

# ERGA Assembly Report

v24.10.15

Tags: ERGA-BGE

TxID	391055
ToLID	<b>drPruRamb1</b>
Species	Prunus ramburei
Class	Magnoliopsida
Order	Rosales

Genome Traits	Expected	Observed
Haploid size (bp)	308,913,532	288,187,127
Haploid Number	8 (source: ancestor)	8
Ploidy	2 (source: ancestor)	2
Sample Sex	Unknown	Unknown

## EBP metrics summary and curation notes

Obtained EBP quality metric for collapsed: 7.7.Q68

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

- . Kmer completeness value is less than 90 for collapsed

### Curator notes

- . Interventions/Gb: 20
- . Contamination notes: ""
- . Other observations: "The assembly of *Prunus ramburei* (drPruRamb1) is based on 42X PacBio data and 170X Omni-C Hi-C data generated as part of the European Reference Genome Atlas (ERGA, <https://www.erga-biodiversity.eu/>) via the Biodiversity Genomics Europe project (BGE, <https://biodiversitygenomics.eu/>). The assembly process included the following steps: initial PacBio assembly generation with Hifiiasm, removal of contaminant sequences using Context, removal of haplotypic duplications using purge\_dups, and Hi-C-based scaffolding with YaHS. In total, 36 contigs were identified as contaminants (bacterial, archaeal, or viral), totaling 3.46 Mb (with the largest being 0.41 Mb). Additionally, 267 regions totaling 14.15 Mb (with the largest being 0.97 Mb) were identified as haplotypic duplications and removed. The mitochondrial and chloroplastic genomes were assembled using OATK. Finally, the primary assembly was analyzed and manually improved using Pretext. During manual curation, 3 haplotypic regions and 23 contaminant sequences were removed, totaling 5.23 Mb and 0.60 Mb, respectively (with the largest being 3.57 Mb and 0.062 Mb). Chromosome-scale scaffolds confirmed by Hi-C data were named in order of size. Chromosomes 3, 4, 5, 6, 7, and 8 are assembled in a single contig, and chromosomes 1 and 2 are composed of two contigs. Chromosomes 1, 2, 3, 4, and 8 have telomeric

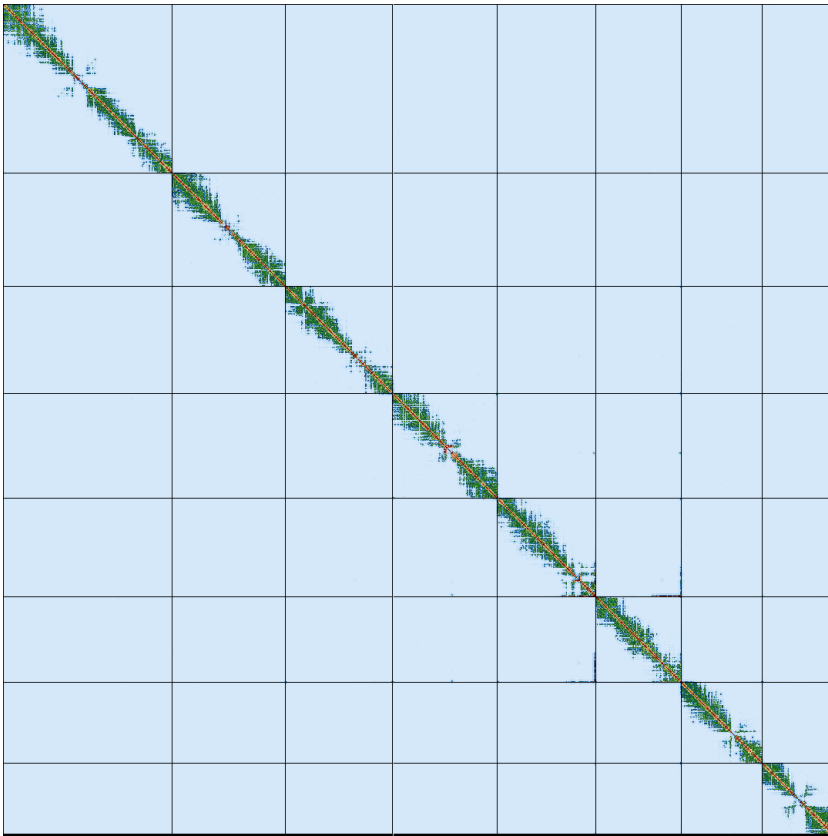
repeats at both extremities, and chromosomes 5, 6, and 7 have them only at the beginning. "

## Quality metrics table

Metrics	Pre-curation collapsed	Curated collapsed
Total bp	293,918,579	288,187,127
GC %	38	37.94
Gaps/Gbp	17.01	6.94
Total gap bp	500	400
Scaffolds	43	12
Scaffold N50	36,288,618	36,288,618
Scaffold L50	4	4
Scaffold L90	8	7
Contigs	48	14
Contig N50	29,554,979	33,968,494
Contig L50	5	4
Contig L90	9	8
QV	64.3195	68.0336
Kmer compl.	88.2555	88.083
BUSCO sing.	96.7%	97.2%
BUSCO dupl.	2.5%	1.9%
BUSCO frag.	0.4%	0.4%
BUSCO miss.	0.4%	0.5%

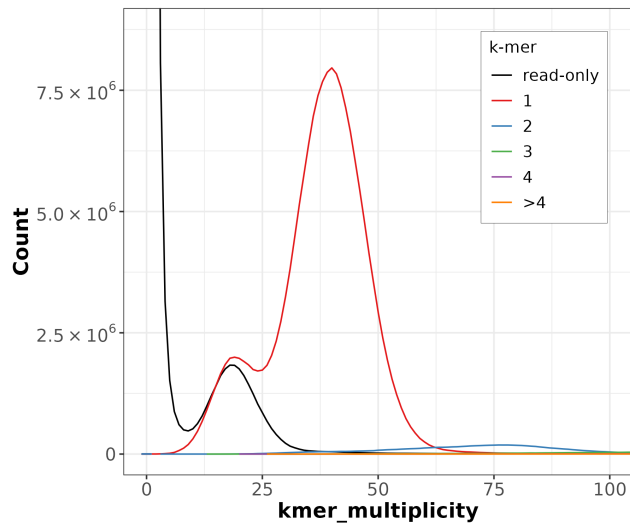
BUSCO: 5.4.3 (euk\_genome\_met, metaeuk) / Lineage: embryophyta\_odb10 (genomes:50, BUSCOs:1614)

# HiC contact map of curated assembly

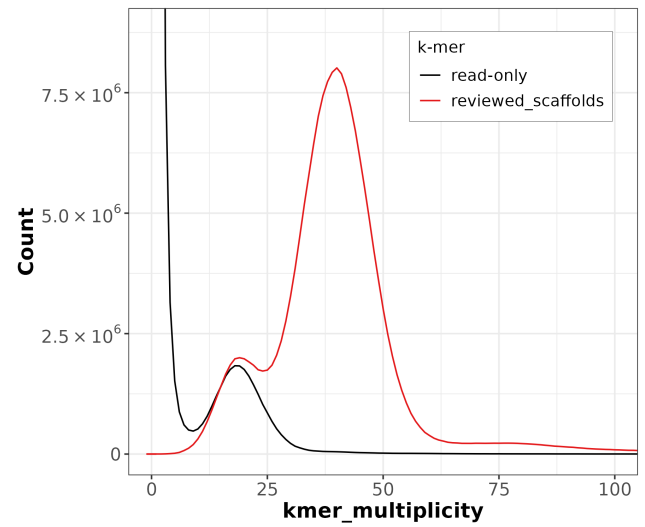


collapsed [\[LINK\]](#)

# K-mer spectra of curated assembly

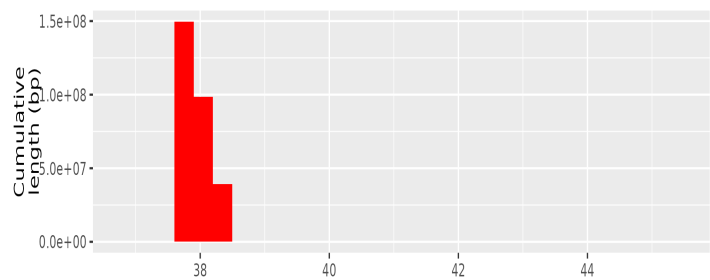


Distribution of k-mer counts per copy numbers found in asm

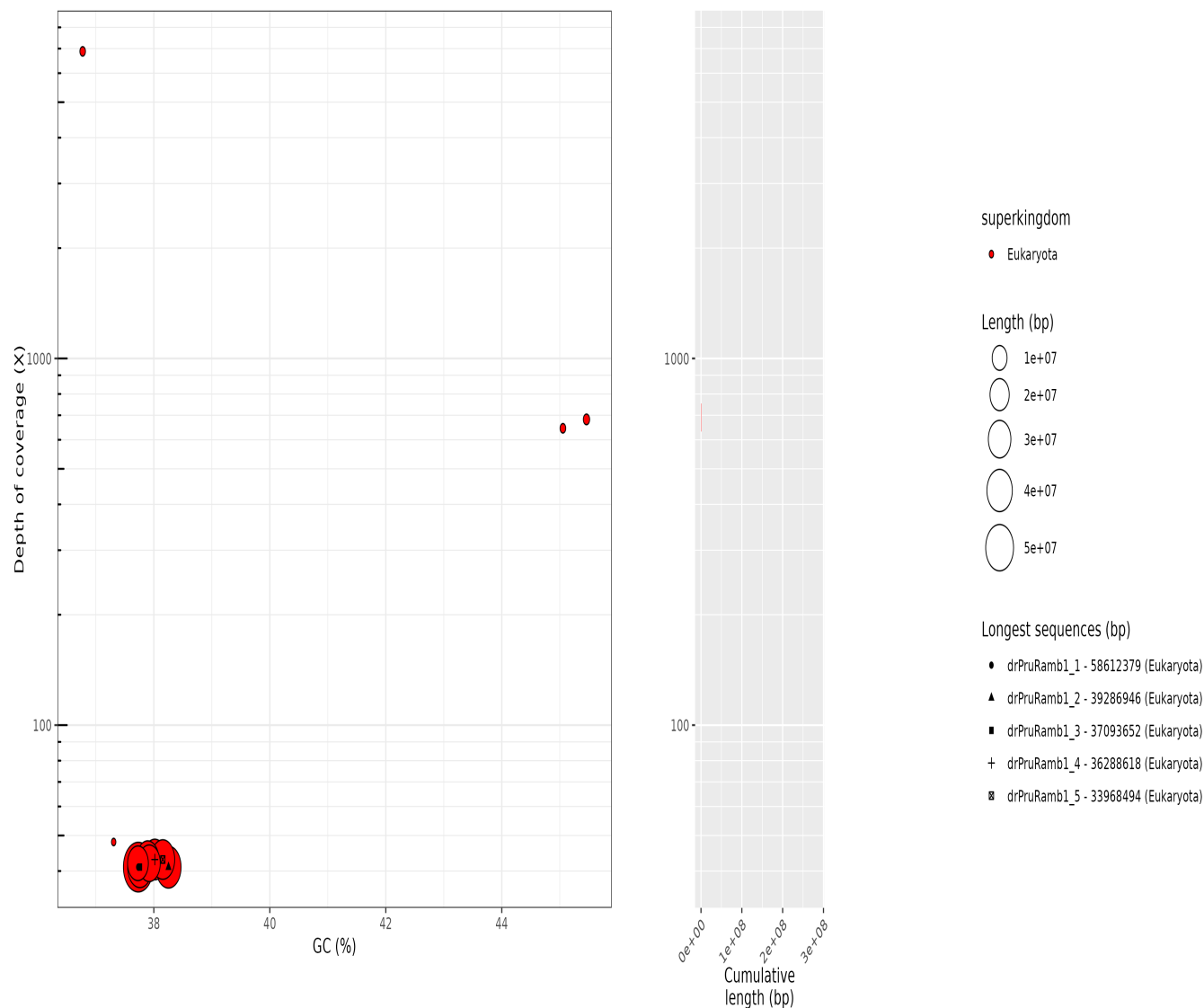


Distribution of k-mer counts coloured by their presence in reads/assemblies

# Post-curation contamination screening



## TAPAs summary Graph



**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	42	163

# 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: 2025-02-21 02:45:59 CET