ERGA Assembly Report

v24.10.15

Tags: ATLASea[INVALID TAG]

TxID	46662
ToLID	**fMurHel1**
Species	Muraena helena
Class	Actinopteri
Order	Anguilliformes

Genome Traits	Expected	Observed
Haploid size (bp)	2,025,090,753	2,057,171,698
Haploid Number	21 (source: direct)	21
Ploidy	2 (source: ancestor)	2
Sample Sex	Unknown	Unknown

EBP metrics summary and curation notes

Obtained EBP quality metric for collapsed: 7.8.Q48

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

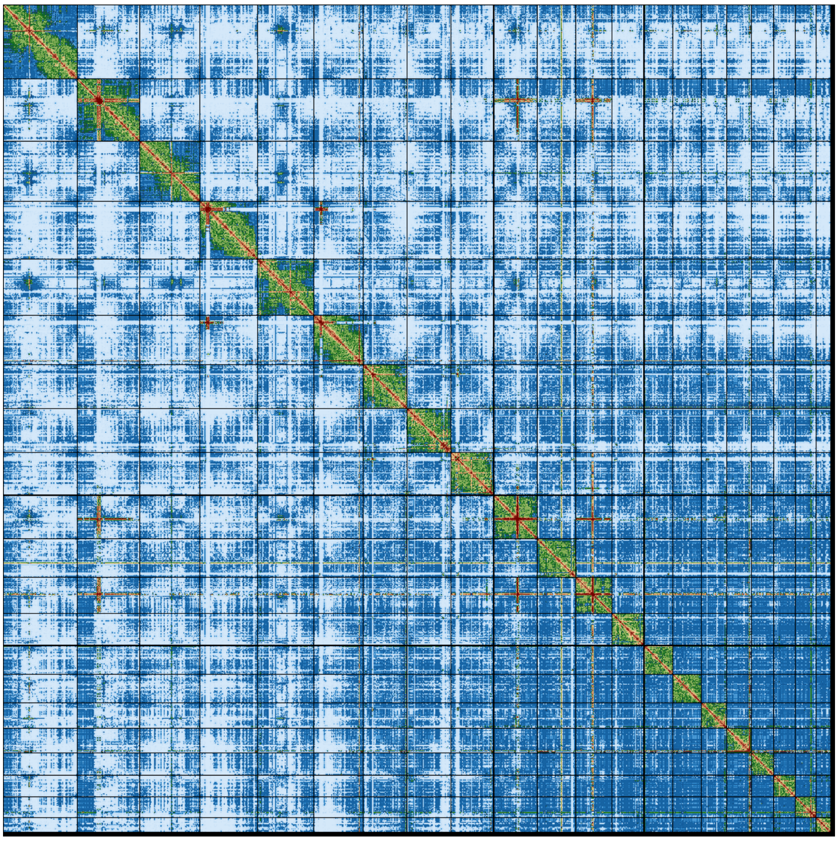
. Interventions/Gb: 16
. Contamination notes: ""
. Other observations: "The assembly of Muraena helena (fMurHel1) is based on 42X PacBio data and 203X Arima Hi-C data generated as part of the ATLASea programme (https://www.atlasea.fr).The assembly process included the following steps: initial PacBio 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, no contigs were identified as contaminants (bacterial,viral,archae). Additionally, 191 regions totaling 85 Mb (with the largest being 8.2 Mb) were identified as haplotypic duplications and removed. The mitochondrial genome was assembled using oatk and the sequence was linear. 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	2,057,183,647	2,057,171,698
GC %	41.64	41.64
Gaps/Gbp	21.87	28.19
Total gap bp	4,500	7,900
Scaffolds	85	75
Scaffold N50	106,420,850	107,794,592
Scaffold L50	8	8
Scaffold L90	18	18
Contigs	130	133
Contig N50	55,409,183	55,409,183
Contig L50	13	13
Contig L90	37	37
QV	48.9597	48.9596
Kmer compl.	85.9776	85.9776
BUSCO sing.	91.4%	91.4%
BUSCO dupl.	6.8%	6.8%
BUSCO frag.	0.3%	0.3%
BUSCO miss.	1.5%	1.5%

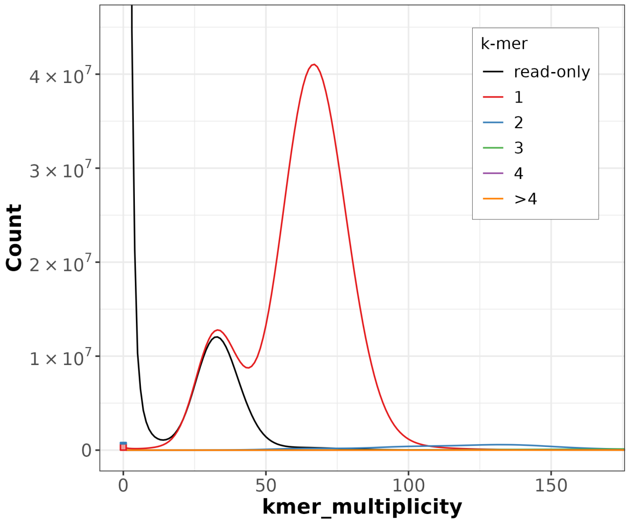
BUSCO: 6.0.0 (euk_genome_min, miniprot) / Lineage: actinopterygii_odb12 (genomes:75, BUSCOs:7207)

HiC contact map of curated assembly

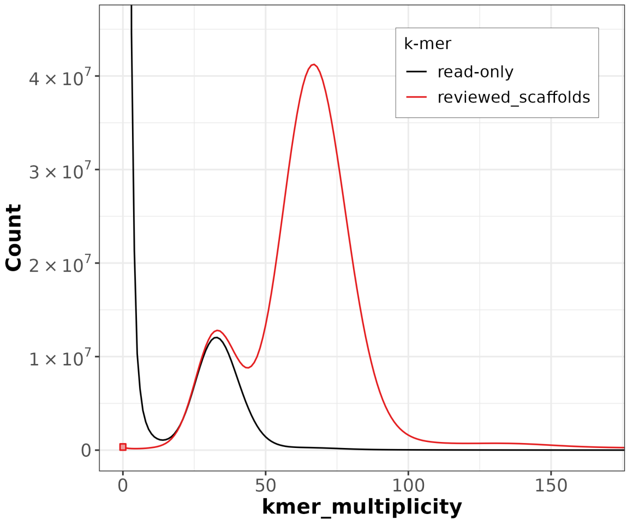


collapsed [LINK]

K-mer spectra of curated assembly

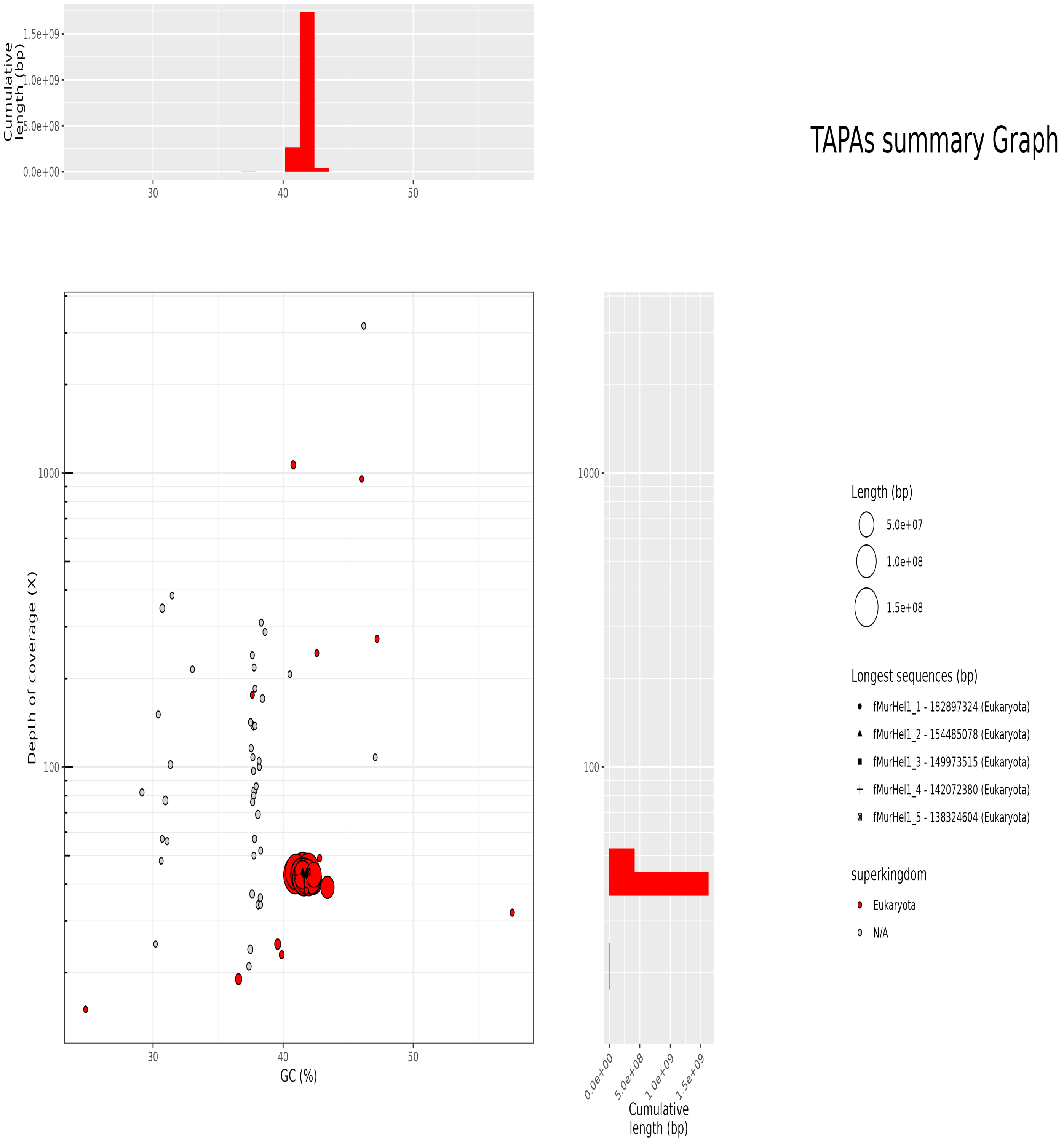


Distribution of k-mer counts per copy numbers found in asm



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	42	203

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: 2026-01-09 19:46:11 CET