ERGA Assembly Report

v24.10.15

Tags: ATLASea[INVALID TAG]

TxID	48913
ToLID	**fDipAnn1**
Species	Diplodus annularis
Class	Actinopteri
Order	Spariformes

Genome Traits	Expected	Observed
Haploid size (bp)	768,003,274	780,840,897
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.Q49

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

Curator notes

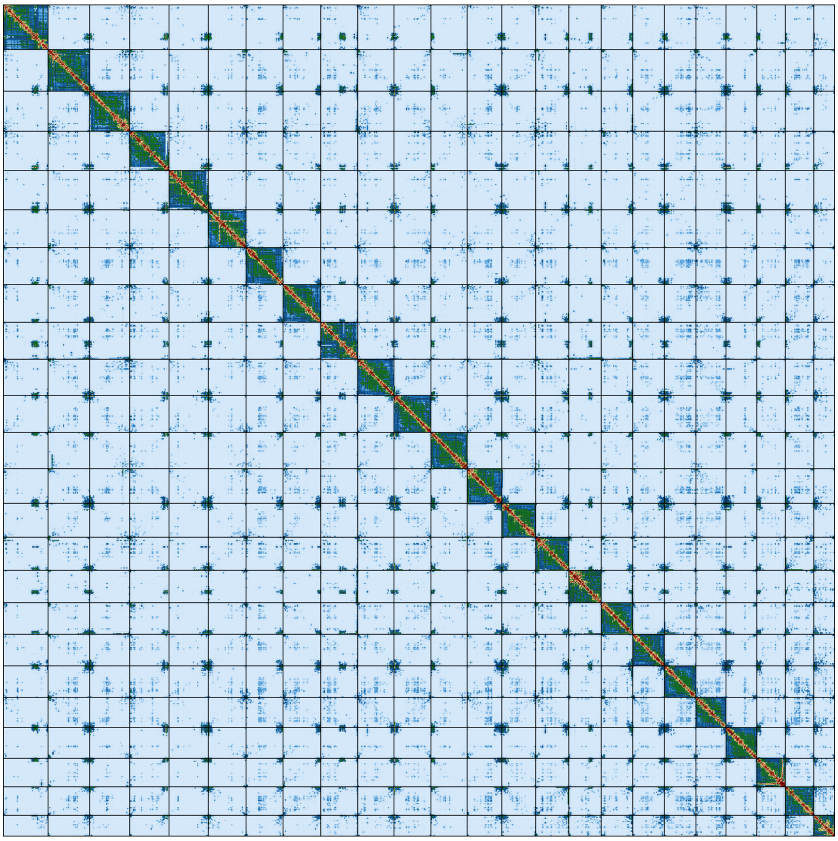
. Interventions/Gb: 12
. Contamination notes: ""
. Other observations: "The assembly of Diplodus annularis (fDipAnn1) is based on 62X PacBio data and 199X Arima Hi-C data generated as part of the ATLASea programme (https://www.atlasea.fr). The assembly process included the following steps: initial PacBio 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, 5 contigs were identified as contaminants (bacterial, archaeal, or viral), totaling 0.132 Mb (with the largest being 0.037 Mb). Additionally, 32 regions totaling 3.7 Mb (with the largest being 1.66 Mb) were identified as haplotypic duplications and removed. During manual curation, 1 haplotypic region was removed, totaling 0.546Mb. 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	781,387,419	780,840,897
GC %	42.57	42.57
Gaps/Gbp	29.43	29.46
Total gap bp	2,300	2,700
Scaffolds	27	27
Scaffold N50	34,512,857	34,512,857
Scaffold L50	11	11
Scaffold L90	21	21
Contigs	50	50
Contig N50	33,175,629	33,175,629
Contig L50	11	11
Contig L90	22	22
QV	49.0448	49.0442
Kmer compl.	83.4019	83.3784
BUSCO sing.	99.1%	
BUSCO dupl.	0.4%	
BUSCO frag.	0.0%	
BUSCO miss.	0.4%	

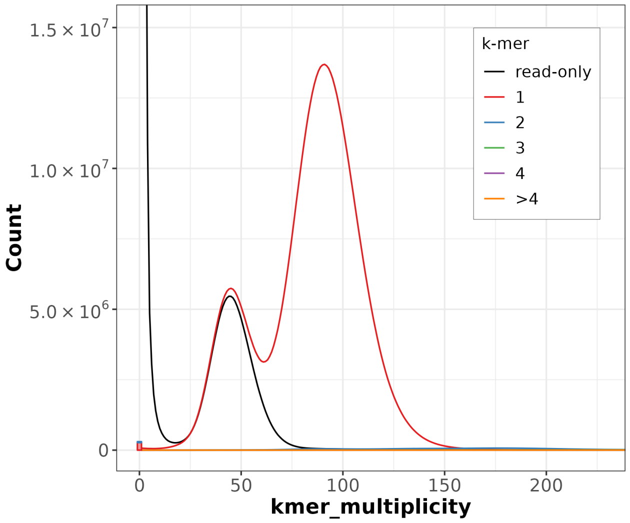
BUSCO: 6.0.0 (euk_genome_min, miniprot) / Lineage: actinopterygii_odb12 (genomes:75, BUSCOs:7207)

HiC contact map of curated assembly

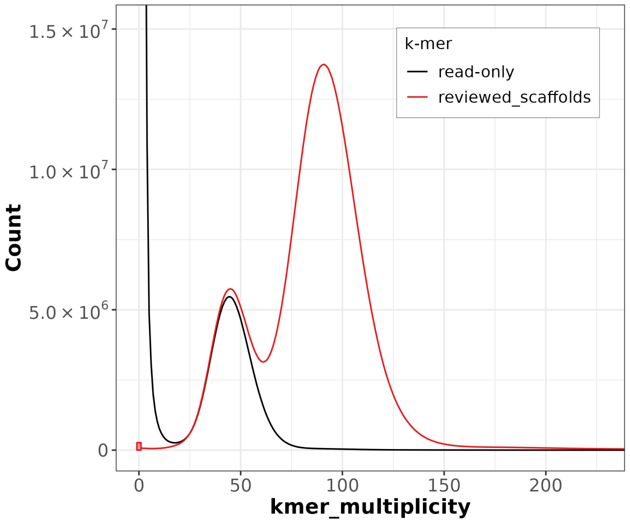


collapsed [LINK]

K-mer spectra of curated assembly

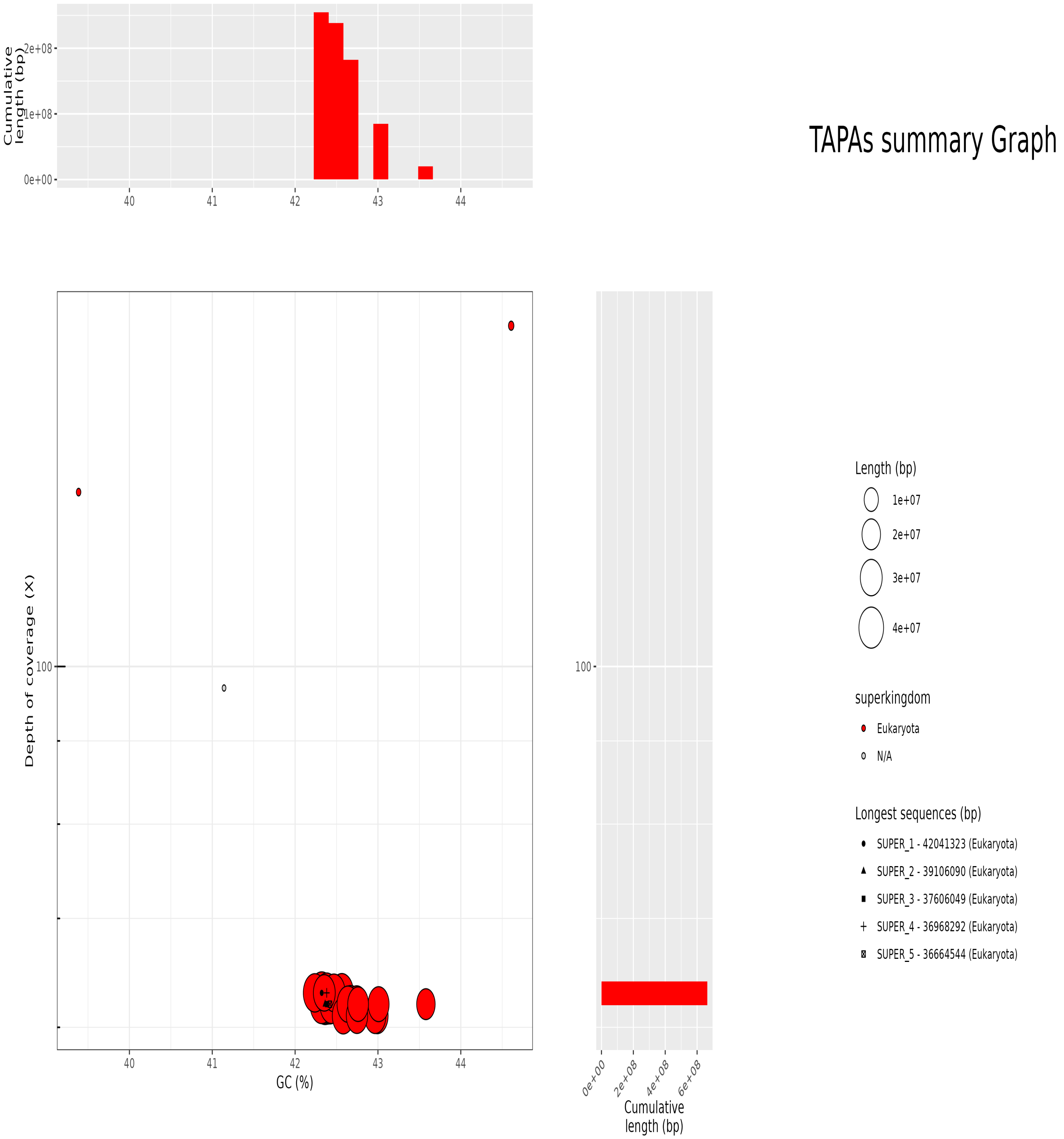


Distribution of k-mer counts per copy numbers found in asm



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	62	199

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: 2025-11-28 06:10:08 CET