

ERGA Assembly Report

v24.10.15

Tags: ERGA-BGE

TxID	1072197
ToLID	ehHolPol1
Species	<i>Holothuria poli</i>
Class	Holothuroidea
Order	Aspidochirotida

Genome Traits	Expected	Observed
Haploid size (bp)	1,578,169,159	1,825,045,599
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.Q65

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: 144
- Contamination notes: ""
- Other observations: "The assembly of HOLOTHURIA POLI (ehHolPol1) is based on 48X PacBio data and 662X Arima Hi-C data generated as part of the European Reference Genome Atlas (ERGA, <https://www.erga-biodiversity.eu/>) via the Biodiversity Genomics Europe project (BGE, <https://biodiversitygenomics.eu/>). The assembly process included the following steps: initial PacBio and Hi-C assembly generation with Hifiasm creating a dual assembly, removal of contaminant sequences using Context, and Hi-C-based scaffolding with YaHS. Both haplotypes were curated, but only haplotype 1 (HAP1) was retained for this EAR report. All subsequent information is based on HAP1. In total, 7 contigs were identified as contaminants (bacterial), totaling 332,045 pb (with the largest being 181,308 pb). The mitochondrial genome was assembled using OATK. Finally, the primary assembly was analyzed and manually improved using Pretext. During manual curation, 33 haplotypic region was removed, totaling 240,764,653 pb (with the largest being 85,225,335 pb). "

Quality metrics table

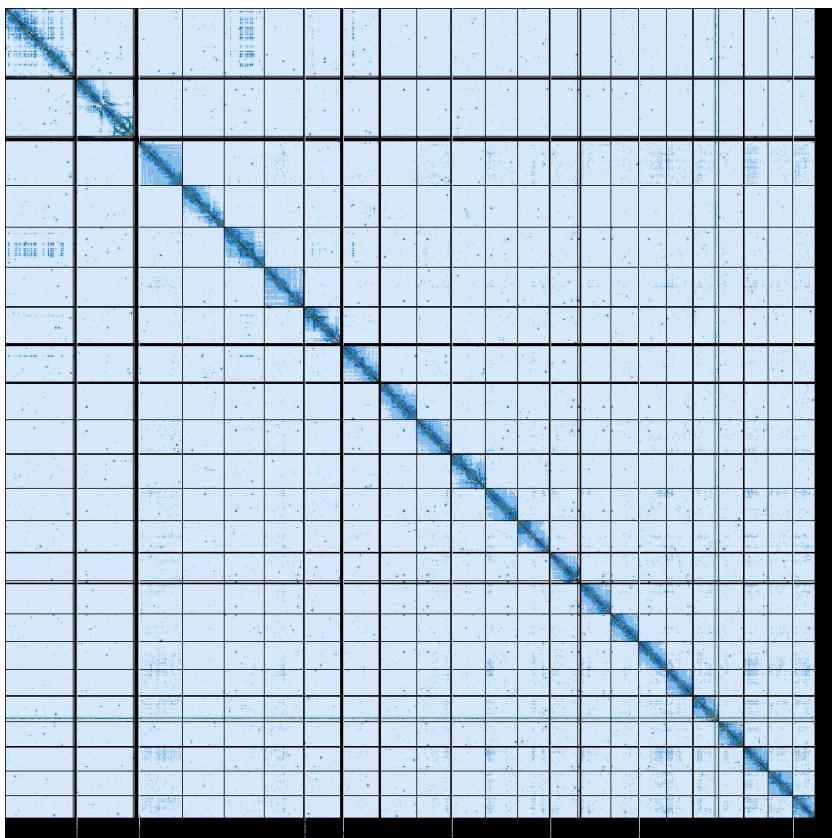
Metrics	Pre-curation collapsed	Curated collapsed
Total bp	1,848,607,575	1,825,045,599
GC %	39.26	39.29
Gaps/Gbp	219.08	470.67
Total gap bp	40,500	88,700
Scaffolds	163	833
Scaffold N50	71,625,132	75,092,538
Scaffold L50	11	10
Scaffold L90	23	22
Contigs	568	1,692
Contig N50	8,197,000	3,578,110
Contig L50	68	151
Contig L90	231	536
QV	65.8204	65.0674
Kmer compl.	70.7546	68.236
BUSCO sing.	80.4%	95.1%
BUSCO dupl.	3.4%	2.7%
BUSCO frag.	14.9%	1.2%
BUSCO miss.	1.3%	1.0%

Warning! BUSCO versions or lineage datasets are not the same across results:

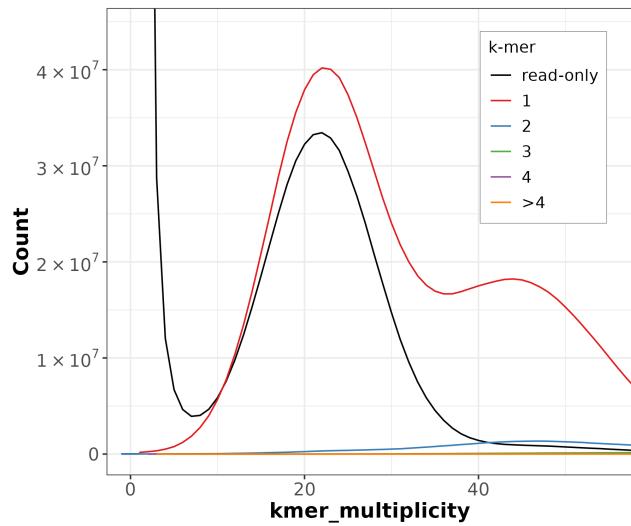
BUSCO: 5.8.2 (euk_genome_met, metaeuk) / Lineage: metazoa_odb12 (genomes:206, BUSCOs:672)

BUSCO: 6.0.0 (euk_genome_min, miniprot) / Lineage: metazoa_odb12 (genomes:206, BUSCOs:672)

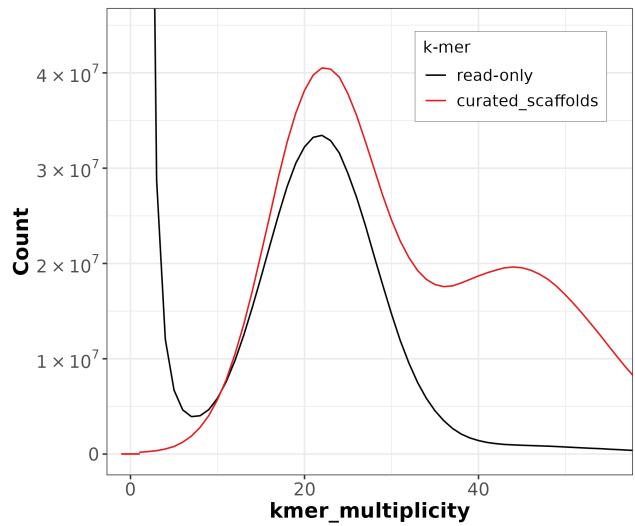
HiC contact map of curated assembly



K-mer spectra of curated assembly

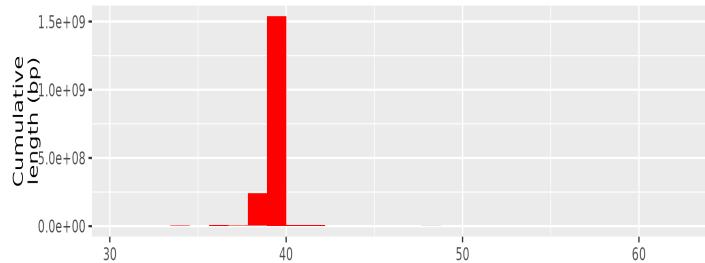


Distribution of k-mer counts per copy numbers found in asm

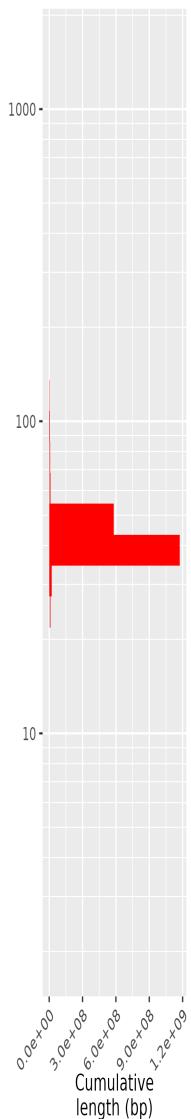
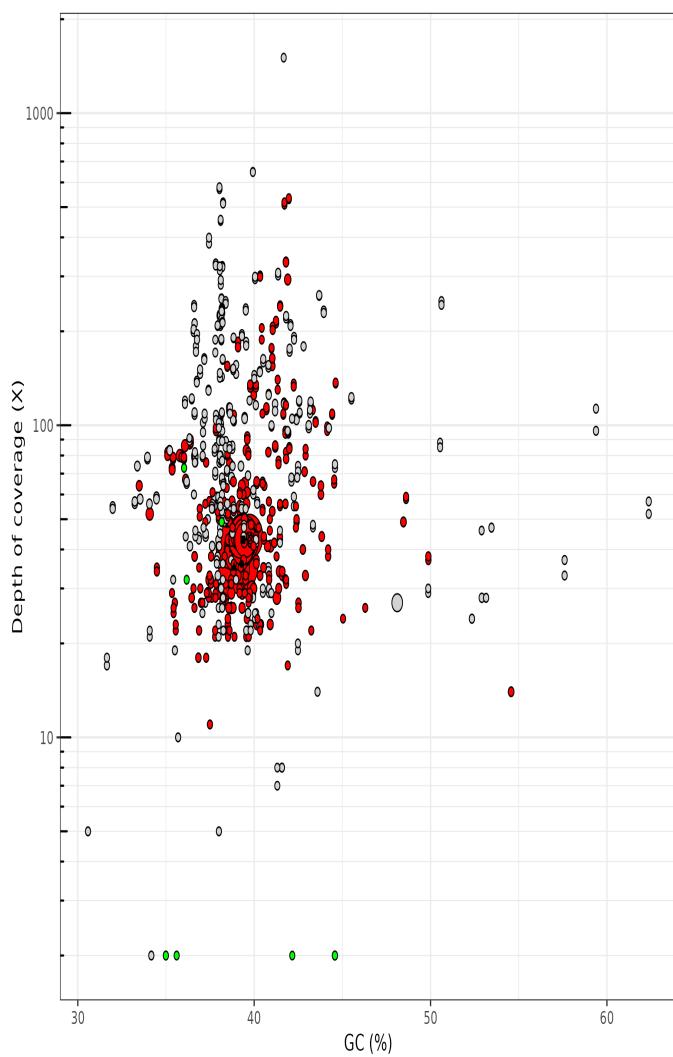


Distribution of k-mer counts coloured by their presence in reads/assemblies

Post-curation contamination screening



TAPAs summary Graph



superkingdom

● Bacteria

● Eukaryota

○ N/A

Longest sequences (bp)

- ehHolPoli1_1 - 149470392 (Eukaryota)
- ▲ ehHolPoli1_2 - 125562657 (Eukaryota)
- ehHolPoli1_3 - 96759529 (Eukaryota)
- + ehHolPoli1_4 - 91822892 (Eukaryota)
- ☒ ehHolPoli1_5 - 88573848 (Eukaryota)

Length (bp)

○ 5.0e+07

○ 1.0e+08

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	Long reads	Arima
Coverage	48	662

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: 2025-12-13 19:16:07 CET