

ERGA Assembly Report

v24.10.15

Tags: ATLASea [INVALID TAG]

TxID	40130
ToLID	xbDonTrun7.1
Species	Donax trunculus
Class	Bivalvia
Order	Cardiida

Genome Traits	Expected	Observed
Haploid size (bp)	1,500,408,723	1,578,564,802
Haploid Number	26 (source: ancestor)	20
Ploidy	2 (source: ancestor)	2
Sample Sex	Unknown	Unknown

EBP metrics summary and curation notes

Obtained EBP quality metric for collapsed: 7.7.Q64

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

Curator notes

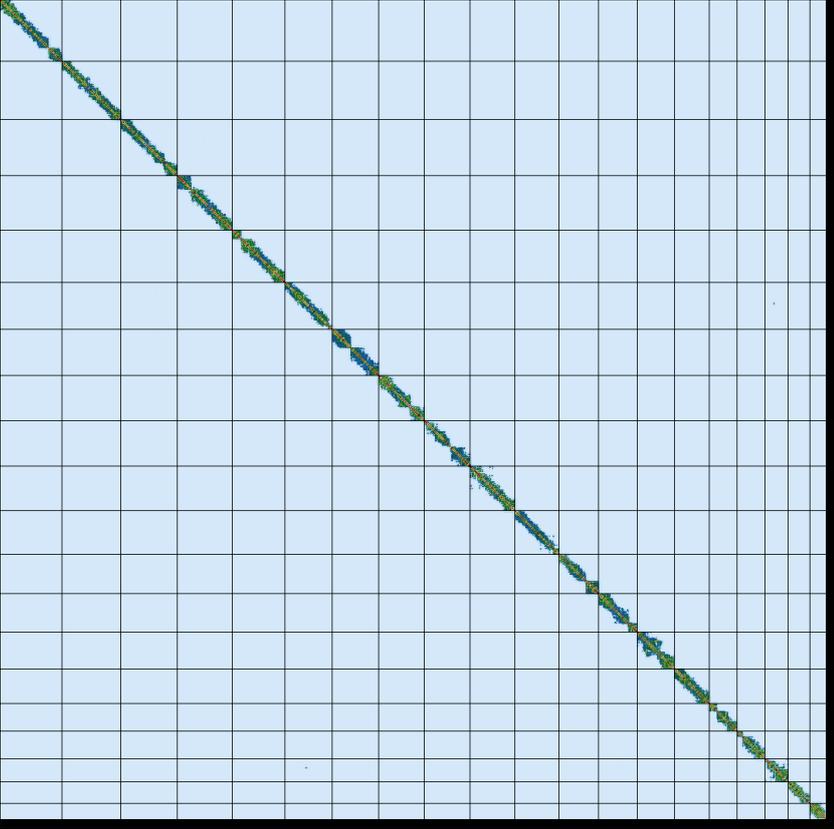
- . Interventions/Gb: 27
- . Contamination notes: ""
- . Other observations: ""

Quality metrics table

Metrics	Pre-curation collapsed	Curated collapsed
Total bp	1,624,261,491	1,578,564,802
GC %	34.51	34.52
Gaps/Gbp	61.57	63.98
Total gap bp	10,000	11,100
Scaffolds	357	247
Scaffold N50	87,739,000	86,814,400
Scaffold L50	8	8
Scaffold L90	18	17
Contigs	457	348
Contig N50	19,615,000	19,973,170
Contig L50	26	25
Contig L90	81	75
QV	47.5452	64.1357
Kmer compl.	61.1543	60.2334
BUSCO sing.	77.6%	79.3%
BUSCO dupl.	2.8%	0.9%
BUSCO frag.	4.5%	4.5%
BUSCO miss.	15.1%	15.3%

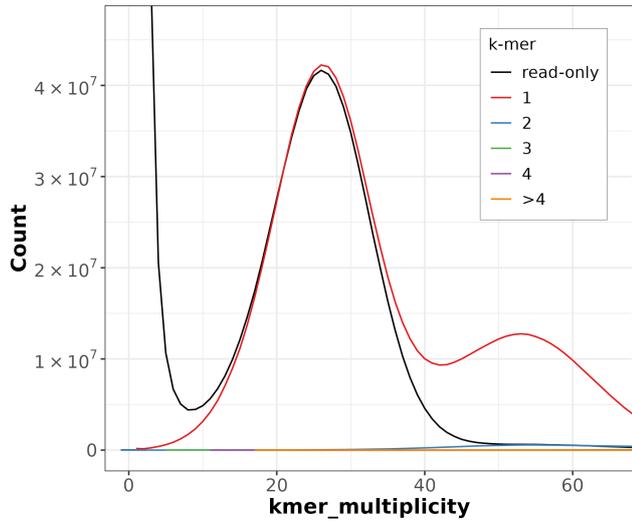
BUSCO: 5.4.3 (euk_genome_met, metaeuk) / Lineage: mollusca_odb10 (genomes:7, BUSCOs:5295)

HiC contact map of curated assembly

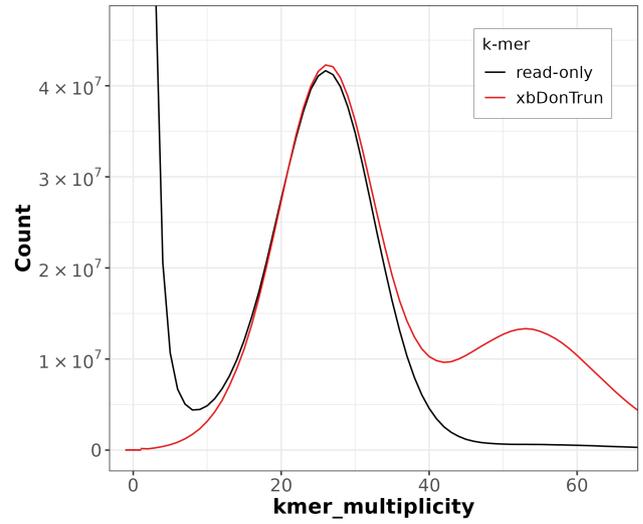


collapsed [\[LINK\]](#)

K-mer spectra of curated assembly

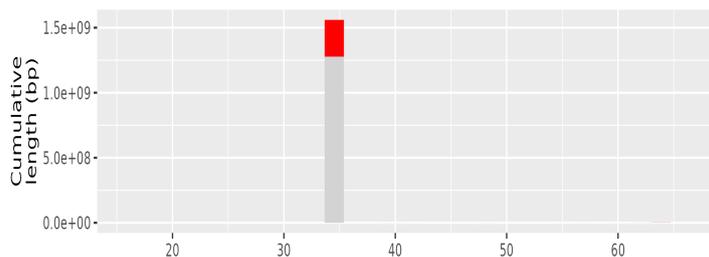


Distribution of k-mer counts per copy numbers found in asm



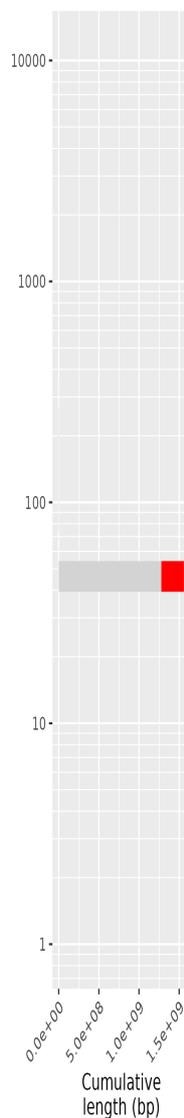
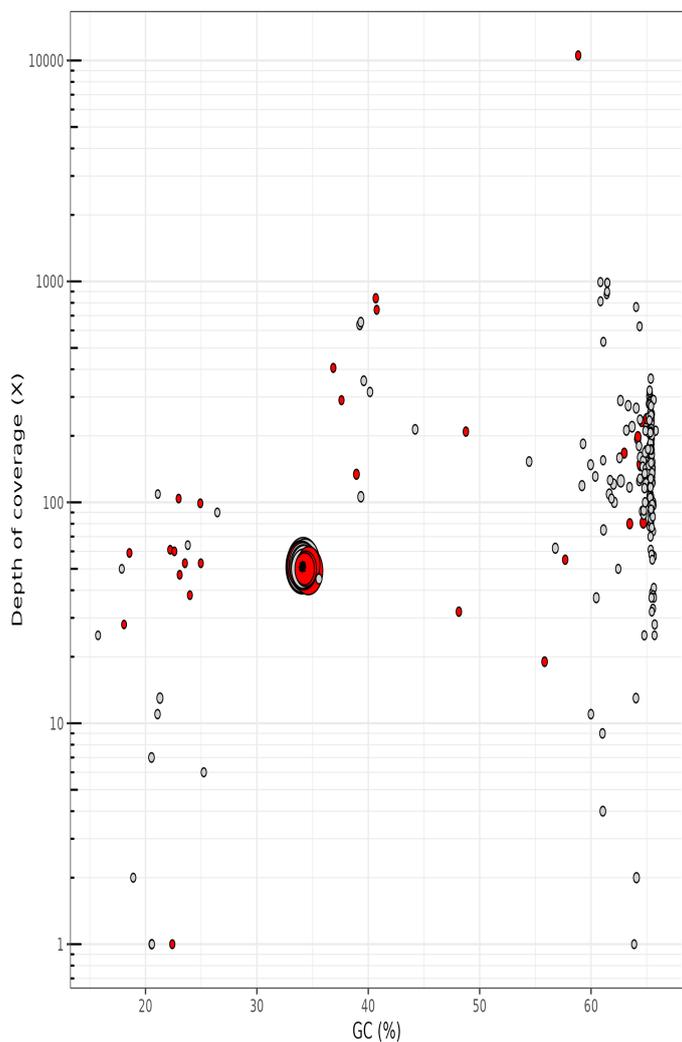
Distribution of k-mer counts coloured by their presence in reads/assemblies

Post-curation contamination screening



TAPAs summary Graph

(13 0X contigs have been hidden)



superkingdom

- Eukaryota
- N/A

Longest sequences (bp)

- SUPER_1 - 118021086 (N/A)
- ▲ SUPER_2 - 111465082 (N/A)
- SUPER_3 - 106386031 (N/A)
- + SUPER_4 - 103182086 (N/A)
- ▣ SUPER_5 - 98941469 (N/A)

Length (bp)

- 3e+07
- 6e+07
- 9e+07

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	53	22

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: Sophie Mangenot

Affiliation: Genoscope

Date and time: 2024-11-25 06:37:05 CET