

# ERGA Assembly Report

v24.04.03\_beta

Tags: ERGA-BGE

TxID	436195
ToLID	<b>daArnMont1</b>
Species	Arnica montana
Class	Magnoliopsida
Order	Asterales

Genome Traits	Expected	Observed
Haploid size (bp)	1,372,599,626	1,368,732,344
Haploid Number	30 (source: direct)	19
Ploidy	4 (source: ancestor)	2
Sample Sex	unknown	unknown

## EBP metrics summary and curation notes

Obtained EBP quality metric for pri: 6.7.Q64

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

- . Observed Haploid Number is different from Expected
- . Observed Ploidy is different from Expected
- . BUSCO single copy value is less than 90% for pri
- . BUSCO duplicated value is more than 5% for pri

### Curator notes

. Interventions/Gb: 54  
. Contamination notes: "Contamination report for assembly labelled hap1; Total length of scaffolds removed: 35,391,045 (2.5 %); Scaffolds removed: 732 (72.1 %); Largest scaffold removed: (244,778); FCS-GX contaminant species (number of scaffolds; total length of scaffolds):Pseudomonas coleopterorum, g-proteobacteria (27; 1,969,142); Pseudomonas sp. UBA6276, g-proteobacteria (8; 642,804); Aureobasidium pullulans, ascomycetes (8; 257,660); Acinetobacter guillouiae, g-proteobacteria (5; 191,036); Sphingobium limneticum, a-proteobacteria (3; 226,823); Sphingomonas faeni, a-proteobacteria (2; 115,342); Aureobasidium vineae, ascomycetes (2; 96,738); Sphingomonas aurantiaca, a-proteobacteria (1; 64,657); Pseudomonas peli, g-proteobacteria (1; 64,353); Pseudomonas cyclaminis, g-proteobacteria (1; 57,154); Sphingomonas albertensis, a-proteobacteria (1; 52,857); Margaritispora aquatica, ascomycetes (1; 46,583); Pseudocercospora eumusae, ascomycetes (1; 45,786); Pseudomonas profundus, g-proteobacteria (1; 35,345); Brevundimonas intermedia, a-proteobacteria (1; 35,076); Sphingomonas panacis, a-proteobacteria (1; 3,000); Mitochondrion (10; 1,118,193); Plastid (658; 30,368,483); Barcodes (1; 13)"

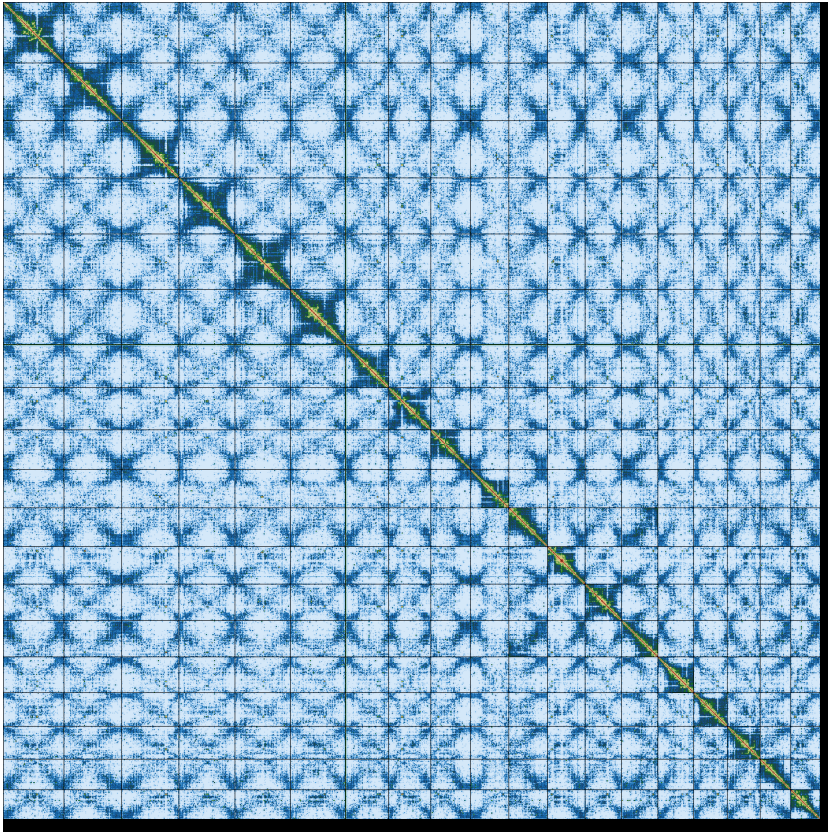
. Other observations: "This assembly is with a new Hi-C library; assembly was Hi-C phased. Heterozygous structural variation was observed in the following regions: chromosome 1, 34.25 Mb to 62.8 Mb; chromosome 4, 25.6 Mb to 29.7 Mb; chromosome 6, 28.8 Mb to 46.3 Mb; chromosome 18; 10.4 Mb to 12.4 Mb and chromosome 19, 14.3 Mb to 21.9 Mb."

# Quality metrics table

Metrics	Pre-curation pri	Curated pri
Total bp	1,407,012,512	1,368,732,344
GC %	38.26	38.21
Gaps/Gbp	467.66	485.85
Total gap bp	65,800	68,200
Scaffolds	1,015	267
Scaffold N50	63,256,956	68,753,894
Scaffold L50	9	8
Scaffold L90	19	17
Contigs	1,673	932
Contig N50	3,498,988	3,642,108
Contig L50	128	122
Contig L90	443	406
QV	60.3	64.4
Kmer compl.	98.64	98.58
BUSCO sing.	86.0%	86.1%
BUSCO dupl.	11.3%	11.2%
BUSCO frag.	0.7%	0.7%
BUSCO miss.	2.0%	2.0%

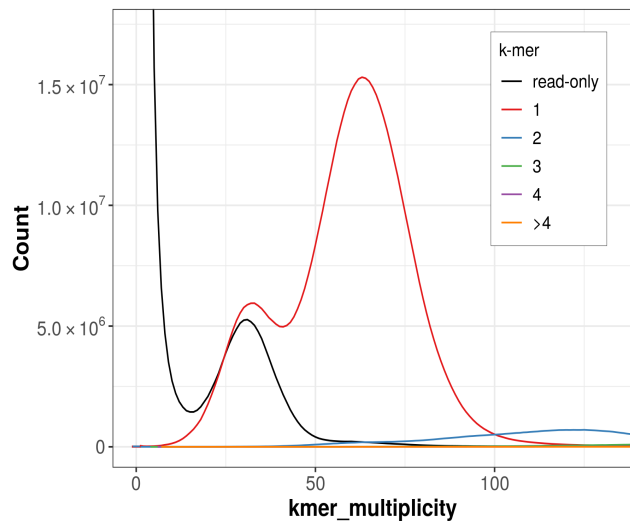
BUSCO 5.3.0 Lineage: eudicots\_odb10 (genomes:31, BUSCOs:2326)

# HiC contact map of curated assembly

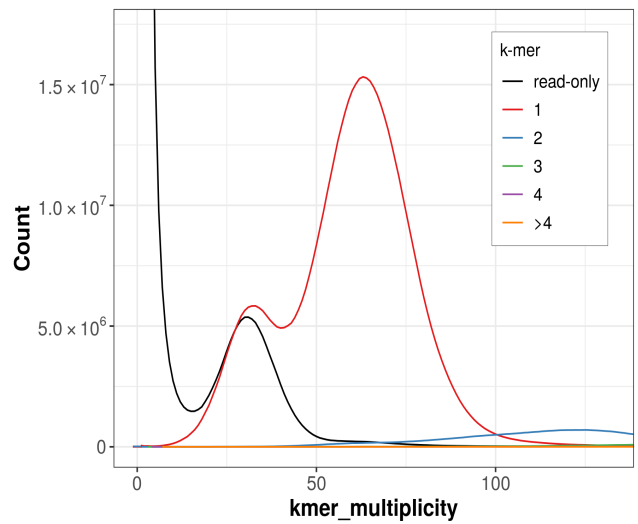


pri [\[LINK\]](#)

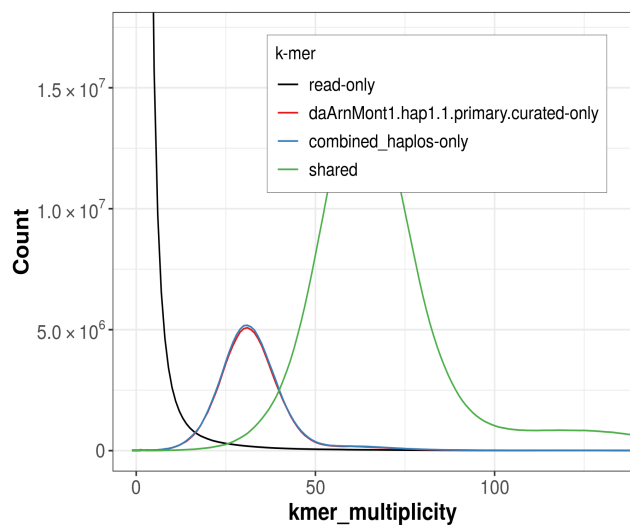
# K-mer spectra of curated assembly



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

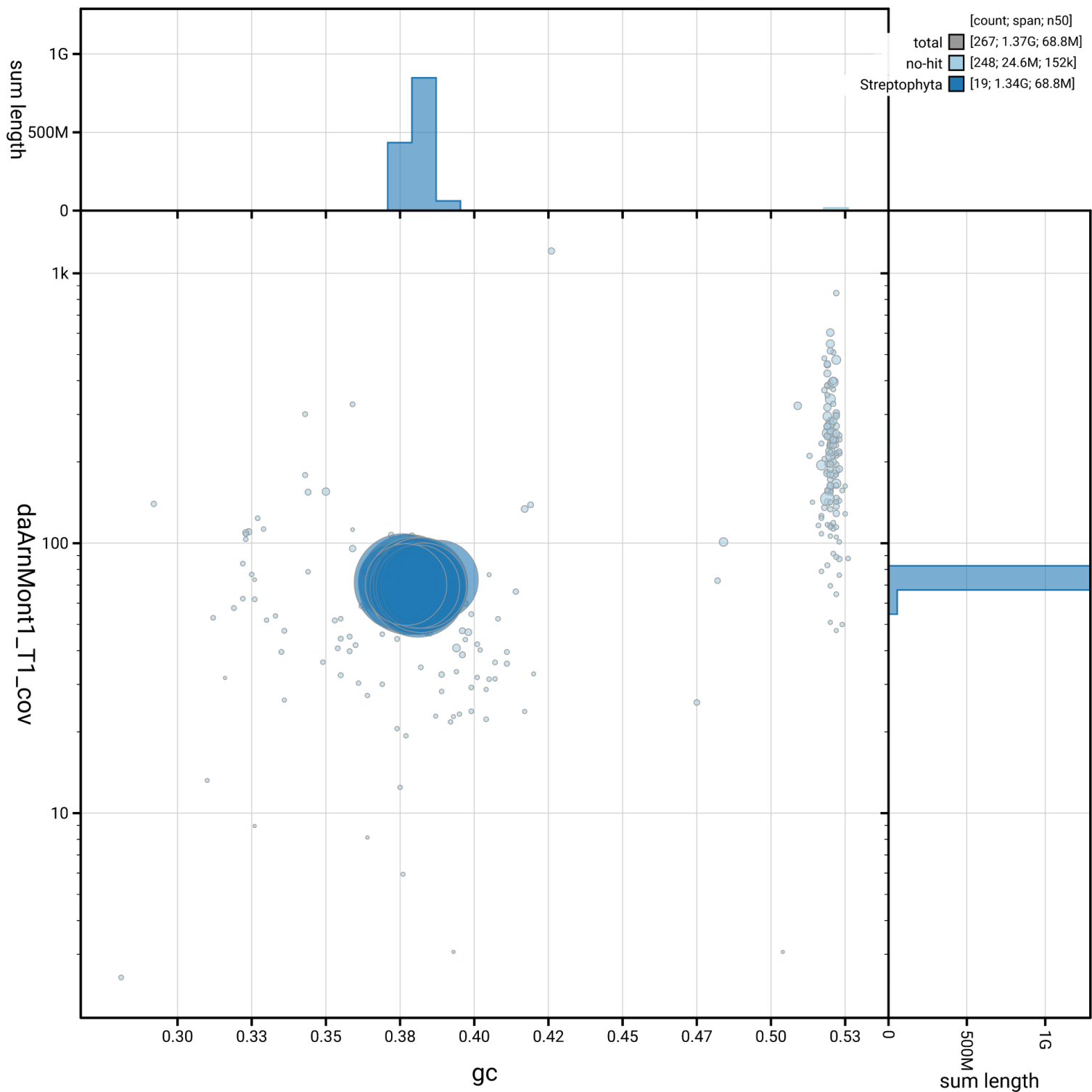


Distribution of k-mer counts per copy numbers found in asm



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

# Post-curation contamination screening



**pri.** 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	Arima v2
Coverage	65x	46x

# Assembly pipeline

- **hifiasm**
  - |\_ *ver*: 0.19.8-r603
  - |\_ *key param*: --h1/--h2
- **yahs**
  - |\_ *ver*: 1.2.2
  - |\_ *key param*: NA

# Curation pipeline

- **hifiasm**
  - |\_ *ver*: 0.19.8-r603
  - |\_ *key param*: --h1/-h2
- **yahs**
  - |\_ *ver*: 1.2.2
  - |\_ *key param*: NA
- **TreeVal**
  - |\_ *ver*: 1.2.1
  - |\_ *key param*: NA

Submitter: Thomas Mathers

Affiliation: WSI

Date and time: 2025-03-03 18:00:25 CET