ERGA Assembly Report

v24.10.15

Tags: ATLASea[INVALID TAG]

TxID	1316109
ToLID	**ehEuaLapp1**
Species	Euapta lappa
Class	Holothuroidea
Order	Apodida

Genome Traits	Expected	Observed
Haploid size (bp)	1,072,889,863	1,144,102,550
Haploid Number	21 (source: ancestor)	19
Ploidy	2 (source: ancestor)	2
Sample Sex	Unknown	Unknown

EBP metrics summary and curation notes

Obtained EBP quality metric for collapsed: 7.7.Q46

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

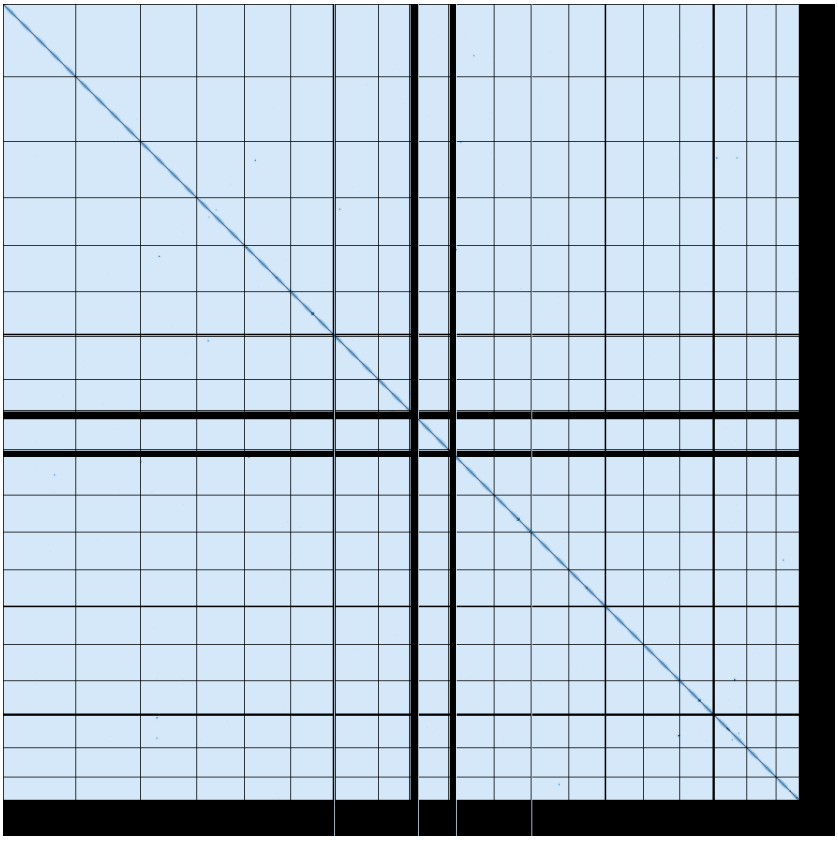
. Interventions/Gb: 76
. Contamination notes: ""
. Other observations: "The assembly of Euapta lappa (ehEuaLapp1) is based on 52X PacBio data and 130X Arima Hi-C data generated as part of the ATLASea programme (https://www.atlasea.fr). The assembly process included the following steps: initial PacBio 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, 8 contigs were identified as contaminants (bacterial, archaeal, or viral), totaling 0.249 Mb (with the largest being 0.084 Mb). Additionally, 80 regions totaling 4.803 Mb (with the largest being 0.686 Mb) were identified as haplotypic duplications and removed. Mitochondrial genome was assembled using OATK. Finally, the primary assembly was analyzed and manually improved using Pretext. During manual curation, 25 haplotypic regions and 13 contaminant sequences were removed, totaling 22.52Mb and 0.924Mb, respectively (with the largest being 4.822Mb and 0.207Mb). The telomeric pattern TTAGGG was identified with TelFinder and used to generate the Pretext Telomeres tracks. 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	1,167,560,831	1,144,102,550
GC %	36.83	36.84
Gaps/Gbp	191	215.89
Total gap bp	22,300	30,000
Scaffolds	742	686
Scaffold N50	51,503,300	51,687,591
Scaffold L50	9	9
Scaffold L90	25	18
Contigs	965	933
Contig N50	15,064,000	15,064,000
Contig L50	24	24
Contig L90	102	95
QV	46.2447	46.2464
Kmer compl.	72.4612	72.0128
BUSCO sing.	93.6%	96.0%
BUSCO dupl.	2.7%	0.7%
BUSCO frag.	2.1%	1.8%
BUSCO miss.	1.6%	1.5%

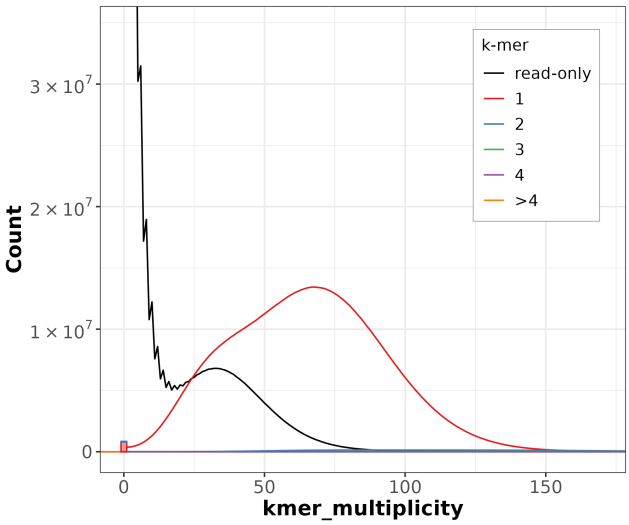
BUSCO: 6.0.0 (euk_genome_min, miniprot) / Lineage: metazoa_odb12 (genomes:206, BUSCOs:672)

HiC contact map of curated assembly

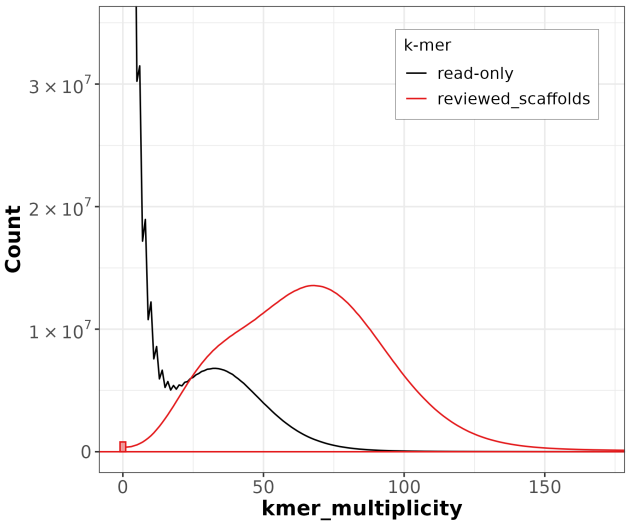


collapsed [LINK]

K-mer spectra of curated assembly

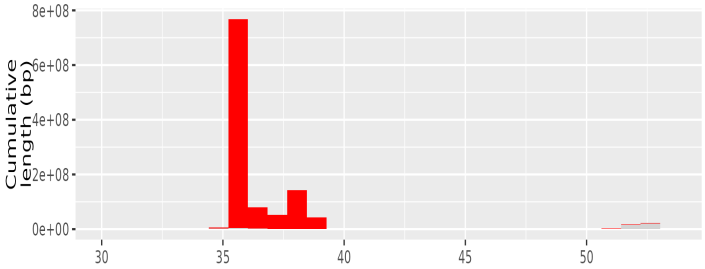


Distribution of k-mer counts per copy numbers found in asm

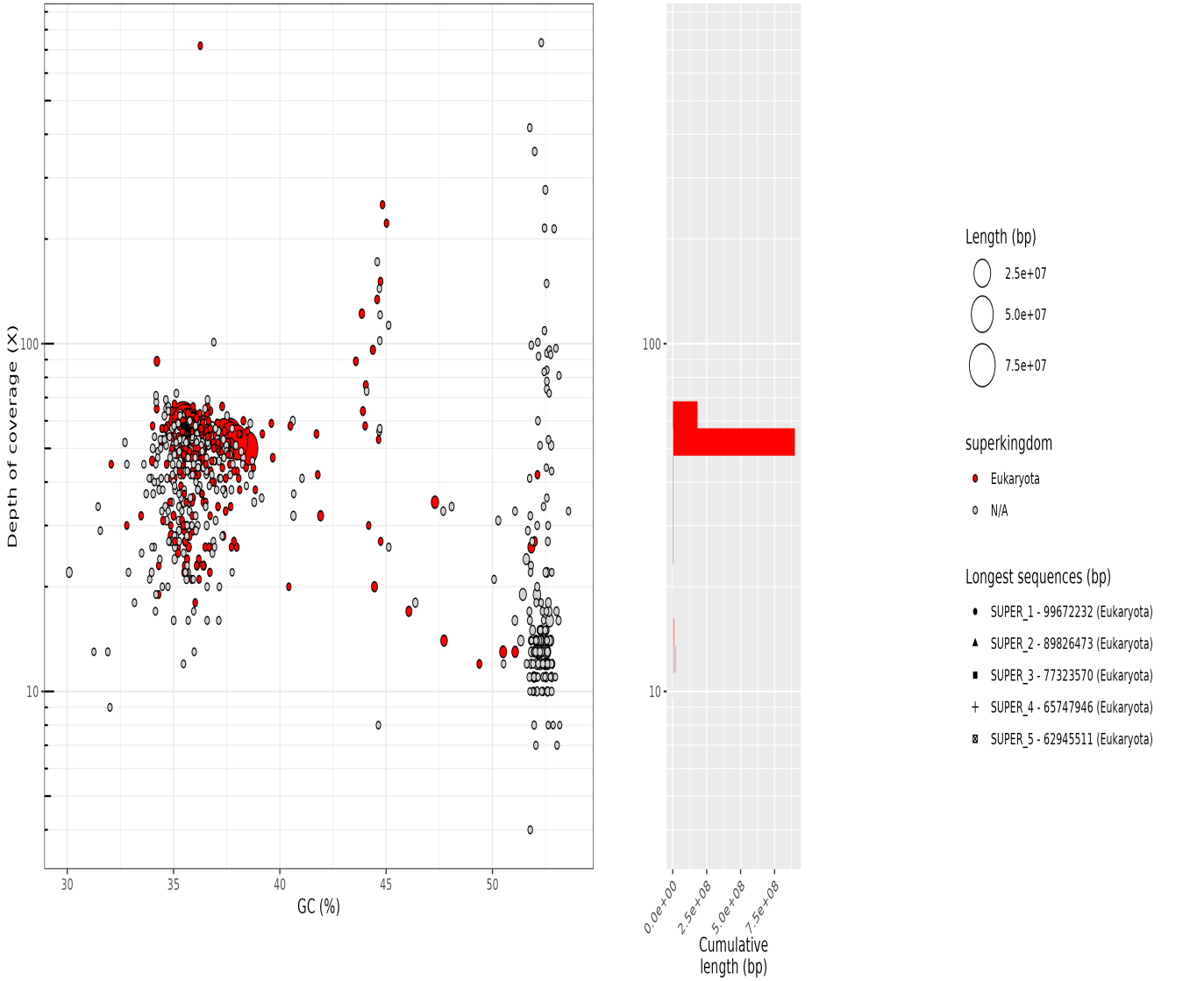


Distribution of k-mer counts coloured by their presence in reads/assemblies

Post-curation contamination screening



TAPAs summary Graph



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	53	133

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: Sophie Layac
Affiliation: Genoscope

Date and time: 2026-05-13 20:05:03 CEST