

ERGA Assembly Report

v24.10.15

Tags: ATLASea[INVALID TAG]

TxID	78549
ToLID	xbCalChio1.1
Species	Callista chione
Class	Bivalvia
Order	Venerida

Genome Traits	Expected	Observed
Haploid size (bp)	1,839,942,493	1,937,208,114
Haploid Number	19 (source: ancestor)	20
Ploidy	2 (source: ancestor)	2
Sample Sex	Unknown	Unknown

EBP metrics summary and curation notes

Obtained EBP quality metric for collapsed: 7.8.Q67

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

- . Observed Haploid Number is different from Expected
- . Assembly length loss > 3% for collapsed

Curator notes

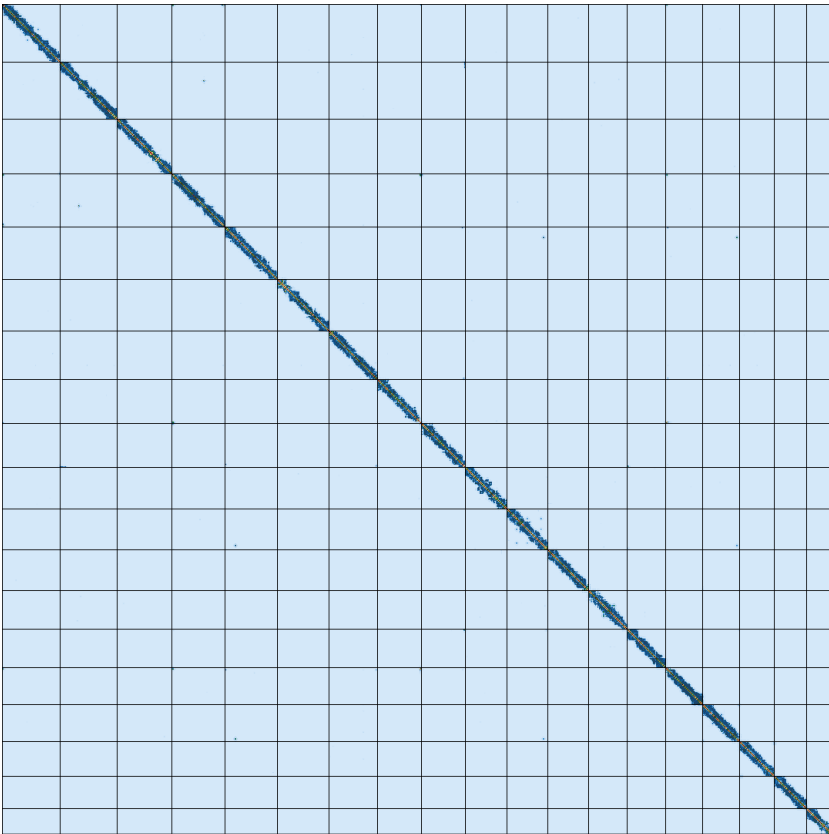
- . Interventions/Gb: 14
- . Contamination notes: "bacterial contigs were removed "
- . Other observations: "few haplotigs were removed "

Quality metrics table

Metrics	Pre-curation collapsed	Curated collapsed
Total bp	2,008,360,745	1,937,208,114
GC %	36.49	36.43
Gaps/Gbp	0	81.04
Total gap bp	0	19,000
Scaffolds	427	61
Scaffold N50	28,489,591	102,394,730
Scaffold L50	23	8
Scaffold L90	68	17
Contigs	427	218
Contig N50	28,489,591	22,152,544
Contig L50	23	29
Contig L90	68	87
QV	65.3265	67.0047
Kmer compl.	69.3901	68.3962
BUSCO sing.	96.1%	
BUSCO dupl.	3.1%	
BUSCO frag.	0.8%	
BUSCO miss.	0.0%	

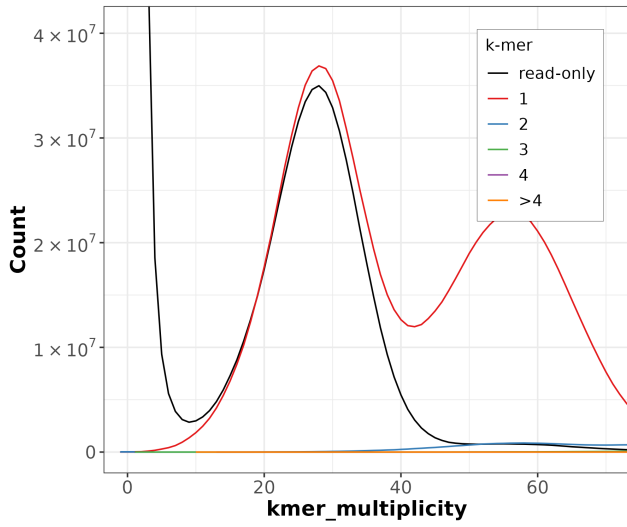
BUSCO: 5.4.3 (euk_genome_met, metaeuk) / Lineage: eukaryota_odb10 (genomes:70, BUSCOs:255)

HiC contact map of curated assembly

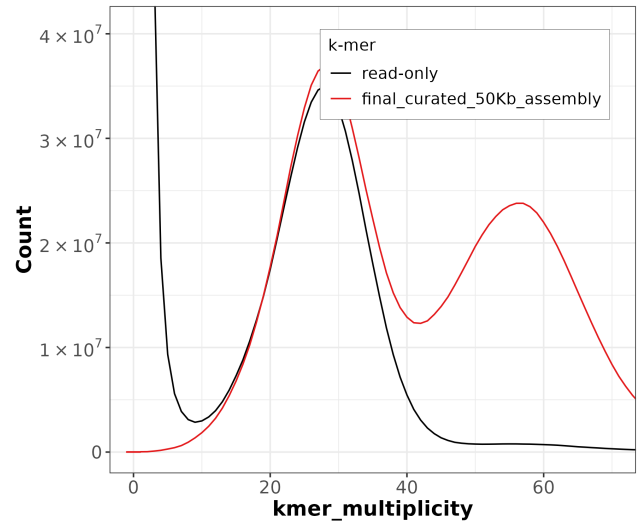


collapsed [\[LINK\]](#)

K-mer spectra of curated assembly

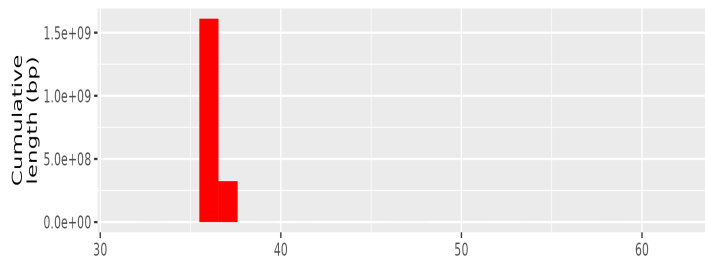


Distribution of k-mer counts per copy numbers found in asm

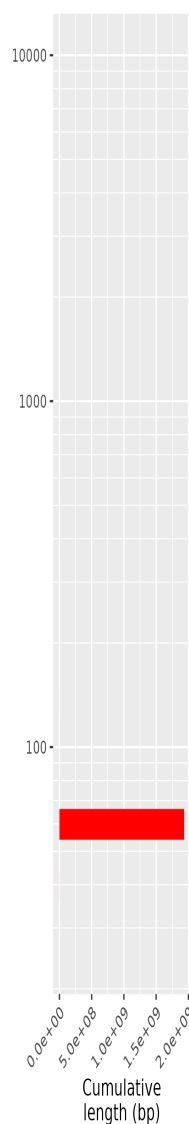
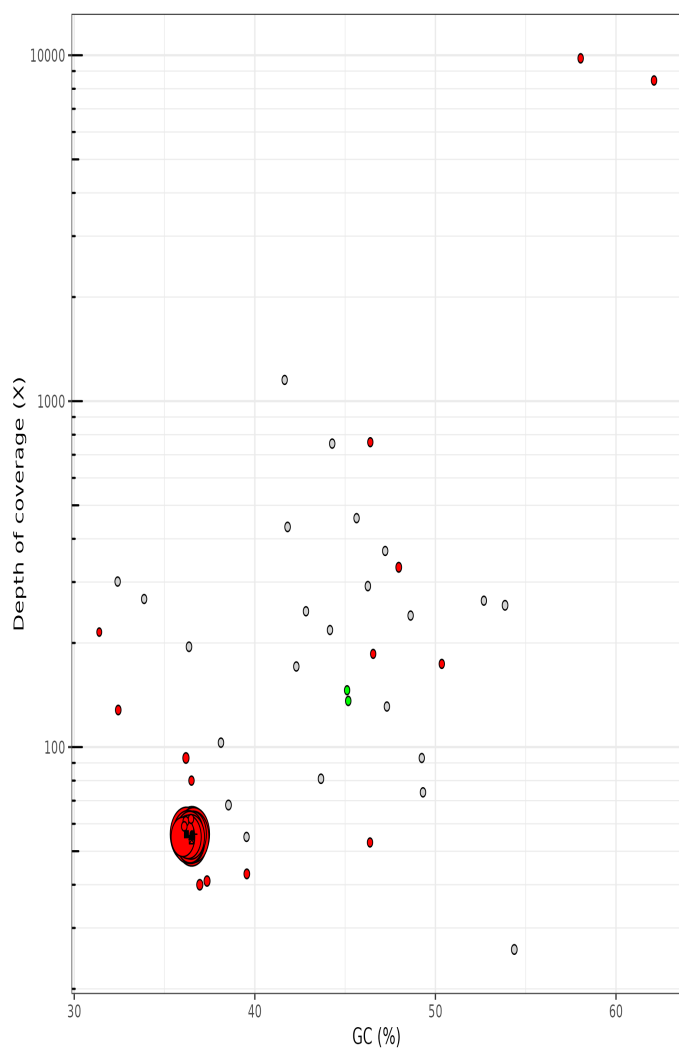


Distribution of k-mer counts coloured by their presence in reads/assemblies

Post-curation contamination screening



TAPAs summary Graph



Longest sequences (bp)

- xbCalChio1_1 - 134772284 (Eukaryota)
- ▲ xbCalChio1_2 - 133438234 (Eukaryota)
- xbCalChio1_3 - 127717548 (Eukaryota)
- + xbCalChio1_4 - 123807757 (Eukaryota)
- ▣ xbCalChio1_5 - 120891552 (Eukaryota)

Length (bp)

- 5e+07
- 1e+08

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	Arima
Coverage	58	175

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: Caroline Belser

Affiliation: Genoscope

Date and time: 2024-12-05 14:31:25 CET