ERGA Assembly Report

v24.10.15

Tags: ATLASea[INVALID TAG]

TxID	241291
ToLID	**fHalBiv1**
Species	Halichoeres bivittatus
Class	Actinopteri
Order	Labriformes

Genome Traits	Expected	Observed
Haploid size (bp)	837,210,768	843,002,783
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

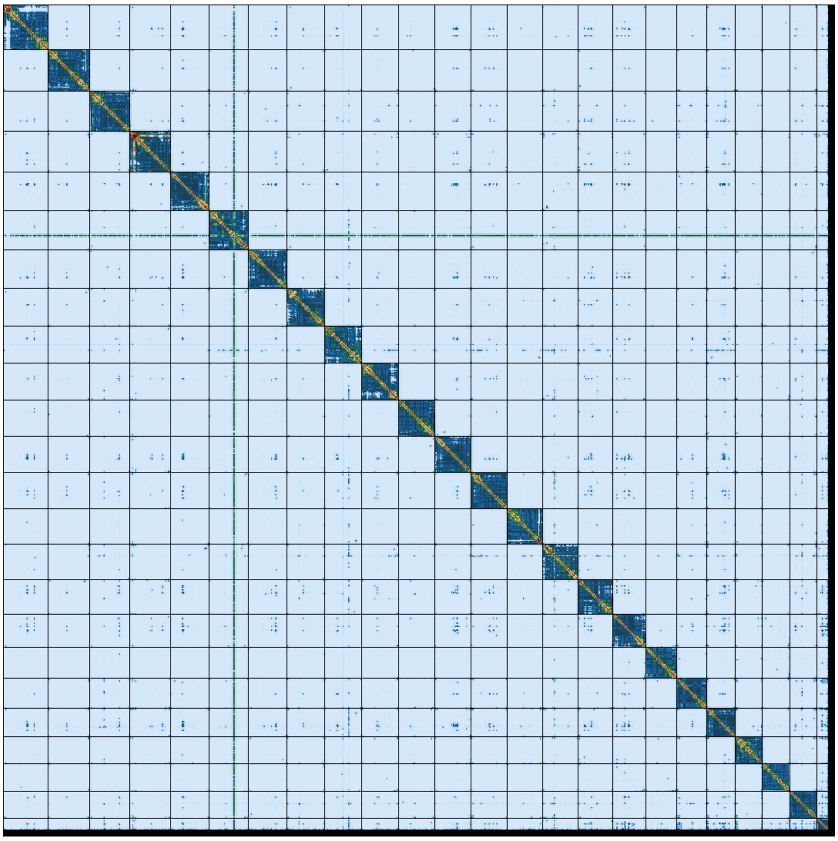
. Interventions/Gb: 65
. Contamination notes: ""
. Other observations: "The assembly of Halichoeres bivittatus (fHalBiv1) is based on 39X PacBio data and 152X Arima Hi-C data generated as part of the ATLASea programme (https://www.atlasea.fr). The assembly process included the following steps: initial 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 0.123 Mb (with the largest being 0.07 Mb). Additionally, 179 regions totaling 24.368 Mb (with the largest being 2.923 Mb) were identified as haplotypic duplications and removed. Finally, the primary assembly was analyzed and manually improved using Pretext.The mitochondrial genome was assembled using OATK. During manual curation, 1 haplotypic regions and 5 contaminant sequences were removed, totaling 0.187Mb and 0.588Mb, respectively (with the largest being 0.187Mb and 0.352Mb). 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	843,458,834	843,002,783
GC %	41.02	41.02
Gaps/Gbp	141.09	168.45
Total gap bp	11,900	18,000
Scaffolds	139	95
Scaffold N50	36,505,208	36,774,543
Scaffold L50	11	11
Scaffold L90	21	21
Contigs	258	237
Contig N50	22,906,117	22,906,117
Contig L50	15	15
Contig L90	40	40
QV	48.4124	48.4355
Kmer compl.	76.0268	76.0168
BUSCO sing.	99.0%	99.1%
BUSCO dupl.	0.3%	0.3%
BUSCO frag.	0.1%	0.0%
BUSCO miss.	0.6%	0.6%

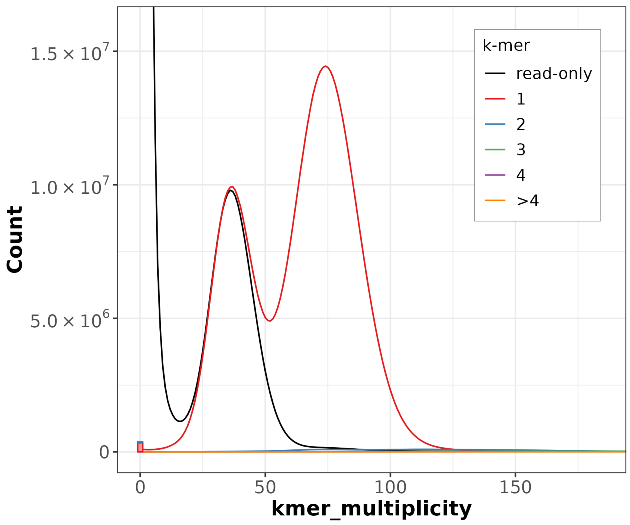
BUSCO: 6.0.0 (euk_genome_min, miniprot) / Lineage: actinopterygii_odb12 (genomes:75, BUSCOs:7207)

HiC contact map of curated assembly

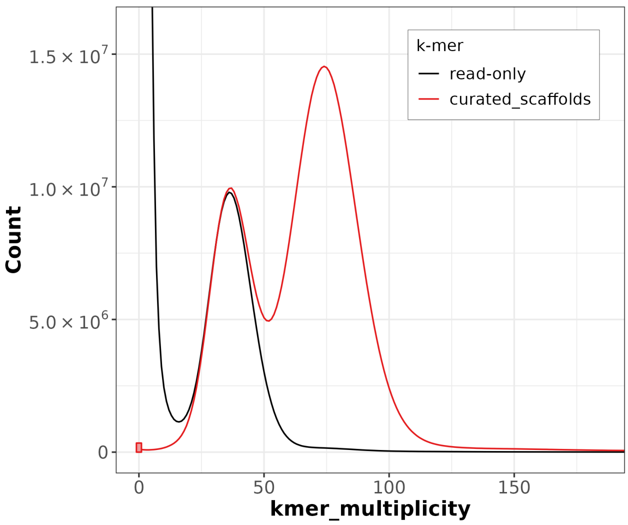


collapsed [LINK]

K-mer spectra of curated assembly

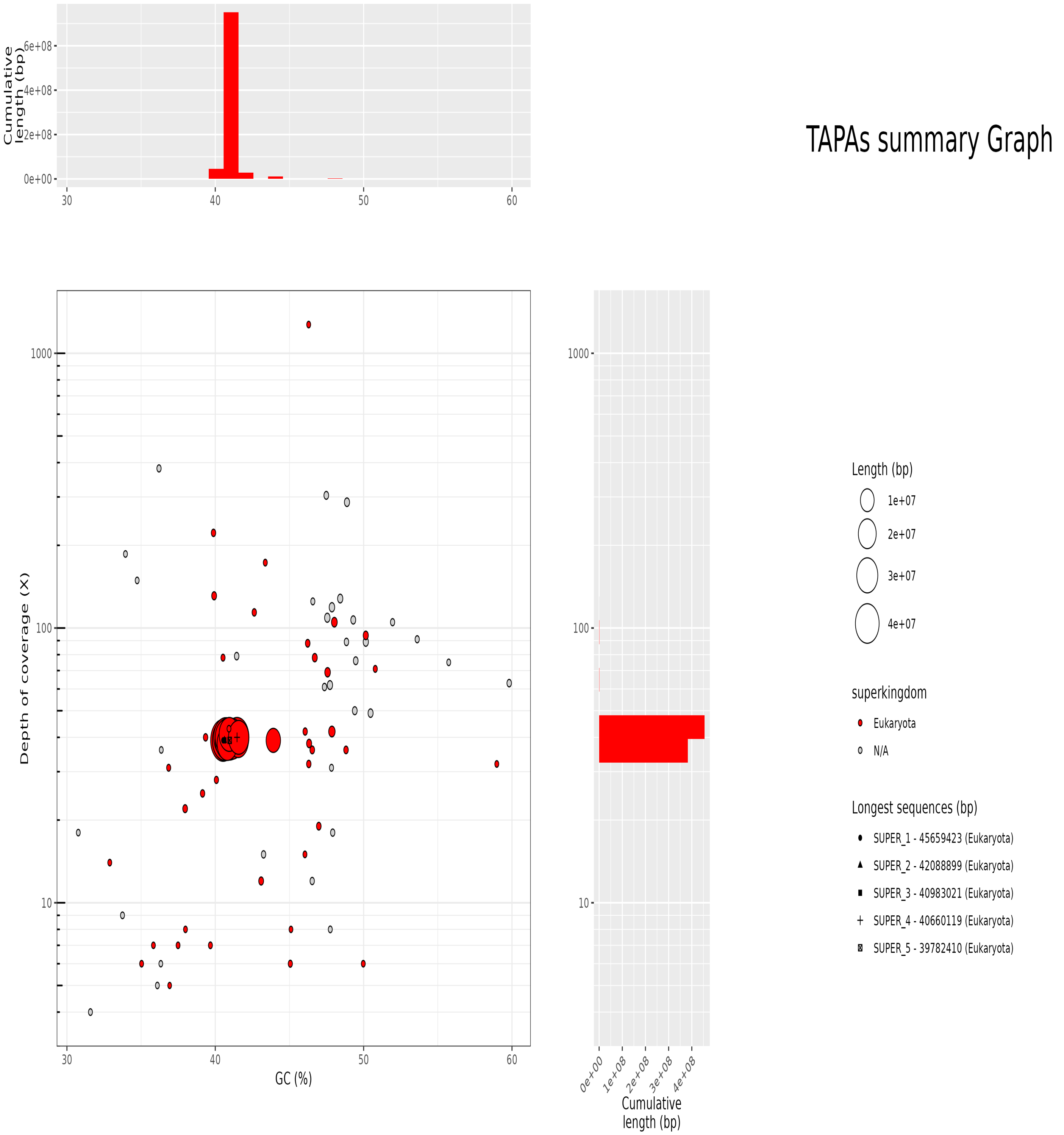


Distribution of k-mer counts per copy numbers found in asm



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	39	152

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: Sophie Layac
Affiliation: Genoscope

Date and time: 2025-12-10 12:03:00 CET