

ERGA Assembly Report

v24.04.03_beta

Tags: ERGA-BGE

| | |
|---------|---------------------------|
| TxID | 65001 |
| ToLID | daCheTaga1 |
| Species | Cheirolophus tagananensis |
| Class | Magnoliopsida |
| Order | Asterales |

| Genome Traits | Expected | Observed |
|-------------------|----------------------|-------------|
| Haploid size (bp) | 645,759,218 | 623,521,733 |
| Haploid Number | 15 (source: direct) | 16 |
| Ploidy | 2 (source: ancestor) | 2 |
| Sample Sex | H | unknown |

EBP metrics summary and curation notes

Obtained EBP quality metric for hap1: 6.7.Q60

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

- . Observed Haploid Number is different from Expected
- . Observed sex is different from Sample sex
- . Assembly length loss > 3% for hap1

Curator notes

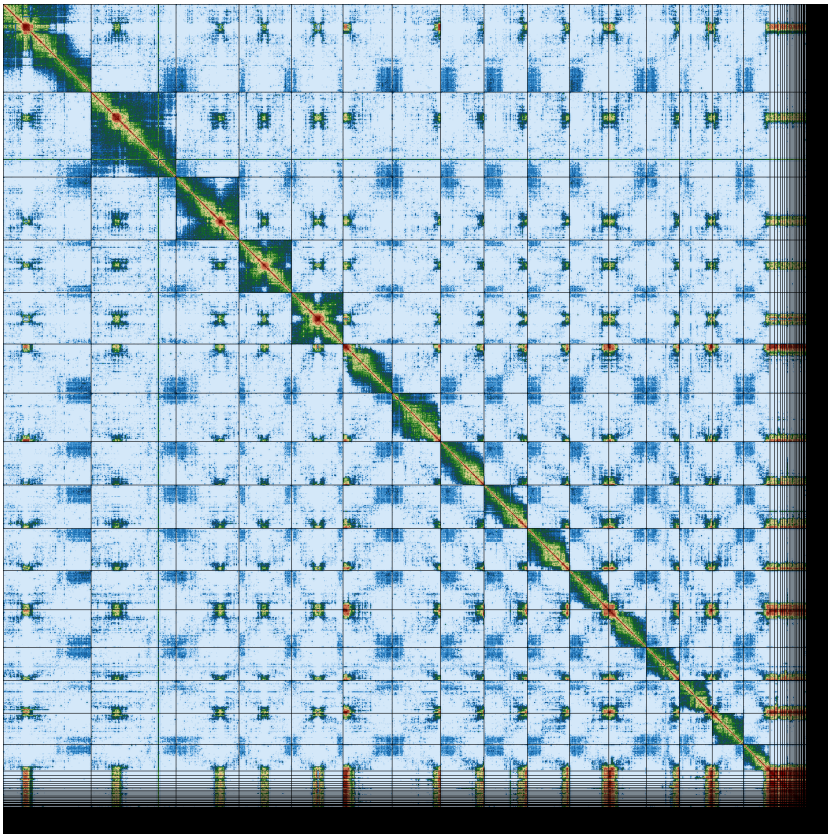
- . Interventions/Gb: 36
- . Contamination notes: "Total length of scaffolds removed: 58,772,027 (8.7 %);Scaffolds removed: 1165 (82.5 %);Largest scaffold removed: (180,771);FCS-GX contaminant species (number of scaffolds; total length of scaffolds):Mitochondrion (44; 1,788,085);Plastid (1121; 56,983,929);Barcodes (1; 13)"
- . Other observations: "The high GC scaffolds on the blobplot are the unplaces repetitive sequence in the bottom right of the Hi-C map."

Quality metrics table

| Metrics | Pre-curation hap1 | Curated hap1 |
|--------------|----------------------|-----------------|
| Total bp | 678,516,303 | 623,521,733 |
| GC % | 36.66 | 36.6 |
| Gaps/Gbp | 284.44 | 298.31 |
| Total gap bp | 38,600 | 37,200 |
| Scaffolds | 1,412 | 235 |
| Scaffold N50 | 35,173,251 | 36,486,517 |
| Scaffold L50 | 7 | 7 |
| Scaffold L90 | 100 | 16 |
| Contigs | 1,605 | 421 |
| Contig N50 | 3,718,255 | 3,968,000 |
| Contig L50 | 56 | 50 |
| Contig L90 | 289 | 151 |
| QV | 56.4 | 60.5 |
| Kmer compl. | 99.18 | 99.07 |
| BUSCO sing. | 91.9% | 92.8% |
| BUSCO dupl. | 5.0% | 5.0% |
| BUSCO frag. | 0.5% | 0.5% |
| BUSCO miss. | 2.6% | 1.7% |

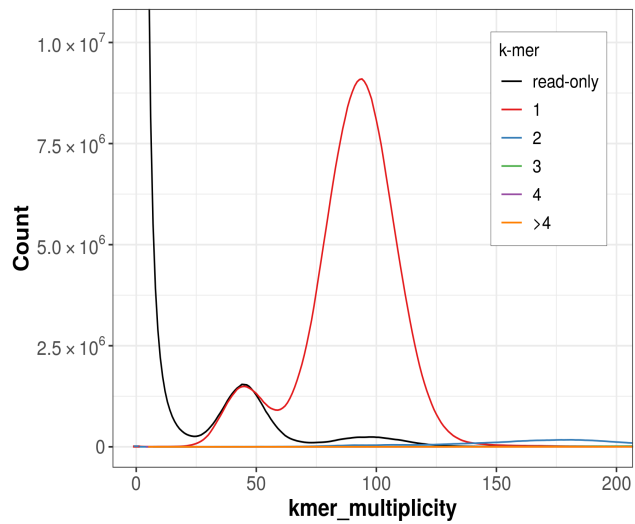
BUSCO 5.3.0 Lineage: eudicots_odb10 (genomes:31, BUSCOs:2326)

HiC contact map of curated assembly

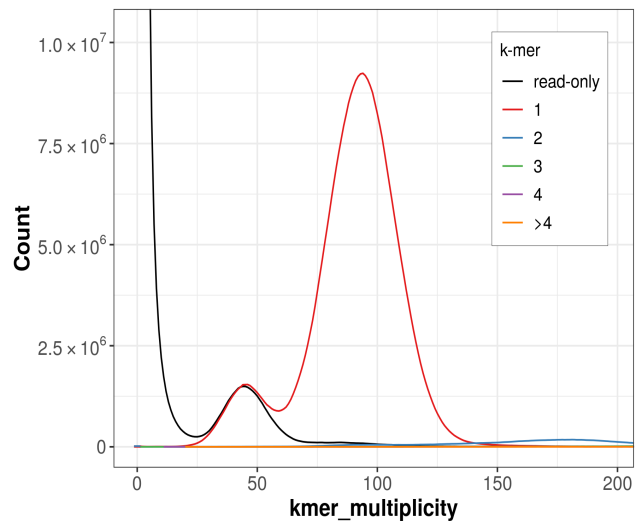


hap1 [\[LINK\]](#)

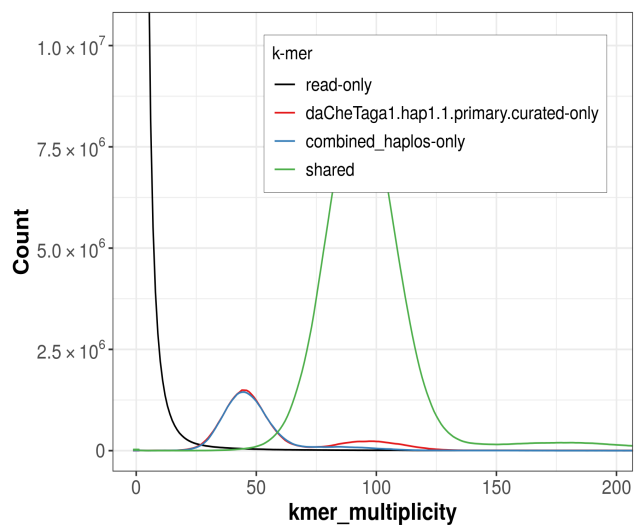
K-mer spectra of curated assembly



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

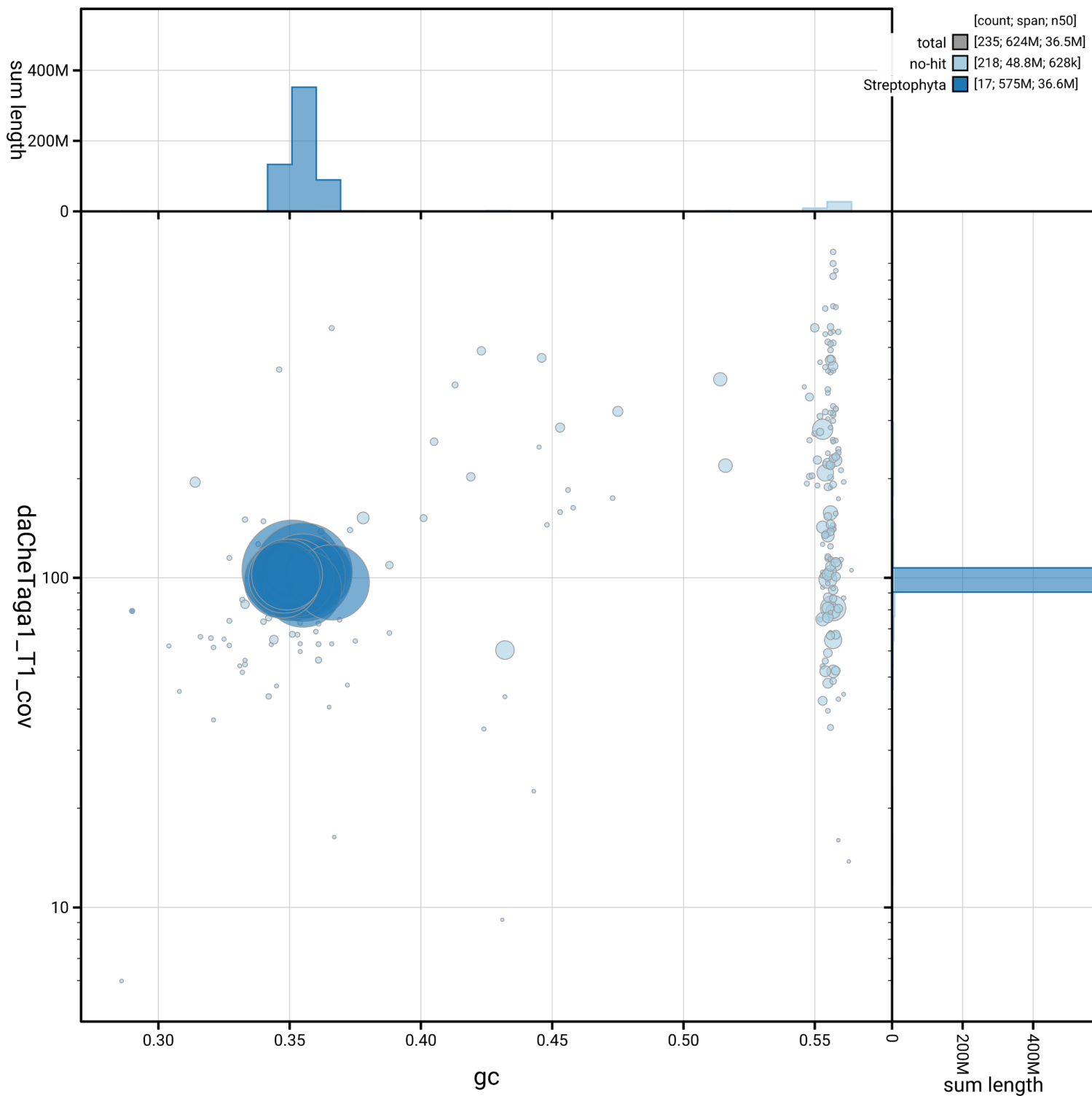


Distribution of k-mer counts per copy numbers found in asm



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

Post-curation contamination screening



hap1. 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 v3 |
| Coverage | 94x | 154x |

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: Camilla Santos

Affiliation: WSI

Date and time: 2024-10-22 13:17:33 CEST