

ERGA Assembly Report

v24.10.15

Tags: ERGA-BGE

TxID	3086330
ToLID	drCotNebr1
Species	Cotoneaster nebrodensis
Class	Magnoliopsida
Order	Rosales

Genome Traits	Expected	Observed
Haploid size (bp)	1,029,637,717	2,837,984,327
Haploid Number	34 (source: direct)	85
Ploidy	4 (source: ancestor)	2
Sample Sex	Unknown	Unknown

EBP metrics summary and curation notes

Obtained EBP quality metric for collapsed: 7.7.Q70

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

- . Observed Haploid size (bp) has >20% difference with Expected
- . Observed Haploid Number is different from Expected
- . Observed Ploidy is different from Expected
- . BUSCO single copy value is less than 90% for collapsed
- . BUSCO duplicated value is more than 5% for collapsed

Curator notes

- . Interventions/Gb: 122
- . Contamination notes: ""
- . Other observations: "The assembly of *Cotoneaster nebrodensis* (drCotNebr1) is based on 20X PacBio data and 98X OmniC 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/>). This genome is pentaploid and expected to have an AAABB structure. The assembly was performed using PacBio and Hi-C data, with the -n_hap parameter set to 2. Several tests were performed with this option, and only with this configuration were we able to retrieve the five chromosomes for each haplotype: in general, hap1 contains two A copies (often fused into a single scaffold) and one B chromosome, while hap2 contains one A and one B chromosome. The assembly process included the following steps: initial PacBio assembly generation with Hifiiasm, removal of contaminant sequences using Context, and Hi-C-based scaffolding with YaHS. The mitochondrial and chloroplastic genomes were assembled using OATK. Finally, the primary assembly was analyzed and

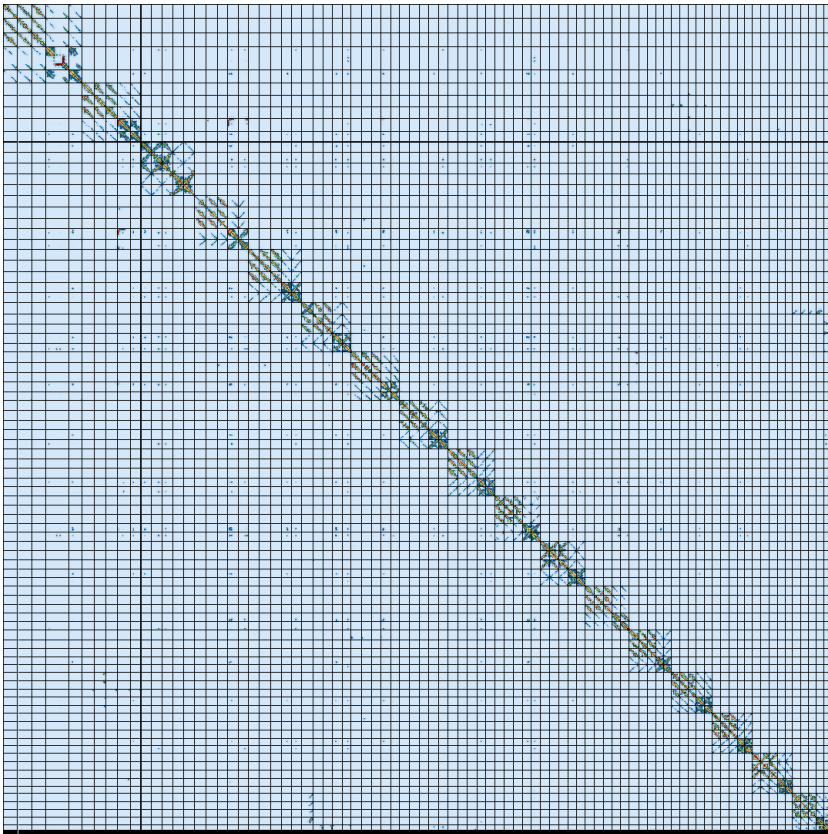
manually improved using Pretext. During manual curation, 895 contaminant sequences were removed, totaling 38.13 Mb (the largest being 0.51 Mb). Chromosome-scale scaffolds confirmed by Hi-C data were named in order of size (average size of the 3 A and 2 B chromosomes), and similarly, A and B copies were named according to their size. "

Quality metrics table

Metrics	Pre-curation collapsed	Curated collapsed
Total bp	2,881,091,729	2,837,984,327
GC %	38.29	38.17
Gaps/Gbp	172.16	150.81
Total gap bp	49,600	65,400
Scaffolds	1,636	420
Scaffold N50	60,616,789	32,541,186
Scaffold L50	19	37
Scaffold L90	55	75
Contigs	2,132	848
Contig N50	11,795,000	11,795,000
Contig L50	74	73
Contig L90	258	254
QV	65.609	70.7191
Kmer compl.	99.4557	99.3787
BUSCO sing.	0.1%	0.1%
BUSCO dupl.	97.6%	97.6%
BUSCO frag.	0.0%	0.0%
BUSCO miss.	2.3%	2.3%

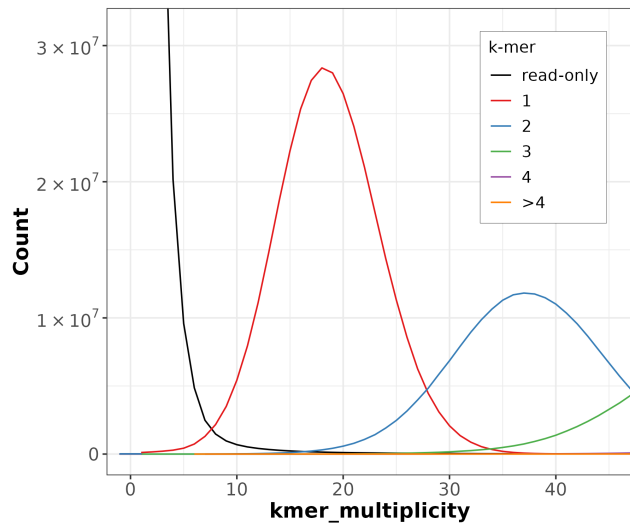
BUSCO: 6.0.0 (euk_genome_min, miniprot) / Lineage: rosaceae_odb12 (genomes:5, BUSCOs:10071)

HiC contact map of curated assembly

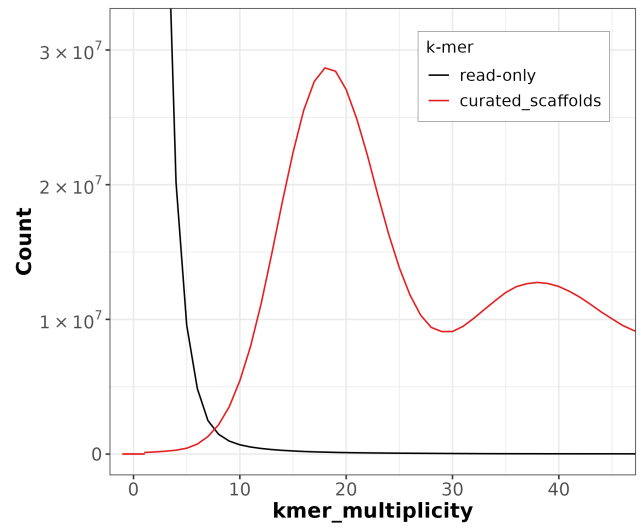


collapsed [\[LINK\]](#)

K-mer spectra of curated assembly

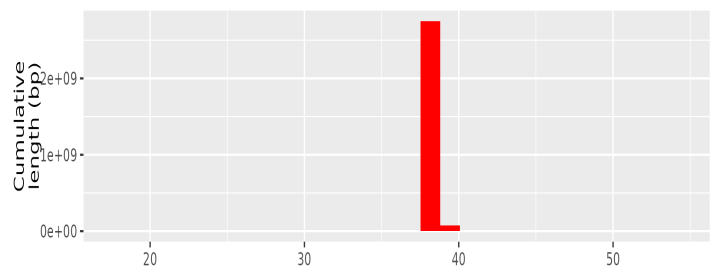


Distribution of k-mer counts per copy numbers found in asm

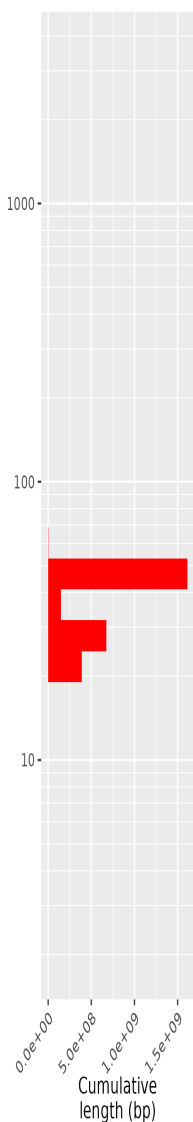
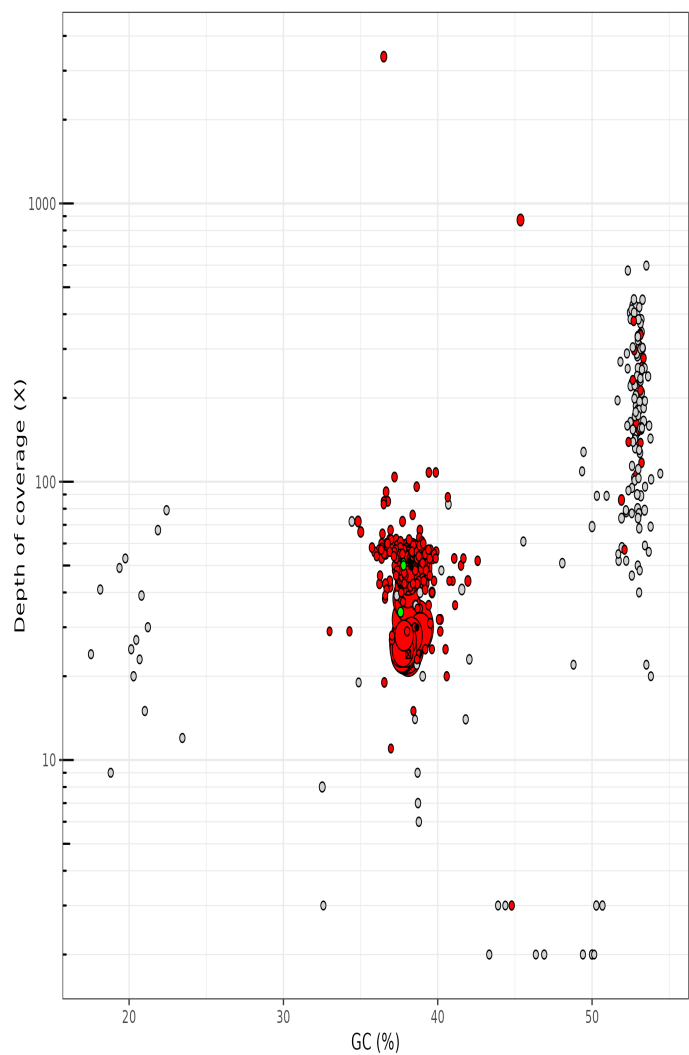


Distribution of k-mer counts coloured by their presence in reads/assemblies

Post-curation contamination screening



TAPAs summary Graph



- superkingdom
- Bacteria
 - Eukaryota
 - N/A
- Length (bp)
- 2e+07
 - 4e+07
 - 6e+07
- Longest sequences (bp)
- drCotNebr1_4 - 77790185 (Eukaryota)
 - ▲ drCotNebr1_1 - 48856001 (Eukaryota)
 - drCotNebr1_2 - 48518616 (Eukaryota)
 - + drCotNebr1_3 - 48140259 (Eukaryota)
 - ▣ drCotNebr1_5 - 45437404 (Eukaryota)

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	Omic
Coverage	53	264

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: 2025-12-05 19:01:47 CET