ERGA Assembly Report

v24.10.15

Tags: ATLASea[INVALID TAG]

TxID	157241
ToLID	**fArnTho1**
Species	Arnoglossus thori
Class	Actinopteri
Order	Pleuronectiformes

Genome Traits	Expected	Observed
Haploid size (bp)	487,526,356	495,104,323
Haploid Number	19 (source: ancestor)	24
Ploidy	2 (source: ancestor)	2
Sample Sex	Unknown	Unknown

EBP metrics summary and curation notes

Obtained EBP quality metric for collapsed: 6.7.Q41

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
. Assembly length loss > 3% for collapsed

Curator notes

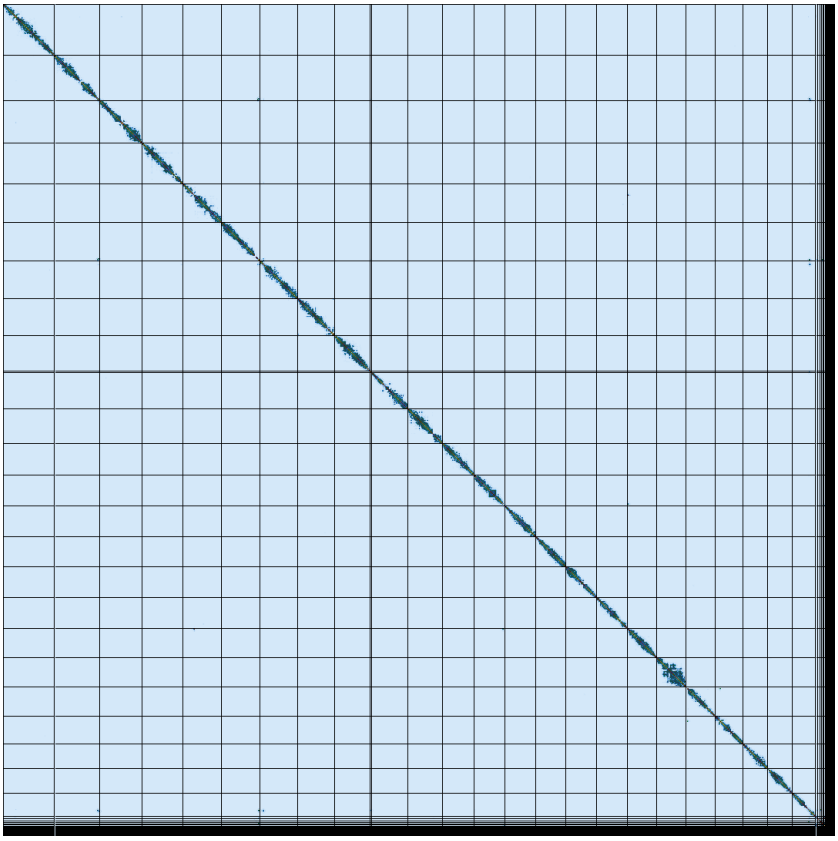
. Interventions/Gb: 130
. Contamination notes: ""
. Other observations: "The assembly of Arnoglossus thori (fArnTho1) is based on 76X ONT data and 90X Arima Hi-C data generated as part of the ATLASea programme (https://www.atlasea.fr). The assembly process included the following steps: initial ONT+HiC assembly generation with Hifiasm, removal of contaminant sequences using Context, and Hi-C-based scaffolding with YaHS of each haplotype. No contigs were found to be contaminants (bacterial, archaeal, or viral). Mitochondrial genome was assembled using ptGAUL. Finally, Both haplotypes were analyzed and manually improved using Pretext. During manual curation, 32 haplotypic regions were removed, totaling 9.646Mb (with the largest being 1.497Mb). 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	514,642,846	495,104,323
GC %	44.94	44.92
Gaps/Gbp	23.32	525.14
Total gap bp	1,200	29,400
Scaffolds	73	87
Scaffold N50	21,071,000	20,376,223
Scaffold L50	11	11
Scaffold L90	23	22
Contigs	85	347
Contig N50	16,132,065	3,013,070
Contig L50	14	54
Contig L90	31	168
QV	44.7334	41.3649
Kmer compl.	82.9655	81.2983
BUSCO sing.	97.4%	96.8%
BUSCO dupl.	1.6%	0.7%
BUSCO frag.	0.1%	0.2%
BUSCO miss.	0.9%	2.2%

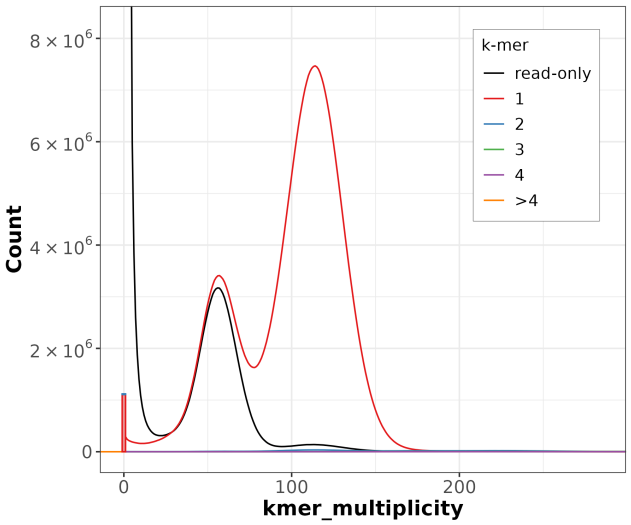
BUSCO: 6.0.0 (euk_genome_min, miniprot) / Lineage: actinopterygii_odb12 (genomes:75, BUSCOs:7207)

HiC contact map of curated assembly

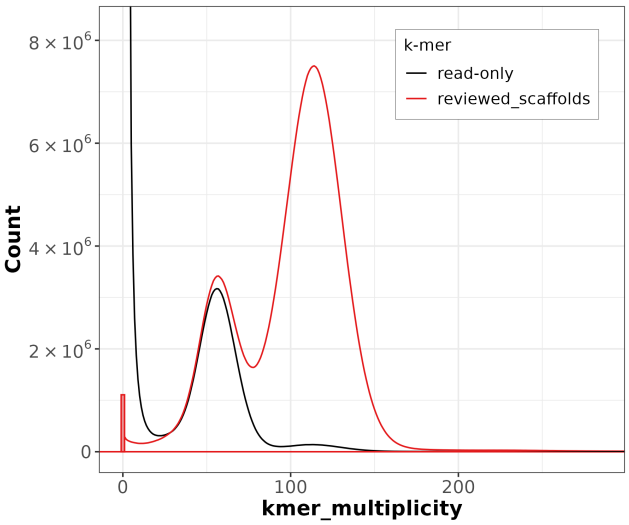


collapsed [LINK]

K-mer spectra of curated assembly

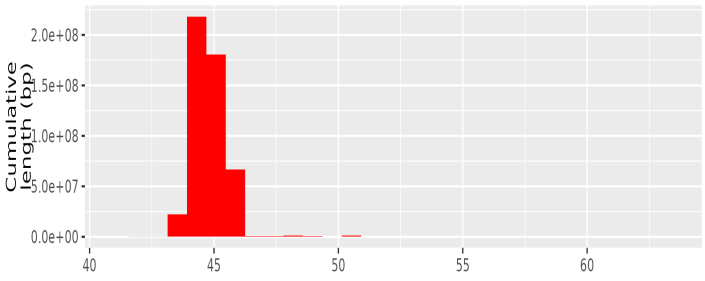


Distribution of k-mer counts per copy numbers found in asm



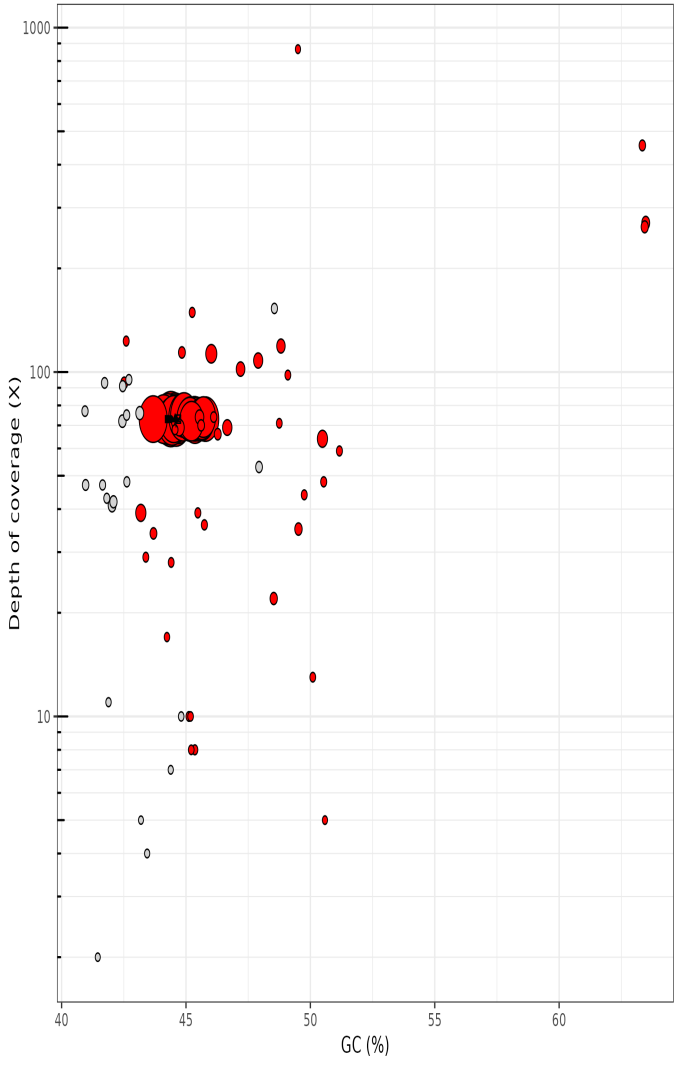
Distribution of k-mer counts coloured by their presence in reads/assemblies

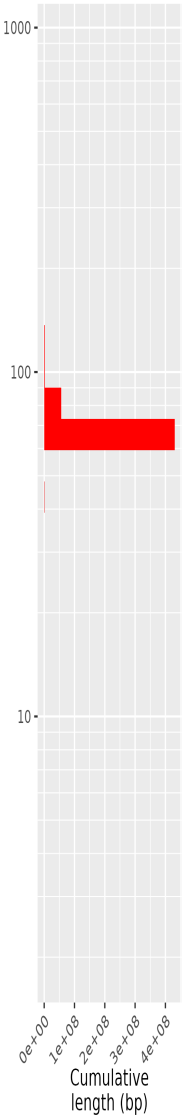
Post-curation contamination screening

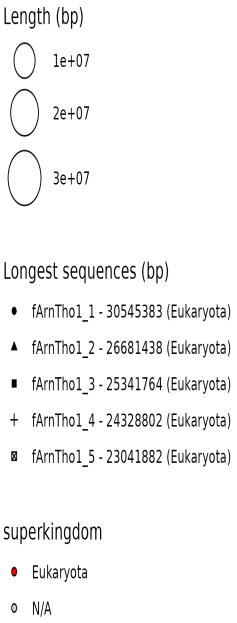


TAPAs summary Graph

(2 0X contigs have been hidden)







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	78	93

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: 2026-06-17 21:56:13 CEST