

# ERGA Assembly Report

v24.04.03\_beta

Tags: ERGA-BGE

TxID	88878
ToLID	<b>drSaxVayr1</b>
Species	Saxifraga vayredana
Class	Magnoliopsida
Order	Saxifragales

Genome Traits	Expected	Observed
Haploid size (bp)	632,079,025	542,591,247
Haploid Number	31 (source: direct)	29
Ploidy	2 (source: ancestor)	2
Sample Sex	H	H

## EBP metrics summary and curation notes

Obtained EBP quality metric for pri: 6.7.Q61

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

- . Observed Haploid Number 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

## Curator notes

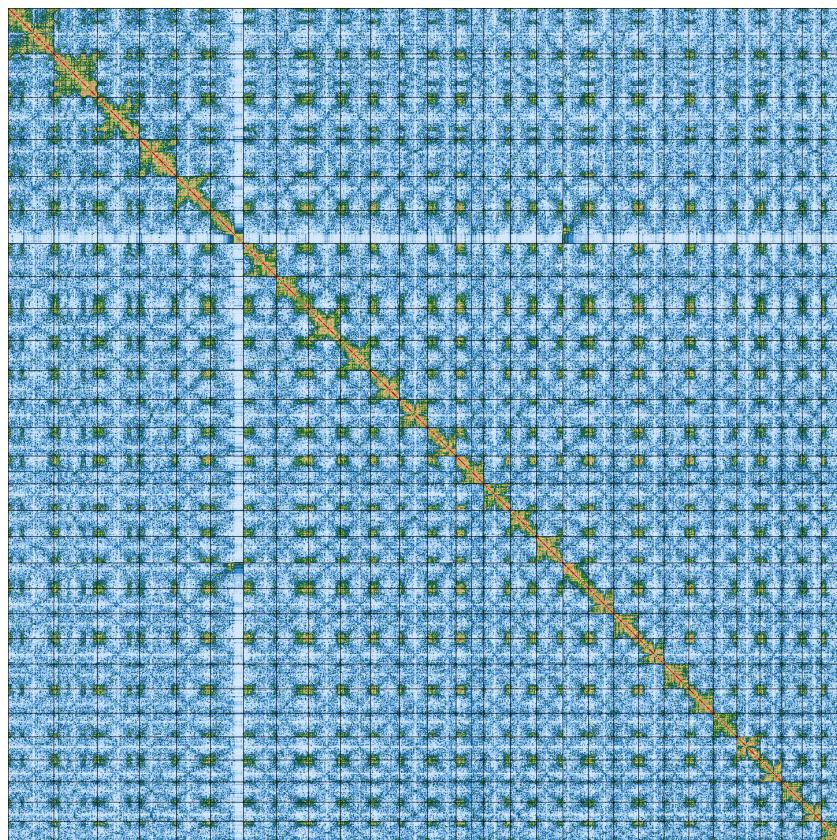
- . Interventions/Gb: 160
- . Contamination notes: "Contamination report for assembly labelled hap1; Total length of scaffolds removed: 106,957,162 (15.3 %); Scaffolds removed: 2477 (75.7 %); Largest scaffold removed: (192,208); FCS-GX contaminant species (number of scaffolds; total length of scaffolds): Aureobasidium vineae, ascomycetes (1; 37,569); Aureobasidium sp. EXF-3400, ascomycetes (1; 7,384); Mitochondrion (1099; 43,974,556); Plastid (1376; 62,937,653)"
- . Other observations: "Assembly was Hi-C phased; The sample appears to be allopolyploid (4n = 58) with the HIC signal indicating highly divergent subgenomes. The primary (HAP 1) assembly contains a single copy of each subgenome. The order and orientation of contigs along chromosome 1 is uncertain between 15 Mb and 21.5 Mb. Heterozygous inversions were observed in the following regions: chromosome 7, 5.3 Mb to 7 Mb.; chromosome 8, 8.2 Mb to 10.9 Mb and 15.6 Mb to 17. 6 Mb."

# Quality metrics table

Metrics	Pre-curation pri	Curated pri
Total bp	783,419,052	542,591,247
GC %	37.86	37.2
Gaps/Gbp	1,018.61	235.91
Total gap bp	79,800	19,900
Scaffolds	4,334	29
Scaffold N50	16,228,285	18,520,738
Scaffold L50	20	12
Scaffold L90	1,968	25
Contigs	5,132	157
Contig N50	5,450,061	8,734,432
Contig L50	38	21
Contig L90	2,687	62
QV	52.6	61.4
Kmer compl.	99.21	98.72
BUSCO sing.	42.1%	52.7%
BUSCO dupl.	54.5%	43.8%
BUSCO frag.	0.2%	0.3%
BUSCO miss.	3.1%	3.2%

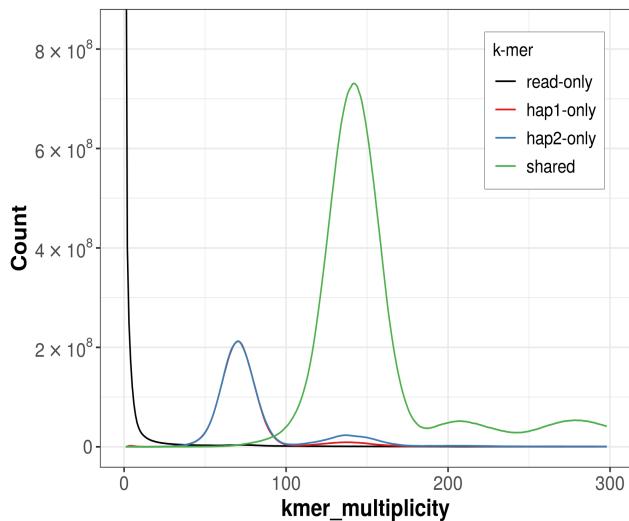
BUSCO 6.0.0 Lineage: eudicots\_odb10 (genomes:31, BUSCOs:2326)

# HiC contact map of curated assembly

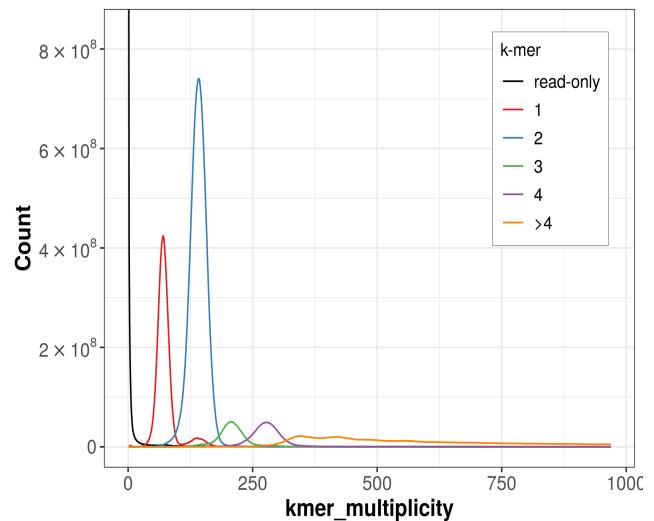


pri [LINK]

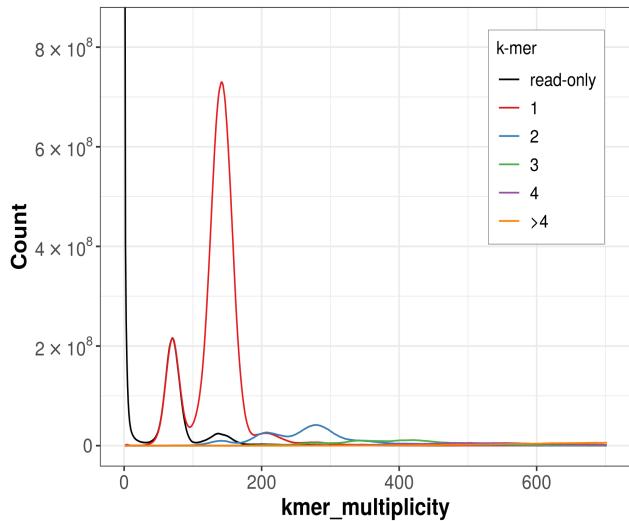
# K-mer spectra of curated assembly



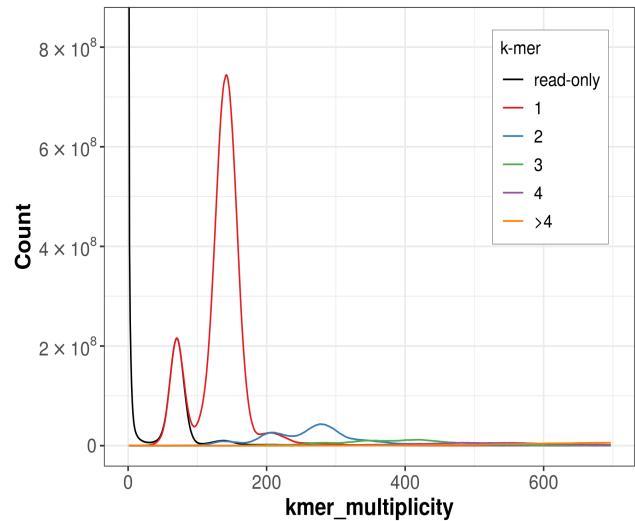
Distribution of k-mer counts coloured by their presence in reads/assemblies



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

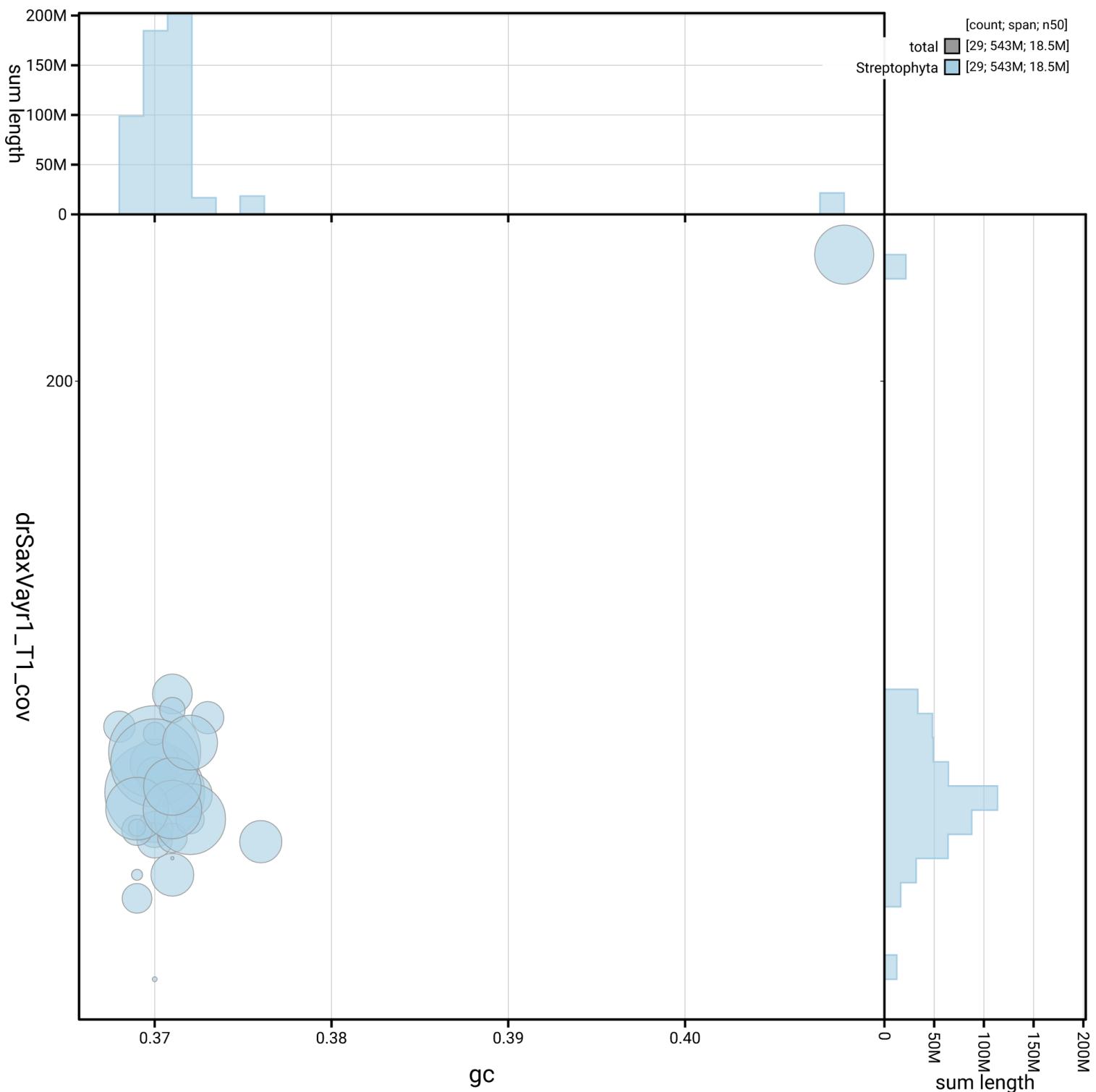


Distribution of k-mer counts per copy numbers found in **hap1** (hapl.)



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

# 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	140x	700x

## 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: Thomas Mathers

Affiliation: WSI

Date and time: 2025-11-21 16:35:23 CET