ERGA Assembly Report

v24.10.15

Tags: ATLASea[INVALID TAG]

TxID	236585
ToLID	**fHaeFla1**
Species	Haemulon flavolineatum
Class	Actinopteri
Order	Lutjaniformes

Genome Traits	Expected	Observed
Haploid size (bp)	700,093,231	709,698,976
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.Q46

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

. Kmer completeness value is less than 90 for collapsed

Curator notes

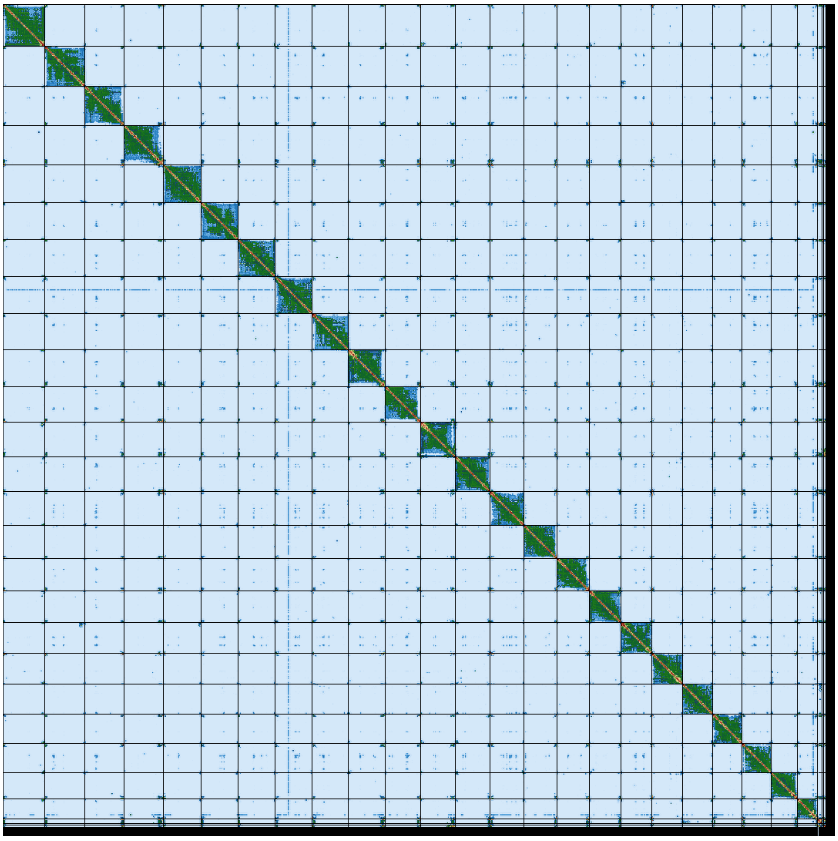
. Interventions/Gb: 7
. Contamination notes: ""
. Other observations: "The assembly of Haemulon flavolineatum (fHaeFla1) is based on 85X PacBio data and 242X Arima Hi-C data generated as part of the ATLASea programme (https://www.atlasea.fr).The assembly process included the following steps: initial 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 identified as contaminants (bacterial, archaeal, or viral). Additionally, 52 regions totaling 9.5 Mb (with the largest being 1.89 Mb) were identified as haplotypic duplications and removed. The mitochondrial genome was assembled using oatk. The obtained sequence is linear. Finally, the primary assembly was analyzed and manually improved using Pretext. During manual curation, no supplementary haplotypic regions were removed. Chromosome-scale scaffolds confirmed by Hi-C data were named in order of size. The end of the fHaeFla1_1 was difficult to organize. "

Quality metrics table

Metrics	Pre-curation collapsed	Curated collapsed
Total bp	709,696,597	709,698,976
GC %	43.28	43.28
Gaps/Gbp	12.68	15.5
Total gap bp	900	1,400
Scaffolds	57	56
Scaffold N50	30,305,000	30,305,000
Scaffold L50	11	11
Scaffold L90	22	22
Contigs	66	67
Contig N50	30,305,000	30,305,000
Contig L50	11	11
Contig L90	23	23
QV	46.9892	46.9892
Kmer compl.	89.2217	89.2217
BUSCO sing.	99.4%	99.4%
BUSCO dupl.	0.3%	0.3%
BUSCO frag.	0.0%	0.0%
BUSCO miss.	0.3%	0.3%

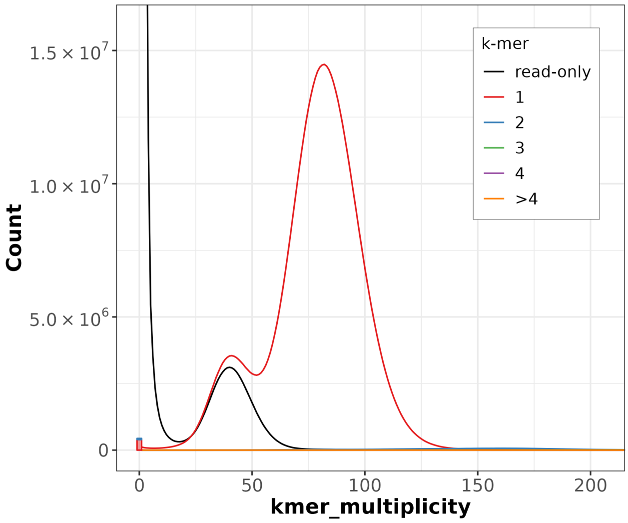
BUSCO: 6.0.0 (euk_genome_min, miniprot) / Lineage: actinopterygii_odb12 (genomes:75, BUSCOs:7207)

HiC contact map of curated assembly

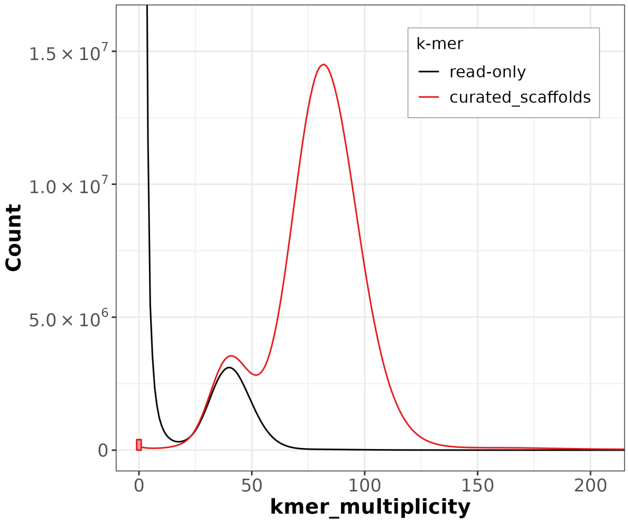


collapsed [LINK]

K-mer spectra of curated assembly

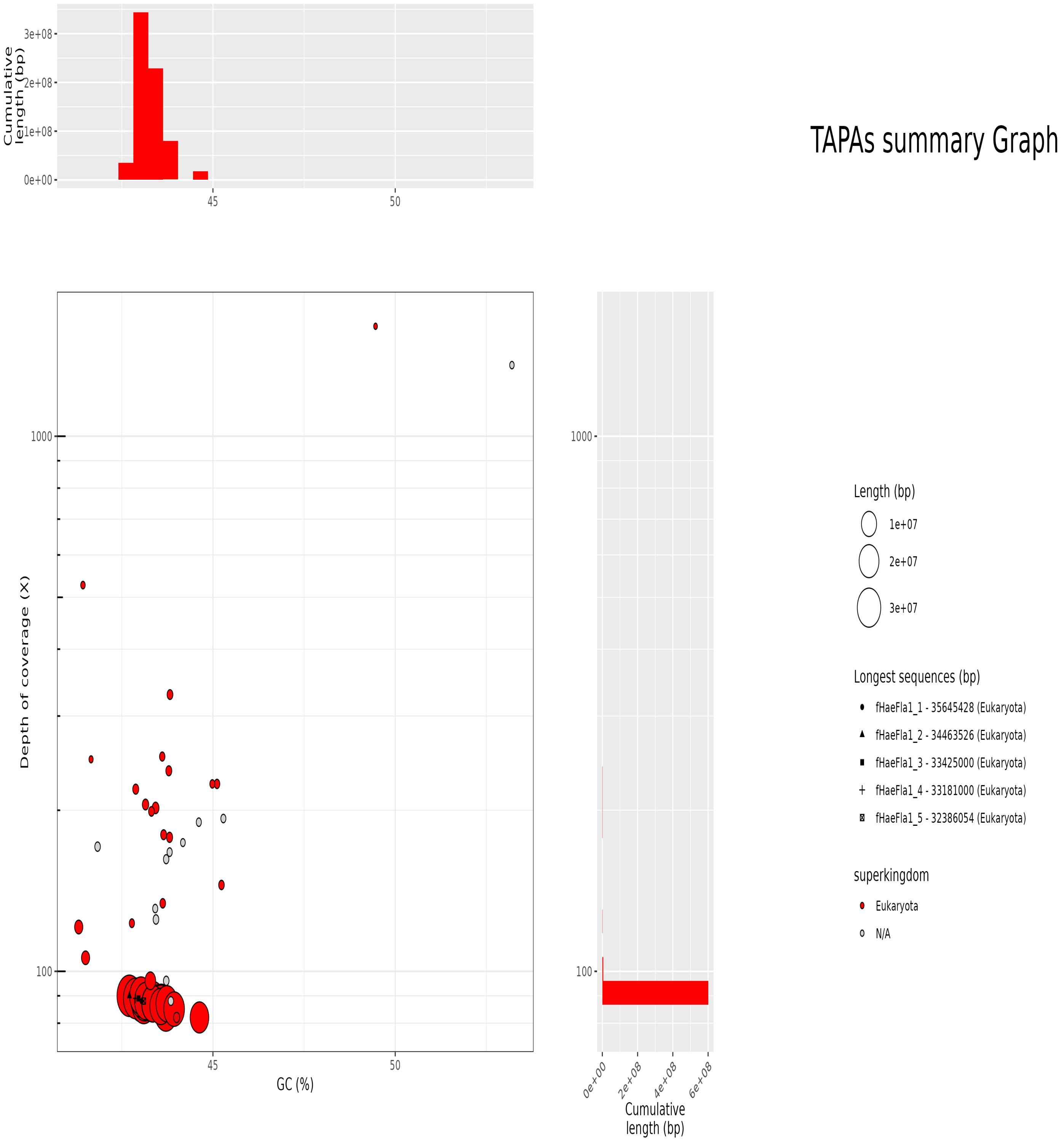


Distribution of k-mer counts per copy numbers found in asm



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	85	242

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: Caroline Menguy
Affiliation: Genoscope

Date and time: 2025-11-25 18:51:09 CET