

ERGA Assembly Report

v24.10.15

Tags: ERGA-BGE

TxID	48712
ToLID	qlOrcRanz
Species	Orchesella ranzii
Class	Collembola
Order	Entomobryomorpha

Genome Traits	Expected	Observed
Haploid size (bp)	244,559,092	254,308,816
Haploid Number	5 (source: ancestor)	6
Ploidy	2 (source: ancestor)	2
Sample Sex	NA	NA

EBP metrics summary and curation notes

Obtained EBP quality metric for hap1: 6.7.Q8

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 hap1
- . Kmer completeness value is less than 90 for hap1

Curator notes

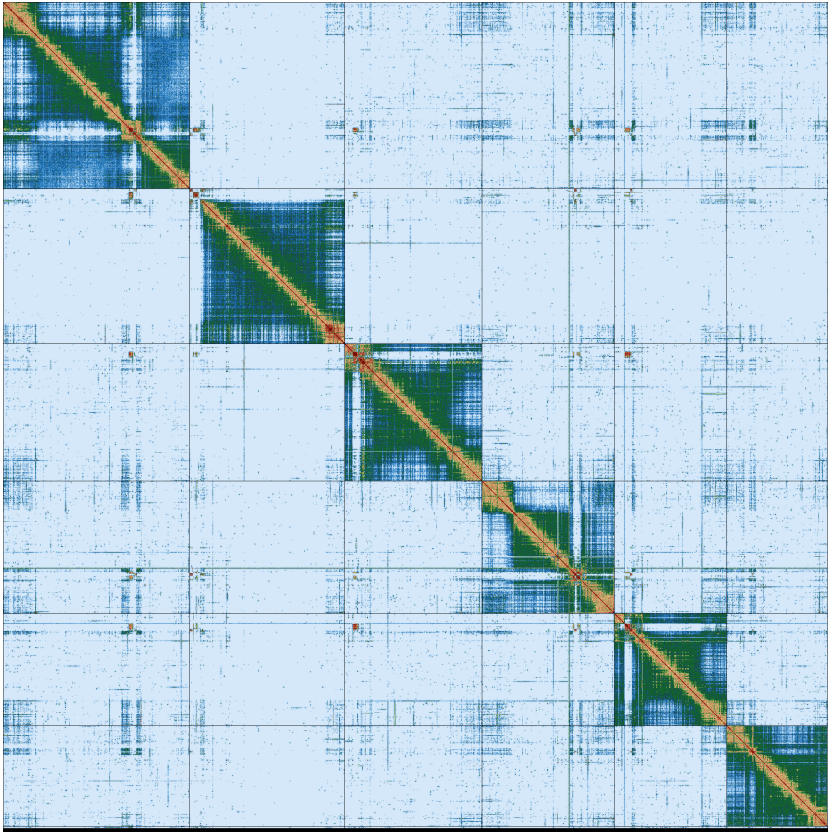
- . Interventions/Gb: None
- . Contamination notes: "No contaminants were found with the pipeline NCBI FCS GX v Galaxy Version 0.5.5+galaxy1. Scaffolds that are flagged as contaminants by BlobToolKit were confirmed with FCS output table and with contact in the map. None were removed."
- . Other observations: "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. 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	253,445,765	254,308,816
GC %	36.07	36.03
Gaps/Gbp	382.72	711.73
Total gap bp	19,400	36,200
Scaffolds	119	62
Scaffold N50	37,959,712	42,002,088
Scaffold L50	3	3
Scaffold L90	6	6
Contigs	216	243
Contig N50	3,944,353	3,623,572
Contig L50	15	17
Contig L90	66	80
QV	61.2585	8.25529
Kmer compl.	83.2072	0.161982
BUSCO sing.	91.0%	91.7%
BUSCO dupl.	4.0%	4.4%
BUSCO frag.	1.9%	1.2%
BUSCO miss.	3.1%	2.7%

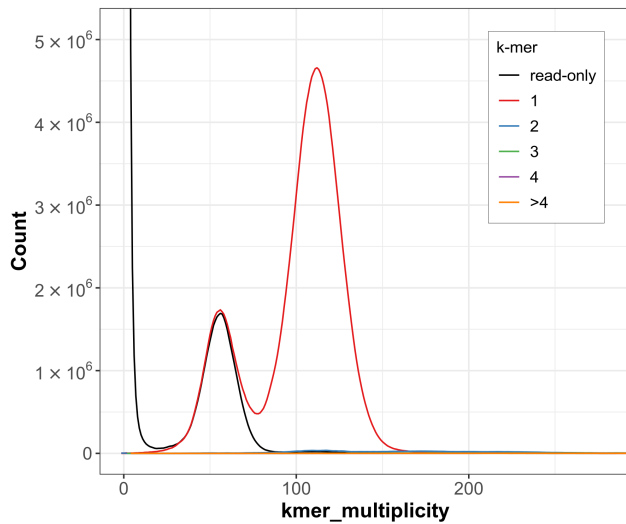
BUSCO: 5.8.0 (euk_genome_met, metaeuk) / Lineage: arthropoda_odb10 (genomes:90, BUSCOs:1013)

HiC contact map of curated assembly

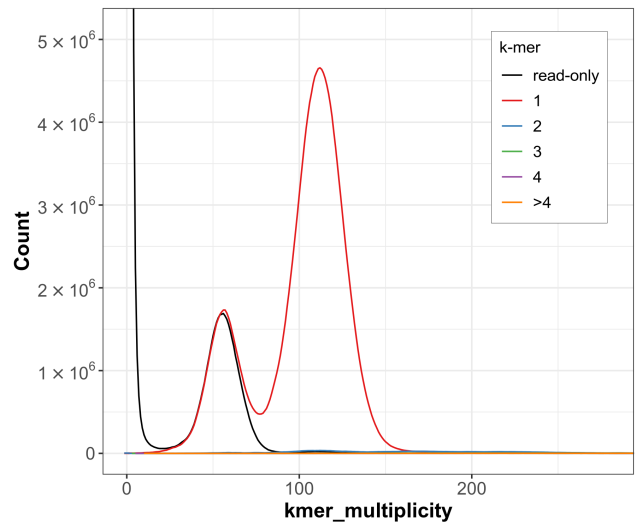


hap1 [\[LINK\]](#)

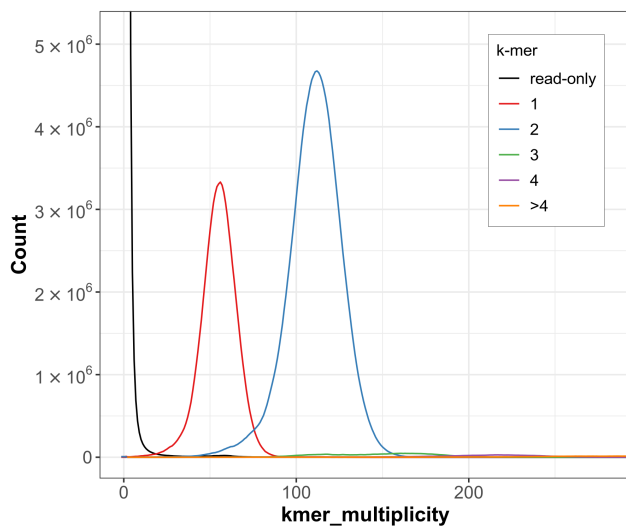
K-mer spectra of curated assembly



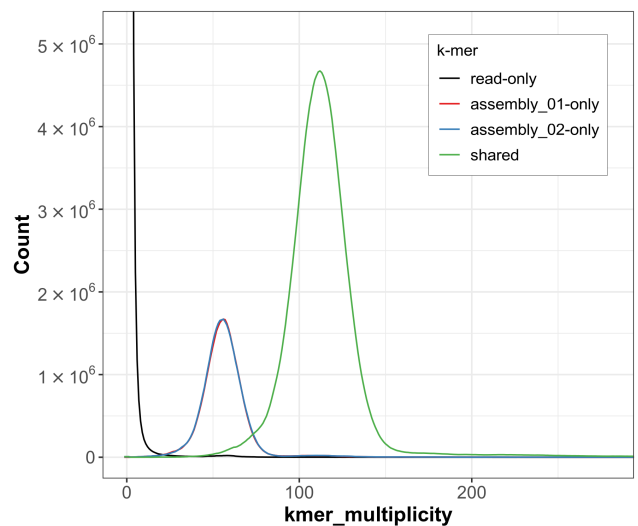
Distribution of k-mer counts per copy numbers found in **assembly_02** (hapl.)



Distribution of k-mer counts per copy numbers found in **assembly_01** (hapl.)

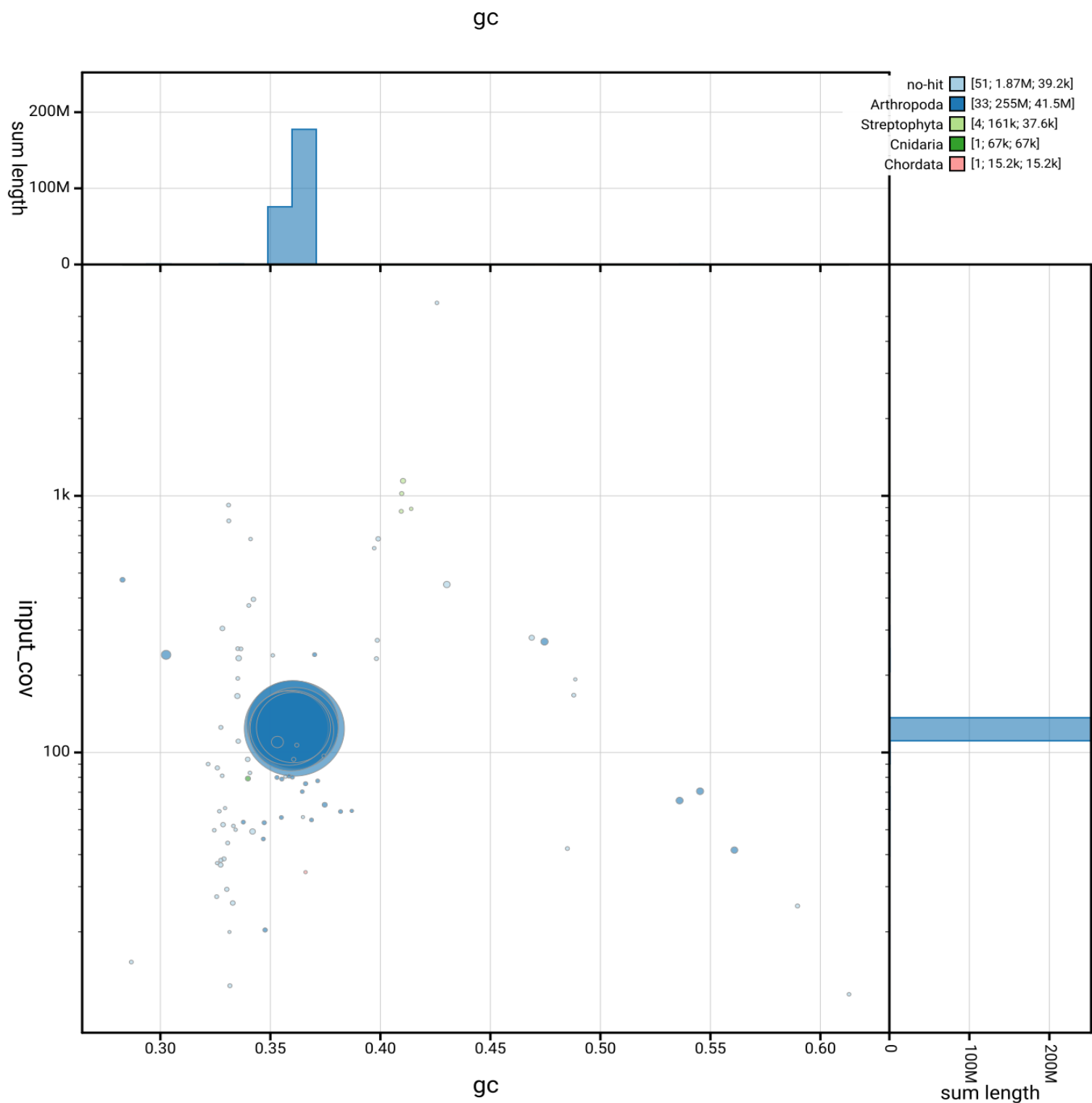


Distribution of k-mer counts per copy numbers found in **asm** (dipl.)



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	113.2X	586.87X

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: 2025-07-09 10:53:44 CEST