ERGA Assembly Report

v24.10.15

Tags: ATLASea[INVALID TAG]

TxID	195073
ToLID	**fAidSph1**
Species	Aidablennius sphynx
Class	Actinopteri
Order	Blenniiformes

Genome Traits	Expected	Observed
Haploid size (bp)	597,080,809	634,926,972
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: 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

. Assembly length loss > 3% for collapsed
. More than 1000 gaps/Gbp for collapsed

Curator notes

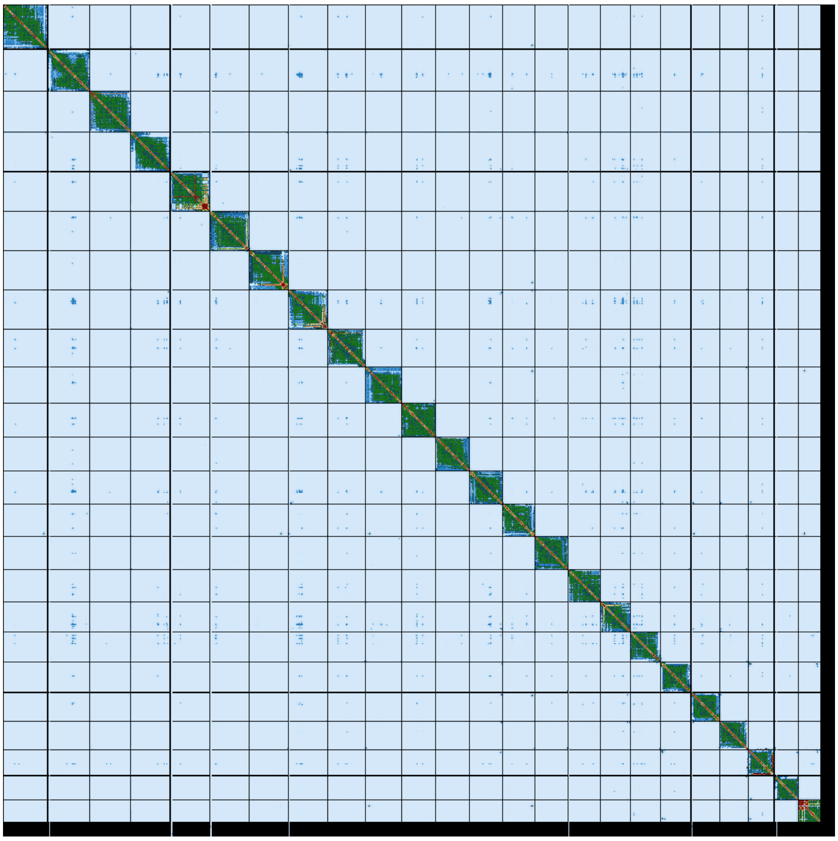
. Interventions/Gb: 505
. Contamination notes: ""
. Other observations: "The assembly of Aidablennius sphynx (fAidSph1) is based on 66X PacBio data and Arima Hi-C data generated as part of the ATLASea programme (https://www.atlasea.fr).The assembly process included the following steps: initial PacBio 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, 95 contigs were identified as contaminants (bacterial, archaeal, or viral), totaling 5.5 Mb (with the largest being 2.8Mb). Additionally, 1,299 regions totaling 33.5 Mb (with the largest being 0.218 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, 13 haplotypic regions were removed, totaling 1.7 Gb (with the largest being 0.49 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	697,351,828	634,926,972
GC %	42.17	42.4
Gaps/Gbp	837.45	1,222.19
Total gap bp	60,500	109,000
Scaffolds	2,920	465
Scaffold N50	24,382,185	26,178,920
Scaffold L50	13	11
Scaffold L90	302	22
Contigs	3,486	1,241
Contig N50	15,420,073	16,191,511
Contig L50	18	16
Contig L90	743	171
QV	32.7108	42.2892
Kmer compl.	92.4777	92.2316
BUSCO sing.	94.1%	97.5%
BUSCO dupl.	0.7%	0.8%
BUSCO frag.	2.0%	0.5%
BUSCO miss.	3.1%	1.2%

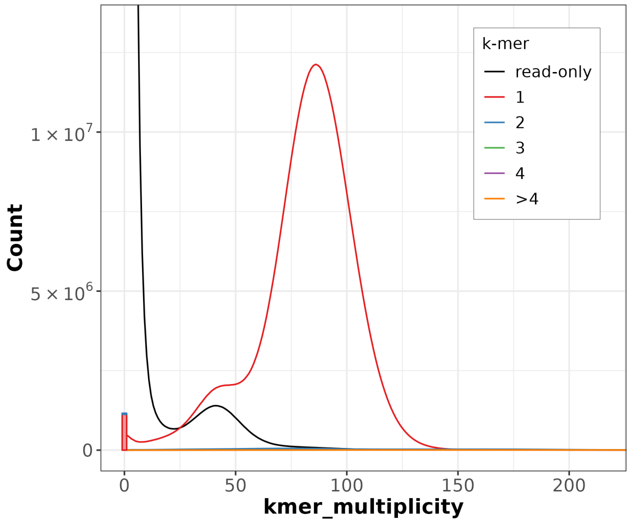
Warning! BUSCO versions or lineage datasets are not the same across results:
BUSCO: 5.8.2 (euk_genome_met, metaeuk) / Lineage: actinopterygii_odb12 (genomes:75, BUSCOs:7207)
BUSCO: 6.0.0 (euk_genome_min, miniprot) / Lineage: actinopterygii_odb12 (genomes:75, BUSCOs:7207)

HiC contact map of curated assembly

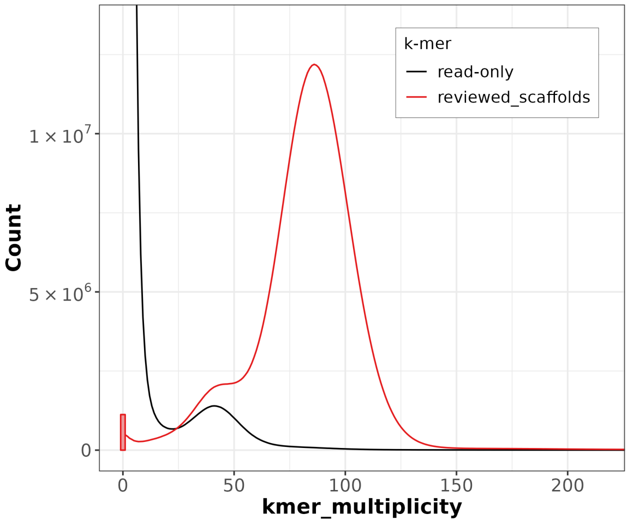


collapsed [LINK]

K-mer spectra of curated assembly

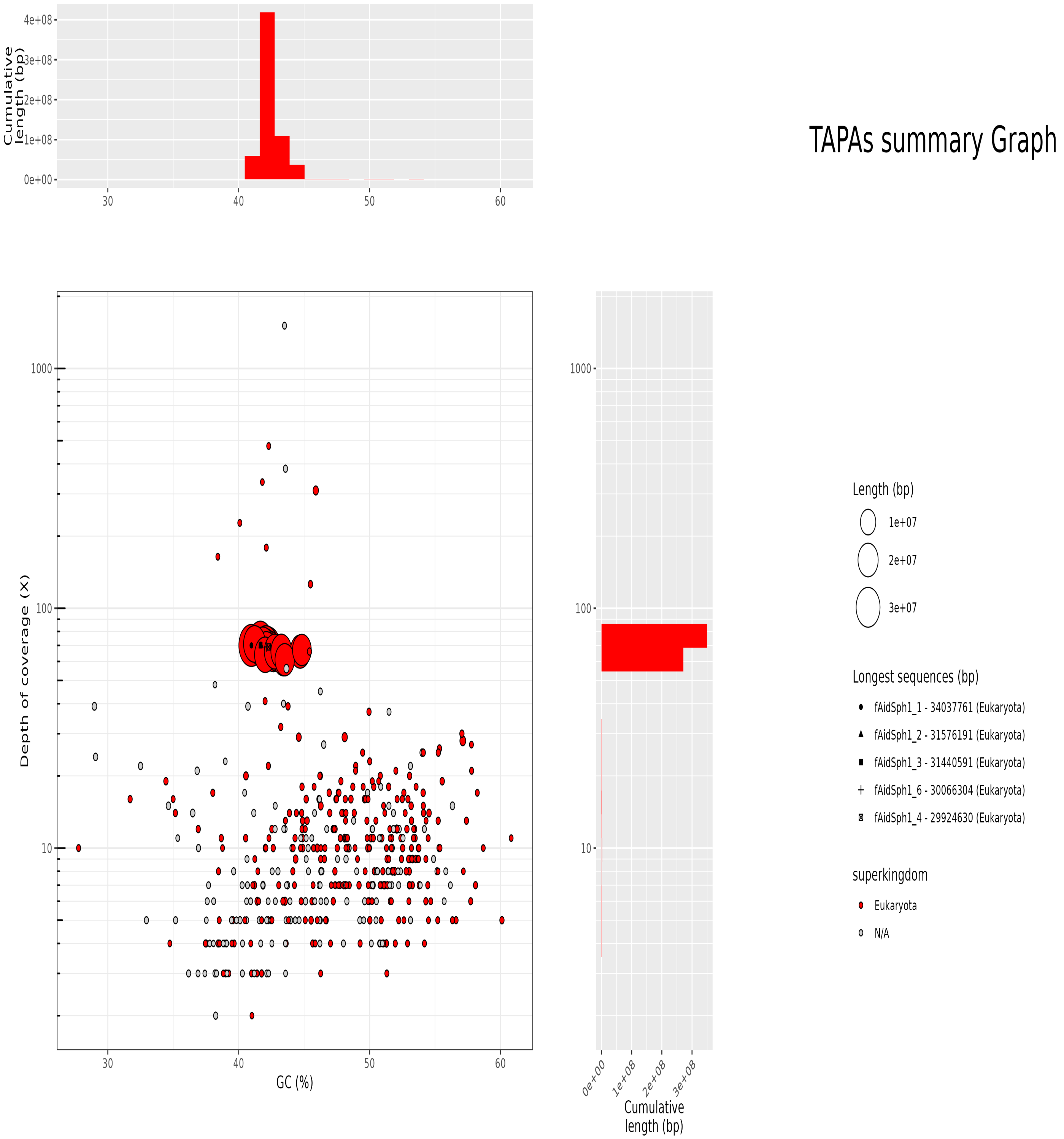


Distribution of k-mer counts per copy numbers found in asm



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

Post-curation contamination screening



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	66	170

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: Caroline Menguy
Affiliation: Genoscope

Date and time: 2025-10-31 21:32:23 CET