

ERGA Assembly Report

v24.09.10

Tags: ERGA-BGE

TxID	754874
ToLID	drEbeCret
Species	Ebenus cretica
Class	Magnoliopsida
Order	Fabales

Genome Traits	Expected	Observed
Haploid size (bp)	989,470,455	944,616,402
Haploid Number	7 (source: direct)	6
Ploidy	2 (source: ancestor)	2
Sample Sex	Unknown	Unknown

EBP metrics summary and curation notes

Obtained EBP quality metric for collapsed: 7.8.Q59

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

- . Observed Haploid Number is different from Expected
- . Kmer completeness value is less than 90 for collapsed
- . BUSCO single copy value is less than 90% for collapsed
- . BUSCO duplicated value is more than 5% for collapsed

Curator notes

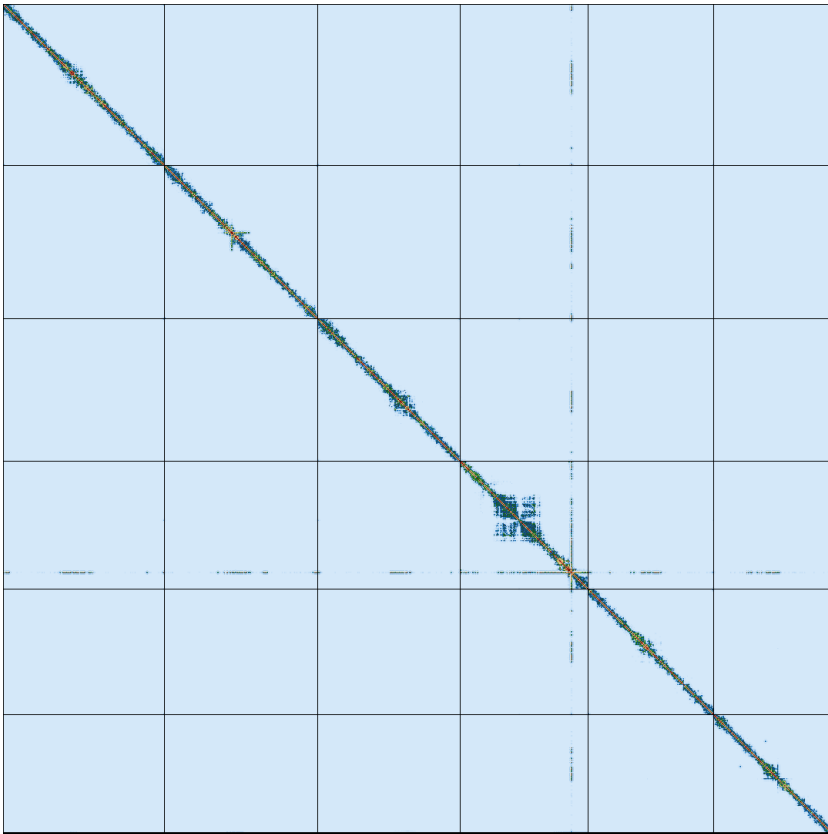
- . Interventions/Gb: 3
- . Contamination notes: "No contaminant detected "
- . Other observations: ""

Quality metrics table

Metrics	Pre-curation collapsed	Curated collapsed
Total bp	946,087,149	944,616,402
GC %	34.54	34.54
Gaps/Gbp	27.48	18
Total gap bp	2,600	1,800
Scaffolds	69	56
Scaffold N50	174,068,480	161,678,577
Scaffold L50	2	3
Scaffold L90	5	6
Contigs	95	73
Contig N50	92,227,207	92,227,207
Contig L50	4	4
Contig L90	10	10
QV	58.9172	59.0295
Kmer compl.	89.3262	89.3094
BUSCO sing.	64.3%	64.4%
BUSCO dupl.	34.6%	34.5%
BUSCO frag.	0.1%	0.1%
BUSCO miss.	1.0%	1.0%

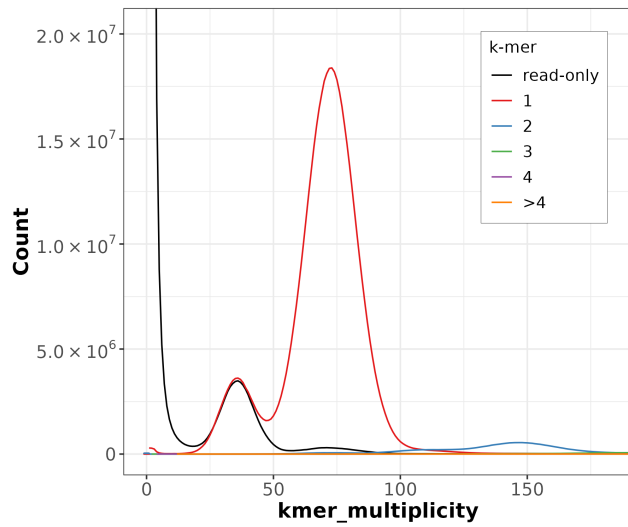
BUSCO 5.4.3 Lineage: embryophyta_odb10 (genomes:50, BUSCOs:1614)

HiC contact map of curated assembly

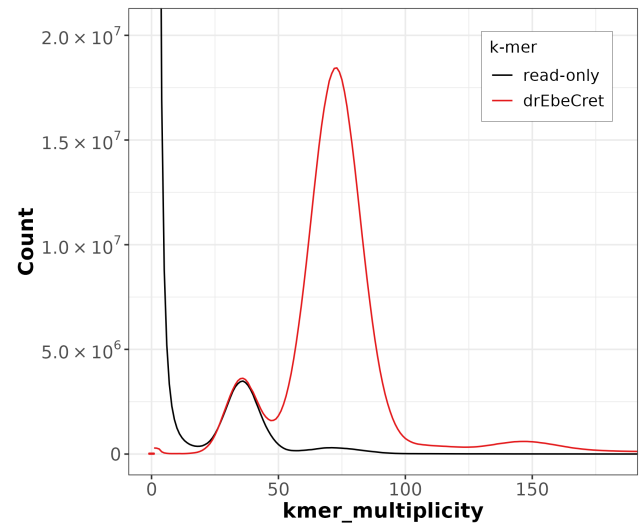


collapsed [\[LINK\]](#)

K-mer spectra of curated assembly

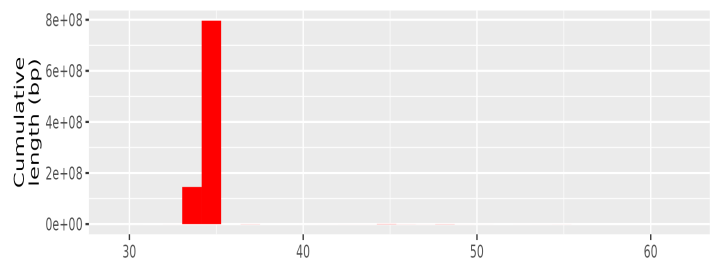


Distribution of k-mer counts per copy numbers found in asm

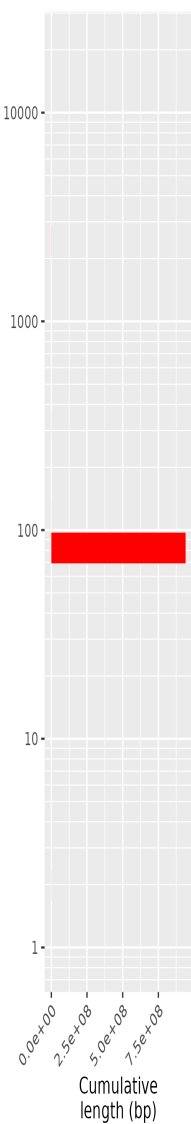
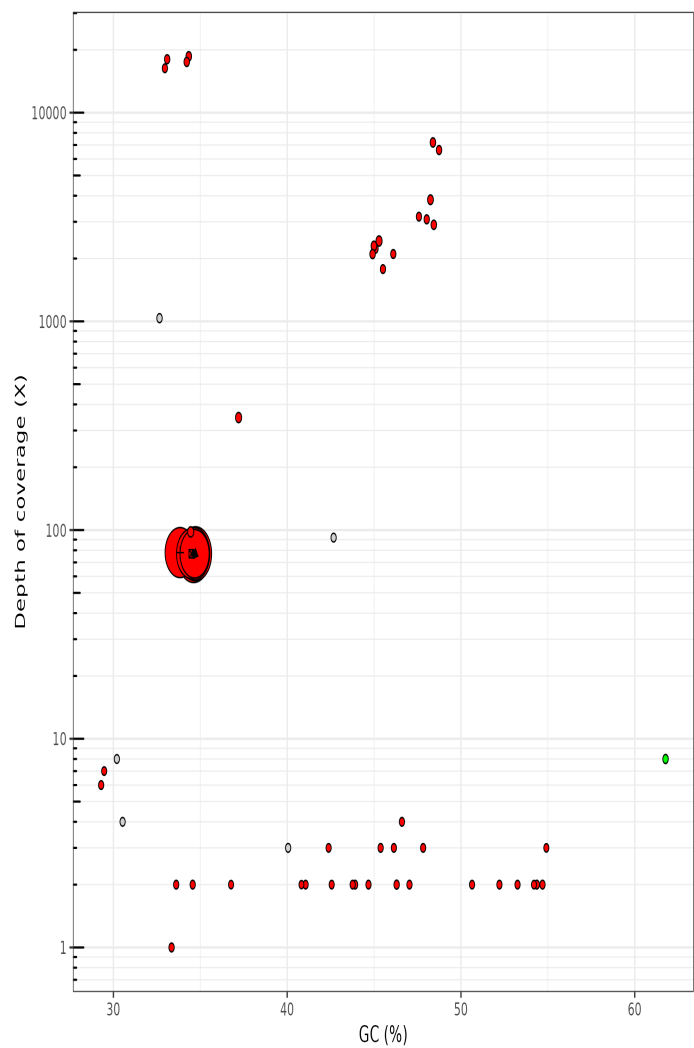


Distribution of k-mer counts coloured by their presence in reads/assemblies

Post-curation contamination screening



TAPAs summary Graph



- Length (bp)
- 5.0e+07
 - 1.0e+08
 - 1.5e+08
- Longest sequences (bp)
- SUPER_1 - 183158860 (Eukaryota)
 - ▲ SUPER_2 - 174068580 (Eukaryota)
 - SUPER_3 - 161678577 (Eukaryota)
 - + SUPER_4 - 145209119 (Eukaryota)
 - ▣ SUPER_5 - 142517981 (Eukaryota)
- superkingdom
- Bacteria
 - Eukaryota
 - N/A

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	75	43

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: Adama Ndar

Affiliation: Genoscope

Date and time: 2024-09-25 10:41:04 CEST