ERGA Assembly Report

v24.10.15

Tags: ATLASea[INVALID TAG]

TxID	57788
ToLID	**fCliArg1**
Species	Clinitrachus argentatus
Class	Actinopteri
Order	Blenniiformes

Genome Traits	Expected	Observed
Haploid size (bp)	258,475,267	574,115,187
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.Q45

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

Curator notes

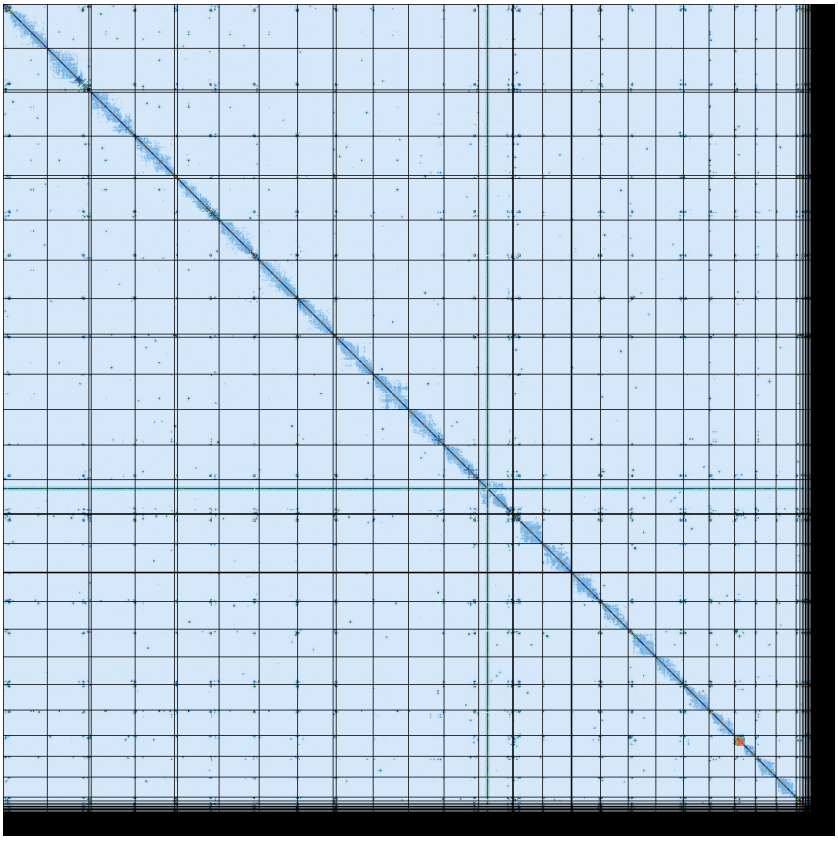
. Interventions/Gb: 16
. Contamination notes: ""
. Other observations: "The assembly of Clinitrachus argentatus (fCliArg1) is based on 29.04X PacBio data and 148.25X Arima Hi-C data generated as part of the ATLASea programme (https://www.atlasea.fr). The assembly process included the following steps: 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, 3 contigs were identified as contaminants (bacterial, archaeal, or viral), totaling 0.127 Mb (with the largest being 0.06 Mb). Additionally, 249 regions totaling 17.024 Mb (with the largest being 0.411 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, 3 haplotypic regions were removed, totaling 0.574Mb (with the largest being 0.257Mb). The telomeric pattern AACCCT was identified with TelFinder and used to generate the Pretext Telomeres tracks. 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	579,371,582	574,115,187
GC %	45.03	45.05
Gaps/Gbp	771.53	801.23
Total gap bp	44,700	48,800
Scaffolds	222	206
Scaffold N50	24,602,090	24,186,005
Scaffold L50	11	11
Scaffold L90	26	23
Contigs	669	666
Contig N50	3,301,000	3,301,000
Contig L50	48	47
Contig L90	227	226
QV	45.5528	45.5353
Kmer compl.	95.0231	94.9513
BUSCO sing.	98.1%	98.6%
BUSCO dupl.	0.8%	0.3%
BUSCO frag.	0.2%	0.2%
BUSCO miss.	0.8%	0.9%

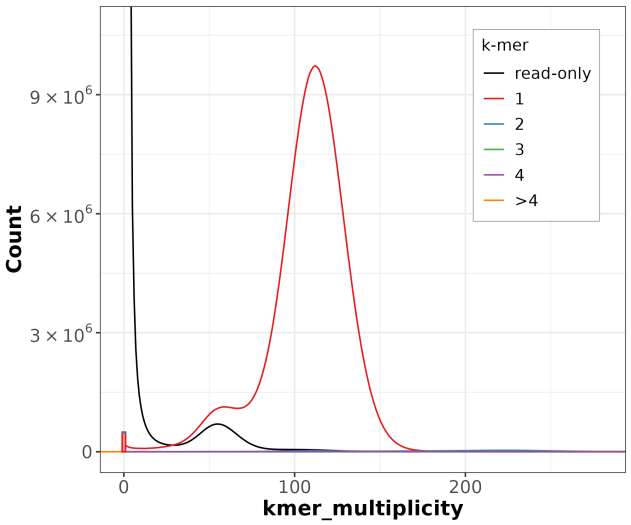
Warning! BUSCO versions or lineage datasets are not the same across results:
BUSCO: 6.0.0 (euk_genome_min, miniprot) / Lineage: actinopterygii_odb12 (genomes:75, BUSCOs:7207)
BUSCO: 6.1.0 (euk_genome_min, miniprot) / Lineage: actinopterygii_odb12 (genomes:75, BUSCOs:7207)

HiC contact map of curated assembly

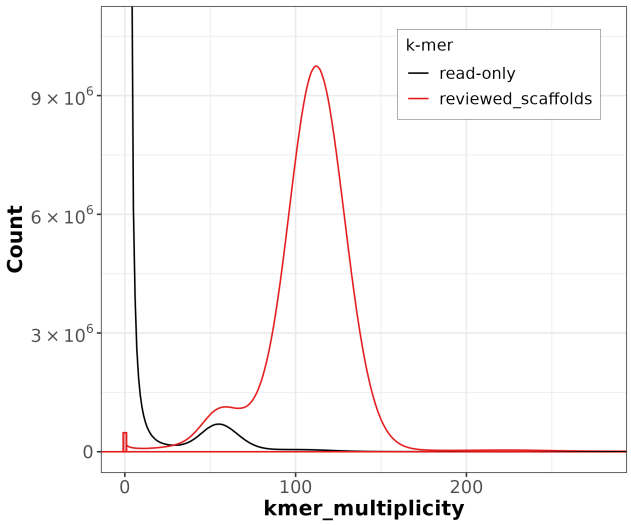


collapsed [LINK]

K-mer spectra of curated assembly

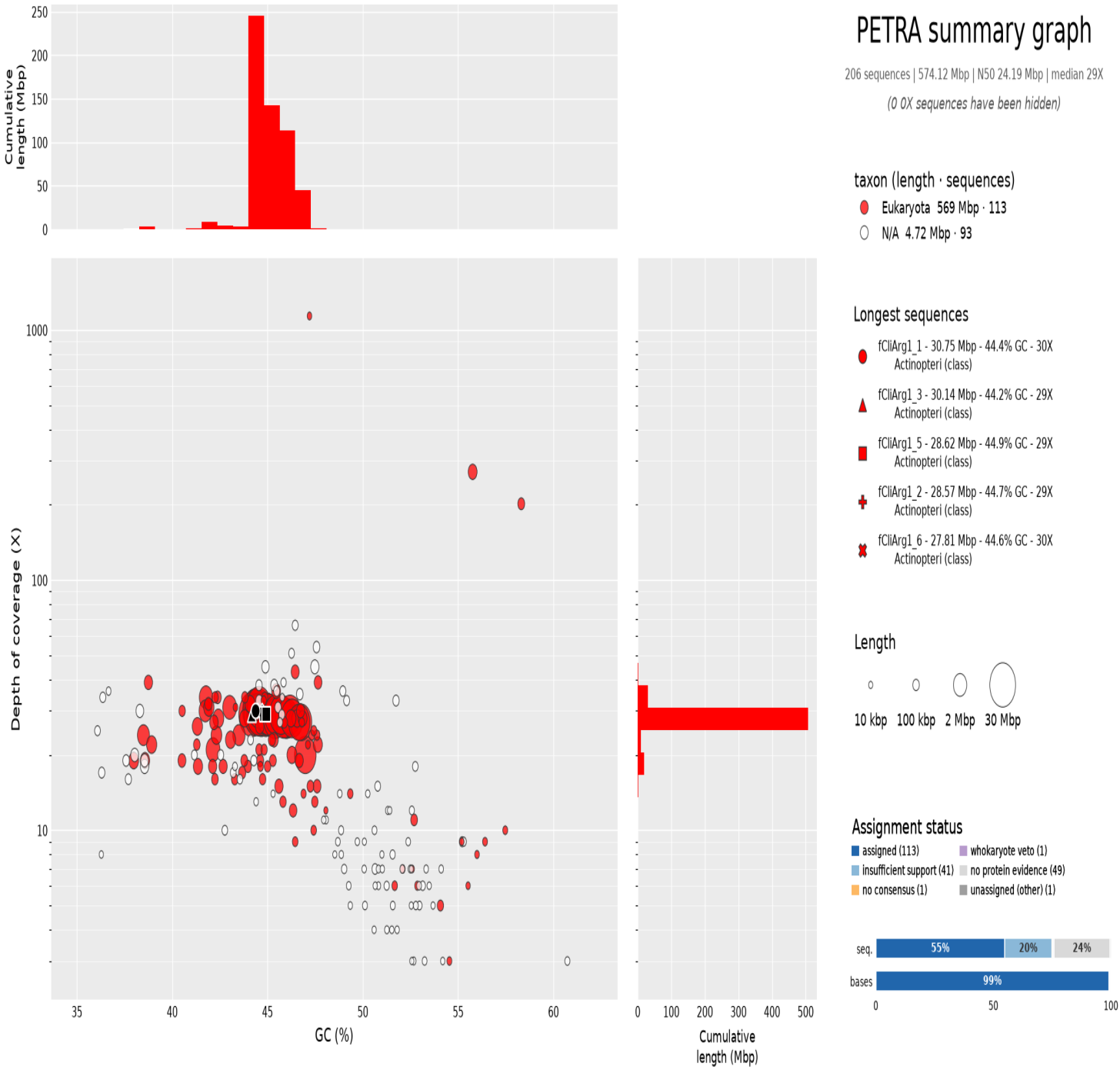


Distribution of k-mer counts per copy numbers found in asm



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	29	149

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:* 0.2.4
  |_ *key param:* NA
- **PretextView**
  |_ *ver:* 1.0.7
  |_ *key param:* NA

Submitter: Eleanore Lacoste
Affiliation: Genoscope

Date and time: 2026-10-07 11:02:31 CEST