

# ERGA Assembly Report

v24.10.15

Tags: ATLASEa[INVALID TAG]

TxID	85418
ToLID	<b>fGobPag2</b>
Species	Gobius paganellus
Class	Actinopteri
Order	Gobiiformes

Genome Traits	Expected	Observed
Haploid size (bp)	785,445,069	831,875,733
Haploid Number	21 (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.Q44

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

### Curator notes

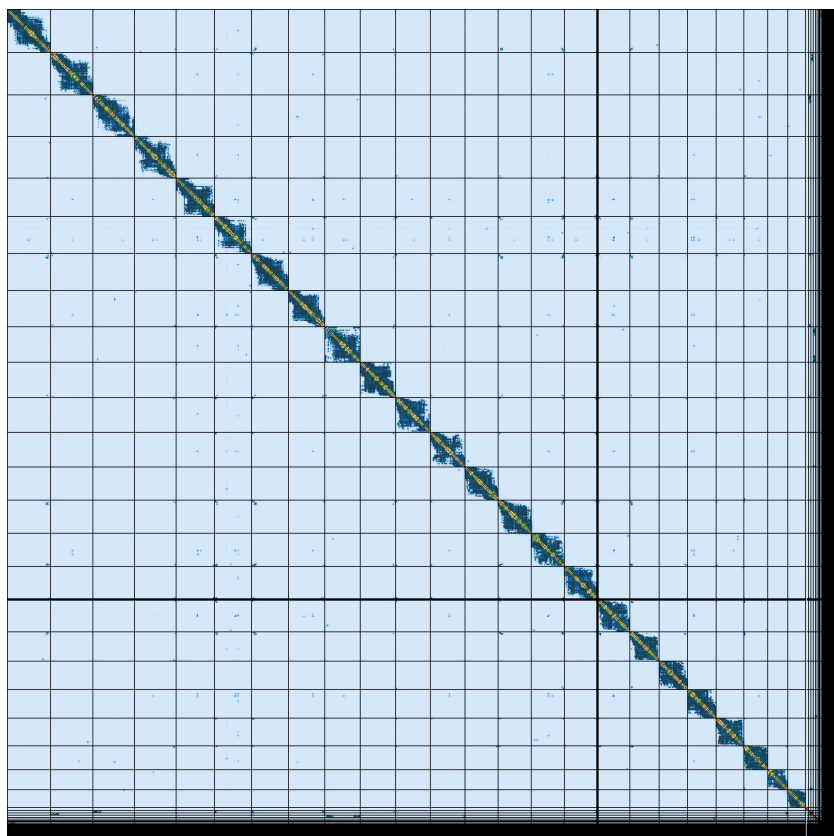
. Interventions/Gb: 11  
. Contamination notes: ""  
. Other observations: "The assembly of *Gobius paganellus* (fGobPag2.1) is based on 64X of PacBio data and 209X of Arima Hi-C data generated as part of the ATLASEa programme (<https://www.atlasea.fr>). The assembly process included the following steps: 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, 2 contigs were identified as contaminants (bacterial), totaling 0.022 Mb (with the largest being 0.018 Mb). Additionally, 167 regions totaling 23.51 Mb (with the largest being 1.06 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, 3 haplotypic regions were removed, totaling 1.64 Mb (with the largest being 1.06 Mb). 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	833,531,336	831,875,733
GC %	42.1	42.1
Gaps/Gbp	85.18	91.36
Total gap bp	7,100	8,200
Scaffolds	137	123
Scaffold N50	34,904,278	34,904,278
Scaffold L50	11	11
Scaffold L90	22	22
Contigs	208	199
Contig N50	29,485,388	29,485,388
Contig L50	13	13
Contig L90	33	33
QV	44.6286	44.7148
Kmer compl.	79.6511	79.6233
BUSCO sing.	93.8%	94.0%
BUSCO dupl.	0.6%	0.4%
BUSCO frag.	1.8%	1.8%
BUSCO miss.	3.7%	3.7%

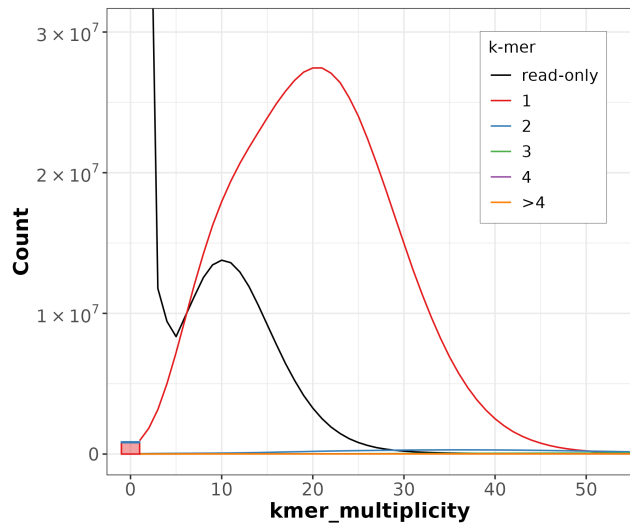
BUSCO: 5.8.2 (euk\_genome\_met, metaeuk) / Lineage: actinopterygii\_odb12 (genomes:75, BUSCOs:7207)

# HiC contact map of curated assembly

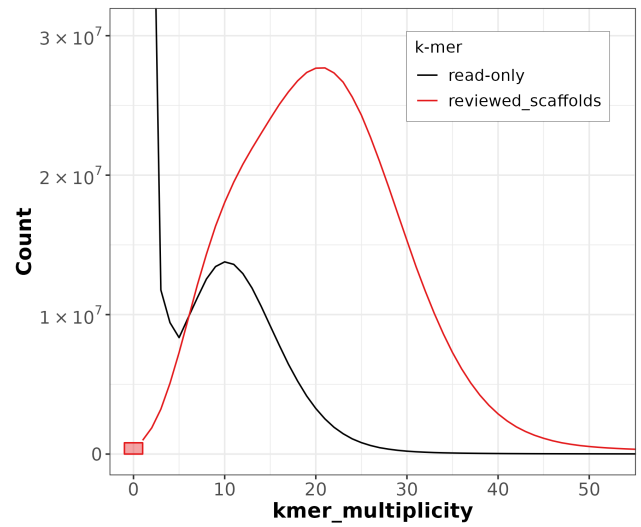


collapsed [\[LINK\]](#)

# K-mer spectra of curated assembly

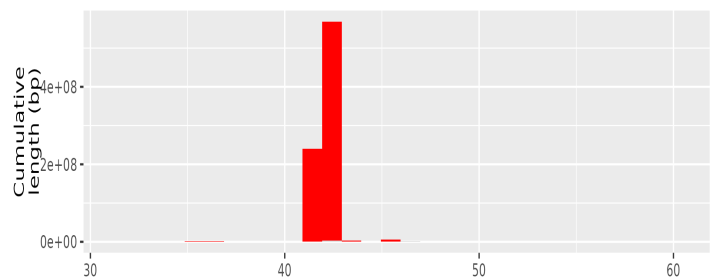


Distribution of k-mer counts per copy numbers found in asm

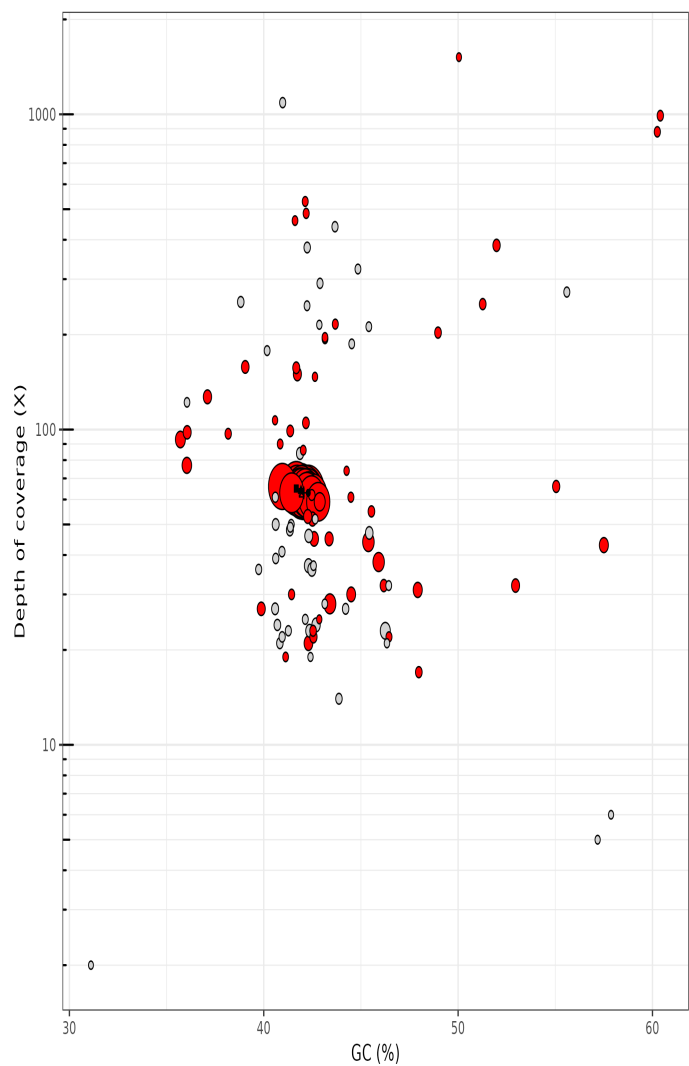


Distribution of k-mer counts coloured by their presence in reads/assemblies

# Post-curation contamination screening



TAPAs summary Graph



**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	Arima
Coverage	64	209

## 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: Adama Ndar

Affiliation: Genoscope

Date and time: 2025-05-05 11:26:19 CEST