ERGA Assembly Report

v24.10.15

Tags: ATLASea[INVALID TAG]

TxID	348445
ToLID	**fMolMac1**
Species	Molva macrophthalma
Class	Actinopteri
Order	Gadiformes

Genome Traits	Expected	Observed
Haploid size (bp)	536,436,979	568,165,719
Haploid Number	23 (source: ancestor)	23
Ploidy	2 (source: ancestor)	2
Sample Sex	Unknown	Unknown

EBP metrics summary and curation notes

Obtained EBP quality metric for collapsed: 7.7.Q44

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

. Assembly length loss > 3% for collapsed

Curator notes

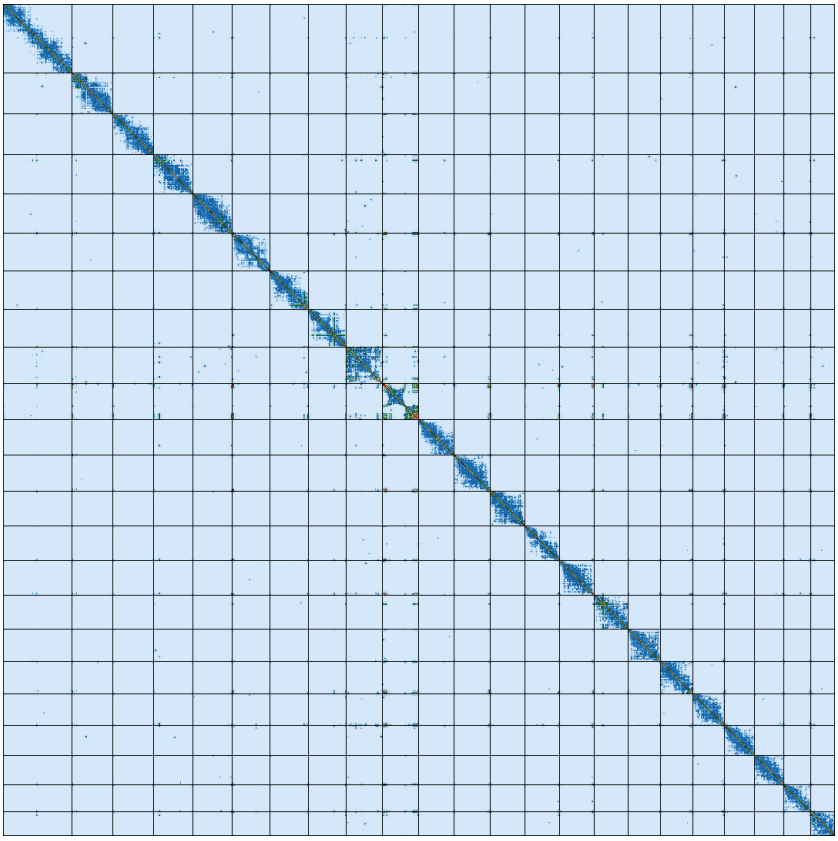
. Interventions/Gb: 73
. Contamination notes: ""
. Other observations: "The assembly of Molva macrophthalma (fMolMac1) is based on 36.3X ONT data and 148.67X Arima Hi-C data generated as part of the ATLASea programme (https://www.atlasea.fr). The assembly process included the following steps: initial 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, 117 regions totaling 24.055 Mb (with the largest being 3.424 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, 8 haplotypic regions and 4 contaminant sequences were removed, totaling 17.804Mb and 0.232Mb, respectively (with the largest being 8.461Mb and 0.104Mb). 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	586,179,449	568,165,719
GC %	44.96	44.96
Gaps/Gbp	35.83	79.2
Total gap bp	2,100	7,800
Scaffolds	58	24
Scaffold N50	23,880,295	24,465,061
Scaffold L50	11	11
Scaffold L90	22	20
Contigs	79	69
Contig N50	23,436,297	23,436,297
Contig L50	11	11
Contig L90	26	24
QV	44.3812	44.4099
Kmer compl.	96.3361	94.8593
BUSCO sing.	96.2%	97.6%
BUSCO dupl.	2.4%	0.5%
BUSCO frag.	0.2%	0.2%
BUSCO miss.	1.2%	1.7%

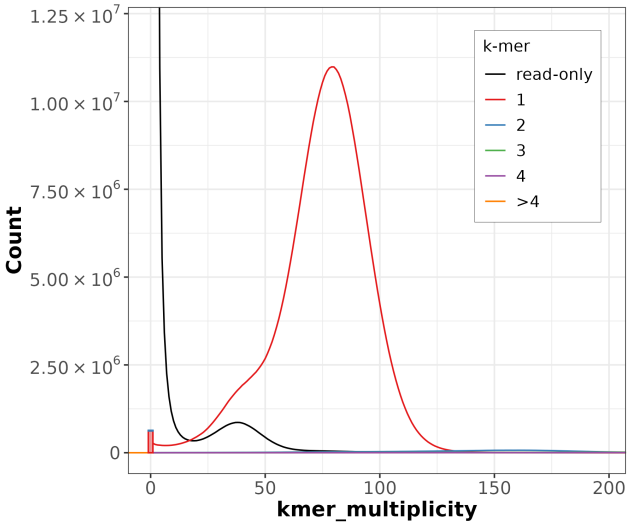
Warning! BUSCO versions or lineage datasets are not the same across results:
BUSCO: 6.0.0 (euk_genome_min, miniprot) / Lineage: actinopterygii_odb12 (genomes:75, BUSCOs:7207)
BUSCO: 6.1.0 (euk_genome_min, miniprot) / Lineage: actinopterygii_odb12 (genomes:75, BUSCOs:7207)

HiC contact map of curated assembly

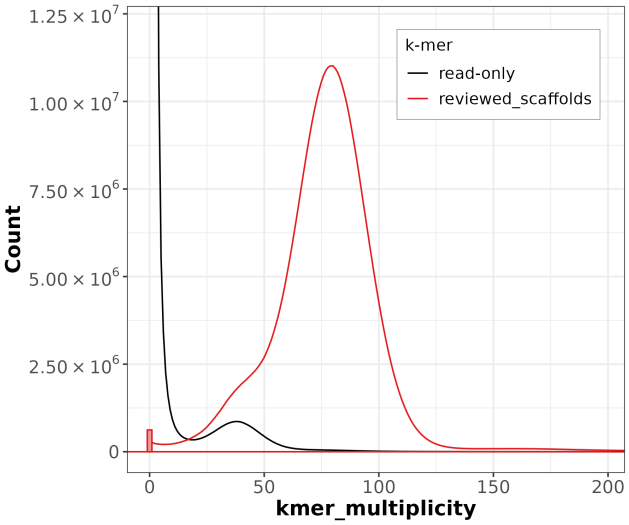


collapsed [LINK]

K-mer spectra of curated assembly

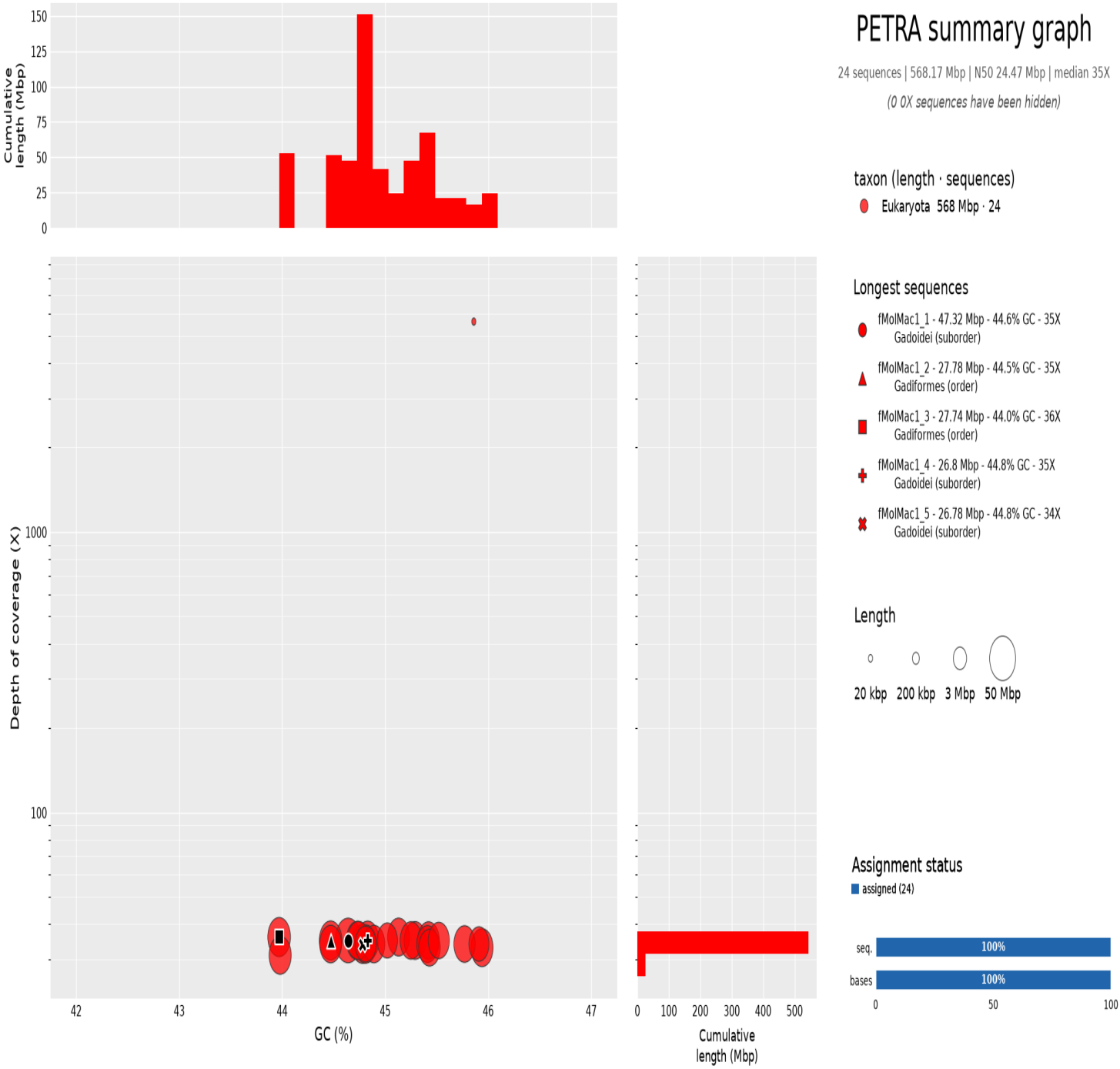


Distribution of k-mer counts per copy numbers found in asm



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	37	153

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: Caroline Menguy
Affiliation: Genoscope

Date and time: 2026-10-13 21:04:30 CEST