

ERGA Assembly Report

v24.10.15

Tags: ERGA-BGE

TxID	581057
ToLID	kaBotViol
Species	<i>Botrylloides violaceus</i>
Class	Asciidiacea
Order	Stolidobranchia

Genome Traits	Expected	Observed
Haploid size (bp)	139,513,824	130,369,453
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 pri: 6.6.Q39

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

- . QV value is less than 40 for pri
- . Kmer completeness value is less than 90 for pri
- . Assembly length loss > 3% for pri

Curator notes

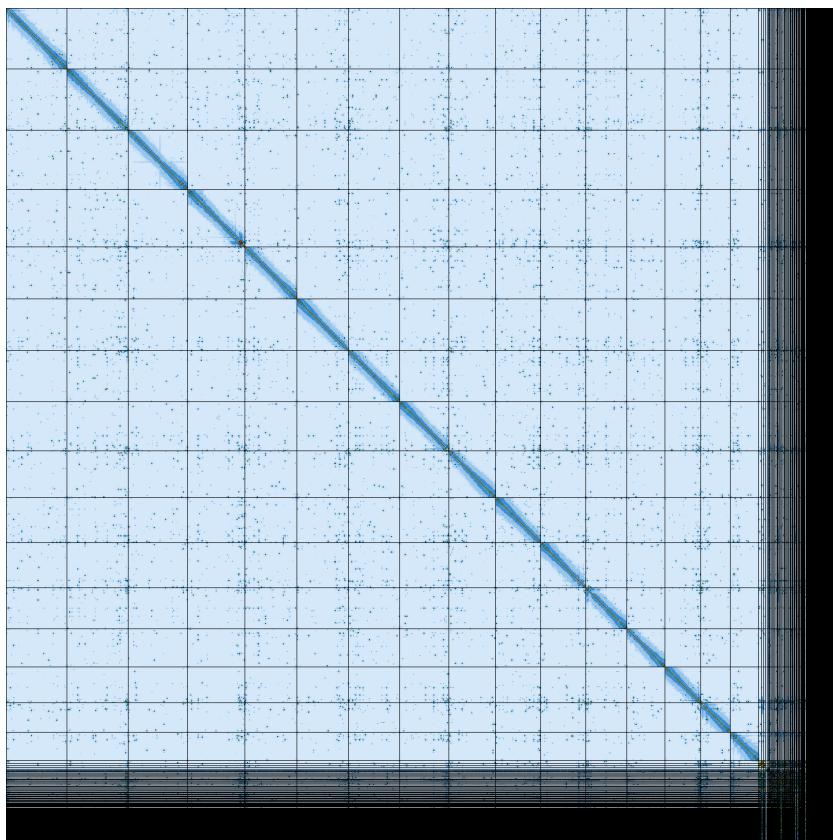
- . Interventions/Gb: 93
- . Contamination notes: "30 sequences were removed as contaminants. In addition, 2 large scaffolds (SUPER_6, SUPER_7) were classified as Mollusca. However, they were not removed because their coverage and GC content were consistent with the rest of the genome, and they showed no Hi-C contacts with other scaffolds."
- . Other observations: "The assembly of species *Botrylloides violaceus* is based on 1493x long read ONT data and 1085x Arima HiC data generated as part of the ERGA initiative via the BGE project. The assembly process included the following steps: ONT reads shorter than 3 kb were filtered out, thus the remaining reads for a total of ~215 Gb were assembled using Necat with polishing performed using NextPolish. Contigs were subsequently filtered with purge_dups. Scaffolding was performed by aligning HiC reads to the purged contigs using the Omni-C mapping pipeline, followed by YaHS for scaffold construction. Contamination was checked using BlobTools. The scaffolds were processed through sanger_tol/curationpretext pipeline to generate the contact map, which was manually improved using PretextView. The presence of a mitochondrial genome contig/scaffold was not assessed. Chromosome-scale scaffolds confirmed by HiC data were named in order of size."

Quality metrics table

Metrics	Pre-curation pri	Curated pri
Total bp	141,960,073	130,369,453
GC %	41.24	41.14
Gaps/Gbp	331.08	460.23
Total gap bp	4,700	9,200
Scaffolds	258	174
Scaffold N50	7,416,636	7,692,581
Scaffold L50	9	8
Scaffold L90	57	16
Contigs	305	234
Contig N50	3,327,369	3,355,943
Contig L50	16	14
Contig L90	101	62
QV	39.8281	39.7479
Kmer compl.	80.5841	78.7004
BUSCO sing.	89.8%	91.8%
BUSCO dupl.	2.8%	0.4%
BUSCO frag.	1.7%	1.9%
BUSCO miss.	5.7%	5.9%

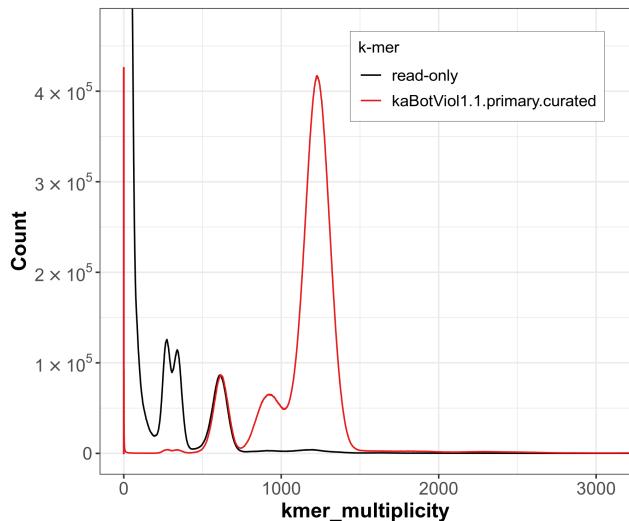
BUSCO: 5.8.2 (euk_genome_aug, augustus) / Lineage: metazoa_odb10 (genomes:65, BUSCOs:954)

HiC contact map of curated assembly

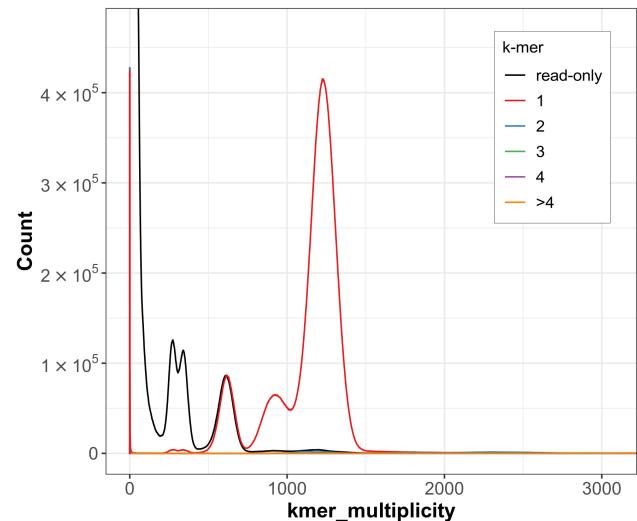


pri [LINK]

K-mer spectra of curated assembly



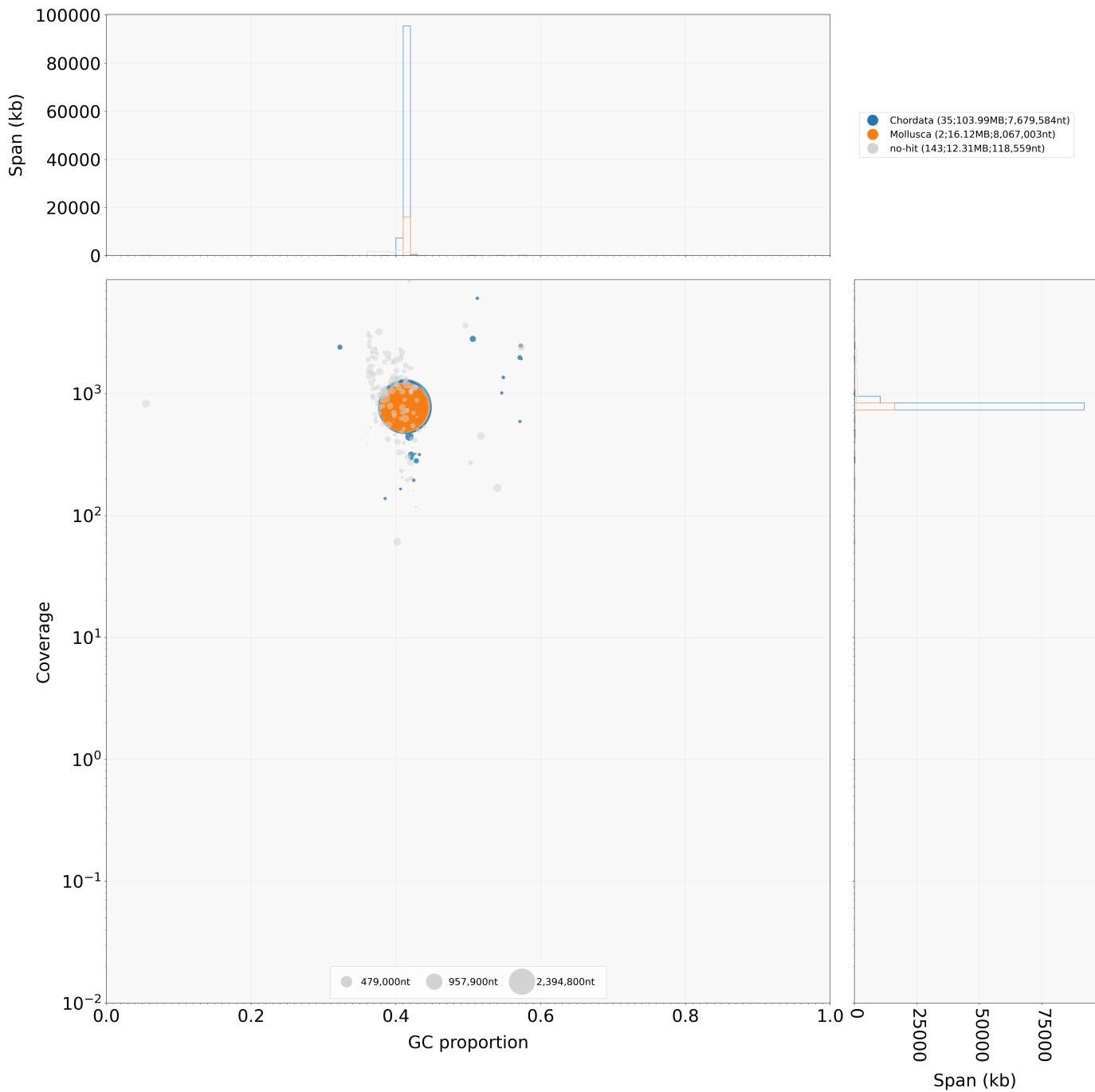
Distribution of k-mer counts coloured by their presence in reads/assemblies



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

Post-curation contamination screening

filename.blobDB.json.bestsum.phylum.p8.span.100.blobplot.bam0



pri. 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	ONT	Arima HiC
Coverage	1493x	1085x

Assembly pipeline

```
- Necat
  |_ ver: 0.0.1
  |_ key param: NA
- purge_dups
  |_ ver: 1.2.5
  |_ key param: NA
- YaHS
  |_ ver: 1.2.2
  |_ key param: NA
```

Curation pipeline

```
- sanger-tol/curationpretext
  |_ ver: 1.5.0
  |_ key param: NA
- PretextView
  |_ ver: 0.2.5
  |_ key param: NA
```

Submitter: Ilenia Urso
Affiliation: UNIBA

Date and time: 2025-12-12 09:52:08 CET