ERGA Assembly Report

v24.10.15

Tags: ATLASea[INVALID TAG]

TxID	177573
ToLID	**odHymPerl2**
Species	Hymeniacidon perlevis
Class	Demospongiae
Order	Suberitida

Genome Traits	Expected	Observed
Haploid size (bp)	130,263,438	133,522,009
Haploid Number	5 (source: ancestor)	19
Ploidy	2 (source: ancestor)	2
Sample Sex	Unknown	Unknown

EBP metrics summary and curation notes

Obtained EBP quality metric for collapsed: 6.6.Q40

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
. BUSCO single copy value is less than 90% for collapsed
. Assembly length loss > 3% for collapsed

Curator notes

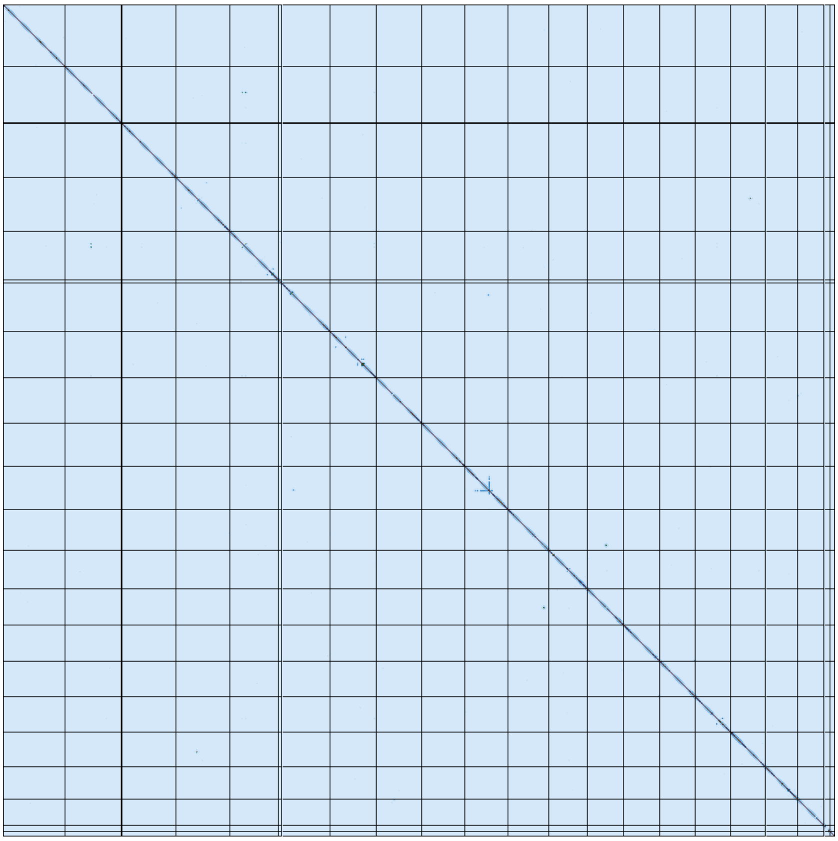
. Interventions/Gb: 303
. Contamination notes: ""
. Other observations: "The assembly of Hymeniacidon perlevis (odHymPerl2) is based on 61X PacBio data and 233X Arima Hi-C data generated as part of the ATLASea programme (https://www.atlasea.fr).The assembly process included the following steps: initial 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, 848 contigs were identified as contaminants (bacterial, archaeal, or viral), totaling 41.2 Mb (with the largest being 0.993Mb). Additionally, 131 regions totaling 7.6 Gb (with the largest being 0.352 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, 16 haplotypic regions were removed, totaling 17.5 Gb (with the largest being 7.1 Mb)and 175 contigs were identified as contaminants (Bacteria or Platyhelminthes). 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	155,917,147	133,522,009
GC %	42.19	42.35
Gaps/Gbp	667.02	606.64
Total gap bp	10,400	9,600
Scaffolds	225	28
Scaffold N50	7,847,659	7,025,240
Scaffold L50	7	9
Scaffold L90	17	17
Contigs	329	109
Contig N50	3,485,282	3,657,595
Contig L50	16	13
Contig L90	65	50
QV	33.5123	40.5428
Kmer compl.	69.8309	63.4219
BUSCO sing.	56.2%	81.7%
BUSCO dupl.	1.8%	1.6%
BUSCO frag.	24.7%	5.2%
BUSCO miss.	17.3%	11.5%

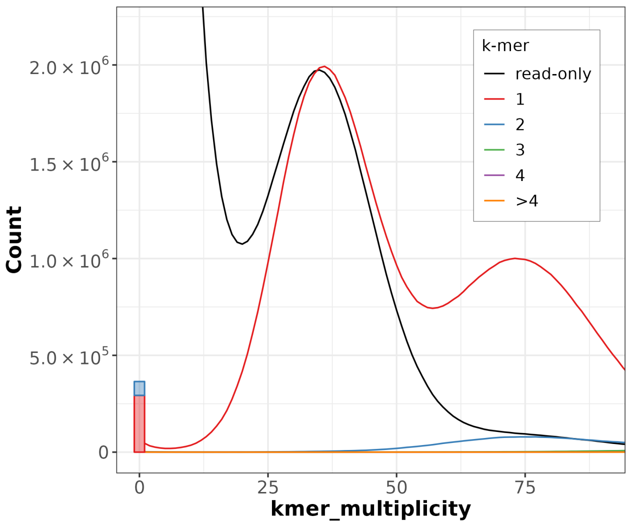
Warning! BUSCO versions or lineage datasets are not the same across results:
BUSCO: 5.8.2 (euk_genome_met, metaeuk) / Lineage: metazoa_odb12 (genomes:206, BUSCOs:672)
BUSCO: 6.0.0 (euk_genome_min, miniprot) / Lineage: metazoa_odb12 (genomes:206, BUSCOs:672)

HiC contact map of curated assembly

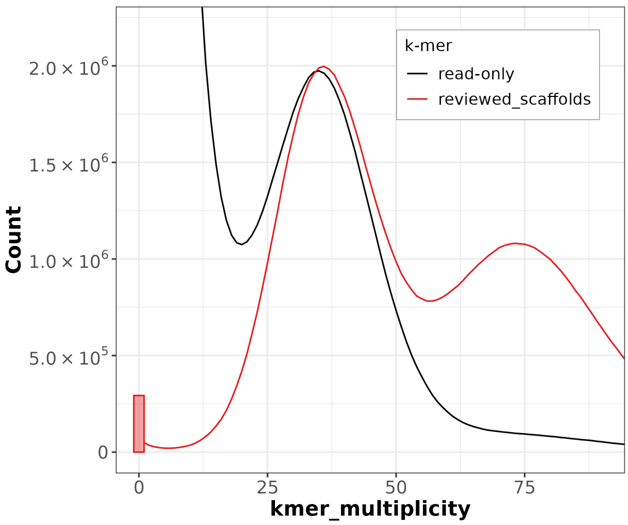


collapsed [LINK]

K-mer spectra of curated assembly

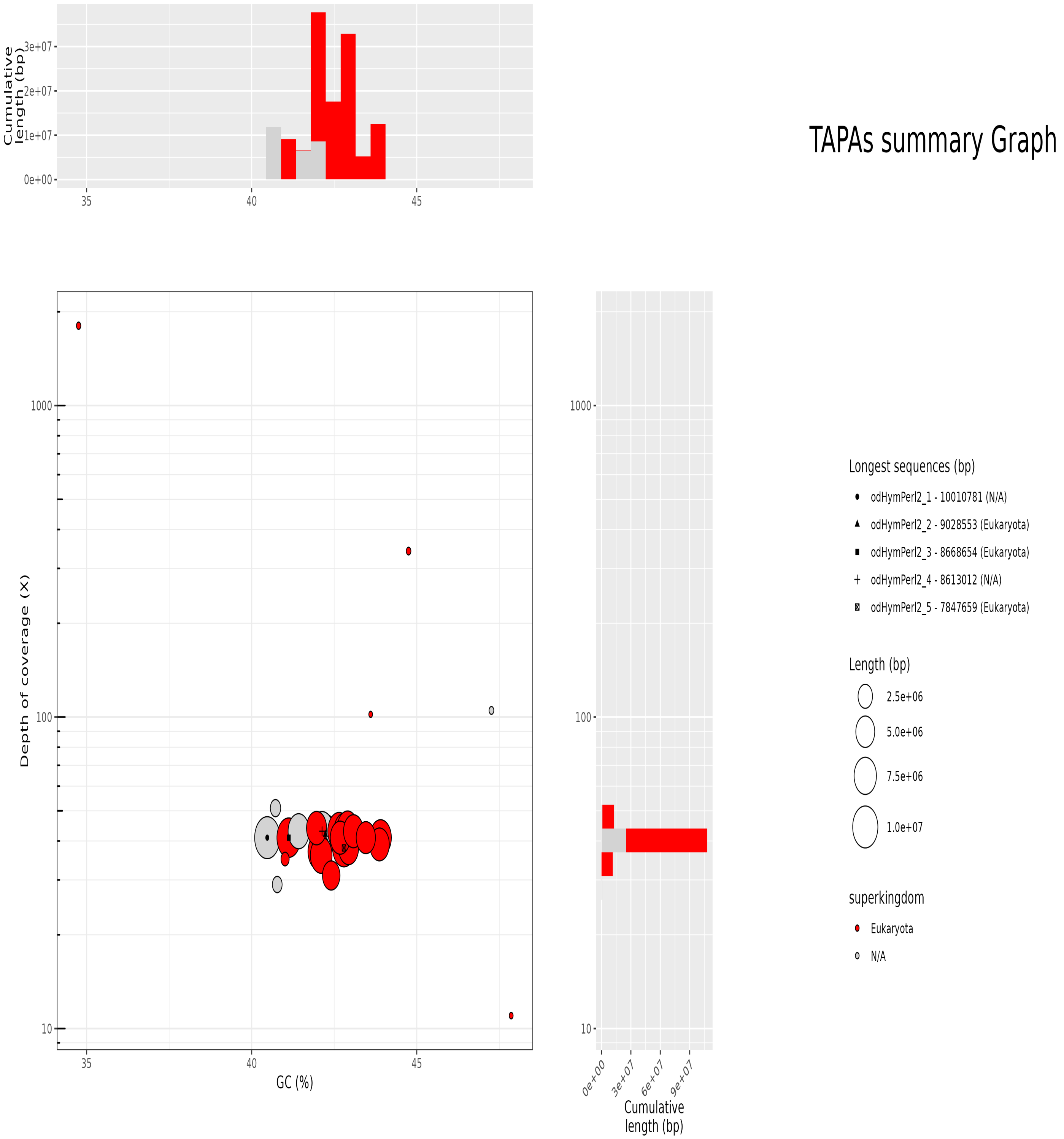


Distribution of k-mer counts per copy numbers found in asm



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	61	233

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 Menguy
Affiliation: Genoscope

Date and time: 2025-10-09 07:11:18 CEST