# ERGA Assembly Report

v24.10.15

Tags: Other[INVALID TAG]

| | |
|---|---|
| TxID | 100826 |
| ToLID | **bScoRus3.2** |
| Species | Scolopax rusticola |
| Class | Aves |
| Order | Charadriiformes |

| Genome Traits | Expected | Observed |
|---|---|---|
| Haploid size (bp) | 1,139,795,121 | 1,199,152,678 |
| Haploid Number | 44 (source: direct) | 44 |
| Ploidy | 3 (source: ancestor) | 2 |
| Sample Sex | Unknown | Unknown |

## EBP metrics summary and curation notes

Obtained EBP quality metric for collapsed: 7.7.Q24

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

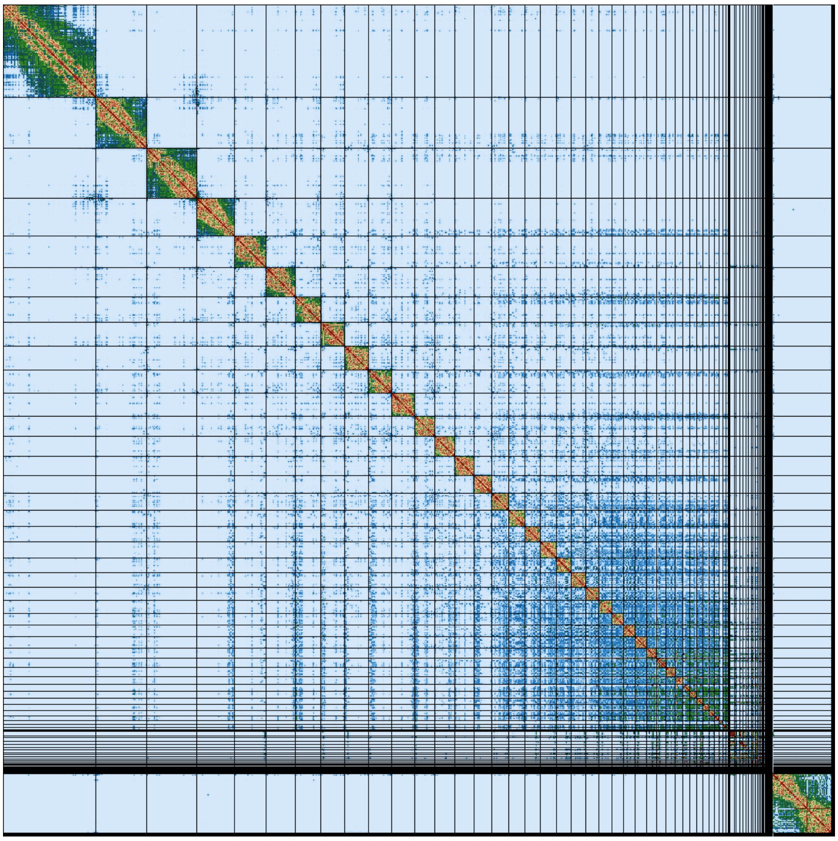
. Interventions/Gb: 10
. Contamination notes: ""
. Other observations: "The assembly of Scolopax rusticola (bScoRus3.2) is based on 110X PacBio and 36X ONT data and 110X OmniC Hi-C data. The assembly process included the following steps: initial PacBio and ONT assembly generation with Hifiasm, removal of contaminant sequences using Context, removal of haplotypic duplications using purge_dups, and Hi-C-based scaffolding with YaHS. In total, XX contigs were identified as contaminants (bacterial, archaeal, or viral), totaling XX Mb (with the largest being XX Mb). Additionally, 96 regions totaling 10.59 Mb (with the largest being 0.88 Mb) were identified as haplotypic duplications and removed. Finally, the primary assembly was analyzed and manually improved using Pretext. Contigs were mostly chromosome-scale and 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 | 1,199,150,978 | 1,199,152,678 |
| GC % | 42.79 | 42.79 |
| Gaps/Gbp | 23.35 | 29.19 |
| Total gap bp | 2,800 | 4,500 |
| Scaffolds | 130 | 123 |
| Scaffold N50 | 33,678,863 | 33,678,863 |
| Scaffold L50 | 10 | 10 |
| Scaffold L90 | 33 | 33 |
| Contigs | 158 | 158 |
| Contig N50 | 33,298,157 | 33,298,157 |
| Contig L50 | 10 | 10 |
| Contig L90 | 35 | 35 |
| QV | 43.8836 | 24.5487 |
| Kmer compl. | 82.8694 | 72.9929 |
| BUSCO sing. | 95.4% | 95.4% |
| BUSCO dupl. | 0.5% | 0.4% |
| BUSCO frag. | 0.4% | 0.4% |
| BUSCO miss. | 3.7% | 3.8% |

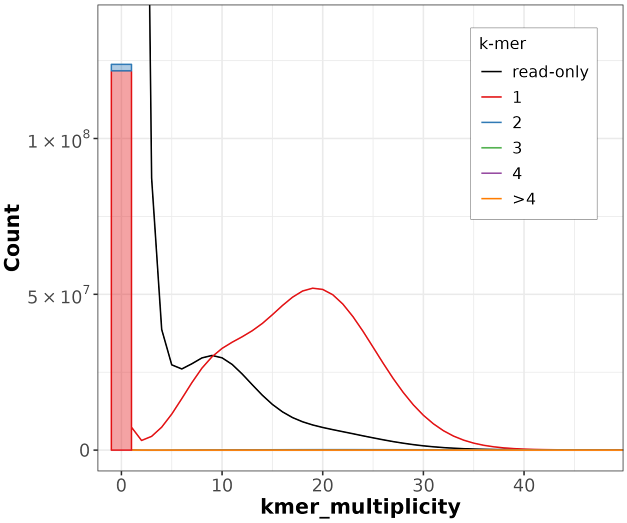
BUSCO: 5.4.3 (euk_genome_met, metaeuk) / Lineage: passeriformes_odb10 (genomes:15, BUSCOs:10844)

# HiC contact map of curated assembly

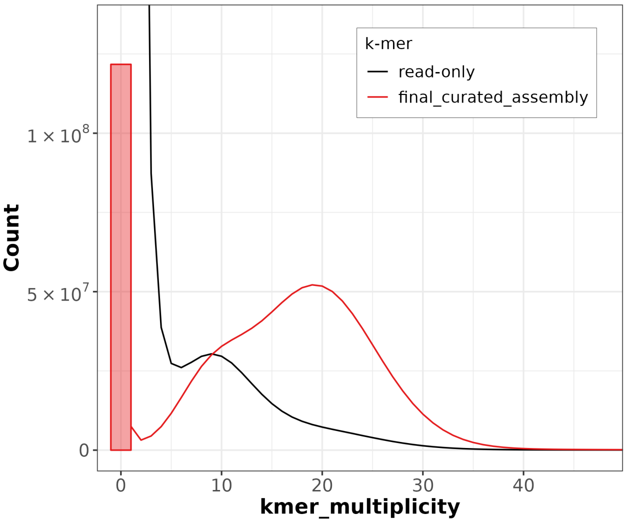


**collapsed** [LINK]

# K-mer spectra of curated assembly

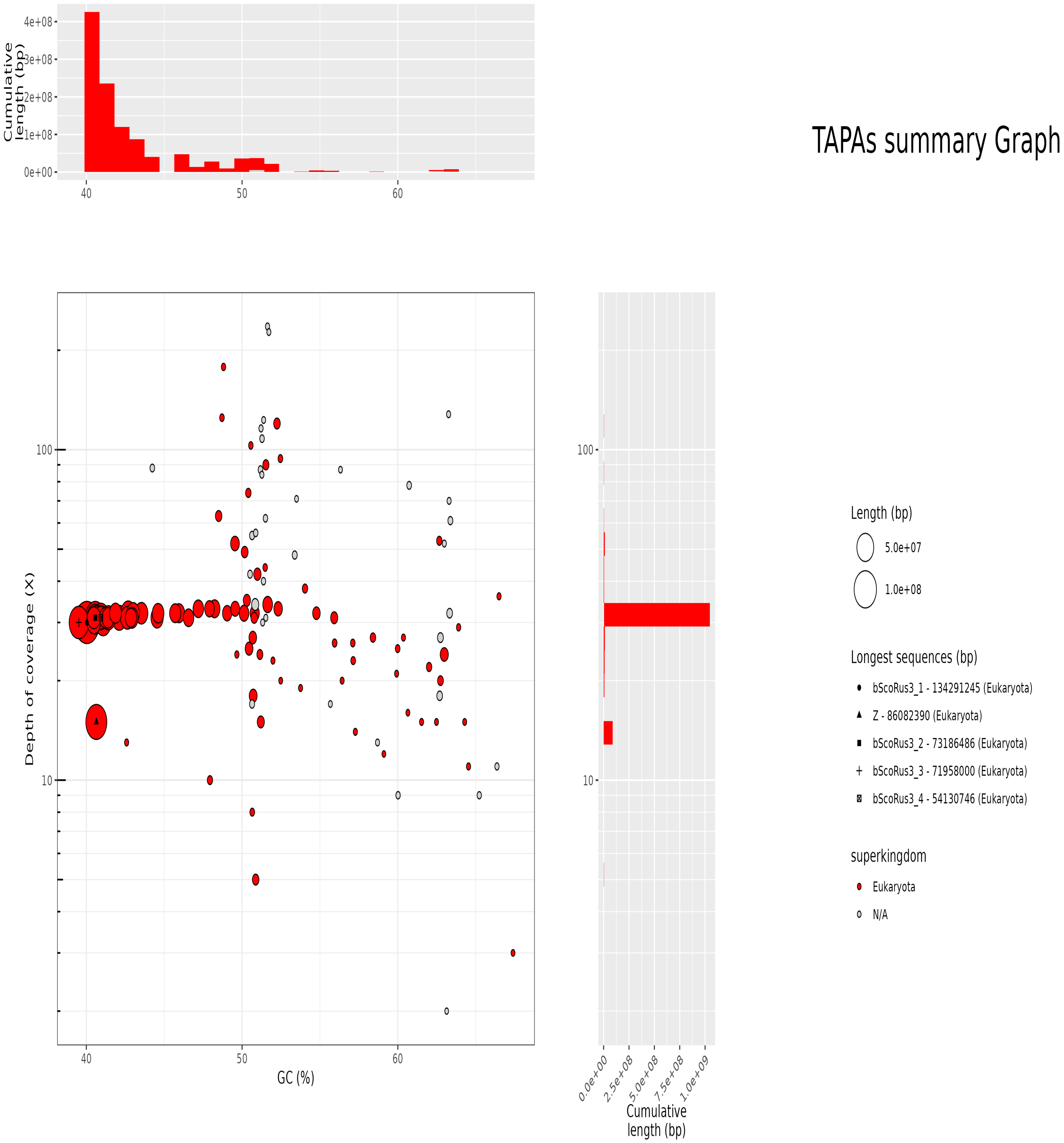


Distribution of k-mer counts per copy numbers found in asm



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 | 146 | 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: Jean-Marc Aury
Affiliation: Genoscope

Date and time: 2025-02-20 06:22:13 CET