ERGA Assembly Report

v24.10.15

Tags: Other[INVALID TAG]

TxID	127946
ToLID	**bTurPhi1**
Species	Turdus philomelos
Class	Aves
Order	Passeriformes

Genome Traits	Expected	Observed
Haploid size (bp)	1,427,399,439	1,090,121,589
Haploid Number	40 (source: direct)	40
Ploidy	3 (source: ancestor)	2
Sample Sex	Unknown	Unknown

EBP metrics summary and curation notes

Obtained EBP quality metric for collapsed: 6.7.Q34

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 Ploidy is different from Expected

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

Curator notes

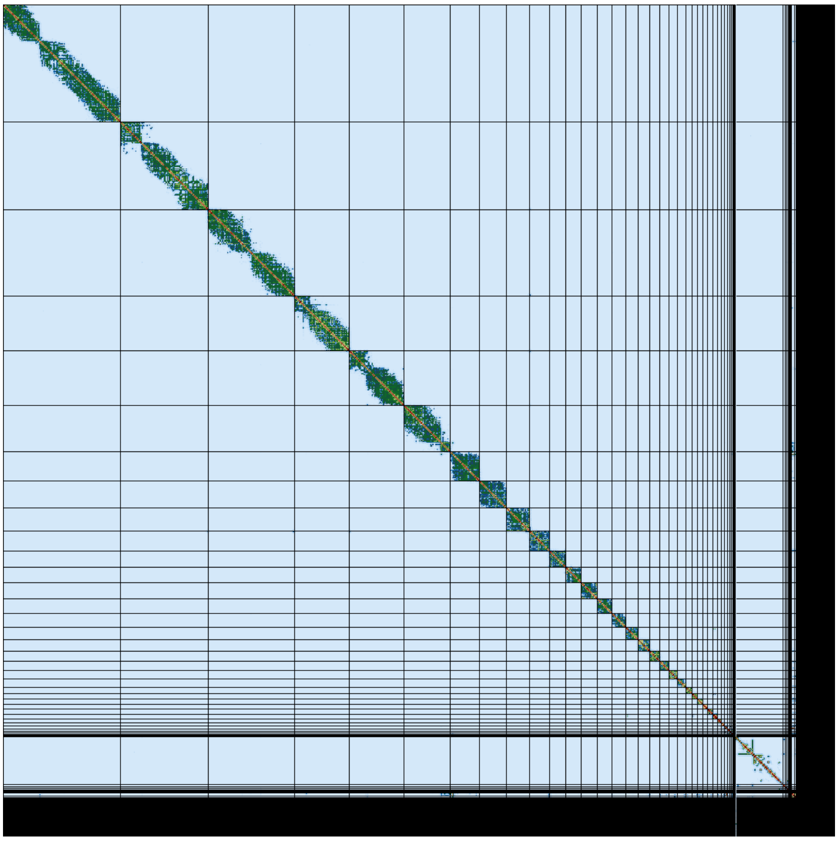
. Interventions/Gb: 95
. Contamination notes: ""
. Other observations: "The assembly of Turdus philomelos (bTurPhi1) is based on 45X ONT data and 56X Arima Hi-C data. The assembly process included the following steps: initial ONT assembly generation with Nextdenovo, removal of contaminant sequences using Context, removal of haplotypic duplications using purge_dups, and Hi-C-based scaffolding with YaHS. Additionally, 1500 regions totaling 394.031 Mb (with the largest being 2.72 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, 6 haplotypic regions, totaling 2.5Mb, (with the largest being 0.789Mb). 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	1,091,155,513	1,090,121,589
GC %	42.26	42.26
Gaps/Gbp	515.97	529.3
Total gap bp	56,300	59,900
Scaffolds	298	281
Scaffold N50	61,036,437	62,560,071
Scaffold L50	6	6
Scaffold L90	23	26
Contigs	861	858
Contig N50	7,452,025	7,452,025
Contig L50	39	39
Contig L90	316	317
QV	26.0652	34.3444
Kmer compl.	76.1414	49.0423
BUSCO sing.	95.3%	95.5%
BUSCO dupl.	0.5%	0.4%
BUSCO frag.	0.6%	0.6%
BUSCO miss.	3.6%	3.5%

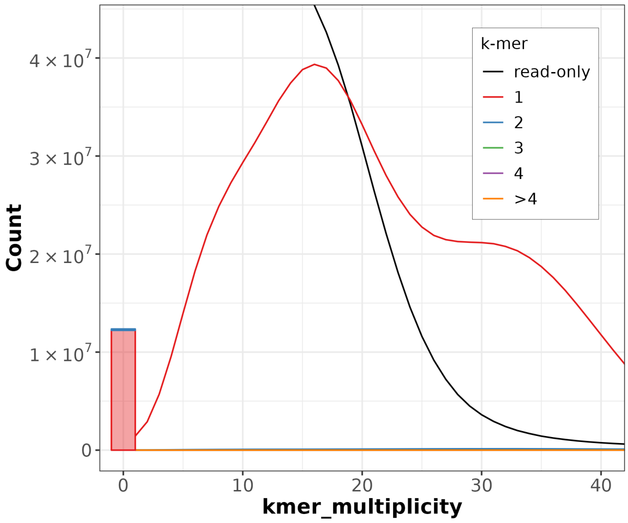
BUSCO: 5.4.3 (euk_genome_met, metaeuk) / Lineage: passeriformes_odb10 (genomes:15, BUSCOs:10844)

HiC contact map of curated assembly

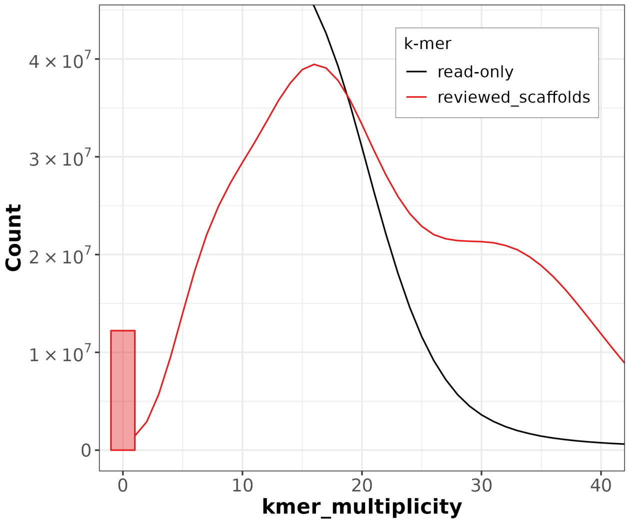


collapsed [LINK]

K-mer spectra of curated assembly

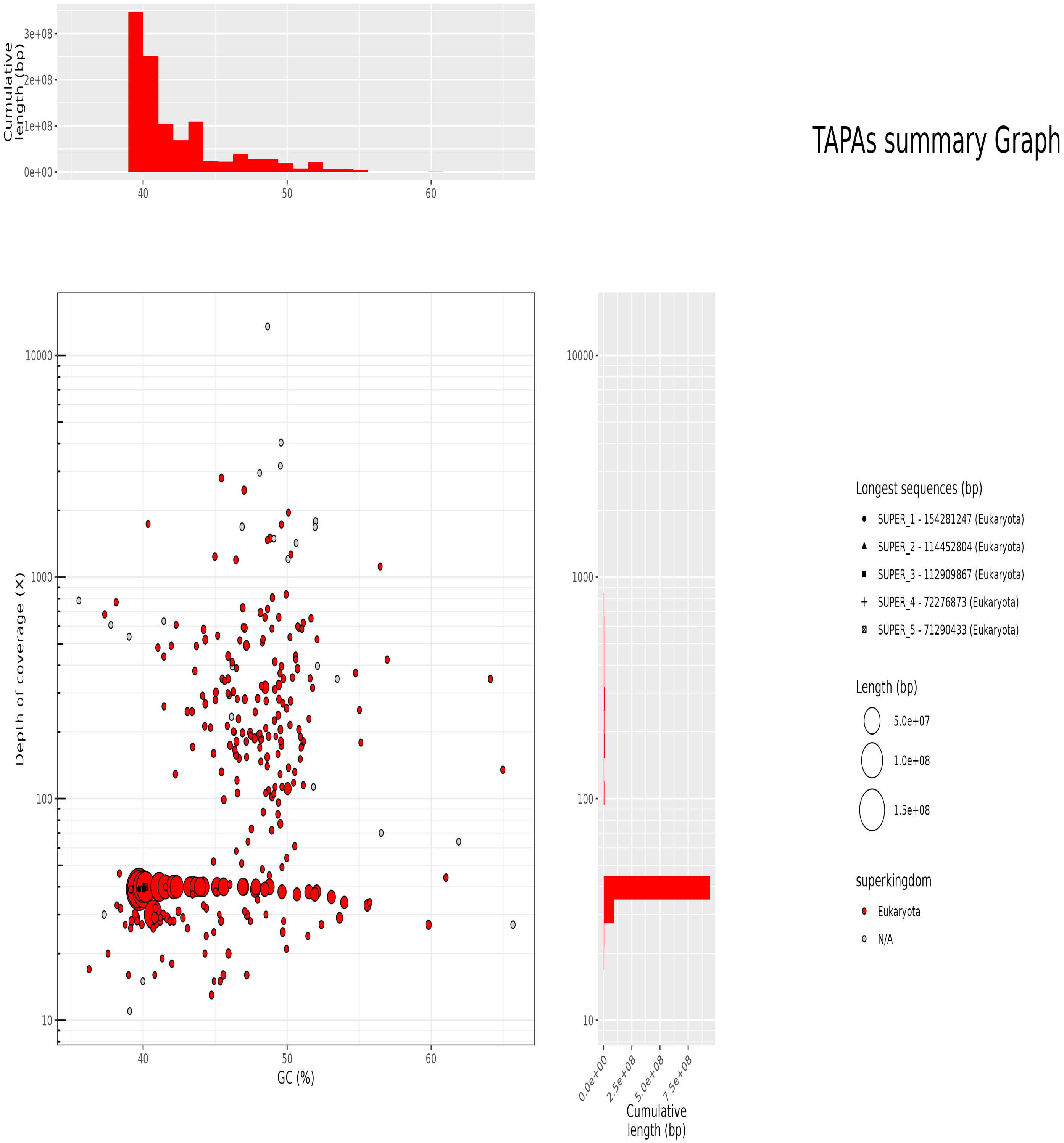


Distribution of k-mer counts per copy numbers found in asm



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	45	44

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: Arnaud Couloux
Affiliation: Genoscope

Date and time: 2025-04-05 19:32:40 CEST