ERGA Assembly Report

v24.10.15

Tags: ATLASea[INVALID TAG]

TxID	13489
ToLID	**fDicLab2**
Species	Dicentrarchus labrax
Class	Actinopteri
Order	NA

Genome Traits	Expected	Observed
Haploid size (bp)	692,710,310	699,158,478
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:

Curator notes

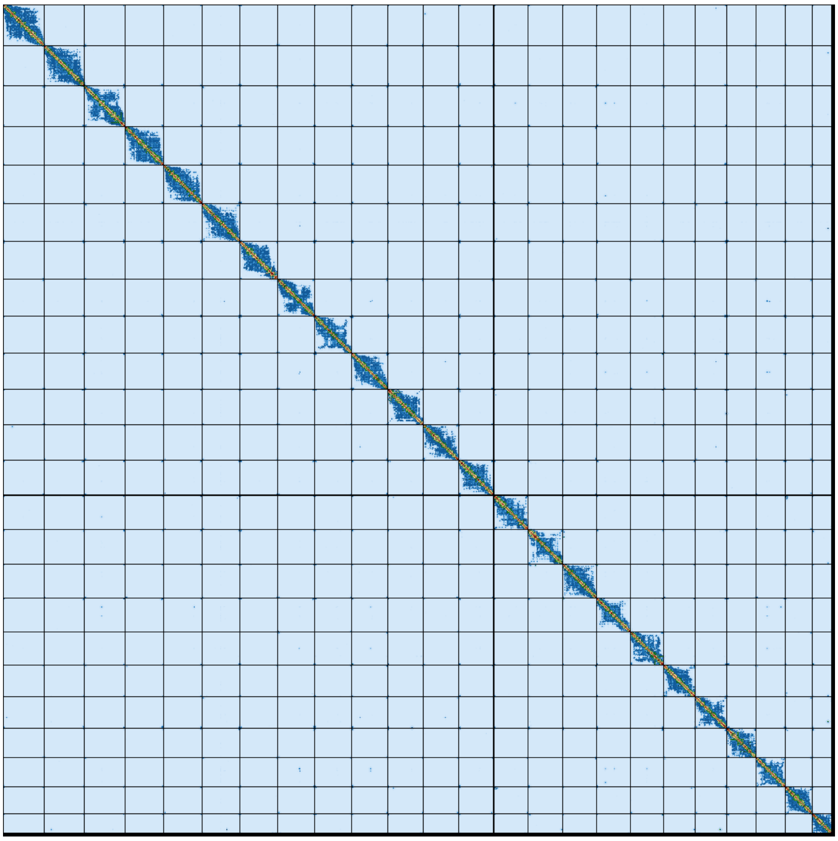
. Interventions/Gb: 4
. Contamination notes: ""
. Other observations: "The assembly of Dicentrarchus labrax (fDicLab2) is based on 58X PacBio data and 164X Arima Hi-C data generated as part of the ATLASea programme (https://www.atlasea.fr). The assembly process included the following steps: initial PacBio 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, 6 contigs were identified as contaminants (bacterial, archaeal, or viral), totaling 0.477 Mb (with the largest being 0.196 Mb). Additionally, 77 regions totaling 4.917 Mb (with the largest being 0.286 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. 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	699,136,713	699,158,478
GC %	40.45	40.45
Gaps/Gbp	12.87	14.3
Total gap bp	900	1,200
Scaffolds	59	59
Scaffold N50	30,046,593	30,046,593
Scaffold L50	11	11
Scaffold L90	21	21
Contigs	68	69
Contig N50	28,475,913	28,475,913
Contig L50	12	12
Contig L90	22	22
QV	48.9414	48.9415
Kmer compl.	91.0977	91.1003
BUSCO sing.	99.6%	99.6%
BUSCO dupl.	0.2%	0.2%
BUSCO frag.	0.0%	0.0%
BUSCO miss.	0.2%	0.2%

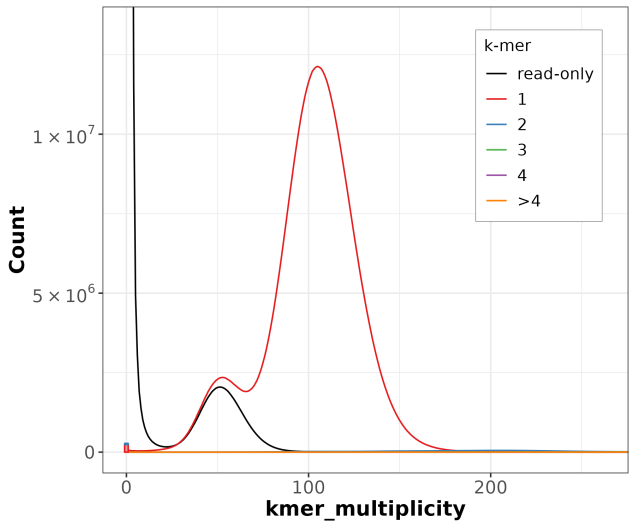
BUSCO: 6.0.0 (euk_genome_min, miniprot) / Lineage: actinopterygii_odb12 (genomes:75, BUSCOs:7207)

HiC contact map of curated assembly

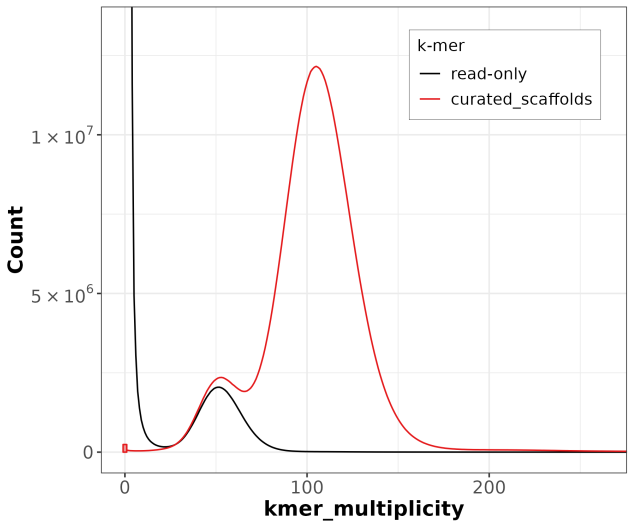


collapsed [LINK]

K-mer spectra of curated assembly

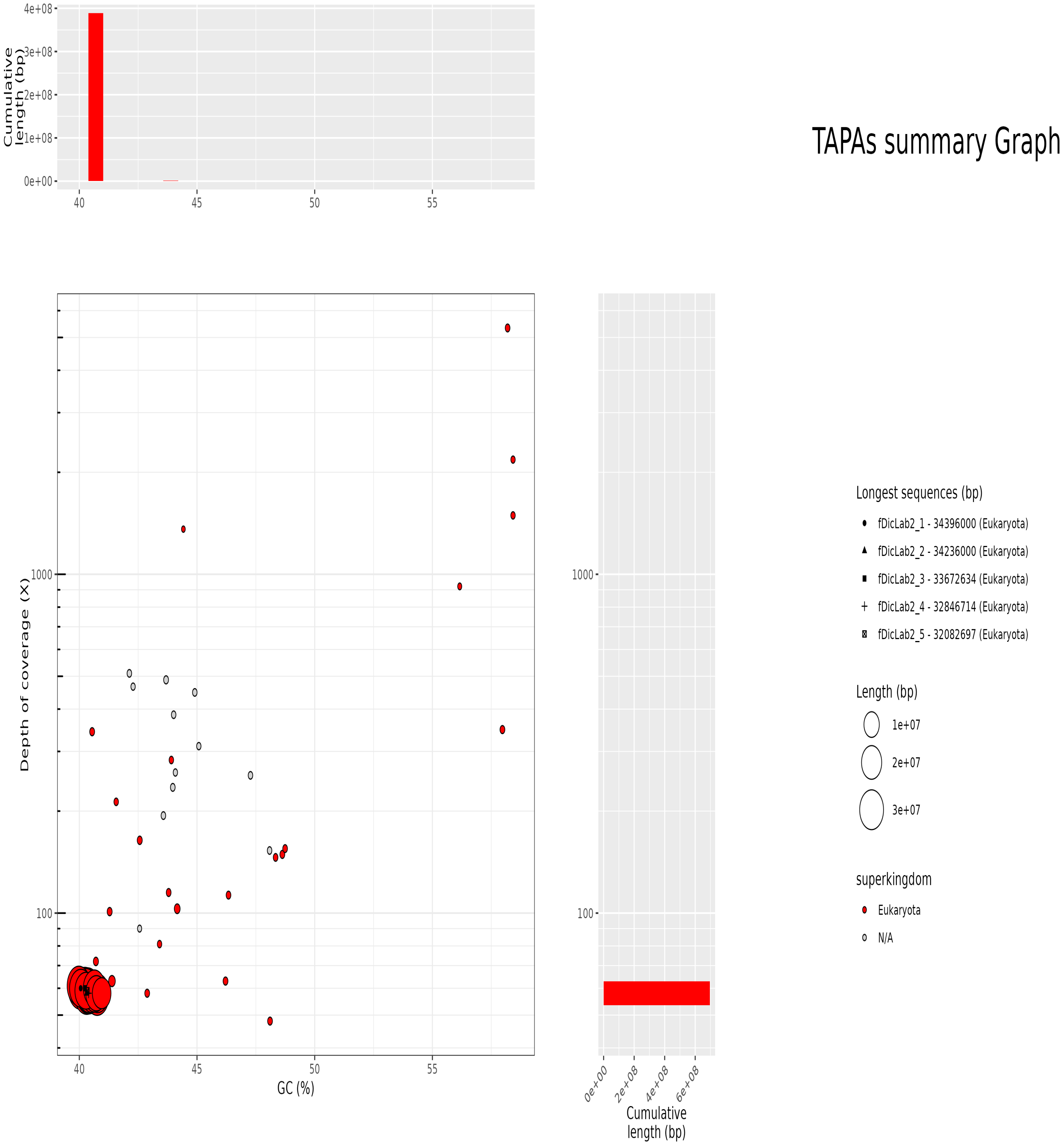


Distribution of k-mer counts per copy numbers found in asm



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	58	164

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: Arnaud Couloux
Affiliation: Genoscope

Date and time: 2025-11-03 11:14:45 CET