ERGA Assembly Report

v24.10.15

Tags: Other[INVALID TAG]

TxID	100826
ToLID	**bScoRus3**
Species	Scolopax rusticola
Class	Aves
Order	Charadriiformes

Genome Traits	Expected	Observed
Haploid size (bp)	1,134,849,087	1,145,070,085
Haploid Number	44 (source: direct)	46
Ploidy	2 (source: ancestor)	2
Sample Sex	Unknown	Unknown

EBP metrics summary and curation notes

Obtained EBP quality metric for collapsed: 7.7.Q23

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

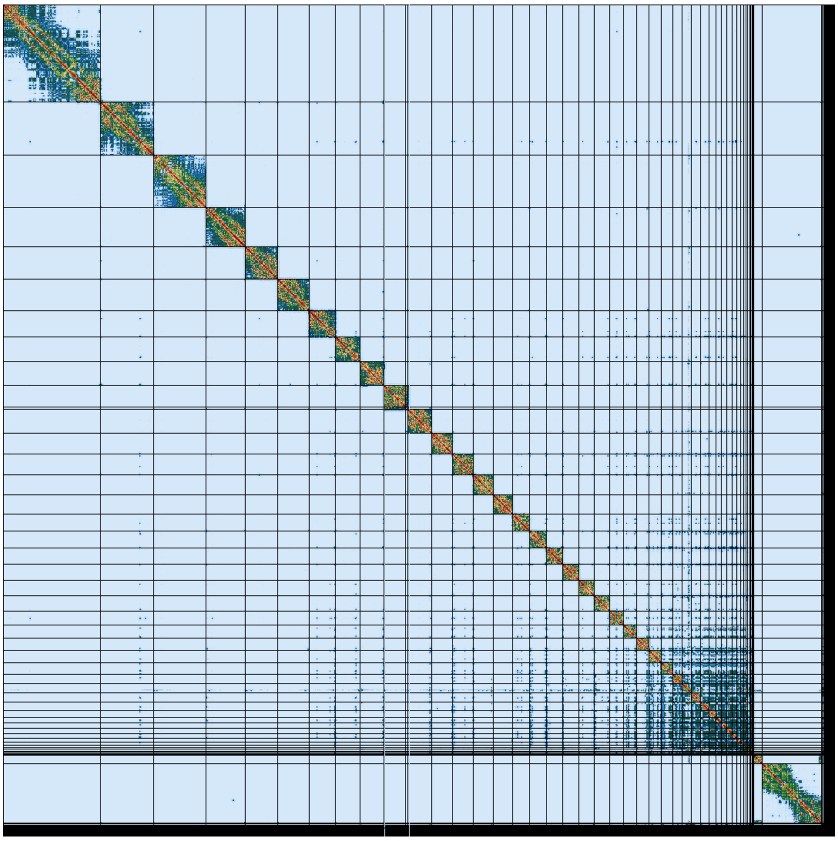
. Interventions/Gb: 7
. Contamination notes: ""
. Other observations: "The assembly of Scolopax rusticola (bScoRus3.3) is based on 107X ONT data and 110X Omnic Hi-C data. The assembly process included the following steps: initial ONT assembly generation with Nextdenovo, removal of contaminant sequences using Context, removal of haplotypic duplications using purge_dups, and Hi-C-based scaffolding with YaHS. No contigs were found to be contaminants (bacterial, archaeal, or viral). Additionally, 108 regions totaling 22.483 Mb (with the largest being 1.248 Mb) were identified as haplotypic duplications and removed. Finally, the primary assembly was analyzed and manually improved using Pretext. During manual curation, 15 contaminant sequences were removed, totaling 11.704 Mb, (with the largest being 2.384 Mb). Chromosome-scale scaffolds confirmed by Hi-C data were named in order of size. Sexual chromosomes Z and W were identified during manual curation and validated using comparative genomics. Organization of microchromosomes was curated using MicroFinder. "

Quality metrics table

Metrics	Pre-curation collapsed	Curated collapsed
Total bp	1,156,772,728	1,145,070,085
GC %	42.29	42.25
Gaps/Gbp	41.49	48.03
Total gap bp	4,800	6,300
Scaffolds	119	96
Scaffold N50	32,951,219	33,623,676
Scaffold L50	10	9
Scaffold L90	31	30
Contigs	167	151
Contig N50	23,579,355	23,579,355
Contig L50	13	13
Contig L90	45	44
QV	35.5926	23.6205
Kmer compl.	83.0295	73.0556
BUSCO sing.	97.7%	98.1%
BUSCO dupl.	0.7%	0.3%
BUSCO frag.	0.0%	0.0%
BUSCO miss.	1.6%	1.6%

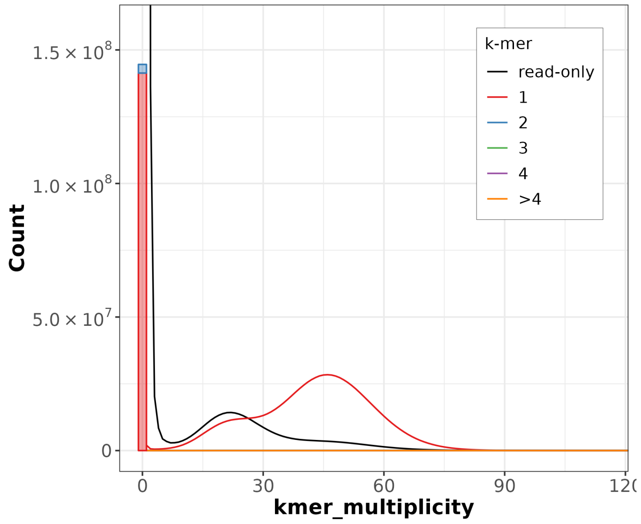
BUSCO: 6.0.0 (euk_genome_min, miniprot) / Lineage: passeriformes_odb12 (genomes:59, BUSCOs:6684)

HiC contact map of curated assembly

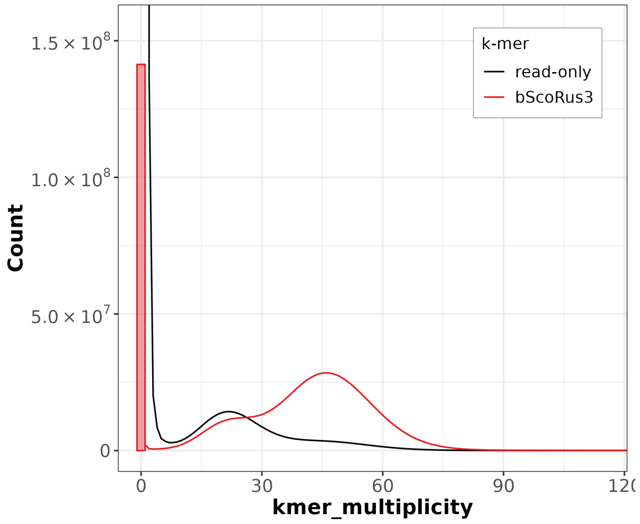


collapsed [LINK]

K-mer spectra of curated assembly

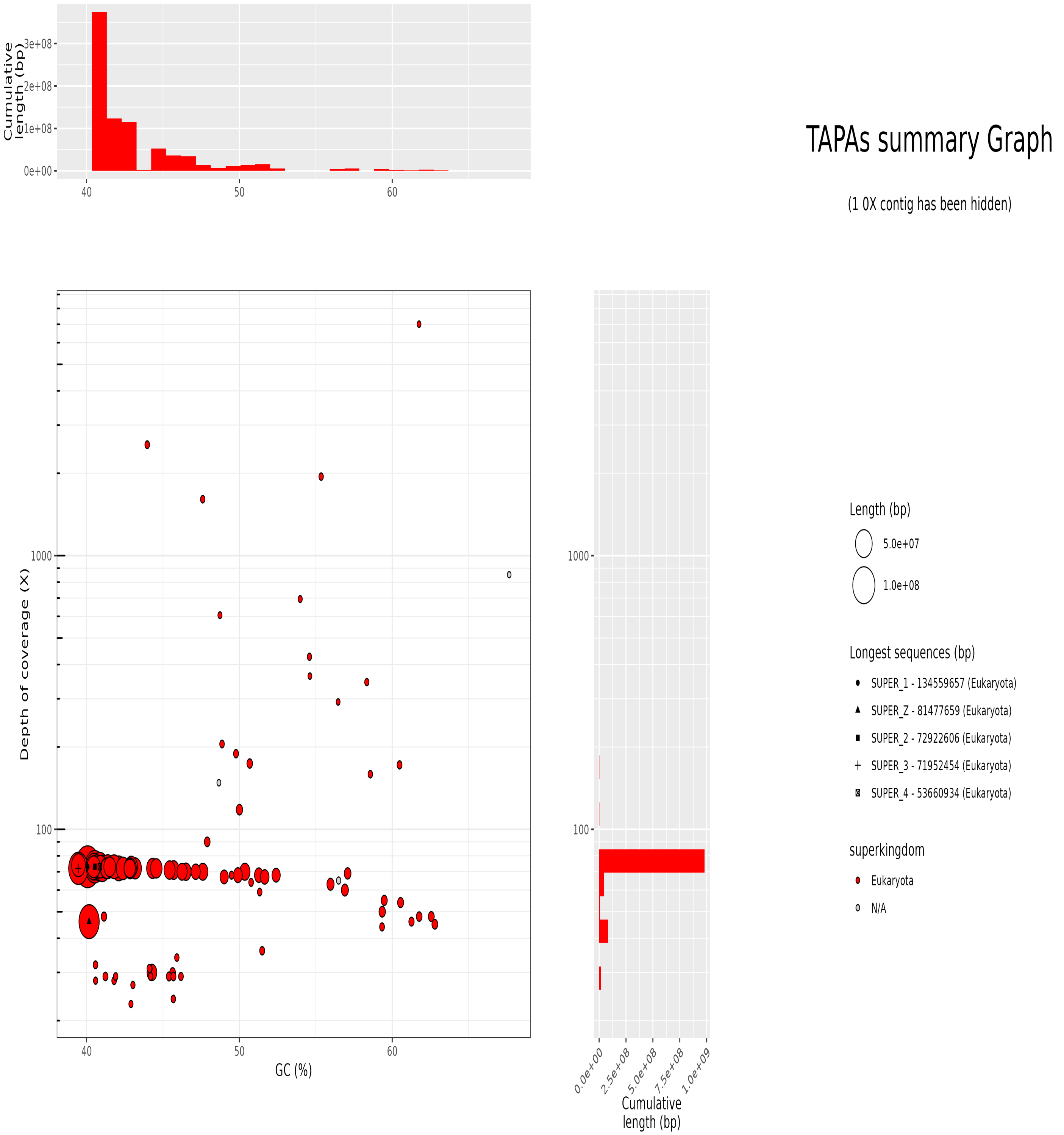


Distribution of k-mer counts per copy numbers found in asm



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	Omnic
Coverage	107	110

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: 2025-10-31 10:42:21 CET