ERGA Assembly Report

v24.10.15

Tags: Other[INVALID TAG]

TxID	411526
ToLID	**bTurIli2.1**
Species	Turdus iliacus
Class	Aves
Order	Passeriformes

Genome Traits	Expected	Observed
Haploid size (bp)	1,538,573,586	1,127,629,219
Haploid Number	40 (source: direct)	41
Ploidy	2 (source: ancestor)	2
Sample Sex	Unknown	Unknown

EBP metrics summary and curation notes

Obtained EBP quality metric for collapsed: 6.7.Q23

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

. QV value is less than 40 for collapsed
. Kmer completeness value is less than 90 for collapsed

Curator notes

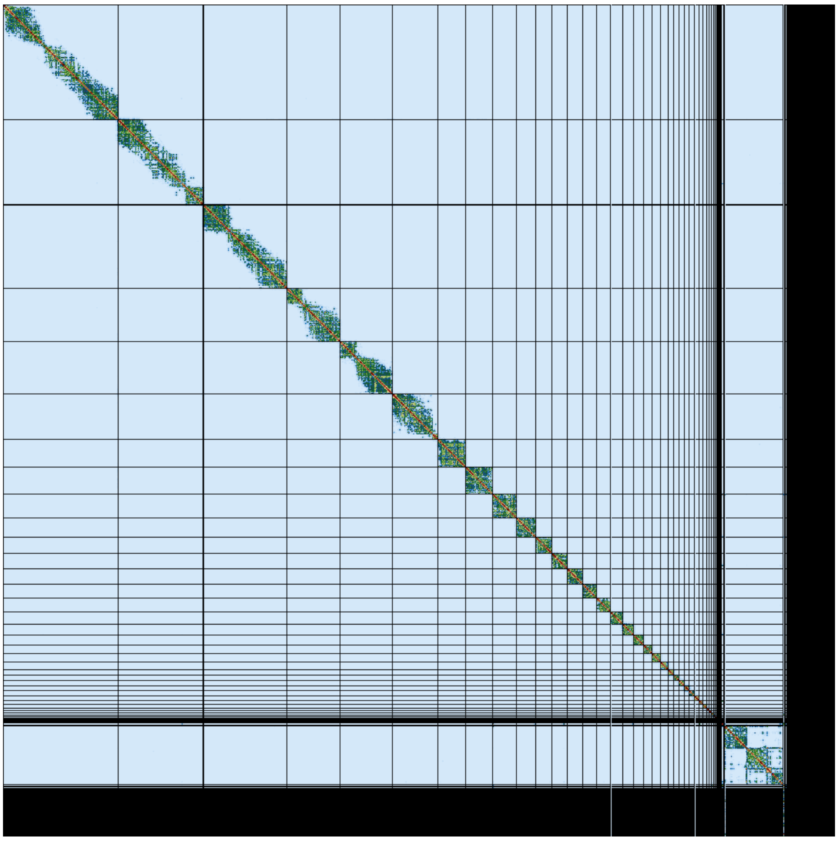
. Interventions/Gb: 96
. Contamination notes: ""
. Other observations: "The assembly of Turdus iliacus (bTurIli2.1) is based on 36X ONT data and Dovetail Hi-C data. The assembly process included the following steps: initial nanopore assembly generation with Nextdenovo, polishing of the consensus using Medaka (with long-reads) and Hapo-G (with short-reads), removal of contaminant sequences using Context, removal of haplotypic duplications using purge_dups, and Hi-C-based scaffolding with YaHS. No contigs were found to be contaminants (bacterial, archaeal, or viral). Additionally, 1,524 regions totaling 492 Mb (with the largest being 3.8 Mb) were identified as haplotypic duplications and removed. The mitochondrial genome was assembled using ptGAUL. Finally, the primary assembly was analyzed and manually improved using Pretext. During manual curation, 14 haplotypic regions were removed, totaling 4.20 Mb (with the largest being 898 Kb). Chromosome-scale scaffolds confirmed by Hi-C data were named in order of size. Sexual chromosomes Z and W were identified using comparative genomics and organization of microchromosomes was curated using MicroFinder. "

Quality metrics table

Metrics	Pre-curation collapsed	Curated collapsed
Total bp	1,131,807,320	1,127,629,219
GC %	42.55	42.54
Gaps/Gbp	312.77	347.63
Total gap bp	35,400	45,400
Scaffolds	388	347
Scaffold N50	61,919,579	70,769,131
Scaffold L50	6	6
Scaffold L90	30	28
Contigs	742	739
Contig N50	8,864,895	8,414,000
Contig L50	35	35
Contig L90	214	217
QV	23.6414	23.696
Kmer compl.	71.6929	71.6424
BUSCO sing.	95.4%	95.9%
BUSCO dupl.	0.9%	0.4%
BUSCO frag.	0.6%	0.6%
BUSCO miss.	3.1%	3.1%

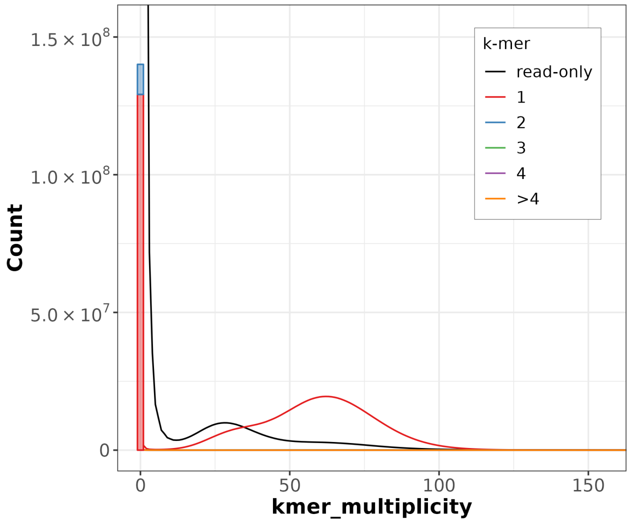
BUSCO: 5.4.3 (euk_genome_met, metaeuk) / Lineage: passeriformes_odb10 (genomes:15, BUSCOs:10844)

HiC contact map of curated assembly

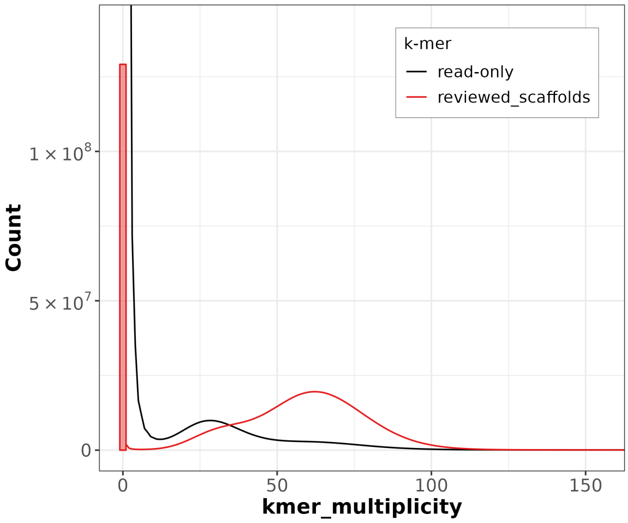


collapsed [LINK]

K-mer spectra of curated assembly

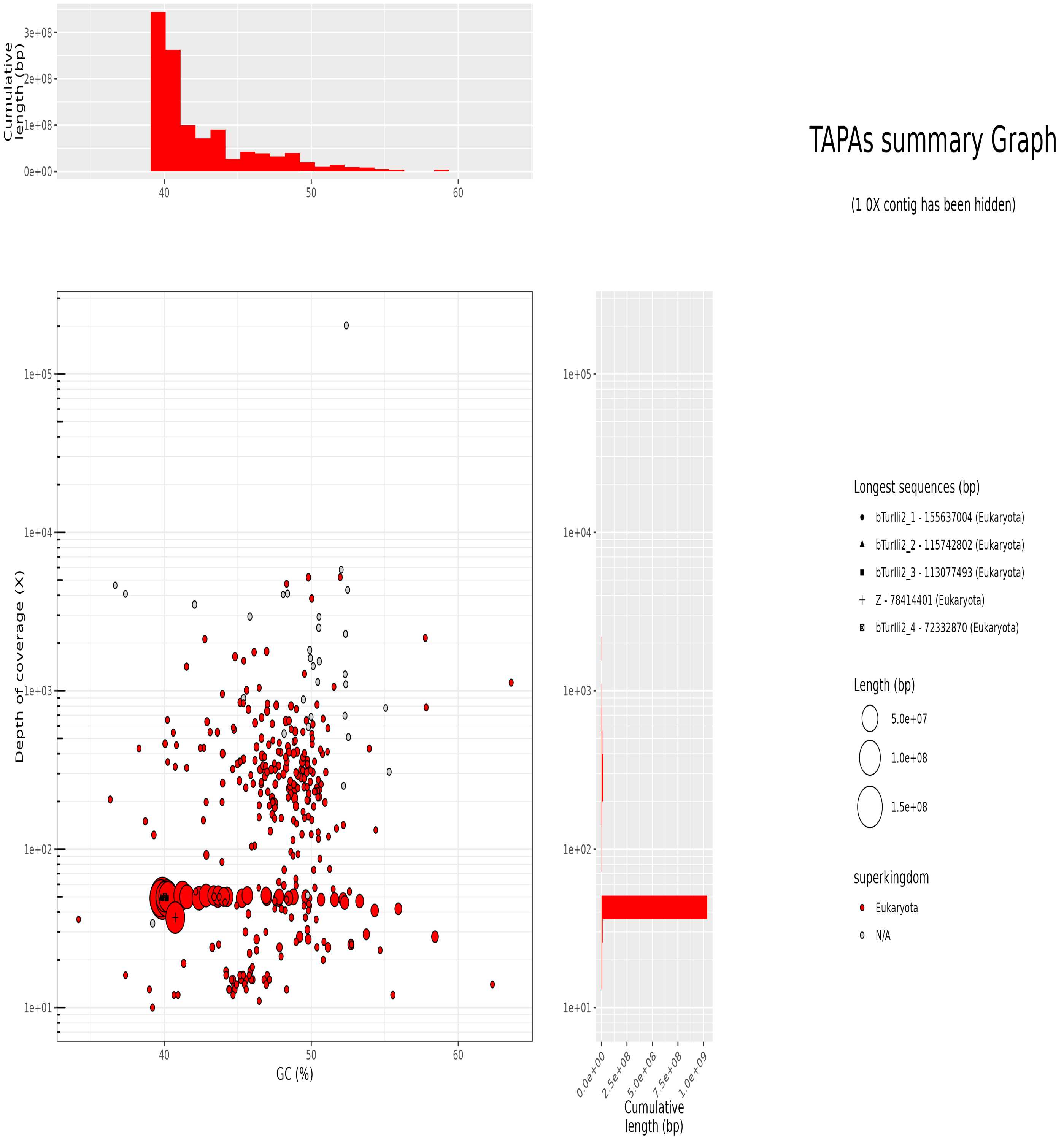


Distribution of k-mer counts per copy numbers found in asm



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	51	35

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: Jean-Marc Aury
Affiliation: Genoscope

Date and time: 2025-05-05 11:40:21 CEST