

# ERGA Assembly Report

v24.10.15

Tags: ERGA-BGE

TxID	100826
ToLID	<b>bScoRus</b>
Species	<i>Scolopax rusticola</i>
Class	Aves
Order	Charadriiformes

Genome Traits	Expected	Observed
Haploid size (bp)	1,031,232,411	1,232,633,887
Haploid Number	44 (source: direct)	44
Ploidy	2 (source: ancestor)	2
Sample Sex	ZW	ZW

## EBP metrics summary and curation notes

Obtained EBP quality metric for hap1: 6.7.Q64

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

- . Kmer completeness value is less than 90 for hap1

### Curator notes

- . Interventions/Gb: None
- . Contamination notes: "No contaminants were found with the pipeline NCBI FCS GX v Galaxy Version 0.5.5+galaxy1. No scaffolds are flagged as contaminants by BlobToolKit."
- . Other observations: "Curation was done in two steps. 1) In a double haplotype contact map, the two haplotypes were checked to resolve haplotigs and construct an Hap1 as complete as possible. The HiC contact was low, so to aid in reconstructing Microchromosomes, an alignment was performed with the scaffold level genome ASM2856531v1 (GCA\_028565315.1) of the same species. Another chromosome-level genome is at the moment available on NCBI (bScoRus3.2, GCA\_964304615.2) 2) The Curation was done only on the most complete haplotype, in this case Hap1."

## Quality metrics table

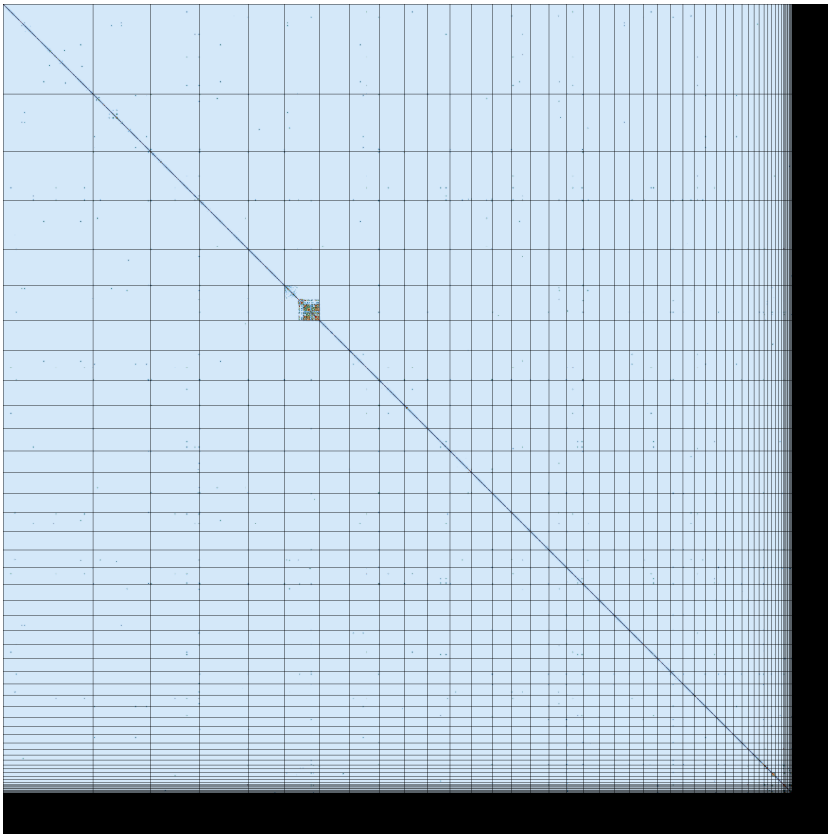
Metrics	Pre-curation hap1	Curated hap1
Total bp	1,150,284,723	1,232,633,887
GC %	42.98	42.86
Gaps/Gbp		544.36
Total gap bp	124,000	134,200
Scaffolds	976	941
Scaffold N50	29,569,713	34,162,297
Scaffold L50	10	10
Scaffold L90	35	33
Contigs	1,596	1,612
Contig N50	5,908,451	5,888,542
Contig L50	48	51
Contig L90	314	320
QV	64.2059	64.5086
Kmer compl.	82.8495	88.2735
BUSCO sing.	92.8%	98.1%
BUSCO dupl.	0.4%	0.5%
BUSCO frag.	0.9%	0.4%
BUSCO miss.	5.9%	1.1%

Warning! BUSCO versions or lineage datasets are not the same across results:

BUSCO: 5.8.0 (euk\_genome\_met, metaeuk) / Lineage: aves\_odb10 (genomes:62, BUSCOs:8338)

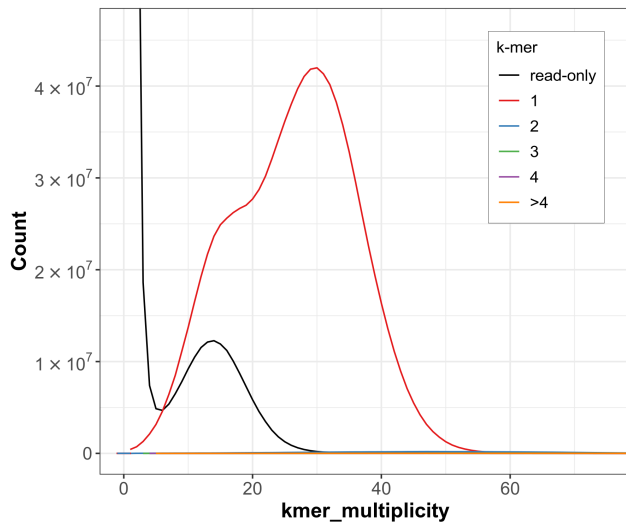
BUSCO: 5.8.0 (euk\_genome\_min, miniprot) / Lineage: aves\_odb10 (genomes:62, BUSCOs:8338)

# HiC contact map of curated assembly

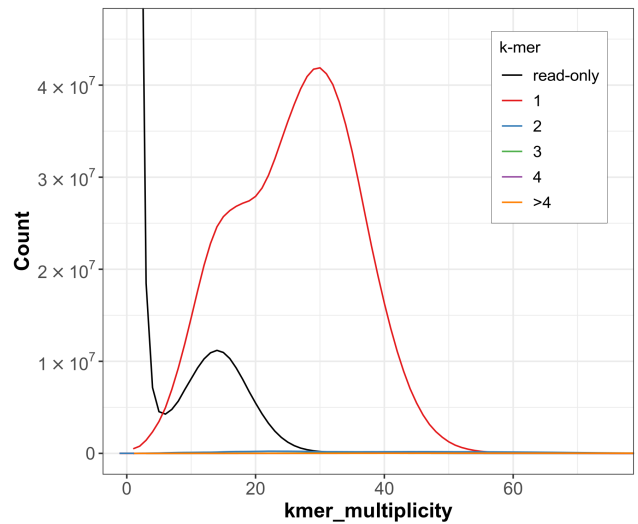


hap1 [\[LINK\]](#)

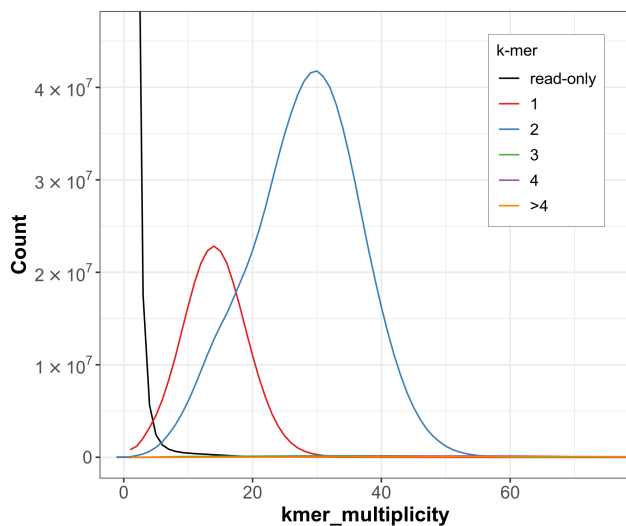
# K-mer spectra of curated assembly



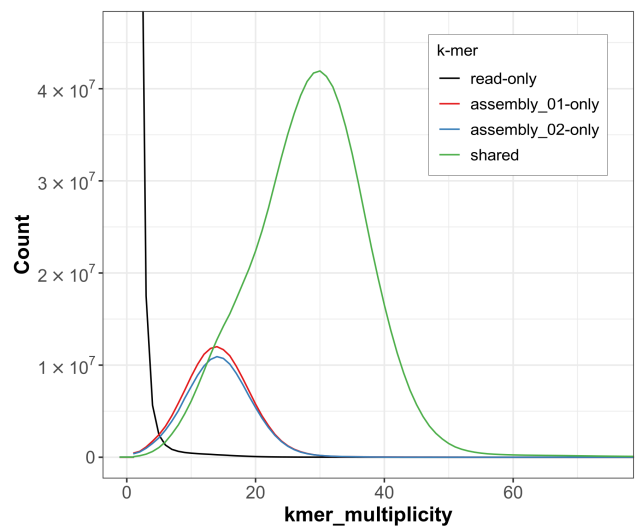
Distribution of k-mer counts per copy numbers found in **assembly\_02** (hapl.)



Distribution of k-mer counts per copy numbers found in **assembly\_01** (hapl.)

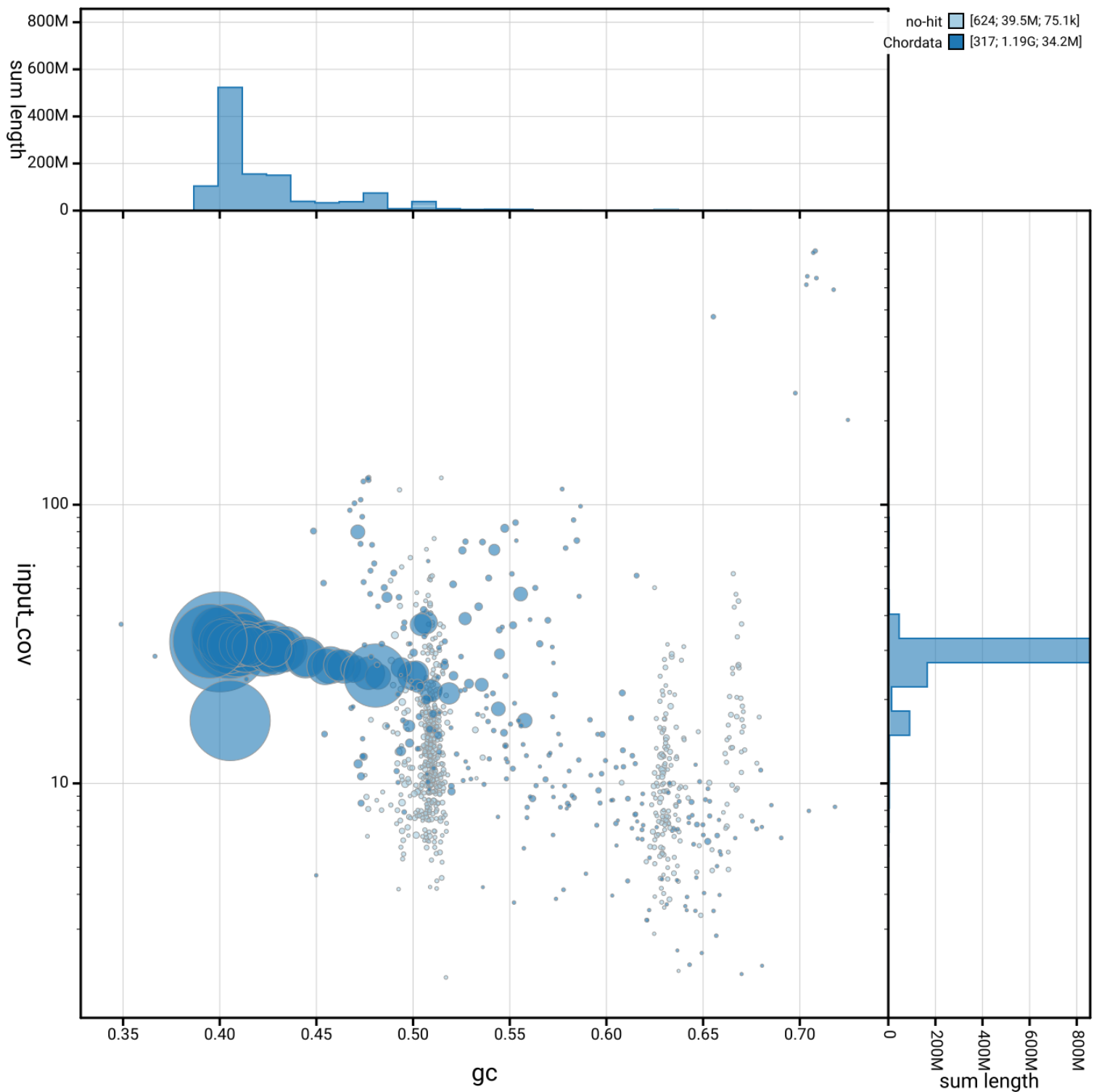


Distribution of k-mer counts per copy numbers found in asm (dipl.)



Distribution of k-mer counts coloured by their presence in reads/assemblies

# Post-curation contamination screening



**hap1.** 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	HiFi	HiC
Coverage	27.90X	66.69X

# Assembly pipeline

- **Hifiasm**
  - |\_ *ver*: Galaxy Version 0.24.0+galaxy0
  - |\_ *key param*: VGP Galaxy workflow VGP3 v 0.2.8
- **purge\_dups**
  - |\_ *ver*: Galaxy Version 1.2.6+galaxy0
  - |\_ *key param*: VGP Galaxy workflow VGP6 v 0.5
- **YaHS**
  - |\_ *ver*: Galaxy Version 1.2a.2+galaxy2
  - |\_ *key param*: VGP Galaxy workflow VGP8 v 1.4

# Curation pipeline

- **PretextMap**
  - |\_ *ver*: Galaxy Version 0.1.9+galaxy1
  - |\_ *key param*: VGP workflow VGP PretextMap generation v2
- **PretextView**
  - |\_ *ver*: 1.0.0
  - |\_ *key param*: NA

Submitter: Marcella Sozzoni

Affiliation: UNIFI

Date and time: 2025-07-10 14:46:27 CEST