

ERGA Assembly Report

v24.10.15

Tags: ATLASEa[INVALID TAG]

TxID	1306753
ToLID	fGymSem1
Species	Gymnammodytes semisquamatus
Class	Actinopteri
Order	Uranoscopiformes

Genome Traits	Expected	Observed
Haploid size (bp)	668,882,634	723,397,971
Haploid Number	13 (source: ancestor)	23
Ploidy	2 (source: ancestor)	2
Sample Sex	Unknown	Unknown

EBP metrics summary and curation notes

Obtained EBP quality metric for collapsed: 7.7.Q45

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

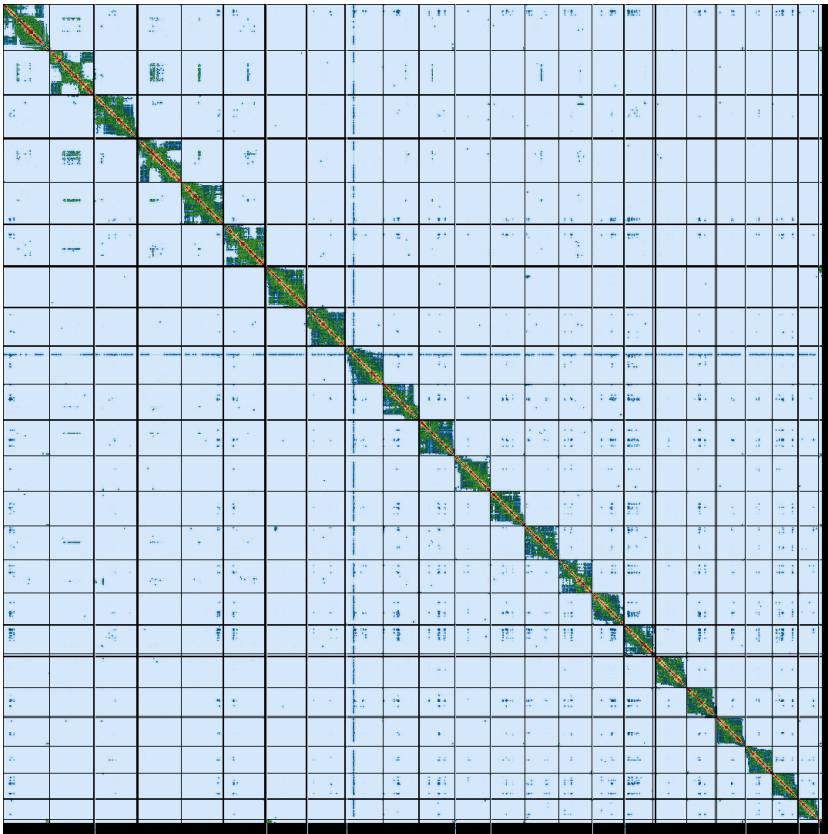
. Interventions/Gb: 32
. Contamination notes: ""
. Other observations: "The assembly of *Gymnammodytes semisquamatus* (fGymSem1) is based on 71X PacBio data and 146X Arima Hi-C data generated as part of the ATLASEa programme (<https://www.atlasea.fr>). The assembly process included the following steps: 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, 3 contigs were identified as contaminants (bacterial, archaeal, or viral), totaling 0.378 Mb (with the largest being 0.234 Mb). Additionally, 468 regions totaling 44.944 Mb (with the largest being 0.535 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, 39 haplotypic regions were removed, totaling 12.9Mb, (with the largest being 1.03Mb). 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	736,324,155	723,397,971
GC %	43.24	43.24
Gaps/Gbp	301.5	337.3
Total gap bp	22,200	27,600
Scaffolds	255	182
Scaffold N50	30,154,693	30,679,050
Scaffold L50	11	11
Scaffold L90	22	21
Contigs	477	426
Contig N50	21,555,529	21,555,529
Contig L50	14	14
Contig L90	73	57
QV	45.0249	45.0847
Kmer compl.	78.522	77.8364
BUSCO sing.	94.1%	94.6%
BUSCO dupl.	1.4%	0.7%
BUSCO frag.	1.5%	1.5%
BUSCO miss.	3.0%	3.2%

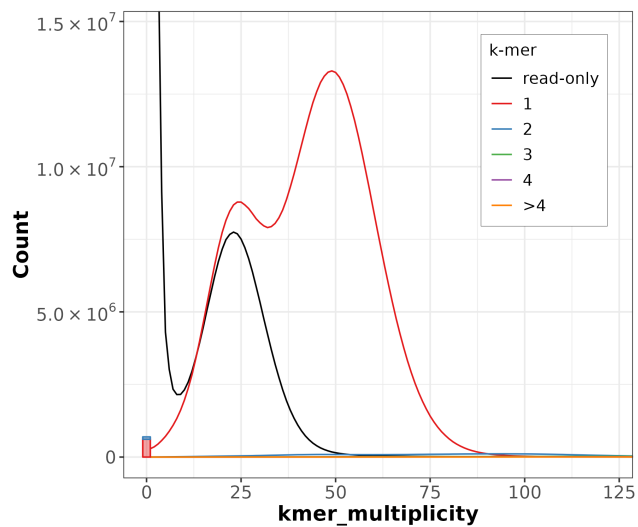
BUSCO: 5.8.2 (euk_genome_met, metaeuk) / Lineage: actinopterygii_odb12 (genomes:75, BUSCOs:7207)

HiC contact map of curated assembly

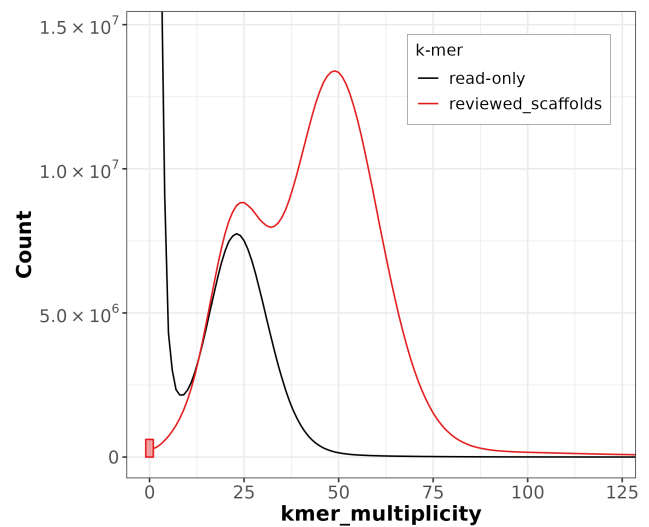


collapsed [\[LINK\]](#)

K-mer spectra of curated assembly

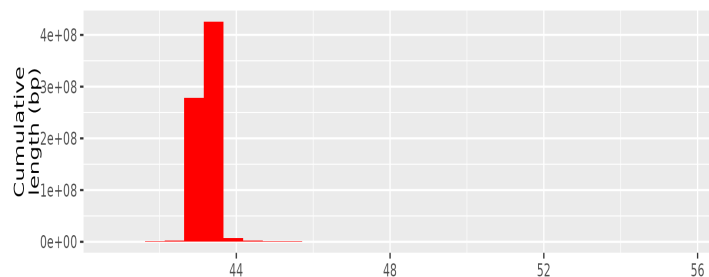


Distribution of k-mer counts per copy numbers found in asm

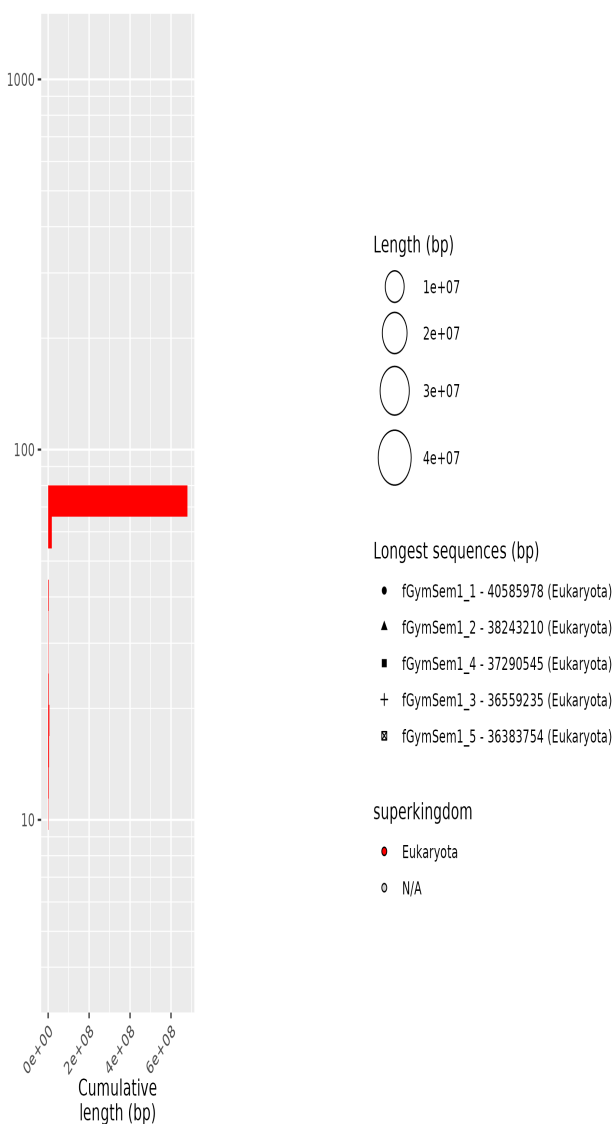
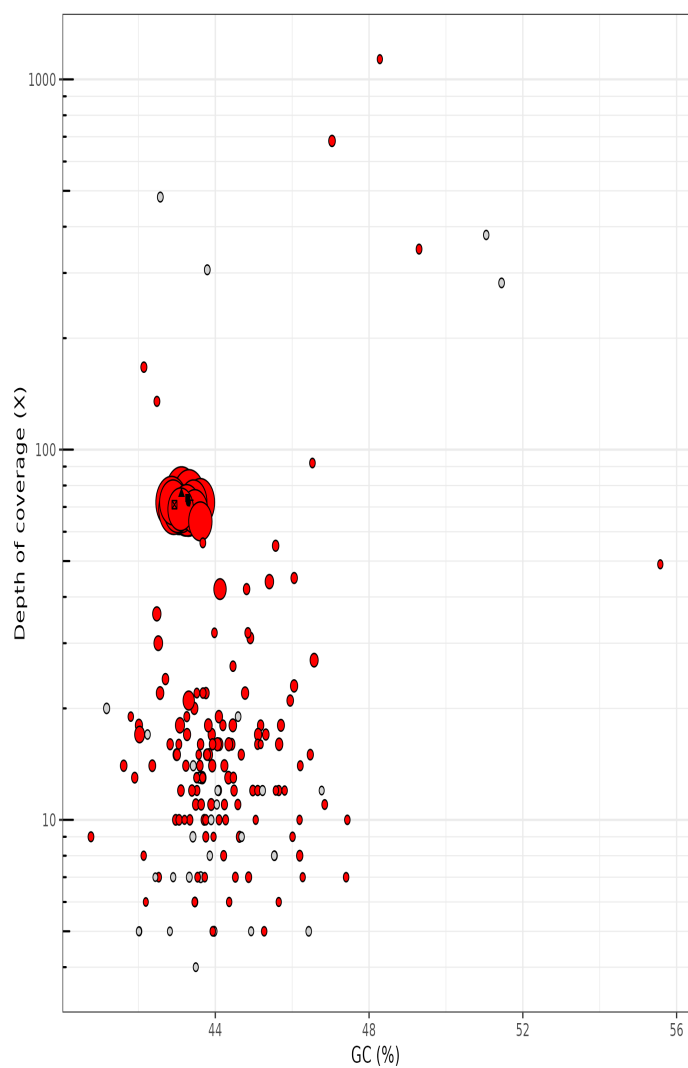


Distribution of k-mer counts coloured by their presence in reads/assemblies

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	71	145

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: Simone Duprat

Affiliation: Genoscope

Date and time: 2025-07-23 18:34:20 CEST