

ERGA Assembly Report

v24.04.03_beta

Tags: ERGA-BGE

TxID	436195
ToLID	daArnMont1
Species	Arnica montana
Class	Magnoliopsida
Order	Asterales

Genome Traits	Expected	Observed
Haploid size (bp)	1,372,599,626	1,406,759,939
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
- . Assembly length loss > 3% for pri
- . More than 1000 gaps/Gbp for pri

Curator notes

- . Interventions/Gb: 328
- . Contamination notes: "Total length of scaffolds removed: 33,962,772 (2.3 %);Scaffolds removed: 685 (38.6 %);Largest scaffold removed: (763,904);FCS-GX contaminant species (number of scaffolds; total length of scaffolds):Pseudomonas coleopterorum, g-proteobacteria (16; 2,098,136);Aureobasidium pullulans, ascomycetes (8; 257,660);Pseudomonas sp. UBA6276, g-proteobacteria (7; 772,951);Acinetobacter guillouiae, g-proteobacteria (6; 222,208);Sphingobium limneticum, a-proteobacteria (2; 167,804);Aureobasidium vineae, ascomycetes (2; 96,738);Pseudocercospora eumusae, ascomycetes (1; 92,569);Sphingomonas aurantiaca, a-proteobacteria (1; 64,657);Pseudomonas cyclaminis, g-proteobacteria (1; 57,154);Sphingomonas faeni, a-proteobacteria (1; 36,368);Pseudomonas profundus, g-proteobacteria (1; 35,345);Brevundimonas intermedia, a-proteobacteria (1; 35,076);Mitochondrion (10; 1,118,193);Plastid (628; 28,907,913)"
- . Other observations: "Hi-C quality not great. Attempting a new library, but

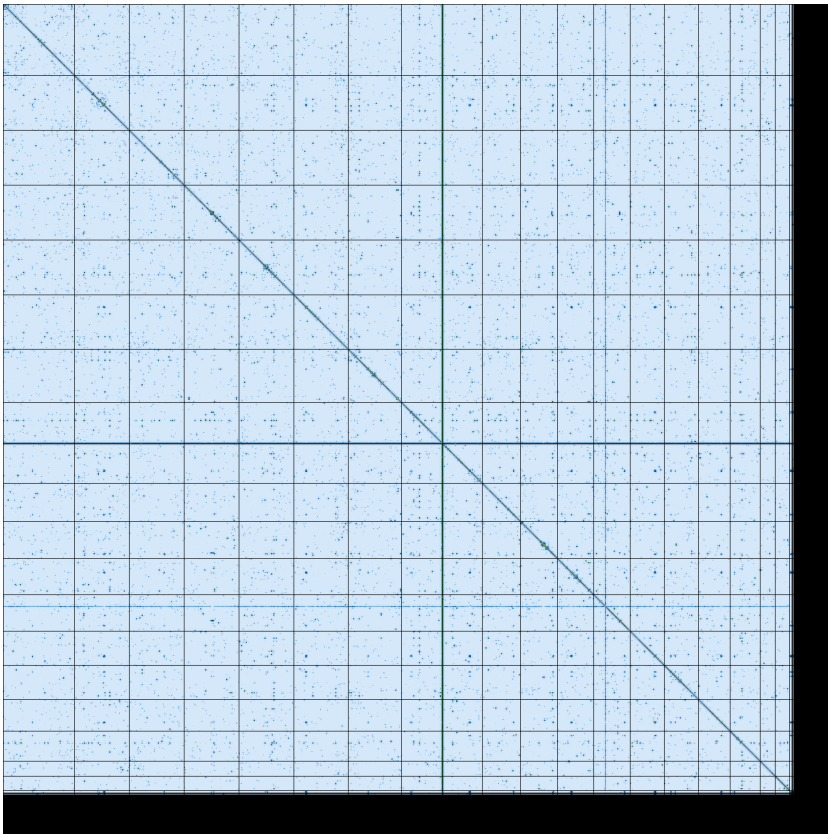
prgressing with this version for now."

Quality metrics table

Metrics	Pre-curation pri	Curated pri
Total bp	1,451,215,518	1,406,759,939
GC %	38.35	38.28
Gaps/Gbp	1,050.15	1,108.93
Total gap bp	304,800	312,000
Scaffolds	1,776	1,032
Scaffold N50	54,488,129	70,274,966
Scaffold L50	10	8
Scaffold L90	38	17
Contigs	3,300	2,592
Contig N50	3,066,000	3,141,178
Contig L50	135	129
Contig L90	611	527
QV	60.6	64.5
Kmer compl.	98.63	98.59
BUSCO sing.	86.2%	85.9%
BUSCO dupl.	11.3%	11.5%
BUSCO frag.	0.7%	0.5%
BUSCO miss.	1.8%	2.1%

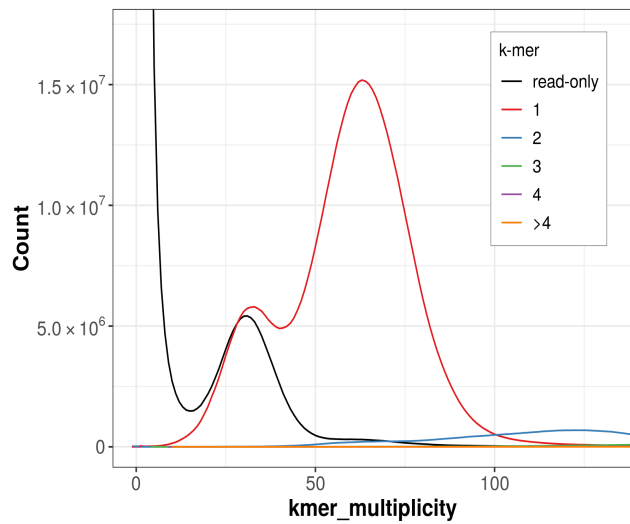
Warning: BUSCO versions or lineage datasets are not the same across results

HiC contact map of curated assembly

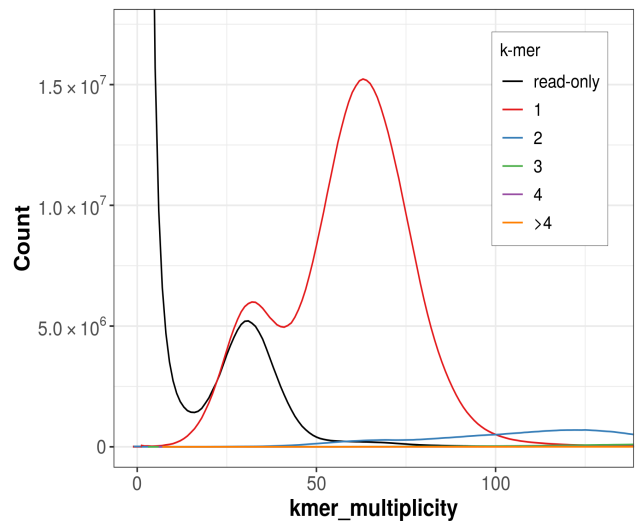


pri [\[LINK\]](#)

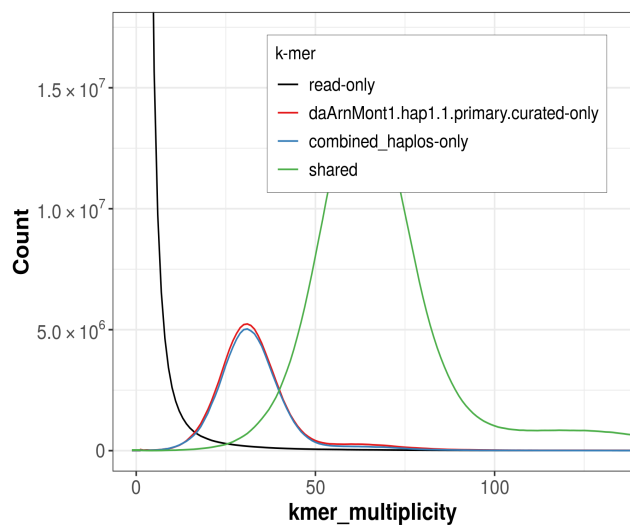
K-mer spectra of curated assembly



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

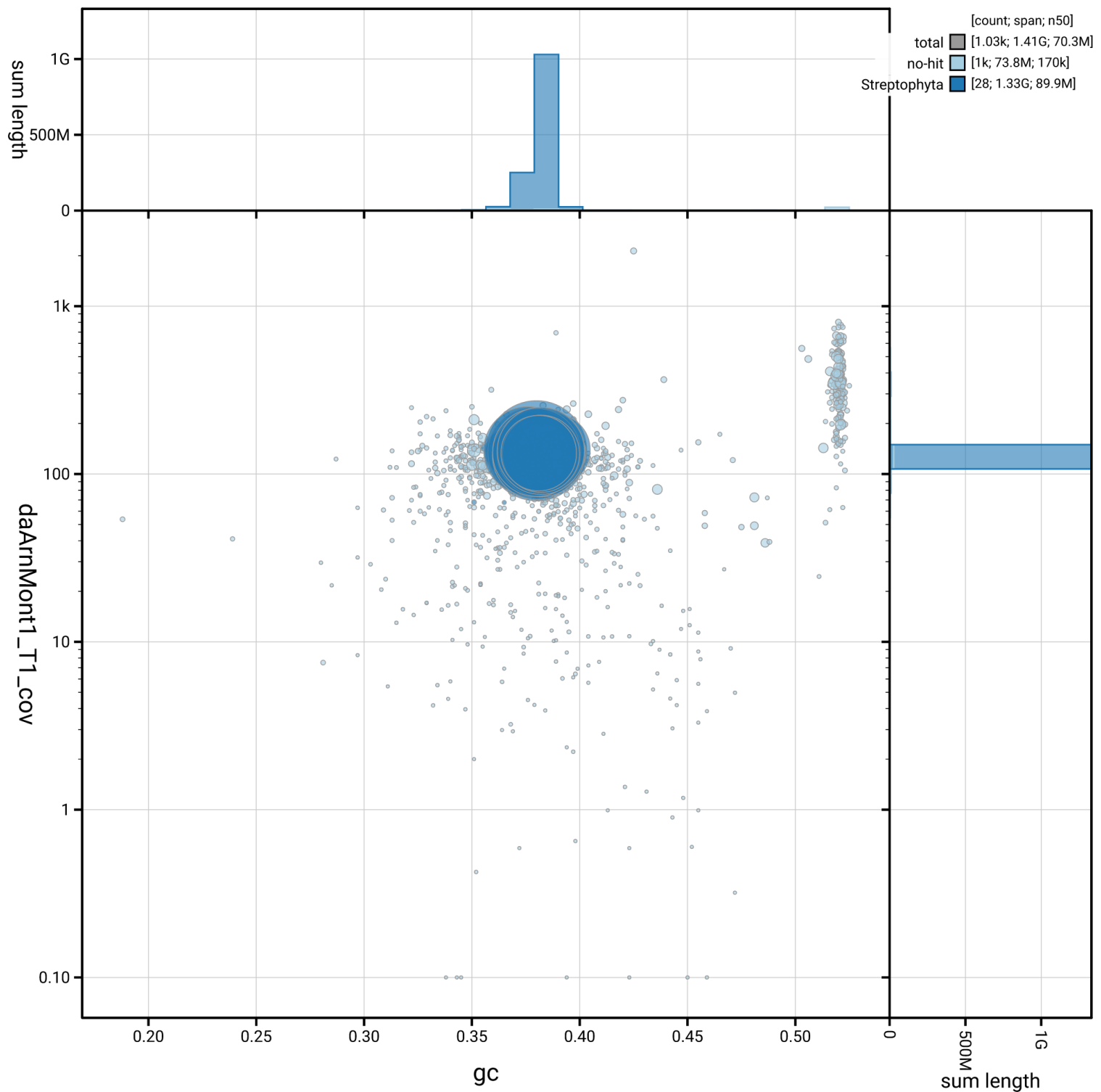


Distribution of k-mer counts per copy numbers found in asm



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*: --primary
- **purge_dups**
 - |_ *ver*: 1.2.5
 - |_ *key param*: -e
- **yahs**
 - |_ *ver*: 1.2a.2
 - |_ *key param*: NA

Curation pipeline

- **hifiasm**
 - |_ *ver*: 0.19.8-r603
 - |_ *key param*: --primary
- **purge_dups**
 - |_ *ver*: 1.2.5
 - |_ *key param*: -e
- **yahs**
 - |_ *ver*: 1.2a.2
 - |_ *key param*: NA
- **TreeVal**
 - |_ *ver*: 1.1.1
 - |_ *key param*: NA

Submitter: Dominic Absolon

Affiliation: WSI

Date and time: 2024-11-14 11:34:22 CET