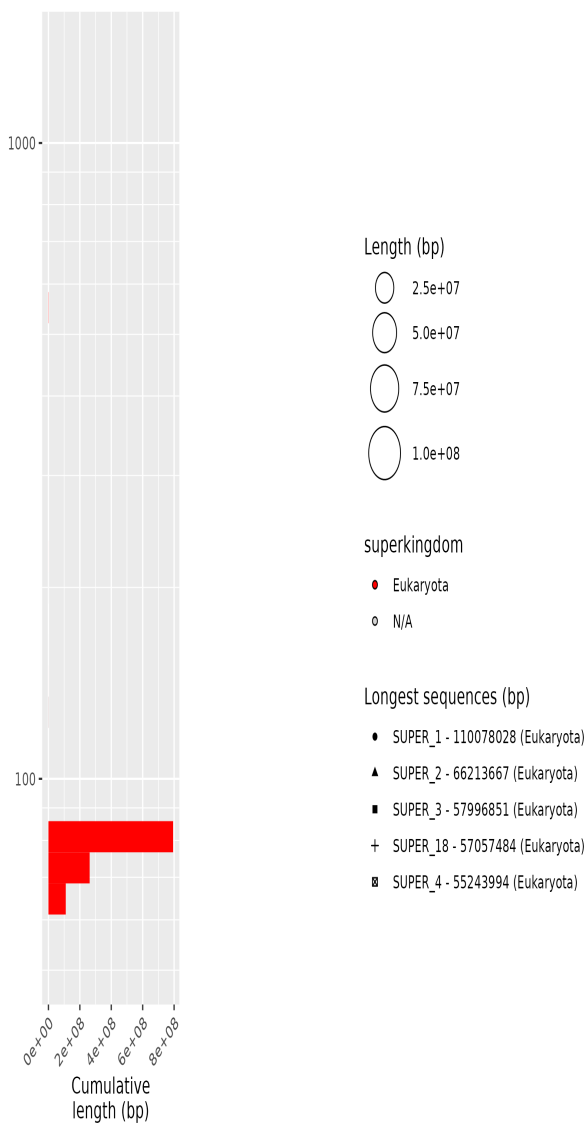
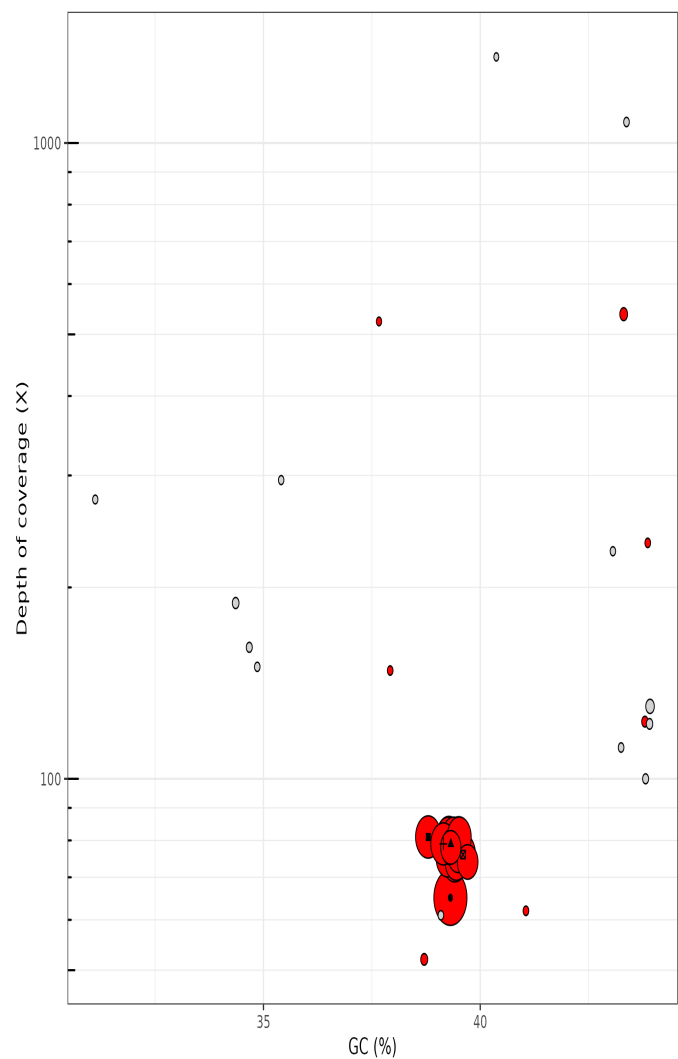


Distribution of k-mer counts coloured by their presence in reads/assemblies

Post-curation contamination screening



TAPAs summary Graph



- Length (bp)
- 2.5e+07
 - 5.0e+07
 - 7.5e+07
 - 1.0e+08
- superkingdom
- Eukaryota
 - N/A
- Longest sequences (bp)
- SUPER_1 - 110078028 (Eukaryota)
 - SUPER_2 - 66213667 (Eukaryota)
 - SUPER_3 - 57996851 (Eukaryota)
 - SUPER_18 - 57057484 (Eukaryota)
 - SUPER_4 - 55243994 (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	90	47

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: Lola Demirdjian

Affiliation: Genoscope

Date and time: 2024-10-22 16:37:51 CEST