

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	577,343,289
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.Q65

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
- . Assembly length loss > 3% for collapsed

Curator notes

- . Interventions/Gb: 667
- . 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 assembly generation with Hifiasm, removal of contaminant sequences using Context, and Hi-C-based scaffolding with YaHS. In total, 20 contigs were identified as contaminants (bacterial), totaling 274,144 pb (with the largest being 31,948 pb). We did not use purge_dups to remove haplotype duplications, as it did not work properly on this genome. The mitochondrial genome was assembled using OATK. Finally, the primary assembly was analyzed and manually improved using Pretext. During manual curation, 45 haplotypic region was removed, totaling 21,851,319 pb (with the largest being 6,007,832 pb) and 38 contaminants was removed. Chromosome-scale scaffolds confirmed by Hi-C data were named in order of size. "

Quality metrics table

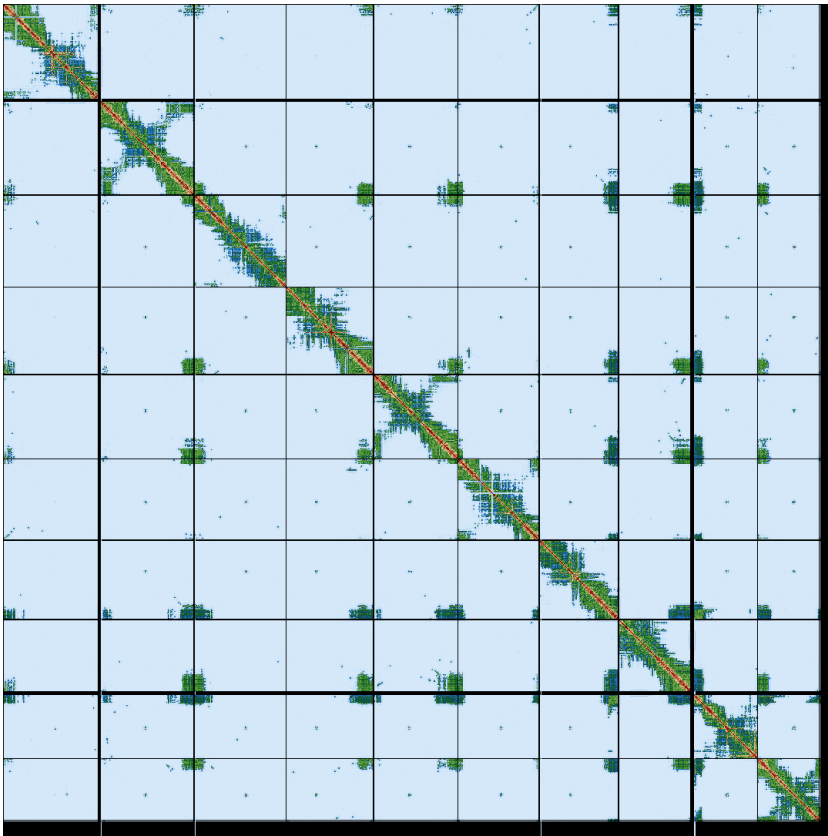
Metrics	Pre-curation collapsed	Curated collapsed
Total bp	607,317,780	577,343,289
GC %	39.56	39.63
Gaps/Gbp	128.43	164.55
Total gap bp	7,800	12,600
Scaffolds	657	252
Scaffold N50	58,417,419	58,417,819
Scaffold L50	5	5
Scaffold L90	10	10
Contigs	735	347
Contig N50	9,366,092	9,057,521
Contig L50	21	22
Contig L90	68	68
QV	61.8594	65.2345
Kmer compl.	95.0546	92.0881
BUSCO sing.	15.4%	20.7%
BUSCO dupl.	72.4%	71.0%
BUSCO frag.	1.4%	0.3%
BUSCO miss.	10.8%	7.9%

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)

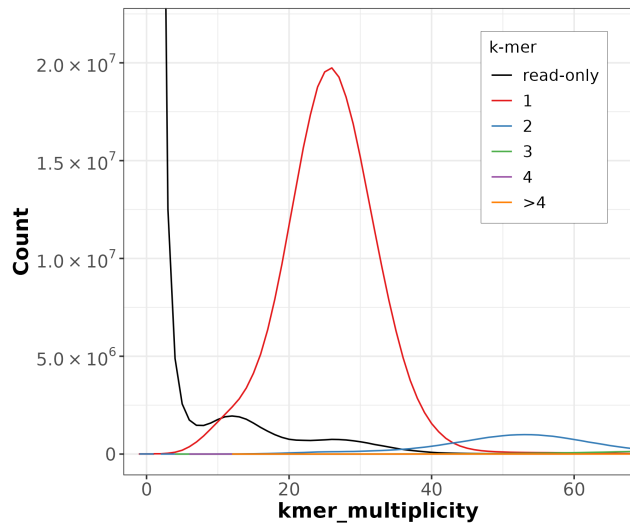
BUSCO: 6.0.0 (euk_genome_min, miniprot) / Lineage: malpighiales_odb12 (genomes:6, BUSCOs:6134)

HiC contact map of curated assembly

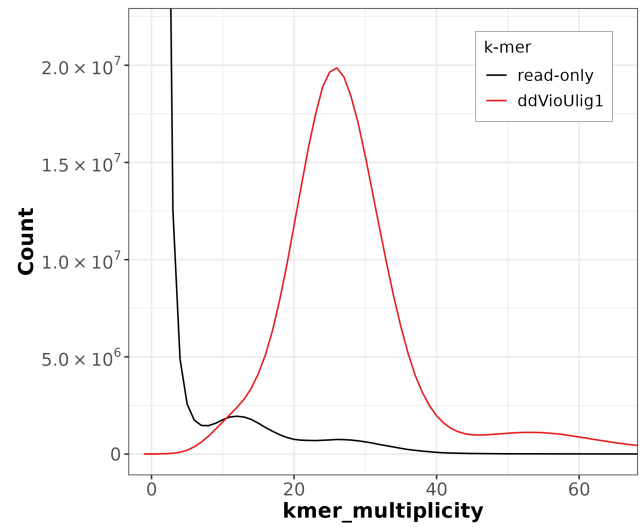


collapsed [\[LINK\]](#)

K-mer spectra of curated assembly

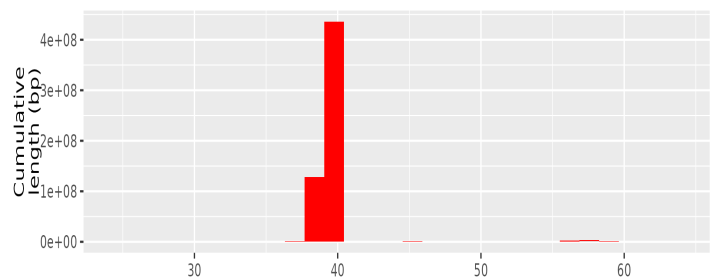


Distribution of k-mer counts per copy numbers found in asm

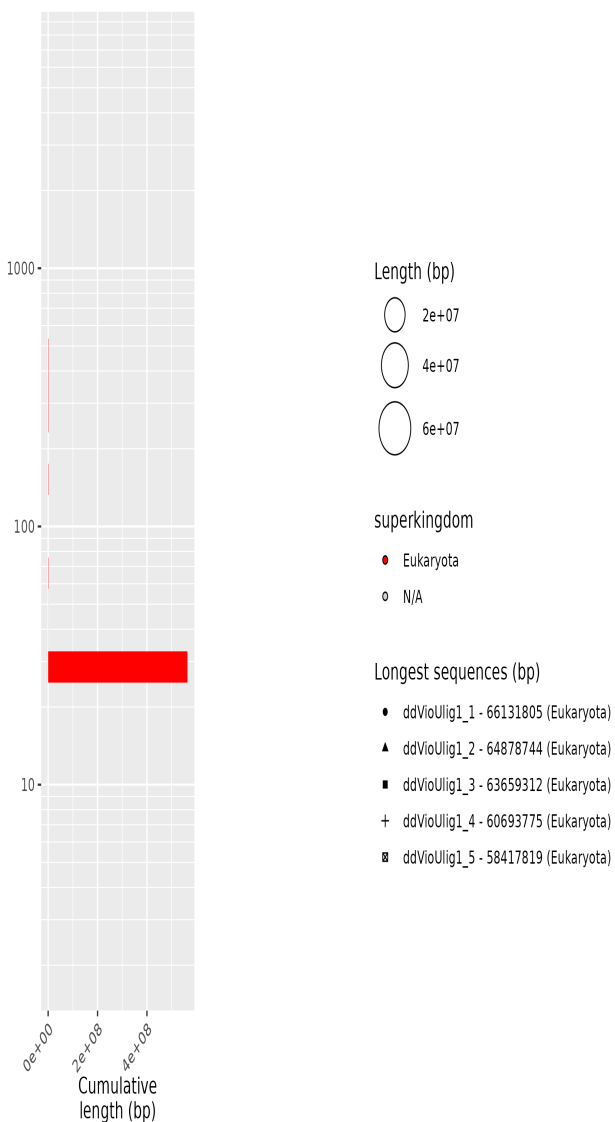
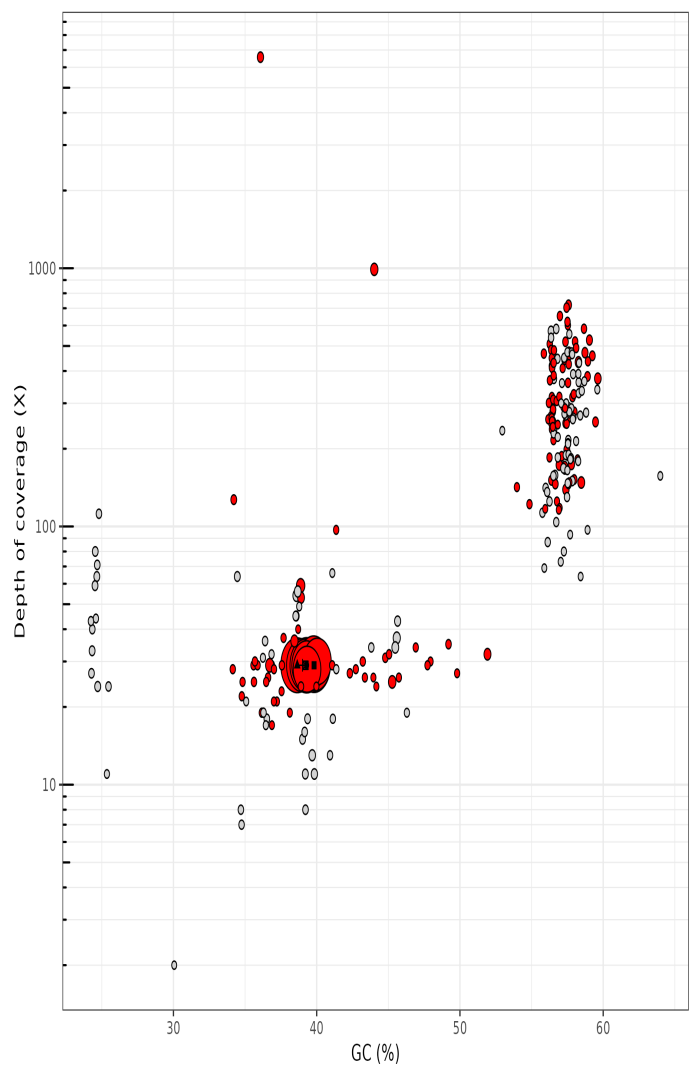


Distribution of k-mer counts coloured by their presence in reads/assemblies

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	347

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: 2025-09-24 14:53:34 CEST