ERGA Assembly Report

v24.10.15

Tags: ATLASea[INVALID TAG]

TxID	87785
ToLID	**fAthBoy1**
Species	Atherina boyeri
Class	Actinopteri
Order	Atheriniformes

Genome Traits	Expected	Observed
Haploid size (bp)	961,313,676	993,936,625
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.Q44

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

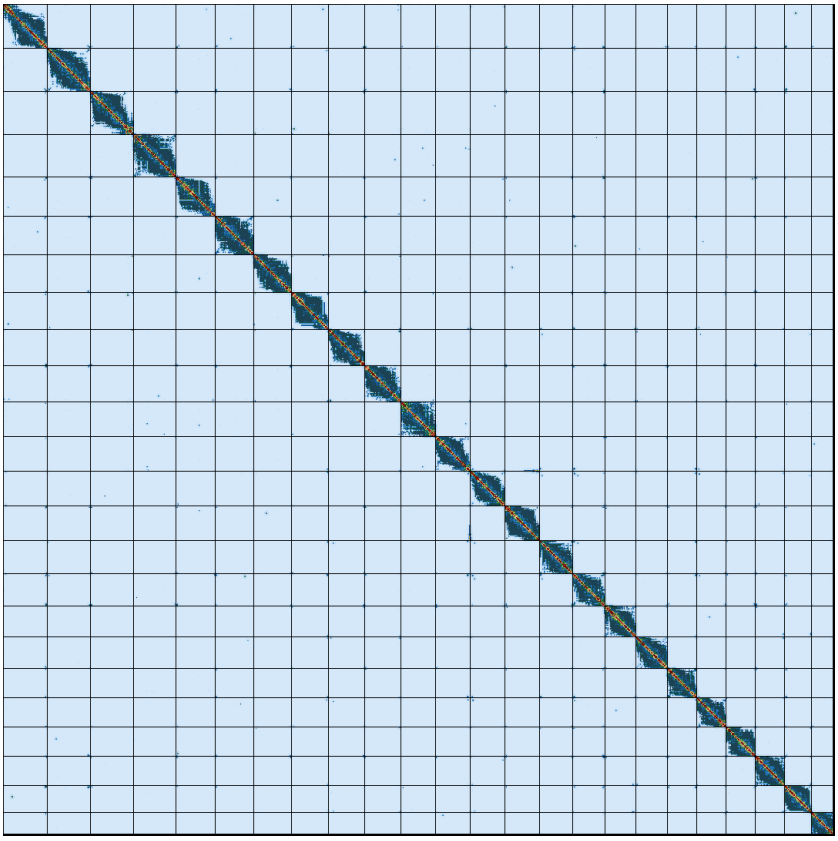
. Interventions/Gb: 29
. Contamination notes: ""
. Other observations: "The assembly of Atherina boyeri (fAthBoy1) is based on 64X PacBio data and 144X Arima Hi-C data generated as part of the ATLASea programme (https://www.atlasea.fr).The assembly process included the following steps: initial PacBio 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 contig weas identified as contaminants (bacterial), totaling 28 Kb. Additionally, 234 regions totaling 29.5 Mb (with the largest being 2.7 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. During manual curation, 5 haplotypic regions were removed, totaling 5.1 Mb (with the largest being 1.5 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	999,147,075	993,936,625
GC %	39.14	39.14
Gaps/Gbp	113.1	128.78
Total gap bp	11,300	14,900
Scaffolds	121	100
Scaffold N50	41,612,380	41,689,329
Scaffold L50	11	11
Scaffold L90	22	21
Contigs	234	228
Contig N50	30,387,814	29,832,748
Contig L50	14	14
Contig L90	42	42
QV	44.265	44.2681
Kmer compl.	73.3093	73.15
BUSCO sing.	99.2%	99.4%
BUSCO dupl.	0.5%	0.2%
BUSCO frag.	0.0%	0.0%
BUSCO miss.	0.3%	0.3%

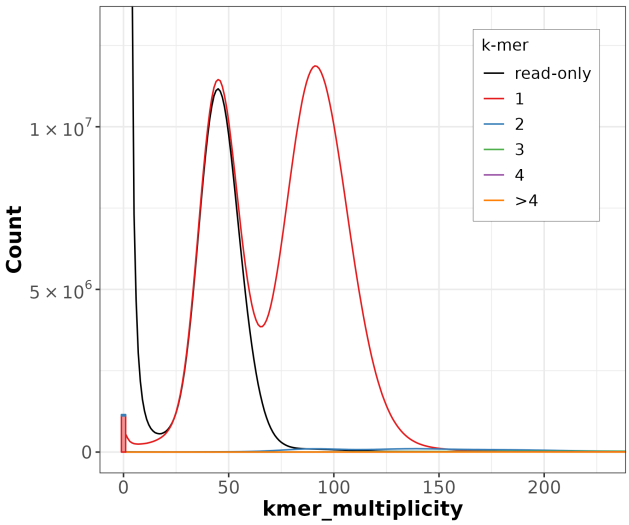
BUSCO: 6.0.0 (euk_genome_min, miniprot) / Lineage: actinopterygii_odb12 (genomes:75, BUSCOs:7207)

HiC contact map of curated assembly

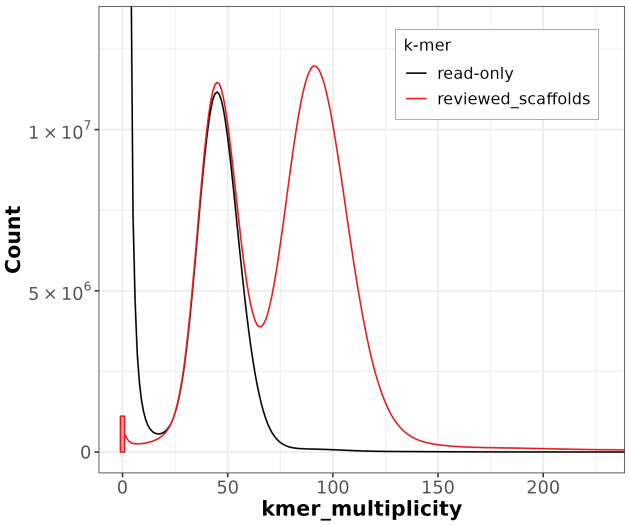


collapsed [LINK]

K-mer spectra of curated assembly

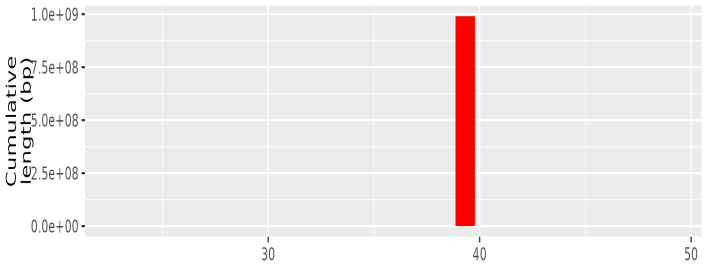


Distribution of k-mer counts per copy numbers found in asm

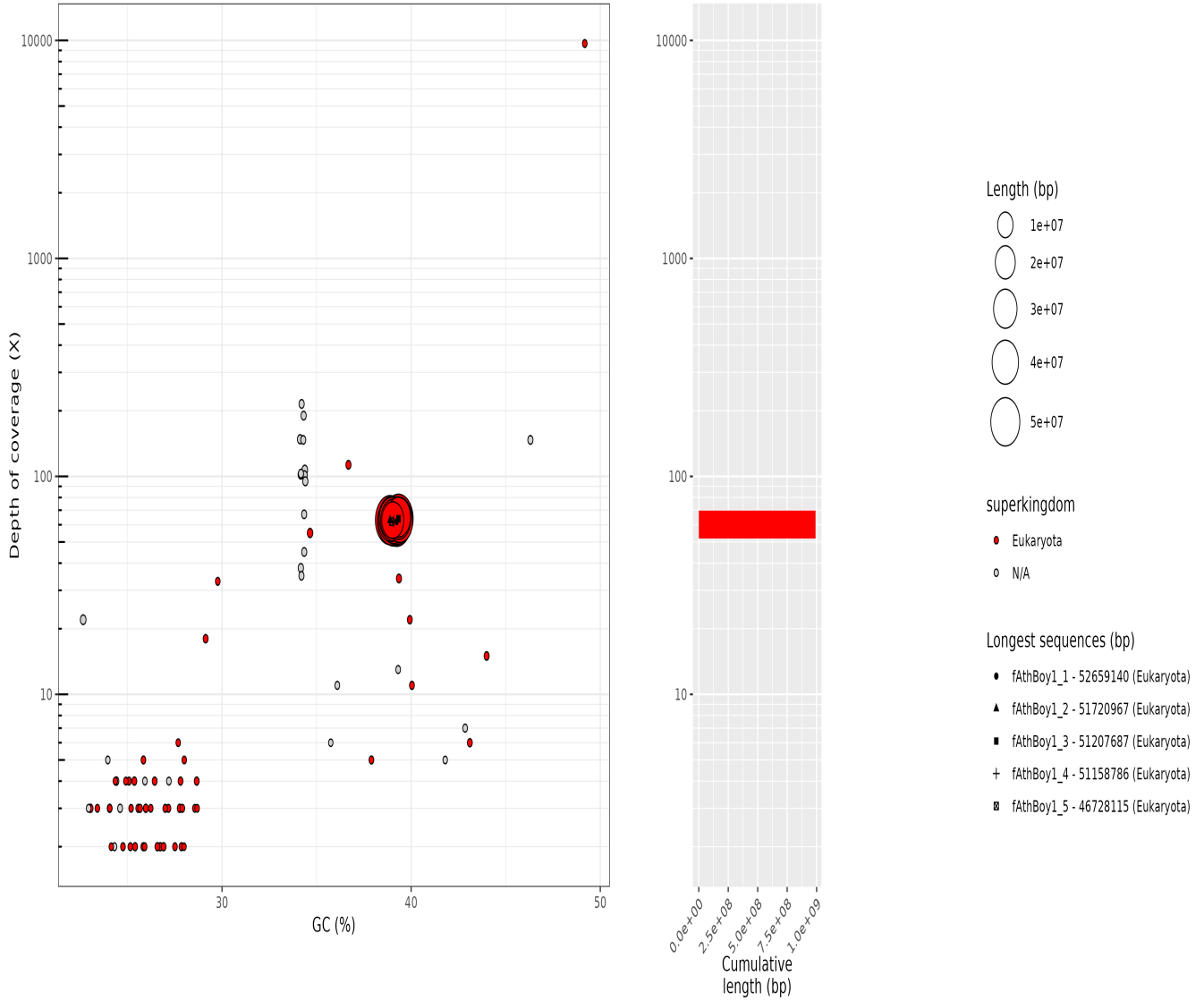


Distribution of k-mer counts coloured by their presence in reads/assemblies

Post-curation contamination screening



TAPAs summary Graph



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	64	144

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: 2026-01-09 07:08:17 CET