ERGA Assembly Report

v24.10.15

Tags: ATLASea[INVALID TAG]

TxID	274700
ToLID	**fScoPlm1**
Species	Scorpaena plumieri
Class	Actinopteri
Order	Perciformes

Genome Traits	Expected	Observed
Haploid size (bp)	1,238,341,046	1,295,977,282
Haploid Number	21 (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.Q50

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

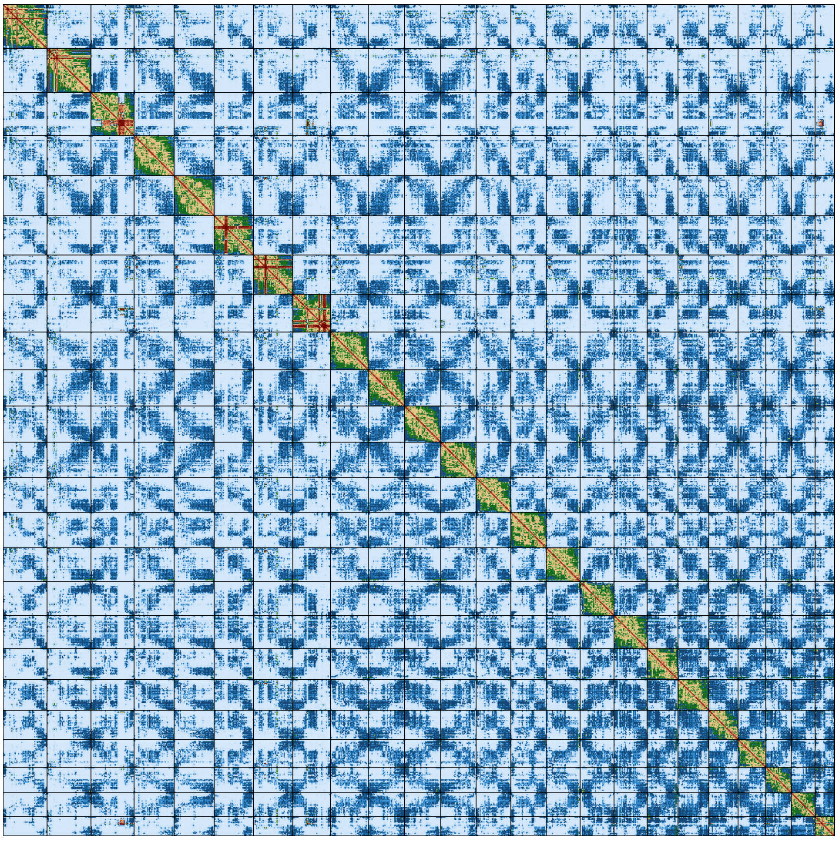
. Interventions/Gb: 21
. Contamination notes: ""
. Other observations: "The assembly of Scorpaena plumieri (fScoPlm1) is based on 65X PacBio data and 173X Arima Hi-C data generated as part of the ATLASea programme (https://www.atlasea.fr). The assembly process included the following steps: initial PacBio 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, 30 regions totaling 23.361 Mb (with the largest being 19.945 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	1,295,982,982	1,295,977,282
GC %	39.05	39.05
Gaps/Gbp	27.01	37.04
Total gap bp	3,500	6,800
Scaffolds	44	30
Scaffold N50	55,501,541	55,658,721
Scaffold L50	11	11
Scaffold L90	21	21
Contigs	79	78
Contig N50	47,234,000	47,234,000
Contig L50	12	12
Contig L90	26	26
QV	50.183	50.1899
Kmer compl.	87.0759	87.0757
BUSCO sing.	99.1%	99.1%
BUSCO dupl.	0.6%	0.6%
BUSCO frag.	0.1%	0.1%
BUSCO miss.	0.2%	0.2%

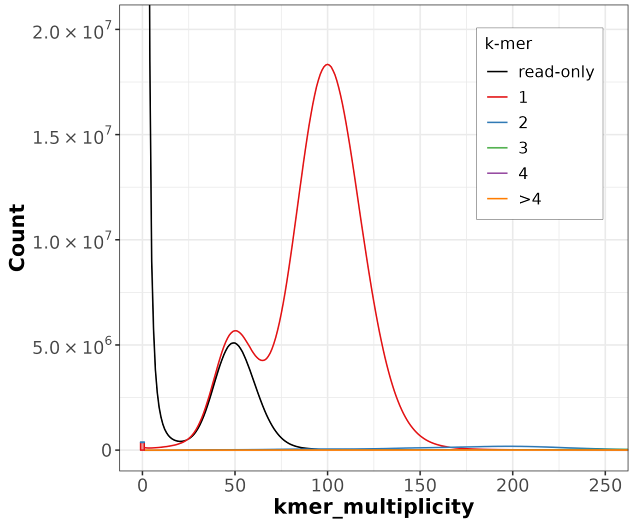
BUSCO: 6.0.0 (euk_genome_min, miniprot) / Lineage: actinopterygii_odb12 (genomes:75, BUSCOs:7207)

HiC contact map of curated assembly

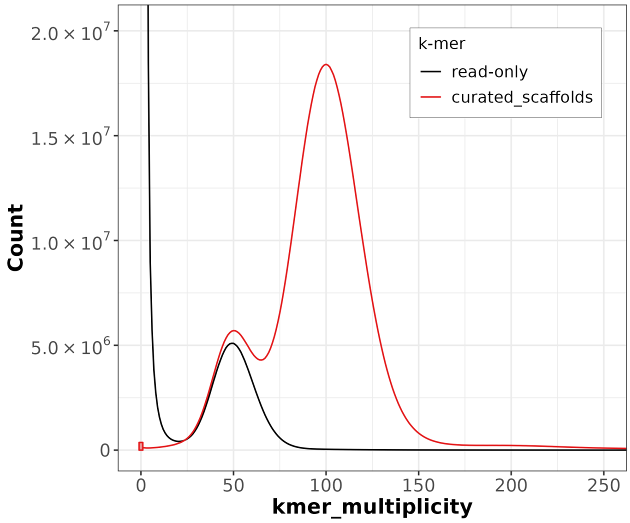


collapsed [LINK]

K-mer spectra of curated assembly

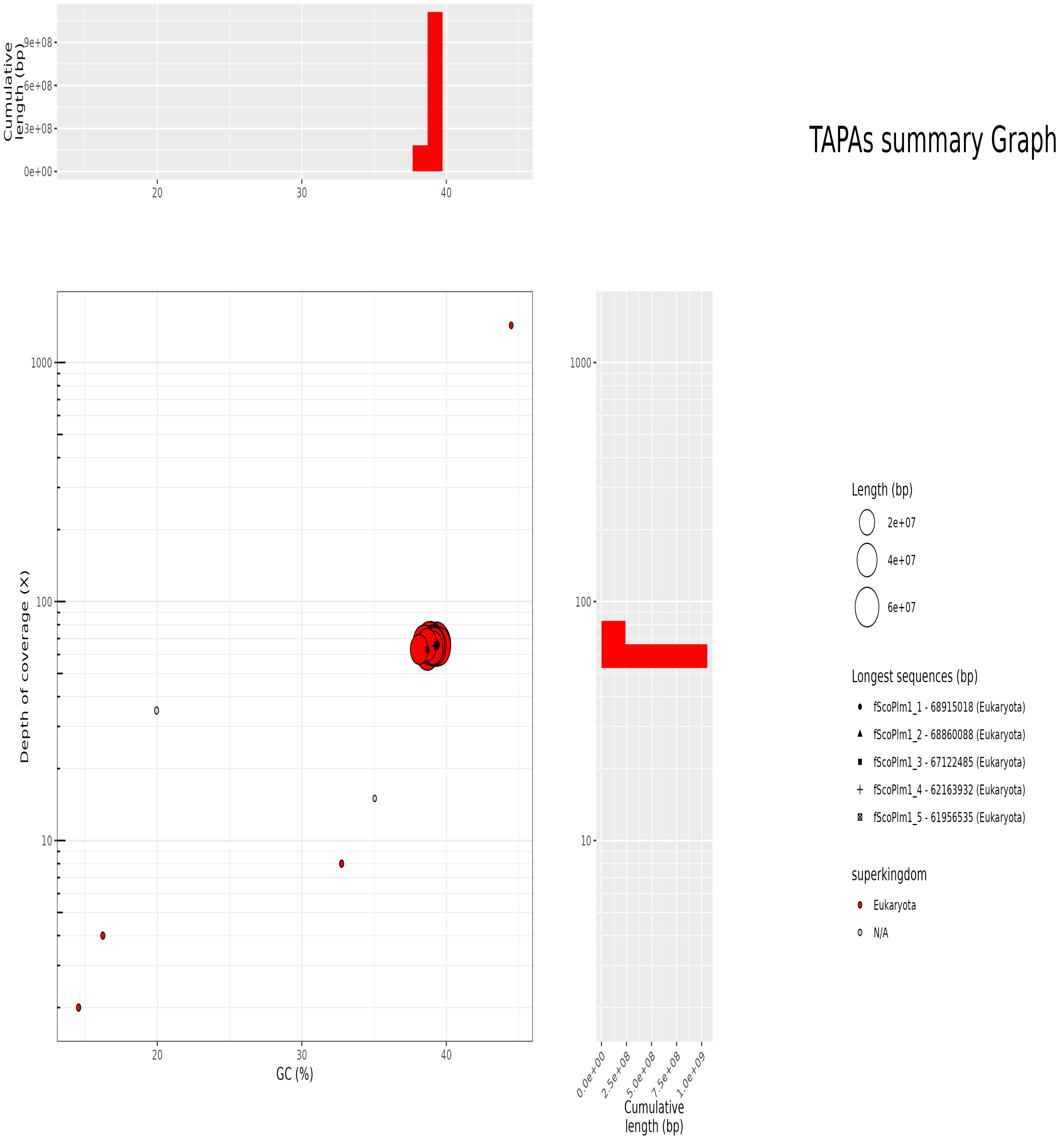


Distribution of k-mer counts per copy numbers found in asm



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	65	173

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: Jean-Marc Aury
Affiliation: Genoscope

Date and time: 2025-11-25 13:55:41 CET