

ERGA Assembly Report

v24.04.03_beta

Tags: ERGA-BGE

TxID	1534734
ToLID	daXanOriel
Species	Xanthium orientale
Class	Magnoliopsida
Order	Asterales

Genome Traits	Expected	Observed
Haploid size (bp)	2,269,223,707	2,258,523,083
Haploid Number	18 (source: direct)	18
Ploidy	4 (source: ancestor)	2
Sample Sex	Unknown	Unknown

EBP metrics summary and curation notes

Obtained EBP quality metric for collapsed: 7.8.Q67

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

- . Observed Ploidy is different from Expected
- . BUSCO duplicated value is more than 5% for collapsed

Curator notes

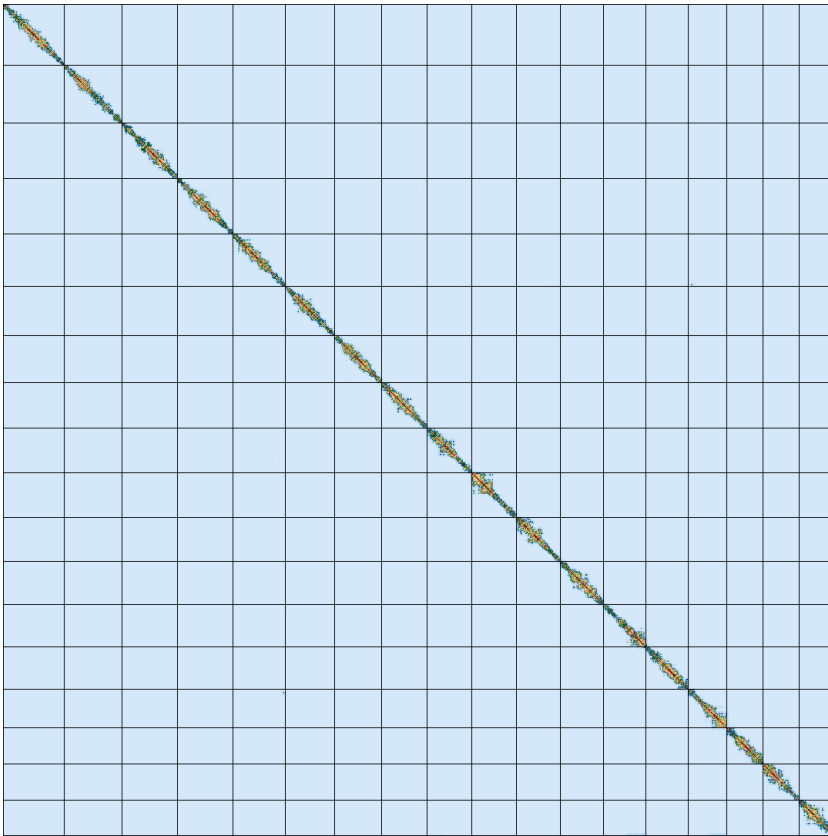
- . Interventions/Gb: 1
- . Contamination notes: "One bacterial contig has been removed "
- . Other observations: ""

Quality metrics table

Metrics	Pre-curation collapsed	Curated collapsed
Total bp	2,258,077,340	2,258,523,083
GC %	37.03	37.03
Gaps/Gbp	7.53	6.2
Total gap bp	1,700	1,400
Scaffolds	27	32
Scaffold N50	134,603,727	124,155,882
Scaffold L50	6	8
Scaffold L90	13	16
Contigs	44	46
Contig N50	98,804,340	98,804,340
Contig L50	9	9
Contig L90	21	21
QV	67.9927	67.9935
Kmer compl.	99.6146	99.6151
BUSCO sing.	93.3%	93.3%
BUSCO dupl.	5.8%	5.8%
BUSCO frag.	0.2%	0.2%
BUSCO miss.	0.7%	0.7%

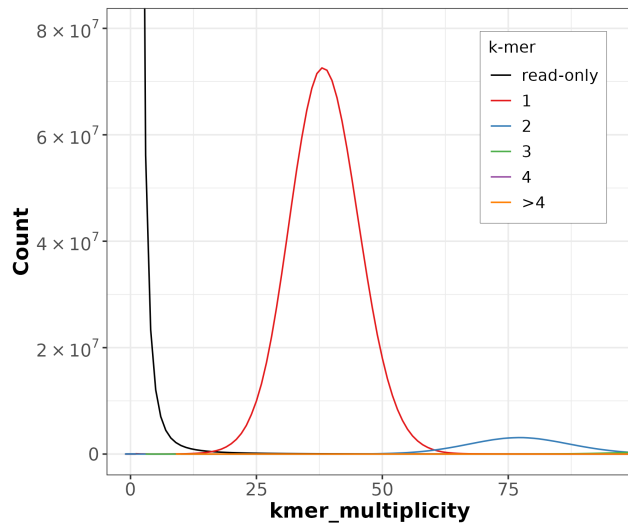
BUSCO 5.4.3 Lineage: embryophyta_odb10 (genomes:50, BUSCOs:1614)

HiC contact map of curated assembly

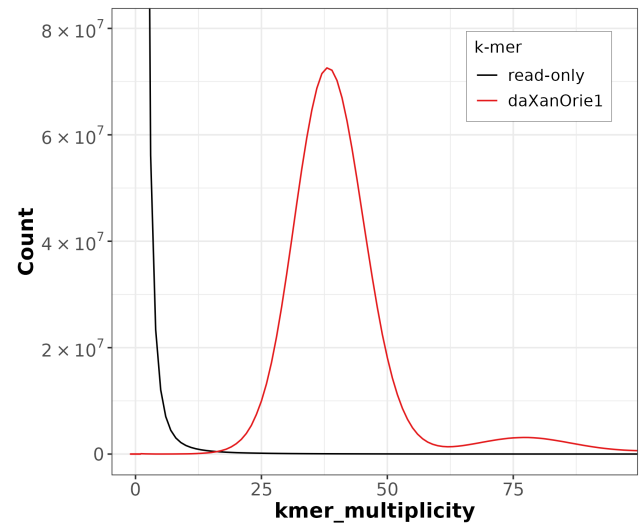


collapsed [\[LINK\]](#)

K-mer spectra of curated assembly

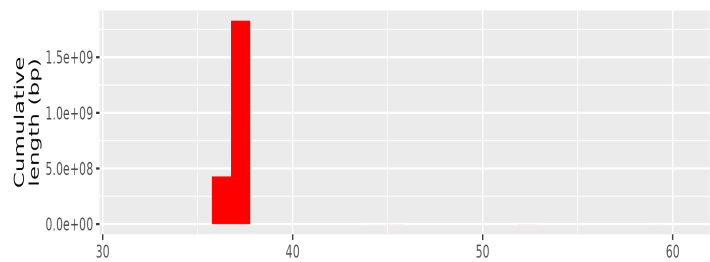


Distribution of k-mer counts per copy numbers found in asm

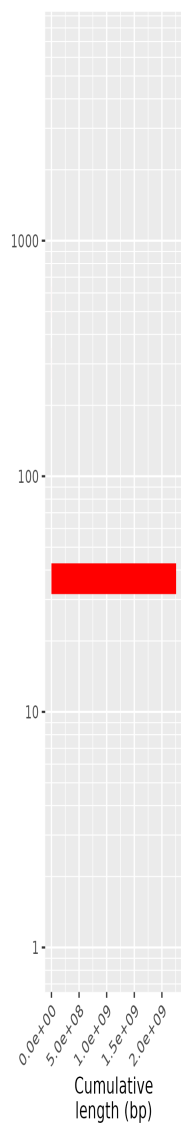
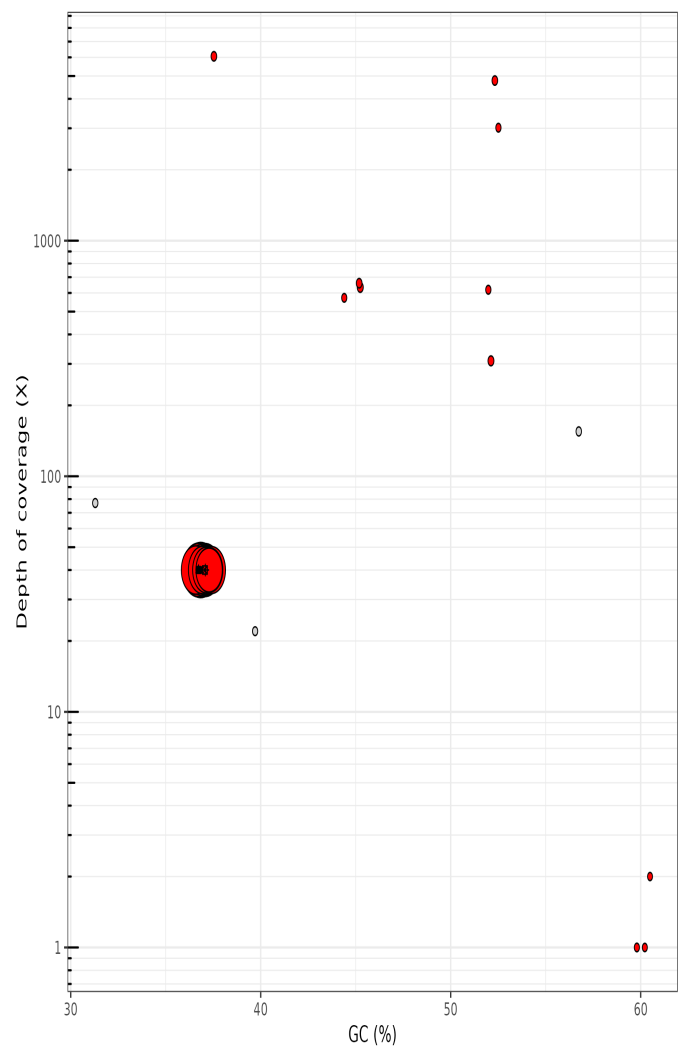


Distribution of k-mer counts coloured by their presence in reads/assemblies

Post-curation contamination screening



TAPAs summary Graph



Longest sequences (bp)

- SUPER_1 - 165985378 (Eukaryota)
- ▲ SUPER_2 - 156505571 (Eukaryota)
- SUPER_3 - 151051287 (Eukaryota)
- + SUPER_4 - 150506000 (Eukaryota)
- ▣ SUPER_5 - 141756184 (Eukaryota)

superkingdom

- Eukaryota
- N/A

Length (bp)

- 4.0e+07
- 8.0e+07
- 1.2e+08
- 1.6e+08

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	Omnic
Coverage	40	10

Assembly pipeline

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

Curation 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

Submitter: Adama Ndar

Affiliation: Genoscope

Date and time: 2024-09-05 16:34:55 CEST