

ERGA Assembly Report

v24.09.10

Tags: ATLASea [INVALID TAG]

| | |
|---------|---------------------|
| TxID | 256108 |
| ToLID | xpChiOliv1.4 |
| Species | Chiton olivaceus |
| Class | Polyplacophora |
| Order | Chitonida |

| Genome Traits | Expected | Observed |
|-------------------|----------------------|---------------|
| Haploid size (bp) | 1,333,525,716 | 1,430,799,647 |
| Haploid Number | 9 (source: ancestor) | 10 |
| Ploidy | 2 (source: ancestor) | 2 |
| Sample Sex | Unknown | Unknown |

EBP metrics summary and curation notes

Obtained EBP quality metric for collapsed: 7.8.Q65

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
- . Assembly length loss > 3% for collapsed
- . Not 90% of assembly in chromosomes for collapsed

Curator notes

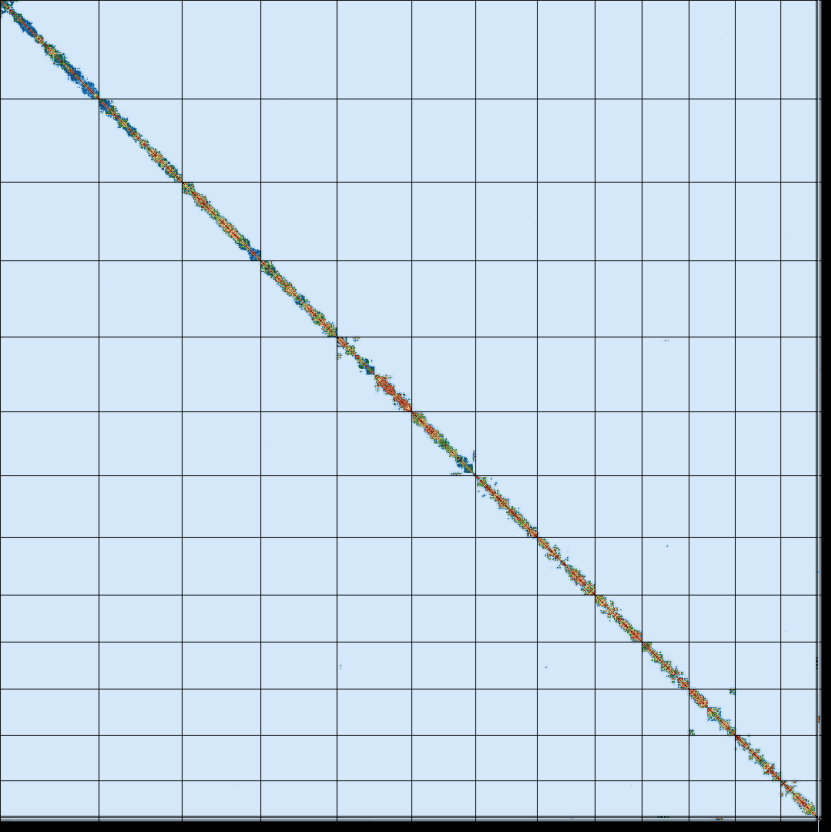
- . Interventions/Gb: 71
- . Contamination notes: ""
- . Other observations: "Low HiC reads coverage "

Quality metrics table

| Metrics | Pre-curation collapsed | Curated collapsed |
|--------------|------------------------|-------------------|
| Total bp | 1,540,050,536 | 1,430,799,647 |
| GC % | 41.63 | 41.64 |
| Gaps/Gbp | 126.62 | 146.77 |
| Total gap bp | 19,500 | 25,300 |
| Scaffolds | 129 | 114 |
| Scaffold N50 | 126,148,331 | 109,787,193 |
| Scaffold L50 | 6 | 6 |
| Scaffold L90 | 12 | 12 |
| Contigs | 324 | 324 |
| Contig N50 | 11,619,751 | 11,619,751 |
| Contig L50 | 39 | 36 |
| Contig L90 | 131 | 125 |
| QV | 65.4522 | 65.3429 |
| Kmer compl. | 63.7984 | 60.473 |
| BUSCO sing. | 91.8% | 98.4% |
| BUSCO dupl. | 7.1% | 0.0% |
| BUSCO frag. | 0.0% | 0.0% |
| BUSCO miss. | 1.1% | 1.6% |

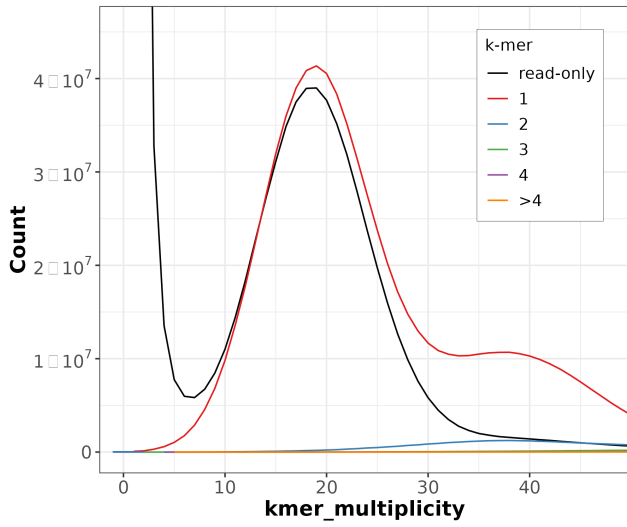
BUSCO 5.4.3 Lineage: eukaryota_odb10 (genomes:70, BUSCOs:255)

HiC contact map of curated assembly

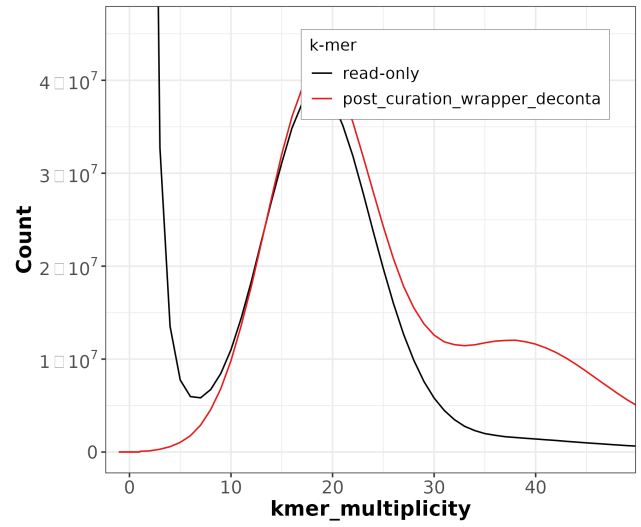


collapsed [\[LINK\]](#)

K-mer spectra of curated assembly

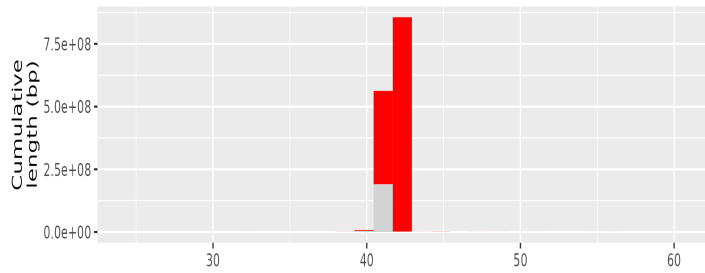


Distribution of k-mer counts per copy numbers found in asm

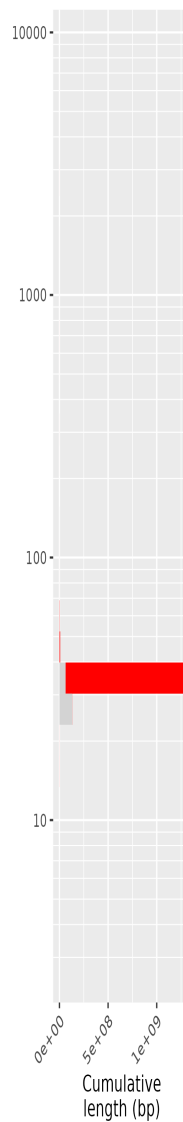
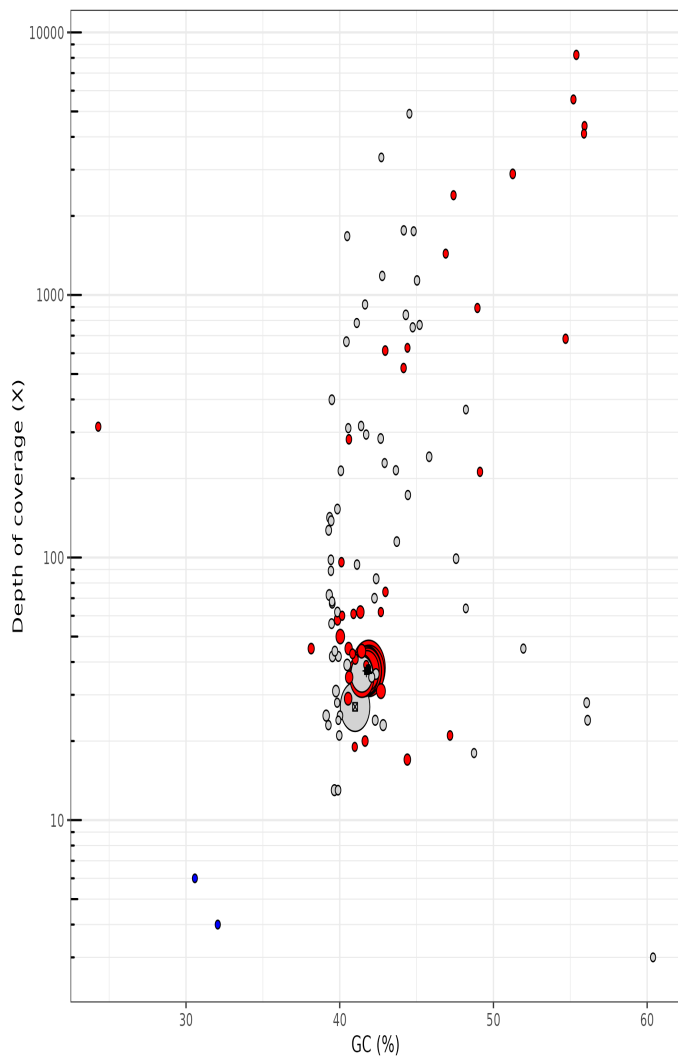


Distribution of k-mer counts coloured by their presence in reads/assemblies

Post-curation contamination screening



TAPAs summary Graph



Longest sequences (bp)

- SUPER_1 - 169973766 (Eukaryota)
- ▲ SUPER_2 - 143102446 (Eukaryota)
- SUPER_3 - 135559188 (Eukaryota)
- + SUPER_4 - 130571375 (Eukaryota)
- ⊠ SUPER_5 - 129065156 (N/A)

superkingdom

- Archaea
- Eukaryota
- N/A

Length (bp)

- 4.0e+07
- 8.0e+07
- 1.2e+08
- 1.6e+08

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 | 23 |

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: Benjamin Istace

Affiliation: Genoscope

Date and time: 2024-09-17 07:47:12 CEST