

ERGA Assembly Report

v24.10.15

Tags: ERGA-BGE

TxID	673926
ToLID	ddHypElod1
Species	Hypericum elodes
Class	Magnoliopsida
Order	Malpighiales

Genome Traits	Expected	Observed
Haploid size (bp)	503,129,940	466,502,948
Haploid Number	9 (source: ancestor)	0
Ploidy	2 (source: ancestor)	2
Sample Sex	Unknown	Unknown

EBP metrics summary and curation notes

Obtained EBP quality metric for collapsed: 6.6.Q58

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 collapsed
- . BUSCO duplicated value is more than 5% for collapsed
- . Not 90% of assembly in chromosomes for collapsed

Curator notes

- . Interventions/Gb: 0
- . Contamination notes: ""
- . Other observations: "The assembly of HYPERICUM ELODES (ddHypElod1) is based on 24X PacBio 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 Hifiiasm, removal of contaminant sequences using Context and removal of haplotypic duplications using purge_dup. In total, 79 contigs were identified as contaminants (bacterial), totaling 2,041,414 pb (with the largest being 83,852 pb). Additionally, 378 regions totaling 67,690,356 pb were identified as haplotypic duplications and removed. No Hi-C available "

Quality metrics table

Metrics	Pre-curation collapsed	Curated collapsed
Total bp	466,502,948	466,502,948
GC %	37.61	37.61
Gaps/Gbp	0	0
Total gap bp	0	0
Scaffolds	494	494
Scaffold N50	3,005,422	3,005,422
Scaffold L50	52	52
Scaffold L90	167	167
Contigs	494	494
Contig N50	3,005,422	3,005,422
Contig L50	52	52
Contig L90	167	167
QV	58.8838	58.8838
Kmer compl.	94.4149	94.4149
BUSCO sing.	86.8%	86.8%
BUSCO dupl.	12.4%	12.4%
BUSCO frag.	0.0%	0.0%
BUSCO miss.	0.8%	0.8%

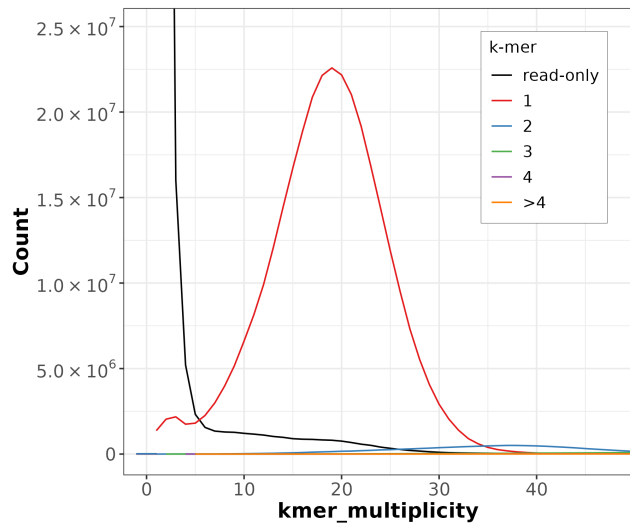
BUSCO: 5.8.2 (euk_genome_met, metaeuk) / Lineage: eukaryota_odb12 (genomes:456, BUSCOs:129)

HiC contact map of curated assembly

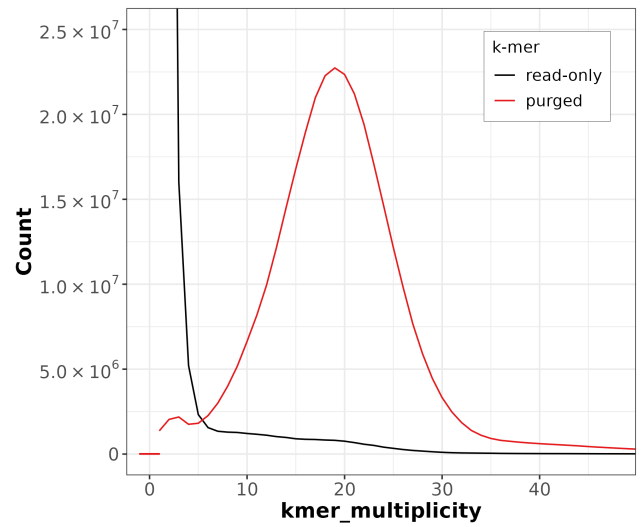


collapsed [\[LINK\]](#)

K-mer spectra of curated assembly

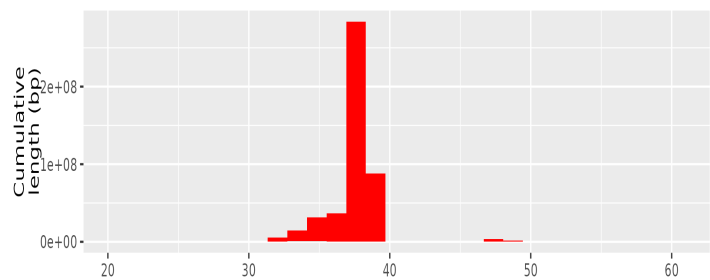


Distribution of k-mer counts per copy numbers found in asm

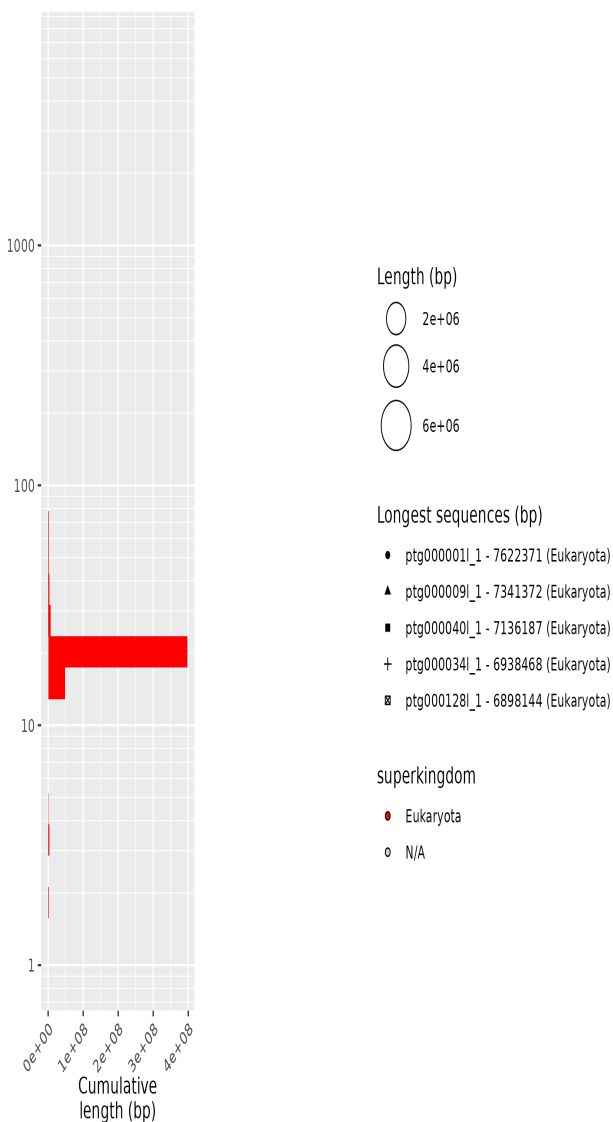
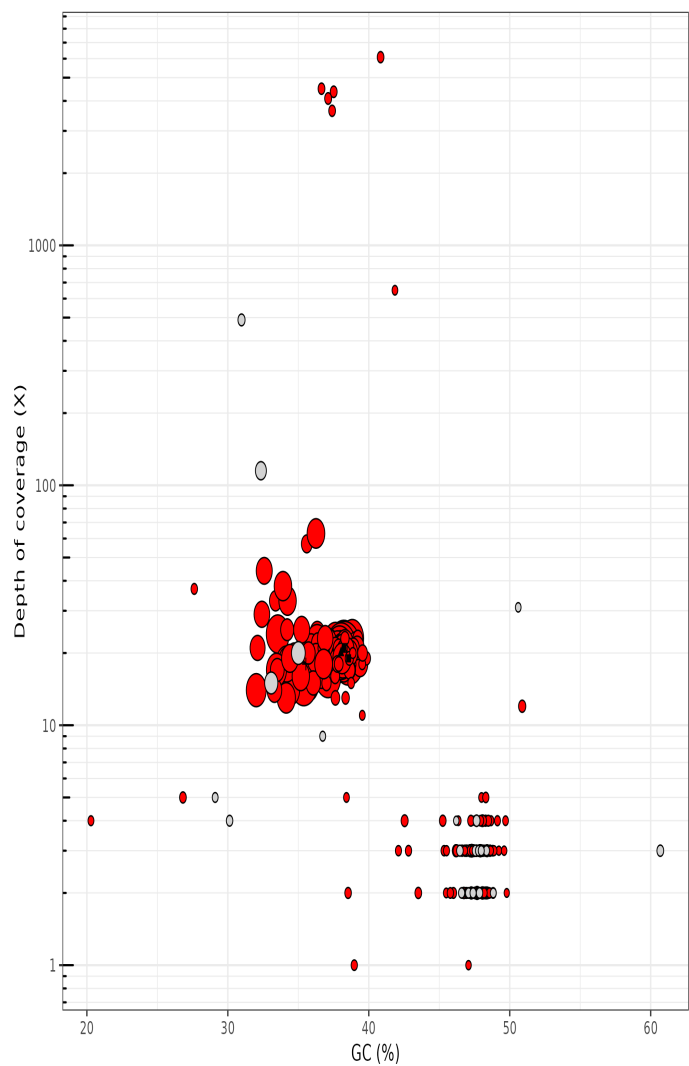


Distribution of k-mer counts coloured by their presence in reads/assemblies

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	PACBIO Hifi	Arima
Coverage	24	NA

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: 2025-09-02 17:49:45 CEST