ERGA Assembly Report

v24.10.15

Tags: ATLASea[INVALID TAG]

TxID	161589
ToLID	**fSynRos1**
Species	Syngnathus rostellatus
Class	Actinopteri
Order	Syngnathiformes

Genome Traits	Expected	Observed
Haploid size (bp)	361,247,141	378,829,410
Haploid Number	22 (source: ancestor)	22
Ploidy	2 (source: ancestor)	2
Sample Sex	Unknown	Unknown

EBP metrics summary and curation notes

Obtained EBP quality metric for collapsed: 6.7.Q48

The following metrics were automatically flagged as below EBP recommended standards or different from expected:

. Kmer completeness value is less than 90 for collapsed

Curator notes

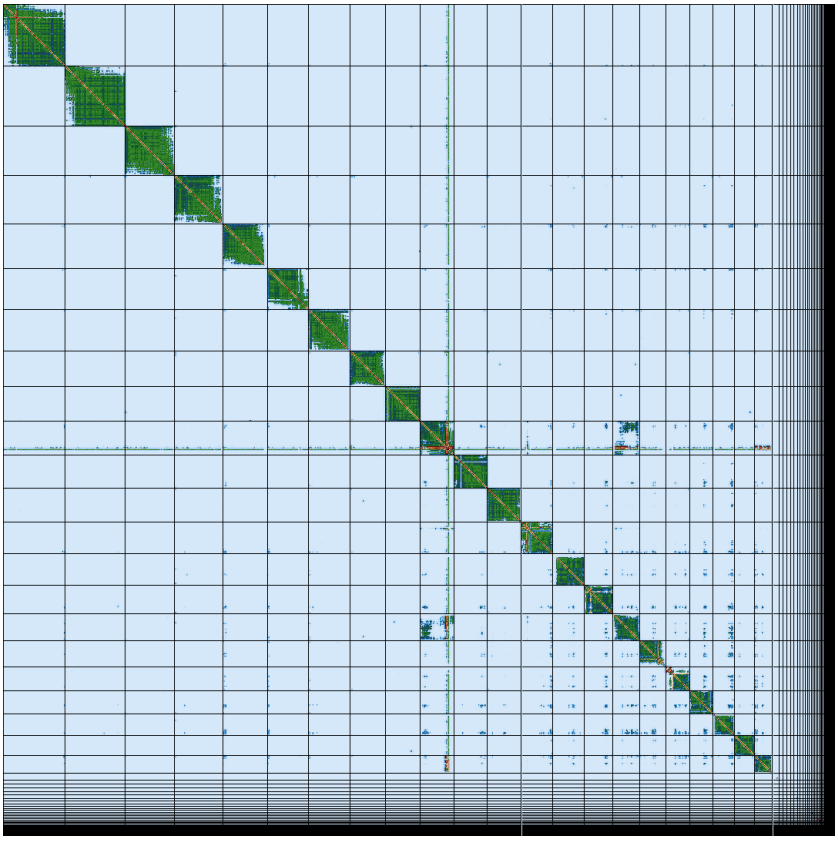
. Interventions/Gb: 28
. Contamination notes: ""
. Other observations: "The assembly of Syngnathus rostellatus (fSynRos1) is based on 47X PacBio data and 192X Arima Hi-C data generated as part of the ATLASea programme (https://www.atlasea.fr). The assembly process included the following steps: initial 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, 1 contig was identified as contaminants (bacterial) , totaling 0.18 Mb. Additionally, 110 regions totaling 24 Mb (with the largest being 1.45 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, 6 haplotypic regions and 2 contaminant sequences were removed, totaling 5.3 Mb and 0.8 Mb, respectively (with the largest being 1.9 Mb and 0.49 Mb). 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	384,208,184	378,829,410
GC %	43.35	43.32
Gaps/Gbp	182.19	195.34
Total gap bp	7,000	8,000
Scaffolds	87	77
Scaffold N50	16,008,141	15,703,910
Scaffold L50	9	9
Scaffold L90	24	21
Contigs	157	151
Contig N50	5,932,802	5,932,802
Contig L50	23	23
Contig L90	76	72
QV	48.4181	48.3861
Kmer compl.	81.5908	81.4923
BUSCO sing.	90.3%	90.9%
BUSCO dupl.	1.2%	0.6%
BUSCO frag.	1.8%	1.8%
BUSCO miss.	6.7%	6.7%

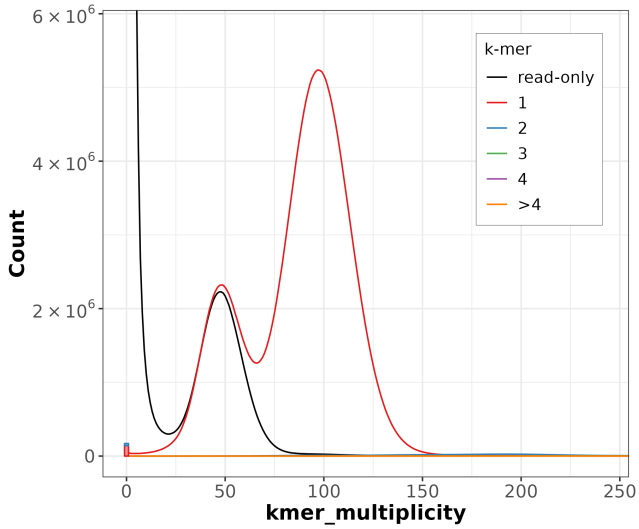
BUSCO: 5.8.2 (euk_genome_met, metaeuk) / Lineage: actinopterygii_odb12 (genomes:75, BUSCOs:7207)

HiC contact map of curated assembly

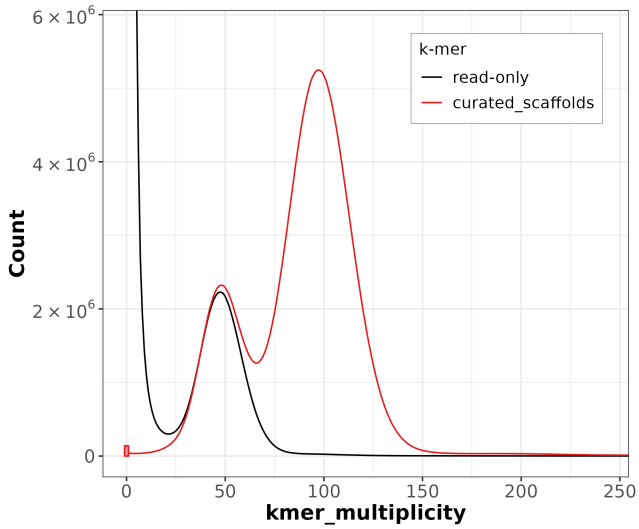


collapsed [LINK]

K-mer spectra of curated assembly

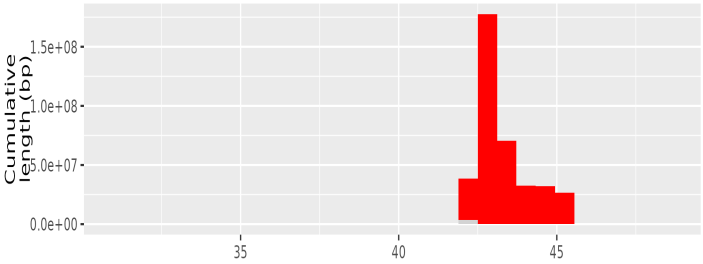


Distribution of k-mer counts per copy numbers found in asm

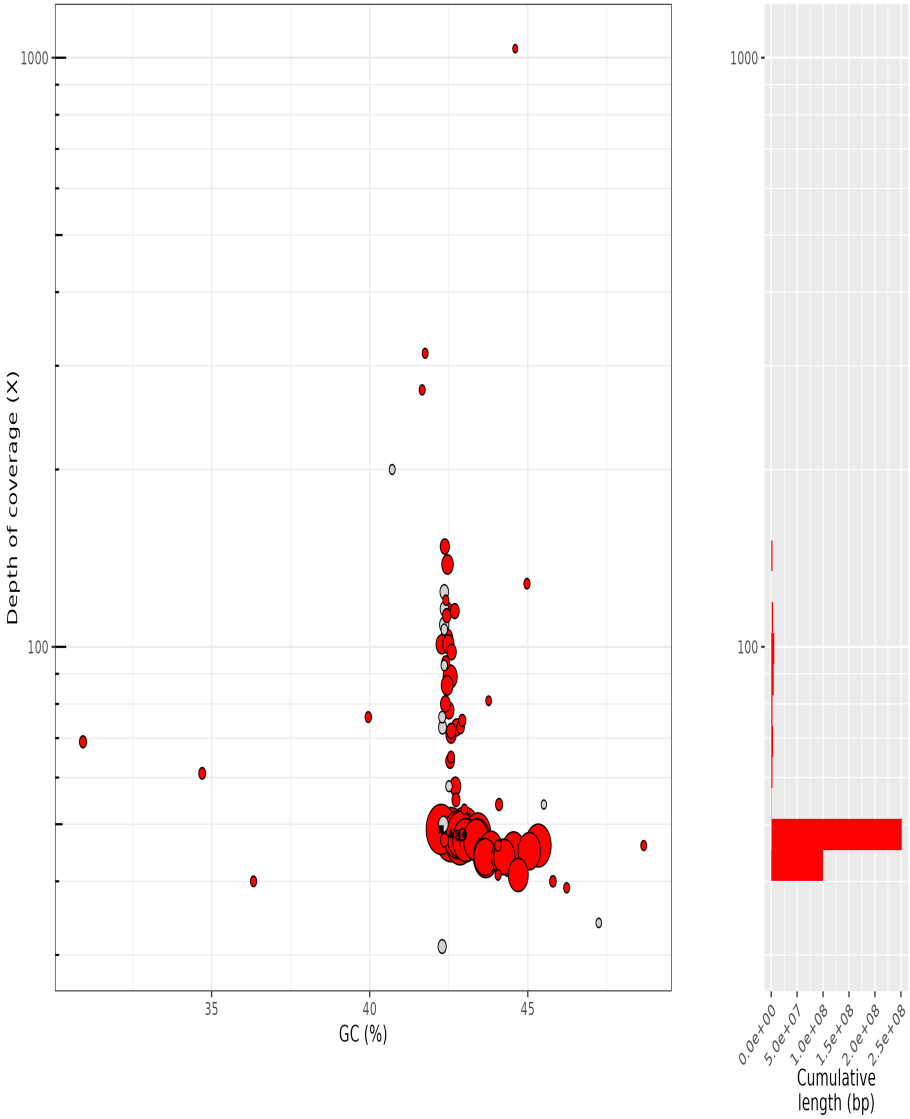


Distribution of k-mer counts coloured by their presence in reads/assemblies

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	PACBIO Hifi	Arima
Coverage	47	192

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: Sophie Layac
Affiliation: Genoscope

Date and time: 2025-06-18 09:23:59 CEST