ERGA Assembly Report

v24.10.15

Tags: ATLASea[INVALID TAG]

TxID	371672
ToLID	**fMyrJac1**
Species	Myripristis jacobus
Class	Actinopteri
Order	Holocentriformes

Genome Traits	Expected	Observed
Haploid size (bp)	765,408,219	793,064,773
Haploid Number	21 (source: ancestor)	24
Ploidy	2 (source: ancestor)	2
Sample Sex	Unknown	Unknown

EBP metrics summary and curation notes

Obtained EBP quality metric for collapsed: 7.7.Q45

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

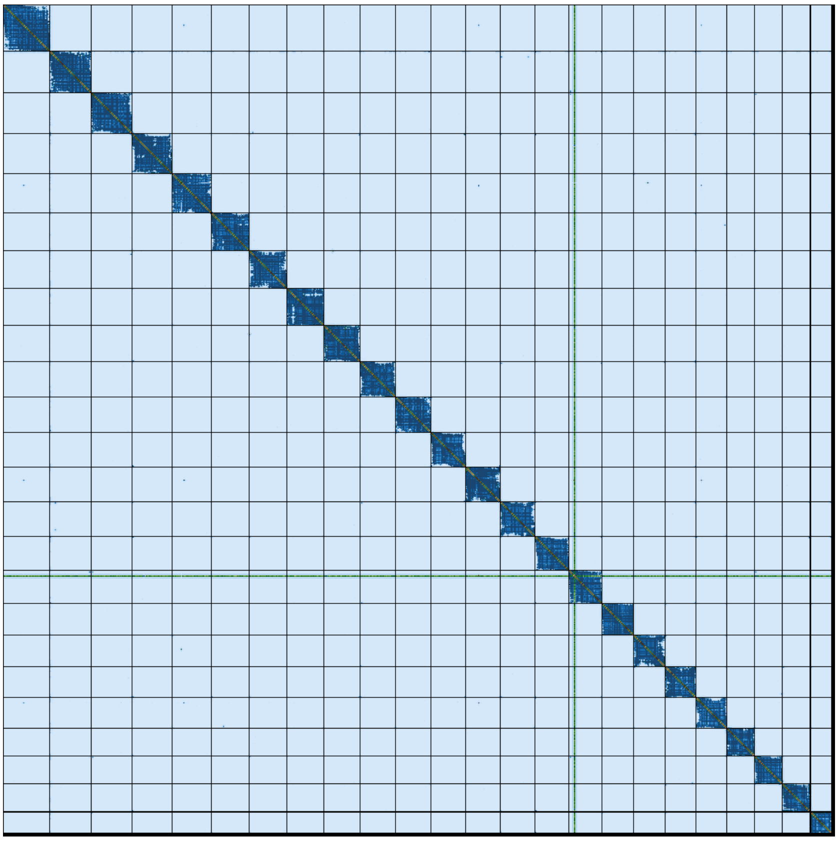
. Interventions/Gb: 37
. Contamination notes: ""
. Other observations: "The assembly of Myripristis jacobus (fMyrJac1) is based on 33X ONT data and 210X Arima Hi-C data generated as part of the ATLASea programme (https://www.atlasea.fr). The assembly process included the following steps: initial ONT assembly generation with Hifiasm, removal of contaminant sequences using Context, removal of haplotypic duplications using purge_dups, and Hi-C-based scaffolding with YaHS. No contigs were found to be contaminants (bacterial, archaeal, or viral). Additionally, 152 regions totaling 13.263 Mb (with the largest being 0.392 Mb) were identified as haplotypic duplications and removed. Mitochondrial genome was assembled using ptGAUL. During manual curation, 6 haplotypic regions were removed, totaling 1.21Mb (with the largest being 0.49Mb). 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	794,301,412	793,064,773
GC %	42.25	42.25
Gaps/Gbp	84.35	93.31
Total gap bp	6,700	8,900
Scaffolds	73	63
Scaffold N50	33,764,184	33,764,184
Scaffold L50	11	11
Scaffold L90	21	21
Contigs	140	137
Contig N50	21,996,288	21,996,288
Contig L50	15	15
Contig L90	41	41
QV	45.0391	45.0809
Kmer compl.	84.2779	84.2381
BUSCO sing.	99.2%	99.3%
BUSCO dupl.	0.5%	0.4%
BUSCO frag.	0.0%	0.0%
BUSCO miss.	0.3%	0.3%

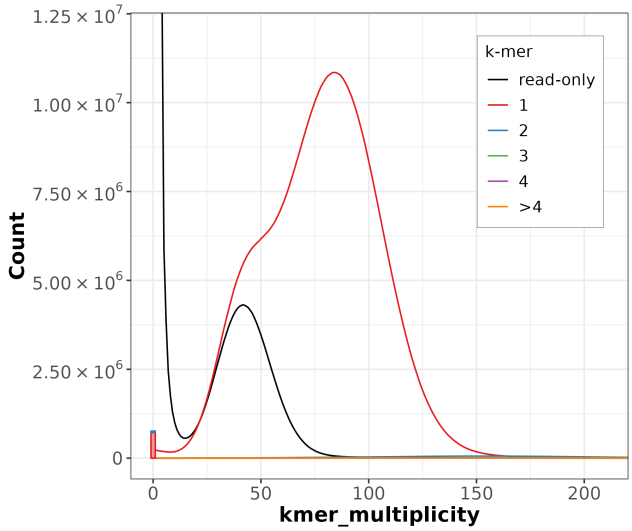
BUSCO: 6.0.0 (euk_genome_min, miniprot) / Lineage: actinopterygii_odb12 (genomes:75, BUSCOs:7207)

HiC contact map of curated assembly

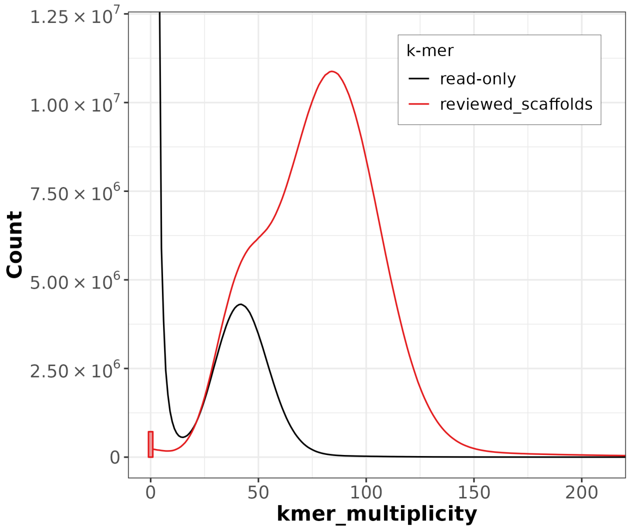


collapsed [LINK]

K-mer spectra of curated assembly

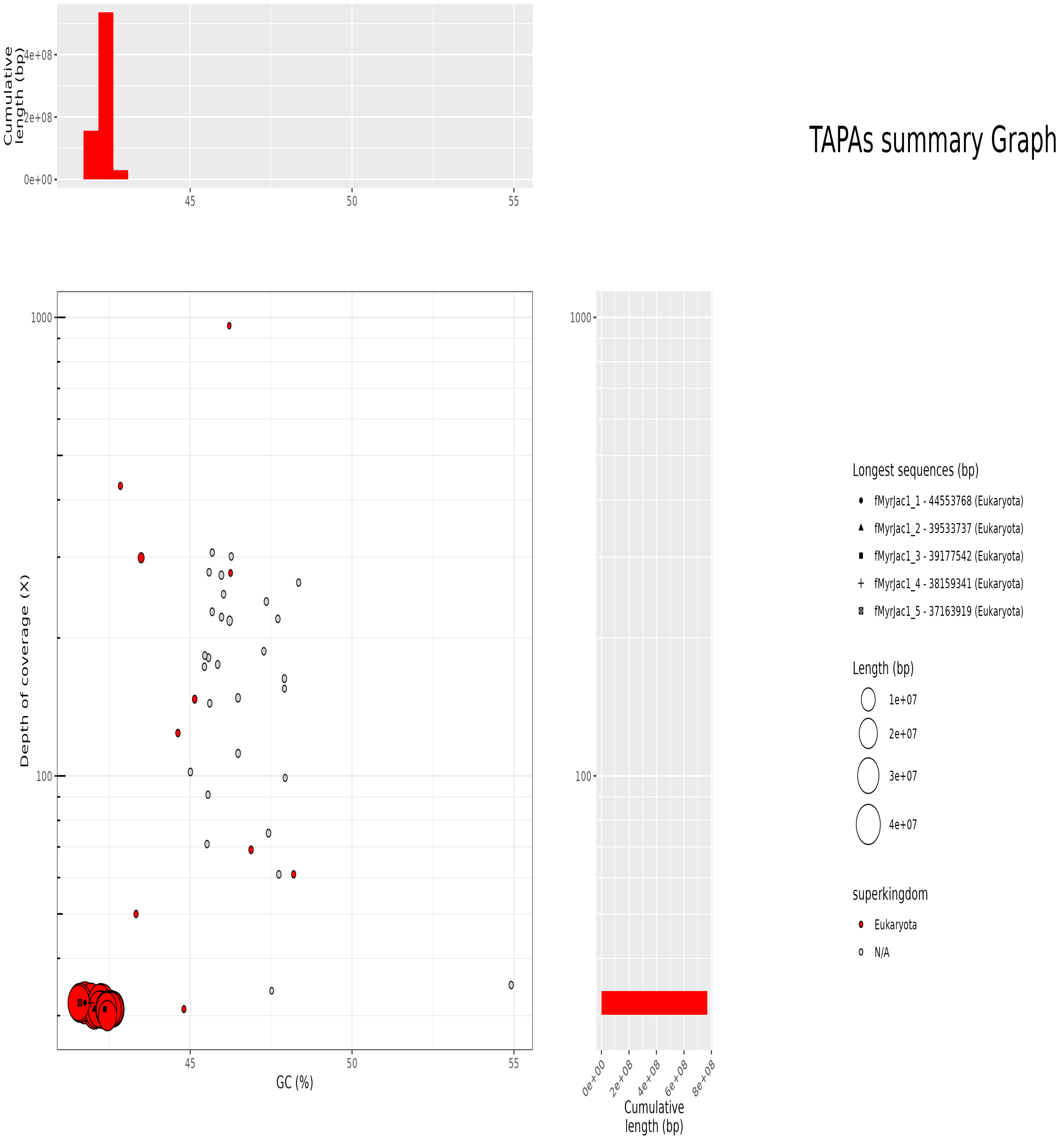


Distribution of k-mer counts per copy numbers found in asm



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	33	210

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: Jean-Marc Aury
Affiliation: Genoscope

Date and time: 2025-12-04 07:35:24 CET