ERGA Assembly Report

v24.10.15

Tags: Other[INVALID TAG]

TxID	7917
ToLID	**fAtrSpa1_hap2**
Species	Alligator gar
Class	Actinopteri
Order	Semionotiformes

Genome Traits	Expected	Observed
Haploid size (bp)	509,031,378	1,146,314,777
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
. Assembly length loss > 3% for collapsed

Curator notes

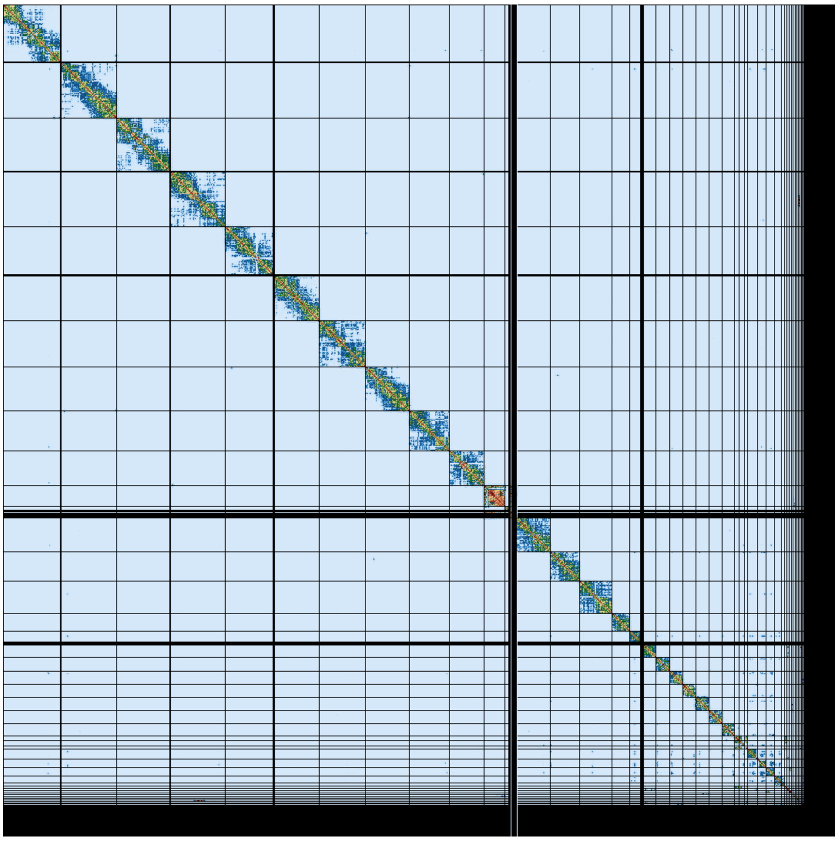
. Interventions/Gb: 31
. Contamination notes: ""
. Other observations: "The assembly of Alligator gar (fAtrSpa1) has been generated as part of the Vertebrate Genomes Project (https://vertebrategenomesproject.org/). Both haplotypes were analyzed and manually improved using Pretext. "

Quality metrics table

Metrics	Pre-curation collapsed	Curated collapsed
Total bp	1,199,401,584	1,146,314,777
GC %	40.69	40.53
Gaps/Gbp	35.02	65.43
Total gap bp	8,400	15,000
Scaffolds	281	236
Scaffold N50	54,681,828	54,681,828
Scaffold L50	9	9
Scaffold L90	33	28
Contigs	323	311
Contig N50	23,006,130	25,914,056
Contig L50	18	16
Contig L90	60	63
QV	61.4239	61.7723
Kmer compl.	97.553	96.955
BUSCO sing.	89.4%	96.1%
BUSCO dupl.	0.7%	0.5%
BUSCO frag.	3.6%	0.6%
BUSCO miss.	6.3%	2.8%

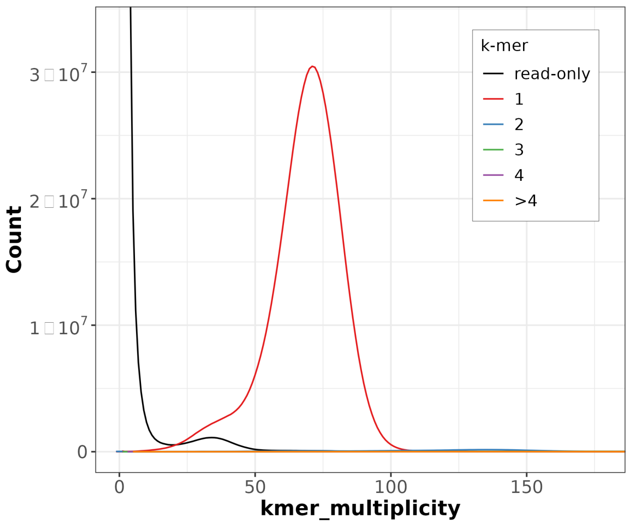
Warning! BUSCO versions or lineage datasets are not the same across results:
BUSCO: 5.8.2 (euk_genome_met, metaeuk) / Lineage: actinopterygii_odb12 (genomes:75, BUSCOs:7207)
BUSCO: 6.0.0 (euk_genome_min, miniprot) / Lineage: actinopterygii_odb12 (genomes:75, BUSCOs:7207)

HiC contact map of curated assembly

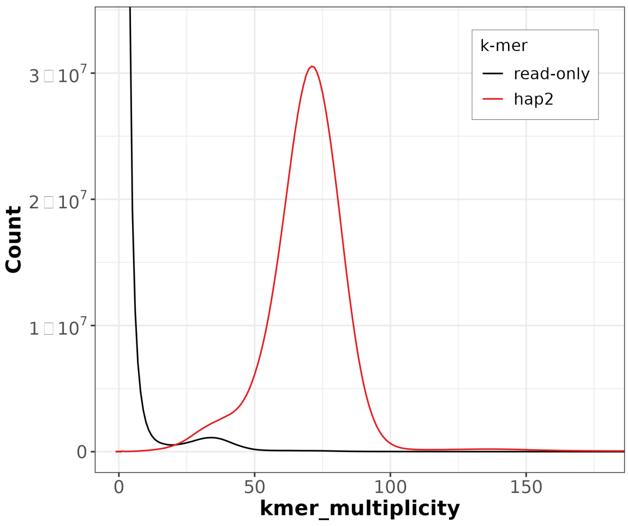


collapsed [LINK]

K-mer spectra of curated assembly

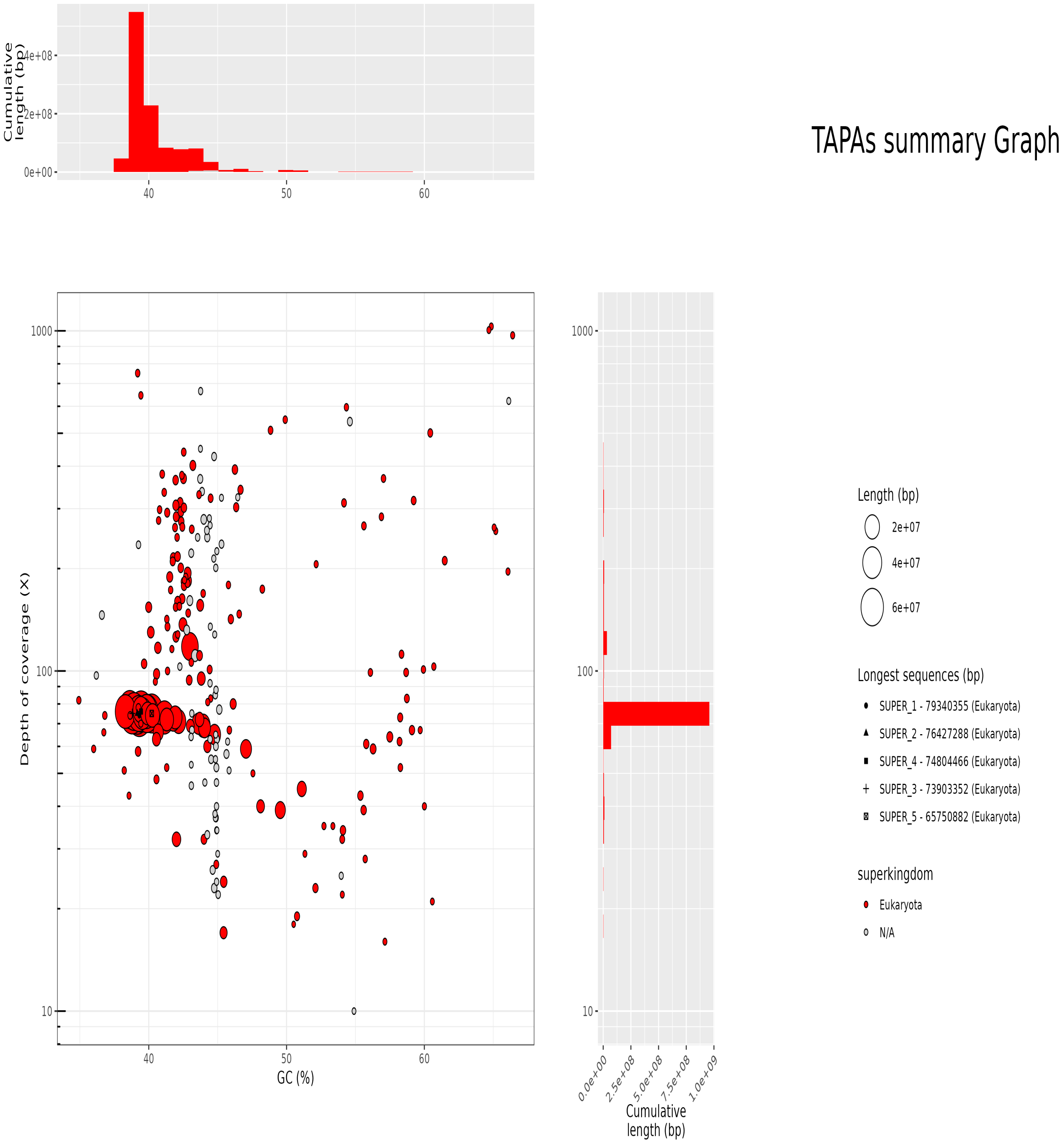


Distribution of k-mer counts per copy numbers found in asm



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: 2025-11-05 02:30:38 CET