ERGA Assembly Report

v24.10.15

Tags: ATLASea[INVALID TAG]

TxID	229062
ToLID	**fDioHol1**
Species	Diodon holocanthus
Class	Actinopteri
Order	Tetraodontiformes

Genome Traits	Expected	Observed
Haploid size (bp)	671,693,241	666,440,257
Haploid Number	21 (source: ancestor)	23
Ploidy	2 (source: ancestor)	2
Sample Sex	Unknown	Unknown

EBP metrics summary and curation notes

Obtained EBP quality metric for collapsed: 7.7.Q44

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

. Observed Haploid Number is different from Expected

. Kmer completeness value is less than 90 for collapsed

Curator notes

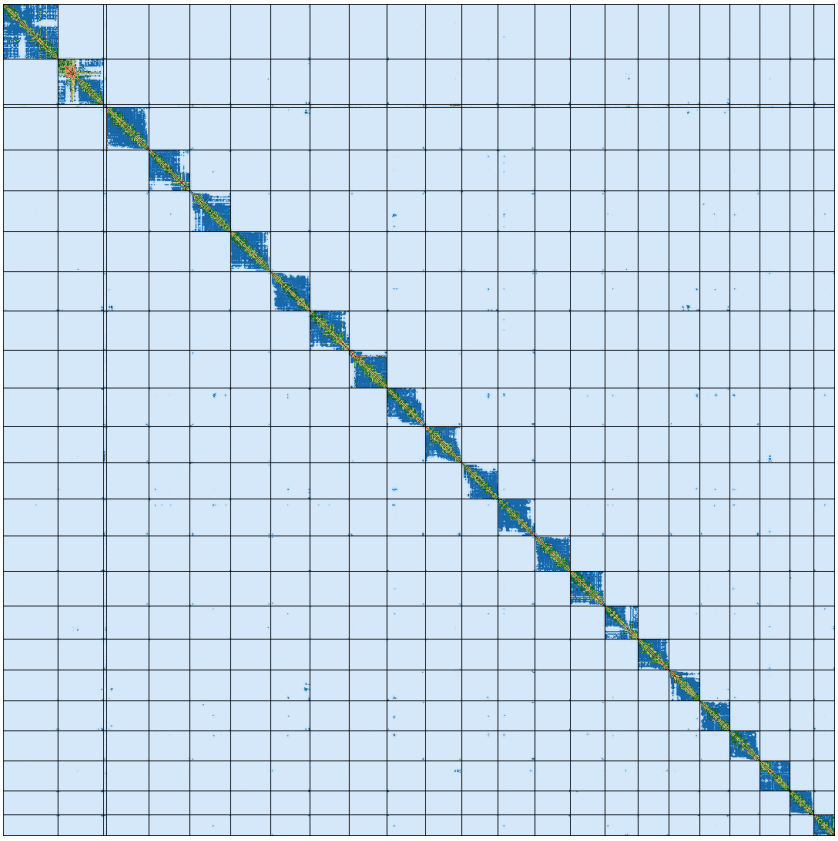
. Interventions/Gb: 134
. Contamination notes: ""
. Other observations: "The assembly of Diodon holocanthus (fDioHol1) is based on 73X PacBio data and 127X Arima Hi-C data generated as part of the ATLASea programme (https://www.atlasea.fr). The assembly process included the following steps: initial PacBio 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, 9 contigs were identified as contaminants (bacterial, archaeal, or viral), totaling 0.244 Mb (with the largest being 0.044 Mb). Additionally, 97 regions totaling 6.417 Mb (with the largest being 1.3 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. 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	666,407,144	666,440,257
GC %	41.78	41.78
Gaps/Gbp	25.51	148.55
Total gap bp	1,700	18,300
Scaffolds	103	25
Scaffold N50	29,103,778	30,560,808
Scaffold L50	11	10
Scaffold L90	21	20
Contigs	120	124
Contig N50	24,191,220	24,191,220
Contig L50	13	13
Contig L90	28	28
QV	44.7982	44.7983
Kmer compl.	85.9866	85.9866
BUSCO sing.	98.7%	98.6%
BUSCO dupl.	0.3%	0.3%
BUSCO frag.	0.1%	0.1%
BUSCO miss.	1.0%	1.0%

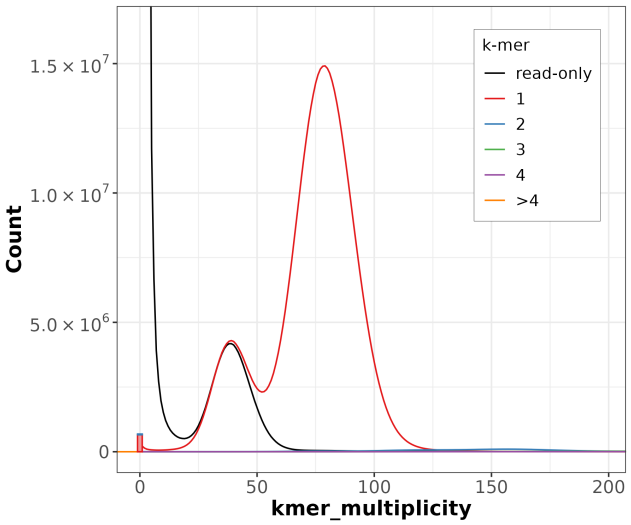
BUSCO: 6.0.0 (euk_genome_min, miniprot) / Lineage: actinopterygii_odb12 (genomes:75, BUSCOs:7207)

HiC contact map of curated assembly

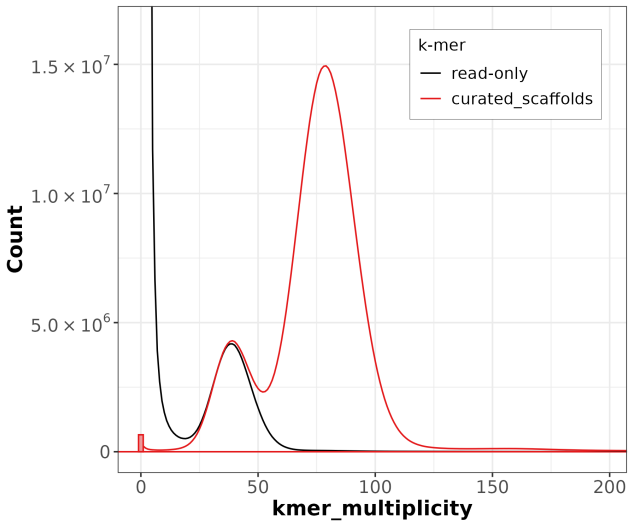


collapsed [LINK]

K-mer spectra of curated assembly

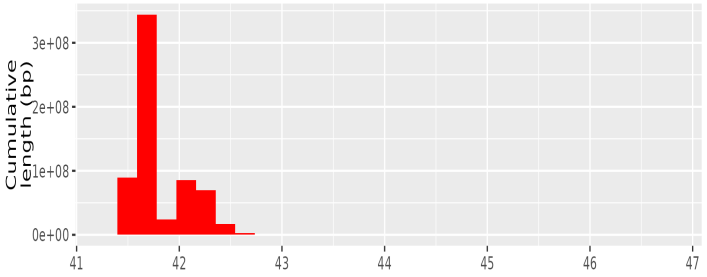


Distribution of k-mer counts per copy numbers found in asm

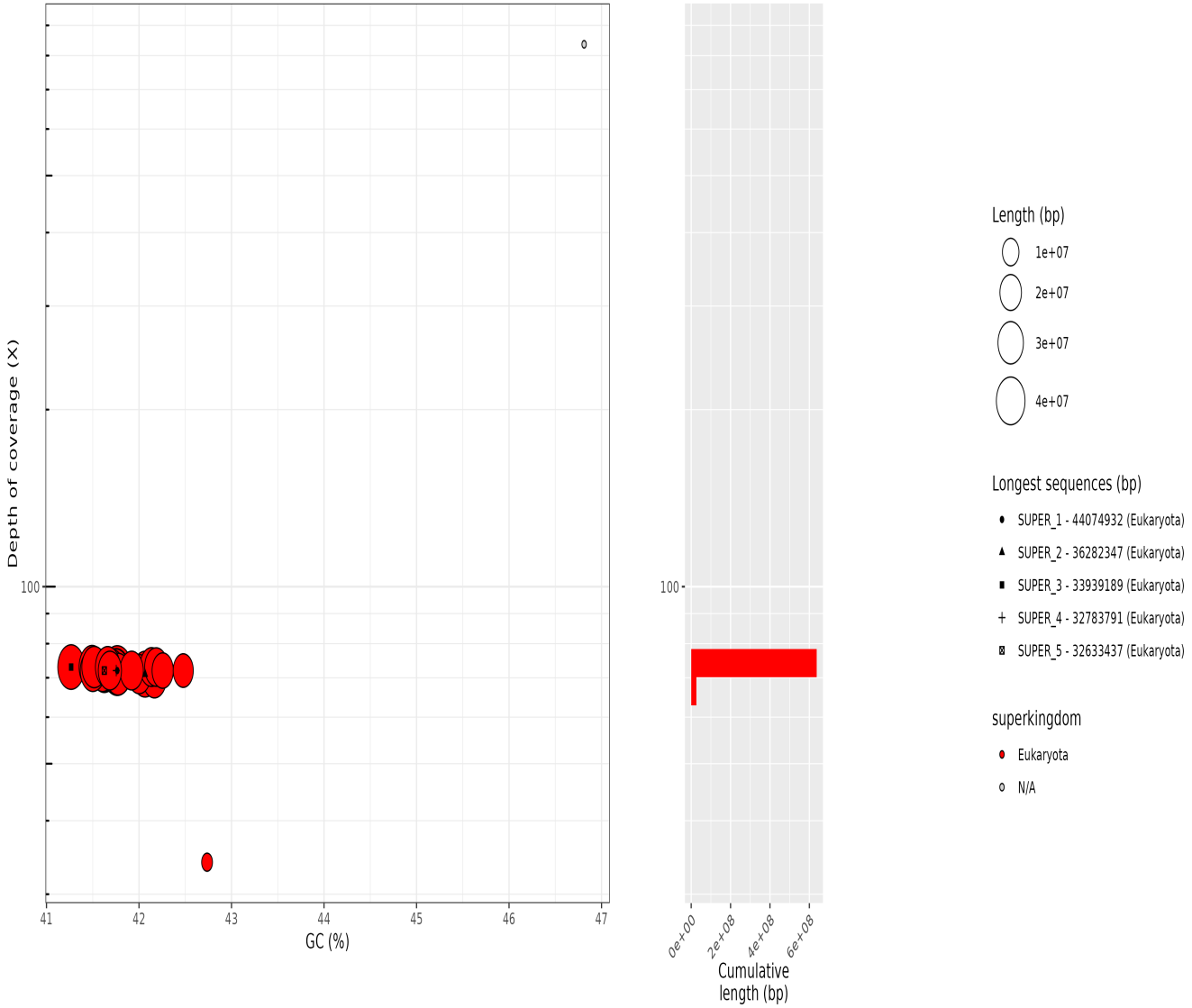


Distribution of k-mer counts coloured by their presence in reads/assemblies

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	73	127

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: 2026-03-17 09:20:40 CET