

ERGA Assembly Report

v24.10.15

Tags: ATLASea[INVALID TAG]

TxID	1182848
ToLID	fScoTrel
Species	Scorpaenodes tredecimspinosus
Class	Actinopteri
Order	Perciformes

Genome Traits	Expected	Observed
Haploid size (bp)	1,046,219,331	1,049,559,207
Haploid Number	21 (source: ancestor)	26
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

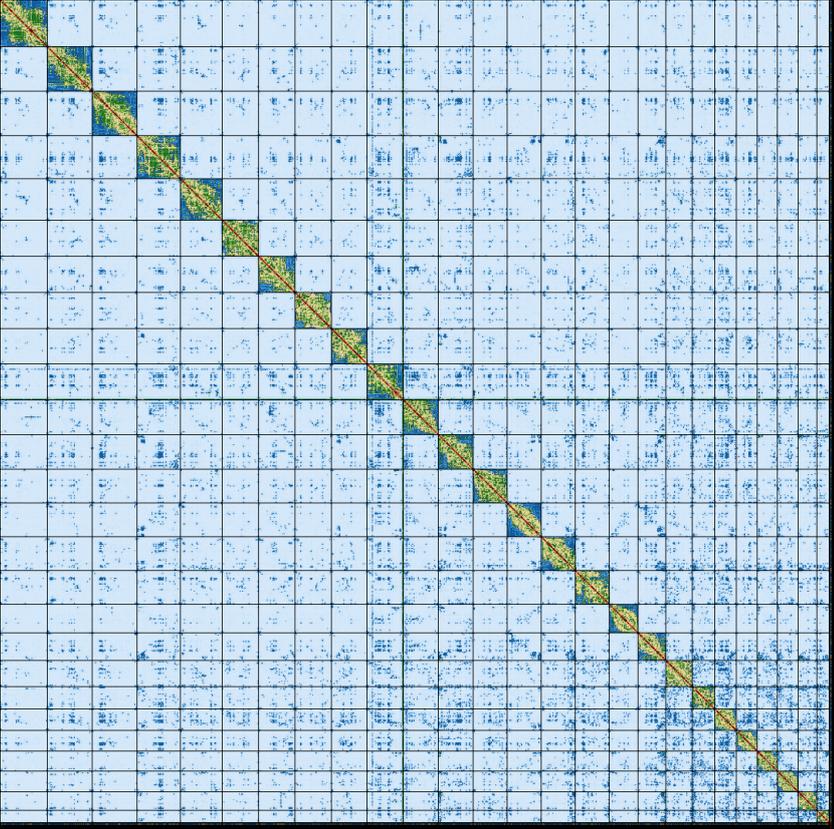
. Interventions/Gb: 31
. Contamination notes: ""
. Other observations: "The assembly of Scorpaenodes tredecimspinosus (fScoTrel) is based on 77X PacBio data and 159X Arima Hi-C data generated as part of the ATLASea programme (<https://www.atlasea.fr>). The assembly process included the following steps: 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 0.154 Mb (with the largest being 0.092 Mb). Additionally, 209 regions totaling 40.918 Mb (with the largest being 2.763 Mb) were identified as haplotypic duplications and removed. Mitochondrial genome was assembled using OATK. During manual curation, 9 haplotypic regions were removed, totaling 20Mb (with the largest being 13Mb). 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	1,071,209,398	1,049,559,207
GC %	41.78	41.79
Gaps/Gbp	85.88	100.99
Total gap bp	9,200	13,200
Scaffolds	76	55
Scaffold N50	43,602,236	43,648,526
Scaffold L50	11	11
Scaffold L90	23	22
Contigs	168	161
Contig N50	31,605,000	33,526,000
Contig L50	14	13
Contig L90	35	35
QV	48.6567	48.7669
Kmer compl.	70.5962	70.0152
BUSCO sing.	96.4%	98.6%
BUSCO dupl.	2.7%	0.5%
BUSCO frag.	0.0%	0.0%
BUSCO miss.	0.8%	0.9%

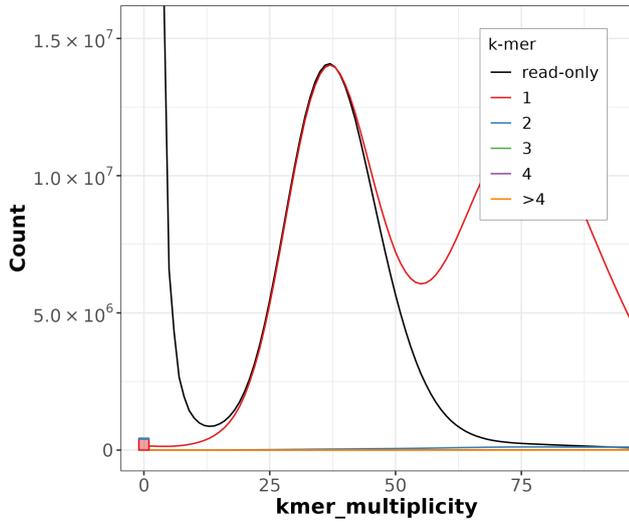
BUSCO: 6.0.0 (euk_genome_min, miniprot) / Lineage: actinopterygii_odb12 (genomes:75, BUSCOs:7207)

HiC contact map of curated assembly

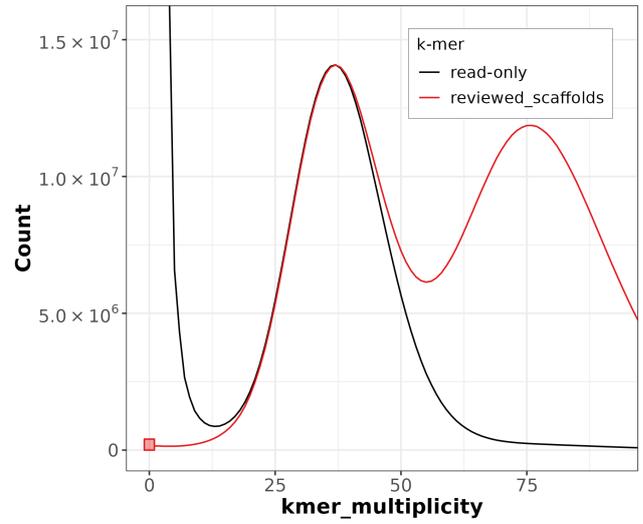


collapsed [\[LINK\]](#)

K-mer spectra of curated assembly

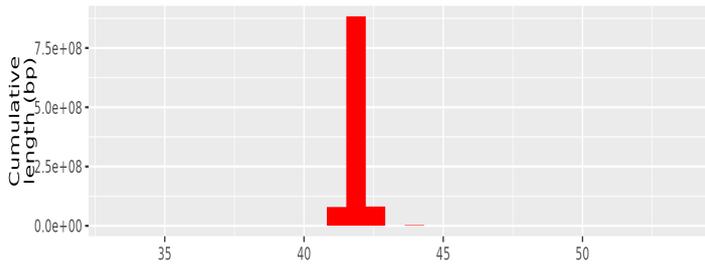


Distribution of k-mer counts per copy numbers found in asm

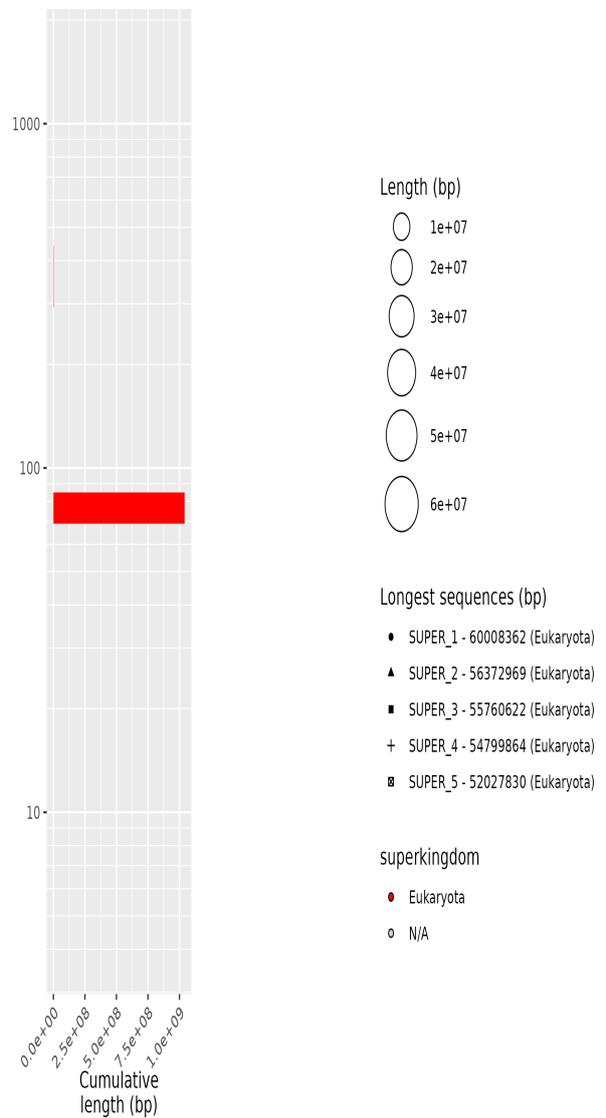
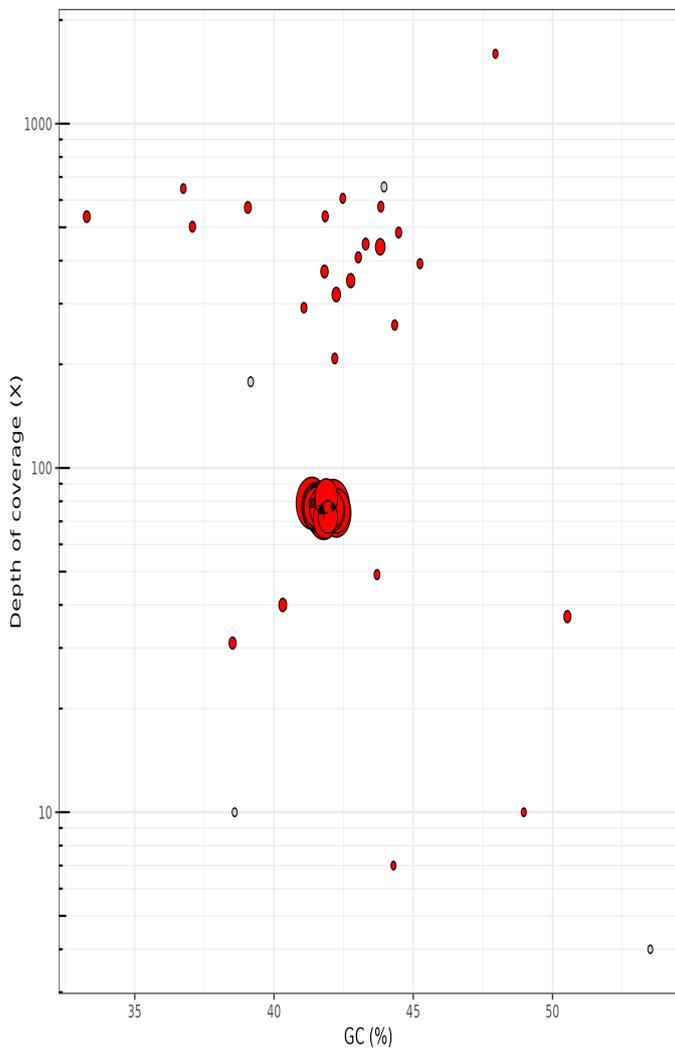


Distribution of k-mer counts coloured by their presence in reads/assemblies

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	77	159

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: Benjamin Istace

Affiliation: Genoscope

Date and time: 2026-01-21 07:04:32 CET