ERGA Assembly Report

v24.10.15

Tags: ATLASea[INVALID TAG]

TxID	721872
ToLID	**fMyrOce1**
Species	Myrichthys ocellatus
Class	Actinopteri
Order	Anguilliformes

Genome Traits	Expected	Observed
Haploid size (bp)	1,472,038,263	1,515,218,214
Haploid Number	23 (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
. BUSCO single copy value is less than 90% for collapsed
. BUSCO duplicated value is more than 5% for collapsed

Curator notes

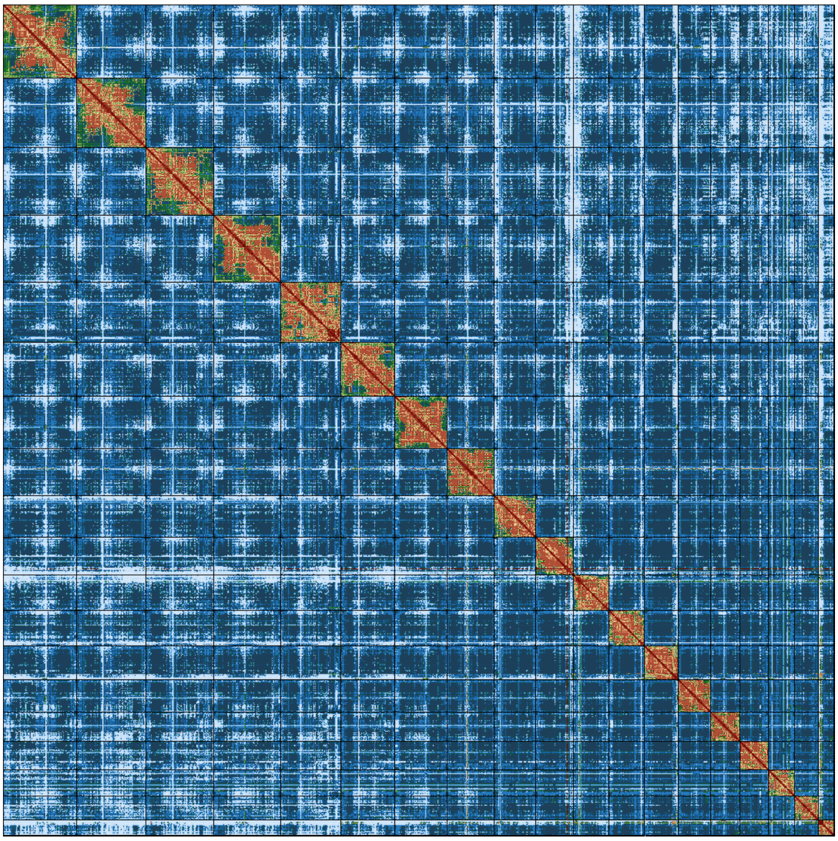
. Interventions/Gb: 3
. Contamination notes: ""
. Other observations: "The assembly of Myrichthys ocellatus (fMyrOce1) is based on 55X PacBio data and 176X Arima Hi-C data generated as part of the ATLASea programme (https://www.atlasea.fr). The assembly process included the following steps: initial PacBio 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 were identified as contaminants (bacterial, archaeal, or viral), totaling 0.045 Mb (with the largest being 0.045 Mb). Additionally, 193 regions totaling 28.781 Mb (with the largest being 3.542 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. 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	1,515,205,644	1,515,218,214
GC %	42.81	42.81
Gaps/Gbp	95.04	97.02
Total gap bp	14,400	15,100
Scaffolds	42	38
Scaffold N50	95,401,461	95,401,461
Scaffold L50	7	7
Scaffold L90	16	16
Contigs	186	185
Contig N50	47,180,267	47,180,267
Contig L50	9	9
Contig L90	51	51
QV	46.5358	46.5363
Kmer compl.	76.3958	76.3967
BUSCO sing.	86.7%	86.7%
BUSCO dupl.	7.3%	7.3%
BUSCO frag.	1.8%	1.8%
BUSCO miss.	4.3%	4.3%

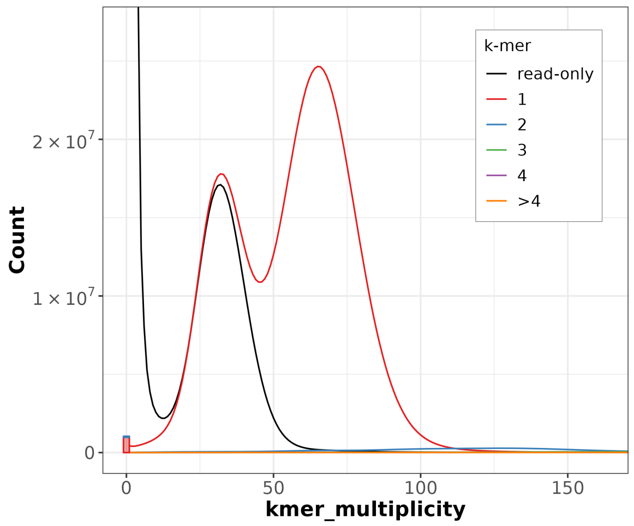
BUSCO: 5.8.2 (euk_genome_met, metaeuk) / Lineage: actinopterygii_odb12 (genomes:75, BUSCOs:7207)

HiC contact map of curated assembly

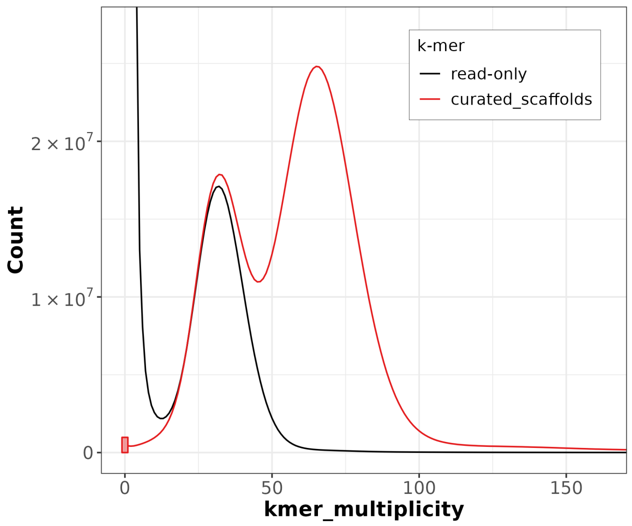


collapsed [LINK]

K-mer spectra of curated assembly

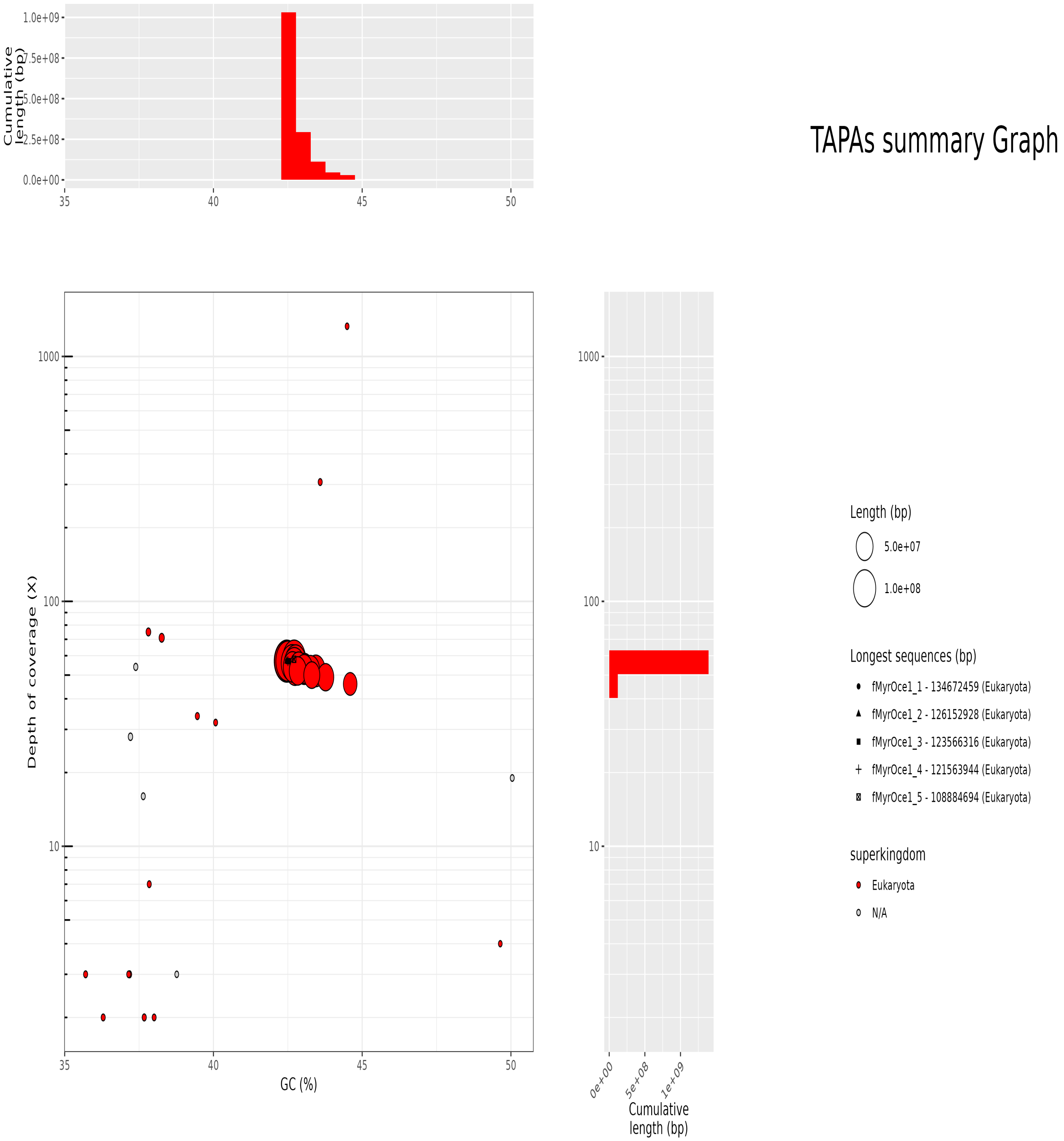


Distribution of k-mer counts per copy numbers found in asm



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	PACBIO Hifi	Arima
Coverage	55	175

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: Simone Duprat
Affiliation: Genoscope

Date and time: 2025-08-31 04:14:41 CEST