ERGA Assembly Report

v24.10.15

Tags: ATLASea[INVALID TAG]

TxID	50595
ToLID	**fSpoCan1**
Species	Spondyliosoma cantharus
Class	Actinopteri
Order	Spariformes

Genome Traits	Expected	Observed
Haploid size (bp)	760,621,324	794,918,341
Haploid Number	24 (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.Q65

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

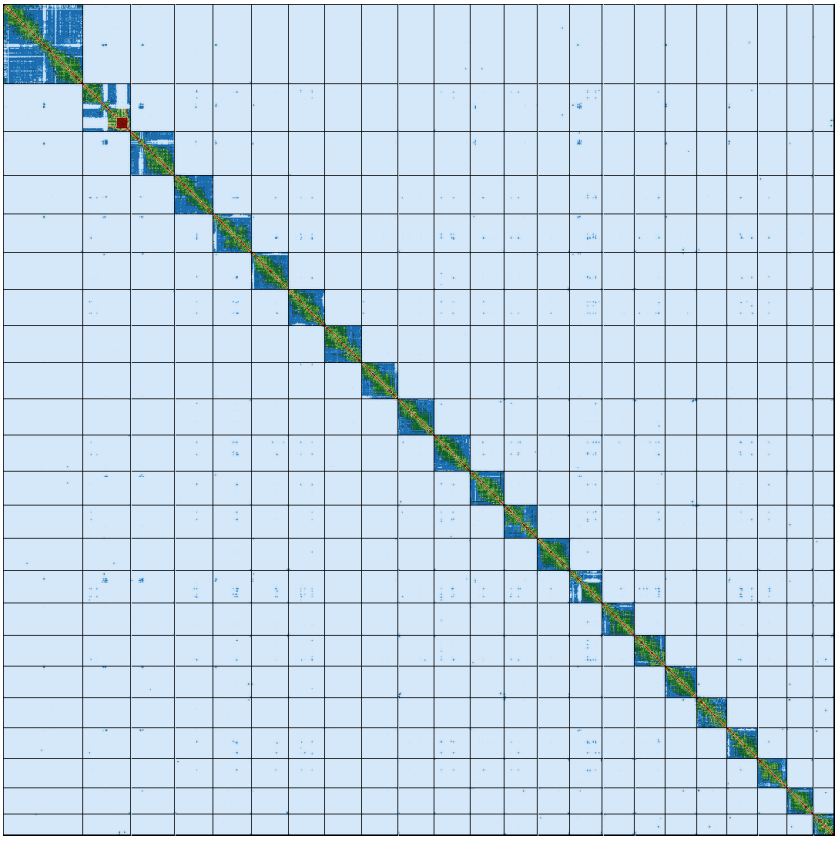
. Interventions/Gb: 3
. Contamination notes: ""
. Other observations: "The assembly of Spondyliosoma cantharus (fSpoCan1) is based on 41X of PacBio data and Arima Hi-C data generated as part of the ATLASea programme (https://www.atlasea.fr). The assembly process included the following steps: initial 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, 51 regions totaling 5.74 Mb (with the largest being 0.85 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, 1 haplotypic region of 0.33 Mb and 1 contaminant sequence of 0.078 Mb 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	795,279,656	794,918,341
GC %	42.13	42.13
Gaps/Gbp	121.97	124.54
Total gap bp	9,700	10,100
Scaffolds	65	59
Scaffold N50	34,487,660	34,487,660
Scaffold L50	10	10
Scaffold L90	21	20
Contigs	162	158
Contig N50	26,522,000	26,522,000
Contig L50	13	13
Contig L90	34	33
QV	48.2892	65.5503
Kmer compl.	88.889	88.8922
BUSCO sing.	97.7%	97.7%
BUSCO dupl.	0.9%	0.9%
BUSCO frag.	0.3%	0.3%
BUSCO miss.	1.1%	1.1%

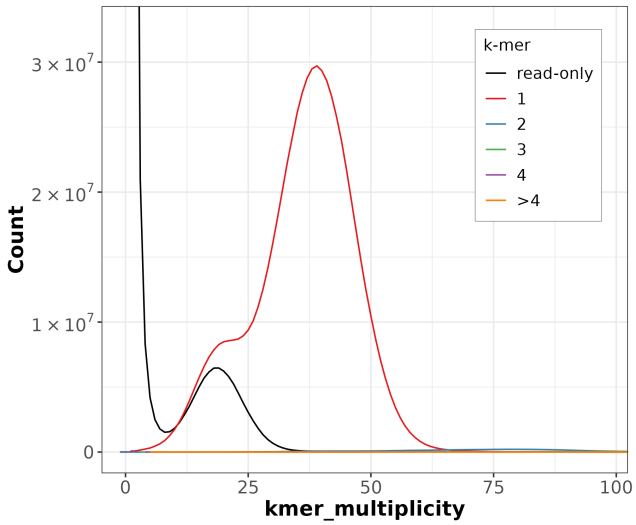
BUSCO: 5.4.3 (euk_genome_met, metaeuk) / Lineage: actinopterygii_odb10 (genomes:26, BUSCOs:3640)

HiC contact map of curated assembly

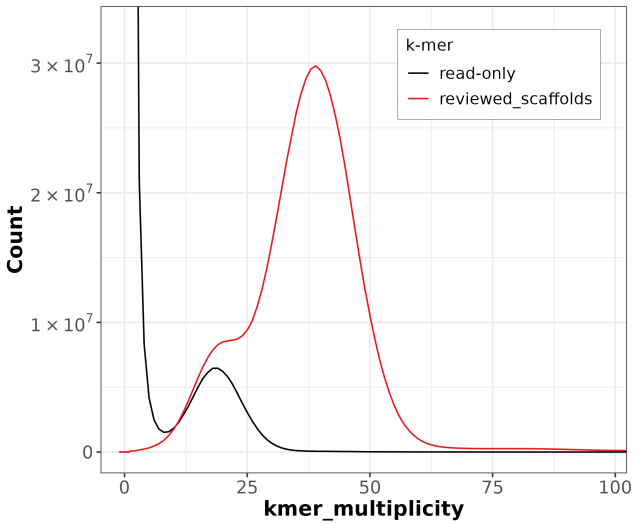


collapsed [LINK]

K-mer spectra of curated assembly

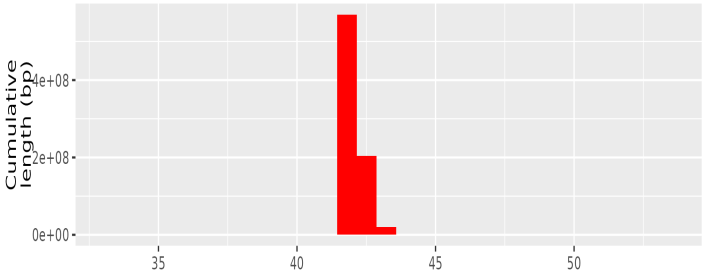


Distribution of k-mer counts per copy numbers found in asm

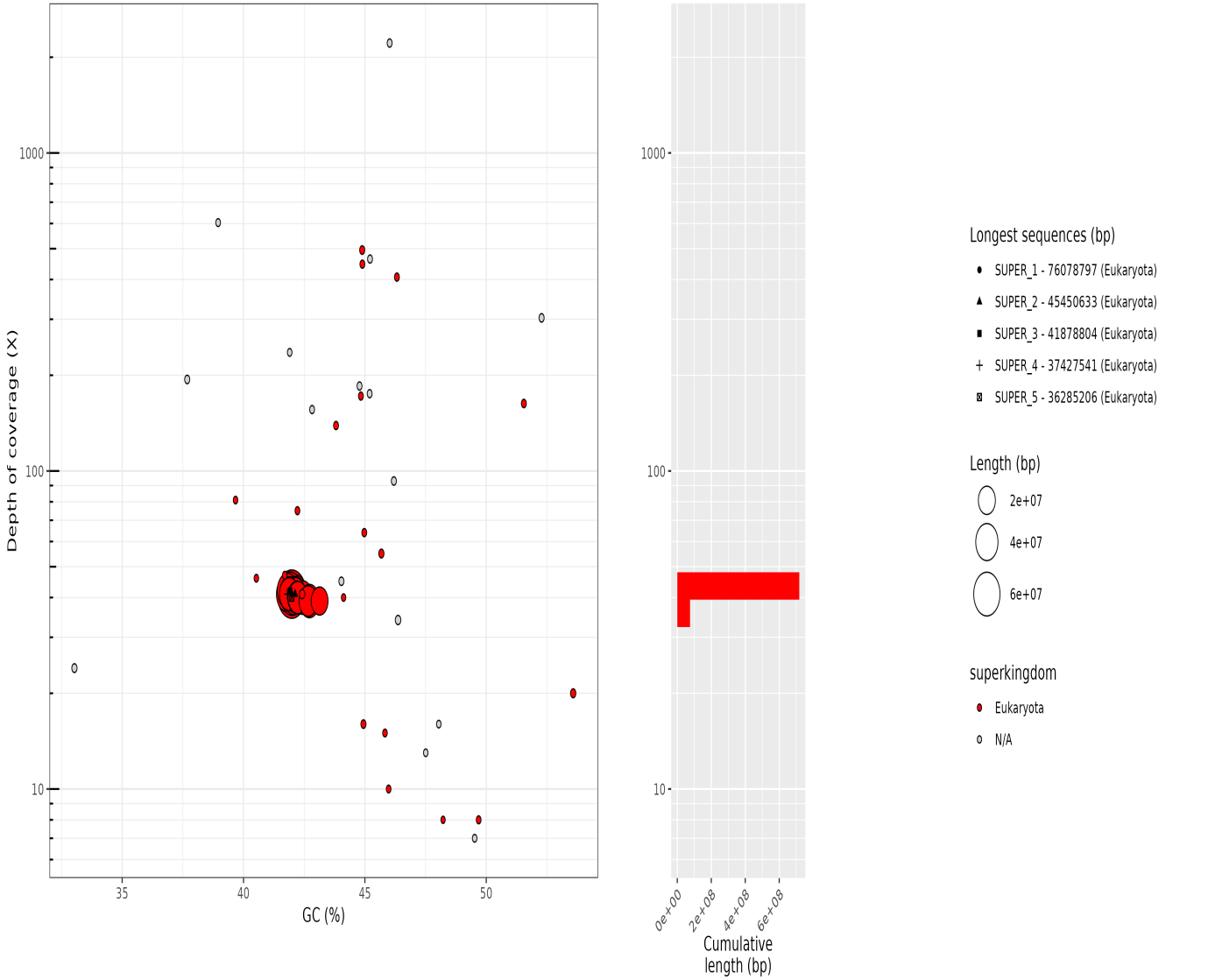


Distribution of k-mer counts coloured by their presence in reads/assemblies

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	PACBIO Hifi	Arima
Coverage	40	136

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: Adama Ndar
Affiliation: Genoscope

Date and time: 2025-04-13 13:13:01 CEST