ERGA Assembly Report

v24.10.15

Tags: Other[INVALID TAG]

TxID	9187
ToLID	**bTurMer3**
Species	Turdus merula
Class	Aves
Order	Passeriformes

Genome Traits	Expected	Observed
Haploid size (bp)	1,304,570,307	1,105,482,267
Haploid Number	40 (source: direct)	41
Ploidy	3 (source: ancestor)	2
Sample Sex	Unknown	Unknown

EBP metrics summary and curation notes

Obtained EBP quality metric for collapsed: 7.7.Q41

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

. Observed Haploid Number is different from Expected
. Observed Ploidy is different from Expected
. Kmer completeness value is less than 90 for collapsed

Curator notes

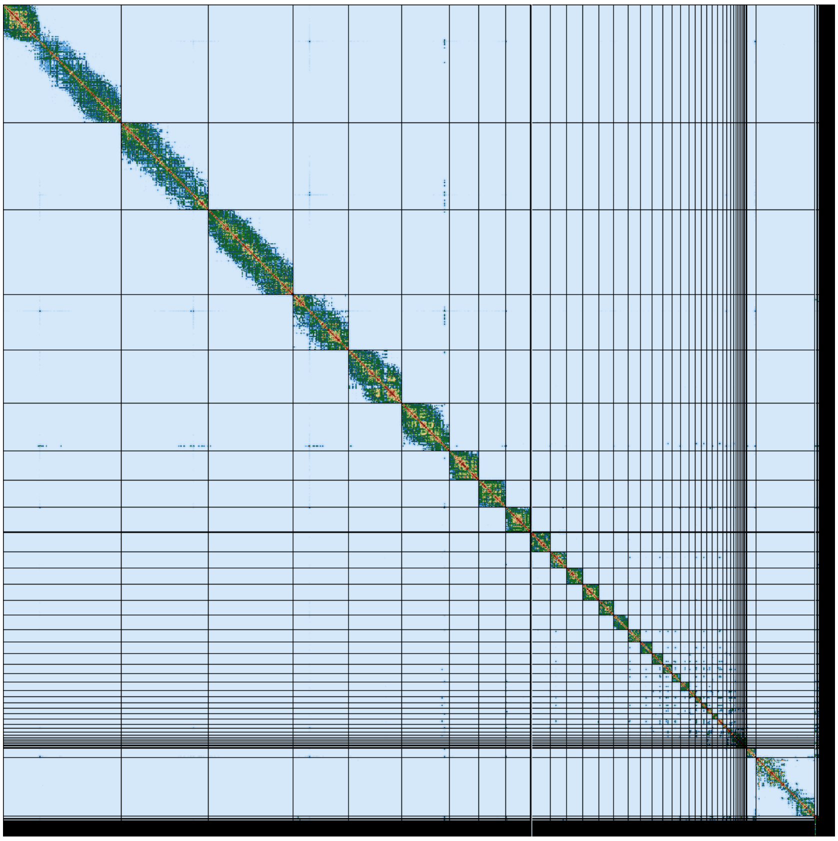
. Interventions/Gb: 21
. Contamination notes: ""
. Other observations: "The assembly process included the following steps: initial ONT assembly generation with Nextdenovo, removal of contaminant sequences using Context, removal of haplotypic duplications using purge_dups, and Hi-C-based scaffolding with YaHS. In total, 1 contig was identified as contaminant (bacterial) totaling 37 kb (with the largest being 37 kb). Additionally, 444 regions totaling 177 Mb (with the largest being 6 Mb) were identified as haplotypic duplications and removed. Finally, the primary assembly was analyzed and manually improved using Pretext. During manual curation, 1 haplotypic region was removed, totaling 573 kb. 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	1,106,047,045	1,105,482,267
GC %	42.47	42.47
Gaps/Gbp	46.11	48.85
Total gap bp	5,100	6,400
Scaffolds	140	137
Scaffold N50	71,343,807	71,343,807
Scaffold L50	6	6
Scaffold L90	21	23
Contigs	191	191
Contig N50	32,937,796	32,937,796
Contig L50	10	10
Contig L90	43	43
QV	40.5322	41.2002
Kmer compl.	86.2077	85.1469
BUSCO sing.	96.2%	96.2%
BUSCO dupl.	0.4%	0.4%
BUSCO frag.	0.6%	0.6%
BUSCO miss.	2.8%	2.8%

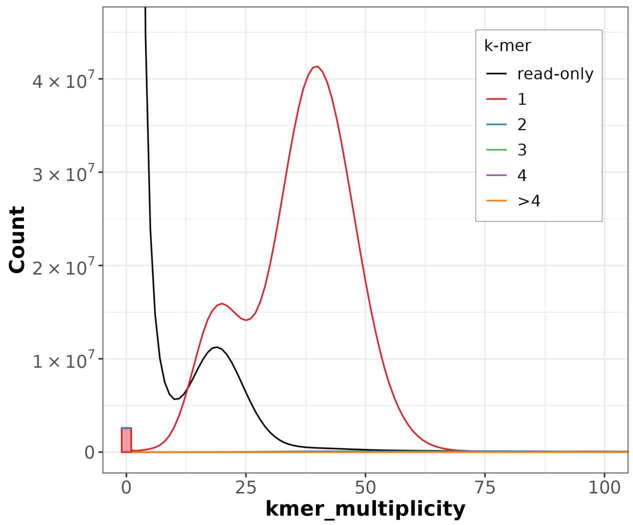
BUSCO: 5.4.3 (euk_genome_met, metaeuk) / Lineage: passeriformes_odb10 (genomes:15, BUSCOs:10844)

HiC contact map of curated assembly

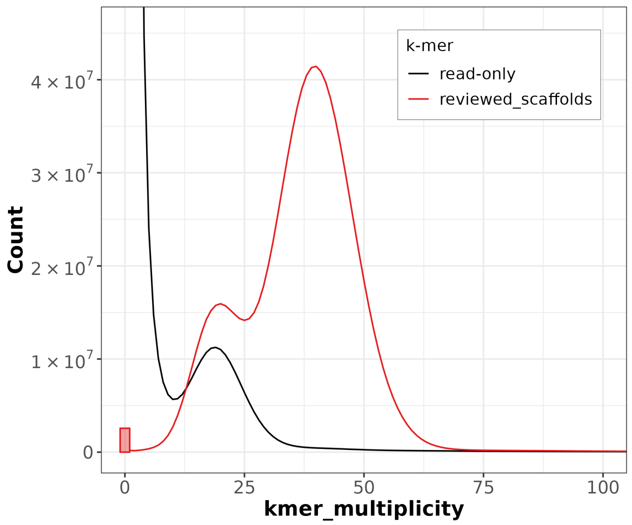


collapsed [LINK]

K-mer spectra of curated assembly

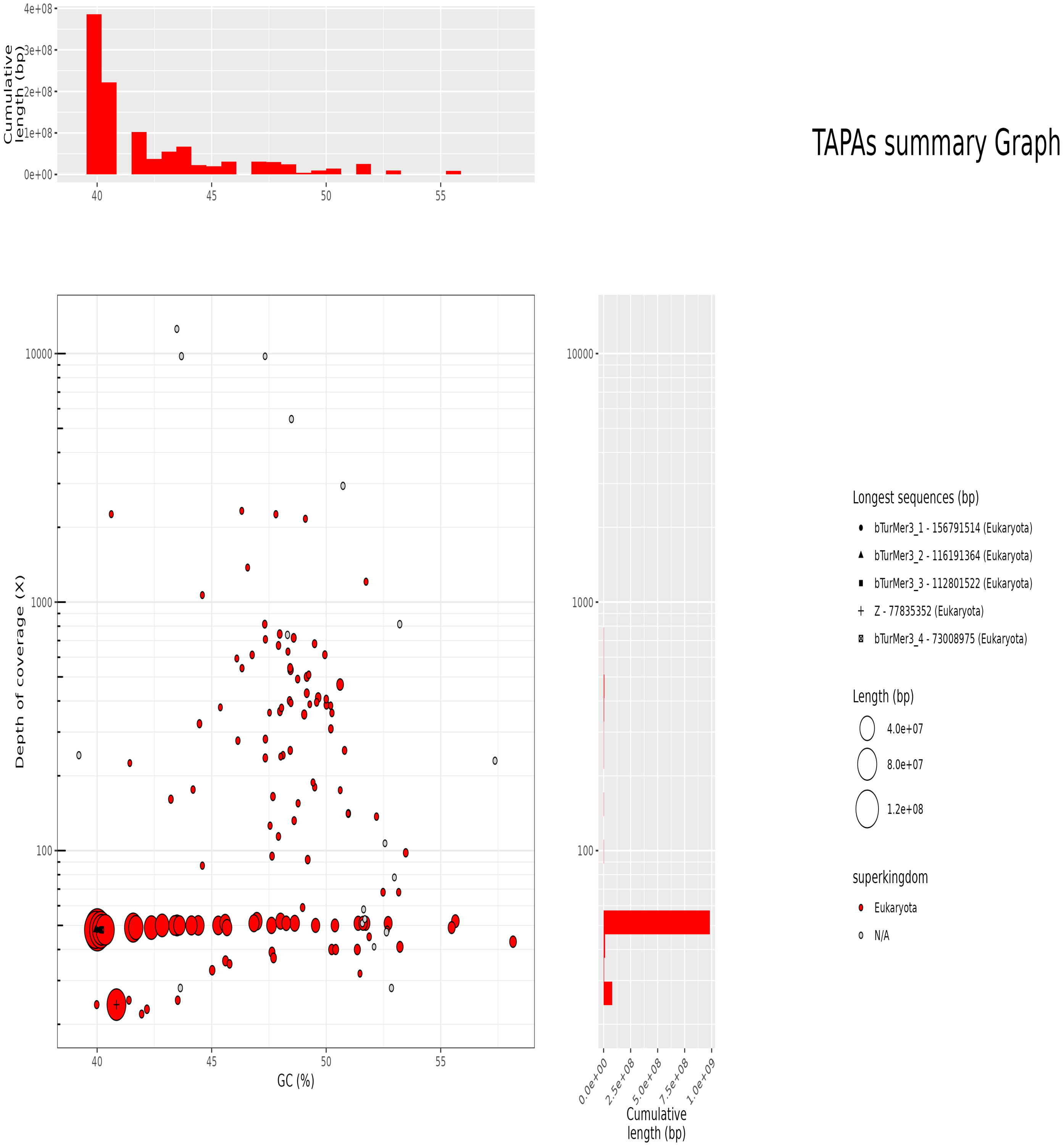


Distribution of k-mer counts per copy numbers found in asm



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	53	51

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: Adama Ndar
Affiliation: Genoscope

Date and time: 2025-04-06 14:55:29 CEST