

# ERGA Assembly Report

v24.10.15

Tags: ATLASea[ INVALID TAG]

TxID	98823
ToLID	<b>fSpiMae1</b>
Species	Spicara maena
Class	Actinopteri
Order	Spariformes

Genome Traits	Expected	Observed
Haploid size (bp)	754,939,922	770,867,518
Haploid Number	22 (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.Q48

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

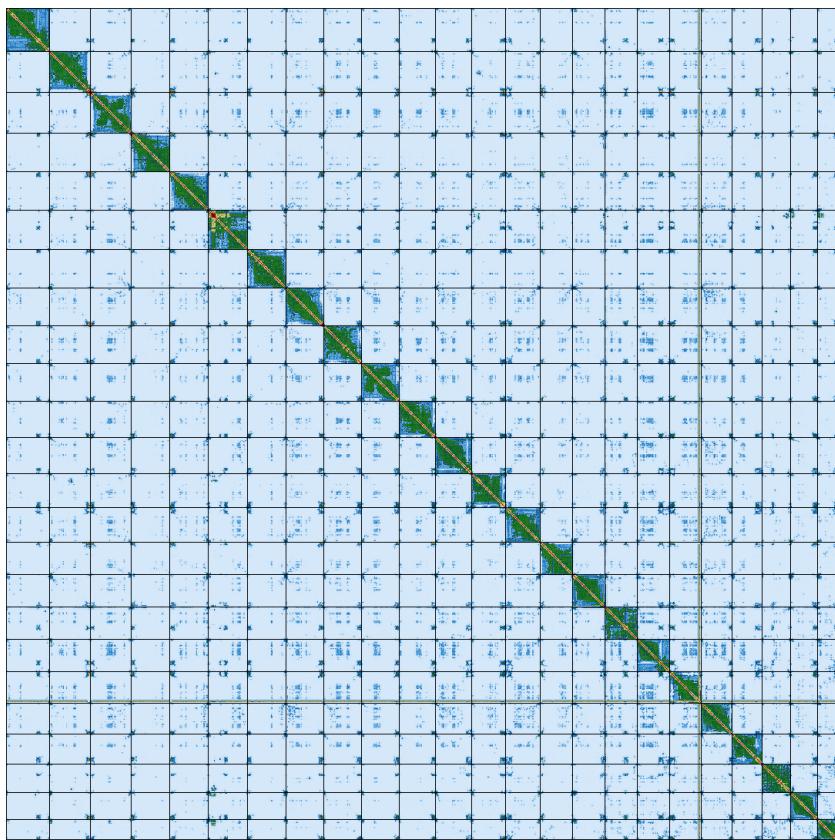
- . Interventions/Gb: 32
- . Contamination notes: ""
- . Other observations: "The assembly of *Spicara maena* (fSpiMae1) is based on 72X PacBio data and 187X Arima Hi-C data generated as part of the ATLASea programme (<https://www.atlasea.fr>). The assembly process included the following steps: initial 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, 43 regions totaling 5.044 Mb (with the largest being 1.759 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. 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	770,848,482	770,867,518
GC %	42.36	42.36
Gaps/Gbp	114.16	124.54
Total gap bp	8,800	11,200
Scaffolds	42	35
Scaffold N50	33,631,152	33,730,052
Scaffold L50	11	11
Scaffold L90	21	21
Contigs	130	131
Contig N50	27,596,238	27,596,238
Contig L50	12	12
Contig L90	27	27
QV	48.72	48.7201
Kmer compl.	85.7144	85.7165
BUSCO sing.	99.5%	99.5%
BUSCO dupl.	0.2%	0.2%
BUSCO frag.	0.0%	0.0%
BUSCO miss.	0.3%	0.3%

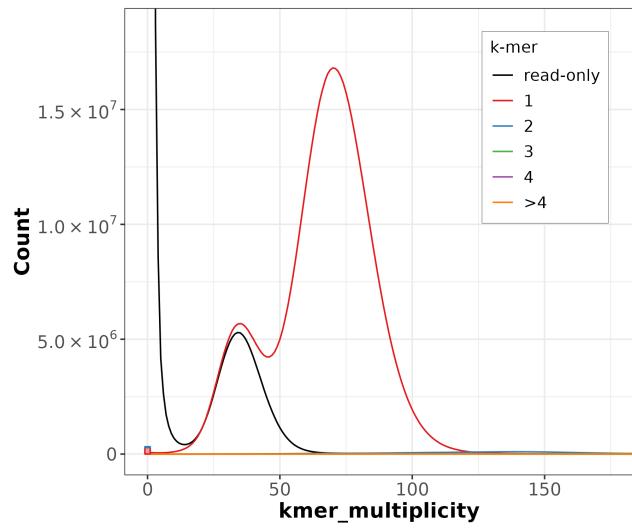
BUSCO: 6.0.0 (euk\_genome\_min, miniprot) / Lineage: actinopterygii\_odb12 (genomes:75, BUSCOs:7207)

# HiC contact map of curated assembly

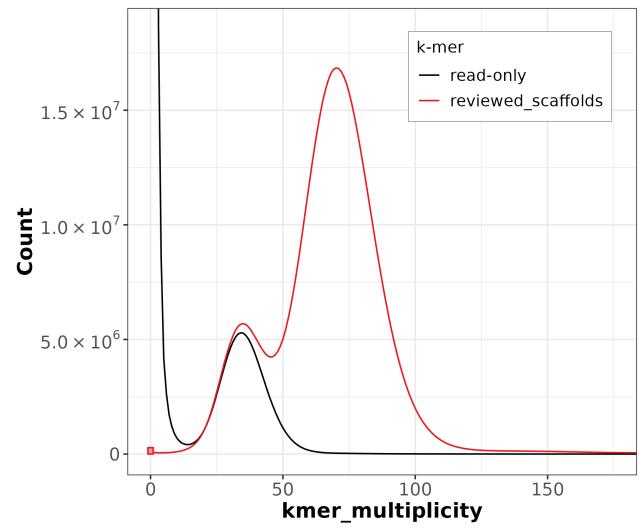


collapsed [\[LINK\]](#)

# K-mer spectra of curated assembly

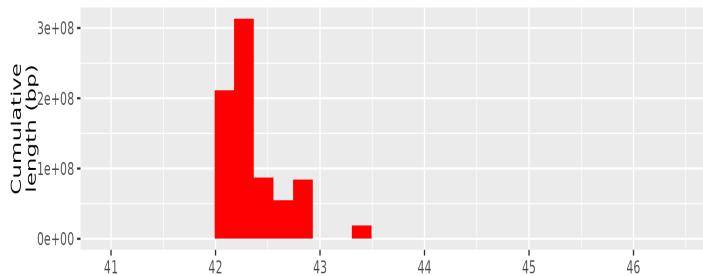


Distribution of k-mer counts per copy numbers found in asm

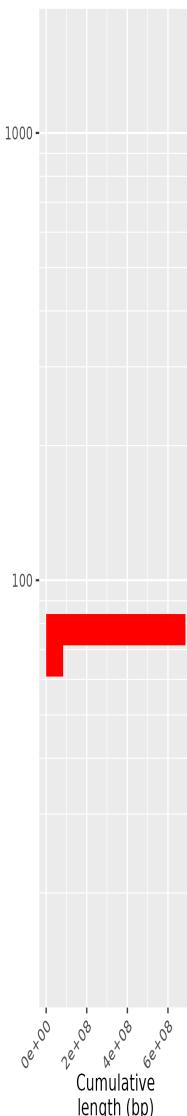
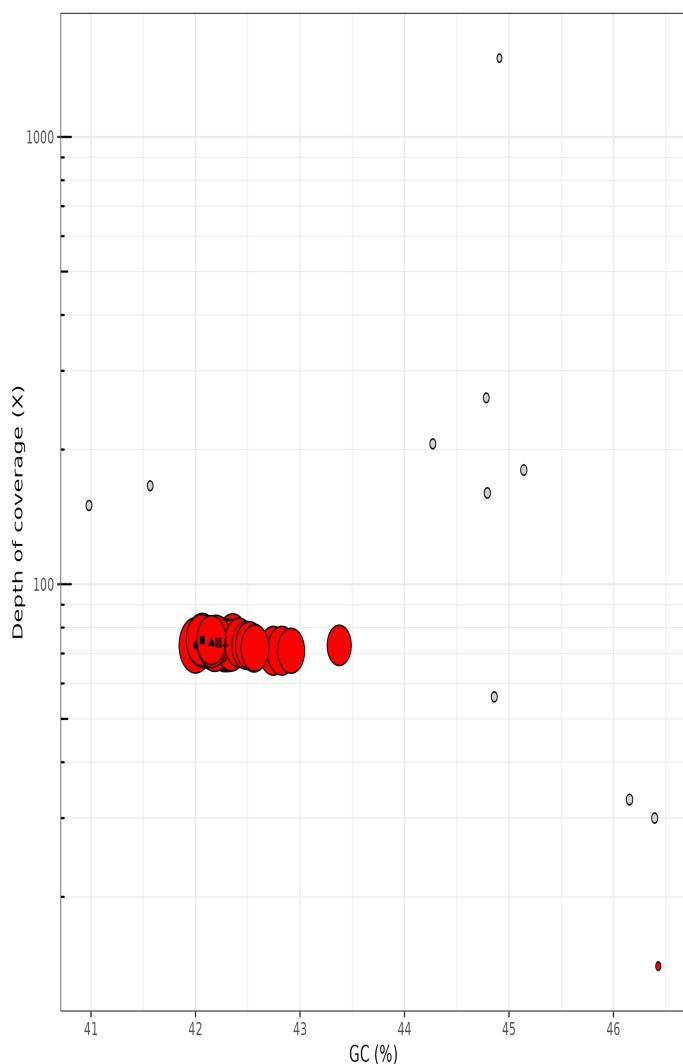


Distribution of k-mer counts coloured by their presence in reads/assemblies

# Post-curation contamination screening



TAPAs summary Graph



Length (bp)

- 1e+07
- 2e+07
- 3e+07
- 4e+07

Longest sequences (bp)

- fSpiMae1\_1 - 40234271 (Eukaryota)
- ▲ fSpiMae1\_2 - 38151549 (Eukaryota)
- fSpiMae1\_3 - 37266167 (Eukaryota)
- + fSpiMae1\_4 - 36023524 (Eukaryota)
- ▣ fSpiMae1\_5 - 35921616 (Eukaryota)

superkingdom

- Eukaryota
- N/A

**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	72	187

## 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: Arnaud Couloux  
Affiliation: Genoscope

Date and time: 2026-01-21 05:11:18 CET