ERGA Assembly Report

v24.10.15

Tags: ATLASea[INVALID TAG]

TxID	77594
ToLID	**fGobCob1**
Species	Gobius cobitis
Class	Actinopteri
Order	Gobiiformes

Genome Traits	Expected	Observed
Haploid size (bp)	890,075,827	859,832,672
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.Q46

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

. Observed Haploid Number is different from Expected

Curator notes

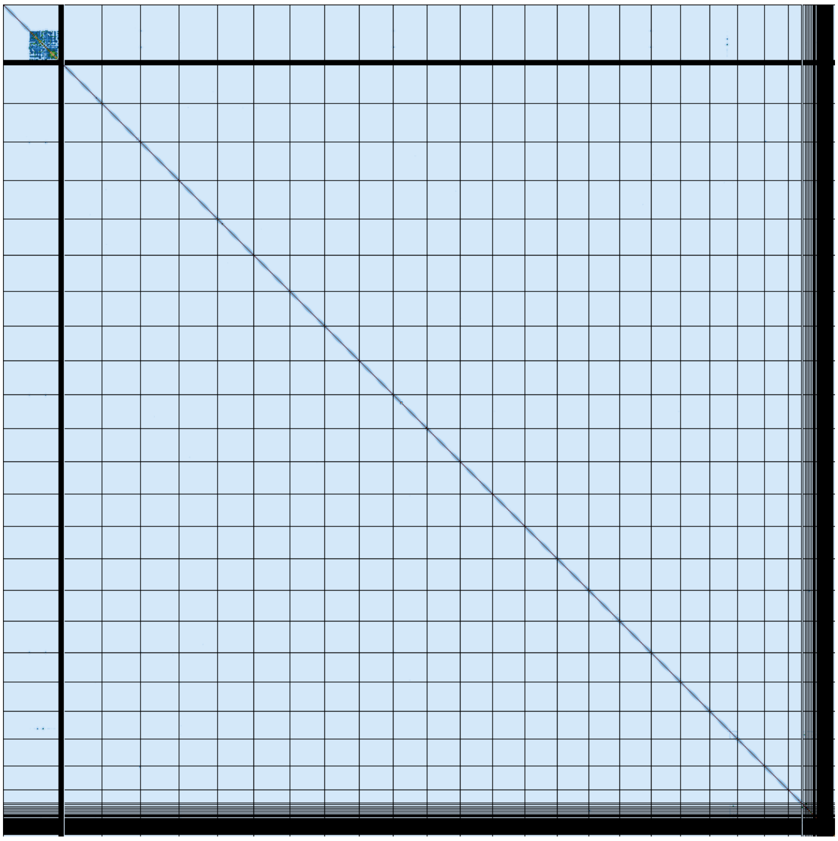
. Interventions/Gb: 58
. Contamination notes: ""
. Other observations: "The assembly of Gobius cobitis (fGobCob1) is based on 52X PacBio data and 176X Arima Hi-C data generated as part of the ATLASea programme (https://www.atlasea.fr). The assembly process included the following steps: initial PacBio assembly generation with Hifiasm, removal of contaminant sequences using Context, removal of haplotypic duplications using purge_dups, and Hi-C-based scaffolding with YaHS. No contigs were found to be contaminants (bacterial, archaeal, or viral). Additionally, 202 regions totaling 57.301 Mb (with the largest being 3.149 Mb) were identified as haplotypic duplications and removed. The mitochondrial genome was assembled using OATK. During manual curation, 5 haplotypic regions and 13 contaminant sequences were removed, totaling 3.8Mb and 0.683 Mb respectively (with the largest being 1.187Mb and 0.229 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	864,354,099	859,832,672
GC %	42.44	42.39
Gaps/Gbp	123.79	152.36
Total gap bp	10,700	16,300
Scaffolds	278	228
Scaffold N50	33,720,239	35,054,144
Scaffold L50	10	11
Scaffold L90	21	22
Contigs	385	359
Contig N50	25,512,000	25,512,000
Contig L50	14	14
Contig L90	55	56
QV	46.6911	46.723
Kmer compl.	96.8552	96.832
BUSCO sing.	94.2%	98.0%
BUSCO dupl.	0.4%	0.4%
BUSCO frag.	1.9%	0.4%
BUSCO miss.	3.4%	1.1%

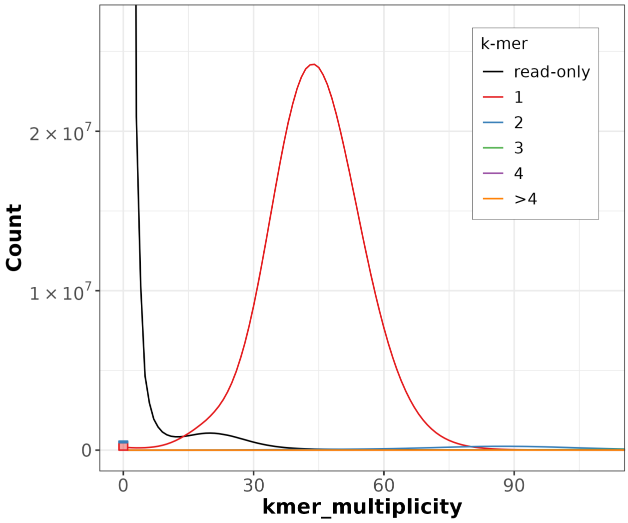
Warning! BUSCO versions or lineage datasets are not the same across results:
BUSCO: 5.8.2 (euk_genome_met, metaeuk) / Lineage: actinopterygii_odb12 (genomes:75, BUSCOs:7207)
BUSCO: 6.0.0 (euk_genome_min, miniprot) / Lineage: actinopterygii_odb12 (genomes:75, BUSCOs:7207)

HiC contact map of curated assembly

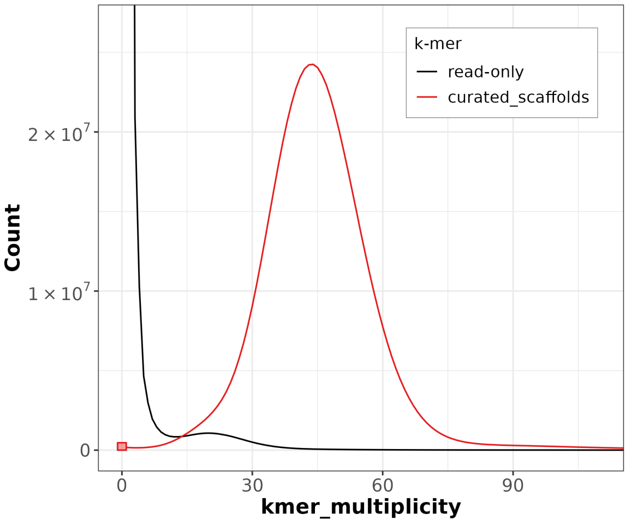


collapsed [LINK]

K-mer spectra of curated assembly

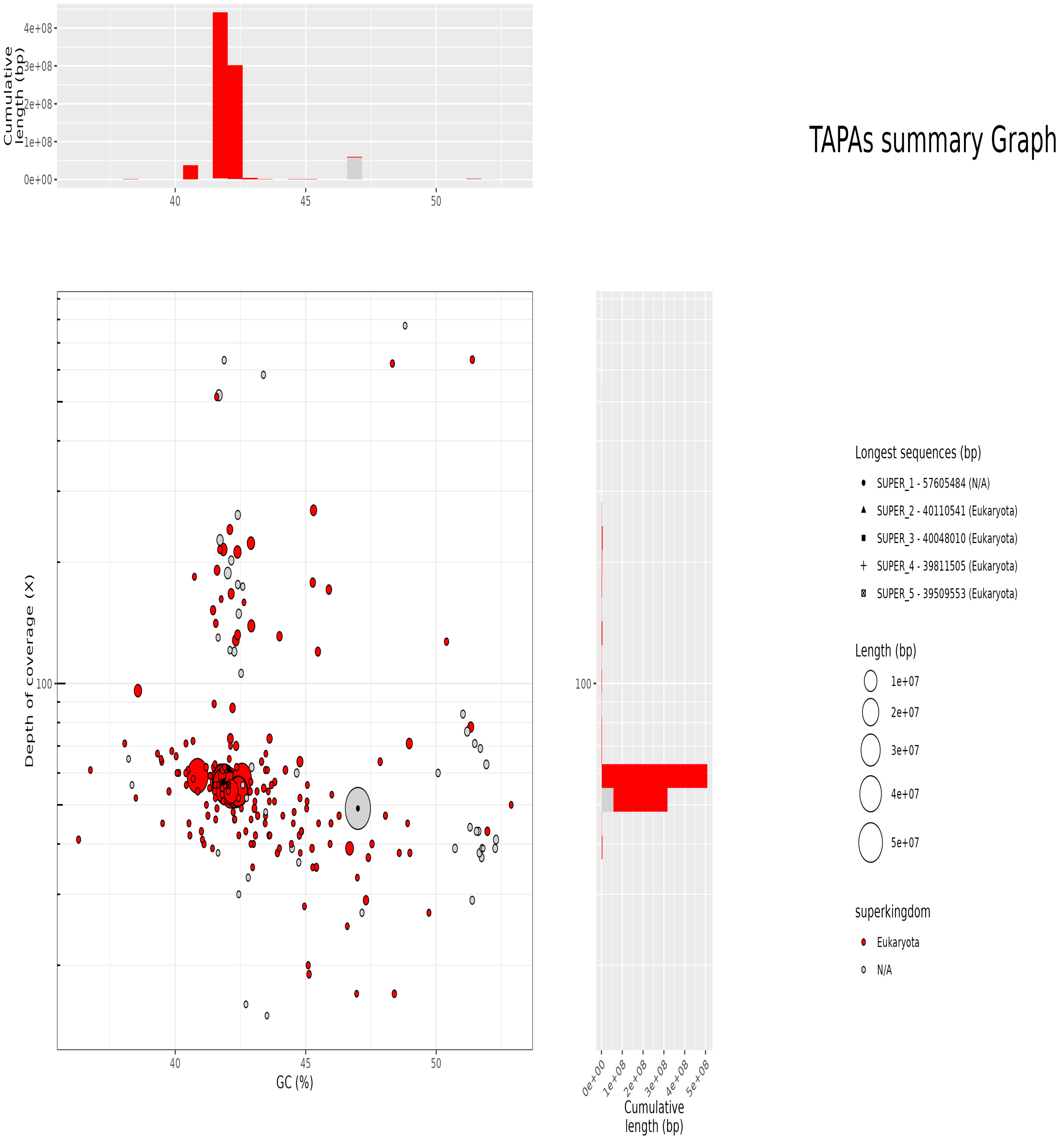


Distribution of k-mer counts per copy numbers found in asm



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

Post-curation contamination screening



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	52	176

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: 2025-11-18 15:21:18 CET