

ERGA Assembly Report

v24.10.15

Tags: ATLASEa[INVALID TAG]

TxID	508607
ToLID	fAgoCat1
Species	Agonus cataphractus
Class	Actinopteri
Order	Perciformes

Genome Traits	Expected	Observed
Haploid size (bp)	606,893,258	673,574,728
Haploid Number	20 (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:

- . Observed Haploid Number is different from Expected
- . More than 1000 gaps/Gbp for collapsed

Curator notes

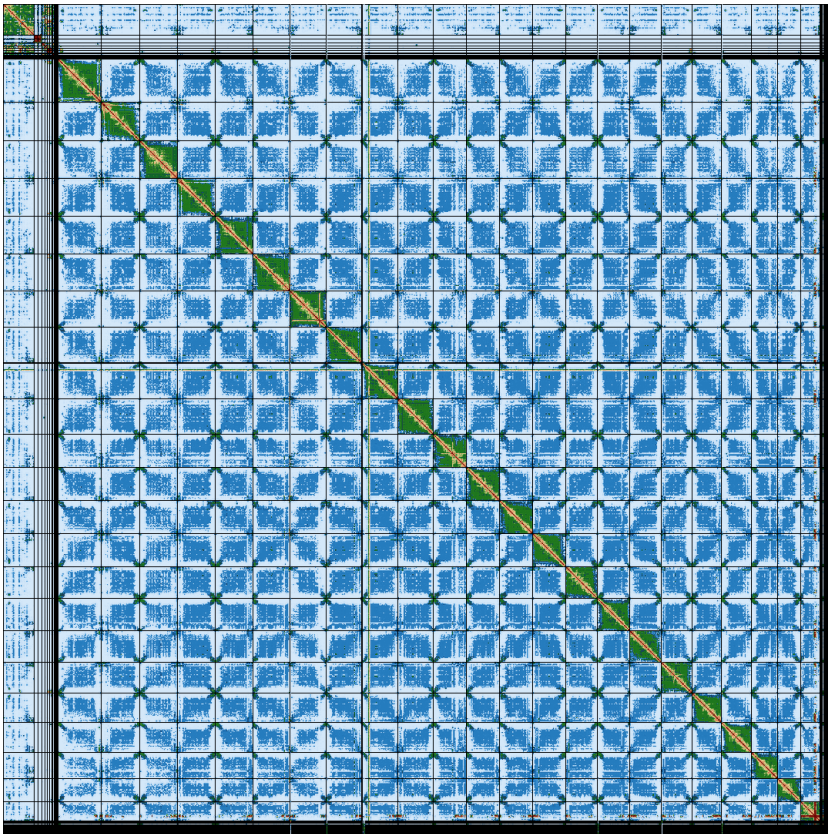
. Interventions/Gb: 322
. Contamination notes: ""
. Other observations: "The assembly of Agonus cataphractus (fAgoCat1) is based on 34.7X PacBio data and 346.7X OmniC Hi-C data generated as part of the ATLASEa programme (<https://www.atlasea.fr>). The assembly process included the following steps: initial 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 0.025 Mb (with the largest being 0.014 Mb). Additionally, 820 regions totaling 42.064 Mb (with the largest being 0.479 Mb) were identified as haplotypic duplications and removed. The mitochondrial genome was assembled using OATK. Finally, the primary assembly was analyzed and manually improved using Pretext. During manual curation, 28 haplotypic regions were removed, totaling 7.8Mb, (with the largest being 2.2Mb). 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	681,681,627	673,574,728
GC %	42.63	42.62
Gaps/Gbp	1,154.5	1,300.52
Total gap bp	78,700	102,800
Scaffolds	503	313
Scaffold N50	26,582,918	27,002,447
Scaffold L50	12	12
Scaffold L90	24	23
Contigs	1,290	1,189
Contig N50	5,243,427	5,286,793
Contig L50	32	31
Contig L90	289	268
QV	44.1727	44.2196
Kmer compl.	91.6635	91.3394
BUSCO sing.	93.4%	94.8%
BUSCO dupl.	2.9%	1.5%
BUSCO frag.	1.2%	1.1%
BUSCO miss.	2.6%	2.6%

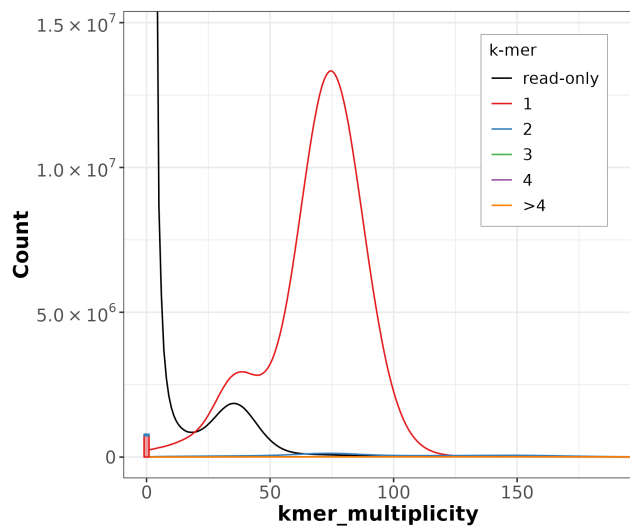
BUSCO: 5.8.2 (euk_genome_met, metaeuk) / Lineage: actinopterygii_odb12 (genomes:75, BUSCOs:7207)

HiC contact map of curated assembly

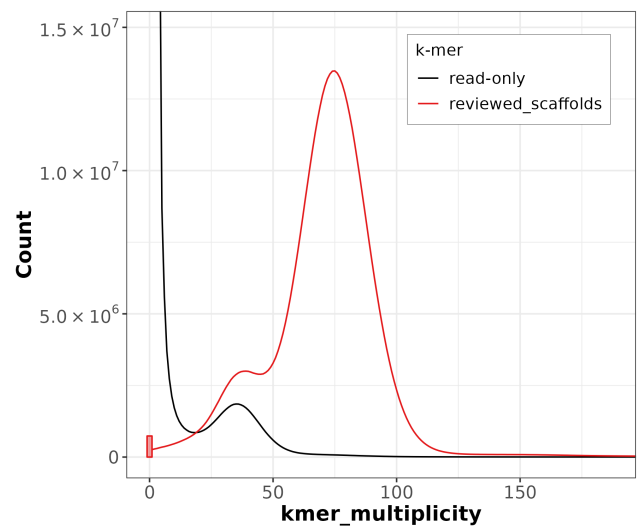


collapsed [\[LINK\]](#)

K-mer spectra of curated assembly

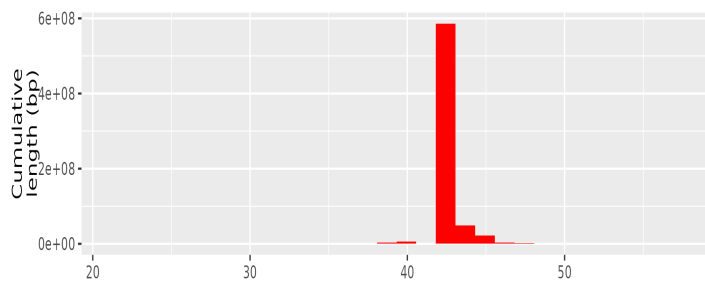


Distribution of k-mer counts per copy numbers found in asm

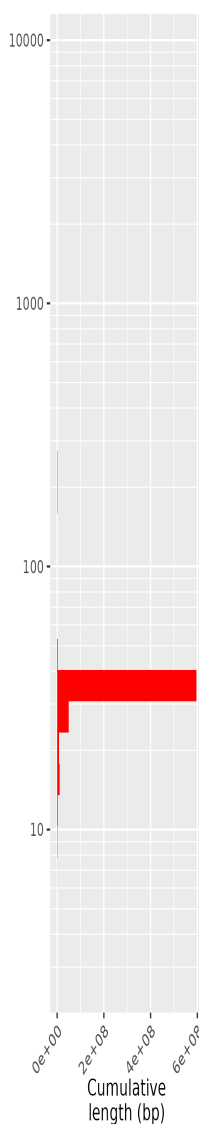
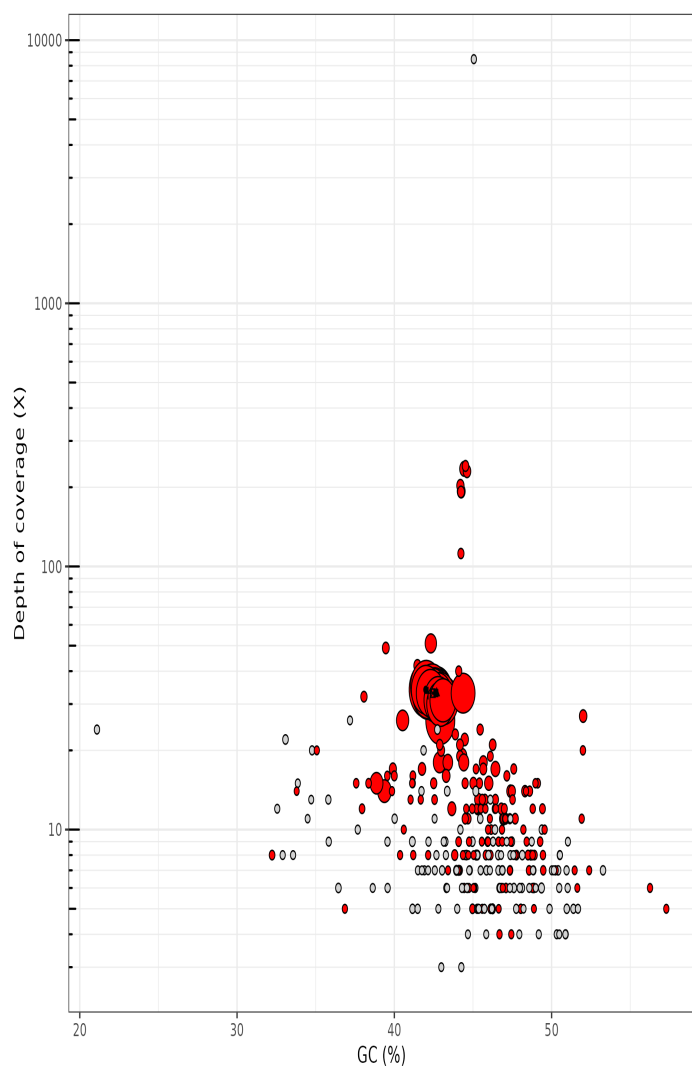


Distribution of k-mer counts coloured by their presence in reads/assemblies

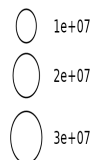
Post-curation contamination screening



TAPAs summary Graph



Length (bp)



Longest sequences (bp)

- fAgoCat1_2 - 34543896 (Eukaryota)
- ▲ fAgoCat1_3 - 31500181 (Eukaryota)
- fAgoCat1_4 - 30643021 (Eukaryota)
- + fAgoCat1_5 - 30533713 (Eukaryota)
- ▣ fAgoCat1_6 - 30238169 (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	PACBIO Hifi	Omic, Arima
Coverage	34	689

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: Emilie Teodori

Affiliation: Genoscope

Date and time: 2025-07-26 07:41:49 CEST