ERGA Assembly Report

v24.10.15

Tags: ATLASea[INVALID TAG]

TxID	191724
ToLID	**fSymCin1**
Species	Symphodus cinereus
Class	Actinopteri
Order	Labriformes

Genome Traits	Expected	Observed
Haploid size (bp)	609,985,920	622,367,766
Haploid Number	24 (source: ancestor)	24
Ploidy	2 (source: ancestor)	2
Sample Sex	Unknown	Unknown

EBP metrics summary and curation notes

Obtained EBP quality metric for collapsed: 7.7.Q45

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

. Kmer completeness value is less than 90 for collapsed

Curator notes

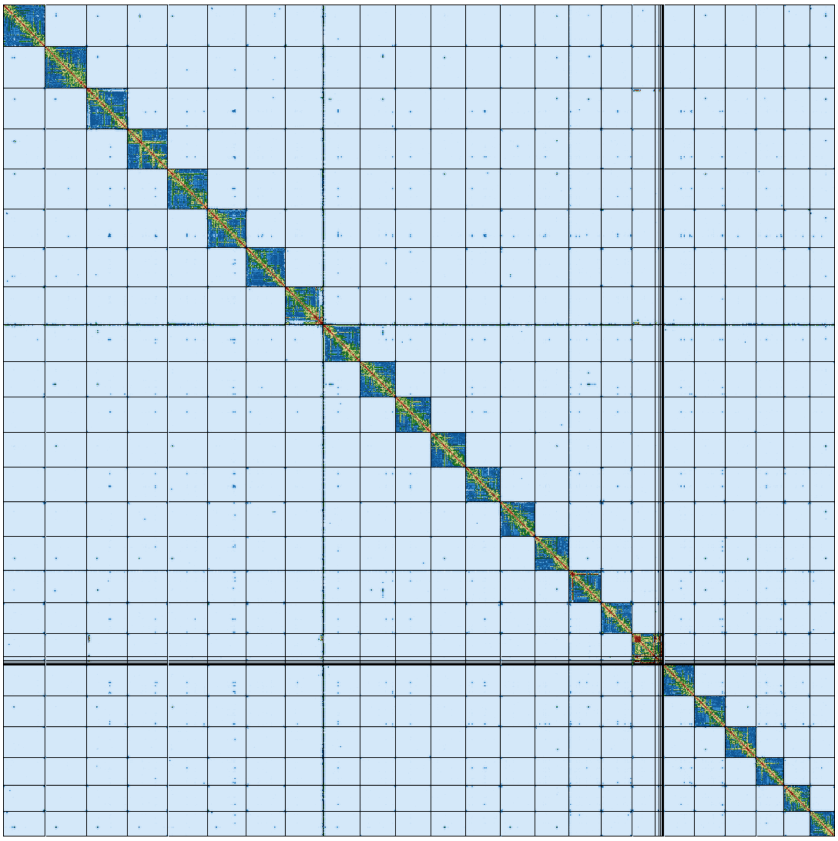
. Interventions/Gb: 123
. Contamination notes: ""
. Other observations: "The assembly of Symphodus cinereus (fSymCin1) is based on 56X PacBio data and 178X Arima Hi-C data generated as part of the ATLASea programme (https://www.atlasea.fr). The assembly process included the following steps: initial PacBio 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, 2 contigs were identified as contaminants (bacterial, archaeal, or viral), totaling 0.129 Mb (with the largest being 0.096 Mb). Additionally, 266 regions totaling 14.761 Mb (with the largest being 0.516 Mb) were identified as haplotypic duplications and removed. Finally, the primary assembly was analyzed and manually improved using Pretext. During manual curation, 2 haplotypic regions and 1 contaminant sequences were removed, totaling 0.962Mb and 0.040Mb, respectively (with the largest being 0.481Mb and 0.040Mb). 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	623,354,072	622,367,766
GC %	41.77	41.77
Gaps/Gbp	564.69	621.82
Total gap bp	35,200	44,000
Scaffolds	83	48
Scaffold N50	26,250,632	26,355,312
Scaffold L50	11	11
Scaffold L90	22	21
Contigs	435	435
Contig N50	13,517,219	13,517,219
Contig L50	16	16
Contig L90	72	71
QV	45.4099	45.4133
Kmer compl.	83.8954	83.8506
BUSCO sing.	99.1%	99.1%
BUSCO dupl.	0.2%	0.2%
BUSCO frag.	0.1%	0.1%
BUSCO miss.	0.6%	0.6%

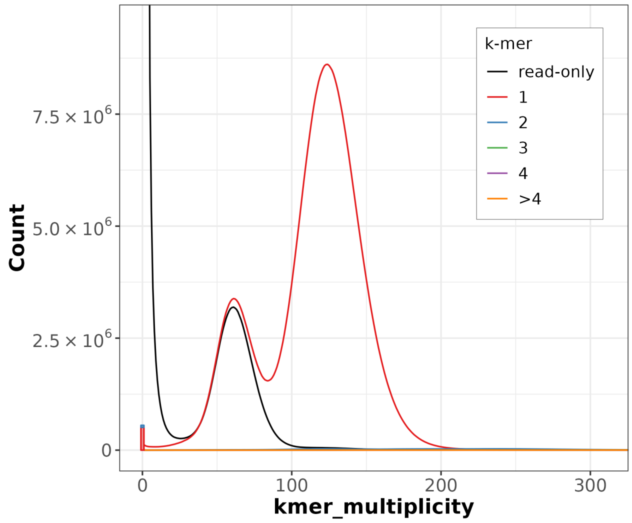
BUSCO: 6.0.0 (euk_genome_min, miniprot) / Lineage: actinopterygii_odb12 (genomes:75, BUSCOs:7207)

HiC contact map of curated assembly

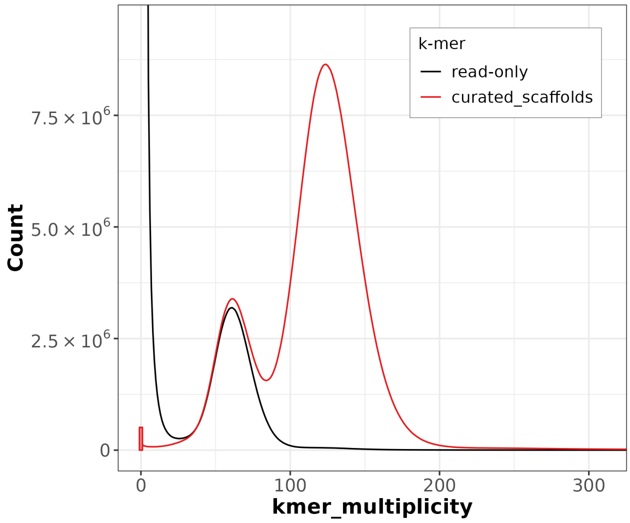


collapsed [LINK]

K-mer spectra of curated assembly

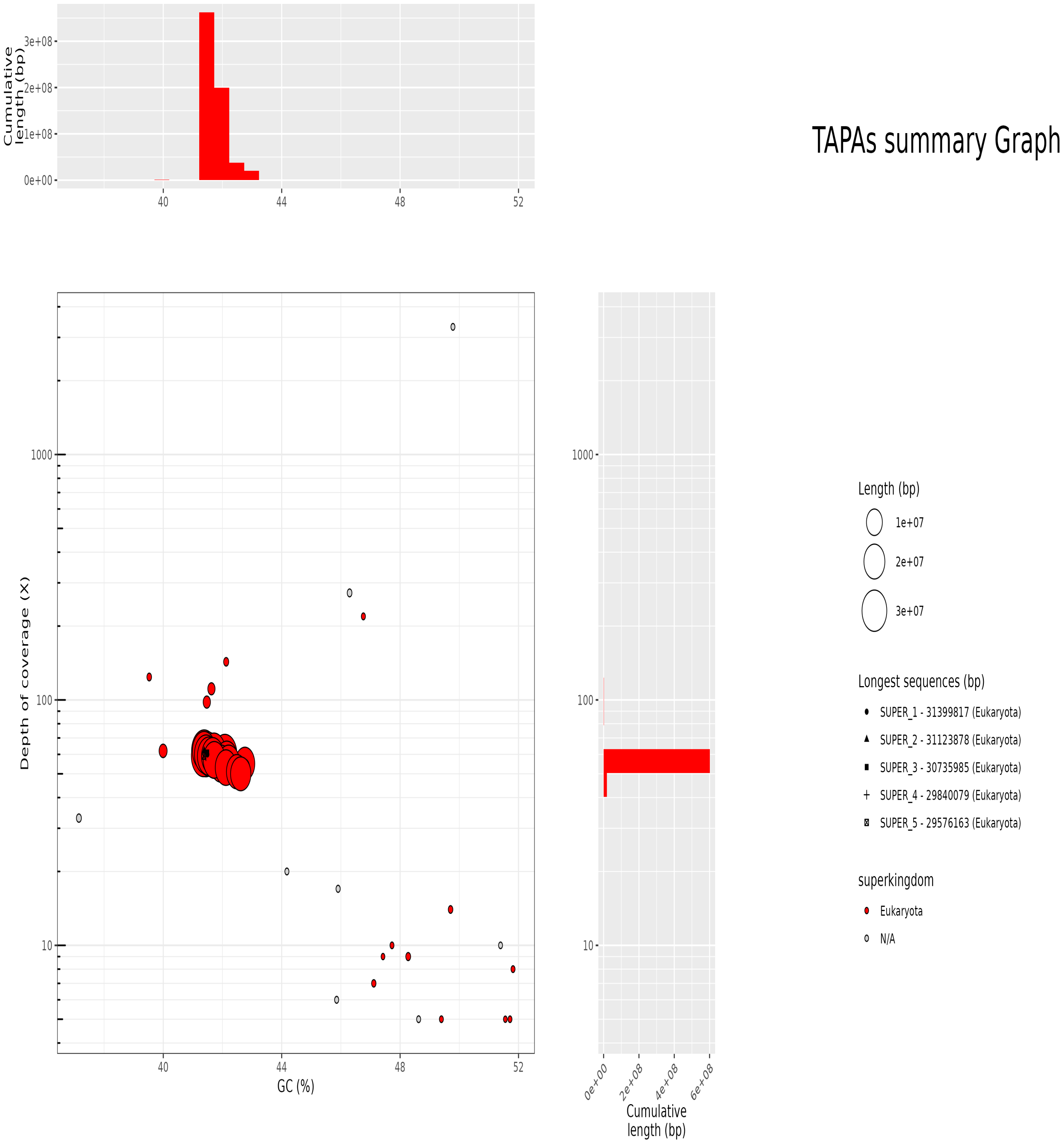


Distribution of k-mer counts per copy numbers found in asm



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	56	178

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: Sophie Layac
Affiliation: Genoscope

Date and time: 2025-12-09 04:58:32 CET