ERGA Assembly Report

v24.10.15

Tags: ATLASea[INVALID TAG]

TxID	1986054
ToLID	**fGobPun1**
Species	Gobiesox punctulatus
Class	Actinopteri
Order	Blenniiformes

Genome Traits	Expected	Observed
Haploid size (bp)	533,416,933	556,726,225
Haploid Number	19 (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.Q48

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

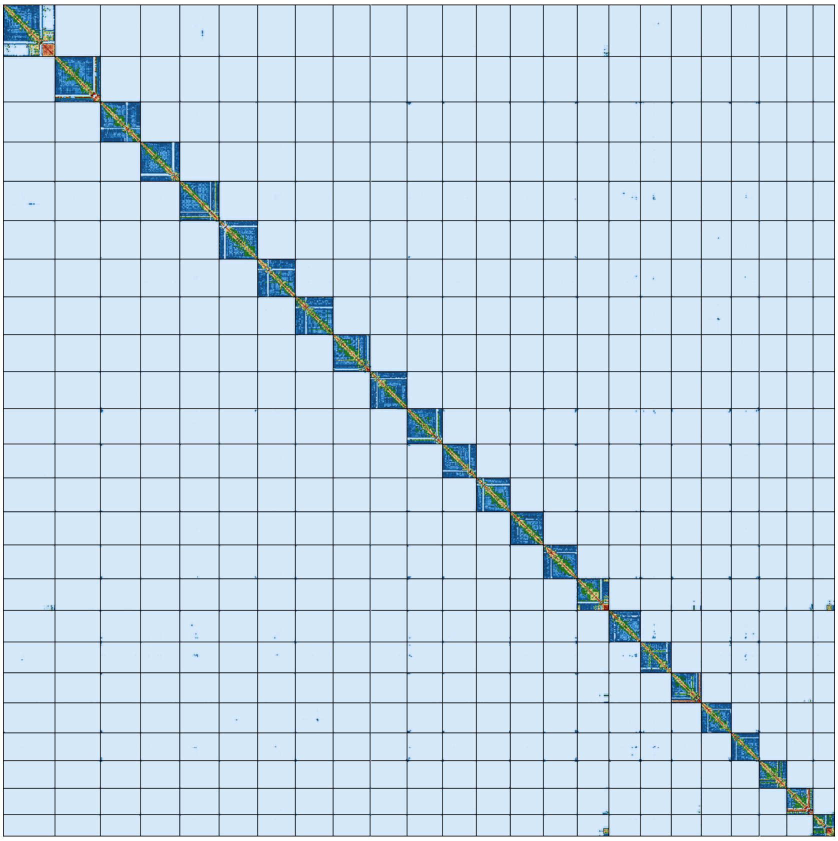
. Interventions/Gb: 31
. Contamination notes: ""
. Other observations: "The assembly of Gobiesox punctulatus (fGobPun1) is based on 56X PacBio data and 143X Arima Hi-C data generated as part of the ATLASea programme (https://www.atlasea.fr). The assembly process included the following steps: initial 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, 82 regions totaling 20.37 Mb (with the largest being 10.748 Mb) were identified as haplotypic duplications and removed. The mitochondrial genome was assembled using OATK. During manual curation, 1 haplotypic regions were removed, totaling 0.171Mb (with the largest being 0.171Mb). 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	556,896,968	556,726,225
GC %	44.06	44.06
Gaps/Gbp	134.67	138.31
Total gap bp	7,500	8,600
Scaffolds	34	32
Scaffold N50	23,545,284	23,545,384
Scaffold L50	11	11
Scaffold L90	22	21
Contigs	109	109
Contig N50	21,185,705	21,185,705
Contig L50	12	12
Contig L90	27	27
QV	48.7834	48.7838
Kmer compl.	89.974	89.9711
BUSCO sing.	98.1%	98.2%
BUSCO dupl.	0.4%	0.4%
BUSCO frag.	0.2%	0.2%
BUSCO miss.	1.3%	1.3%

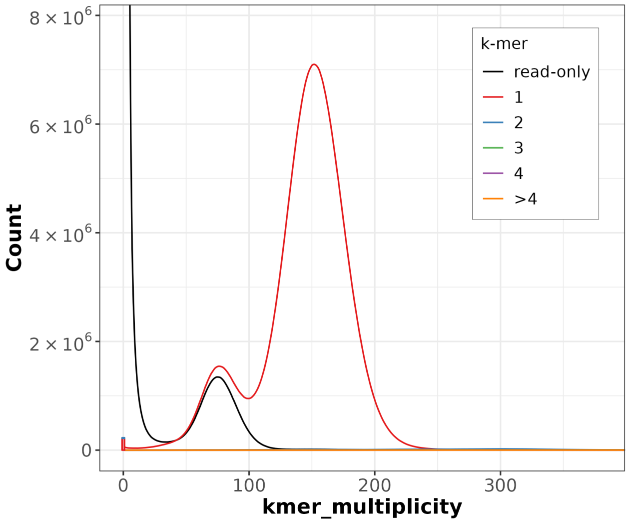
BUSCO: 6.0.0 (euk_genome_min, miniprot) / Lineage: actinopterygii_odb12 (genomes:75, BUSCOs:7207)

HiC contact map of curated assembly

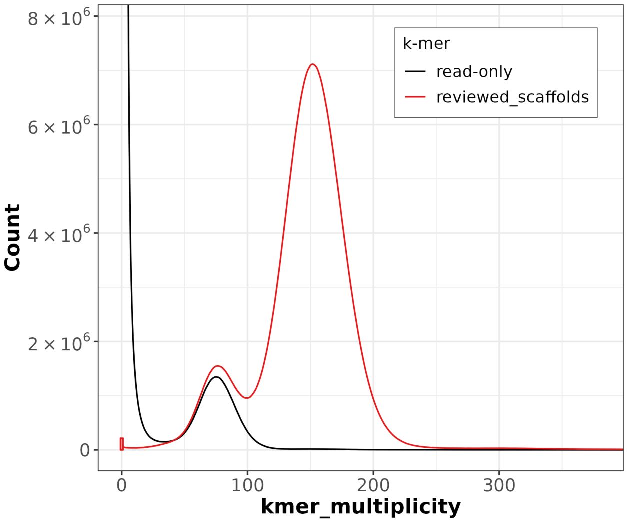


collapsed [LINK]

K-mer spectra of curated assembly

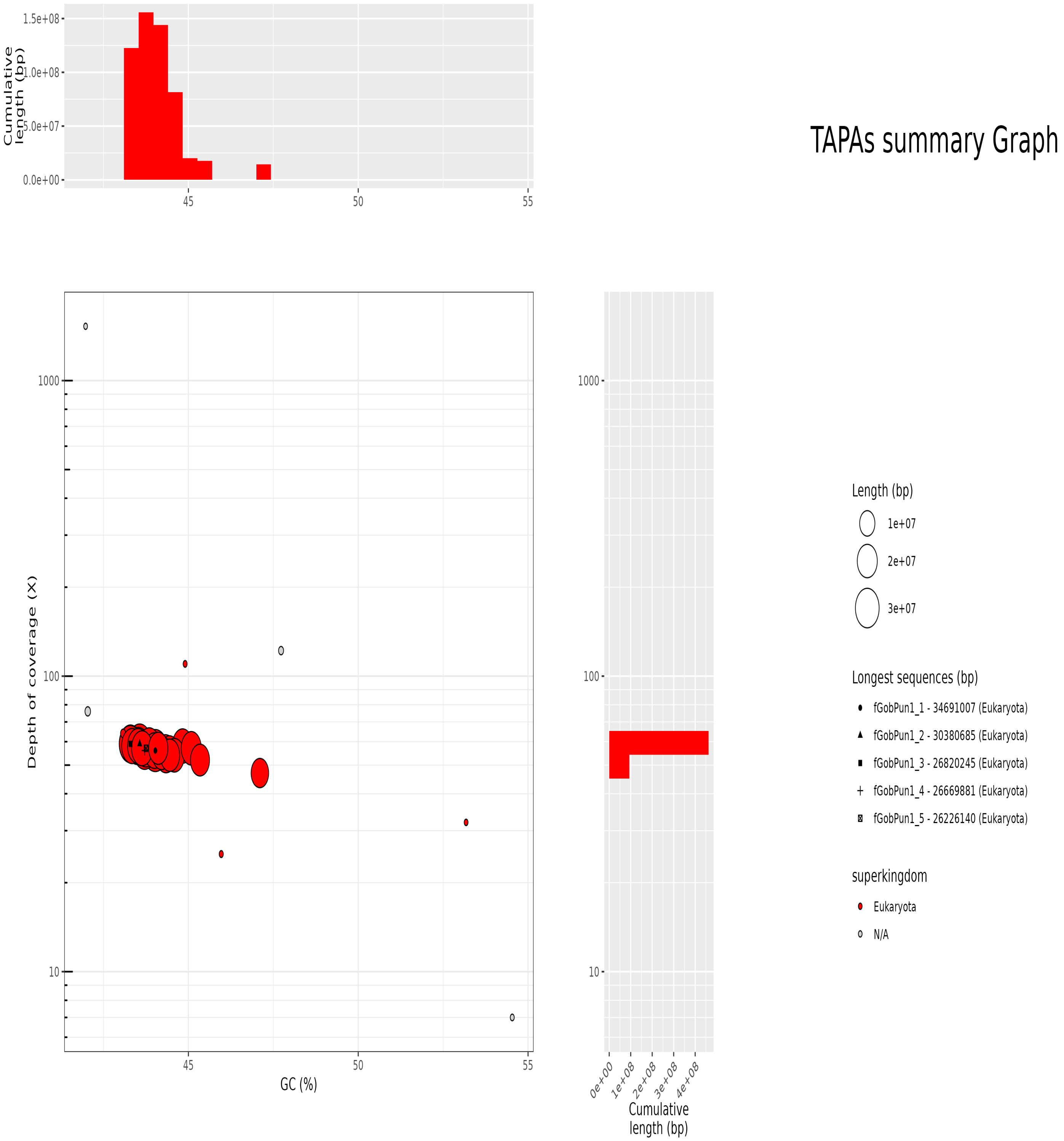


Distribution of k-mer counts per copy numbers found in asm



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	56	143

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: 2025-10-25 01:30:57 CEST