ERGA Assembly Report

v24.10.15

Tags: Other[INVALID TAG]

TxID	7917
ToLID	**fAtrSpa1_combined**
Species	Alligator gar
Class	Actinopteri
Order	Semionotiformes

Genome Traits	Expected	Observed
Haploid size (bp)	509,031,378	2,323,214,533
Haploid Number	34 (source: ancestor)	28
Ploidy	2 (source: ancestor)	2
Sample Sex	Unknown	Unknown

EBP metrics summary and curation notes

Obtained EBP quality metric for collapsed: 7.7.Q61

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

. BUSCO single copy value is less than 90% for collapsed
. BUSCO duplicated value is more than 5% for collapsed
. Not 90% of assembly in chromosomes for collapsed

Curator notes

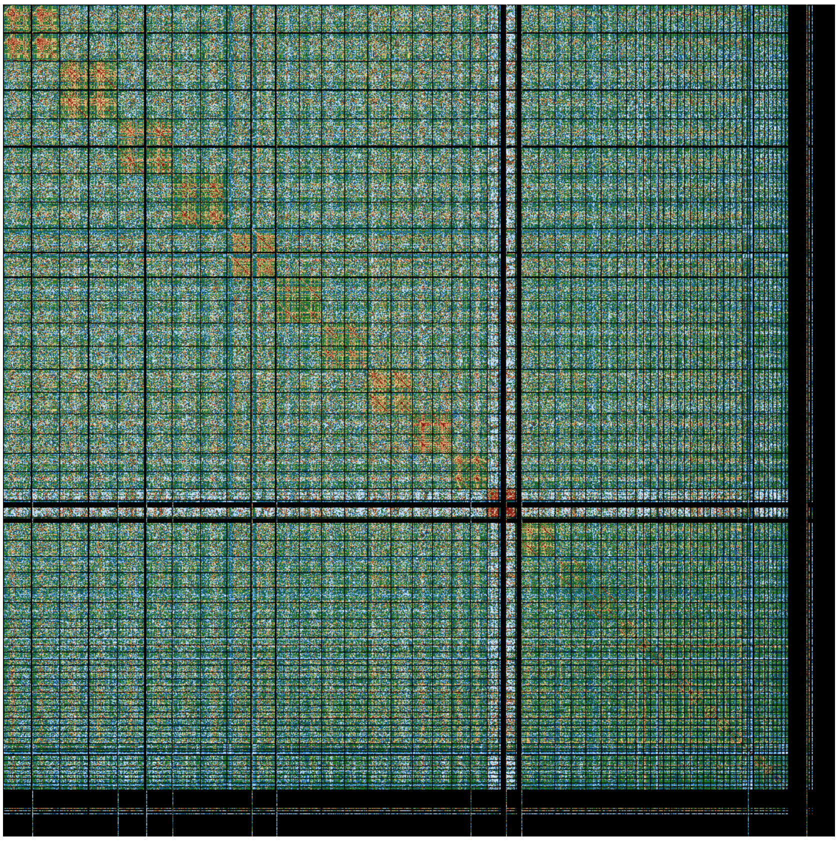
. Interventions/Gb: 31
. Contamination notes: ""
. Other observations: "The assembly of Alligator gar (fAtrSpa1) has been generated as part of the Vertebrate Genomes Project (https://vertebrategenomesproject.org/). Both haplotypes were analyzedand manually improved using Pretext. "

Quality metrics table

Metrics	Pre-curation collapsed	Curated collapsed
Total bp	2,323,190,435	2,323,214,533
GC %	40.59	40.59
Gaps/Gbp	46.06	62.84
Total gap bp	21,400	29,200
Scaffolds	626	589
Scaffold N50	55,314,901	55,314,901
Scaffold L50	17	17
Scaffold L90	60	52
Contigs	733	735
Contig N50	24,401,090	25,314,049
Contig L50	34	32
Contig L90	116	115
QV	61.5883	61.5812
Kmer compl.	99.4264	99.4282
BUSCO sing.	2.6%	2.0%
BUSCO dupl.	87.8%	95.9%
BUSCO frag.	3.5%	0.5%
BUSCO miss.	6.1%	1.6%

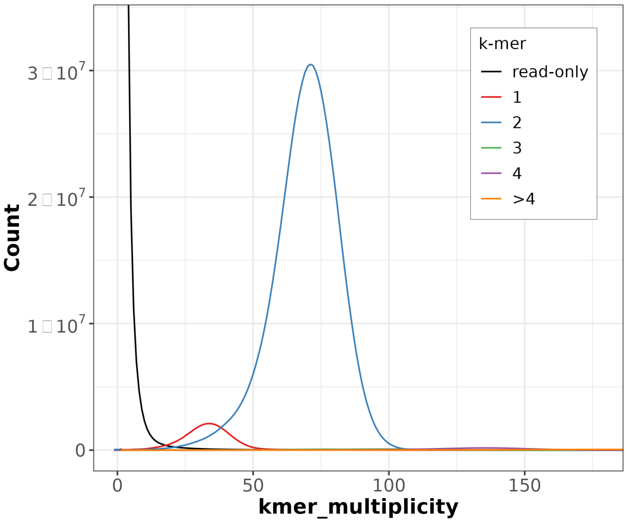
Warning! BUSCO versions or lineage datasets are not the same across results:
BUSCO: 5.8.2 (euk_genome_met, metaeuk) / Lineage: actinopterygii_odb12 (genomes:75, BUSCOs:7207)
BUSCO: 6.0.0 (euk_genome_min, miniprot) / Lineage: actinopterygii_odb12 (genomes:75, BUSCOs:7207)

HiC contact map of curated assembly

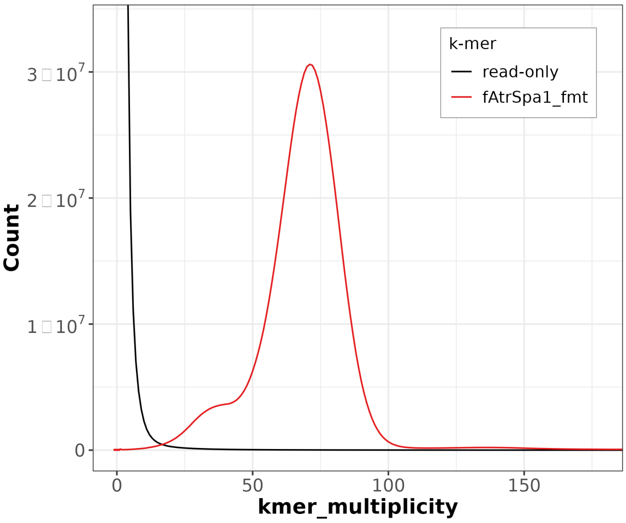


collapsed [LINK]

K-mer spectra of curated assembly

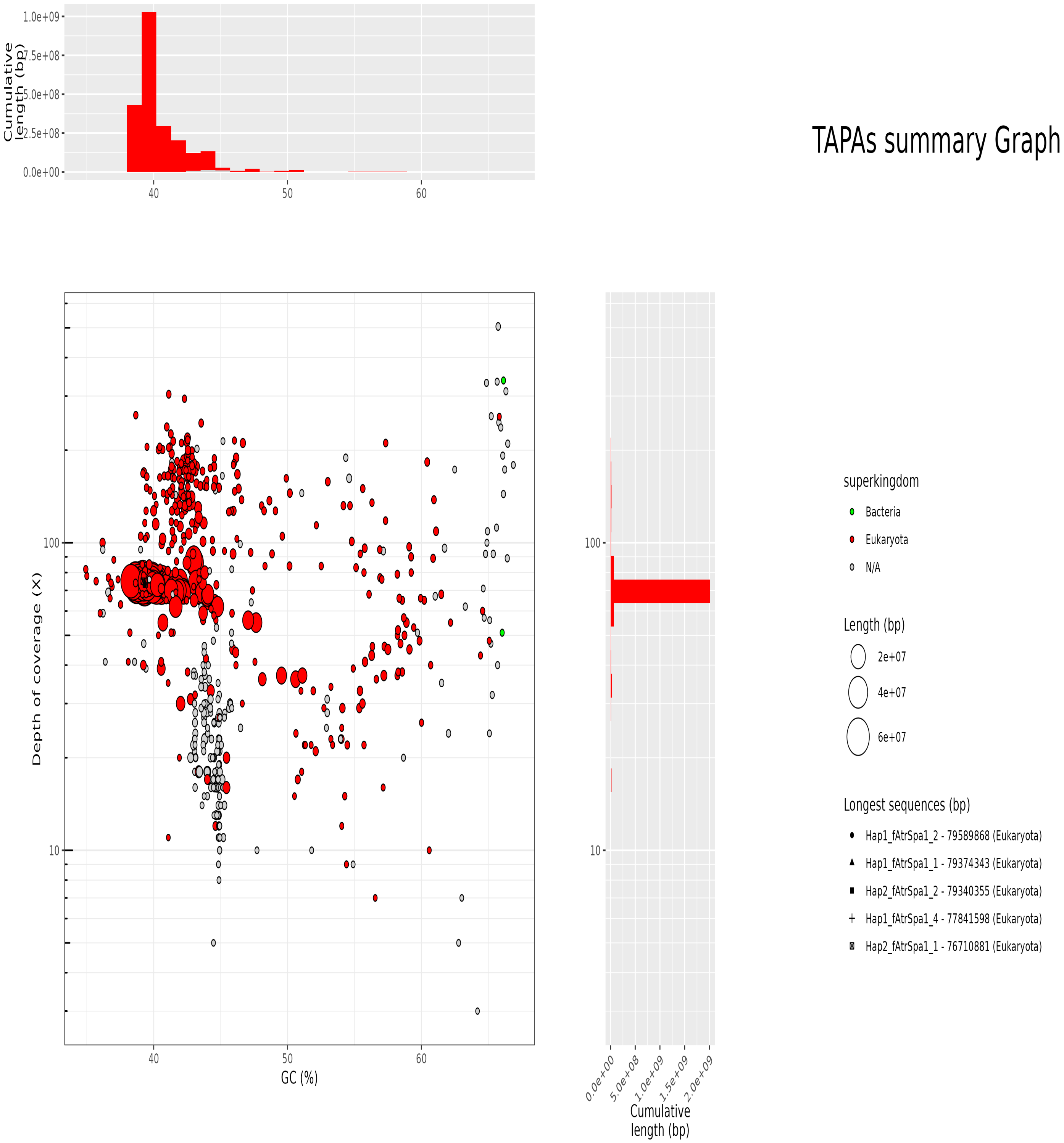


Distribution of k-mer counts per copy numbers found in asm



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	NA	NA

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: Benjamin Istace
Affiliation: Genoscope

Date and time: 2025-10-20 22:36:43 CEST