ERGA Assembly Report

v24.10.15

Tags: ERGA-BGE

TxID	502525
ToLID	**ddVioUlig1**
Species	Viola uliginosa
Class	Magnoliopsida
Order	Malpighiales

Genome Traits	Expected	Observed
Haploid size (bp)	314,524,744	597,386,493
Haploid Number	10 (source: direct)	10
Ploidy	2 (source: direct)	2
Sample Sex	Unknown	Unknown

EBP metrics summary and curation notes

Obtained EBP quality metric for collapsed: 6.7.Q64

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

. Observed Haploid size (bp) has >20% difference with Expected

. BUSCO single copy value is less than 90% for collapsed
. BUSCO duplicated value is more than 5% for collapsed

Curator notes

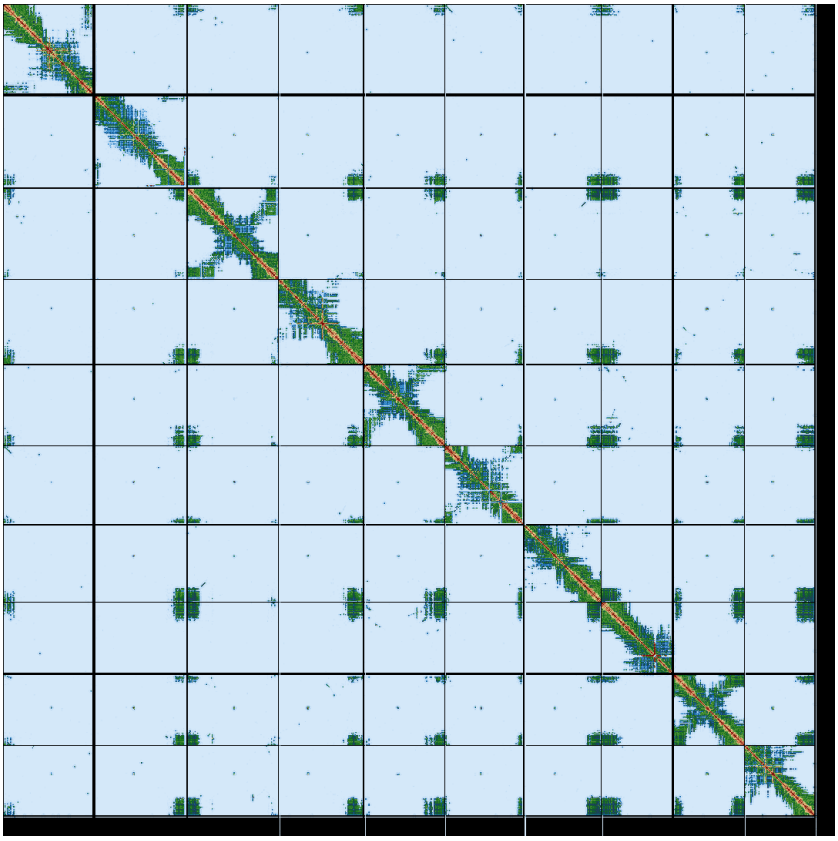
. Interventions/Gb: 821
. Contamination notes: ""
. Other observations: "The assembly of VIOLA ULIGINOSA (ddVioUlig1) is based on 57X PacBio data and 347X Arima Hi-C data generated as part of the European Reference Genome Atlas (ERGA, https://www.erga-biodiversity.eu/) via the Biodiversity Genomics Europe project (BGE, https://biodiversitygenomics.eu/). The assembly process included the following steps: initial PacBio and Hi-C assembly generation with Hifiasm creating a dual assembly, removal of contaminant sequences using Context, and Hi-C-based scaffolding with YaHS. Both haplotypes were curated, but only haplotype 1 (HAP1) was retained for this EAR report. All subsequent information is based on HAP1. In total, 22 contigs were identified as contaminants (bacterial), totaling 550,002 pb (with the largest being 227,651 pb). The mitochondrial genome was assembled using OATK. Finally, the primary assembly was analyzed and manually improved using Pretext. During manual curation, 3 haplotypic region was removed, totaling 794,349 pb (with the largest being 589,537 pb). Chromosome-scale scaffolds confirmed by Hi-C data were named in order of size. "

Quality metrics table

Metrics	Pre-curation collapsed	Curated collapsed
Total bp	607,309,980	597,386,493
GC %	39.56	39.65
Gaps/Gbp	0	455.32
Total gap bp	0	37,200
Scaffolds	700	326
Scaffold N50	10,831,798	58,224,354
Scaffold L50	16	5
Scaffold L90	55	10
Contigs	700	598
Contig N50	10,831,798	4,898,686
Contig L50	16	39
Contig L90	55	131
QV	61.8594	64.1193
Kmer compl.	95.0546	94.6478
BUSCO sing.	15.1%	12.4%
BUSCO dupl.	72.7%	79.8%
BUSCO frag.	1.4%	0.3%
BUSCO miss.	10.9%	7.5%

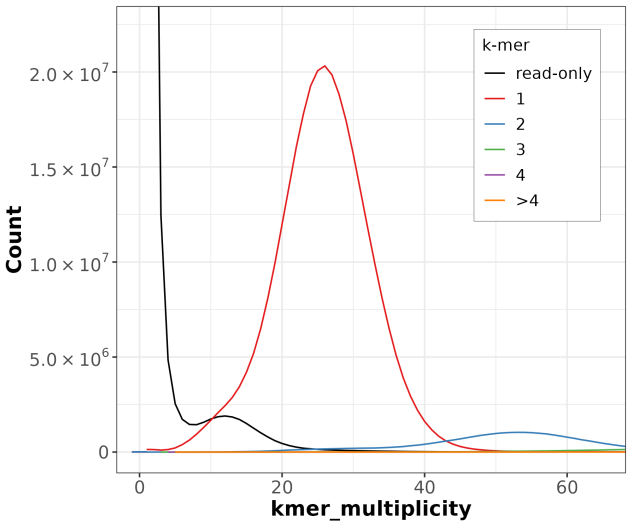
Warning! BUSCO versions or lineage datasets are not the same across results:
BUSCO: 5.8.2 (euk_genome_met, metaeuk) / Lineage: malpighiales_odb12 (genomes:6, BUSCOs:6134)
BUSCO: 6.0.0 (euk_genome_min, miniprot) / Lineage: malpighiales_odb12 (genomes:6, BUSCOs:6134)

HiC contact map of curated assembly

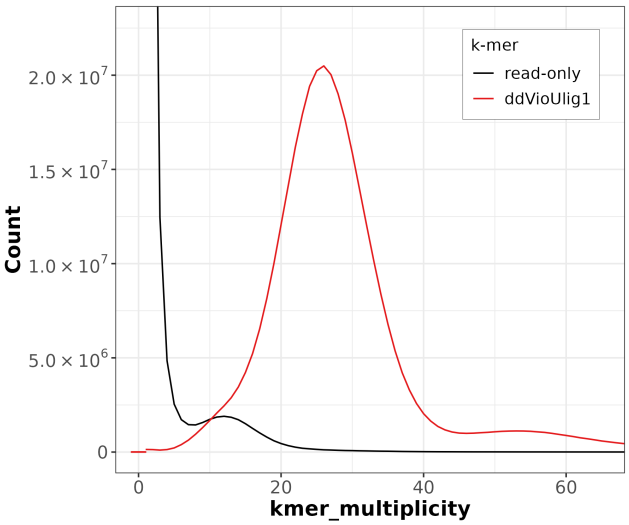


collapsed [LINK]

K-mer spectra of curated assembly

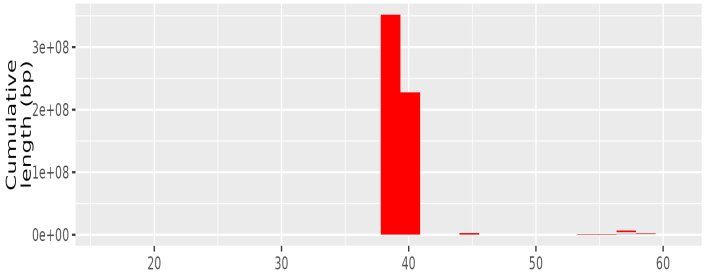


Distribution of k-mer counts per copy numbers found in asm

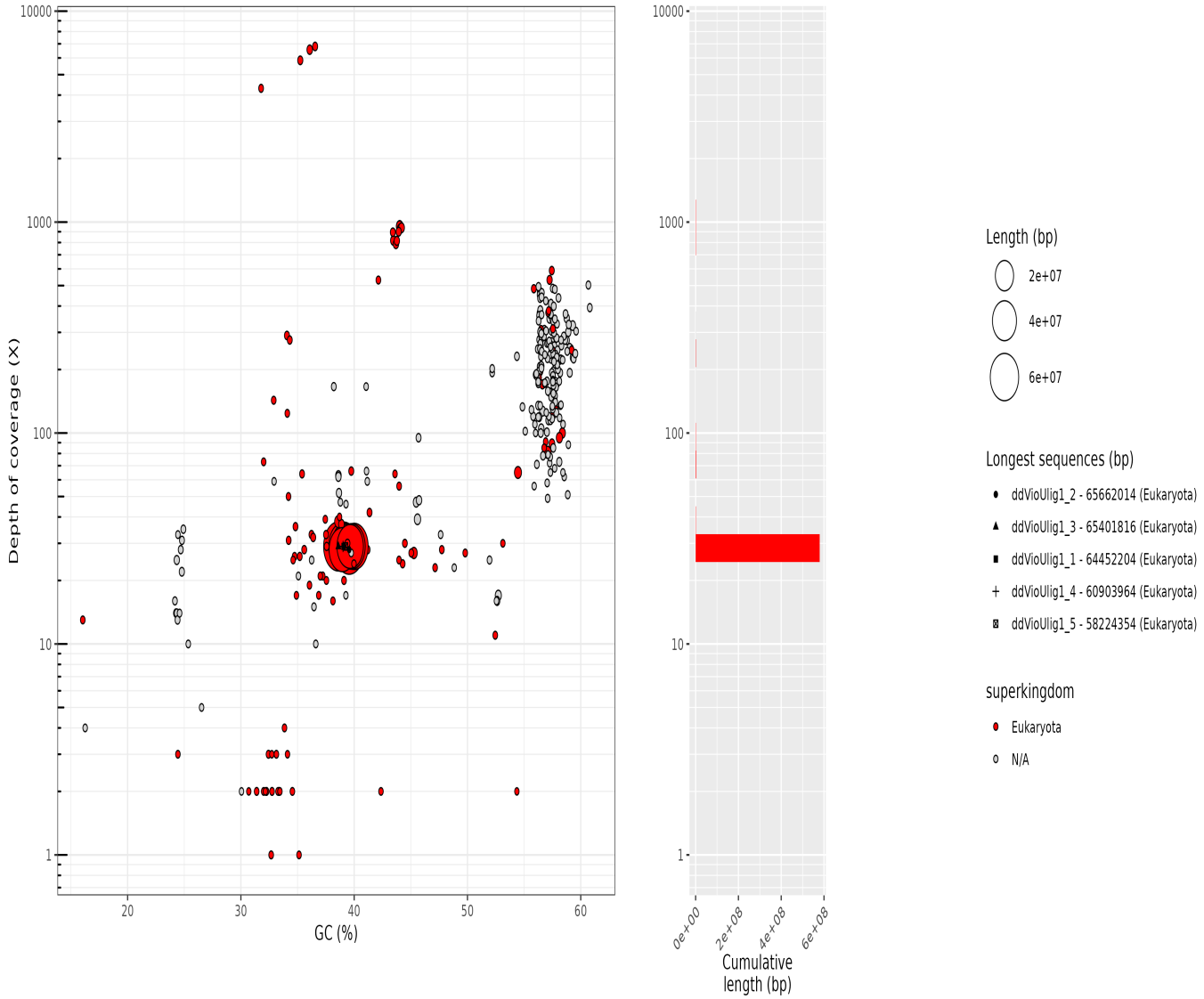


Distribution of k-mer counts coloured by their presence in reads/assemblies

Post-curation contamination screening



TAPAs summary Graph



collapsed. 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	Long reads	Arima
Coverage	57	346

Assembly pipeline

- **Hifiasm**
  |_ *ver:* 0.19.5-r593
  |_ *key param:* NA
- **purge_dups**
  |_ *ver:* 1.2.5
  |_ *key param:* NA
- **YaHS**
  |_ *ver:* 1.2
  |_ *key param:* NA

Curation pipeline

- **PretextMap**
  |_ *ver:* 0.1.9
  |_ *key param:* NA
- **PretextView**
  |_ *ver:* 0.2.5
  |_ *key param:* NA

Submitter: Lola Demirdjian
Affiliation: Genoscope

Date and time: 2025-12-16 20:27:16 CET