ERGA Assembly Report

v24.10.15

Tags: ATLASea[INVALID TAG]

TxID	151728
ToLID	**fGnaTho1**
Species	Gnatholepis thompsoni
Class	Actinopteri
Order	Gobiiformes

Genome Traits	Expected	Observed
Haploid size (bp)	1,220,778,809	1,211,077,838
Haploid Number	21 (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.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

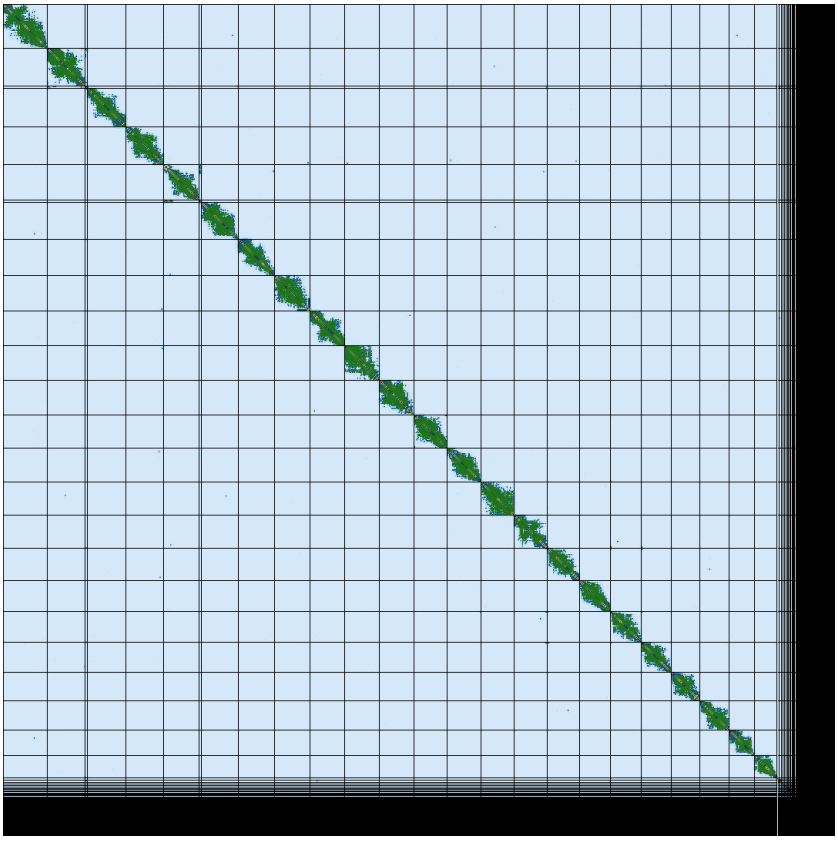
. Interventions/Gb: 13
. Contamination notes: ""
. Other observations: "The assembly of Gnatholepis thompsoni (fGnaTho1) is based on 64X PacBio data and 99X Arima Hi-C data generated as part of the ATLASea programme (https://www.atlasea.fr). The assembly process included the following steps: initial 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, 497 regions totaling 63.236 Mb (with the largest being 1.579 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. During manual curation, 5 haplotypic regions and 4 contaminant sequences were removed, totaling 2.393Mb and 0.389Mb, respectively (with the largest being 0.864Mb and 0.178Mb). 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	1,213,894,736	1,211,077,838
GC %	40.1	40.1
Gaps/Gbp	183.71	185.78
Total gap bp	22,300	23,200
Scaffolds	339	330
Scaffold N50	49,001,316	49,001,316
Scaffold L50	12	12
Scaffold L90	23	23
Contigs	562	555
Contig N50	35,102,583	35,102,583
Contig L50	14	14
Contig L90	73	72
QV	50.0478	50.0635
Kmer compl.	69.2313	69.1623
BUSCO sing.	96.3%	96.4%
BUSCO dupl.	1.0%	0.9%
BUSCO frag.	0.4%	0.4%
BUSCO miss.	2.4%	2.4%

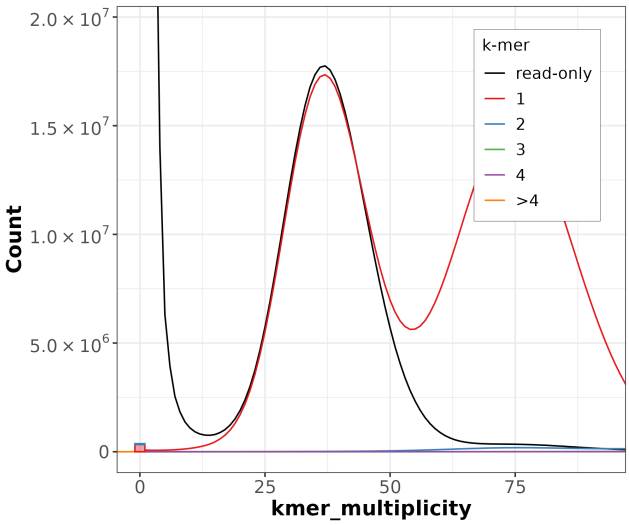
BUSCO: 6.0.0 (euk_genome_min, miniprot) / Lineage: actinopterygii_odb12 (genomes:75, BUSCOs:7207)

HiC contact map of curated assembly

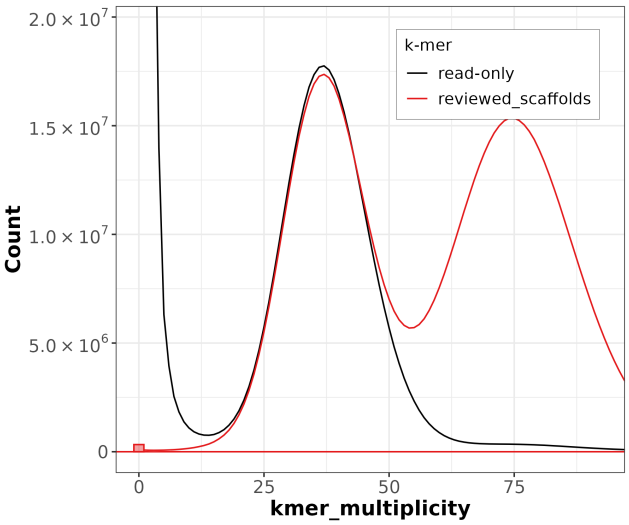


collapsed [LINK]

K-mer spectra of curated assembly

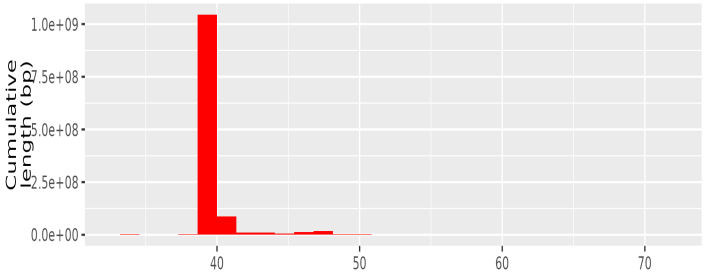


Distribution of k-mer counts per copy numbers found in asm

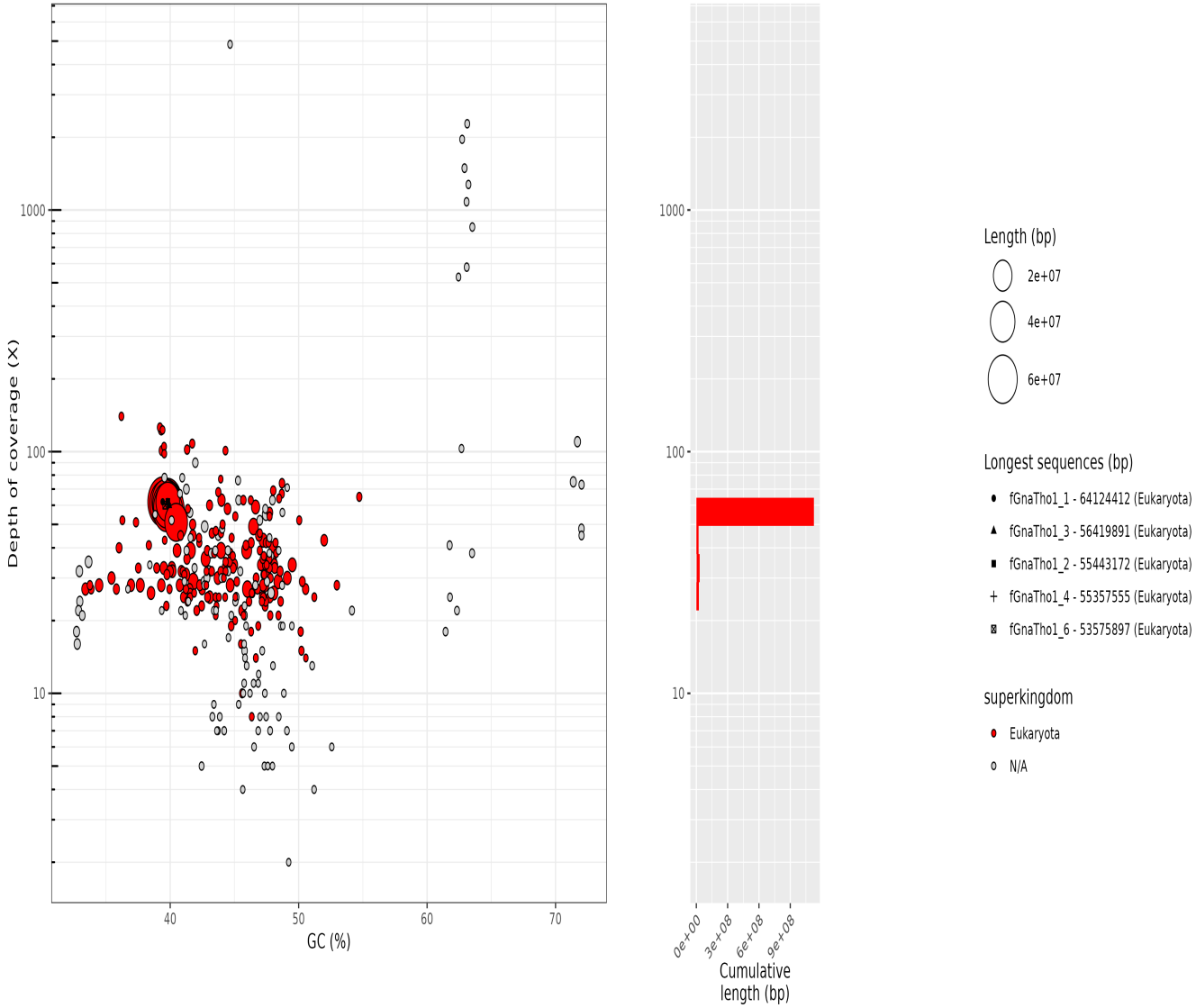


Distribution of k-mer counts coloured by their presence in reads/assemblies

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	64	100

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: Emilie Teodori
Affiliation: Genoscope

Date and time: 2026-05-03 13:21:11 CEST