ERGA Assembly Report

v24.10.15

Tags: ATLASea[INVALID TAG]

TxID	38941
ToLID	**fDipSar1**
Species	Diplodus sargus
Class	Actinopteri
Order	Spariformes

Genome Traits	Expected	Observed
Haploid size (bp)	1,502,489,855	807,045,846
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.Q46

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

. Observed Haploid size (bp) has >20% difference with Expected

. Kmer completeness value is less than 90 for collapsed

Curator notes

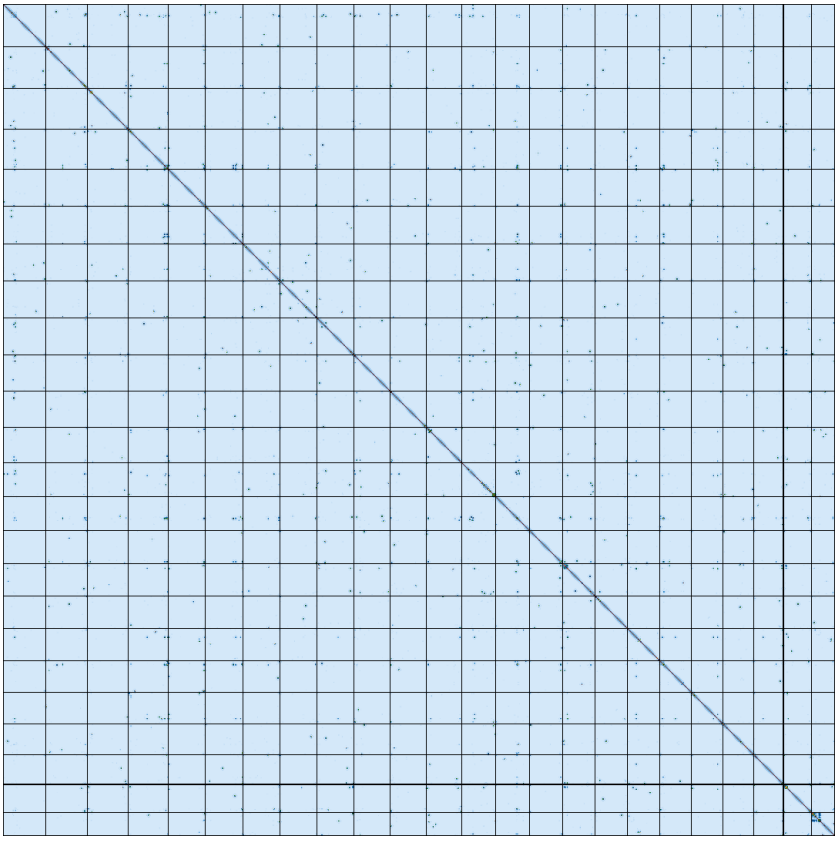
. Interventions/Gb: 26
. Contamination notes: ""
. Other observations: "The assembly of \'Diplodus sargus\' (fDipSar1) is based on 92X PacBio data and 113X Arima Hi-C data generated as part of the ATLASea programme (https://www.atlasea.fr). The assembly process included the following steps: initial PacBio 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, 47 regions totaling 2.426 Mb (with the largest being 0.276 Mb) were identified as haplotypic duplications and removed. Mitochondrial genome was assembled using OATK. Finally, the primary assembly was analyzed and manually improved using Pretext. During manual curation, 7 haplotypic regions were removed, totaling 1.99Mb (with the largest being 1.154Mb). 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	809,049,556	807,045,846
GC %	42.19	42.19
Gaps/Gbp	49.44	63.19
Total gap bp	4,000	6,600
Scaffolds	50	34
Scaffold N50	35,070,228	35,070,228
Scaffold L50	11	11
Scaffold L90	21	21
Contigs	90	85
Contig N50	31,322,003	31,322,003
Contig L50	12	12
Contig L90	24	24
QV	46.4291	46.4388
Kmer compl.	45.134	45.119
BUSCO sing.	99.4%	98.4%
BUSCO dupl.	0.2%	0.2%
BUSCO frag.	0.0%	0.7%
BUSCO miss.	0.3%	0.7%

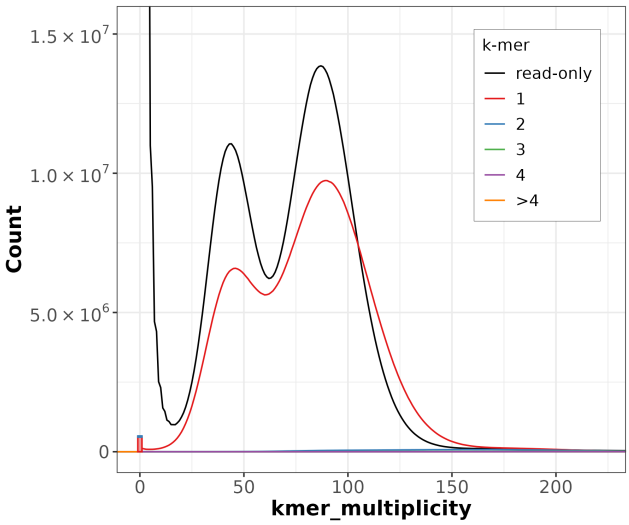
Warning! BUSCO versions or lineage datasets are not the same across results:
BUSCO: 6.0.0 (euk_genome_min, miniprot) / Lineage: actinopterygii_odb12 (genomes:75, BUSCOs:7207)
BUSCO: 5.8.2 (euk_genome_min, miniprot) / Lineage: actinopterygii_odb12 (genomes:75, BUSCOs:7207)

HiC contact map of curated assembly

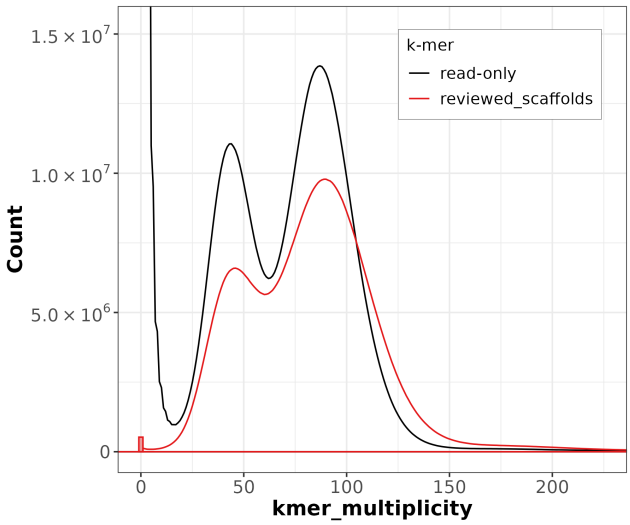


collapsed [LINK]

K-mer spectra of curated assembly

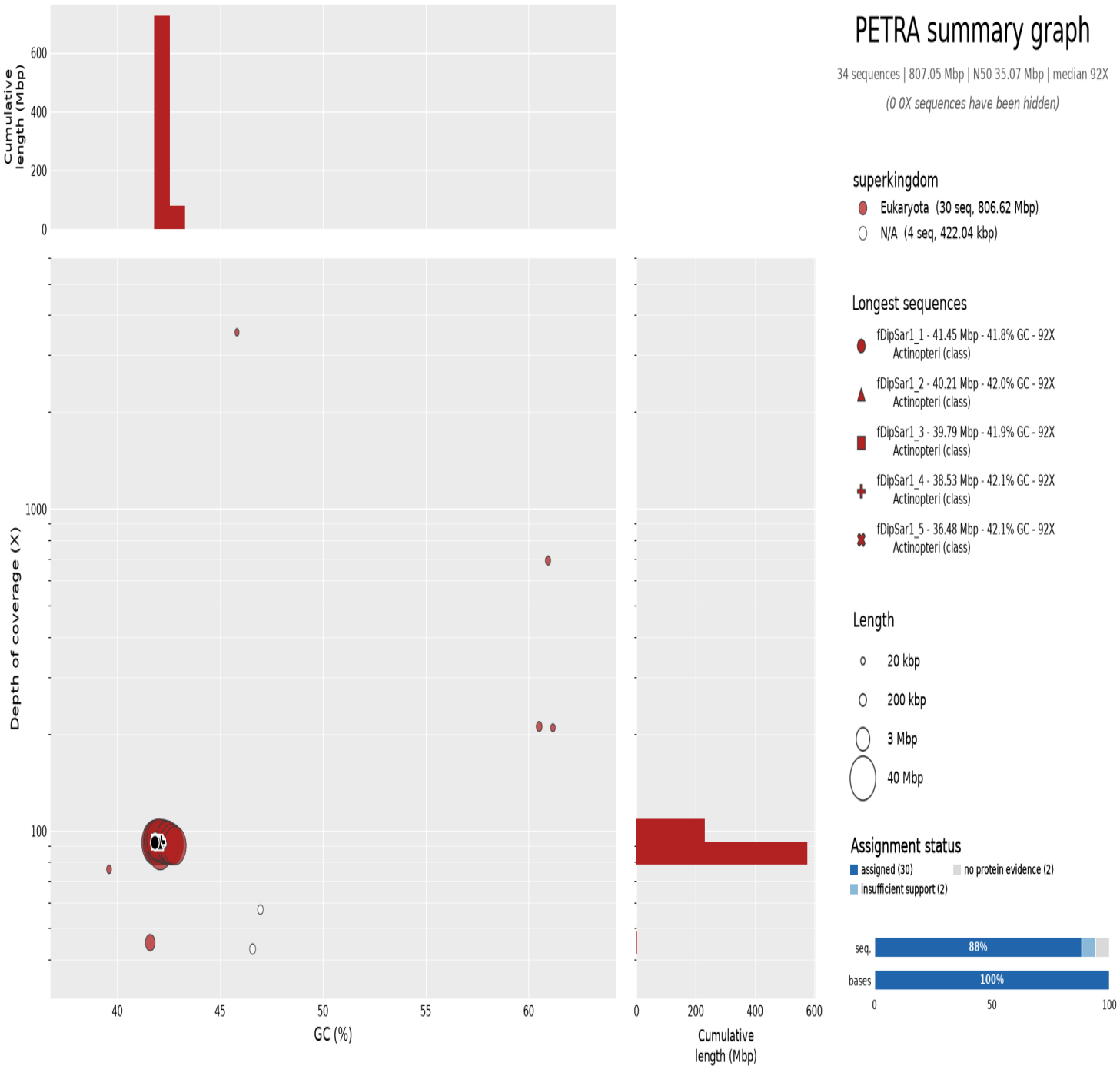


Distribution of k-mer counts per copy numbers found in asm



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	92	113

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:* 0.2.4
  |_ *key param:* NA
- **PretextView**
  |_ *ver:* 1.0.7
  |_ *key param:* NA

Submitter: Simone Duprat
Affiliation: Genoscope

Date and time: 2026-09-12 09:15:42 CEST