ERGA Assembly Report
v24.10.15

Tags: ERGA-Pilot

TxID	30301
ToLID	**kaBotSchl1**
Species	Botryllus schlosseri
Class	Ascidiacea
Order	Stolidobranchia

Genome Traits	Expected	Observed
Haploid size (bp)	536,801,721	531,369,912
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: 6.7.Q41

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

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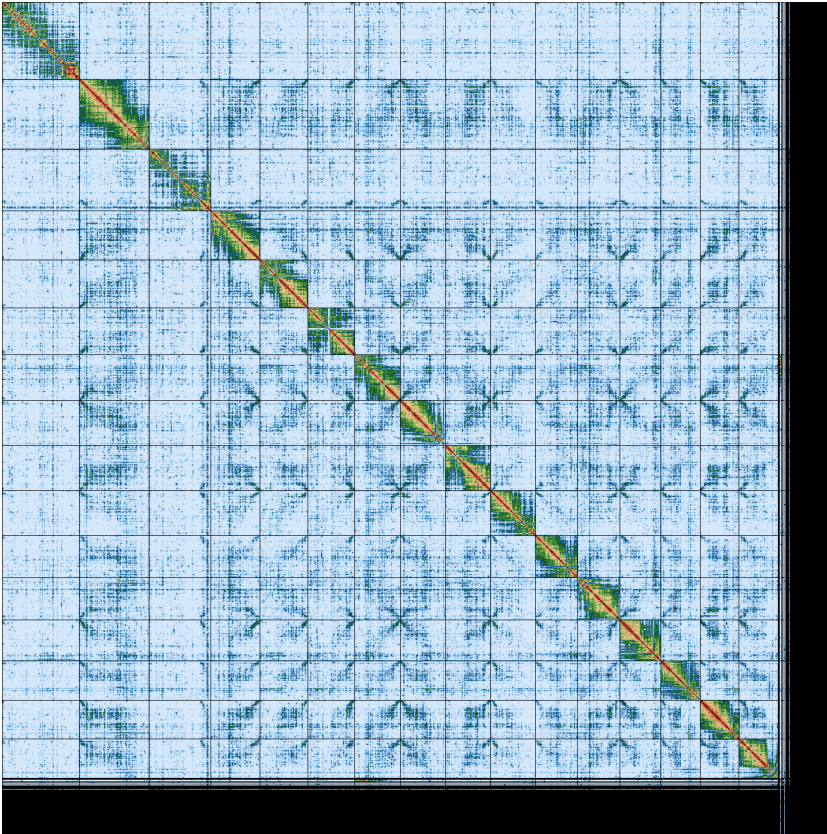. Assembly length loss > 3% for pri

Curator notes

. Interventions/Gb: 228
. Contamination notes: "No presence of contaminants."
. Other observations: "The assembly of species Botryllus schlosseri (kaBotSchl) is based on 135X long read ONT data and 141X Dovetail OmniC data generated as part of the European Reference Genome Atlas (ERGA, https://www.erga-biodiversity.eu/) via the Pilot 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 ~81 Gb were assembled using Necat with polishing performed using NextPolish. Contigs were subsequently filtered with purge_dups to remove haplotigs and obtain a more accurate, non-redundant assembly. Scaffolding was performed by aligning Omni-C 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 Omni-C data were named in order of size."

Quality metrics table

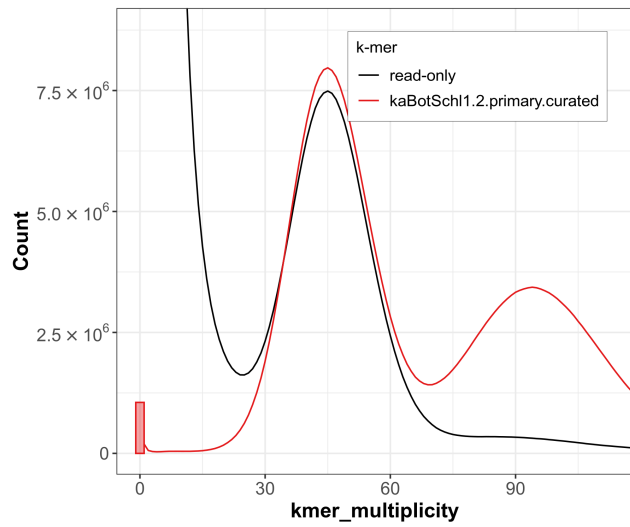| Metrics	Pre-curation pri	Curated pri
Total bp	669,506,172	531,369,912
GC %	40.46	40.46
Gaps/Gbp	648.24	907.09
Total gap bp	86,800	62,200
Scaffolds	560	423
Scaffold N50	29,857,706	28,922,816
Scaffold L50	10	8
Scaffold L90	49	16
Contigs	994	905
Contig N50	1,847,095	1,770,787
Contig L50	112	92
Contig L90	373	317
QV	42.6514	41.9236
Kmer compl.	75.4032	63.5014
BUSCO sing.	84.5%	90.0%
BUSCO dupl.	7.8%	0.8%
BUSCO frag.	1.5%	1.5%
BUSCO miss.	6.3%	7.7%

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

HiC contact map of curated assembly

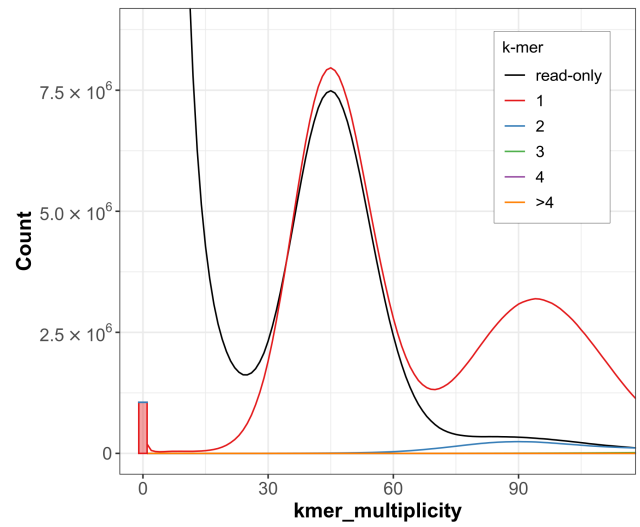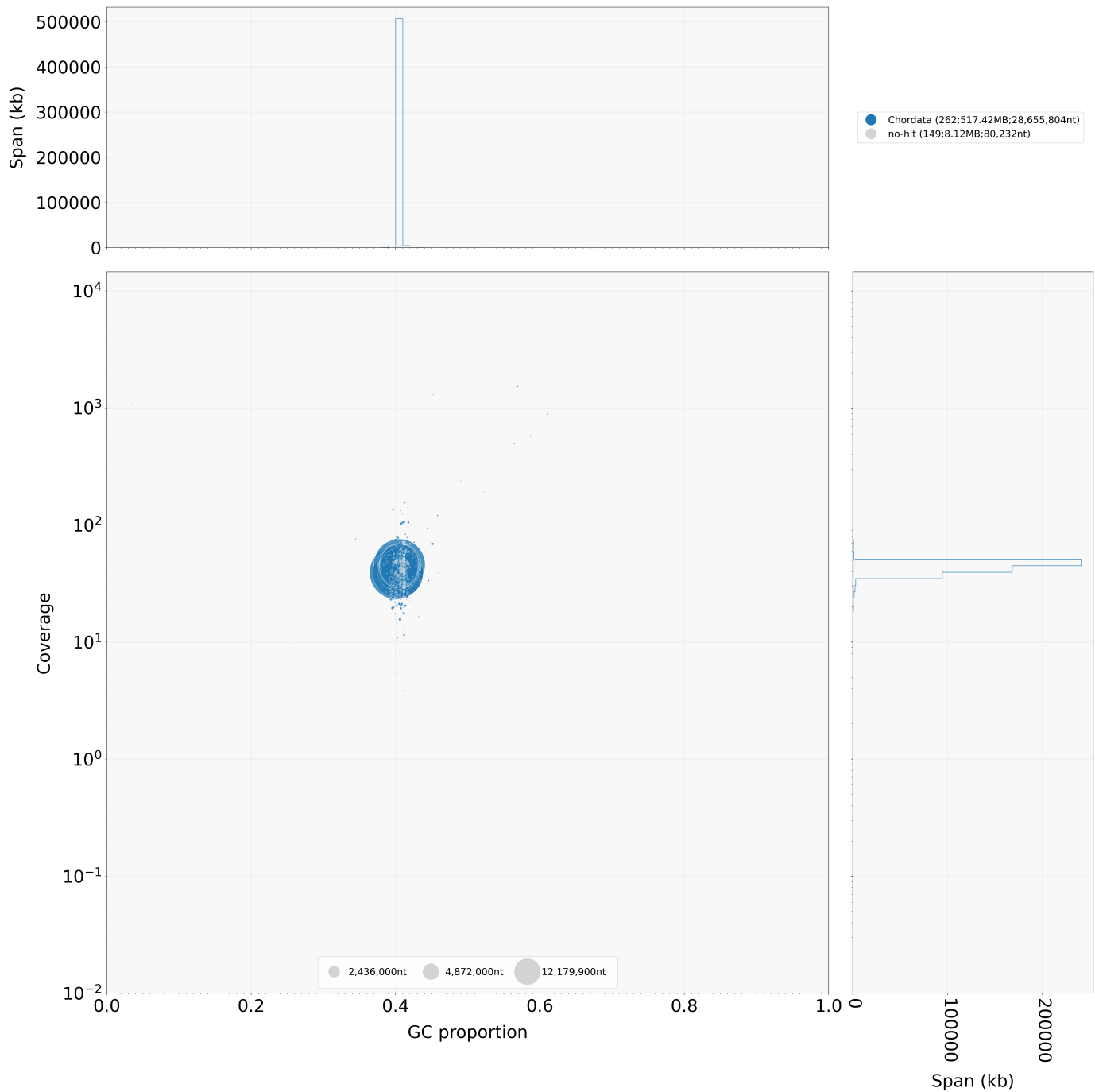**pri** [LINK]

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

Legend:
- Chordata (262;517.42MB;28,655,804nt)
- no-hit (149;8.12MB;80,232nt)

Bubble sizes: 2,436,000nt 4,872,000nt 12,179,900nt

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	OmniC
Coverage	135x	141x

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:* 1.0.5
    |_ *key param:* NA

Submitter: Ilenia Urso
Affiliation: UNIBA

Date and time: 2025-12-22 09:35:22 CET