

ERGA Assembly Report

v24.04.03_beta

Tags: ERGA-BGE

TxID	1940801
ToLID	daAveFist1
Species	Avellara fistulosa
Class	Magnoliopsida
Order	Asterales

Genome Traits	Expected	Observed
Haploid size (bp)	612,541,738	1,223,654,298
Haploid Number	7 (source: direct)	7
Ploidy	2 (source: ancestor)	4
Sample Sex	H	H

EBP metrics summary and curation notes

Obtained EBP quality metric for pri: 6.8.Q60

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 Ploidy is different from Expected

Curator notes

. Interventions/Gb: 77
. Contamination notes: "Contamination report for assembly labelled hap1; Total length of scaffolds removed: 49,238,713 (3.9 %); Scaffolds removed: 1133 (92.3 %); Largest scaffold removed: (263,852); FCS-GX contaminant species (number of scaffolds; total length of scaffolds): Filobasidium floriforme, basidiomycetes (26; 1,144,140); Pseudomonas coleopterorum, g-proteobacteria (14; 737,313); Aureobasidium pullulans, ascomycetes (13; 541,546); Lelliottia amnigena, g-proteobacteria (5; 254,275); Macroventuria anomochaeta, ascomycetes (4; 166,635); Pseudomonas sp. UBA6276, g-proteobacteria (3; 155,854); Brevundimonas sp. UBA6550, a-proteobacteria (3; 168,410); Alternaria alternata, ascomycetes (3; 110,610); Aureobasidium vineae, ascomycetes (2; 160,875); Ascochyta lentis, ascomycetes (2; 63,087); Thrips palmi, insects (2; 53,697); Sphingomonas sp. UBA978, a-proteobacteria (1; 59,444); Sclerotiophoma versabilis, ascomycetes (1; 58,308); Ascochyta rabiei, ascomycetes (1; 56,158); Pseudomonas profundus, g-proteobacteria (1; 53,444); Pseudomonas peli, g-proteobacteria (1; 46,132); Ascochyta koolunga, ascomycetes (1; 45,307); Naganishia liquefaciens, basidiomycetes (1; 44,149); Cladosporium phlei, ascomycetes (1; 43,608); Melanodothia sp. VS III C KN 2, ascomycetes (1; 42,524); Rhodotorula babjevae, basidiomycetes (1; 35,239); Corallococcus sicarius, d-proteobacteria (1;

32,100); *Sarcomyxa edulis*, basidiomycetes (1; 22,244); *Brachyptera putata*, insects (1; 19,121); Mitochondrion (23; 1,503,637); Plastid (1020; 43,620,856)"

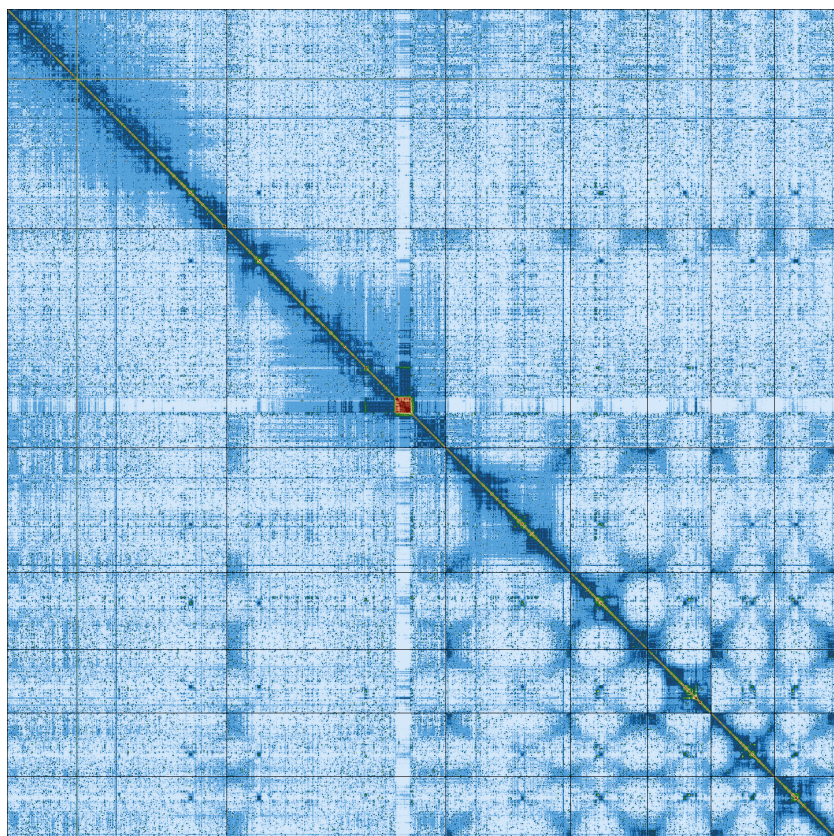
. Other observations: "This genome has been assembled using PacBio and HiC data and phased. The result is two curated haplotypes. The exact order and orientation of the contigs on chromosome 2 (245 - 275 Mbp) are unknown."

Quality metrics table

Metrics	Pre-curation pri	Curated pri
Total bp	1,257,861,400	1,223,654,298
GC %	36.63	36.67
Gaps/Gbp	219.42	275.4
Total gap bp	27,600	40,400
Scaffolds	1,228	65
Scaffold N50	183,878,289	321,656,683
Scaffold L50	3	2
Scaffold L90	7	6
Contigs	1,504	402
Contig N50	6,302,318	6,377,107
Contig L50	63	61
Contig L90	199	184
QV	57.8	60.1
Kmer compl.	99.17	99.48
BUSCO sing.	92.9%	92.9%
BUSCO dupl.	4.0%	4.1%
BUSCO frag.	0.5%	0.3%
BUSCO miss.	2.6%	2.7%

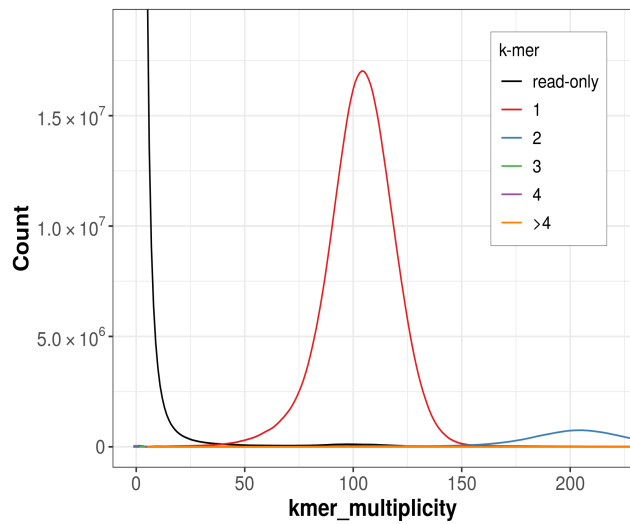
Warning: BUSCO versions or lineage datasets are not the same across results

HiC contact map of curated assembly

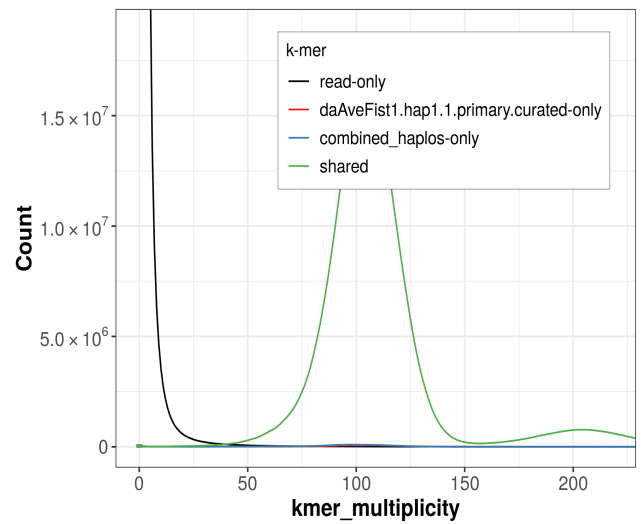


pri [\[LINK\]](#)

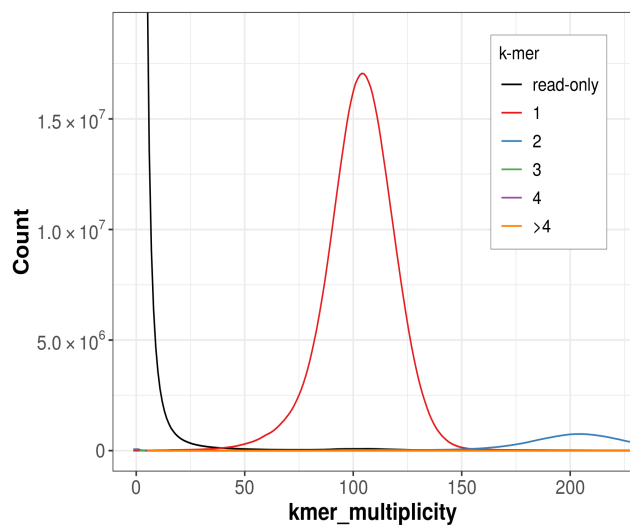
K-mer spectra of curated assembly



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

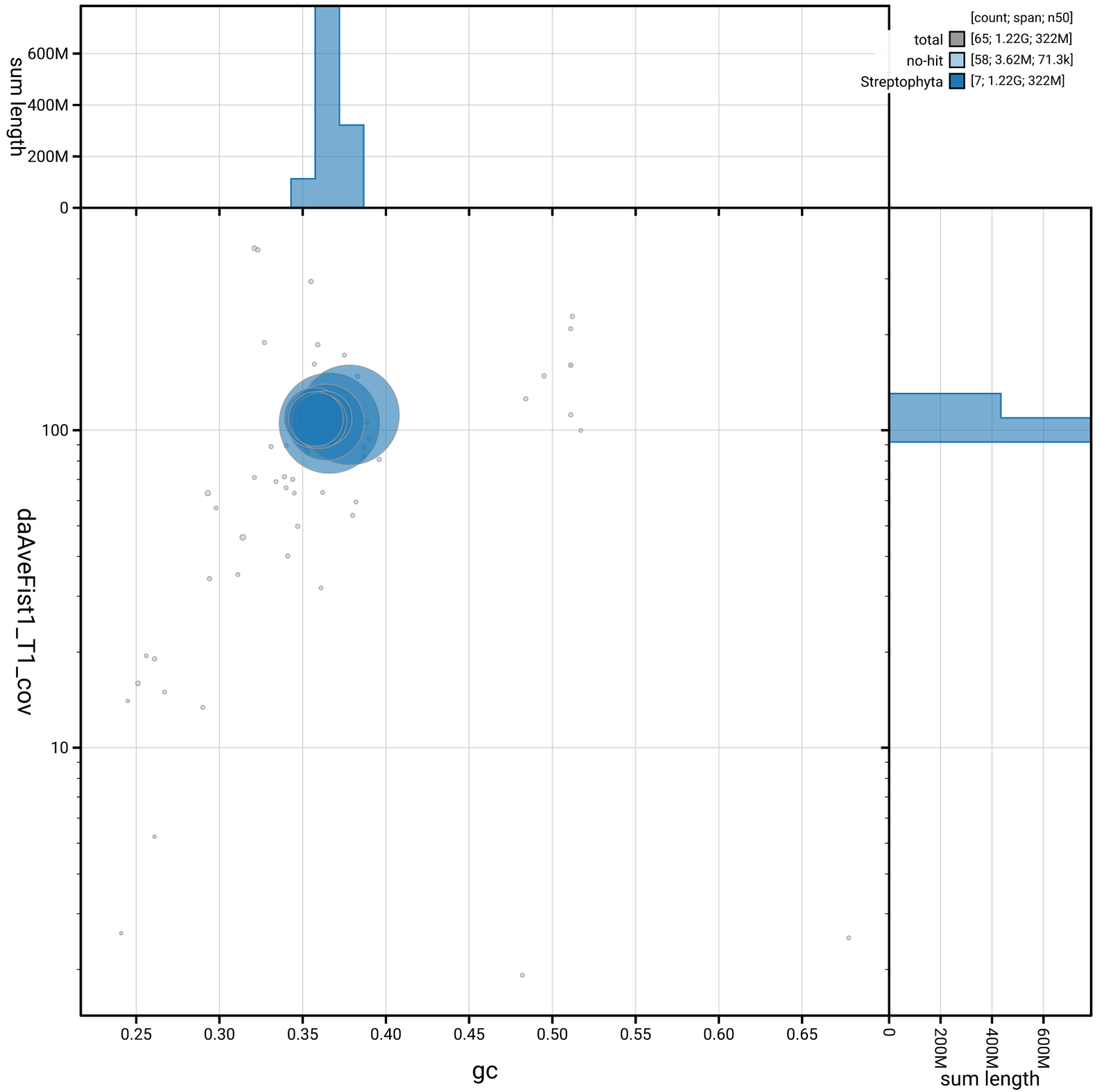


Distribution of k-mer counts coloured by their presence in reads/assemblies



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

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	51x	213x

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*: 1.1.1
 - |_ *key param*: NA

Submitter: Danil Zilov

Affiliation: WSI

Date and time: 2025-01-31 14:23:56 CET