ERGA Assembly Report

v24.10.15

Tags: ATLASea[INVALID TAG]

TxID	13225
ToLID	**fPomMin1**
Species	Pomatoschistus minutus
Class	Actinopteri
Order	Gobiiformes

Genome Traits	Expected	Observed
Haploid size (bp)	775,039,688	861,159,696
Haploid Number	22 (source: ancestor)	23
Ploidy	2 (source: ancestor)	2
Sample Sex	Unknown	Unknown

EBP metrics summary and curation notes

Obtained EBP quality metric for collapsed: 7.7.Q63

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

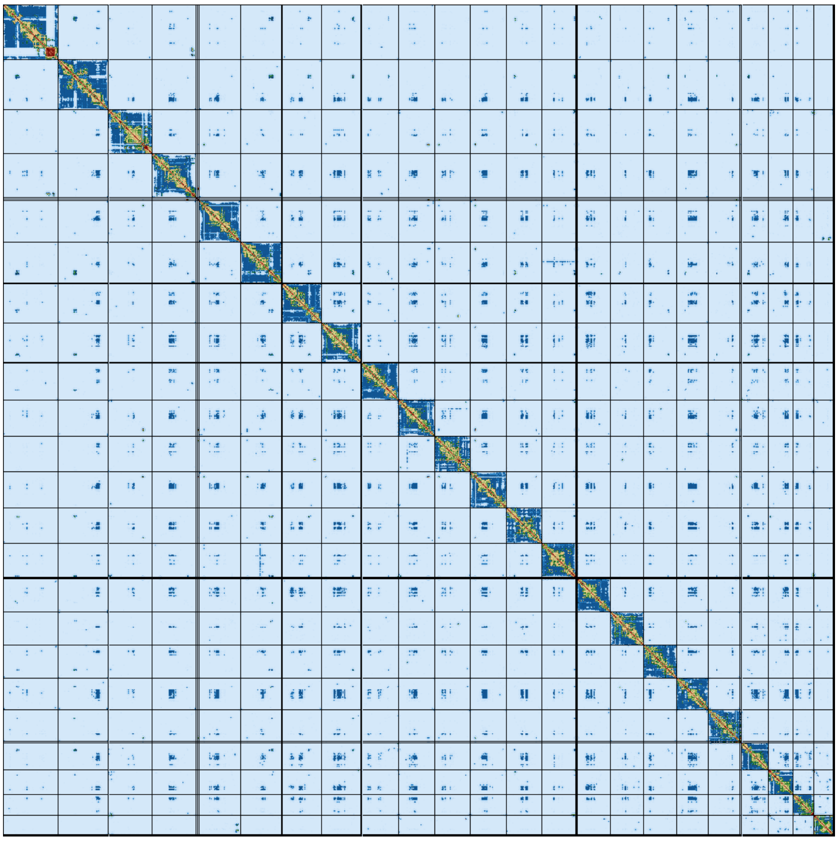
. Interventions/Gb: 95
. Contamination notes: ""
. Other observations: "The assembly of Pomatoschistus minutus (fPomMin1.1) is based on 54X of PacBio data and Arima Hi-C data generated as part of the ATLASea programme (https://www.atlasea.fr). The assembly process included the following steps: initial 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. In total, 23 contigs were identified as contaminants (bacterial, archaeal, or viral), totaling 0.48 Mb (with the largest being 0.047 Mb). Additionally, 484 regions totaling 43 Mb were identified as haplotypic duplications and removed (with the largest being 2.49 Mb). The mitochondrial genome was assembled using OATK. Finally, the primary assembly was analyzed and manually improved using Pretext. During manual curation, 12 haplotypic regions and 0 contaminant sequences were removed, totaling 4.9 Mb (with the largest being 1.9 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	866,169,693	861,159,696
GC %	42.64	42.63
Gaps/Gbp	404.08	448.23
Total gap bp	35,000	43,900
Scaffolds	182	105
Scaffold N50	36,954,725	37,552,910
Scaffold L50	11	10
Scaffold L90	21	20
Contigs	532	491
Contig N50	14,259,656	14,259,656
Contig L50	19	19
Contig L90	116	112
QV	48.8059	63.3027
Kmer compl.	78.9345	79.1647
BUSCO sing.	94.8%	95.0%
BUSCO dupl.	1.3%	1.1%
BUSCO frag.	0.7%	0.7%
BUSCO miss.	3.2%	3.2%

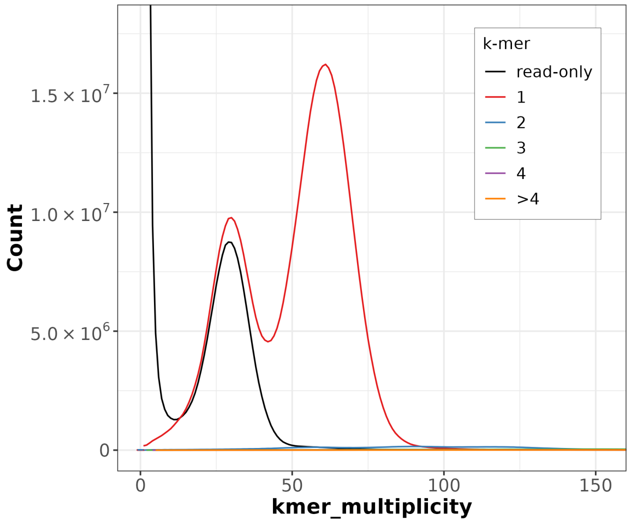
BUSCO: 5.4.3 (euk_genome_met, metaeuk) / Lineage: actinopterygii_odb10 (genomes:26, BUSCOs:3640)

HiC contact map of curated assembly

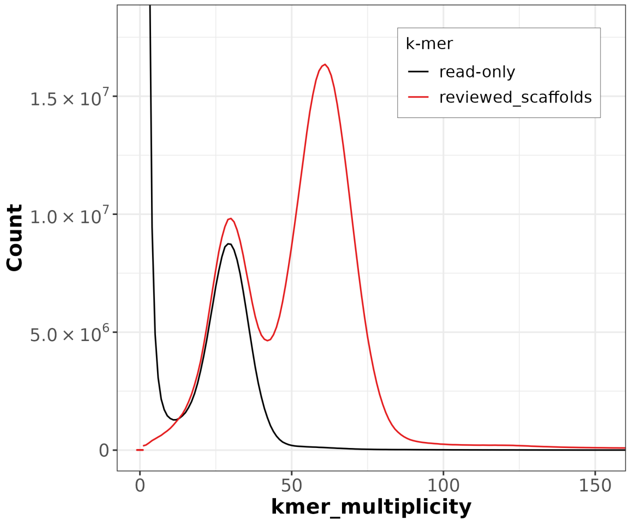


collapsed [LINK]

K-mer spectra of curated assembly

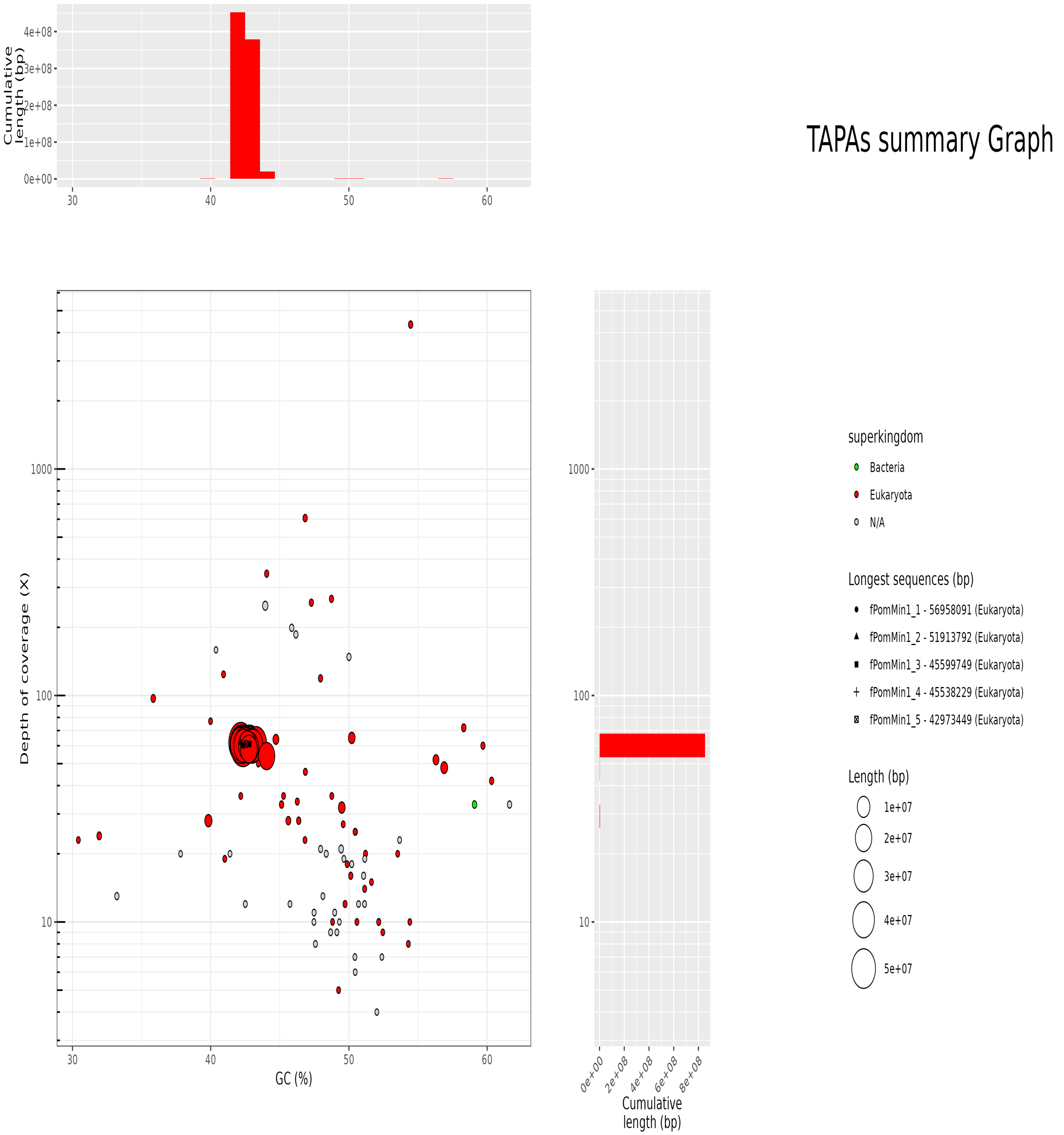


Distribution of k-mer counts per copy numbers found in asm



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	PACBIO Hifi	Arima
Coverage	61	185

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: Arnaud Couloux
Affiliation: Genoscope

Date and time: 2025-04-18 01:09:56 CEST