

ERGA Assembly Report

v24.10.15

Tags: ATLASEa[INVALID TAG]

TxID	191725
ToLID	fSymBai1
Species	<i>Symphodus bailloni</i>
Class	Actinopteri
Order	Labriiformes

Genome Traits	Expected	Observed
Haploid size (bp)	608,994,662	627,958,947
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.Q45

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

Curator notes

. Interventions/Gb: 135
. Contamination notes: ""
. Other observations: "The assembly of *Symphodus bailloni* (fSymBai1) is based on 59X PacBio data and 174X 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 Hifiiasm, removal of contaminant sequences using Context, removal of haplotypic duplications using purge_dups, and Hi-C-based scaffolding with YaHS. In total, 6 contigs were identified as contaminants (bacterial, archaeal, or viral), totaling 0.152 Mb (with the largest being 0.045 Mb). Additionally, 264 regions totaling 20.709 Mb (with the largest being 1.819 Mb) were identified as haplotypic duplications and removed. The mitochondrial genome was assembled using OATK. During manual curation, 5 haplotypic regions were removed, totaling 0.39Mb (with the largest being 0.105Mb). Chromosome-scale scaffolds confirmed by Hi-C data were named in order of size. "

Quality metrics table

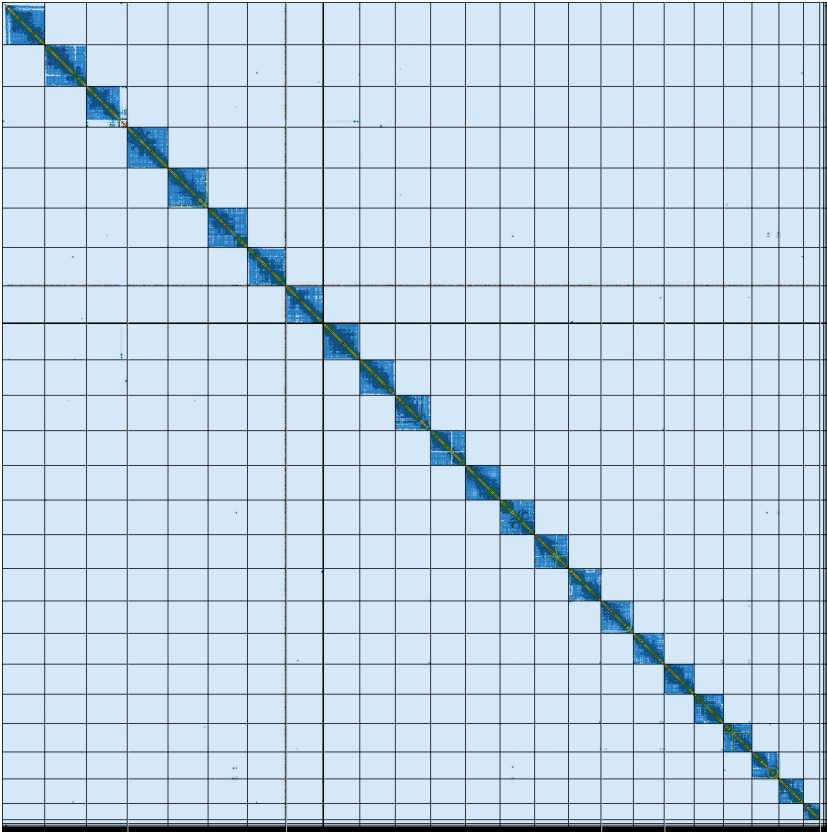
Metrics	Pre-curation collapsed	Curated collapsed
Total bp	628,396,800	627,958,947
GC %	42.03	42.03
Gaps/Gbp	817.95	874.26
Total gap bp	53,000	63,100
Scaffolds	161	99
Scaffold N50	26,469,832	26,570,420
Scaffold L50	11	11
Scaffold L90	22	21
Contigs	669	648
Contig N50	6,830,908	6,830,908
Contig L50	28	28
Contig L90	159	159
QV	45.2878	45.2999
Kmer compl.	91.3183	91.2791
BUSCO sing.	96.7%	99.0%
BUSCO dupl.	0.2%	0.2%
BUSCO frag.	0.9%	0.2%
BUSCO miss.	2.2%	0.6%

Warning! BUSCO versions or lineage datasets are not the same across results:

BUSCO: 5.8.2 (euk_genome_met, metaeuk) / Lineage: actinopterygii_odb12 (genomes:75, BUSCOs:7207)

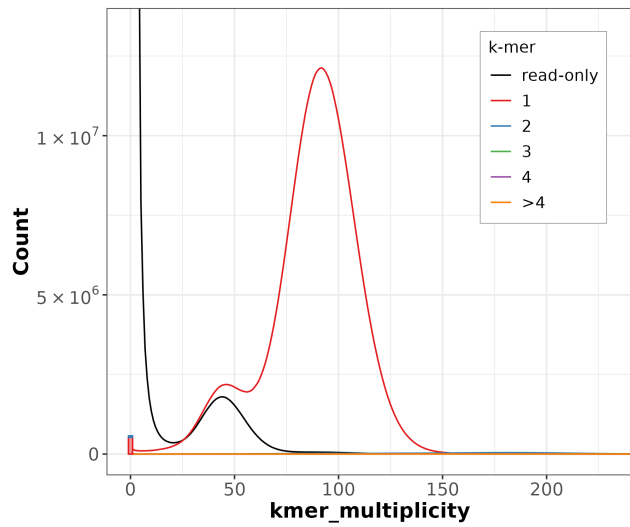
BUSCO: 6.0.0 (euk_genome_min, miniprot) / Lineage: actinopterygii_odb12 (genomes:75, BUSCOs:7207)

HiC contact map of curated assembly

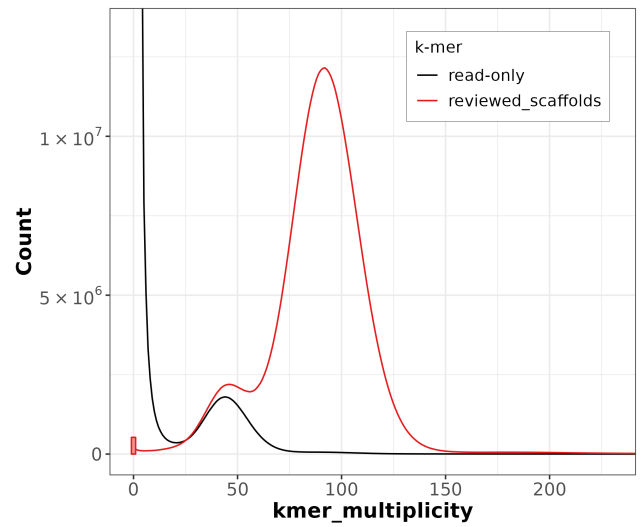


collapsed [\[LINK\]](#)

K-mer spectra of curated assembly

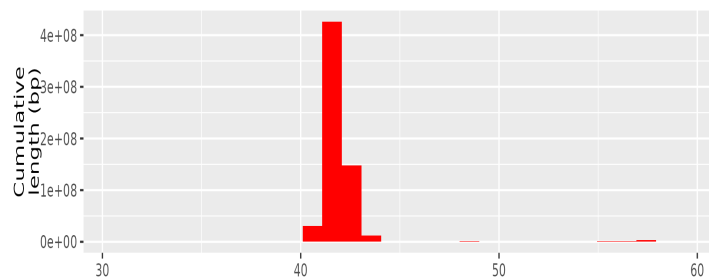


Distribution of k-mer counts per copy numbers found in asm

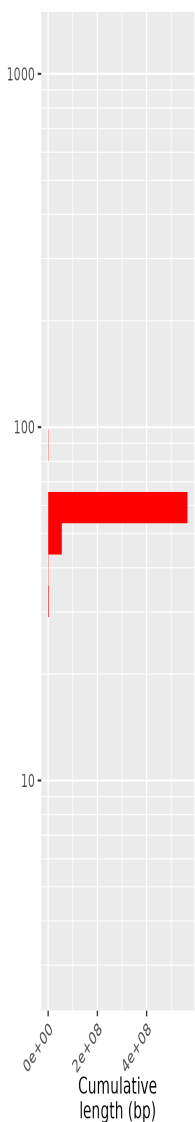
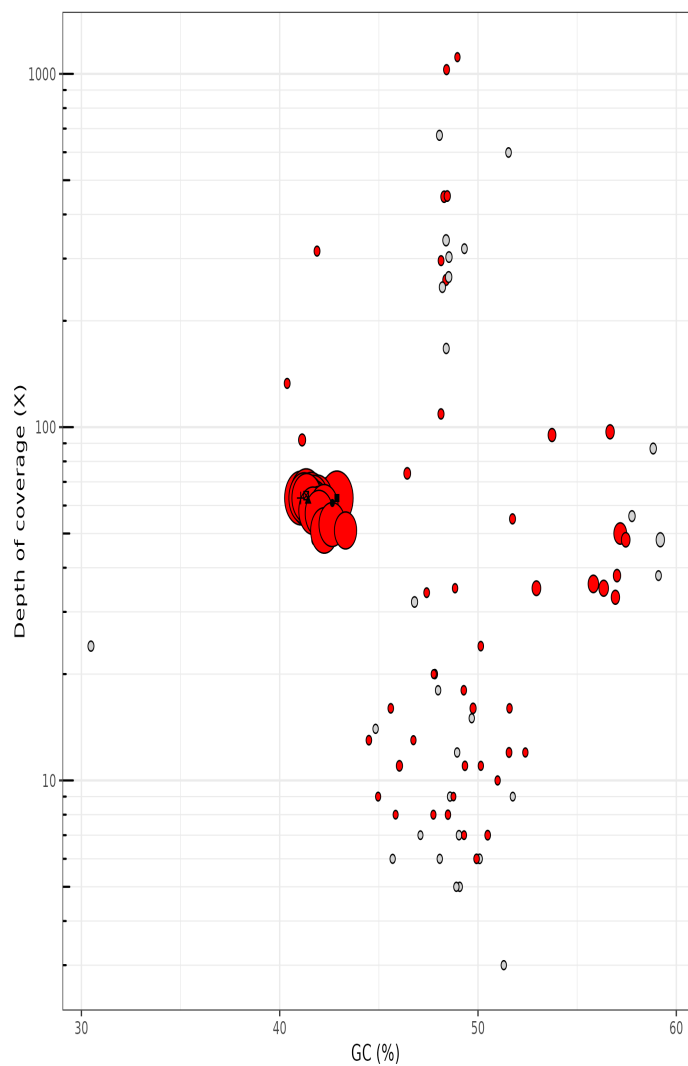


Distribution of k-mer counts coloured by their presence in reads/assemblies

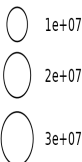
Post-curation contamination screening



TAPAs summary Graph



Length (bp)



Longest sequences (bp)

- fSymBai1_1 - 32565741 (Eukaryota)
- ▲ fSymBai1_2 - 31075949 (Eukaryota)
- fSymBai1_3 - 30938799 (Eukaryota)
- + fSymBai1_4 - 30509193 (Eukaryota)
- ⊠ fSymBai1_5 - 30334820 (Eukaryota)

superkingdom

- Eukaryota
- N/A

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	58	173

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: Jean-Marc Aury

Affiliation: Genoscope

Date and time: 2025-10-13 07:04:11 CEST