ERGA Assembly Report

v24.10.15

Tags: ATLASea[INVALID TAG]

TxID	112728
ToLID	**fPolAme2**
Species	Polyprion americanus
Class	Actinopteri
Order	Acropomatiformes

Genome Traits	Expected	Observed
Haploid size (bp)	768,293,502	767,342,601
Haploid Number	24 (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:

Curator notes

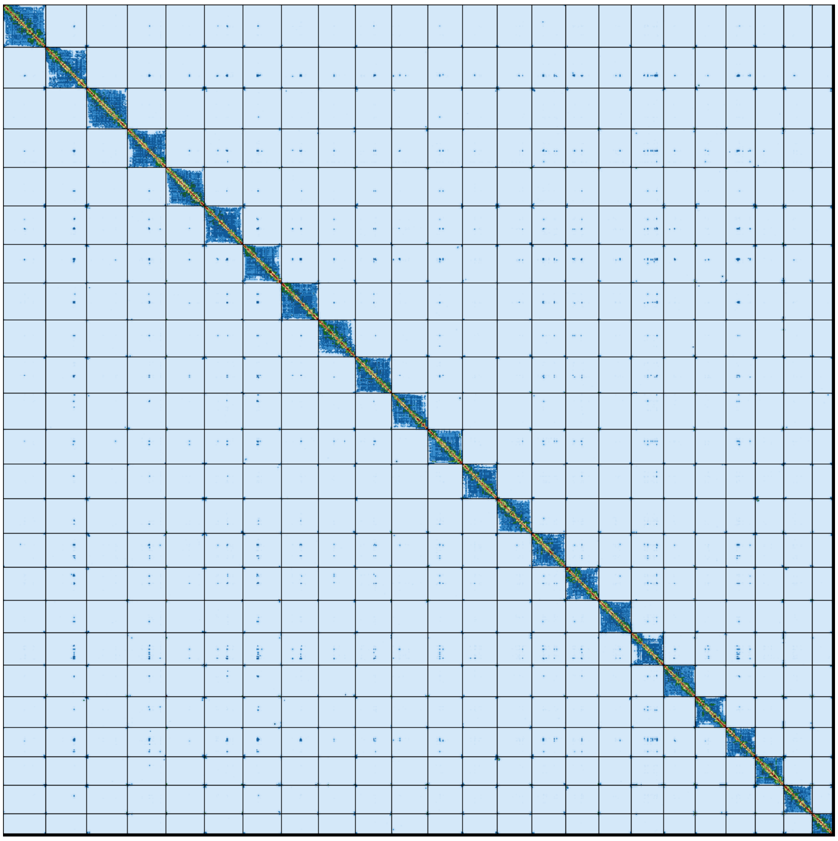
. Interventions/Gb: 8
. Contamination notes: ""
. Other observations: "The assembly of Polyprion americanus (fPolAme2) is based on 70X PacBio data and 187X Arima Hi-C data generated as part of the ATLASea programme (https://www.atlasea.fr). The assembly process included the following steps: initial PacBio 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, 64 regions totaling 8.695 Mb (with the largest being 1.955 Mb) were identified as haplotypic duplications and removed. Mitochondrial genome was assembled using OATK. Finally, the primary assembly was analyzed and manually improved using Pretext. 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	767,366,565	767,342,601
GC %	40.4	40.4
Gaps/Gbp	27.37	29.97
Total gap bp	2,100	2,700
Scaffolds	50	47
Scaffold N50	33,312,100	33,312,100
Scaffold L50	11	11
Scaffold L90	21	21
Contigs	71	70
Contig N50	31,516,000	31,516,000
Contig L50	11	11
Contig L90	23	23
QV	48.8898	48.9136
Kmer compl.	97.646	97.6459
BUSCO sing.	99.4%	99.4%
BUSCO dupl.	0.2%	0.2%
BUSCO frag.	0.0%	0.0%
BUSCO miss.	0.4%	0.4%

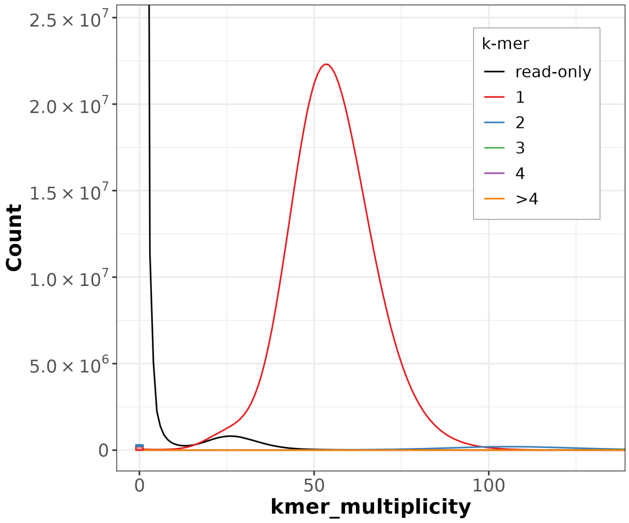
BUSCO: 6.0.0 (euk_genome_min, miniprot) / Lineage: actinopterygii_odb12 (genomes:75, BUSCOs:7207)

HiC contact map of curated assembly

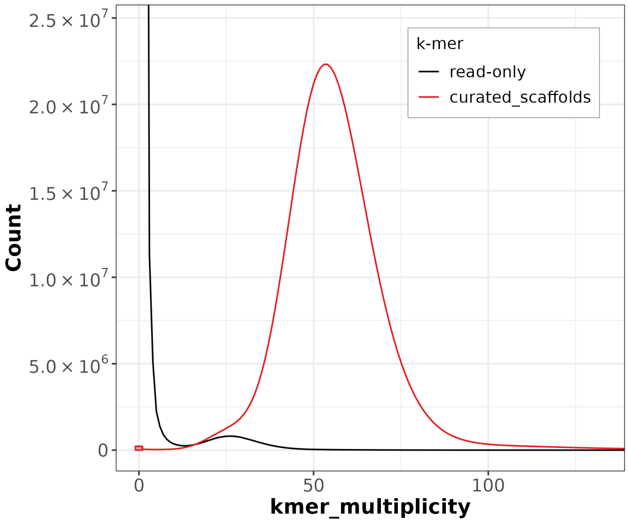


collapsed [LINK]

K-mer spectra of curated assembly

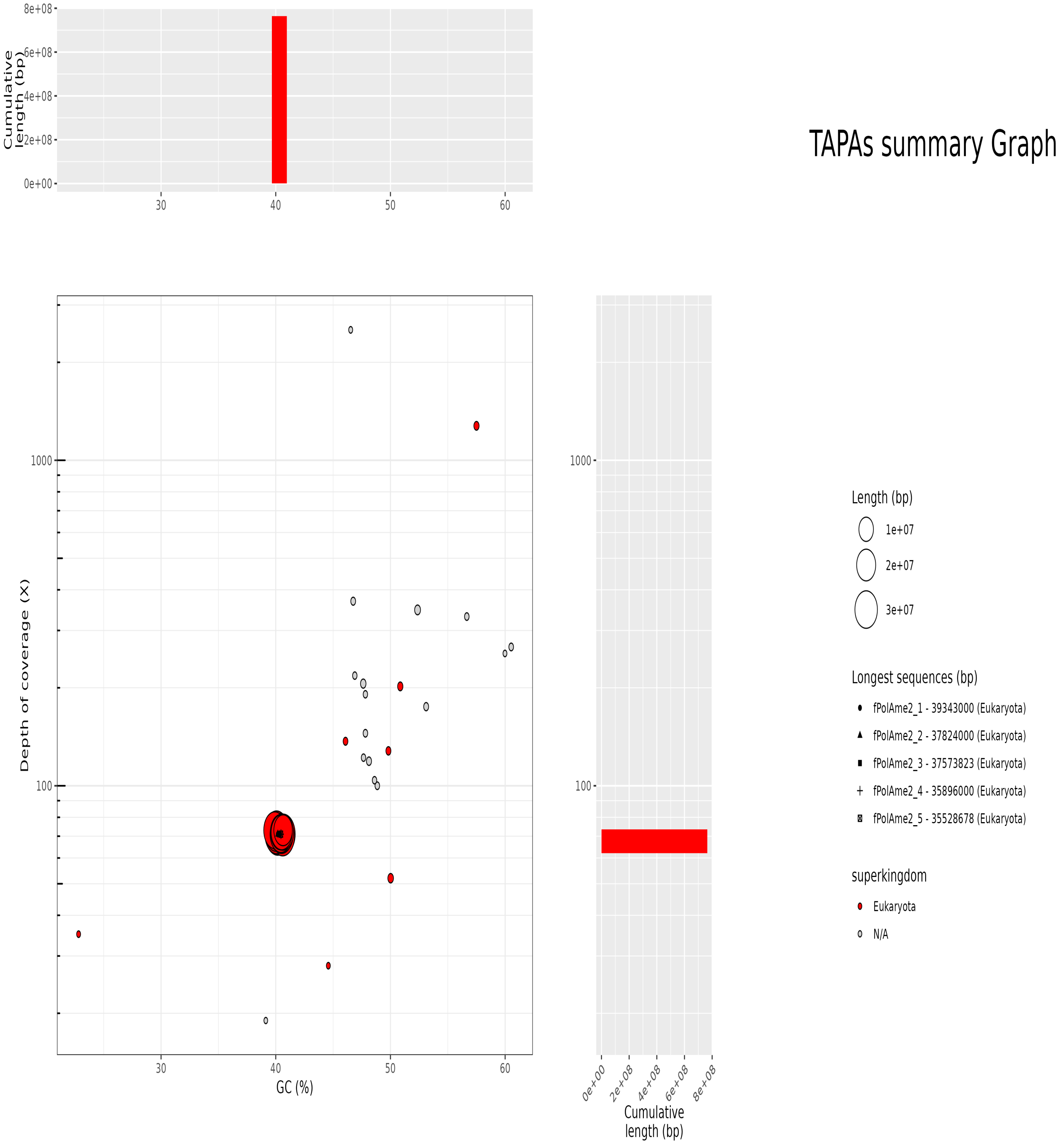


Distribution of k-mer counts per copy numbers found in asm



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	70	187

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: Jean-Marc Aury
Affiliation: Genoscope

Date and time: 2025-12-19 21:44:45 CET