# ERGA Assembly Report

v24.10.15

Tags: ATLASea[INVALID TAG]

| | |
|---|---|
| TxID | 248765 |
| ToLID | **fMorMor1** |
| Species | Mora moro |
| Class | Actinopteri |
| Order | Gadiformes |

| Genome Traits | Expected | Observed |
|---|---|---|
| Haploid size (bp) | 569,005,204 | 549,828,115 |
| Haploid Number | 23 (source: ancestor) | 24 |
| Ploidy | 2 (source: ancestor) | 2 |
| Sample Sex | Unknown | Unknown |

## EBP metrics summary and curation notes

Obtained EBP quality metric for collapsed: 6.7.Q42

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

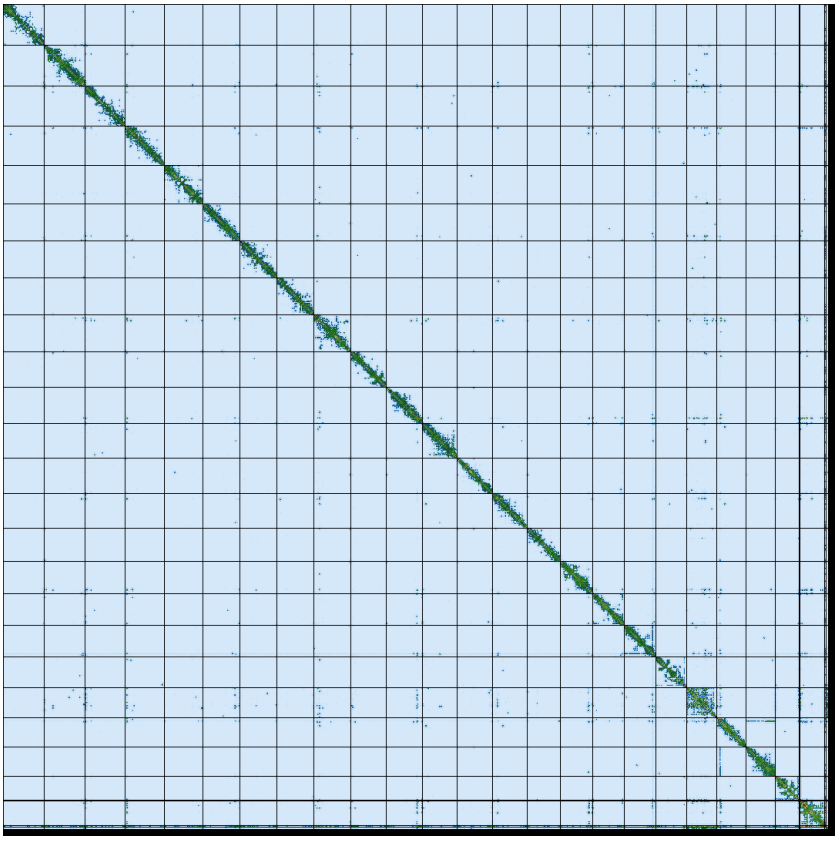
. Interventions/Gb: 15
. Contamination notes: ""
. Other observations: "The assembly of Mora moro (fMorMor1) is based on 30X ONT data and 152X Arima Hi-C data generated as part of the ATLASea programme (https://www.atlasea.fr). The assembly process included the following steps: initial ONT assembly generation with Nextdenovo, removal of contaminant sequences using Context, removal of haplotypic duplications using purge_dups, and Hi-C-based scaffolding with YaHS. In total, 1 contigs was identified as contaminants (bacterial, archaeal, or viral), totaling 0.032 Mb (with the largest being 0.032 Mb). Additionally, 190 regions totaling 25.806 Mb (with the largest being 0.418 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. 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 | 549,820,802 | 549,828,115 |
| GC % | 46.97 | 46.97 |
| Gaps/Gbp | 458.33 | 461.96 |
| Total gap bp | 25,200 | 25,900 |
| Scaffolds | 51 | 49 |
| Scaffold N50 | 23,482,639 | 23,482,639 |
| Scaffold L50 | 11 | 11 |
| Scaffold L90 | 22 | 22 |
| Contigs | 303 | 303 |
| Contig N50 | 7,705,494 | 7,705,494 |
| Contig L50 | 20 | 20 |
| Contig L90 | 95 | 95 |
| QV | 42.0568 | 42.0568 |
| Kmer compl. | 88.2325 | 88.2357 |
| BUSCO sing. | 95.4% | 95.4% |
| BUSCO dupl. | 1.0% | 1.0% |
| BUSCO frag. | 0.1% | 0.1% |
| BUSCO miss. | 3.5% | 3.5% |

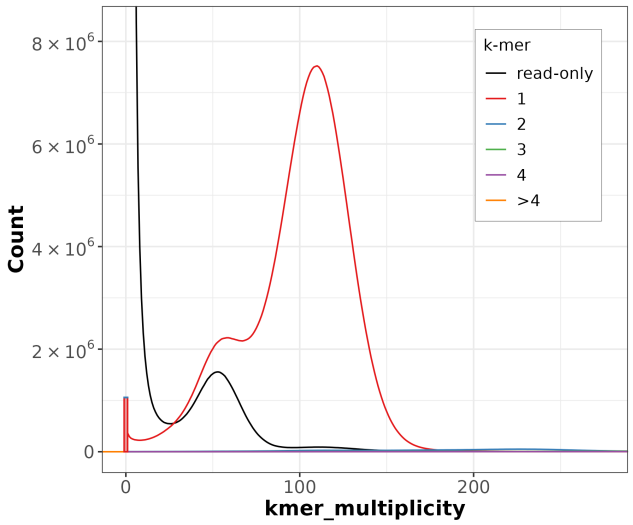
BUSCO: 6.1.0 (euk_genome_min, miniprot) / Lineage: actinopterygii_odb12.2 (genomes:79, BUSCOs:9048)

# HiC contact map of curated assembly

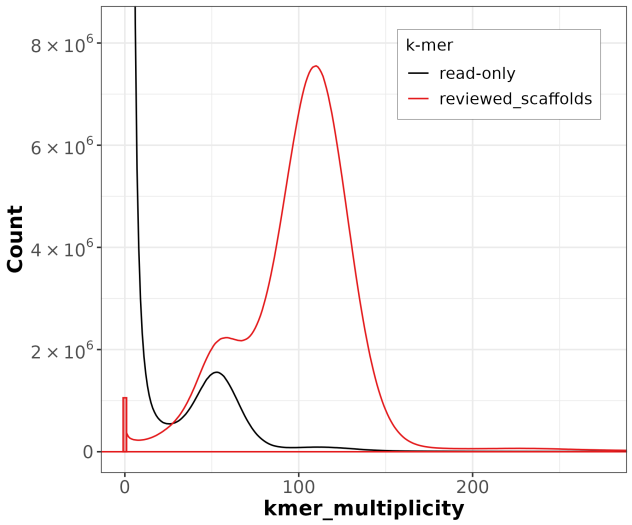


**collapsed** [LINK]

# K-mer spectra of curated assembly

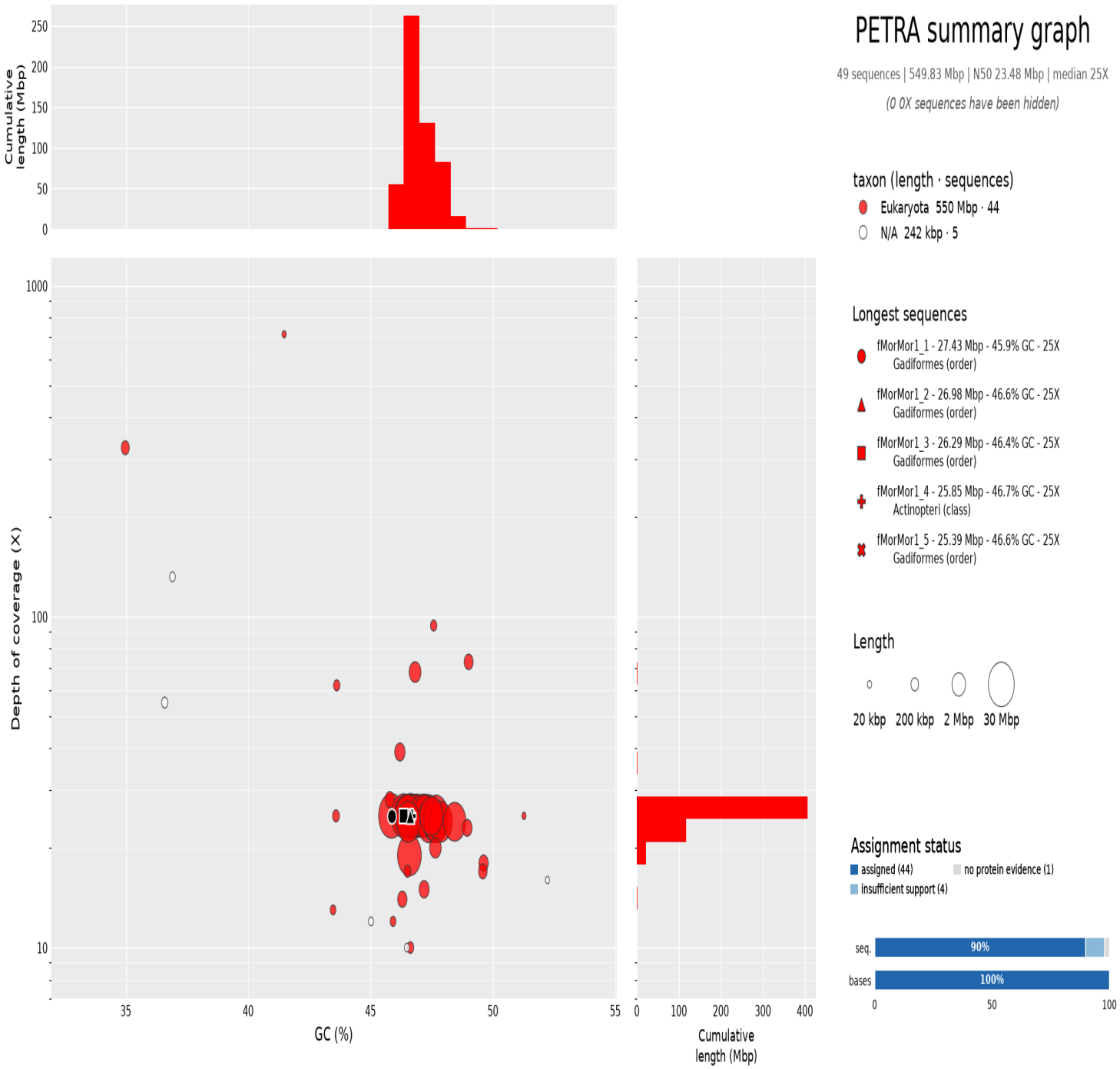



Distribution of k-mer counts per copy numbers found in asm




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 | 30 | 152 |

# 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: Arnaud Couloux
Affiliation: Genoscope

Date and time: 2026-11-04 01:47:55 CET