

ERGA Assembly Report

v24.10.15

Tags: ATLASEa[INVALID TAG]

TxID	206117
ToLID	fAp1Den1
Species	Apletodon dentatus
Class	Actinopteri
Order	Blenniiformes

Genome Traits	Expected	Observed
Haploid size (bp)	817,080,913	863,952,356
Haploid Number	19 (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.Q48

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

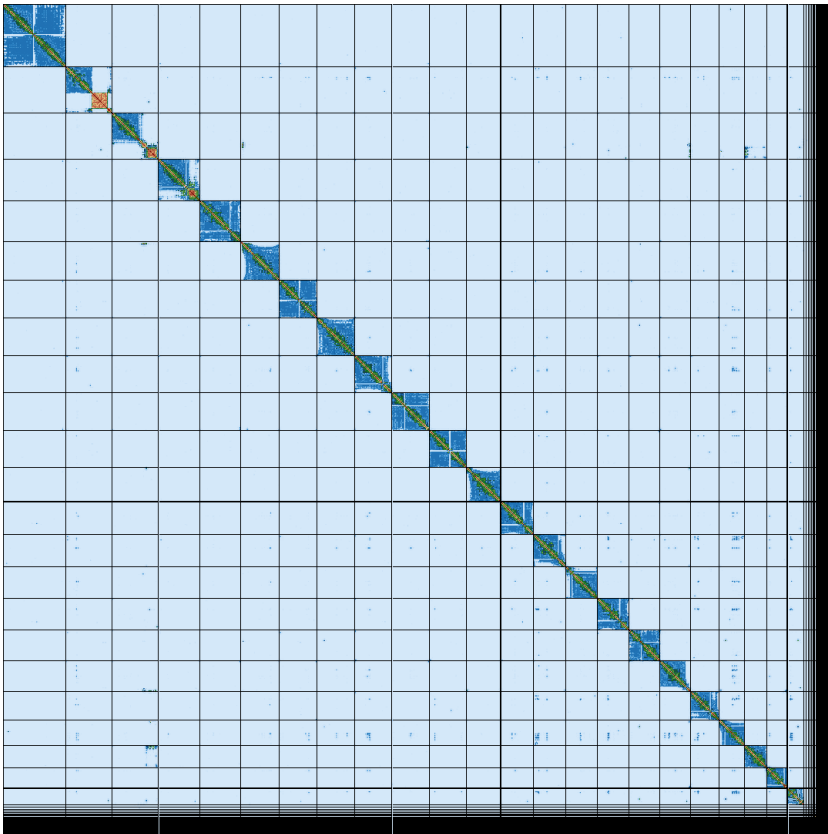
. Interventions/Gb: 37
. Contamination notes: ""
. Other observations: "The assembly of Apletodon dentatus (fAp1Den1) is based on 95X PacBio data and 164X Arima Hi-C data generated as part of the ATLASEa programme (<https://www.atlasea.fr>). The assembly process included the following steps: initial PacBio assembly generation with Hifiiasm, removal of contaminant sequences using Context, removal of haplotypic duplications using purge_dups, and Hi-C-based scaffolding with YaHS. In total, 48 contigs were identified as contaminants (bacterial, archaeal, or viral), totaling 2.807 Mb (with the largest being 0.594 Mb). Additionally, 129 regions totaling 48.169 Mb (with the largest being 10.697 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, 1 haplotypic region was removed, totaling 1.9Mb. 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	866,278,169	863,952,356
GC %	39.47	39.47
Gaps/Gbp	54.26	78.71
Total gap bp	4,700	9,500
Scaffolds	149	114
Scaffold N50	35,288,922	38,704,038
Scaffold L50	11	10
Scaffold L90	23	21
Contigs	196	182
Contig N50	20,711,000	20,711,000
Contig L50	17	17
Contig L90	48	47
QV	47.6856	48.4991
Kmer compl.	74.9706	74.964
BUSCO sing.	93.9%	93.9%
BUSCO dupl.	0.7%	0.7%
BUSCO frag.	1.5%	1.5%
BUSCO miss.	3.9%	3.9%

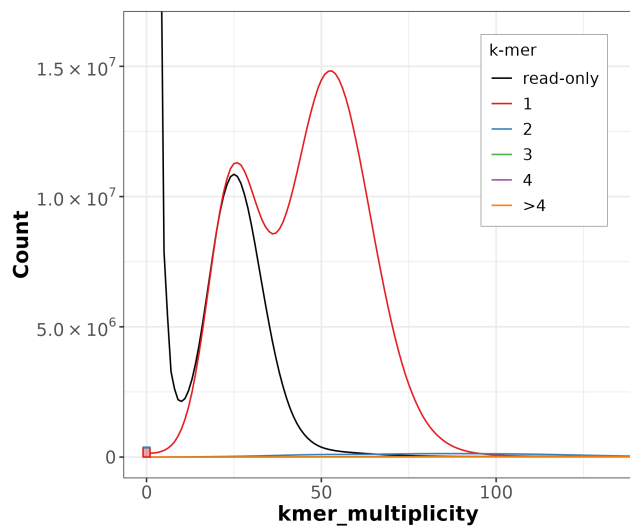
BUSCO: 5.8.2 (euk_genome_met, metaeuk) / Lineage: actinopterygii_odb12 (genomes:75, BUSCOs:7207)

HiC contact map of curated assembly

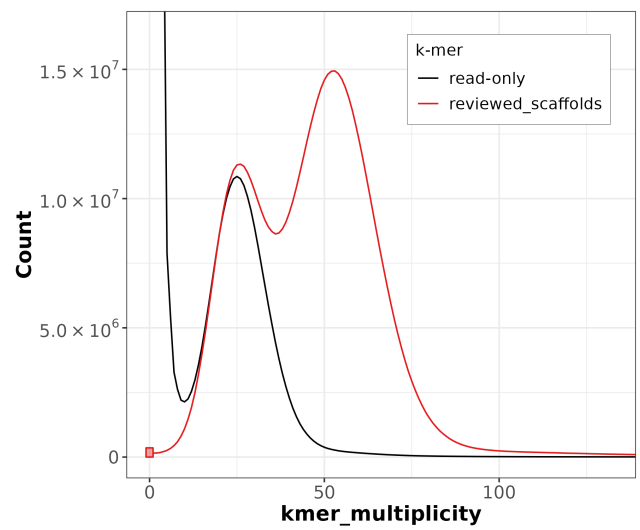


collapsed [\[LINK\]](#)

K-mer spectra of curated assembly

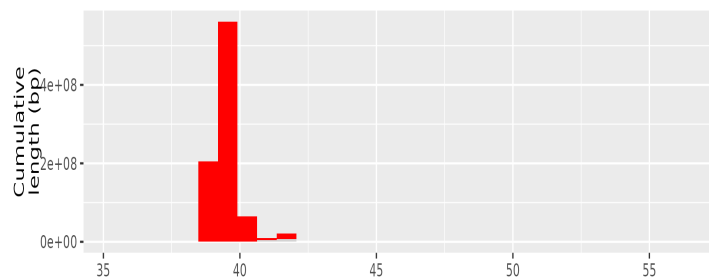


Distribution of k-mer counts per copy numbers found in asm

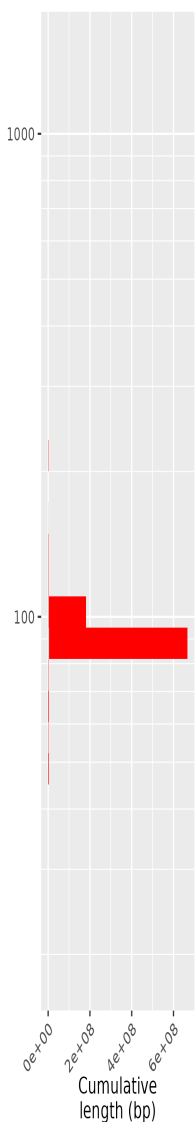
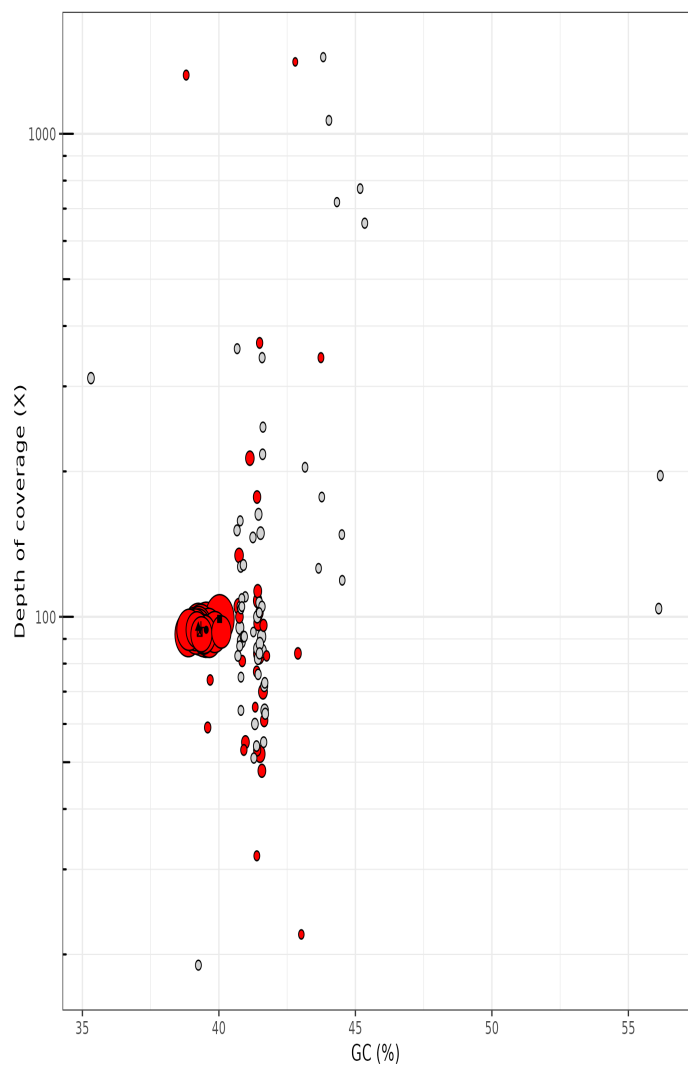


Distribution of k-mer counts coloured by their presence in reads/assemblies

Post-curation contamination screening



TAPAs summary Graph



Length (bp)

2e+07

4e+07

6e+07

Longest sequences (bp)

• SUPER_1 - 65034855 (Eukaryota)

▲ SUPER_2 - 47951832 (Eukaryota)

■ SUPER_3 - 47888904 (Eukaryota)

+ SUPER_4 - 43233548 (Eukaryota)

▣ SUPER_5 - 42652503 (Eukaryota)

superkingdom

• Eukaryota

○ N/A

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	95	164

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: 2025-06-11 09:12:29 CEST