ERGA Assembly Report

v24.10.15

Tags: ATLASea[INVALID TAG]

TxID	3256806
ToLID	**ehHolPrin1**
Species	Holothuria princeps
Class	Holothuroidea
Order	Aspidochirotida

Genome Traits	Expected	Observed
Haploid size (bp)	1,185,064,294	1,226,494,088
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: 6.7.Q48

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
. Not 90% of assembly in chromosomes for collapsed

Curator notes

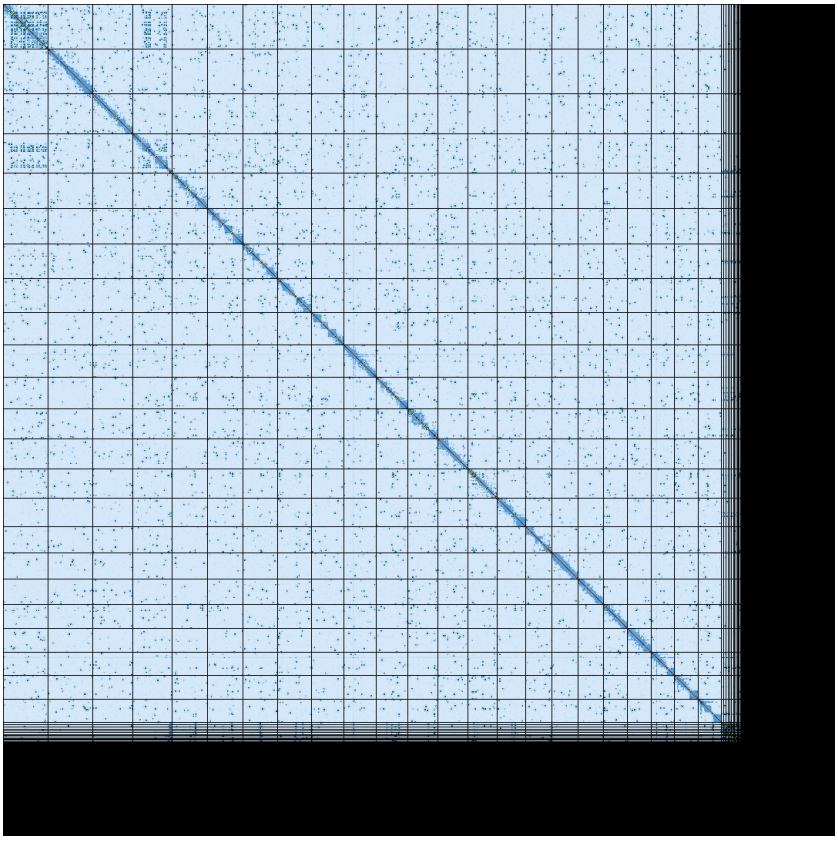
. Interventions/Gb: 131
. Contamination notes: ""
. Other observations: "The assembly of Holothuria princeps (ehHolPrin1) is based on 29.39X PacBio data and 145.14X Arima Hi-C data generated as part of the ATLASea programme (https://www.atlasea.fr). The assembly process included the following steps: initial PacBio assembly generation with Hifiasm, removal of contaminant sequences using Context, removal of haplotypic duplications using purge_dups, and Hi-C-based scaffolding with YaHS. In total, 1 contigs was identified as contaminants (bacterial, archaeal, or viral), totaling 0.028 Mb (with the largest being 0.028 Mb). Additionally, 1696 regions totaling 96.815 Mb (with the largest being 0.893 Mb) were identified as haplotypic duplications and removed. The mitochondrial genome was assembled using Oatk. Finally, the primary assembly was analyzed and manually improved using Pretext. During manual curation, 31 haplotypic regions and 3 contaminant sequences were removed, totaling 17.435Mb and 0.135Mb, respectively (with the largest being 1.448Mb and 0.072Mb). Chromosome-scale scaffolds confirmed by Hi-C data were named in order of size. "

Quality metrics table

Metrics	Pre-curation collapsed	Curated collapsed
Total bp	1,244,039,851	1,226,494,088
GC %	38.98	38.98
Gaps/Gbp	479.08	542.2
Total gap bp	59,600	76,800
Scaffolds	1,130	1,053
Scaffold N50	40,424,310	44,997,604
Scaffold L50	13	12
Scaffold L90	108	52
Contigs	1,726	1,718
Contig N50	6,236,270	6,457,638
Contig L50	61	60
Contig L90	344	336
QV	48.7202	48.7171
Kmer compl.	84.8821	84.4746
BUSCO sing.	98.0%	99.2%
BUSCO dupl.	1.8%	0.5%
BUSCO frag.	0.2%	0.2%
BUSCO miss.	0.0%	0.0%

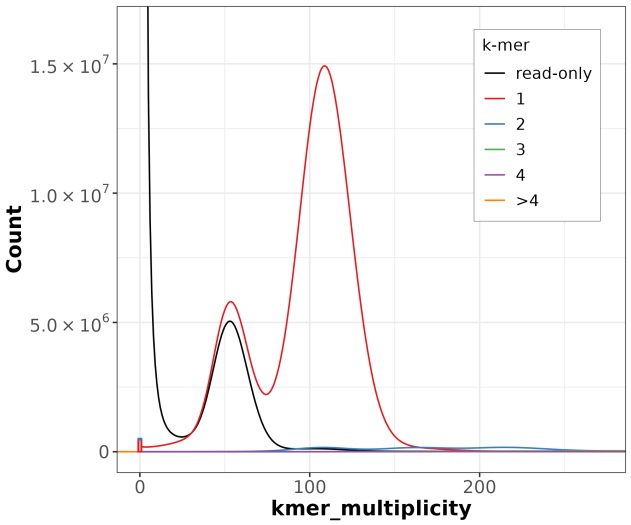
BUSCO: 6.1.0 (euk_genome_min, miniprot) / Lineage: metazoa_odb12.2 (genomes:182, BUSCOs:932)

HiC contact map of curated assembly

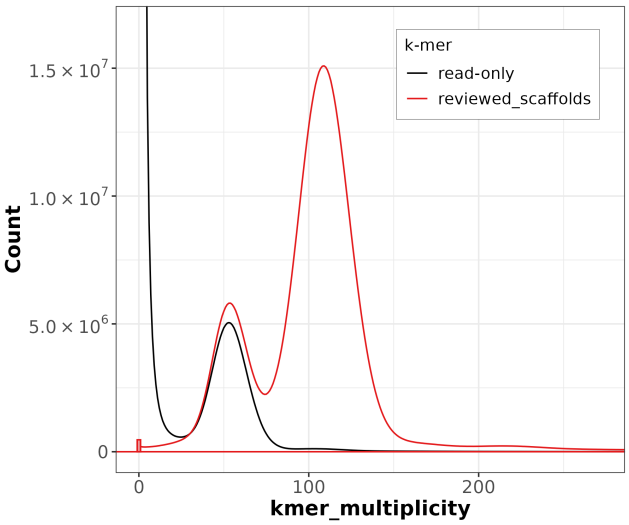


collapsed [LINK]

K-mer spectra of curated assembly

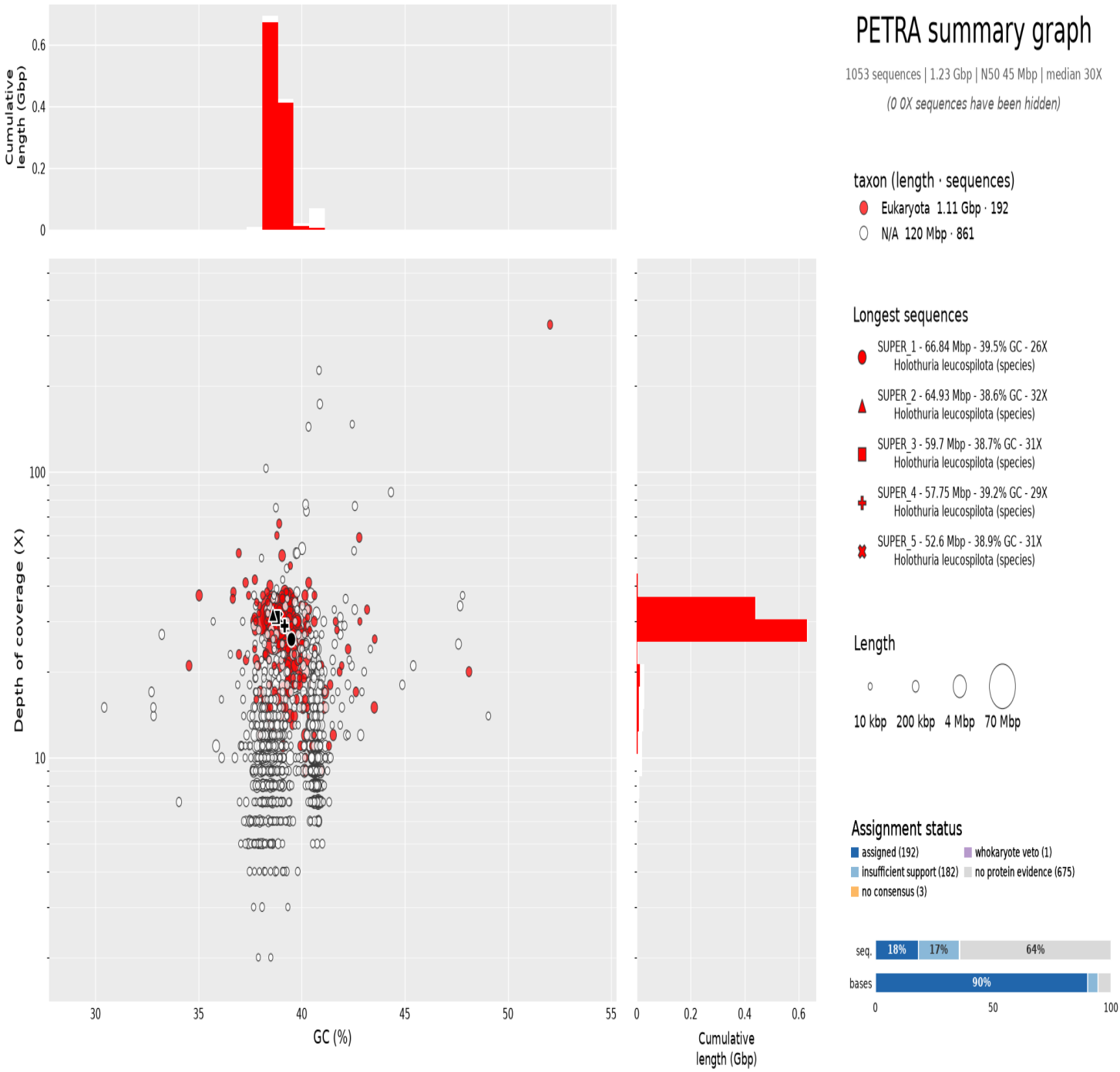


Distribution of k-mer counts per copy numbers found in asm



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	Long reads	Arima
Coverage	29	147

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:* 0.2.4
    |_ *key param:* NA
- **PretextView**
    |_ *ver:* 1.0.7
    |_ *key param:* NA

Submitter: Caroline Menguy
Affiliation: Genoscope

Date and time: 2026-10-08 11:16:57 CEST