ERGA Assembly Report

v24.10.15

Tags: ATLASea[INVALID TAG]

TxID	87756
ToLID	**fDipVul1**
Species	Diplodus vulgaris
Class	Actinopteri
Order	Spariformes

Genome Traits	Expected	Observed
Haploid size (bp)	820,641,738	825,880,091
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

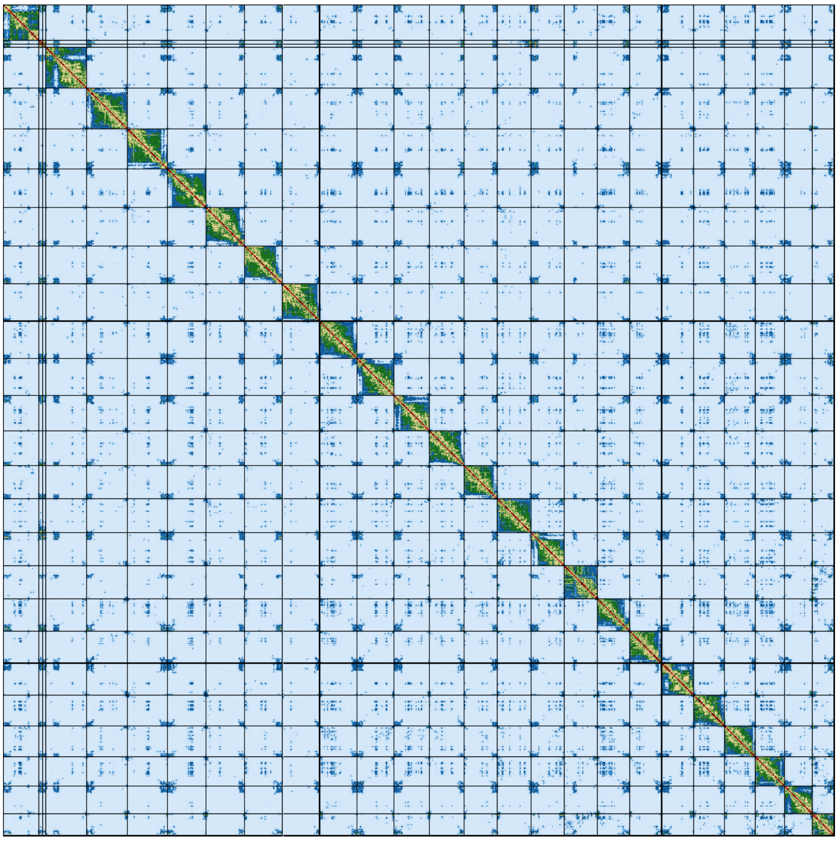
. Interventions/Gb: 10
. Contamination notes: ""
. Other observations: "The assembly of Diplodus vulgaris (fDipVul1) is based on 52X PacBio data and 185X Arima Hi-C data generated as part of the ATLASea programme (https://www.atlasea.fr).The assembly process included the following steps: initial PacBio 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, no contigs were identified as contaminants (bacterial, archaeal, or viral). Additionally, 40 regions totaling 13.2 Mb (with the largest being 3.7 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, no supplementary haplotypic regions were removed. 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	825,862,545	825,880,091
GC %	42.1	42.1
Gaps/Gbp	50.86	52.07
Total gap bp	4,200	5,100
Scaffolds	41	41
Scaffold N50	34,807,499	35,011,911
Scaffold L50	11	11
Scaffold L90	22	22
Contigs	83	84
Contig N50	27,800,000	27,800,000
Contig L50	13	13
Contig L90	31	31
QV	48.616	48.6161
Kmer compl.	79.142	79.1439
BUSCO sing.	99.2%	99.1%
BUSCO dupl.	0.2%	0.2%
BUSCO frag.	0.0%	0.0%
BUSCO miss.	0.6%	0.6%

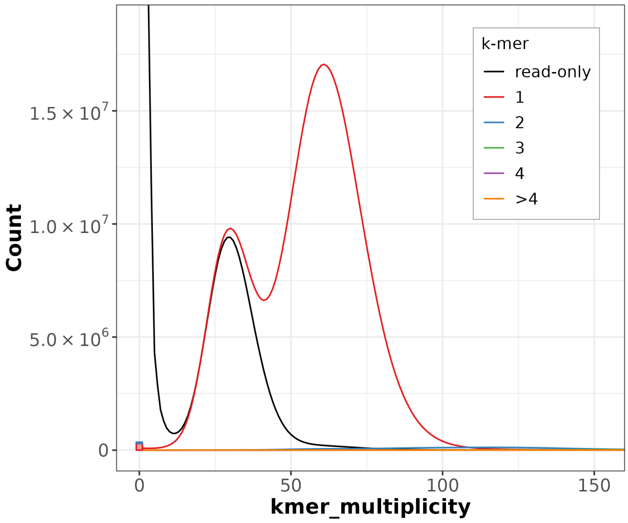
BUSCO: 6.0.0 (euk_genome_min, miniprot) / Lineage: actinopterygii_odb12 (genomes:75, BUSCOs:7207)

HiC contact map of curated assembly

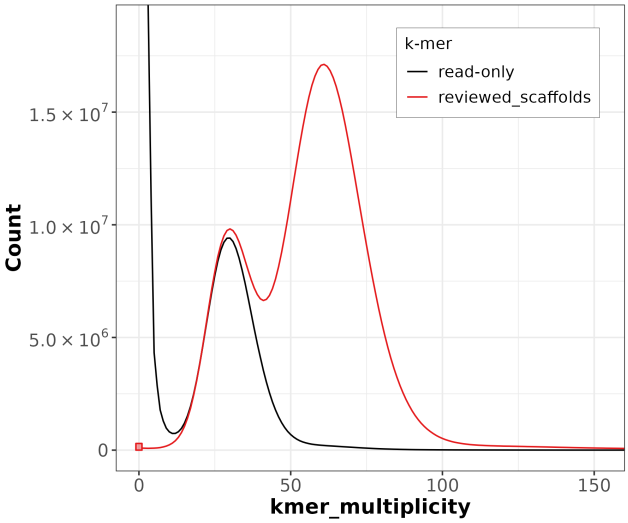


collapsed [LINK]

K-mer spectra of curated assembly

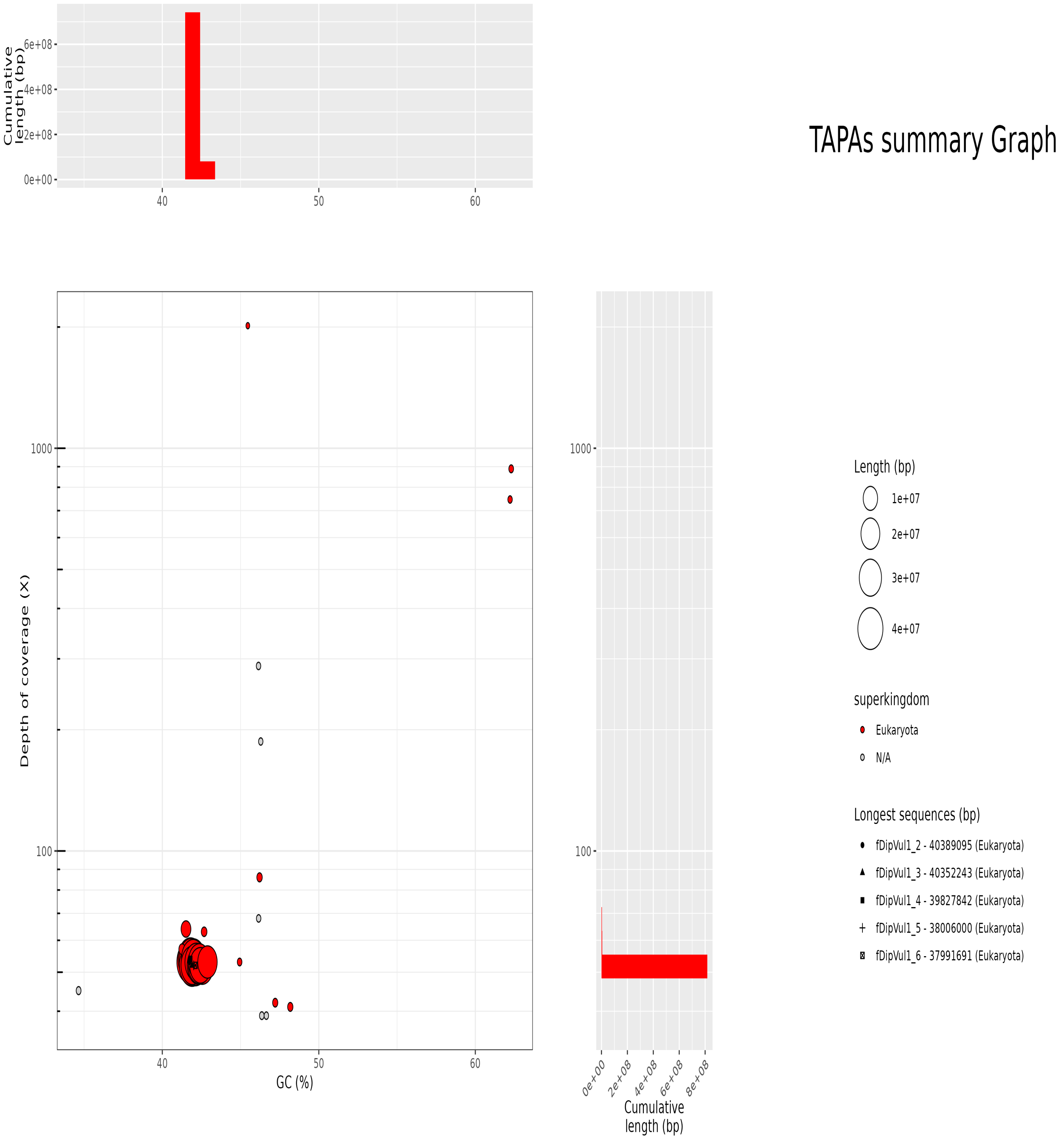


Distribution of k-mer counts per copy numbers found in asm



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	52	185

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: Mohamed Meftah Saoues
Affiliation: Genoscope

Date and time: 2026-01-13 09:44:06 CET