ERGA Assembly Report

v24.10.15

Tags: ATLASea[INVALID TAG]

TxID	202597
ToLID	**fSymDod1**
Species	Symphodus doderleini
Class	Actinopteri
Order	Labriformes

Genome Traits	Expected	Observed
Haploid size (bp)	622,961,832	658,301,760
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: 6.7.Q44

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

. Kmer completeness value is less than 90 for collapsed
. More than 1000 gaps/Gbp for collapsed

Curator notes

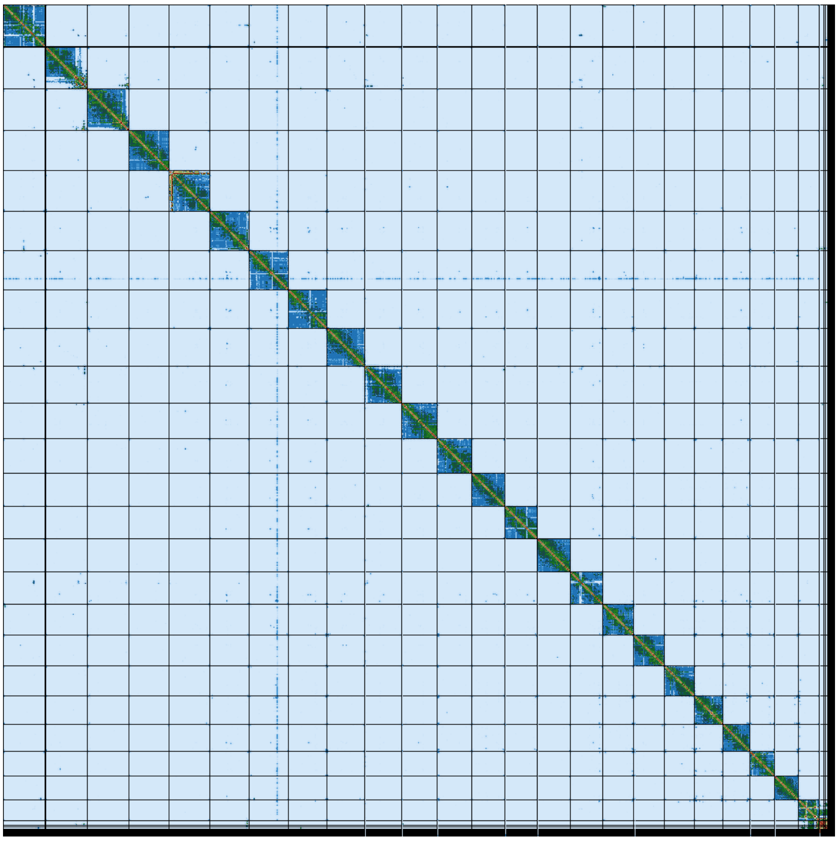
. Interventions/Gb: 141
. Contamination notes: ""
. Other observations: "The assembly of \'Symphodus doderleini\' (fSymDod1) is based on 40X PacBio data and 181X Arima Hi-C data generated as part of the ATLASea programme (https://www.atlasea.fr). The assembly process included the following steps: initial 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, 1 contigs was identified as contaminants (bacterial, archaeal, or viral), totaling 0.013 Mb (with the largest being 0.013 Mb). Additionally, 327 regions totaling 13.537 Mb (with the largest being 0.331 Mb) were identified as haplotypic duplications and removed. Mitochondrial genome was assembled using OATK. During manual curation, 13 haplotypic regions were removed, totaling 1.074Mb (with the largest being 0.294Mb). 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	659,388,822	658,301,760
GC %	41.95	41.94
Gaps/Gbp	1,172.3	1,250.19
Total gap bp	77,300	89,200
Scaffolds	236	147
Scaffold N50	27,181,611	28,071,209
Scaffold L50	11	11
Scaffold L90	22	22
Contigs	1,009	970
Contig N50	5,778,720	5,778,720
Contig L50	36	36
Contig L90	229	227
QV	44.8991	44.9369
Kmer compl.	87.9377	87.8586
BUSCO sing.	98.6%	98.6%
BUSCO dupl.	0.3%	0.3%
BUSCO frag.	0.3%	0.2%
BUSCO miss.	0.8%	0.9%

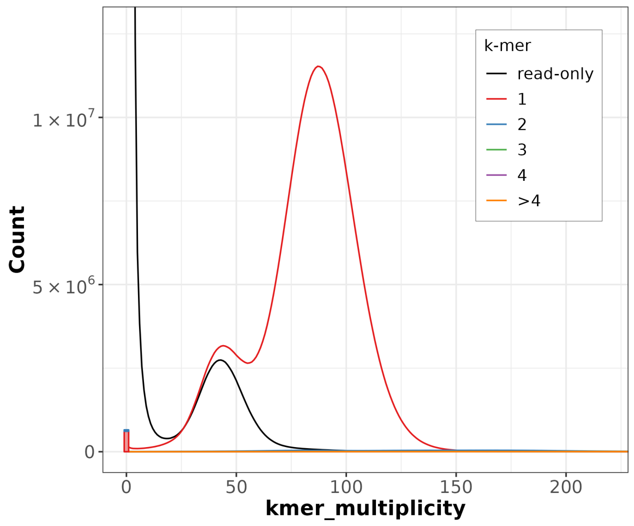
BUSCO: 6.0.0 (euk_genome_min, miniprot) / Lineage: actinopterygii_odb12 (genomes:75, BUSCOs:7207)

HiC contact map of curated assembly

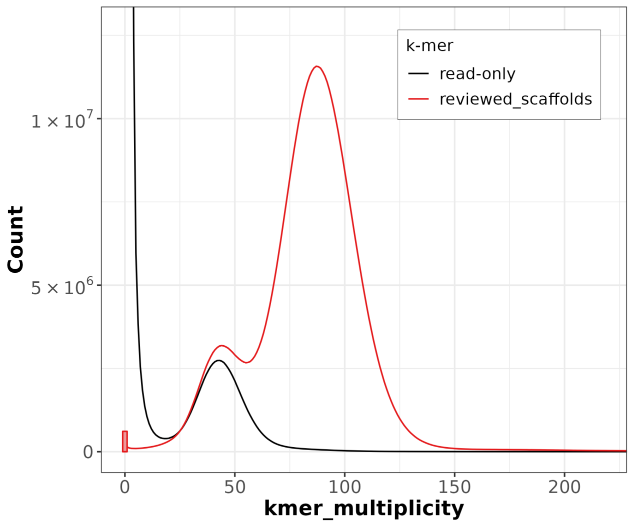


collapsed [LINK]

K-mer spectra of curated assembly

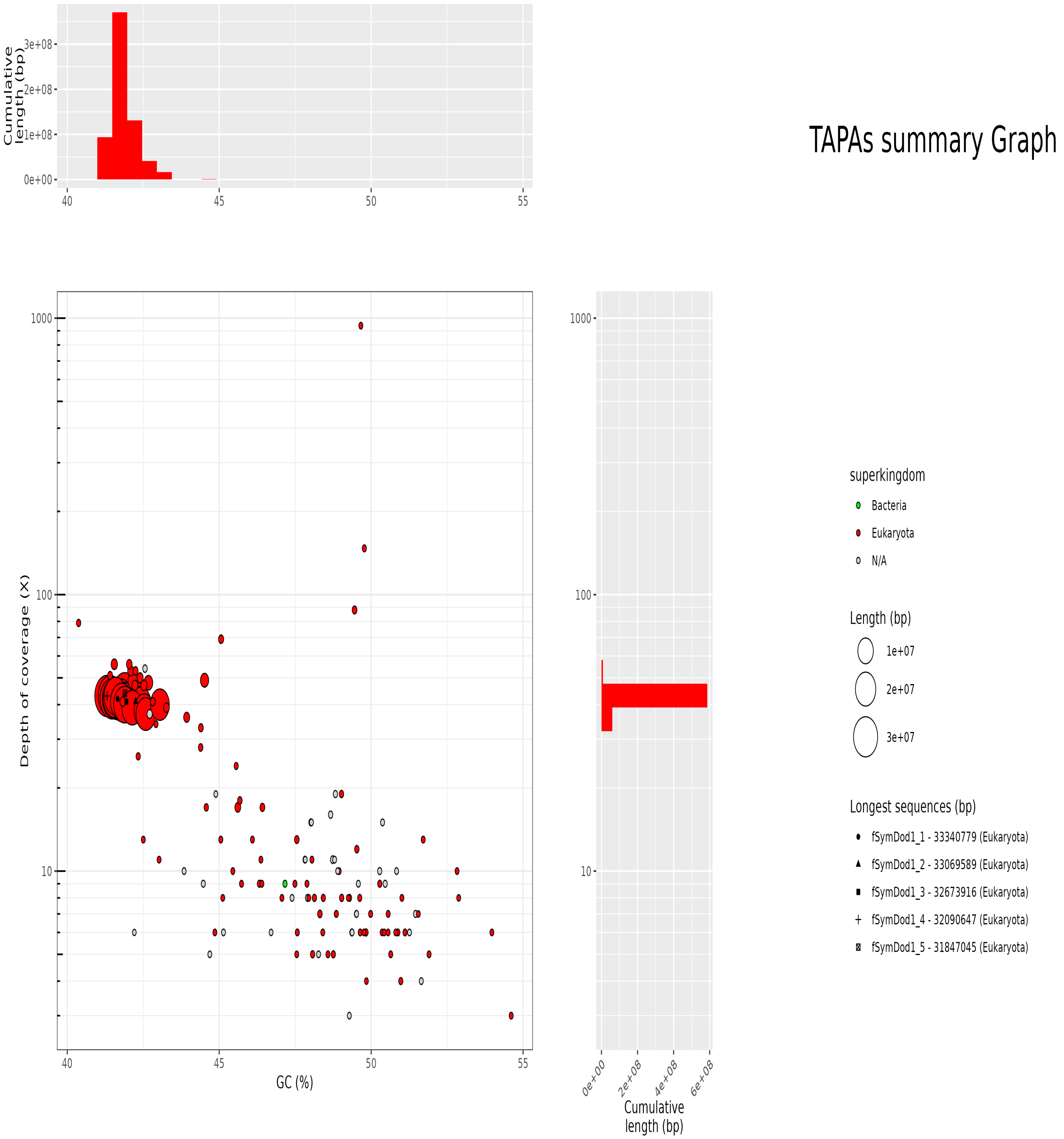


Distribution of k-mer counts per copy numbers found in asm



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	40	181

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 05:56:37 CET