ERGA Assembly Report

v24.10.15

Tags: ATLASea[INVALID TAG]

TxID	241299
ToLID	**fHalRad1**
Species	Halichoeres radiatus
Class	Actinopteri
Order	Labriformes

Genome Traits	Expected	Observed
Haploid size (bp)	805,318,011	818,519,959
Haploid Number	24 (source: direct)	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

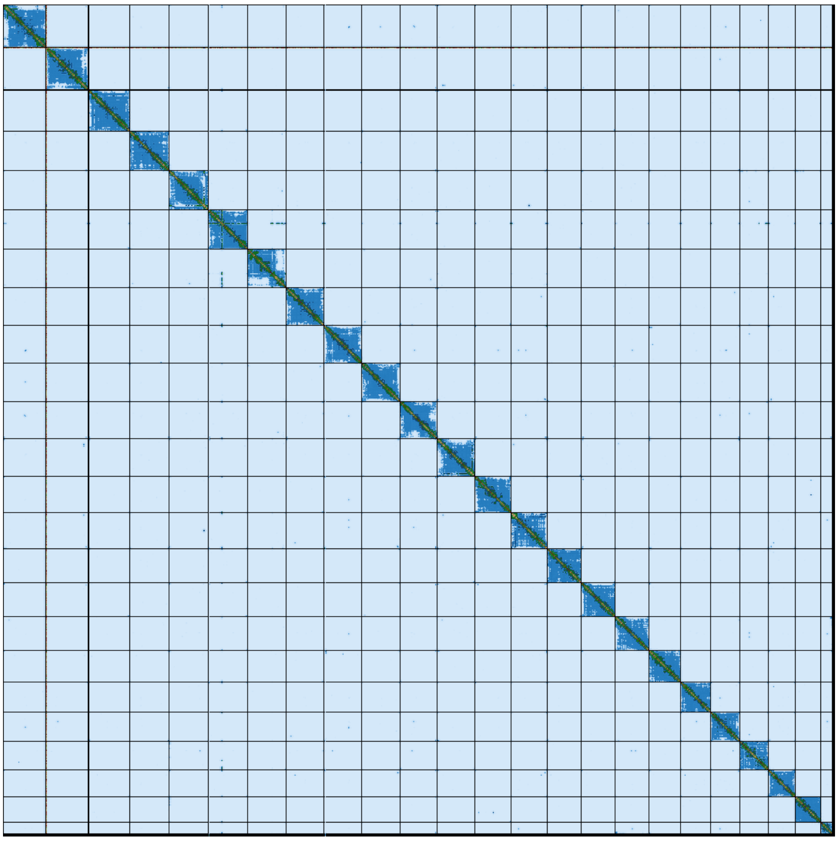
. Interventions/Gb: 65
. Contamination notes: ""
. Other observations: "The assembly of Halichoeres radiatus (fHalRad1) is based on 45X PacBio data and 198X Arima Hi-C data generated as part of the ATLASea programme (https://www.atlasea.fr). The assembly process included the following steps: initial 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, 2 contigs were identified as contaminants (bacterial, archaeal, or viral), totaling 0.09 Mb (with the largest being 0.048 Mb). Additionally, 183 regions totaling 18.993 Mb (with the largest being 2.164 Mb) were identified as haplotypic duplications and removed. The mitochondrial genome was assembled using OATK. During manual curation, 6 haplotypic regions were removed, totaling 0.939Mb (with the largest being 0.300Mb). 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	819,523,458	818,519,959
GC %	40.9	40.9
Gaps/Gbp	284.31	320.09
Total gap bp	23,300	30,000
Scaffolds	125	80
Scaffold N50	36,901,111	37,043,606
Scaffold L50	10	11
Scaffold L90	20	21
Contigs	358	342
Contig N50	22,936,681	22,272,120
Contig L50	14	14
Contig L90	38	41
QV	48.2183	48.2261
Kmer compl.	89.7256	89.7003
BUSCO sing.	99.1%	99.0%
BUSCO dupl.	0.3%	0.3%
BUSCO frag.	0.1%	0.1%
BUSCO miss.	0.5%	0.6%

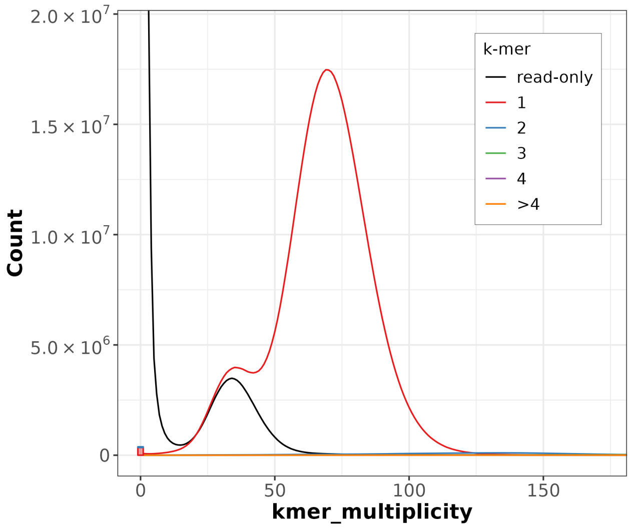
BUSCO: 6.0.0 (euk_genome_min, miniprot) / Lineage: actinopterygii_odb12 (genomes:75, BUSCOs:7207)

HiC contact map of curated assembly

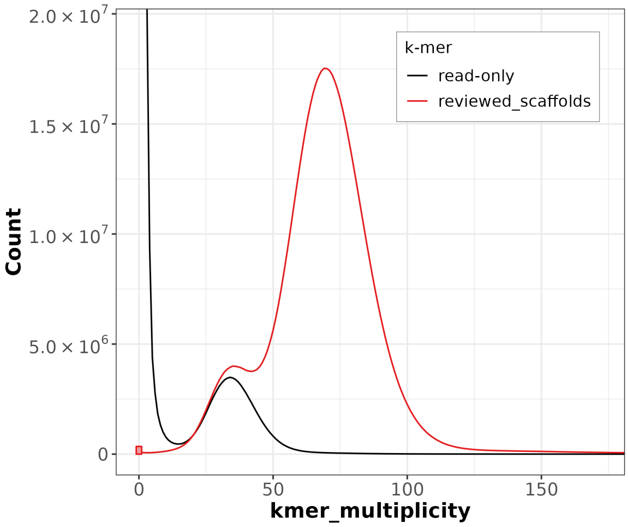


collapsed [LINK]

K-mer spectra of curated assembly

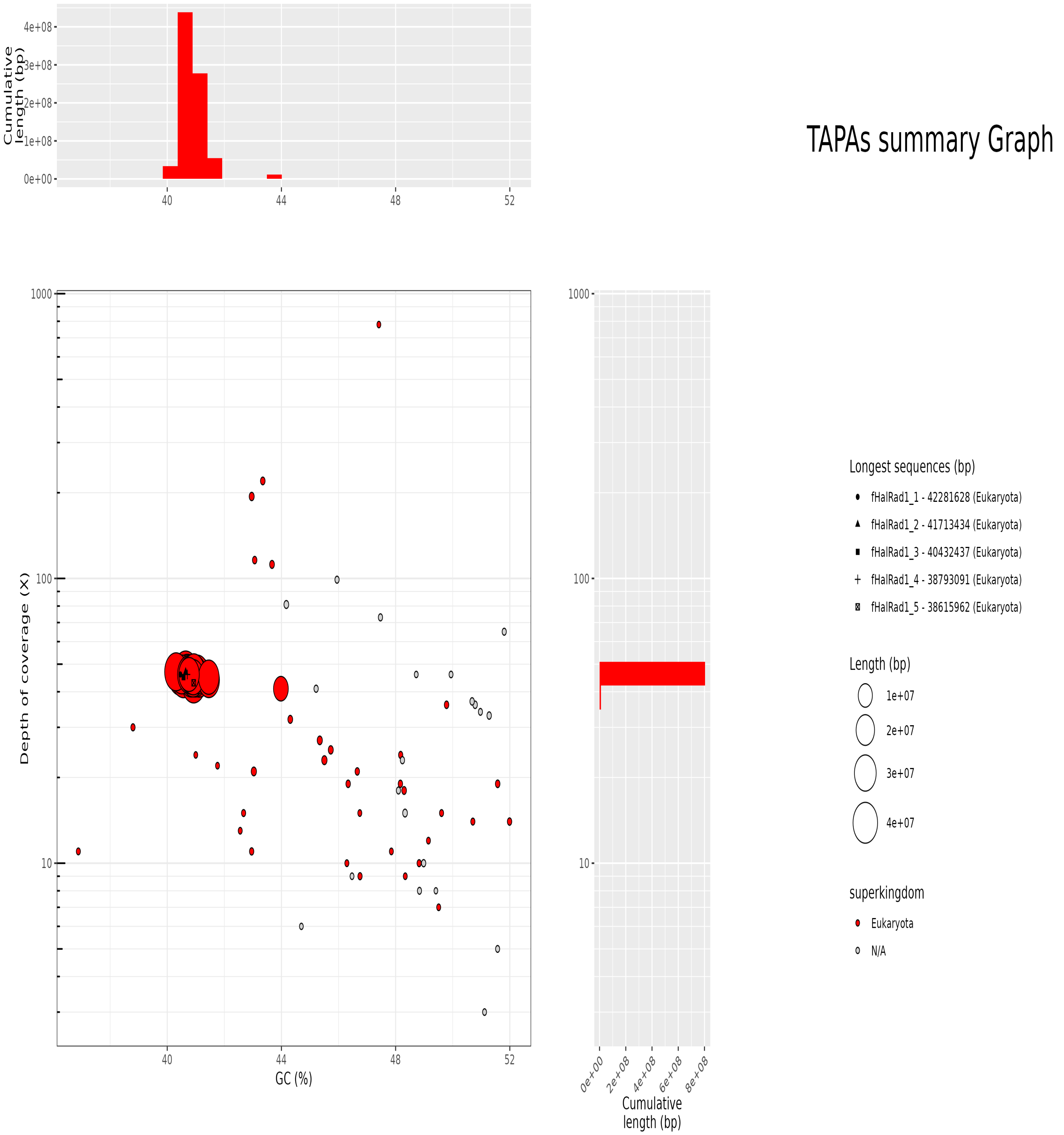


Distribution of k-mer counts per copy numbers found in asm



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	45	198

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: Benjamin Istace
Affiliation: Genoscope

Date and time: 2025-12-10 09:56:15 CET