

ERGA Assembly Report

v24.09.10

Tags: ERGA-BGE

TxID	69427
ToLID	xgAlbTers5
Species	Albinaria teres
Class	Gastropoda
Order	Stylommatophora

Genome Traits	Expected	Observed
Haploid size (bp)	1,601,490,426	1,710,615,619
Haploid Number	20 (source: ancestor)	31
Ploidy	2 (source: ancestor)	2
Sample Sex	Unknown	Unknown

EBP metrics summary and curation notes

Obtained EBP quality metric for collapsed: 6.7.Q56

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

Curator notes

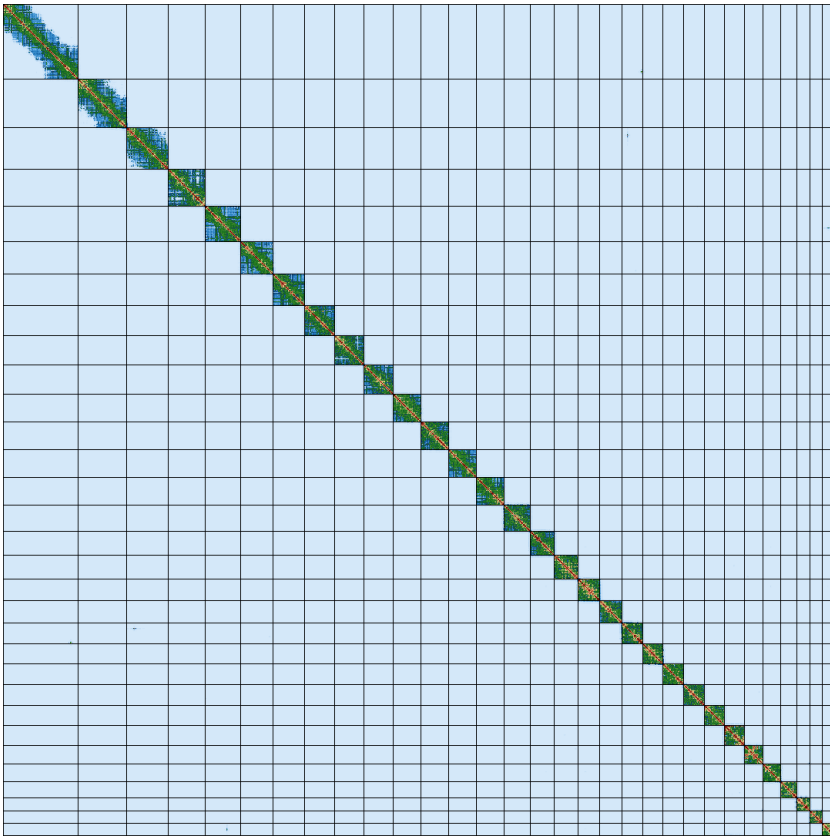
- . Interventions/Gb: 1
- . Contamination notes: ""
- . Other observations: "A large number of duplications have been removed "

Quality metrics table

Metrics	Pre-curation collapsed	Curated collapsed
Total bp	1,710,601,376	1,710,615,619
GC %	38.99	38.99
Gaps/Gbp	218.05	218.05
Total gap bp	43,000	43,100
Scaffolds	43	44
Scaffold N50	57,558,842	57,558,842
Scaffold L50	11	11
Scaffold L90	26	26
Contigs	416	417
Contig N50	7,880,036	7,880,036
Contig L50	74	74
Contig L90	226	226
QV	56.8495	56.8495
Kmer compl.	77.4389	77.4397
BUSCO sing.	96.1%	96.1%
BUSCO dupl.	3.5%	3.5%
BUSCO frag.	0.0%	0.0%
BUSCO miss.	0.4%	0.4%

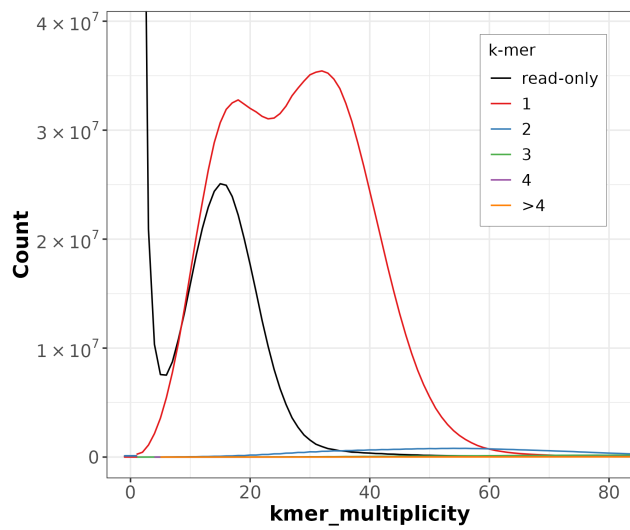
BUSCO 5.4.3 Lineage: eukaryota_odb10 (genomes:70, BUSCOs:255)

HiC contact map of curated assembly

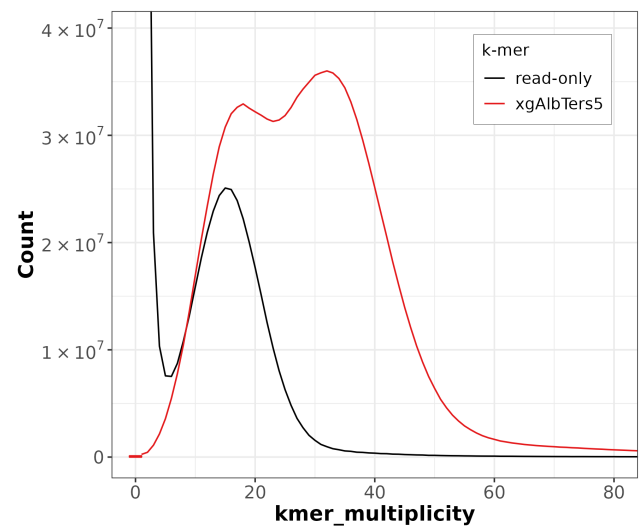


collapsed [\[LINK\]](#)

K-mer spectra of curated assembly

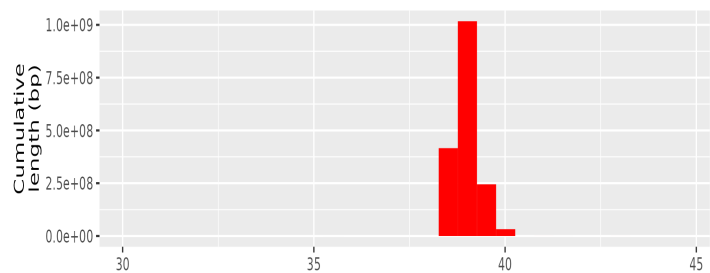


Distribution of k-mer counts per copy numbers found in asm

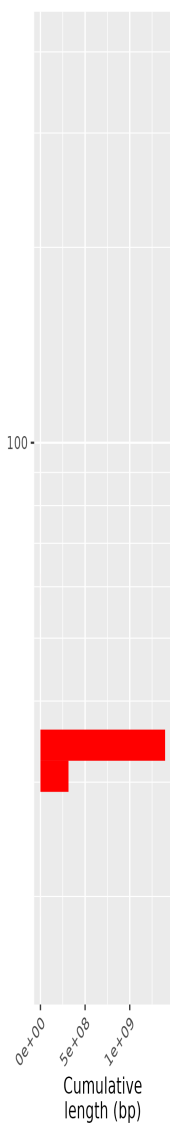
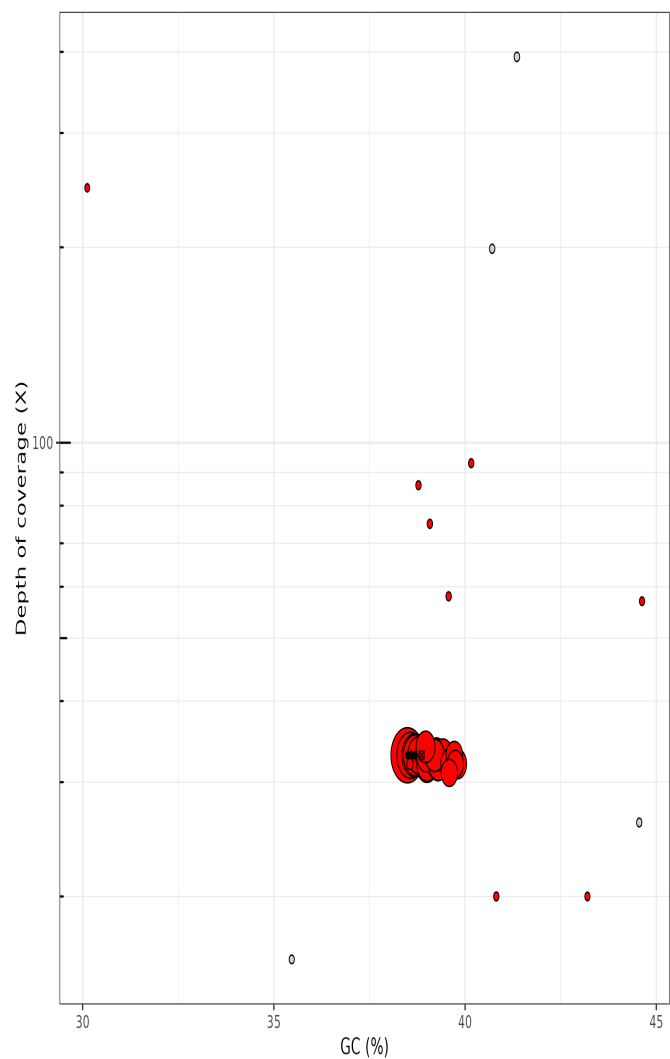


Distribution of k-mer counts coloured by their presence in reads/assemblies

Post-curation contamination screening



TAPAs summary Graph



- Longest sequences (bp)
- xgAlbTers5_1 - 154472788 (Eukaryota)
 - ▲ xgAlbTers5_2 - 100181018 (Eukaryota)
 - xgAlbTers5_3 - 84720331 (Eukaryota)
 - + xgAlbTers5_4 - 76100557 (Eukaryota)
 - ▣ xgAlbTers5_5 - 72645944 (Eukaryota)
- Length (bp)
- 5.0e+07
 - 1.0e+08
 - 1.5e+08
- superkingdom
- 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	Arima (4-enz)
Coverage	36	33

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: 2024-09-13 15:35:10 CEST