

ERGA Assembly Report

v24.10.15

Tags: ATLASEa[INVALID TAG]

TxID	236586
ToLID	fHaeAur1
Species	Haemulon aurolineatum
Class	Actinopteri
Order	Lutjaniformes

Genome Traits	Expected	Observed
Haploid size (bp)	720,970,852	729,967,414
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:

- . Kmer completeness value is less than 90 for collapsed

Curator notes

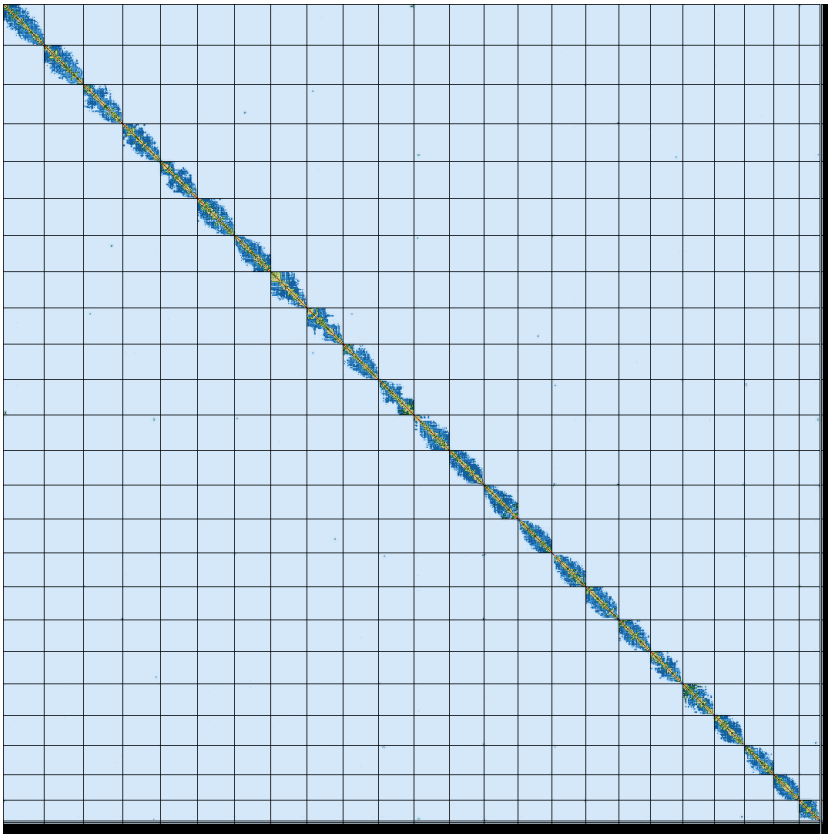
- . Interventions/Gb: 4
- . Contamination notes: ""
- . Other observations: "The assembly of Haemulon aurolineatum (fHaeAur1) is based on 47X PacBio data and 127X Arima Hi-C data generated as part of the ATLASEa programme (<https://www.atlasea.fr>). The assembly process included the following steps: 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, 13 contigs were identified as contaminants (bacterial, archaeal, or viral), totaling 0.464 Mb (with the largest being 0.07 Mb). Additionally, 44 regions totaling 10.252 Mb (with the largest being 2.187 Mb) were identified as haplotypic duplications and removed. The 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	730,013,472	729,967,414
GC %	43.1	43.1
Gaps/Gbp	16.44	17.81
Total gap bp	1,200	1,500
Scaffolds	88	86
Scaffold N50	30,839,100	30,839,200
Scaffold L50	12	12
Scaffold L90	22	22
Contigs	100	99
Contig N50	29,649,258	29,649,258
Contig L50	12	12
Contig L90	23	23
QV	48.7864	48.7986
Kmer compl.	80.1578	80.1585
BUSCO sing.	99.5%	99.5%
BUSCO dupl.	0.2%	0.2%
BUSCO frag.	0.0%	0.0%
BUSCO miss.	0.2%	0.2%

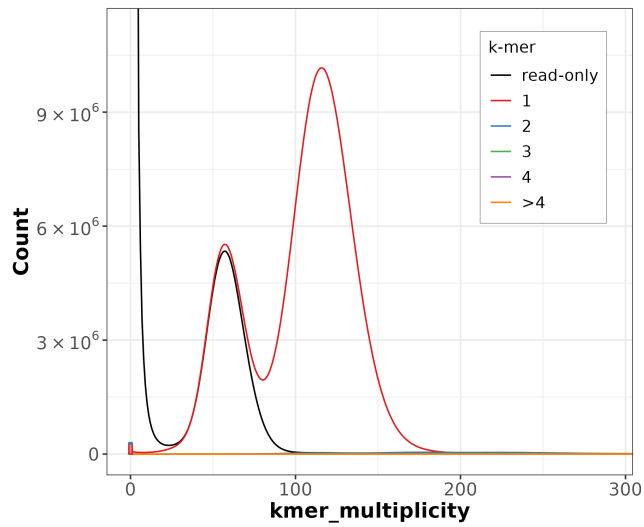
BUSCO: 6.0.0 (euk_genome_min, miniprot) / Lineage: actinopterygii_odb12 (genomes:75, BUSCOs:7207)

HiC contact map of curated assembly

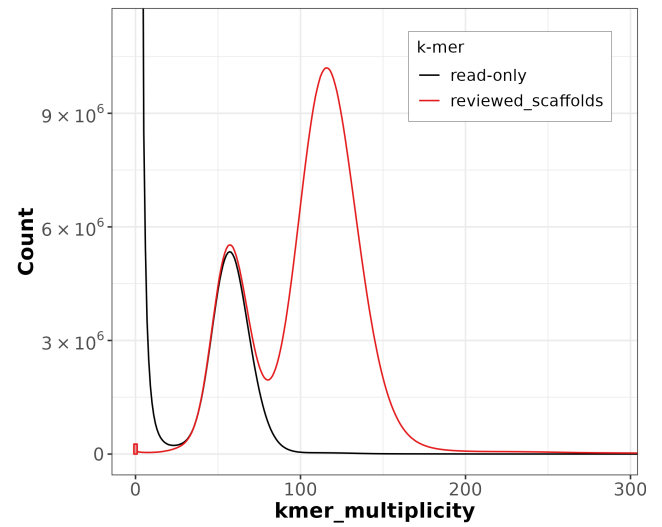


collapsed [\[LINK\]](#)

K-mer spectra of curated assembly

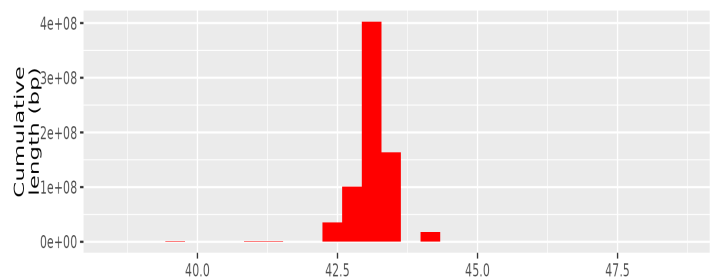


Distribution of k-mer counts per copy numbers found in asm

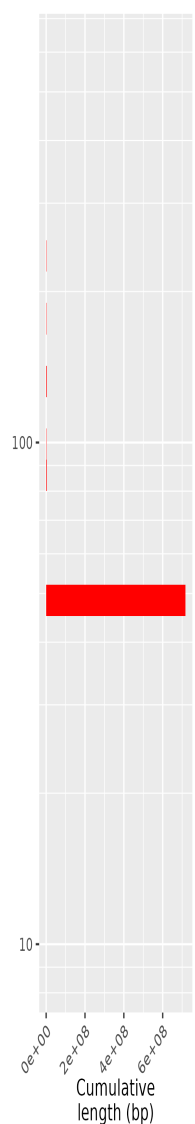
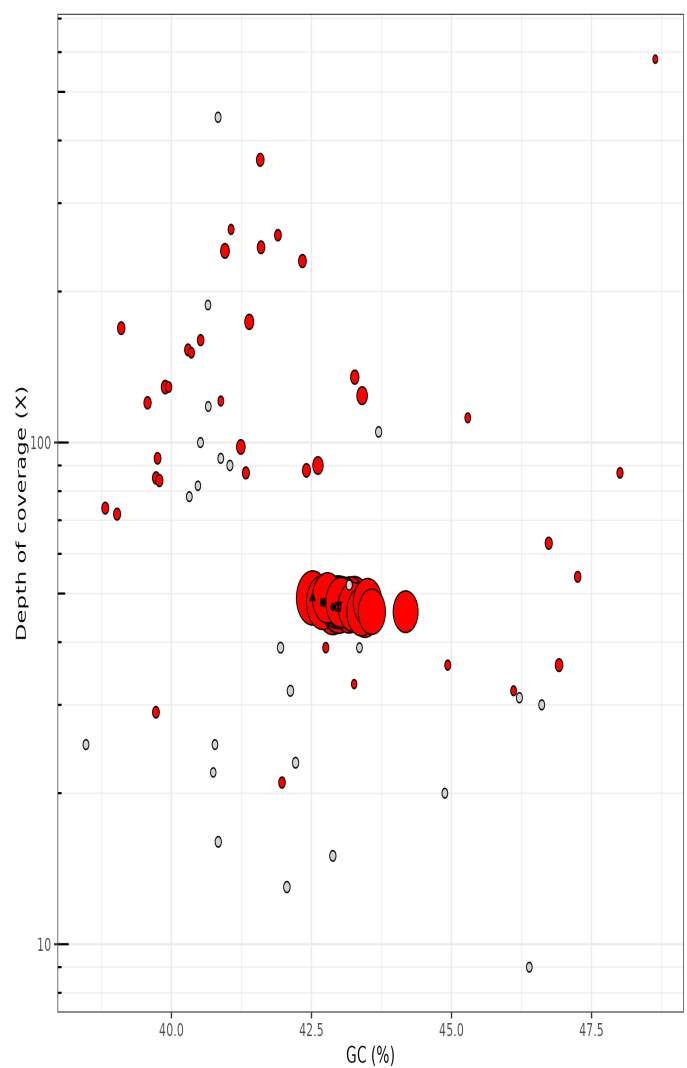


Distribution of k-mer counts coloured by their presence in reads/assemblies

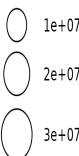
Post-curation contamination screening



TAPAs summary Graph



Length (bp)



Longest sequences (bp)

- fHaeAur1_1 - 36070000 (Eukaryota)
- ▲ fHaeAur1_2 - 34437464 (Eukaryota)
- fHaeAur1_3 - 34338000 (Eukaryota)
- + fHaeAur1_4 - 33087000 (Eukaryota)
- ▣ fHaeAur1_5 - 32672000 (Eukaryota)

superkingdom

- Eukaryota
- N/A

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	47	127

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 14:21:01 CEST