

ERGA Assembly Report

v24.10.15

Tags: ATLASEa[INVALID TAG]

TxID	94946
ToLID	fDicPun1
Species	Dicentrarchus punctatus
Class	Actinopteri
Order	NA

Genome Traits	Expected	Observed
Haploid size (bp)	679,711,098	688,301,903
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:

- . Kmer completeness value is less than 90 for collapsed

Curator notes

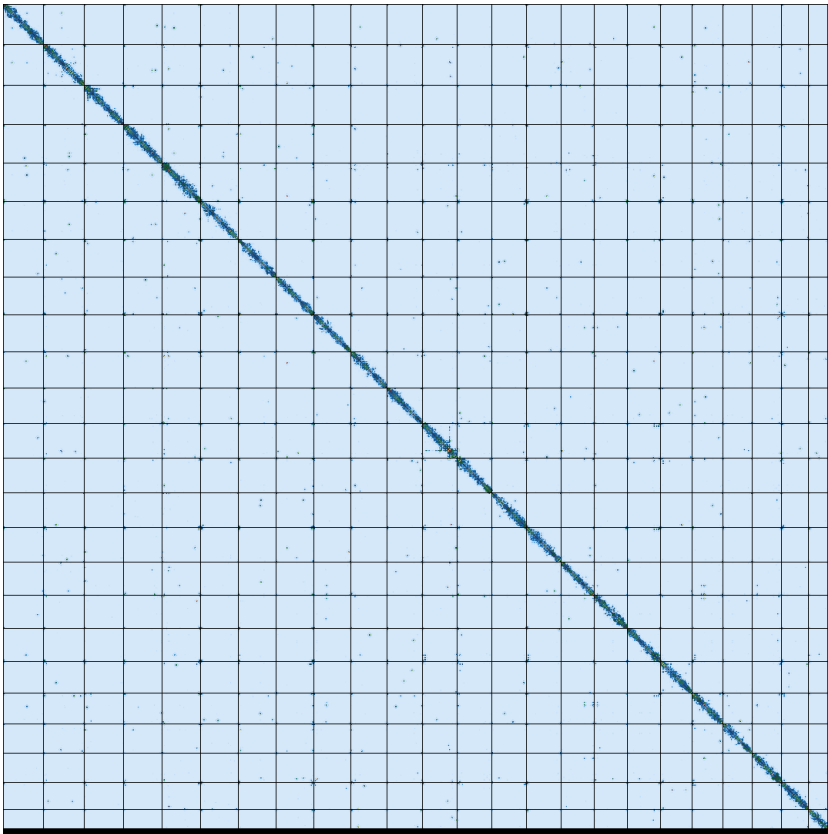
- . Interventions/Gb: 13
- . Contamination notes: ""
- . Other observations: "The assembly of *Dicentrarchus punctatus* (fDicPun1) is based on 28X PacBio data and 130X Arima Hi-C data generated as part of the ATLASEa programme (<https://www.atlasea.fr>). The assembly process included the following steps: initial PacBio assembly generation with Hifiiasm, 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, 92 regions totaling 6.64 Mb (with the largest being 1.178 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, 1 haplotypic region and 1 contaminant sequence were removed, totaling 0.152Mb and 0.047Mb, respectively (with the largest being 0.152Mb and 0.047Mb). The telomeric pattern TTAGGG was identified with TelFinder and used to generate the Pretext Telomeres tracks. 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	688,483,138	688,301,903
GC %	40.38	40.38
Gaps/Gbp	85.7	88.62
Total gap bp	5,900	6,500
Scaffolds	81	81
Scaffold N50	29,203,659	29,203,659
Scaffold L50	11	11
Scaffold L90	21	21
Contigs	140	142
Contig N50	24,124,559	24,124,559
Contig L50	13	13
Contig L90	31	31
QV	48.3013	48.3128
Kmer compl.	89.2942	89.289
BUSCO sing.	99.4%	99.4%
BUSCO dupl.	0.3%	0.3%
BUSCO frag.	0.0%	0.0%
BUSCO miss.	0.3%	0.3%

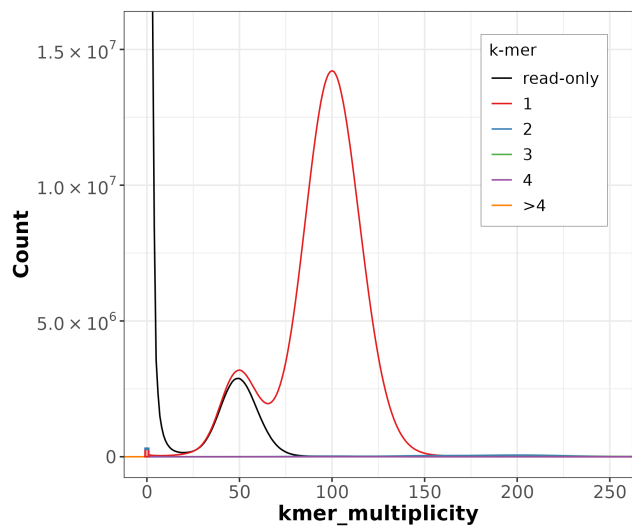
BUSCO: 6.0.0 (euk_genome_min, miniprot) / Lineage: actinopterygii_odb12 (genomes:75, BUSCOs:7207)

HiC contact map of curated assembly

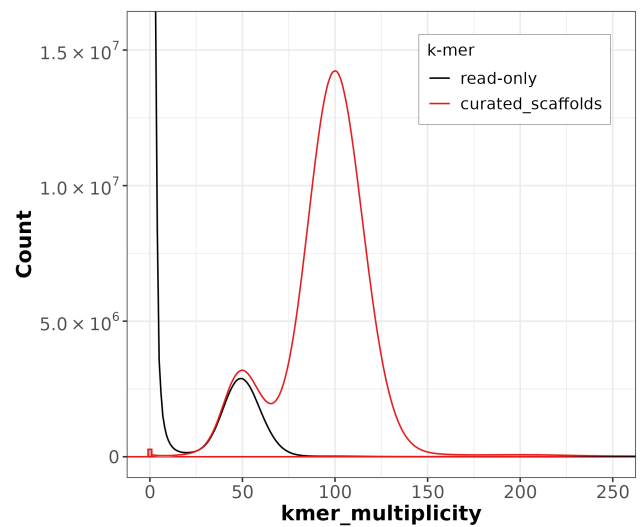


collapsed [\[LINK\]](#)

K-mer spectra of curated assembly

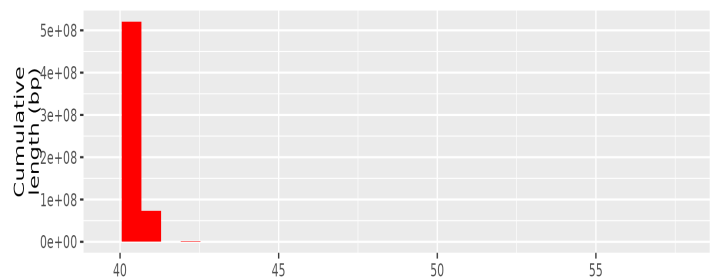


Distribution of k-mer counts per copy numbers found in asm

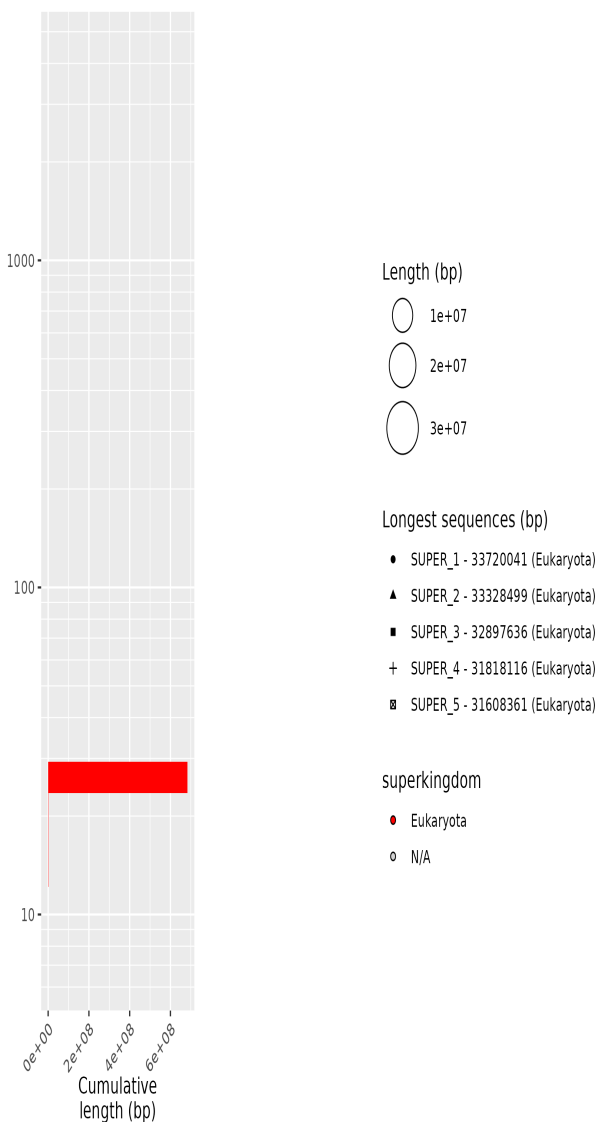
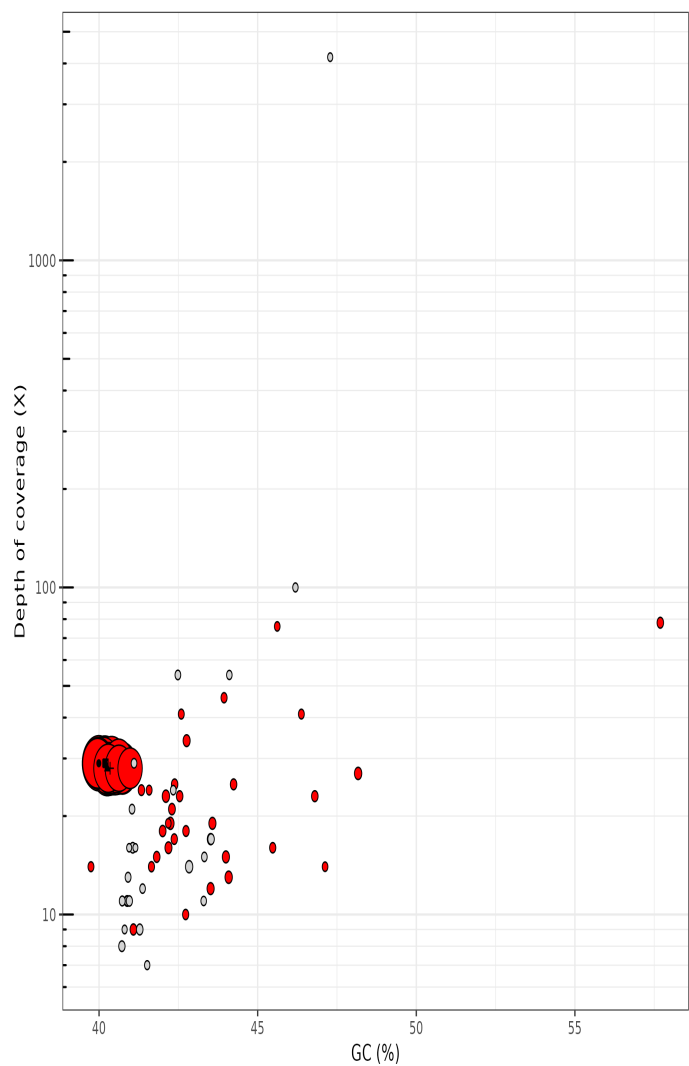


Distribution of k-mer counts coloured by their presence in reads/assemblies

Post-curation contamination screening



TAPAs summary Graph



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	28	130

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: 2026-07-17 12:19:19 CEST