

ERGA Assembly Report

v24.10.15

Tags: ATLASea[INVALID TAG]

TxID	349666
ToLID	fPhyPhy1
Species	Phycis phycis
Class	Actinopteri
Order	Gadiformes

Genome Traits	Expected	Observed
Haploid size (bp)	524,692,262	565,357,171
Haploid Number	23 (source: ancestor)	25
Ploidy	2 (source: ancestor)	2
Sample Sex	Unknown	Unknown

EBP metrics summary and curation notes

Obtained EBP quality metric for collapsed: 7.7.Q42

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

- . Observed Haploid Number is different from Expected

Curator notes

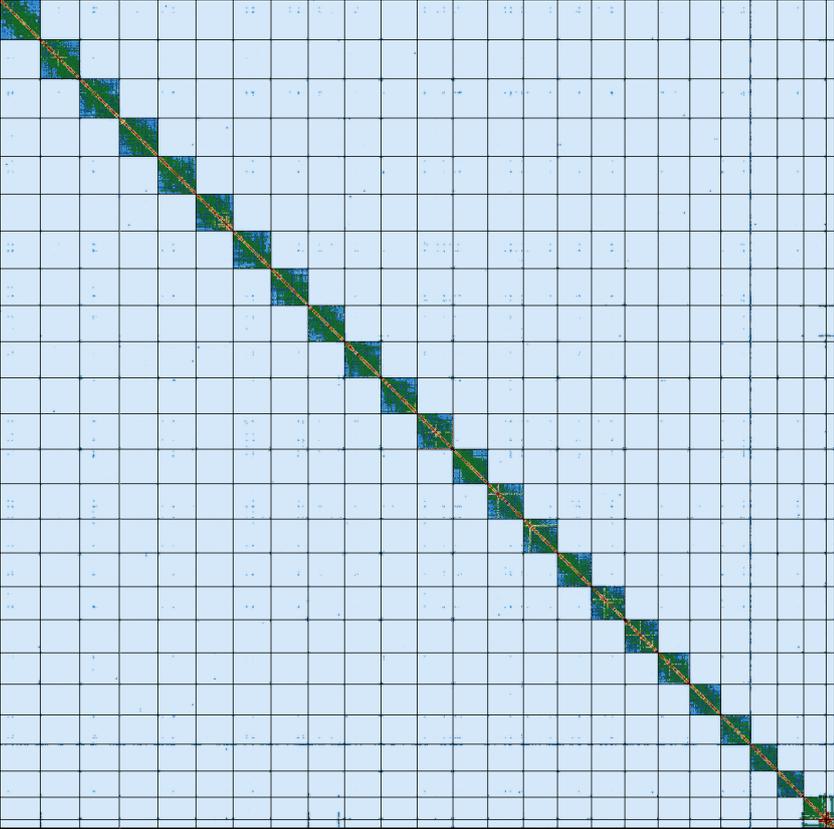
- . Interventions/Gb: 9
- . Contamination notes: ""
- . Other observations: "The assembly of Phycis phycis (fPhyPhy1) is based on 46X ONT data and 205X Arima Hi-C data generated as part of the ATLASea programme (<https://www.atlasea.fr>). The assembly process included the following steps: initial ONT 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, 71 regions totaling 14.251 Mb (with the largest being 5.69 Mb) were identified as haplotypic duplications and removed. The mitochondrial genome was assembled using OATK and PacBio data initially generated for this assembly. During manual curation, one haplotypic region totaling 1.05 Mb was removed. Chromosomes fPhyPhy1_24 and fPhyPhy1_25 appear to be sex chromosomes, as they show approximately half the coverage observed for other chromosomes. 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	566,504,102	565,357,171
GC %	48.35	48.36
Gaps/Gbp	15.89	17.69
Total gap bp	900	1,300
Scaffolds	40	34
Scaffold N50	24,000,441	24,071,641
Scaffold L50	12	12
Scaffold L90	22	22
Contigs	49	44
Contig N50	22,804,303	22,804,303
Contig L50	12	12
Contig L90	23	23
QV	42.0454	42.0408
Kmer compl.	93.7882	93.7499
BUSCO sing.	98.0%	98.0%
BUSCO dupl.	0.5%	0.5%
BUSCO frag.	0.3%	0.3%
BUSCO miss.	1.3%	1.3%

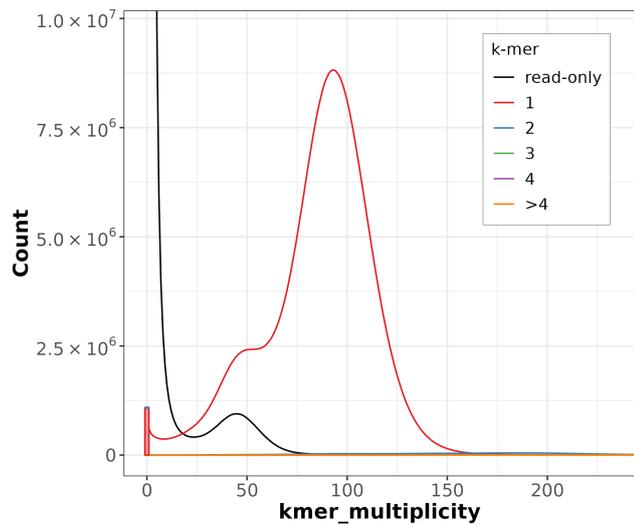
BUSCO: 6.0.0 (euk_genome_min, miniprot) / Lineage: actinopterygii_odb12 (genomes:75, BUSCOs:7207)

HiC contact map of curated assembly

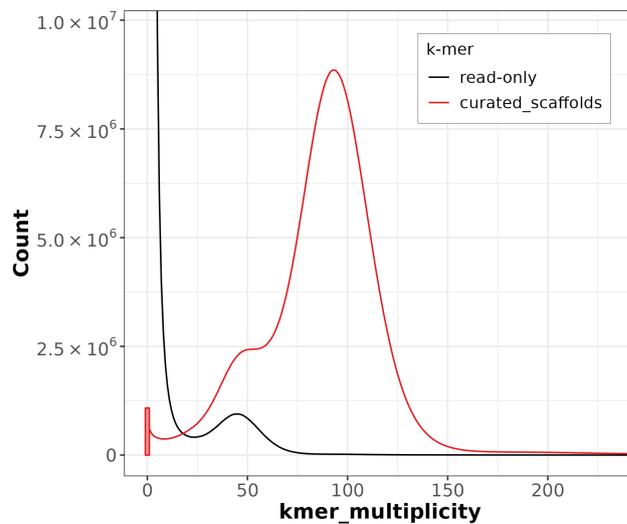


collapsed [\[LINK\]](#)

K-mer spectra of curated assembly

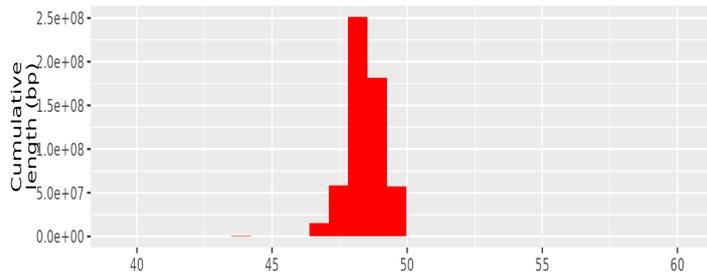


Distribution of k-mer counts per copy numbers found in asm

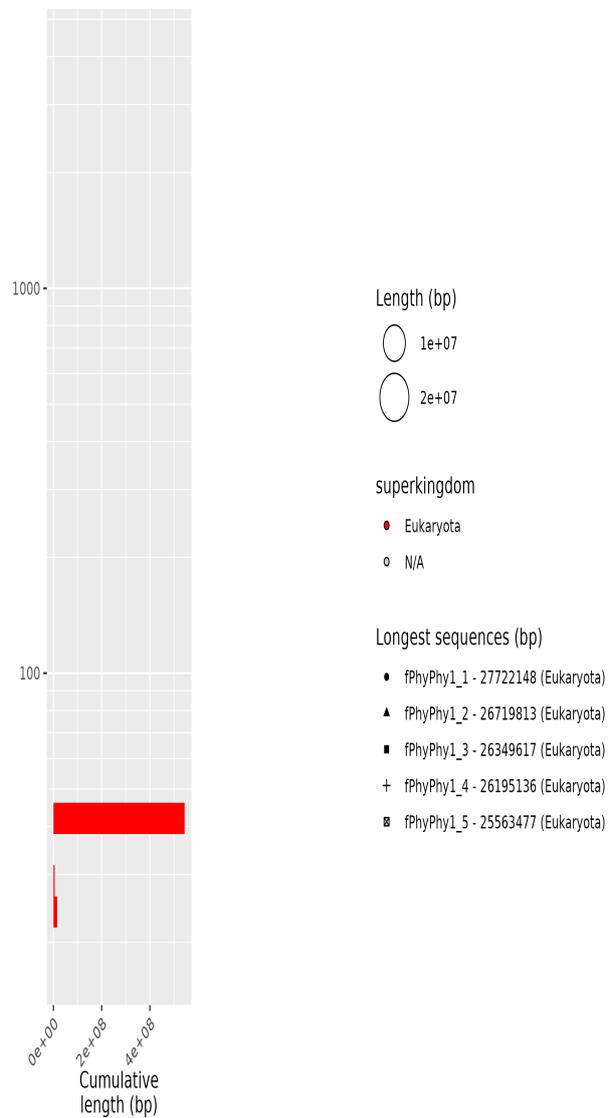
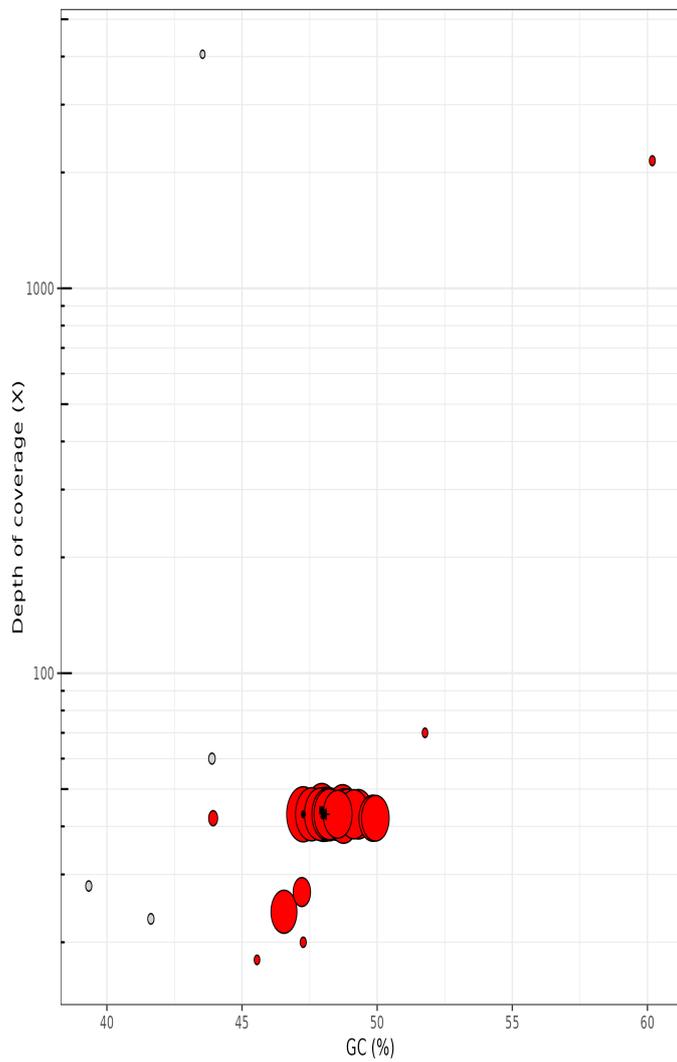


Distribution of k-mer counts coloured by their presence in reads/assemblies

Post-curation contamination screening



TAPAs summary Graph



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	50	221

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: 2026-01-30 15:44:11 CET