ERGA Assembly Report

v24.10.15

Tags: ATLASea[INVALID TAG]

TxID	202598
ToLID	**fSymMed1**
Species	Symphodus mediterraneus
Class	Actinopteri
Order	Labriformes

Genome Traits	Expected	Observed
Haploid size (bp)	626,723,098	653,105,232
Haploid Number	23 (source: direct)	24
Ploidy	2 (source: ancestor)	2
Sample Sex	Unknown	Unknown

EBP metrics summary and curation notes

Obtained EBP quality metric for collapsed: 6.7.Q43

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
. More than 1000 gaps/Gbp for collapsed

Curator notes

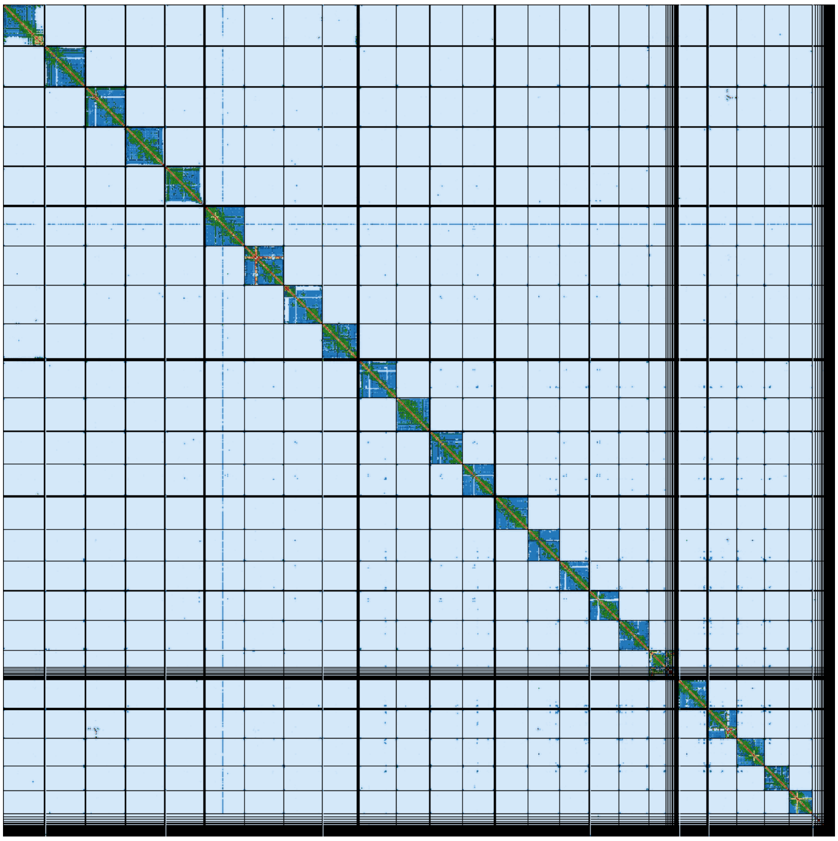
. Interventions/Gb: 199
. Contamination notes: ""
. Other observations: "The assembly of \'Symphodus mediterraneus\' (fSymMed1) is based on 35X PacBio data and 155X Arima Hi-C data generated as part of the ATLASea programme (https://www.atlasea.fr). The assembly process included the following steps: 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, 646 regions totaling 26.266 Mb (with the largest being 0.428 Mb) were identified as haplotypic duplications and removed. Mitochondrial genome was assembled using OATK. During manual curation, 9 haplotypic regions were removed, totaling 1.163Mb (with the largest being 0.252Mb). 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	654,272,804	653,105,232
GC %	41.86	41.86
Gaps/Gbp	1,487.15	1,609.24
Total gap bp	97,300	115,000
Scaffolds	326	224
Scaffold N50	25,244,266	25,972,075
Scaffold L50	12	11
Scaffold L90	25	23
Contigs	1,299	1,275
Contig N50	4,153,000	4,121,121
Contig L50	44	45
Contig L90	329	329
QV	43.3638	43.3893
Kmer compl.	79.93	79.8566
BUSCO sing.	97.0%	97.1%
BUSCO dupl.	0.8%	0.7%
BUSCO frag.	0.4%	0.4%
BUSCO miss.	1.8%	1.8%

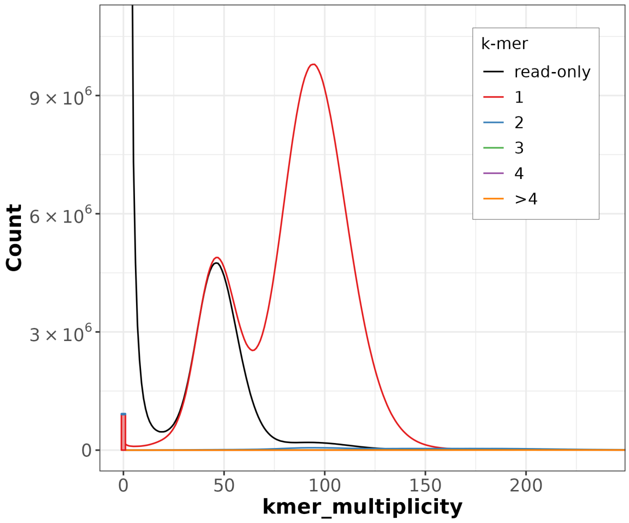
BUSCO: 6.0.0 (euk_genome_min, miniprot) / Lineage: actinopterygii_odb12 (genomes:75, BUSCOs:7207)

HiC contact map of curated assembly

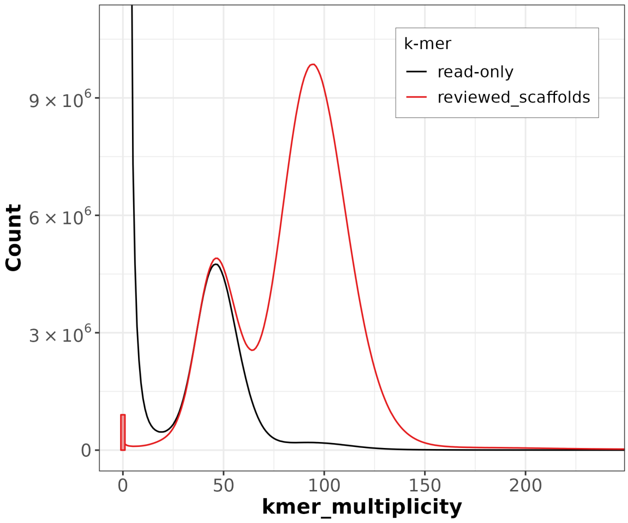


collapsed [LINK]

K-mer spectra of curated assembly

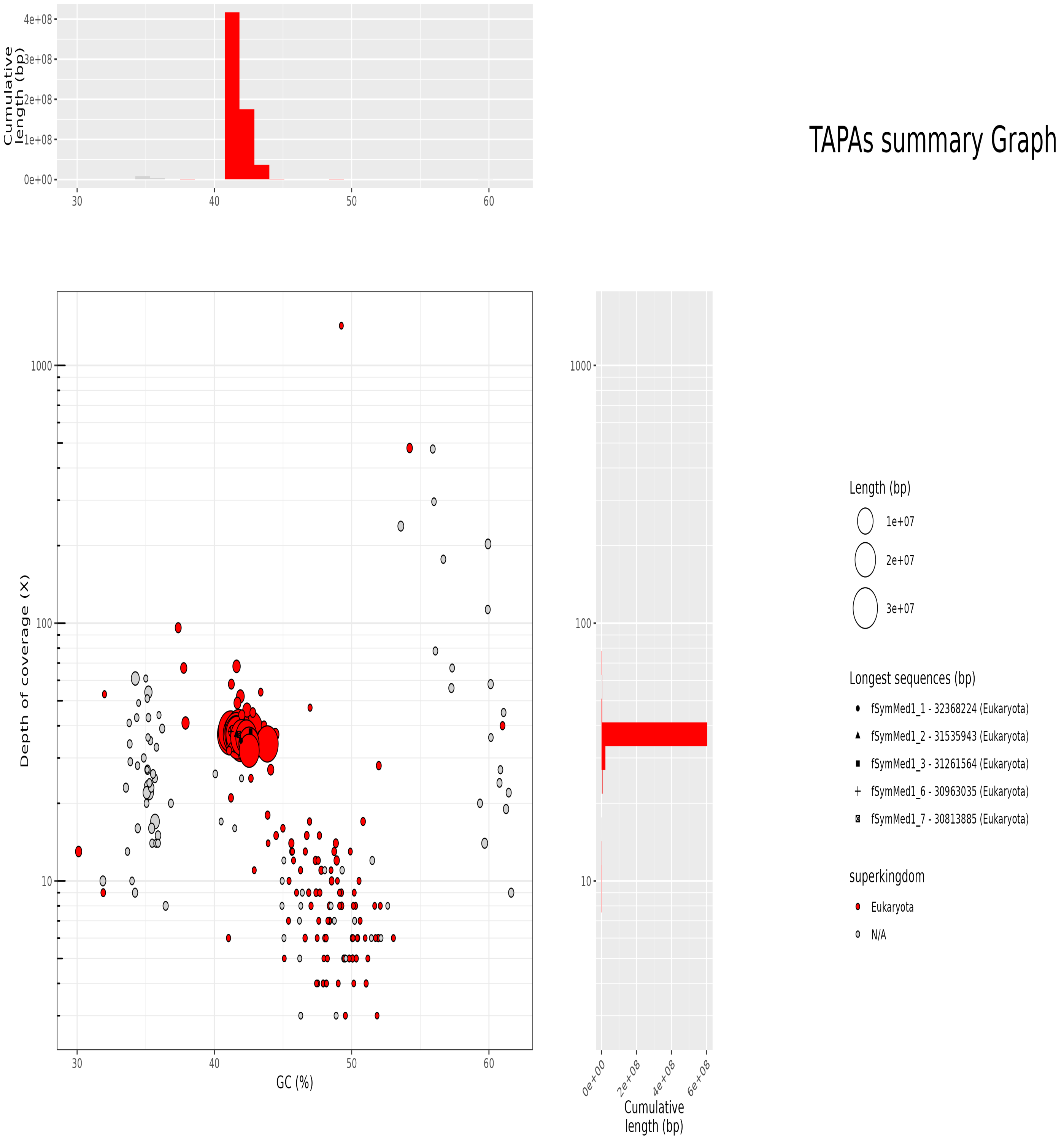


Distribution of k-mer counts per copy numbers found in asm



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	35	155

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: Simone Duprat
Affiliation: Genoscope

Date and time: 2026-01-21 05:59:22 CET