ERGA Assembly Report

v24.10.15

Tags: ATLASea[INVALID TAG]

TxID	202588
ToLID	**fLabVir1**
Species	Labrus viridis
Class	Actinopteri
Order	Labriformes

Genome Traits	Expected	Observed
Haploid size (bp)	696,713,904	708,878,304
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.Q46

The following metrics were automatically flagged as below EBP recommended standards or different from expected:

Curator notes

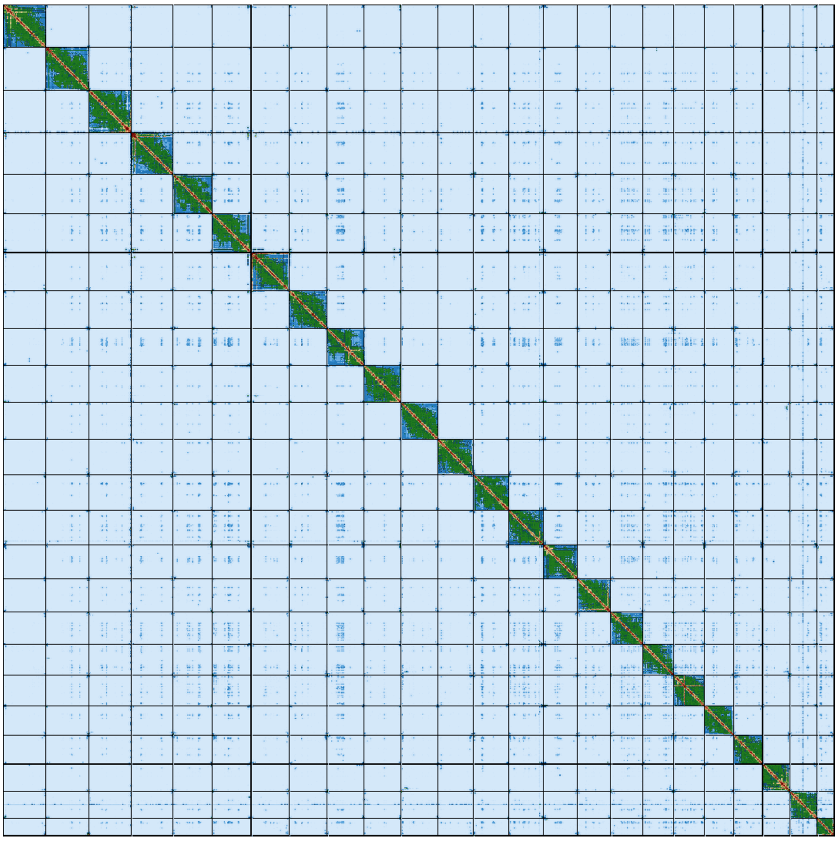
. Interventions/Gb: 59
. Contamination notes: ""
. Other observations: "The assembly of Labrus viridis (fLabVir1) is based on 43X PacBio data and 176X Arima Hi-C data generated as part of the ATLASea programme (https://www.atlasea.fr). The assembly process included the following steps: initial PacBio assembly generation with Hifiasm, removal of contaminant sequences using Context, removal of haplotypic duplications using purge_dups, and Hi-C-based scaffolding with YaHS. No contigs were found to be contaminants (bacterial, archaeal, or viral). Additionally, 152 regions totaling 11.578 Mb (with the largest being 1.02 Mb) were identified as haplotypic duplications and removed. Finally, the primary assembly was analyzed and manually improved using Pretext. During manual curation, 1 haplotypic region was removed, totaling 0.31Mb. 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	709,165,984	708,878,304
GC %	40.68	40.68
Gaps/Gbp	463.93	488.1
Total gap bp	32,900	37,400
Scaffolds	93	77
Scaffold N50	30,876,382	31,263,669
Scaffold L50	11	11
Scaffold L90	21	21
Contigs	422	423
Contig N50	17,732,158	17,732,158
Contig L50	15	15
Contig L90	51	51
QV	46.8116	46.8116
Kmer compl.	91.7514	91.7487
BUSCO sing.	99.1%	99.2%
BUSCO dupl.	0.2%	0.2%
BUSCO frag.	0.2%	0.2%
BUSCO miss.	0.5%	0.5%

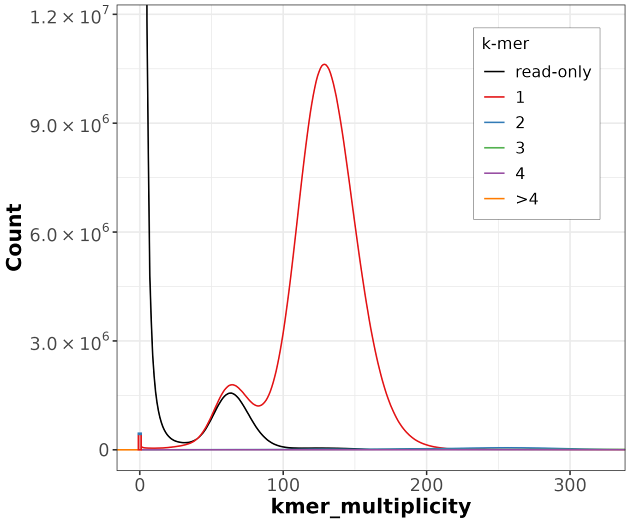
BUSCO: 6.0.0 (euk_genome_min, miniprot) / Lineage: actinopterygii_odb12 (genomes:75, BUSCOs:7207)

HiC contact map of curated assembly

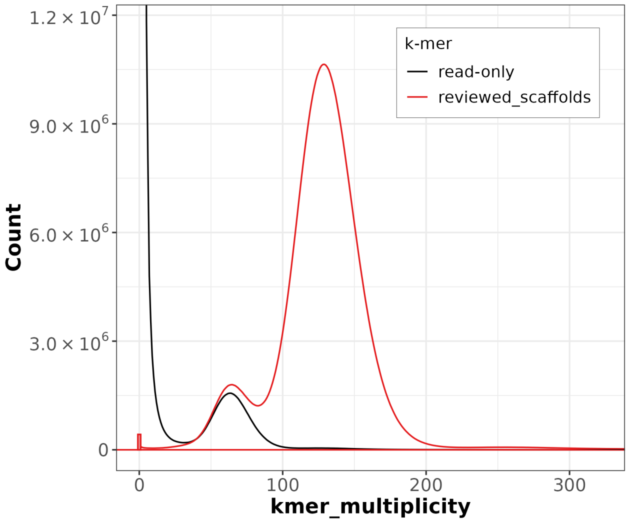


collapsed [LINK]

K-mer spectra of curated assembly

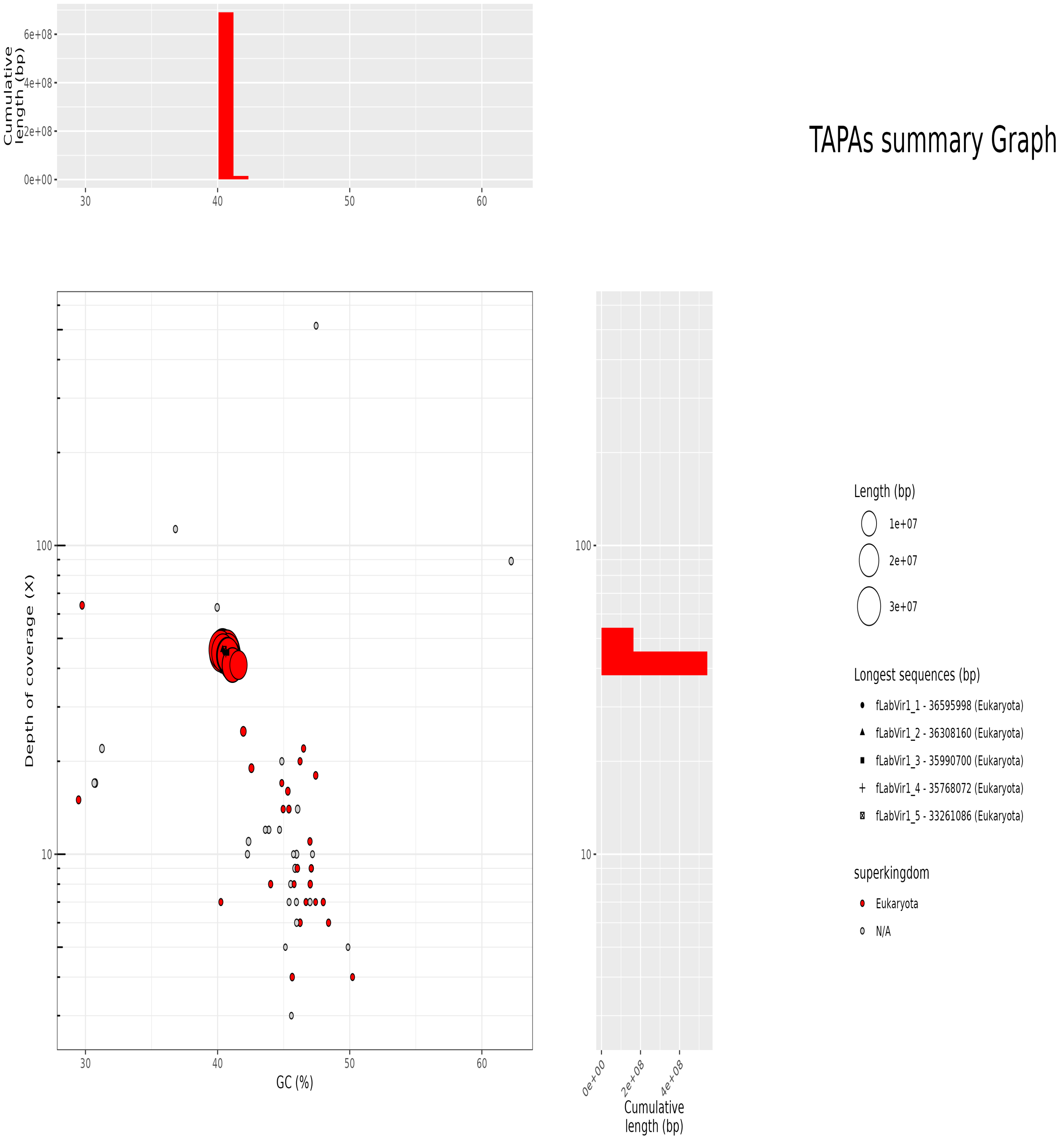


Distribution of k-mer counts per copy numbers found in asm



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	43	176

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: Jean-Marc Aury
Affiliation: Genoscope

Date and time: 2026-02-18 12:14:37 CET