ERGA Assembly Report

v24.10.15

Tags: ATLASea[INVALID TAG]

TxID	387458
ToLID	**xgStrGall1**
Species	Aliger gallus
Class	Gastropoda
Order	Littorinimorpha

Genome Traits	Expected	Observed
Haploid size (bp)	1,661,707,610	1,877,661,135
Haploid Number	17 (source: ancestor)	34
Ploidy	2 (source: ancestor)	2
Sample Sex	Unknown	Unknown

EBP metrics summary and curation notes

Obtained EBP quality metric for collapsed: 6.7.Q39

The following metrics were automatically flagged as below EBP recommended standards or different from expected:

. Observed Haploid Number is different from Expected

. QV value is less than 40 for collapsed
. Kmer completeness value is less than 90 for collapsed
. BUSCO single copy value is less than 90% for collapsed
. BUSCO duplicated value is more than 5% for collapsed

Curator notes

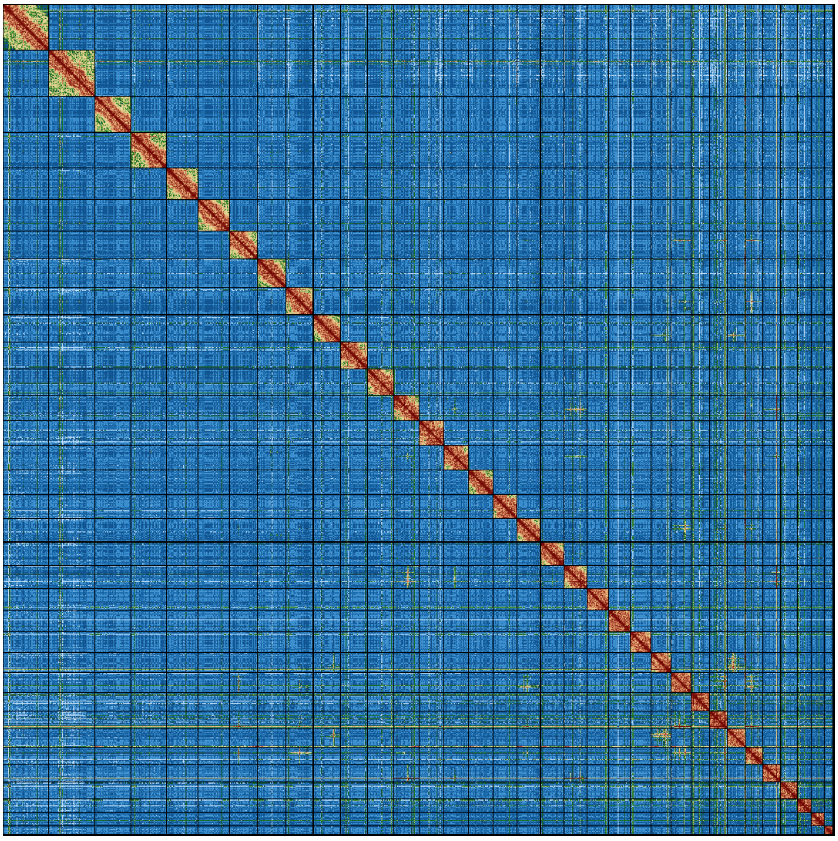
. Interventions/Gb: 33
. Contamination notes: ""
. Other observations: "The assembly of Aliger gallus (xgStrGall1) is based on 44.6X ONT data and 221X Arima Hi-C data generated as part of the ATLASea programme (https://www.atlasea.fr). The assembly process included the followig steps: initial ONT assembly 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 17.839 Mb (with the largest being 13.597 Mb). Additionally, 844 regions totaling 104.33 Mb (with the largest being 1.071 M) were identified as haplotypic duplications and removed. Mitochondrial genome was assembled using mitoHifi. During manual curation, 8 haplotypic regions were removed, totaling 4.05Mb (with the largest being1.2Mb).. 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	1,881,734,830	1,877,661,135
GC %	46.04	46.04
Gaps/Gbp	365.09	373.87
Total gap bp	68,700	73,800
Scaffolds	132	109
Scaffold N50	56,282,416	56,986,569
Scaffold L50	14	13
Scaffold L90	29	29
Contigs	819	811
Contig N50	4,056,144	4,086,803
Contig L50	138	137
Contig L90	426	424
QV	39.8332	39.8339
Kmer compl.	85.8546	85.8279
BUSCO sing.	78.9%	79.0%
BUSCO dupl.	18.6%	18.5%
BUSCO frag.	0.9%	0.8%
BUSCO miss.	1.6%	1.6%

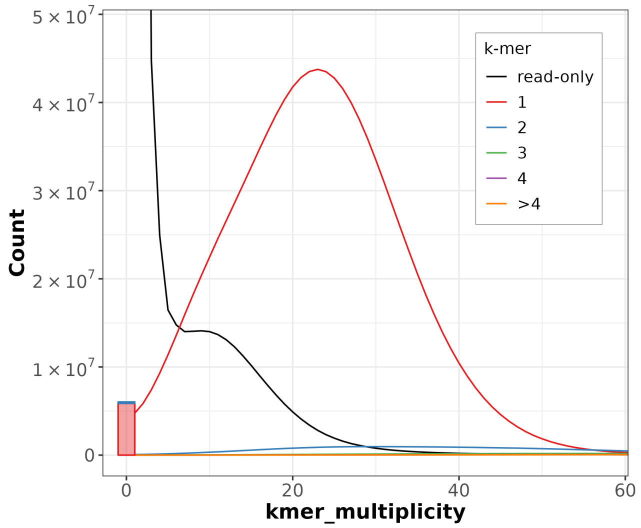
BUSCO: 6.0.0 (euk_genome_min, miniprot) / Lineage: mollusca_odb12 (genomes:36, BUSCOs:4421)

HiC contact map of curated assembly

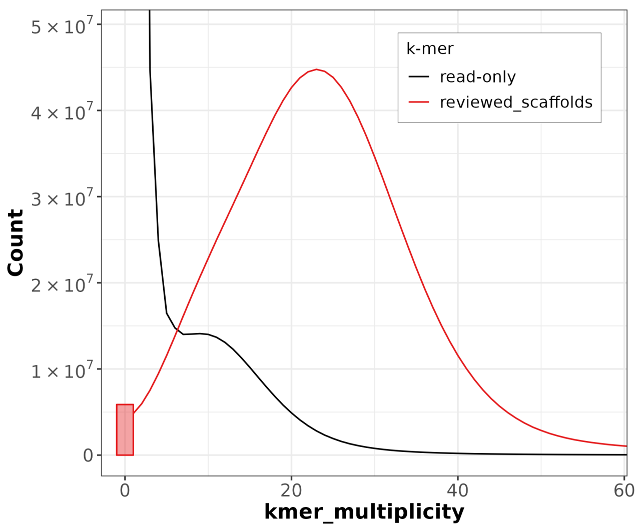


collapsed [LINK]

K-mer spectra of curated assembly

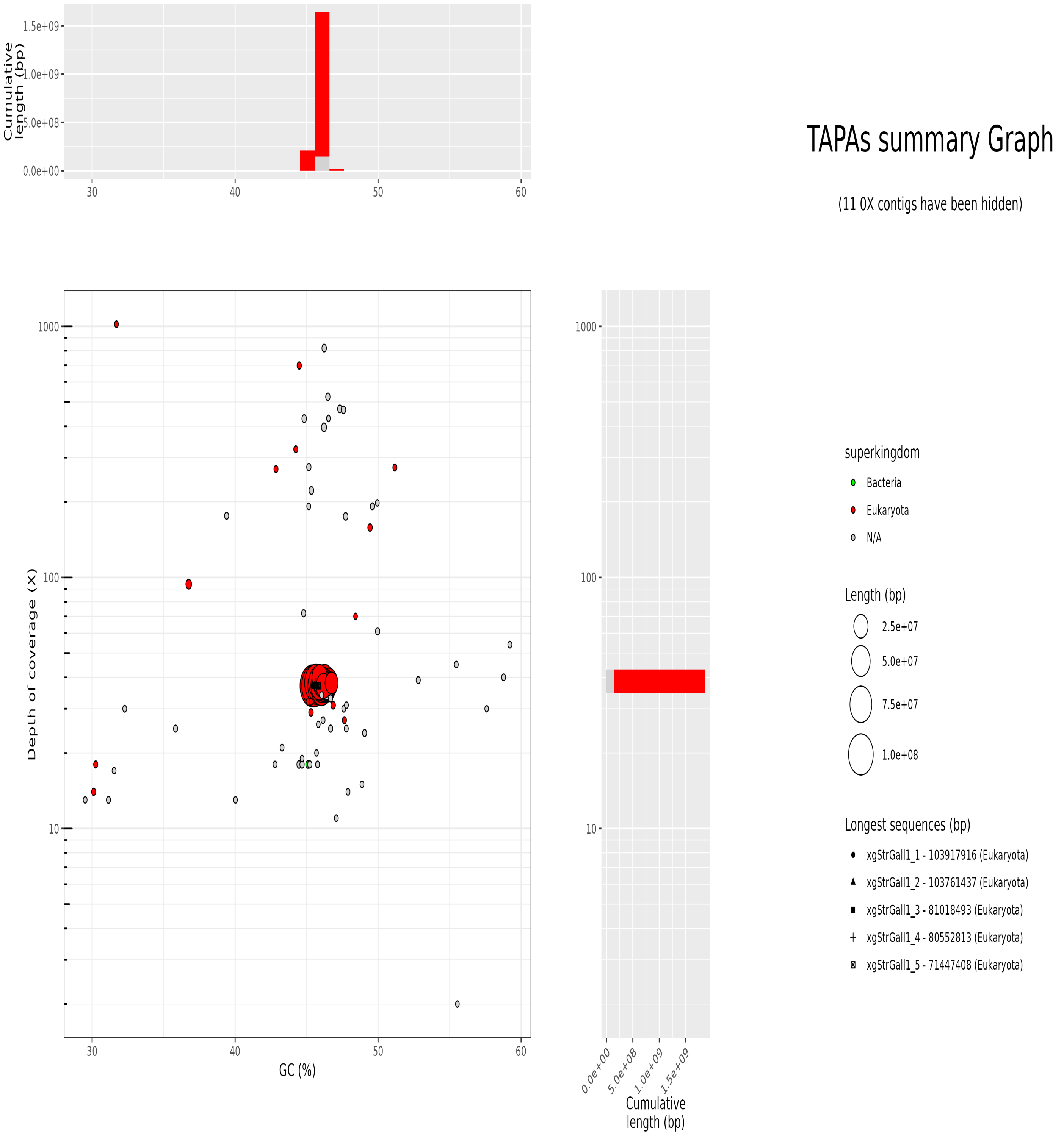


Distribution of k-mer counts per copy numbers found in asm



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	44	221

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: 2026-01-14 22:42:18 CET