ERGA Assembly Report

v24.10.15

Tags: ATLASea[INVALID TAG]

TxID	475381
ToLID	**tzMyrTrun3**
Species	Myriapora truncata
Class	Gymnolaemata
Order	Cheilostomatida

Genome Traits	Expected	Observed
Haploid size (bp)	1,025,082,291	1,071,630,790
Haploid Number	11 (source: ancestor)	11
Ploidy	2 (source: ancestor)	2
Sample Sex	Unknown	Unknown

EBP metrics summary and curation notes

Obtained EBP quality metric for collapsed: 6.7.Q44

The following metrics were automatically flagged as below EBP recommended standards or different from expected:

. Kmer completeness value is less than 90 for collapsed

Curator notes

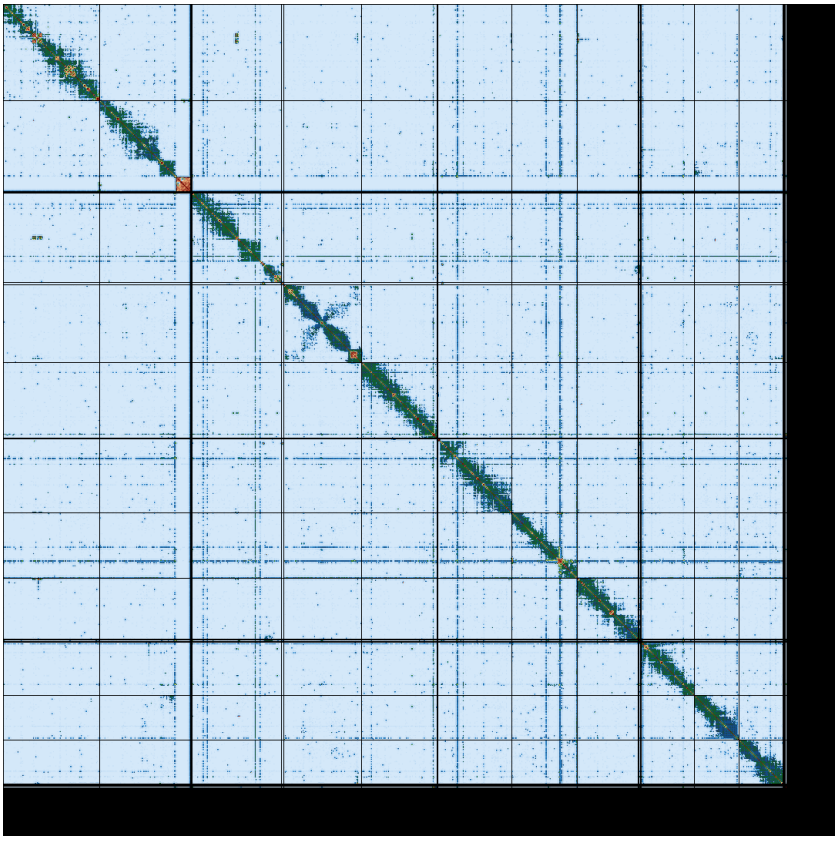
. Interventions/Gb: 133
. Contamination notes: ""
. Other observations: "The assembly of Myriapora truncata (tzMyrTrun3) is based on 53.02X PacBio data and 131.67X Arima Hi-C data generated as part of the ATLASea programme (https://www.atlasea.fr). The assembly process included the following steps: initial 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, 309 contigs were identified as contaminants (bacterial, archaeal, or viral), totaling 28.448 Mb (with the largest being 5.552 Mb). Additionally, 732 regions totaling 68.67 Mb (with the largest being 1.279 Mb) were identified as haplotypic duplications and removed. Finally, the primary assembly was analyzed and manually improved using Pretext. The mitochondrial genome was assembled using OATK. During manual curation, 22 haplotypic regions and 444 contaminant sequences were removed, totaling 26.617Mb and 71.267Mb, respectively (with the largest being 4.245Mb and 1.53Mb). 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	1,102,561,451	1,071,630,790
GC %	36.13	36.19
Gaps/Gbp	179.58	258.48
Total gap bp	19,800	37,700
Scaffolds	551	430
Scaffold N50	40,875,722	96,719,149
Scaffold L50	10	5
Scaffold L90	40	11
Contigs	749	707
Contig N50	9,263,811	9,148,470
Contig L50	35	34
Contig L90	141	135
QV	44.5831	44.8584
Kmer compl.	77.7498	76.9712
BUSCO sing.	92.0%	93.1%
BUSCO dupl.	2.7%	1.7%
BUSCO frag.	2.1%	2.1%
BUSCO miss.	3.2%	3.2%

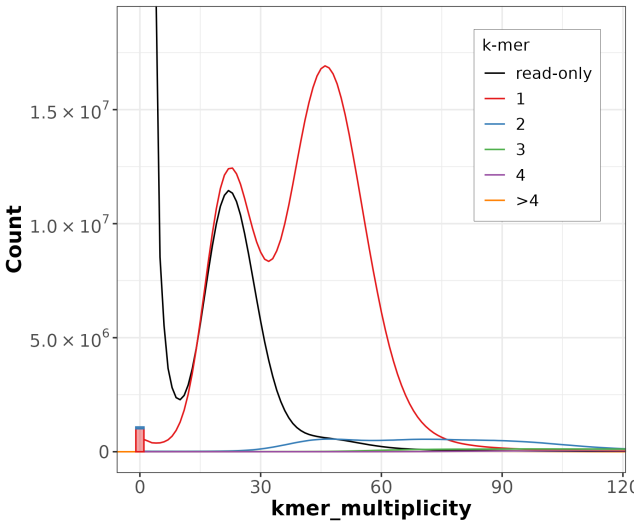
Warning! BUSCO versions or lineage datasets are not the same across results:
BUSCO: 6.0.0 (euk_genome_min, miniprot) / Lineage: lophotrochozoa_odb12 (genomes:75, BUSCOs:1252)
BUSCO: 6.1.0 (euk_genome_min, miniprot) / Lineage: lophotrochozoa_odb12 (genomes:75, BUSCOs:1252)

HiC contact map of curated assembly

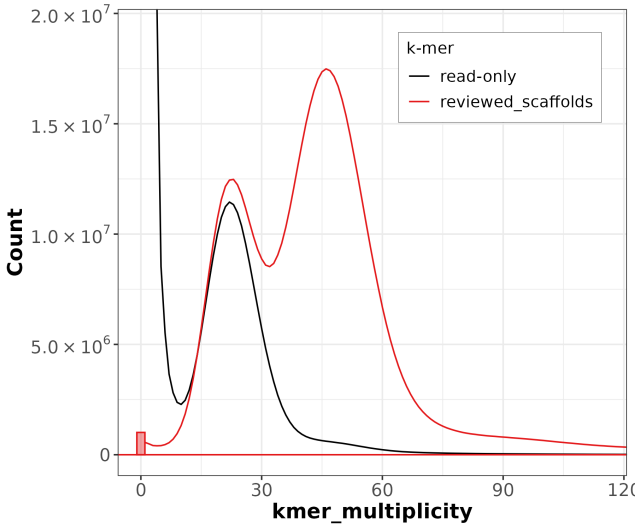


collapsed [LINK]

K-mer spectra of curated assembly

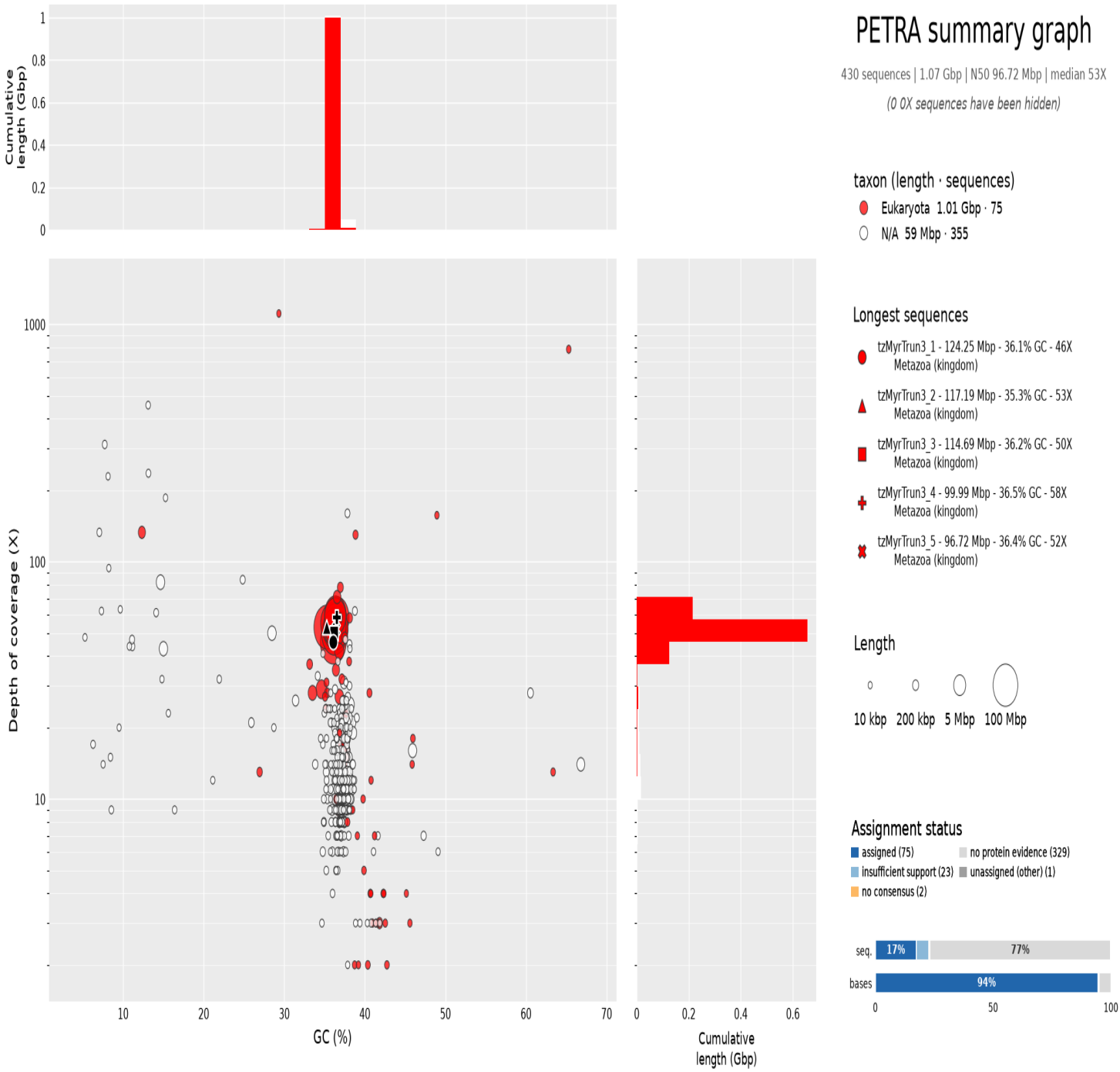


Distribution of k-mer counts per copy numbers found in asm



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	54	135

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: 2026-10-20 06:02:33 CEST