ERGA Assembly Report

v24.10.15

Tags: ATLASea[INVALID TAG]

TxID	210003
ToLID	**fScoNot1**
Species	Scorpaena notata
Class	Actinopteri
Order	Perciformes

Genome Traits	Expected	Observed
Haploid size (bp)	1,083,905,723	1,111,146,844
Haploid Number	21 (source: ancestor)	17
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:

. Observed Haploid Number is different from Expected

. Kmer completeness value is less than 90 for collapsed
. Assembly length loss > 3% for collapsed

Curator notes

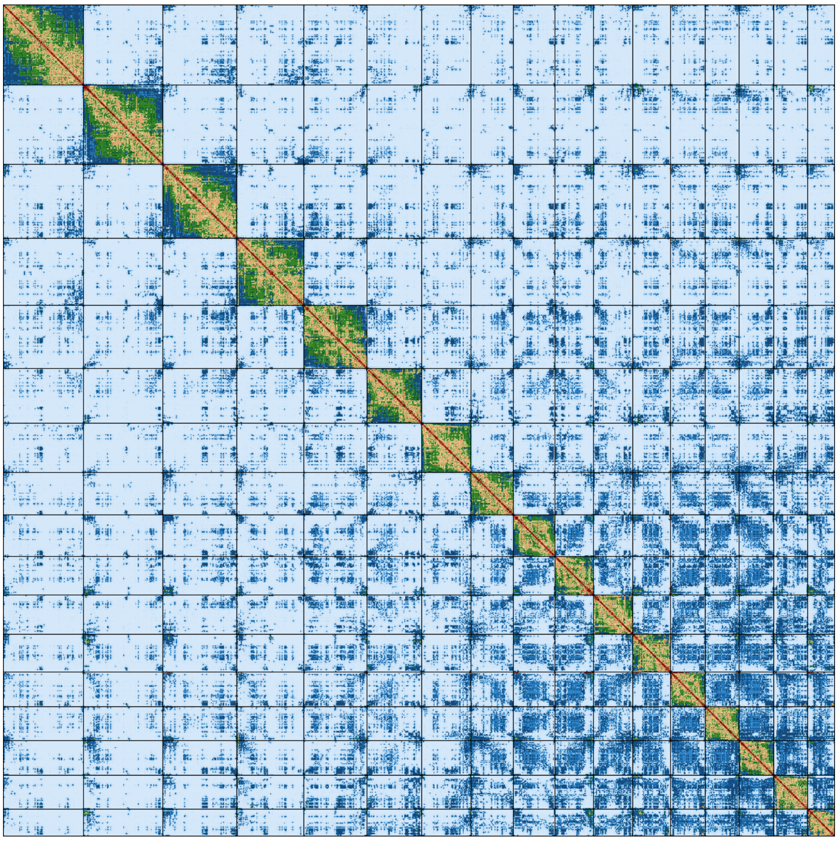
. Interventions/Gb: 8
. Contamination notes: ""
. Other observations: "The assembly of Scorpaena notata (fScoNot1) is based on 82X PacBio data and 220X Arima Hi-C data generated as part of the ATLASea programme (https://www.atlasea.fr). The assembly process included the following steps: initial PacBio 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, 2 contigs were identified as contaminants (bacterial, archaeal, or viral), totaling 0.102 Mb (with the largest being 0.076 Mb). Additionally, 160 regions totaling 9.826 Mb (with the largest being 2.099 Mb) were identified as haplotypic duplications and removed. Mitochondrial genome was assembled using OATK. During manual curation, 1 haplotypic region was removed, totaling 5.5Mb. 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	1,154,814,009	1,111,146,844
GC %	39.12	39.43
Gaps/Gbp	241.6	26.1
Total gap bp	27,900	3,600
Scaffolds	169	33
Scaffold N50	65,689,828	73,160,922
Scaffold L50	7	6
Scaffold L90	16	15
Contigs	448	62
Contig N50	61,225,004	61,225,004
Contig L50	7	7
Contig L90	16	16
QV	44.3071	49.7092
Kmer compl.	77.9562	77.7627
BUSCO sing.	98.9%	99.0%
BUSCO dupl.	0.7%	0.6%
BUSCO frag.	0.0%	0.0%
BUSCO miss.	0.3%	0.4%

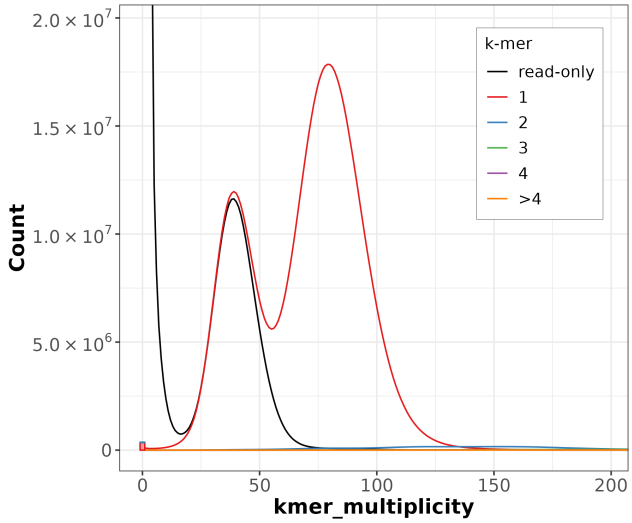
BUSCO: 6.0.0 (euk_genome_min, miniprot) / Lineage: actinopterygii_odb12 (genomes:75, BUSCOs:7207)

HiC contact map of curated assembly

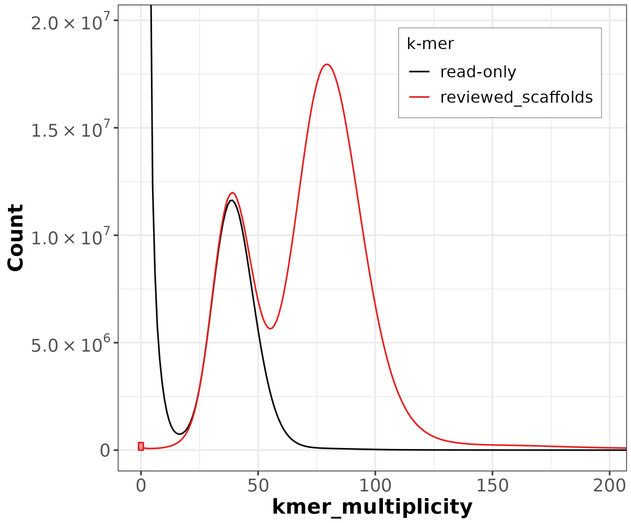


collapsed [LINK]

K-mer spectra of curated assembly

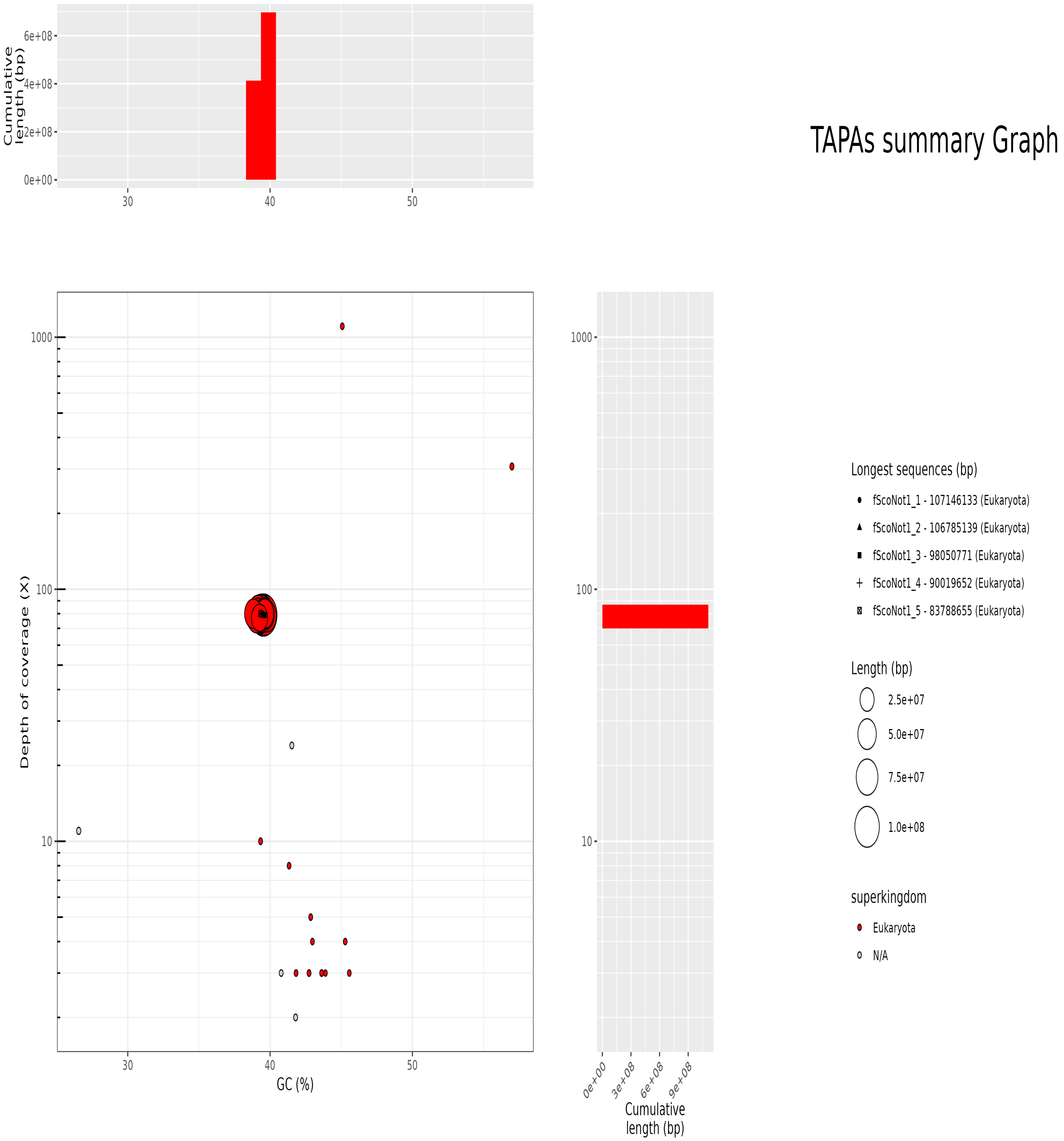


Distribution of k-mer counts per copy numbers found in asm



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	82	220

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: Emilie Teodori
Affiliation: Genoscope

Date and time: 2026-01-28 23:44:51 CET