

ERGA Assembly Report

v24.10.15

Tags: ATLASEa[INVALID TAG]

TxID	303715
ToLID	fMonCil1
Species	Monacanthus ciliatus
Class	Actinopteri
Order	Tetraodontiformes

Genome Traits	Expected	Observed
Haploid size (bp)	485,582,816	476,314,676
Haploid Number	17 (source: ancestor)	22
Ploidy	2 (source: ancestor)	2
Sample Sex	Unknown	Unknown

EBP metrics summary and curation notes

Obtained EBP quality metric for collapsed: 6.7.Q34

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

- . Observed Haploid Number is different from Expected
- . QV value is less than 40 for collapsed
- . Kmer completeness value is less than 90 for collapsed

Curator notes

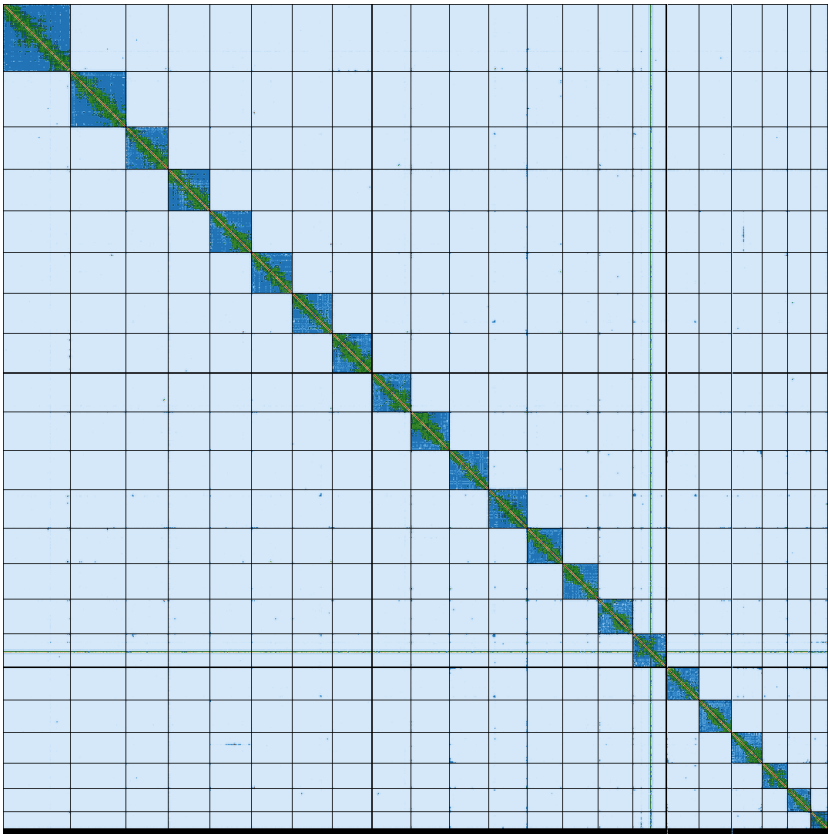
. Interventions/Gb: 105
. Contamination notes: ""
. Other observations: "The assembly of *Monacanthus ciliatus* (fMonCil1) is based on 34X ONT data and 225X Arima Hi-C data generated as part of the ATLASEa programme (<https://www.atlasea.fr>). The assembly process included the following steps: initial PacBio assembly generation with Hifiiasm, removal of contaminant sequences using Context, removal of haplotypic duplications using purge_dups, and Hi-C-based scaffolding with YaHS. In total, no contigs were identified as contaminants (bacterial, archaeal, or viral). Additionally, 317 regions totaling 8.4 Mb (with the largest being 0.229 Mb) were identified as haplotypic duplications and removed. The mitochondrial genome was assembled using ptGAUL. Finally, the primary assembly was analyzed and manually improved using Pretext. During manual curation, no supplementary haplotypic regions were removed. 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	476,339,152	476,314,676
GC %	48.09	48.09
Gaps/Gbp	848.14	894.37
Total gap bp	40,400	46,200
Scaffolds	174	146
Scaffold N50	22,210,013	22,210,013
Scaffold L50	10	10
Scaffold L90	19	19
Contigs	578	572
Contig N50	3,389,092	3,389,092
Contig L50	42	42
Contig L90	170	170
QV	34.6849	34.685
Kmer compl.	87.0954	87.094
BUSCO sing.	98.2%	98.2%
BUSCO dupl.	0.4%	0.4%
BUSCO frag.	0.2%	0.2%
BUSCO miss.	1.2%	1.2%

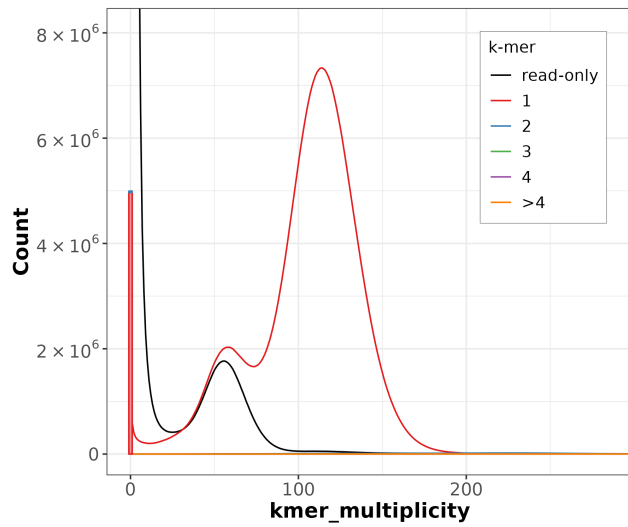
BUSCO: 6.0.0 (euk_genome_min, miniprot) / Lineage: actinopterygii_odb12 (genomes:75, BUSCOs:7207)

HiC contact map of curated assembly

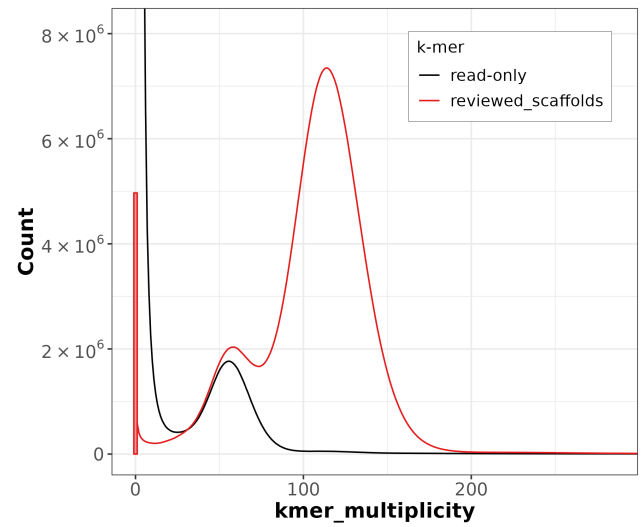


collapsed [\[LINK\]](#)

K-mer spectra of curated assembly

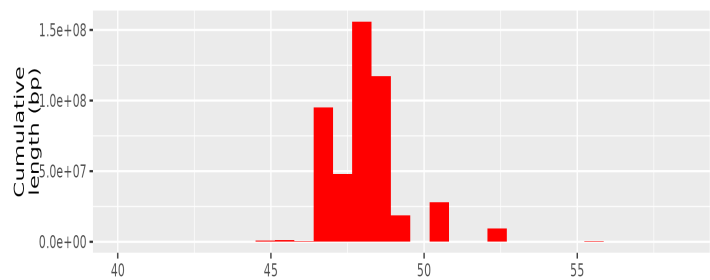


Distribution of k-mer counts per copy numbers found in asm

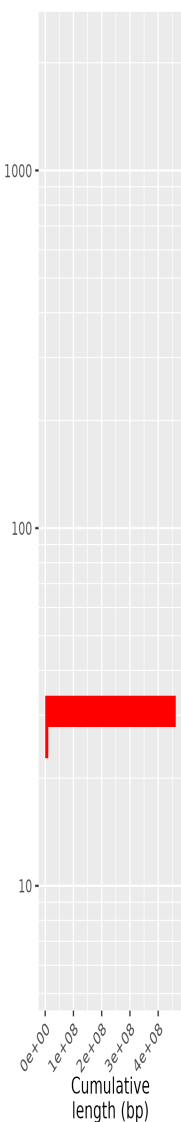
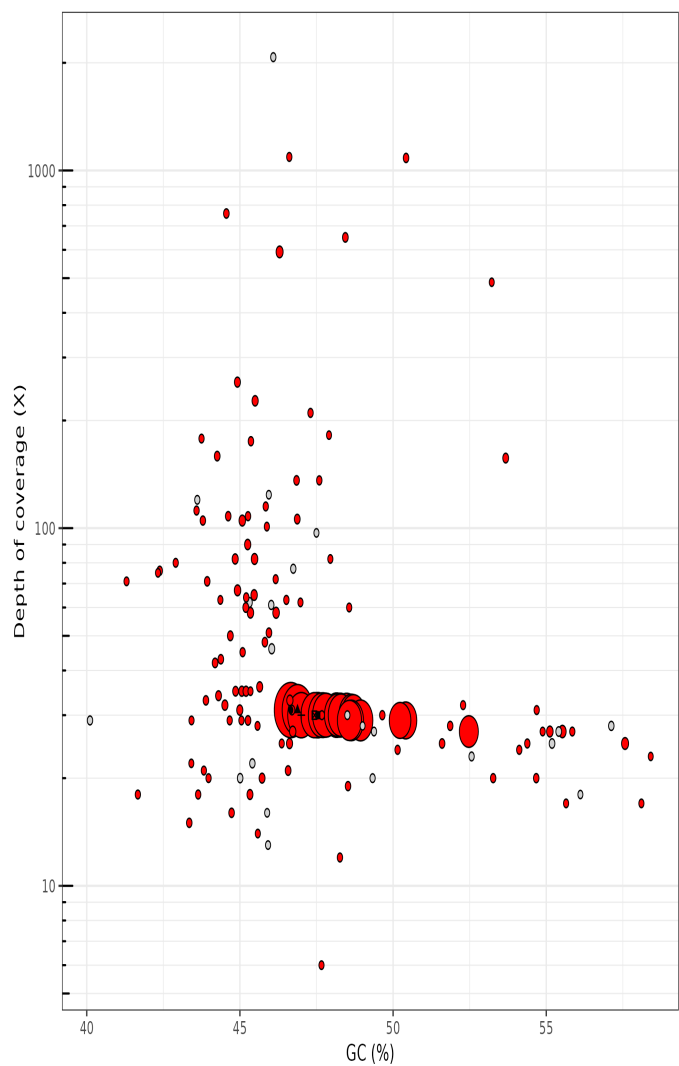


Distribution of k-mer counts coloured by their presence in reads/assemblies

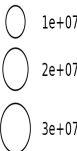
Post-curation contamination screening



TAPAs summary Graph



Length (bp)



superkingdom



Longest sequences (bp)

- fMonTuc1_1 - 38468255 (Eukaryota)
- ▲ fMonTuc1_2 - 32100646 (Eukaryota)
- fMonTuc1_3 - 24081325 (Eukaryota)
- + fMonTuc1_4 - 23980431 (Eukaryota)
- ⊠ fMonTuc1_5 - 23781557 (Eukaryota)

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	34	225

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: 2025-11-01 08:30:02 CET