ERGA Assembly Report

v24.10.15

Tags: ATLASea[INVALID TAG]

TxID	171735
ToLID	**fCteRup1**
Species	Ctenolabrus rupestris
Class	Actinopteri
Order	Labriformes

Genome Traits	Expected	Observed
Haploid size (bp)	735,565,183	767,190,745
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: 6.7.Q46

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

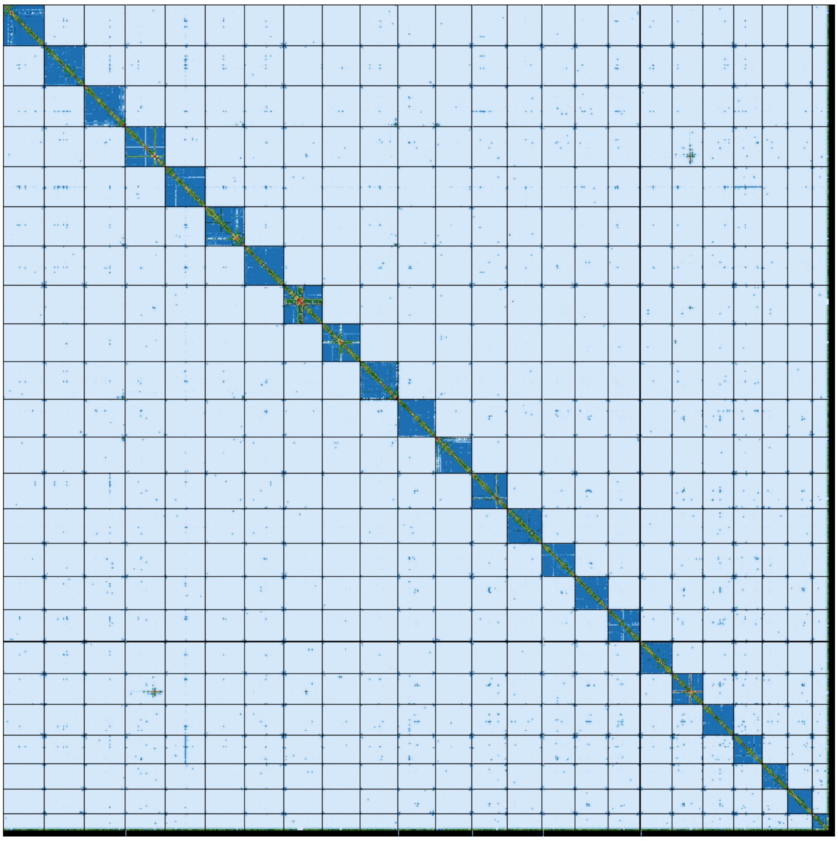
. Interventions/Gb: 114
. Contamination notes: ""
. Other observations: "The assembly of Ctenolabrus rupestris (fCteRup1) is based on 42X PacBio data and 214X Arima Hi-C data generated as part of the ATLASea programme (https://www.atlasea.fr). The assembly process included the following steps: initial 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, 281 regions totaling 16.662 Mb (with the largest being 0.655 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, 12 haplotypic regions and 1 contaminant sequences were removed, totaling 2.99 Mb and 0.015 Mb, respectively (with the largest being 0.748 Mb and 0.015 Mb). Chromosome-scale scaffolds confirmed by Hi-C data were named in order of size. 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	770,321,090	767,190,745
GC %	41.64	41.64
Gaps/Gbp	639.99	688.23
Total gap bp	49,300	58,800
Scaffolds	234	147
Scaffold N50	32,210,603	34,173,648
Scaffold L50	11	11
Scaffold L90	23	21
Contigs	727	675
Contig N50	8,814,000	9,703,000
Contig L50	26	25
Contig L90	115	111
QV	46.2819	46.3028
Kmer compl.	78.507	78.4311
BUSCO sing.	96.6%	97.0%
BUSCO dupl.	0.6%	0.3%
BUSCO frag.	0.7%	0.7%
BUSCO miss.	2.1%	2.1%

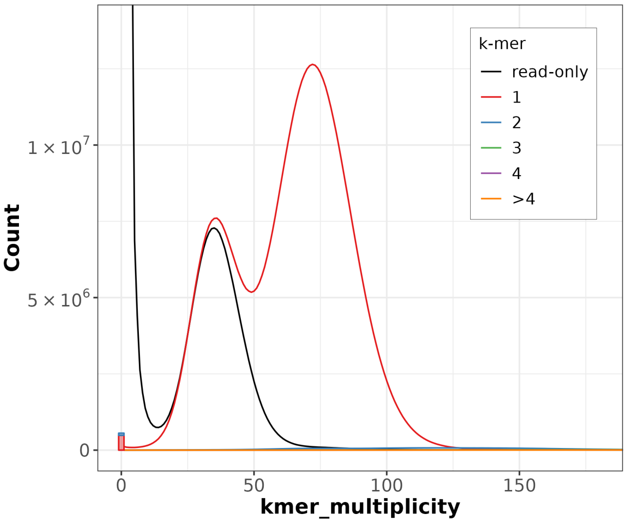
BUSCO: 5.8.2 (euk_genome_met, metaeuk) / Lineage: actinopterygii_odb12 (genomes:75, BUSCOs:7207)

HiC contact map of curated assembly

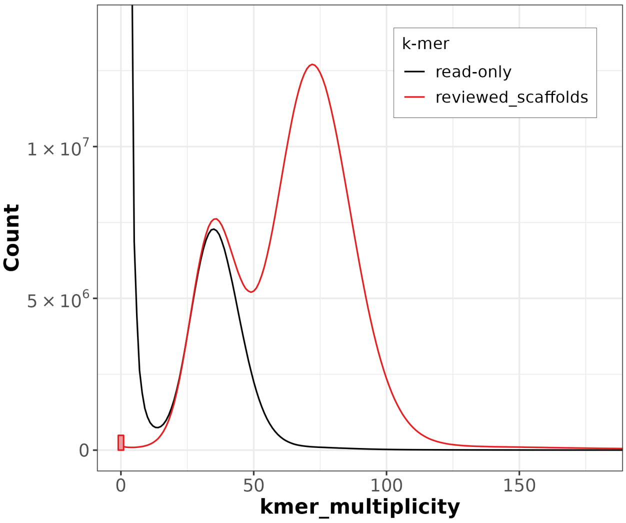


collapsed [LINK]

K-mer spectra of curated assembly

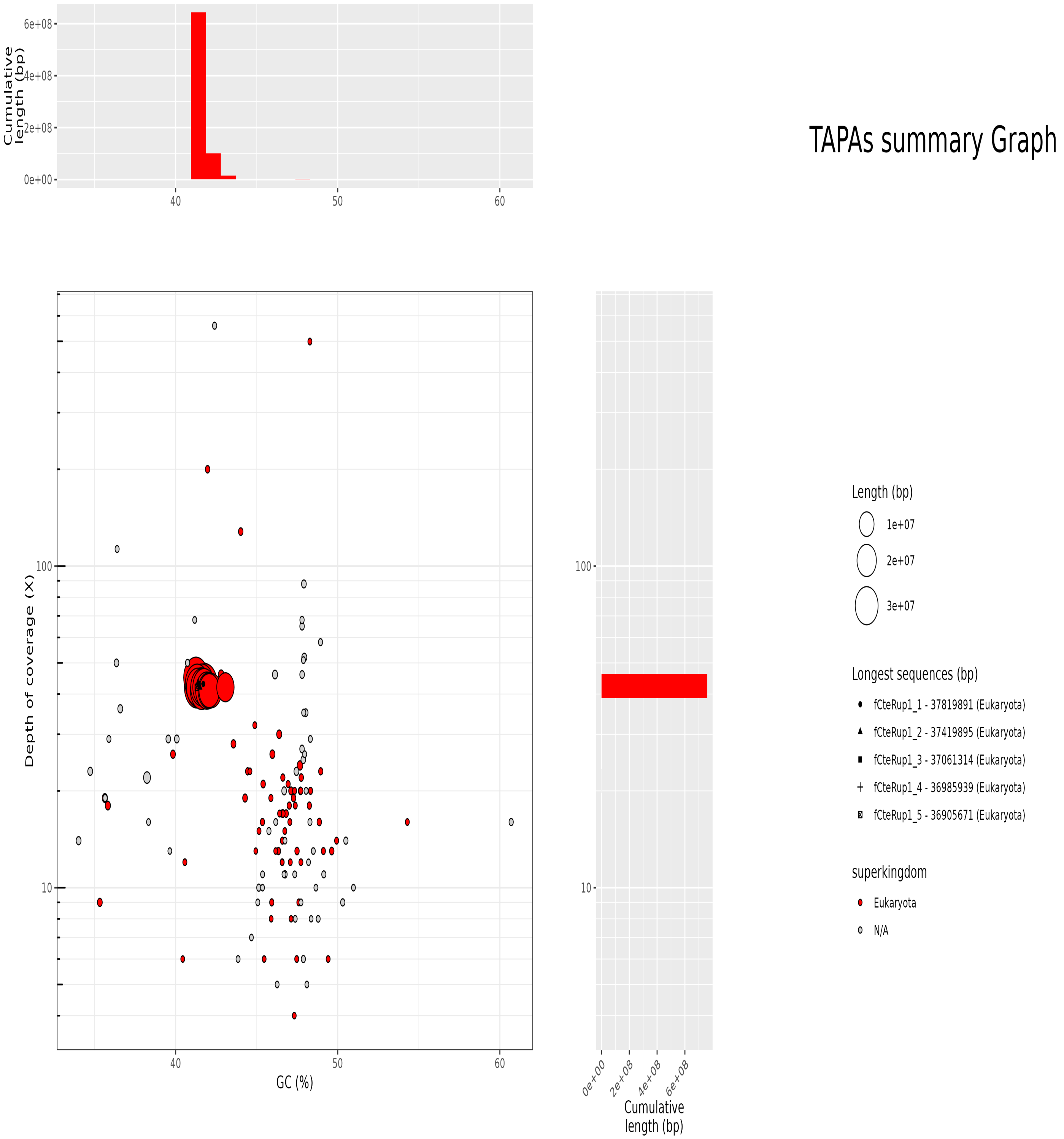


Distribution of k-mer counts per copy numbers found in asm



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	PACBIO Hifi	Arima
Coverage	42	213

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: Simone Duprat
Affiliation: Genoscope

Date and time: 2025-05-29 16:11:30 CEST