ERGA Assembly Report

v24.10.15

Tags: ATLASea[INVALID TAG]

TxID	182438
ToLID	**fGymMor1**
Species	Gymnothorax moringa
Class	Actinopteri
Order	Anguilliformes

Genome Traits	Expected	Observed
Haploid size (bp)	1,926,500,382	1,994,637,259
Haploid Number	21 (source: ancestor)	21
Ploidy	2 (source: ancestor)	2
Sample Sex	Unknown	Unknown

EBP metrics summary and curation notes

Obtained EBP quality metric for collapsed: 7.7.Q49

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

. Kmer completeness value is less than 90 for collapsed
. BUSCO duplicated value is more than 5% for collapsed

Curator notes

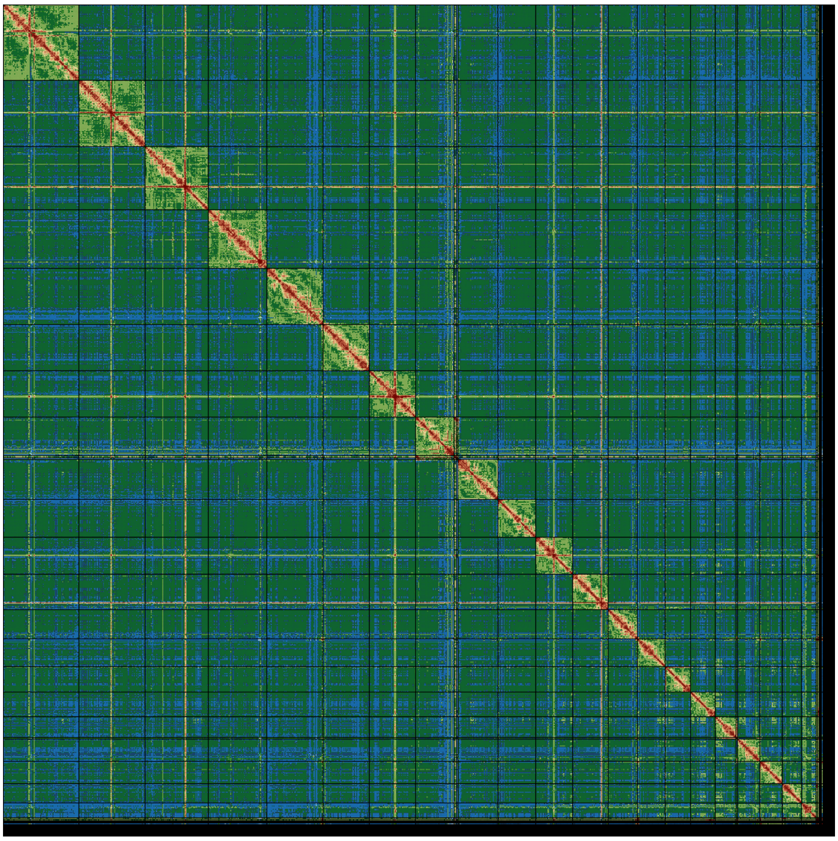
. Interventions/Gb: 16
. Contamination notes: ""
. Other observations: "The assembly of Gymnothorax moringa (fGymMor1) is based on 45X PacBio data and 441X Arima Hi-C data generated as part of the ATLASea programme (https://www.atlasea.fr). The assembly process included the following steps: initial PacBio assembly generation with Hifiasm, removal of contaminant sequences using Context, removal of haplotypic duplications using purge_dups, and Hi-C-based scaffolding with YaHS. No contigs were found to be contaminants (bacterial, archaeal, or viral). Additionally, 99 regions totaling 42.933 Mb (with the largest being 6.496 Mb) were identified as haplotypic duplications and removed. Mitochondrial genome was assembled using OATK. During manual curation, 2 haplotypic regions were removed, totaling 5.29Mb (with the largest being 3.59Mb).. 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	1,999,950,407	1,994,637,259
GC %	41.7	41.71
Gaps/Gbp	29.5	35.6
Total gap bp	5,900	9,200
Scaffolds	131	115
Scaffold N50	88,074,907	94,562,300
Scaffold L50	9	8
Scaffold L90	22	18
Contigs	190	186
Contig N50	51,686,000	51,686,000
Contig L50	15	15
Contig L90	52	51
QV	49.5907	49.5933
Kmer compl.	75.4971	75.4434
BUSCO sing.	91.5%	91.6%
BUSCO dupl.	6.9%	6.8%
BUSCO frag.	0.2%	0.2%
BUSCO miss.	1.5%	1.4%

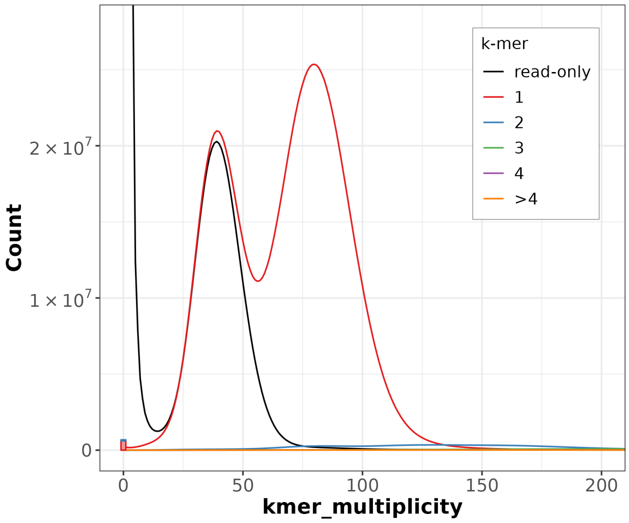
BUSCO: 6.0.0 (euk_genome_min, miniprot) / Lineage: actinopterygii_odb12 (genomes:75, BUSCOs:7207)

HiC contact map of curated assembly

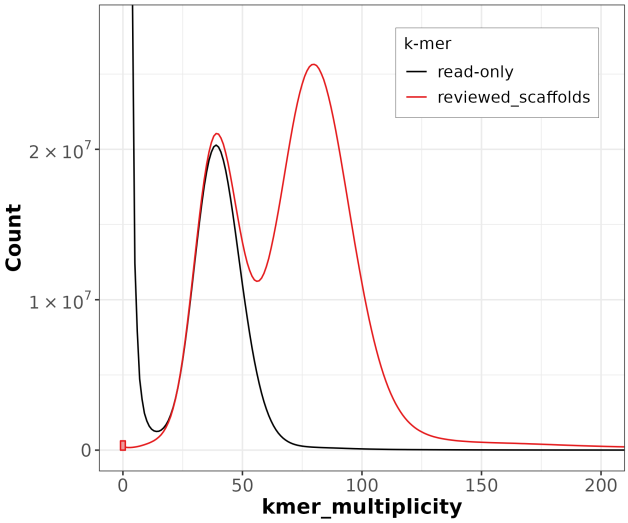


collapsed [LINK]

K-mer spectra of curated assembly

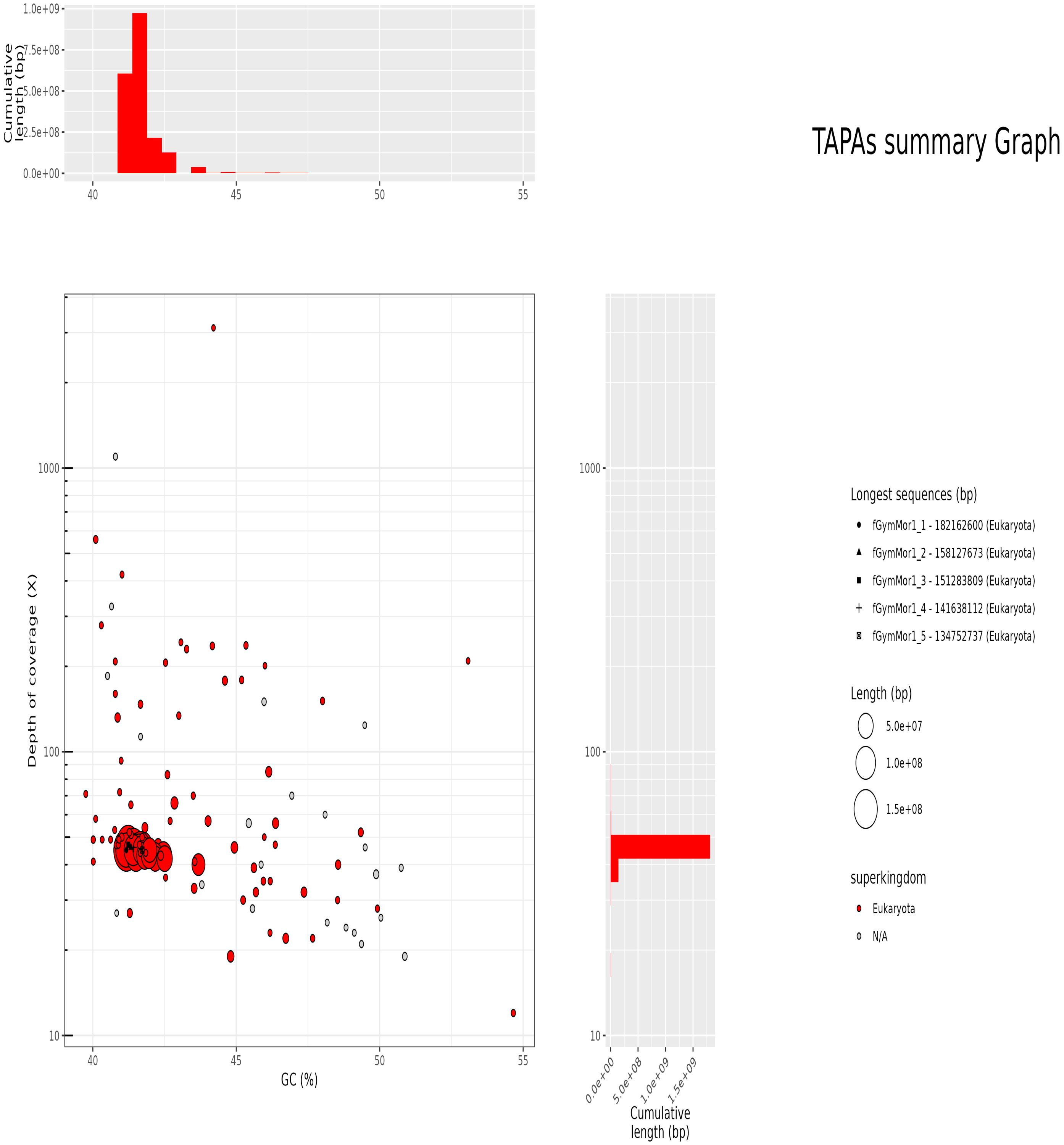


Distribution of k-mer counts per copy numbers found in asm



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	45	441

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: Arnaud Couloux
Affiliation: Genoscope

Date and time: 2026-01-11 21:37:44 CET