

ERGA Assembly Report

v24.10.15

Tags: ERGA-BGE

| | |
|---------|----------------------|
| TxID | 57642 |
| ToLID | xgSipPect3 |
| Species | Siphonaria pectinata |
| Class | Gastropoda |
| Order | Siphonariida |

| Genome Traits | Expected | Observed |
|-------------------|-----------------------|-------------|
| Haploid size (bp) | 674,619,290 | 819,350,751 |
| Haploid Number | 16 (source: ancestor) | 16 |
| Ploidy | 2 (source: ancestor) | 2 |
| Sample Sex | unknown | unknown |

EBP metrics summary and curation notes

Obtained EBP quality metric for collapsed: 6.7.Q55

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

- . Observed Haploid size (bp) has >20% difference with Expected
- . More than 1000 gaps/Gbp for collapsed

Curator notes

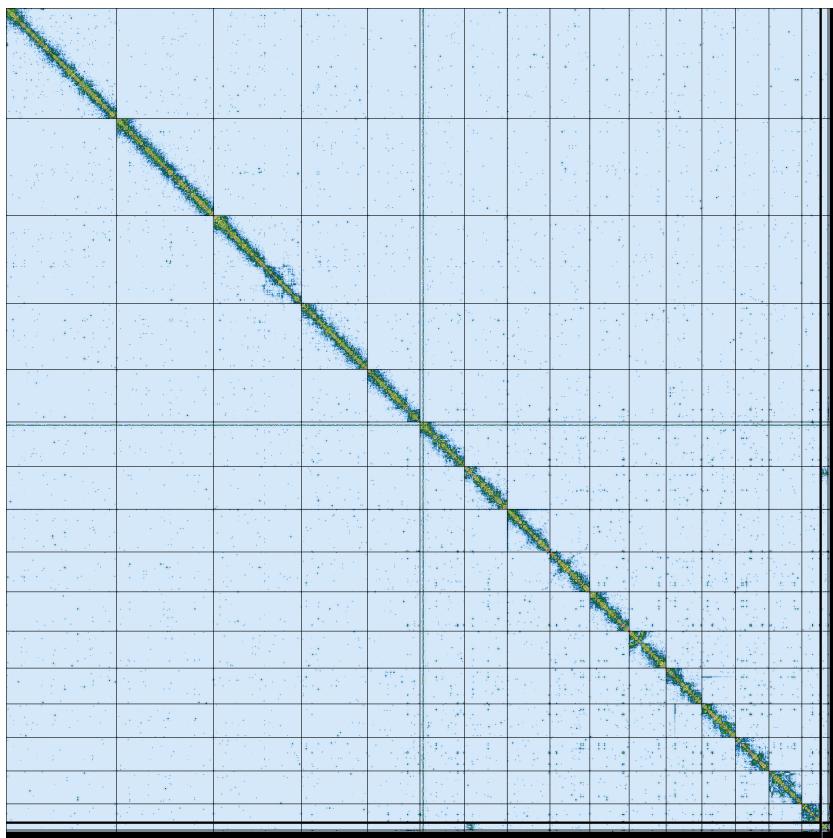
- . Interventions/Gb: 69
- . Contamination notes: "Few contigs (5, bacterial) were detected by FCS-GX or blob toolkit and were removed."
- . Other observations: "The assembly was produced with hifiasm using HiFi data. The resulting collapsetd *.bp.p_ctg.gfa was selected, screened for contamination with FCS-GX, and purged of haplotigs with purge_dups. Scaffolding was performed with YaHS. The HiC data was not of great use for curation, pairtools reported the fraction frac_cis_40kb+ being 0.042. 57 changes were made using PretextView."

Quality metrics table

| Metrics | Pre-curation collapsed | Curated collapsed |
|--------------|------------------------|-------------------|
| Total bp | 830,088,125 | 819,350,751 |
| GC % | 37.18 | 37.17 |
| Gaps/Gbp | 1,017.96 | 1,048.39 |
| Total gap bp | 169,000 | 171,800 |
| Scaffolds | 255 | 227 |
| Scaffold N50 | 43,728,786 | 43,574,499 |
| Scaffold L50 | 6 | 6 |
| Scaffold L90 | 14 | 14 |
| Contigs | 1,100 | 1,086 |
| Contig N50 | 1,870,862 | 1,839,735 |
| Contig L50 | 134 | 134 |
| Contig L90 | 460 | 459 |
| QV | 55.4 | 55.7 |
| Kmer compl. | 81.42 | 99.99 |
| BUSCO sing. | 96.0% | 96.8% |
| BUSCO dupl. | 2.1% | 1.2% |
| BUSCO frag. | 0.3% | 0.3% |
| BUSCO miss. | 1.6% | 1.7% |

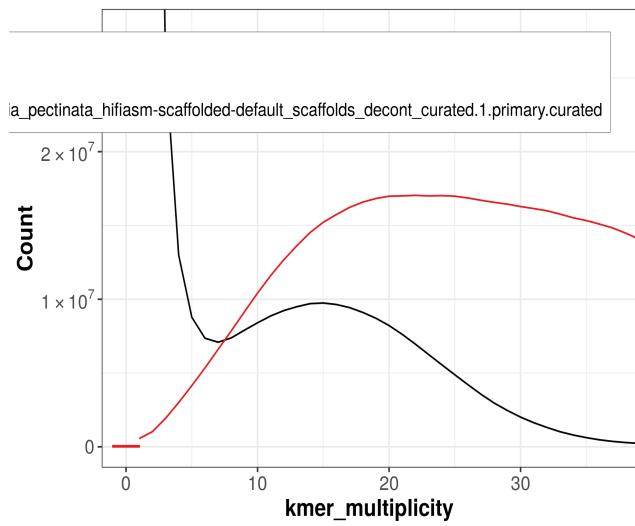
BUSCO: 6.0.0 (euk_genome_min, miniprot) / Lineage: mollusca_odb10 (genomes:7, BUSCOs:5295)

HiC contact map of curated assembly

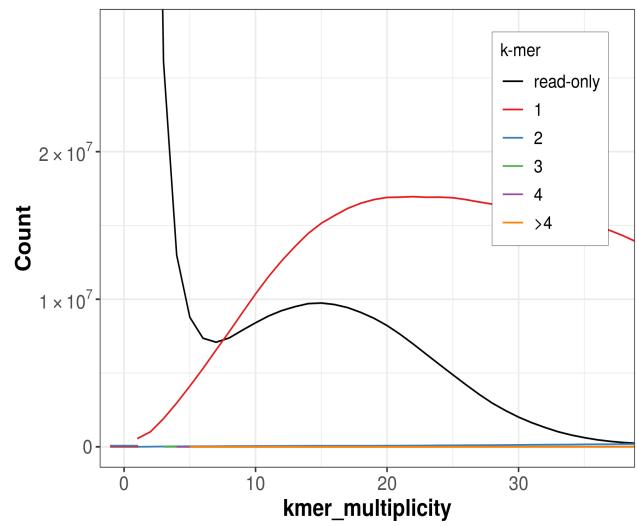


collapsed [\[LINK\]](#)

K-mer spectra of curated assembly

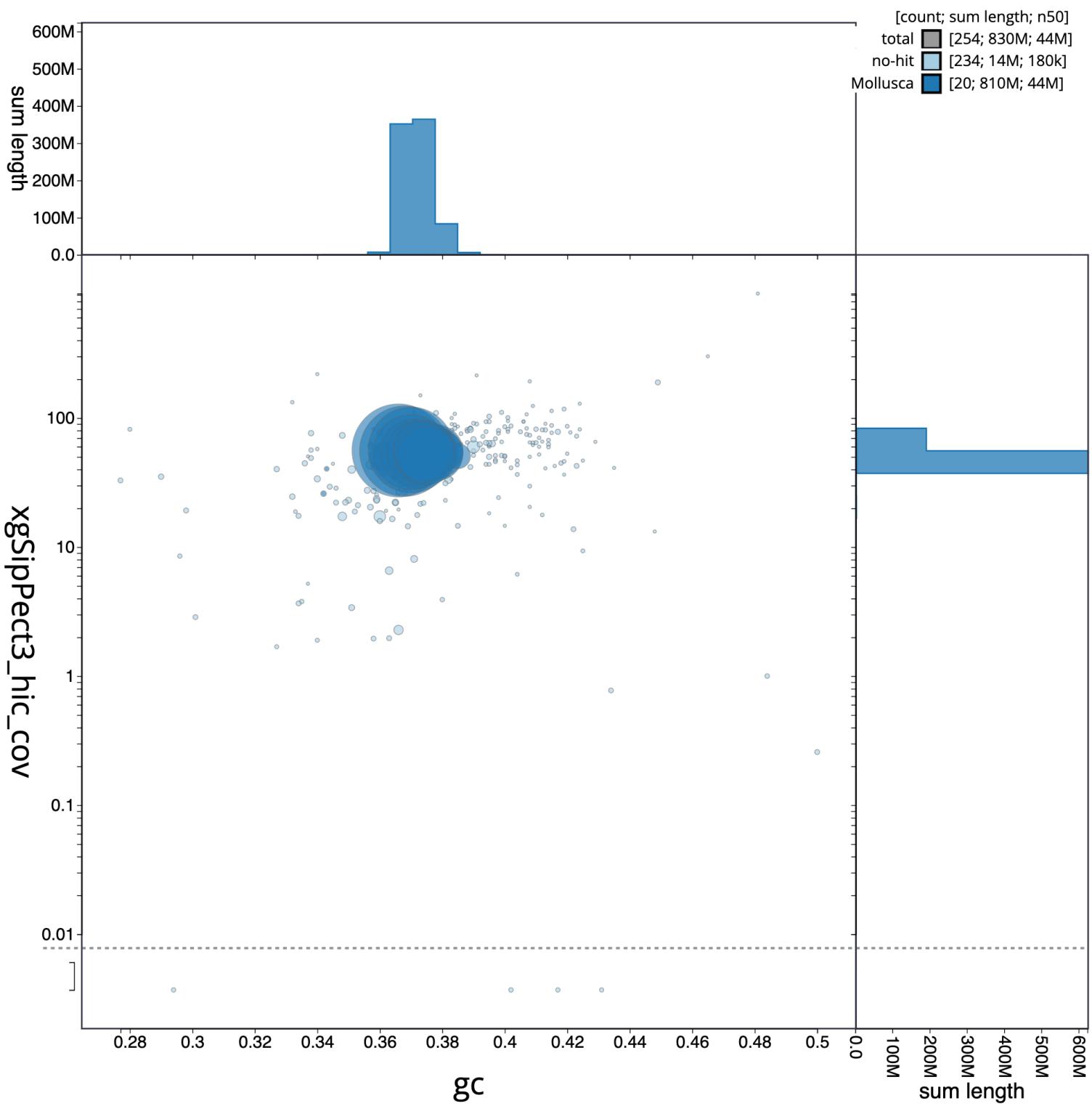


Distribution of k-mer counts coloured by their presence in reads/assemblies



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

Post-curation contamination screening



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 HiC v2 |
|----------|-------------|--------------|
| Coverage | 35X | 72X |

Assembly pipeline

```
- hifiasm
  |_ ver: 0.25.0-r726
  |_ key param: NA
- purge_dups
  |_ ver: 1.2.6
  |_ key param: NA
- YaHS
  |_ ver: 1.2a.2
  |_ key param: NA
- NBIS/Earth-Biogenome-Project-pilot
  |_ ver: e6e9c1621b
  |_ key param: NA
- sanger-tol/blob toolkit
  |_ ver: 0.9.0
  |_ key param: NA
```

Curation pipeline

```
- sanger-tol/curationpretext
  |_ ver: 1.5.1
  |_ key param: NA
```

Submitter: Estelle Proux-Wéra

Affiliation: SciLifeLab

Date and time: 2025-11-25 14:27:45 CET