

ERGA Assembly Report

v24.04.03_beta

Tags: ERGA-BGE

| | |
|---------|---------------------------|
| TxID | 1929230 |
| ToLID | daGyrOppo1 |
| Species | Gyrocaryum oppositifolium |
| Class | Magnoliopsida |
| Order | Boraginales |

| Genome Traits | Expected | Observed |
|-------------------|-----------------------|-------------|
| Haploid size (bp) | 334,951,947 | 621,021,262 |
| Haploid Number | 12 (source: ancestor) | 24 |
| Ploidy | 4 (source: ancestor) | 2 |
| Sample Sex | U | U |

EBP metrics summary and curation notes

Obtained EBP quality metric for pri: 6.7.Q63

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 pri
- . BUSCO duplicated value is more than 5% for pri
- . Assembly length loss > 3% for pri
- . Not 90% of assembly in chromosomes for pri

Curator notes

. Interventions/Gb: 35
. Contamination notes: "Contamination report for assembly labelled hap1;Total length of scaffolds removed: 142,386,935 (18.7 %);Scaffolds removed: 6527 (67.5 %);Largest scaffold removed: (379,346);FCS-GX contaminant species (number of scaffolds; total length of scaffolds):Pseudomonas viridiflava, g-proteobacteria (98; 3,163,964);Myzus persicae, insects (65; 3,364,089);Pseudomonas lactis, g-proteobacteria (26; 483,773);Acyrtosiphon pisum, insects (20; 533,071);Pseudomonas lactucae, g-proteobacteria (16; 203,164);Pseudomonas coleopterorum, g-proteobacteria (6; 240,855);Duganella phyllosphaerae, b-proteobacteria (5; 101,492);Methylobacterium bullatum, a-proteobacteria (5; 111,448);Buchnera aphidicola, g-proteobacteria (3; 149,086);Aphis gossypii, insects (3; 122,405);Pseudomonas salomonii, g-proteobacteria (2; 143,427);Pseudomonas sp. UBA6276, g-proteobacteria (2; 145,612);Candidatus Regiella insecticola, g-proteobacteria (2; 69,635);Diuraphis noxia, insects (2;

64,706);Pseudomonas petroselini, g-proteobacteria (1; 69,094);Pseudomonas zeae, g-proteobacteria (1; 21,645);Sphingomonas bisphenolicum, a-proteobacteria (1; 20,791);Pseudomonas ovata, g-proteobacteria (1; 18,000);Methylobacterium mesophilicum, a-proteobacteria (1; 16,382);Streptomyces phage SF3, prokaryotic viruses (1; 13,311);Pseudomonas tremae, g-proteobacteria (1; 7,000);Pseudomonas cyclaminis, g-proteobacteria (1; 3,000);Mitochondrion (20; 388,974);Plastid (6244; 132,931,987);Barcodes (2; 24)"

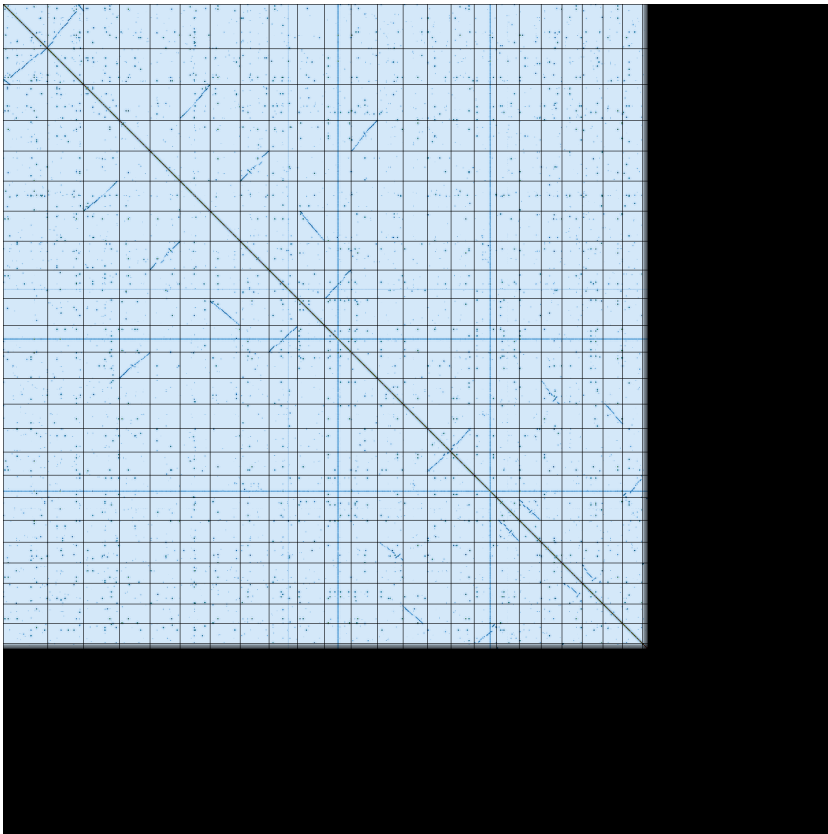
. Other observations: "Assembly was Hi-C phased; "

Quality metrics table

| Metrics | Pre-curation pri | Curated pri |
|--------------|---------------------|----------------|
| Total bp | 762,383,583 | 621,021,262 |
| GC % | 36.45 | 36.13 |
| Gaps/Gbp | 628.29 | 739.11 |
| Total gap bp | 47,900 | 47,400 |
| Scaffolds | 9,665 | 3,128 |
| Scaffold N50 | 16,365,971 | 18,143,900 |
| Scaffold L50 | 18 | 14 |
| Scaffold L90 | 5,075 | 828 |
| Contigs | 10,144 | 3,587 |
| Contig N50 | 2,328,000 | 3,046,908 |
| Contig L50 | 76 | 50 |
| Contig L90 | 5,554 | 1,194 |
| QV | 54.6 | 63.2 |
| Kmer compl. | 99.70 | 96.07 |
| BUSCO sing. | 10.1% | 6.9% |
| BUSCO dupl. | 85.6% | 90.2% |
| BUSCO frag. | 1.3% | 0.3% |
| BUSCO miss. | 3.0% | 2.6% |

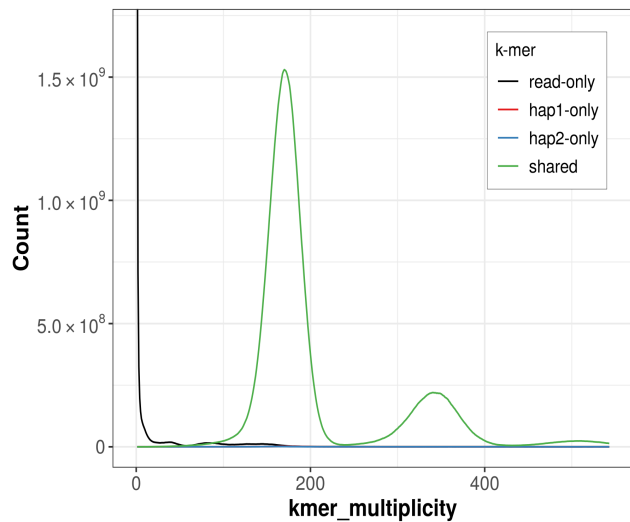
Warning: BUSCO versions or lineage datasets are not the same across results

HiC contact map of curated assembly

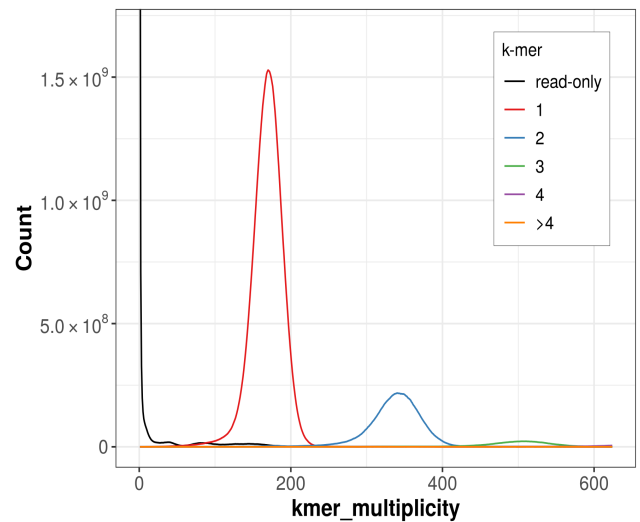


pri [\[LINK\]](#)

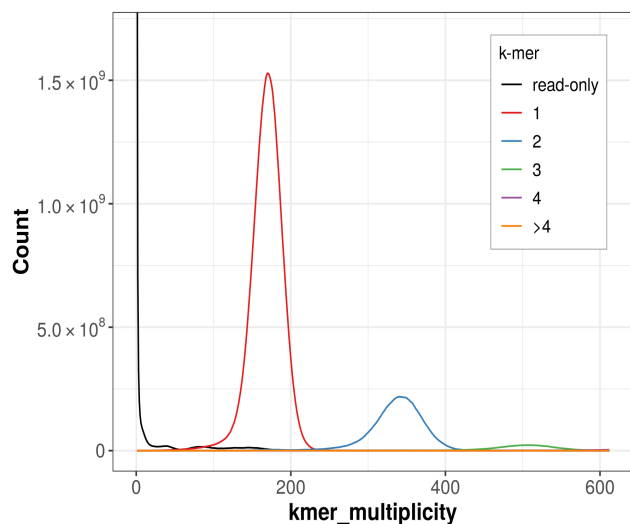
K-mer spectra of curated assembly



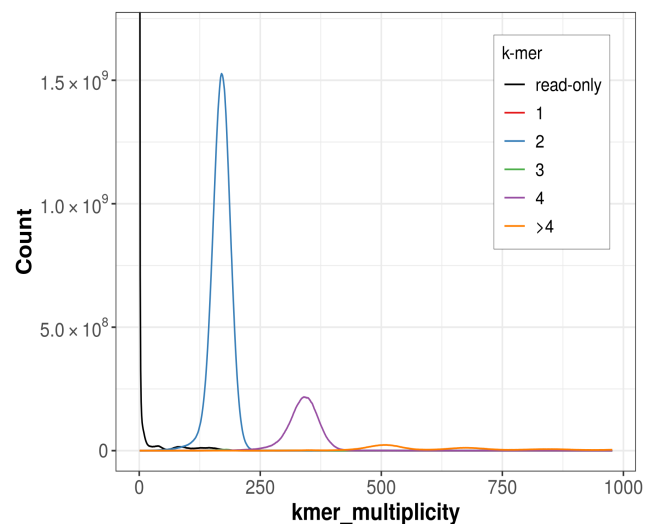
Distribution of k-mer counts coloured by their presence in reads/assemblies



Distribution of k-mer counts per copy numbers found in **hap2** (hap1.)

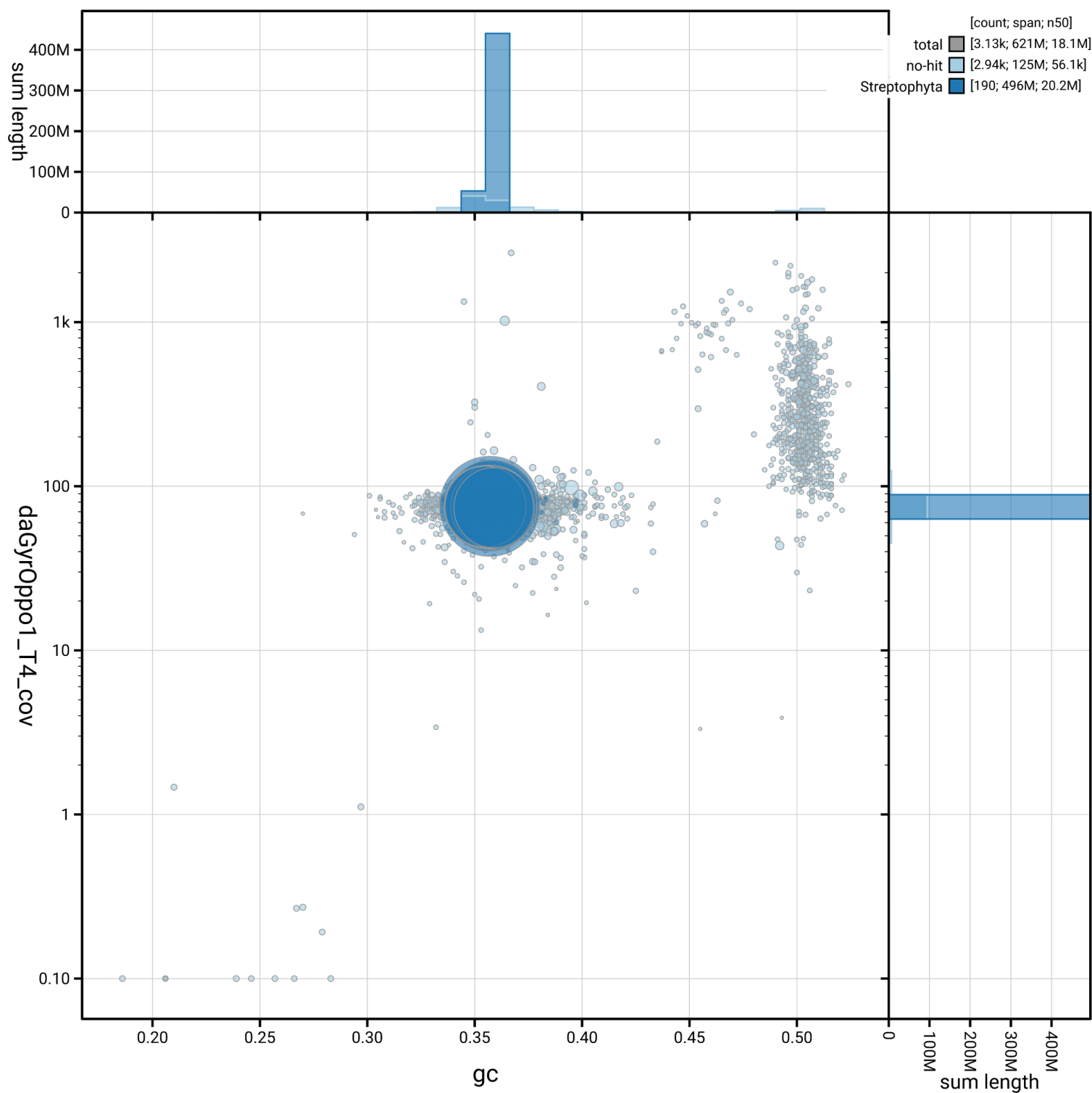


Distribution of k-mer counts per copy numbers found in **hap1** (hap1.)



Distribution of k-mer counts per copy numbers found in asm (dipl.)

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 | 169x | 452x |

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*: 1.2.2
 - |_ *key param*: NA

Submitter: Dominic Absolon

Affiliation: WSI

Date and time: 2025-12-10 13:49:15 CET