ERGA Assembly Report

v24.10.15

Tags: ATLASea[INVALID TAG]

TxID	208529
ToLID	**fArnLat1**
Species	Arnoglossus laterna
Class	Actinopteri
Order	Pleuronectiformes

Genome Traits	Expected	Observed
Haploid size (bp)	312,061,154	497,693,160
Haploid Number	19 (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.Q56

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

. Observed Haploid size (bp) has >20% difference with Expected
. Observed Haploid Number is different from Expected

. Kmer completeness value is less than 90 for collapsed
. Assembly length loss > 3% for collapsed
. More than 1000 gaps/Gbp for collapsed

Curator notes

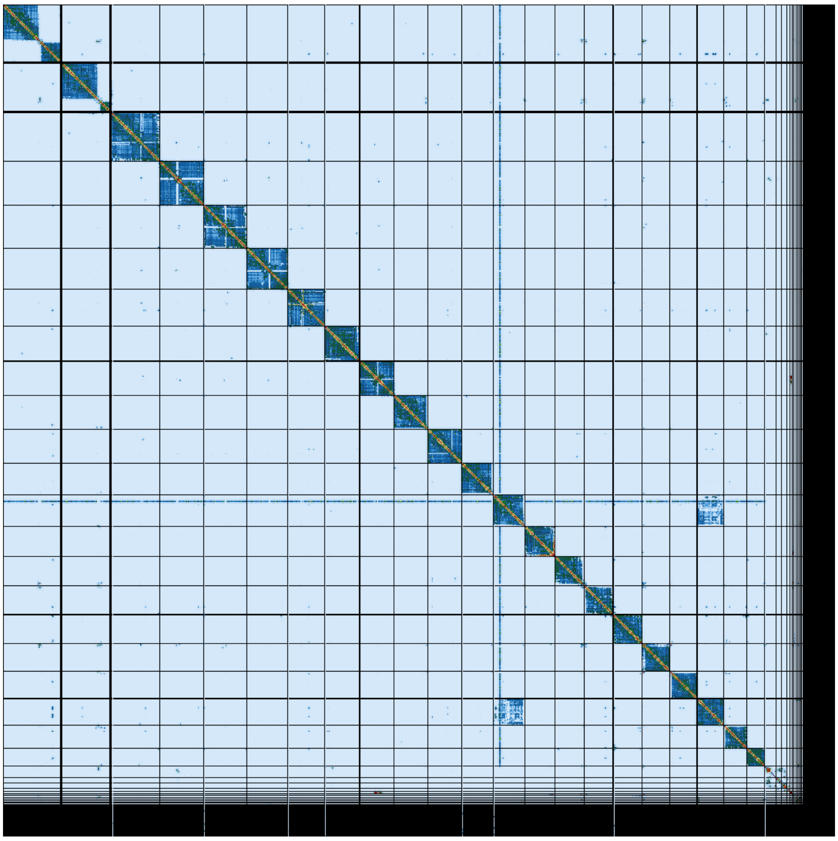
. Interventions/Gb: 177
. Contamination notes: ""
. Other observations: "The assembly of Arnoglossus laterna (fArnLat1.1) is based on 41X of PacBio data and 210X of Arima Hi-C data generated as part of the ATLASea programme (https://www.atlasea.fr). The assembly process included YaHS. In total, 8 contigs were identified as contaminants (bacterial), totaling 0.84 Mb (with the largest being 0.52 Mb). Additionally, 578 regions totaling 19.41 Mb (with the largest being 0.085 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, 261 contaminant sequences and 32 haplotypic regions were removed, totaling 13.87 Mb and 6.78 Mb (with the largest being 0.12 Mb and 2 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	518,604,327	497,693,160
GC %	44.79	45.45
Gaps/Gbp	1,920.54	2,073.57
Total gap bp	99,600	109,100
Scaffolds	719	368
Scaffold N50	18,764,940	20,332,259
Scaffold L50	12	10
Scaffold L90	31	22
Contigs	1,715	1,400
Contig N50	1,658,301	1,717,611
Contig L50	71	65
Contig L90	644	515
QV	35.2953	56.8311
Kmer compl.	79.1525	86.6887
BUSCO sing.	91.8%	93.1%
BUSCO dupl.	2.3%	1.0%
BUSCO frag.	1.7%	1.7%
BUSCO miss.	4.2%	4.2%

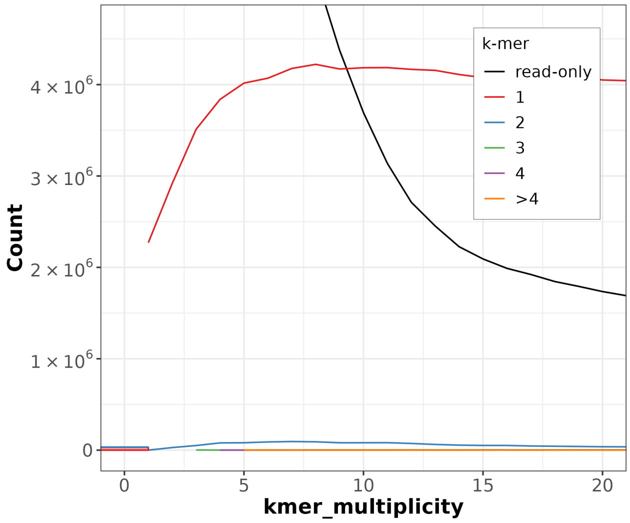
BUSCO: 5.8.2 (euk_genome_met, metaeuk) / Lineage: actinopterygii_odb12 (genomes:75, BUSCOs:7207)

HiC contact map of curated assembly

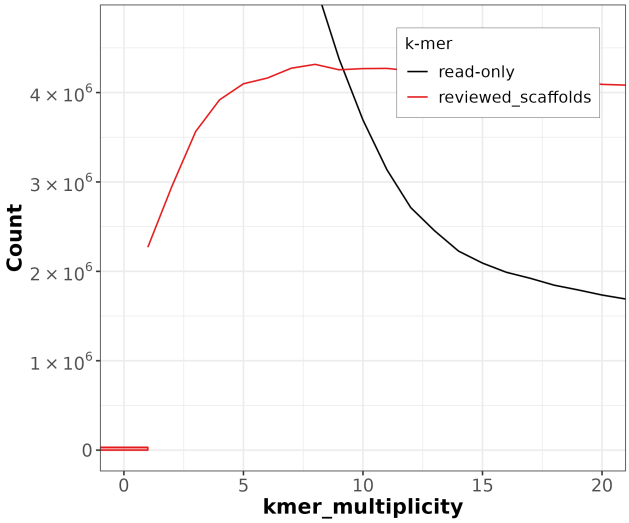


collapsed [LINK]

K-mer spectra of curated assembly

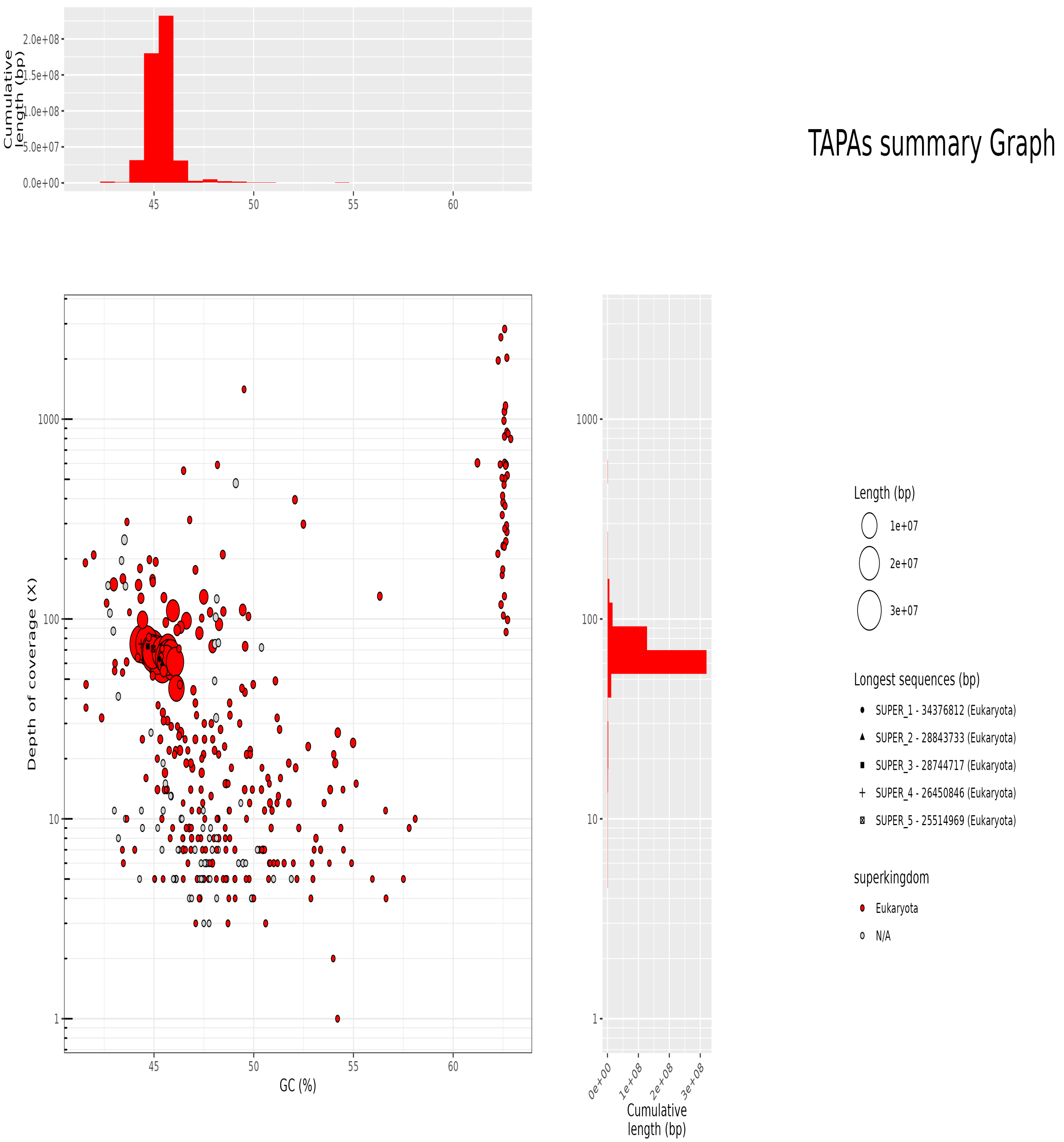


Distribution of k-mer counts per copy numbers found in asm



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

Post-curation contamination screening



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	70	210

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: Adama Ndar
Affiliation: Genoscope

Date and time: 2025-04-23 09:23:52 CEST