ERGA Assembly Report

v24.10.15

Tags: ATLASea[INVALID TAG]

TxID	238709
ToLID	**fAthHes1**
Species	Atherina hepsetus
Class	Actinopteri
Order	Atheriniformes

Genome Traits	Expected	Observed
Haploid size (bp)	1,005,375,545	1,033,458,085
Haploid Number	18 (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:

. Observed Haploid Number is different from Expected

. Kmer completeness value is less than 90 for collapsed

Curator notes

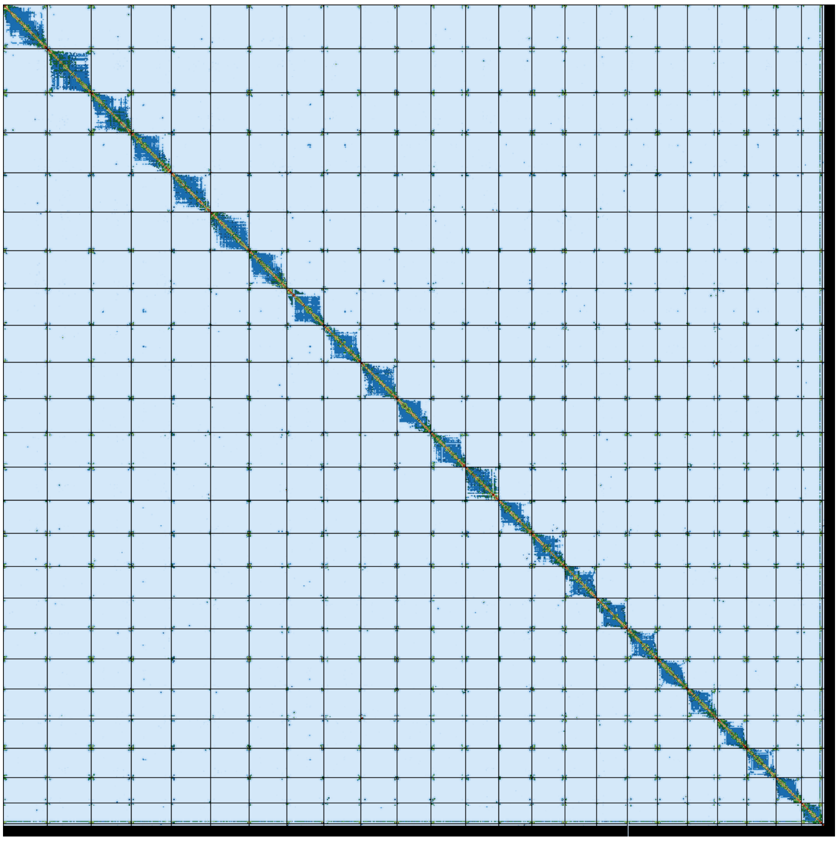
. Interventions/Gb: 39
. Contamination notes: ""
. Other observations: "The assembly of Atherina hepsetus (fAthHes1) is based on 53X PacBio data and 183X Arima Hi-C data generated as part of the ATLASea programme (https://www.atlasea.fr). The assembly process included the following steps: initial PacBio 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, 1 contigs were identified as contaminants (bacterial, archaeal, or viral), totaling 0.054 Mb (with the largest being 0.054 Mb). Additionally, 265 regions totaling 31.353 Mb (with the largest being 1.249 Mb) were identified as haplotypic duplications and removed. The mitochondrial genome was assembled using OATK. During manual curation, 7 haplotypic regions were removed, totaling 2.27Mb (with the largest being 0.648Mb). 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	1,035,791,441	1,033,458,085
GC %	39.39	39.38
Gaps/Gbp	234.6	245.78
Total gap bp	24,300	27,600
Scaffolds	162	133
Scaffold N50	43,406,156	42,607,871
Scaffold L50	10	11
Scaffold L90	21	22
Contigs	405	387
Contig N50	27,725,460	27,725,460
Contig L50	15	15
Contig L90	67	68
QV	46.4887	46.4981
Kmer compl.	82.5358	82.4944
BUSCO sing.	96.9%	99.2%
BUSCO dupl.	0.5%	0.3%
BUSCO frag.	0.7%	0.0%
BUSCO miss.	1.9%	0.4%

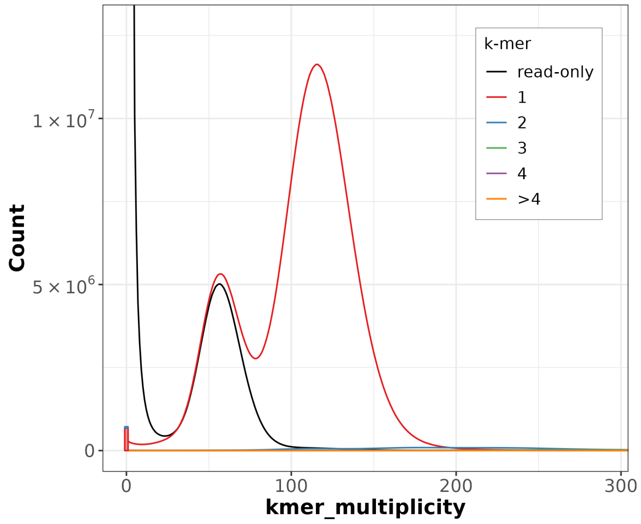
Warning! BUSCO versions or lineage datasets are not the same across results:
BUSCO: 5.8.2 (euk_genome_met, metaeuk) / Lineage: actinopterygii_odb12 (genomes:75, BUSCOs:7207)
BUSCO: 6.0.0 (euk_genome_min, miniprot) / Lineage: actinopterygii_odb12 (genomes:75, BUSCOs:7207)

HiC contact map of curated assembly

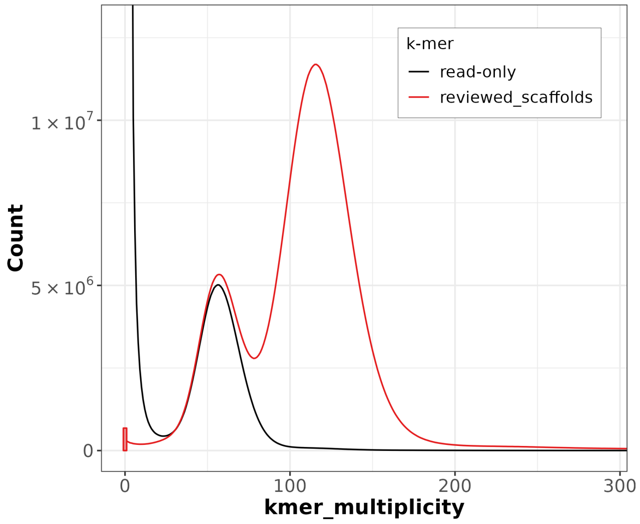


collapsed [LINK]

K-mer spectra of curated assembly

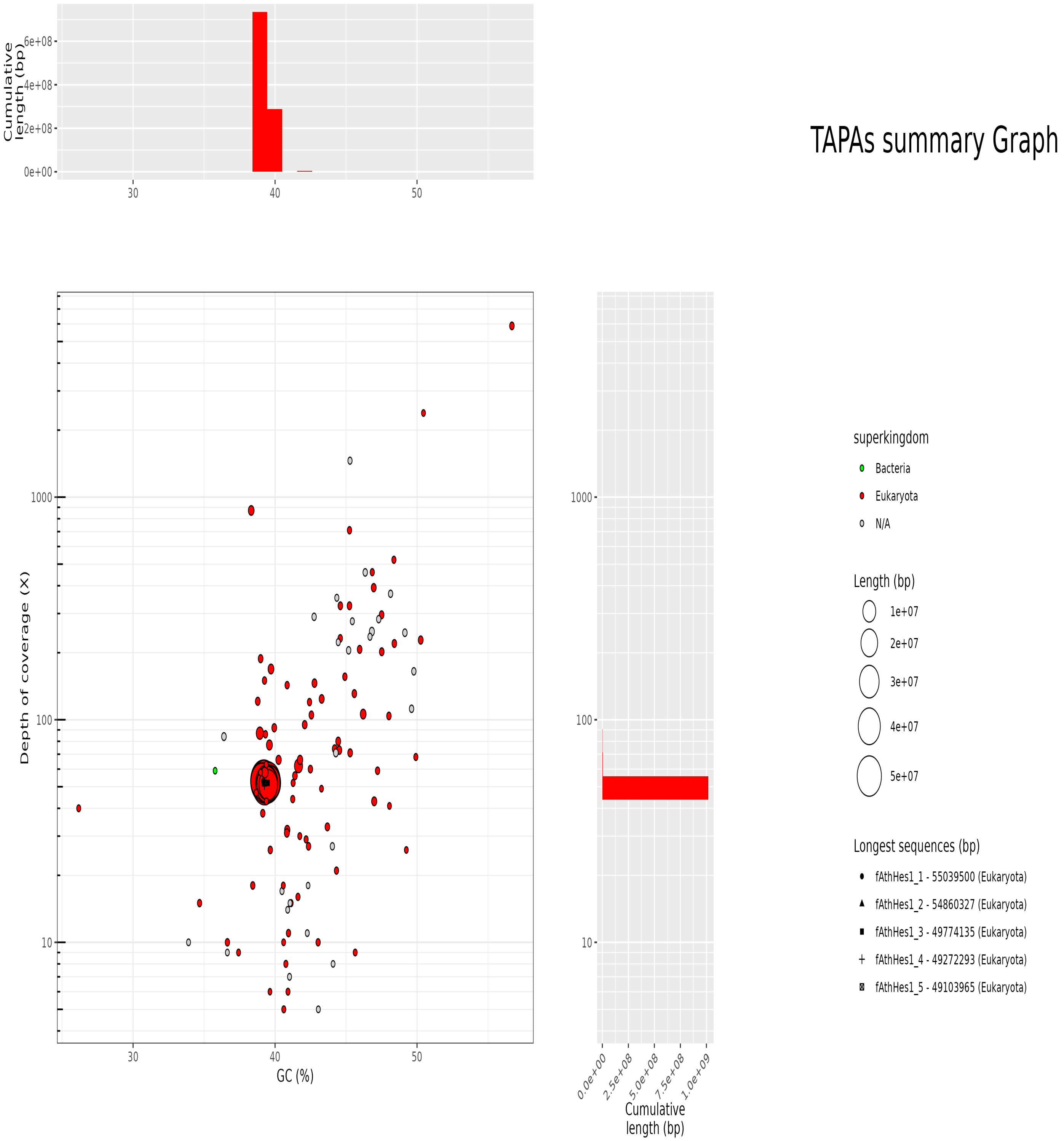


Distribution of k-mer counts per copy numbers found in asm



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	52	182

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: 2025-11-05 15:03:59 CET