ERGA Assembly Report

v24.10.15

Tags: Other[INVALID TAG]

TxID	107379
ToLID	**bLymMin2**
Species	Lymnocryptes minimus
Class	Aves
Order	Charadriiformes

Genome Traits	Expected	Observed
Haploid size (bp)	1,357,199,191	1,320,528,828
Haploid Number	44 (source: ancestor)	49
Ploidy	2 (source: ancestor)	2
Sample Sex	Unknown	Unknown

EBP metrics summary and curation notes

Obtained EBP quality metric for collapsed: 7.7.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

Curator notes

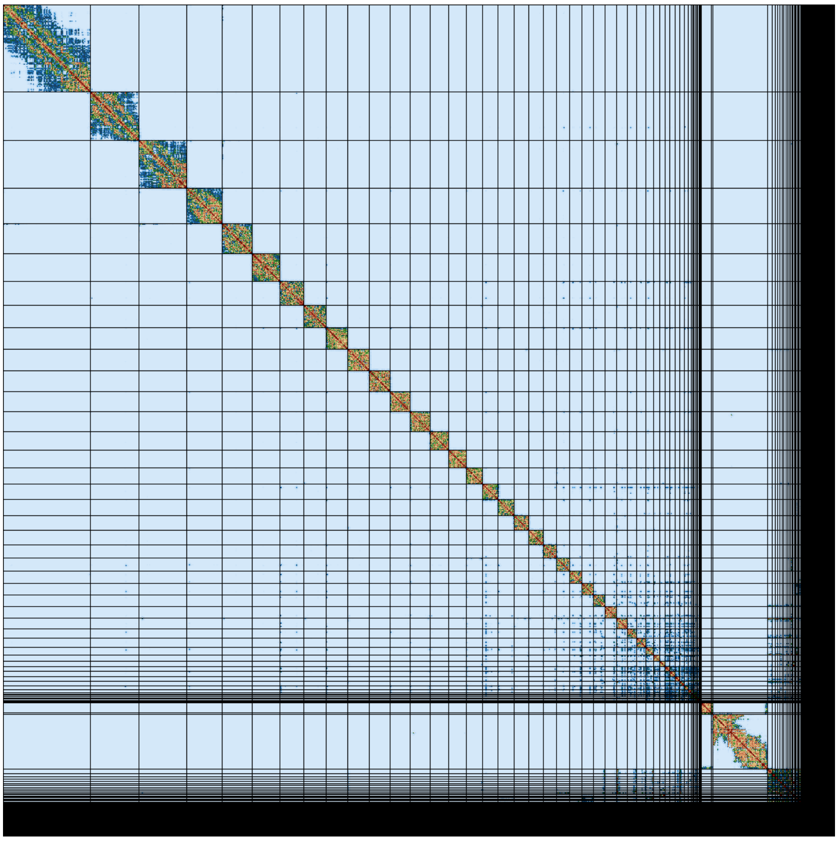
. Interventions/Gb: 23
. Contamination notes: ""
. Other observations: "The assembly of Lymnocryptes minimus (bLymMin2) is based on 57X ONT data and 37X OmniC Hi-C data. The assembly process included the following steps: initial ONT assembly generation with Nextdenovo, removal of contaminant sequences using Context, removal of haplotypic duplications using purge_dups, and Hi-C-based scaffolding with YaHS. Additionally, 434 regions totaling 133.574 Mb (with the largest being 0.997 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 regions, totaling 1.1Mb, (with the largest being 1.1Mb). 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	1,321,705,628	1,320,528,828
GC %	43.71	43.69
Gaps/Gbp	71.12	75.73
Total gap bp	9,400	11,600
Scaffolds	243	240
Scaffold N50	33,500,732	33,872,632
Scaffold L50	12	11
Scaffold L90	40	39
Contigs	337	340
Contig N50	21,678,437	21,678,437
Contig L50	19	19
Contig L90	76	77
QV	39.9487	39.9797
Kmer compl.	89.5048	89.4836
BUSCO sing.	96.9%	96.9%
BUSCO dupl.	0.6%	0.6%
BUSCO frag.	0.5%	0.5%
BUSCO miss.	2.0%	2.0%

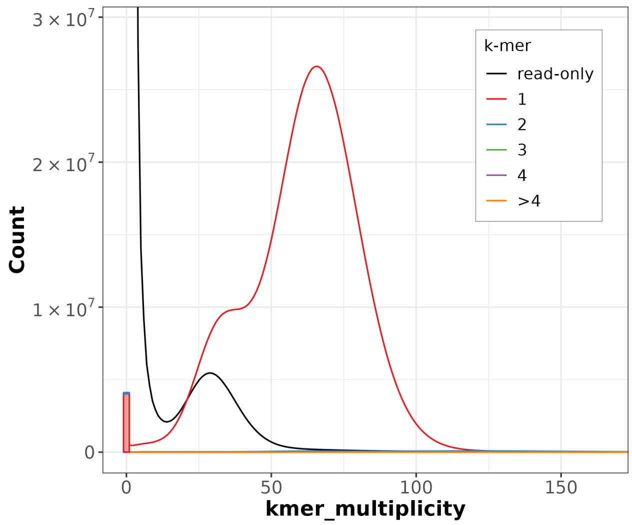
BUSCO: 5.4.3 (euk_genome_met, metaeuk) / Lineage: aves_odb10 (genomes:62, BUSCOs:8338)

HiC contact map of curated assembly

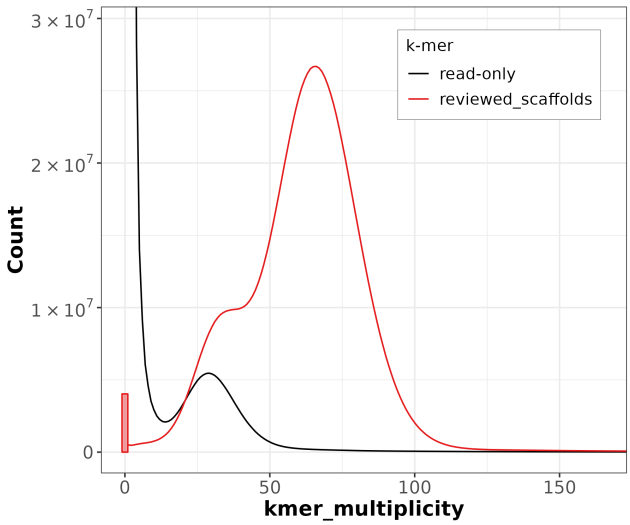


collapsed [LINK]

K-mer spectra of curated assembly

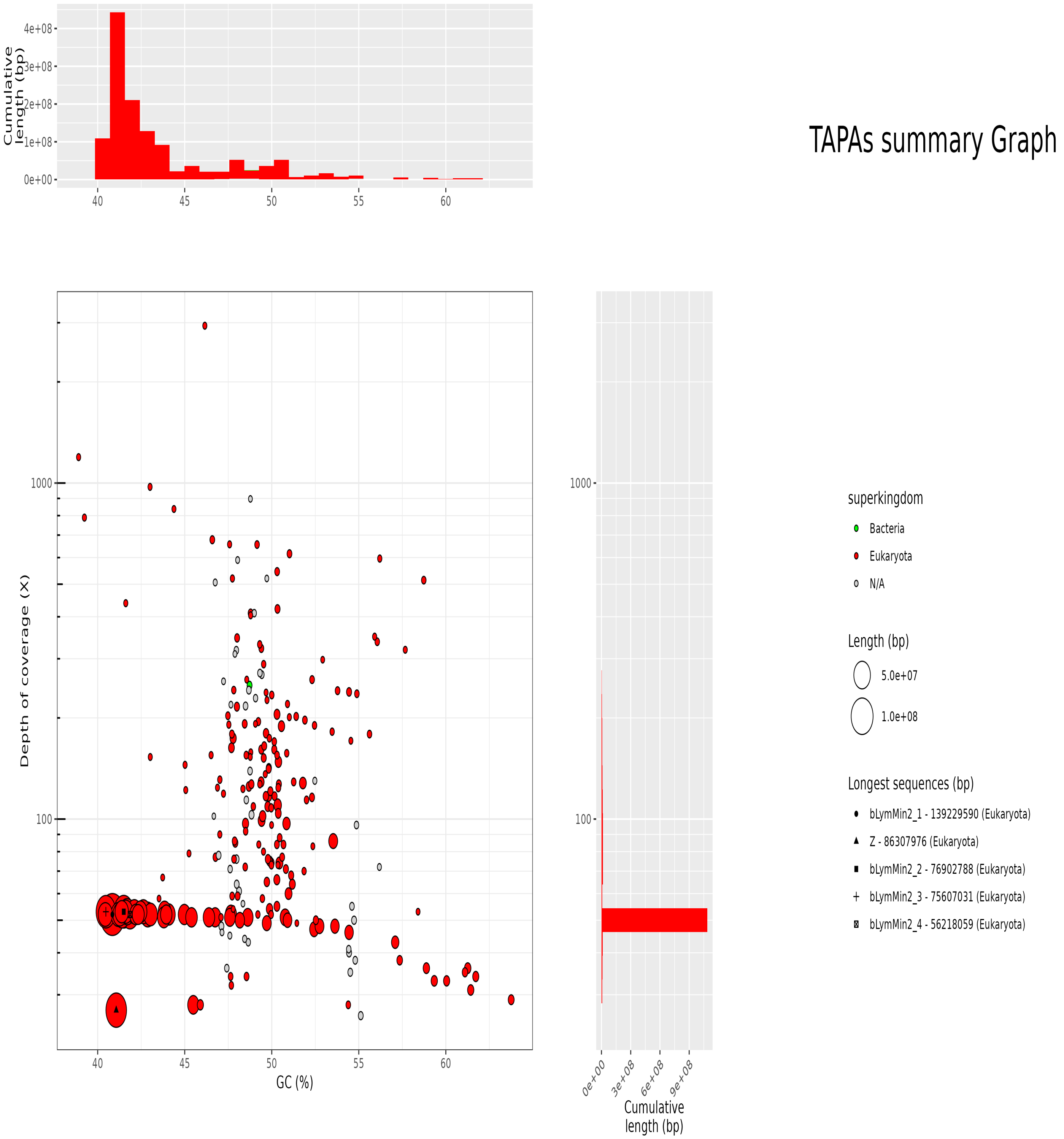


Distribution of k-mer counts per copy numbers found in asm



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	PACBIO Hifi	Omnic
Coverage	57	39

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: 2025-05-06 19:04:20 CEST