

ERGA Assembly Report

v24.04.03_beta

Tags: ERGA-BGE

TxID	55448
ToLID	1aCymNodo2
Species	Cymodocea nodosa
Class	Magnoliopsida
Order	Alismatales

Genome Traits	Expected	Observed
Haploid size (bp)	71,013,296	359,948,421
Haploid Number	7 (source: direct)	18
Ploidy	4 (source: direct)	2
Sample Sex	Unknown	Unknown

EBP metrics summary and curation notes

Obtained EBP quality metric for collapsed: 6.7.Q52

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

- . Observed Haploid size (bp) has >20% difference with Expected
- . Observed Haploid Number is different from Expected
- . Observed Ploidy is different from Expected

Curator notes

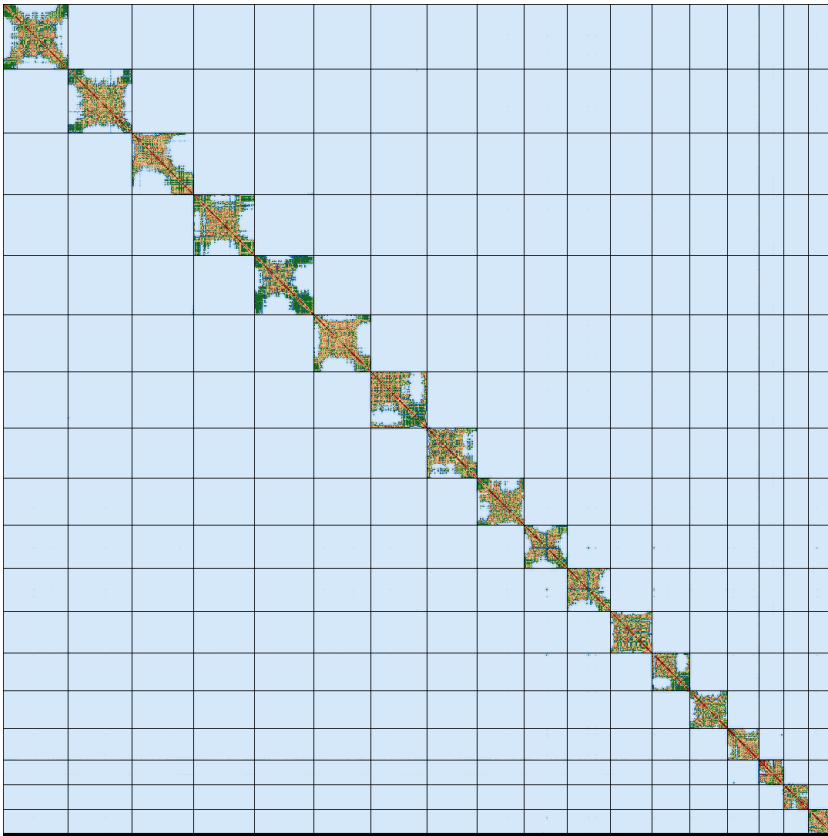
- . Interventions/Gb: 71
- . Contamination notes: "A low number of bacterial contigs were removed post-assembly "
- . Other observations: "Several low-contacts regions were detected and identified as heterozygous insertions so we decided to leave them in the assembly "

Quality metrics table

Metrics	Pre-curation collapsed	Curated collapsed
Total bp	360,243,928	359,948,421
GC %	33.22	33.22
Gaps/Gbp	113.81	125.02
Total gap bp	4,100	5,400
Scaffolds	31	31
Scaffold N50	20,283,553	24,195,584
Scaffold L50	7	7
Scaffold L90	15	15
Contigs	72	76
Contig N50	9,103,177	9,103,177
Contig L50	14	14
Contig L90	34	35
QV	6.53197	52.5369
Kmer compl.	0.0108303	92.0943
BUSCO sing.	95.3%	95.3%
BUSCO dupl.	3.9%	3.9%
BUSCO frag.	0.0%	0.0%
BUSCO miss.	0.8%	0.8%

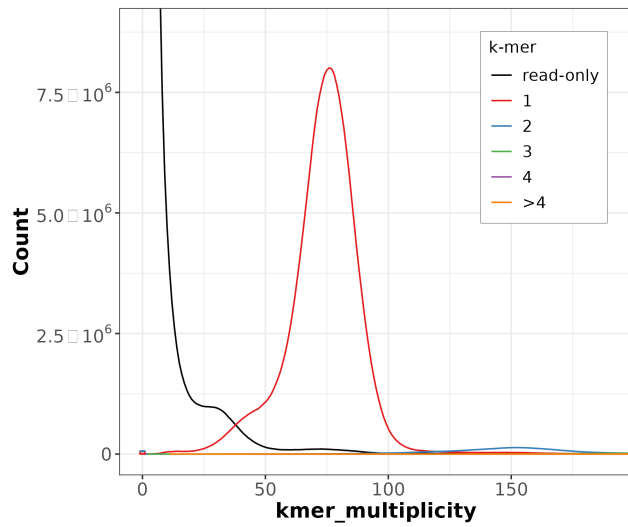
BUSCO 5.4.3 Lineage: eukaryota_odb10 (genomes:70, BUSCOs:255)

HiC contact map of curated assembly

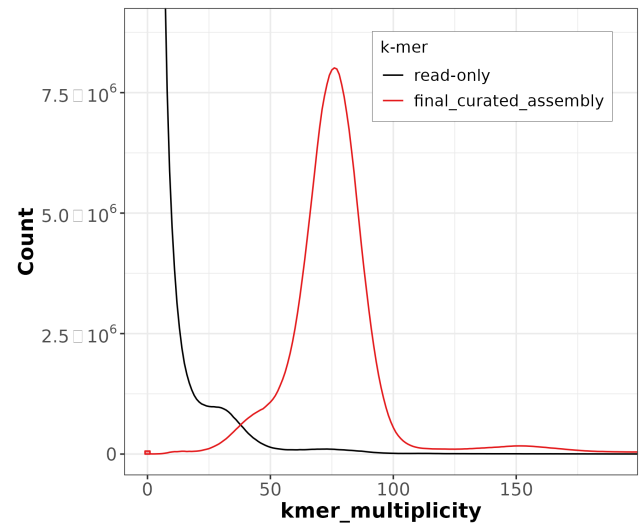


collapsed [\[LINK\]](#)

K-mer spectra of curated assembly

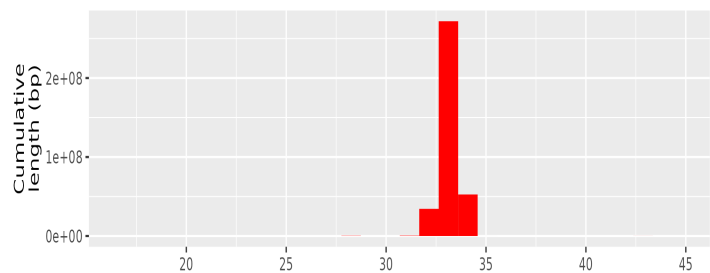


Distribution of k-mer counts per copy numbers found in asm

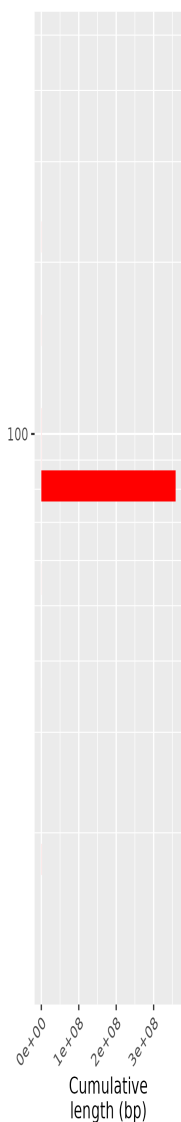
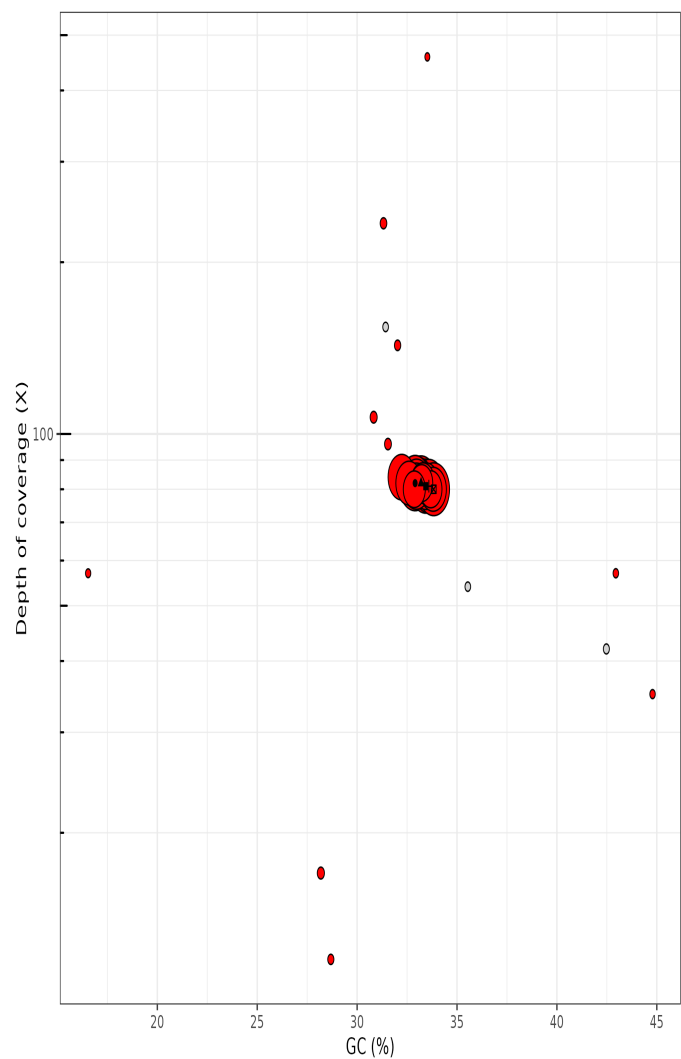


Distribution of k-mer counts coloured by their presence in reads/assemblies

Post-curation contamination screening



TAPAs summary Graph



- Length (bp)
- 1e+07
 - 2e+07
- superkingdom
- Eukaryota
 - N/A
- Longest sequences (bp)
- SUPER_1 - 28285218 (Eukaryota)
 - ▲ SUPER_2 - 27445384 (Eukaryota)
 - SUPER_3 - 26910892 (Eukaryota)
 - + SUPER_4 - 26344093 (Eukaryota)
 - ▣ SUPER_5 - 25505551 (Eukaryota)

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	40	100

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: Benjamin Istace

Affiliation: Genoscope

Date and time: 2024-09-09 07:37:25 CEST