ERGA Assembly Report

v24.10.15

Tags: ATLASea[INVALID TAG]

TxID	62933
ToLID	**kaBotDieg4**
Species	Botrylloides diegensis
Class	Ascidiacea
Order	Stolidobranchia

Genome Traits	Expected	Observed
Haploid size (bp)	177,063,050	202,711,179
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: 6.6.Q47

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

. Kmer completeness value is less than 90 for collapsed
. Not 90% of assembly in chromosomes for collapsed

Curator notes

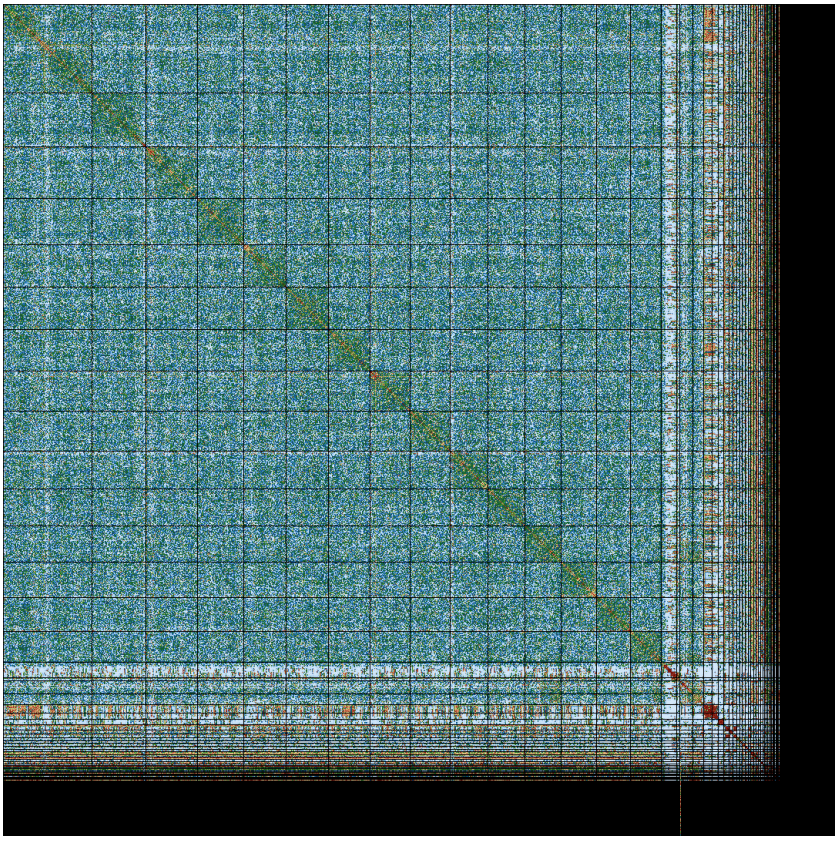
. Interventions/Gb: 34
. Contamination notes: ""
. Other observations: "The assembly of Botrylloides diegensis (kaBotDieg4) is based on 74X PacBio data and 382X OmniC Hi-C data generated as part of the ATLASea programme (https://www.atlasea.fr). The assembly process included the following steps: initial PacBio assembly generation with Hifiasm, removal of contaminant sequences using Context, removal of haplotypic duplications using purge_dups, and Hi-C-based scaffolding with YaHS. In total, 40 contigs were identified as contaminants (bacterial, archaeal, or viral), totaling 1.217 Mb (with the largest being 0.073 Mb). Additionally, 108 regions totaling 5.72 Mb (with the largest being 0.236 Mb) were identified as haplotypic duplications and removed. Mitochondrial genome was assembled using OATK. Finally, the primary assembly was analyzed and manually improved using Pretext. Chromosome-scale scaffolds confirmed by Hi-C data were named in order of size. "

Quality metrics table

Metrics	Pre-curation collapsed	Curated collapsed
Total bp	203,354,507	202,711,179
GC %	40.78	40.8
Gaps/Gbp	157.36	152.93
Total gap bp	3,200	3,300
Scaffolds	234	221
Scaffold N50	22,573,799	9,728,604
Scaffold L50	4	9
Scaffold L90	25	31
Contigs	266	252
Contig N50	9,169,000	9,128,188
Contig L50	10	10
Contig L90	45	44
QV	47.2247	47.2338
Kmer compl.	78.8509	78.8489
BUSCO sing.	96.0%	96.0%
BUSCO dupl.	0.6%	0.6%
BUSCO frag.	1.0%	1.0%
BUSCO miss.	2.4%	2.4%

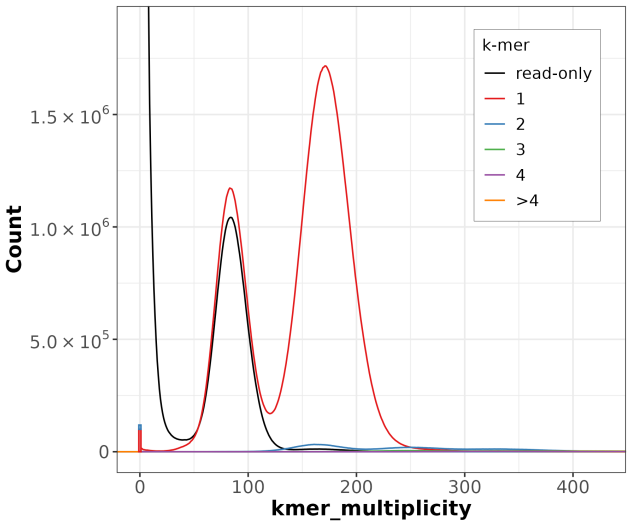
BUSCO: 6.0.0 (euk_genome_min, miniprot) / Lineage: metazoa_odb12 (genomes:206, BUSCOs:672)

HiC contact map of curated assembly

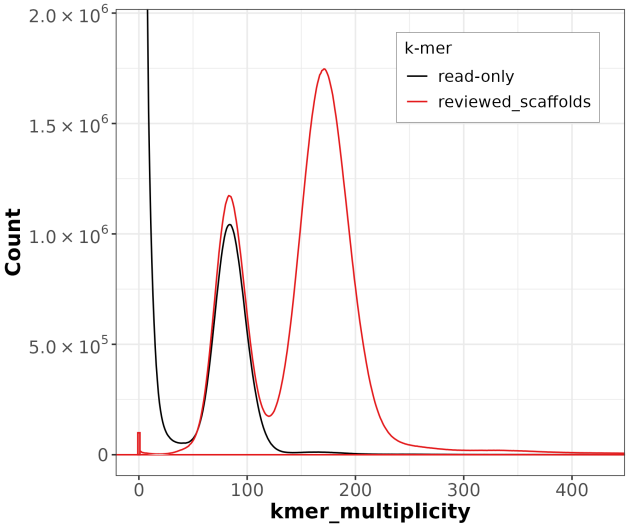


collapsed [LINK]

K-mer spectra of curated assembly

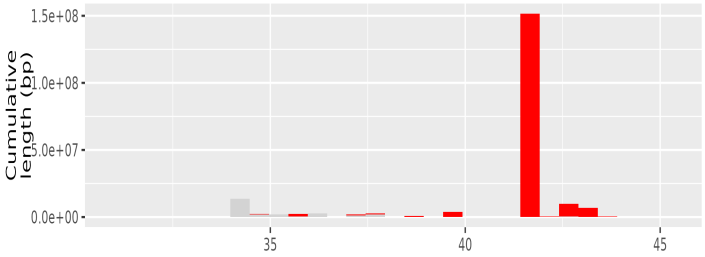


Distribution of k-mer counts per copy numbers found in asm

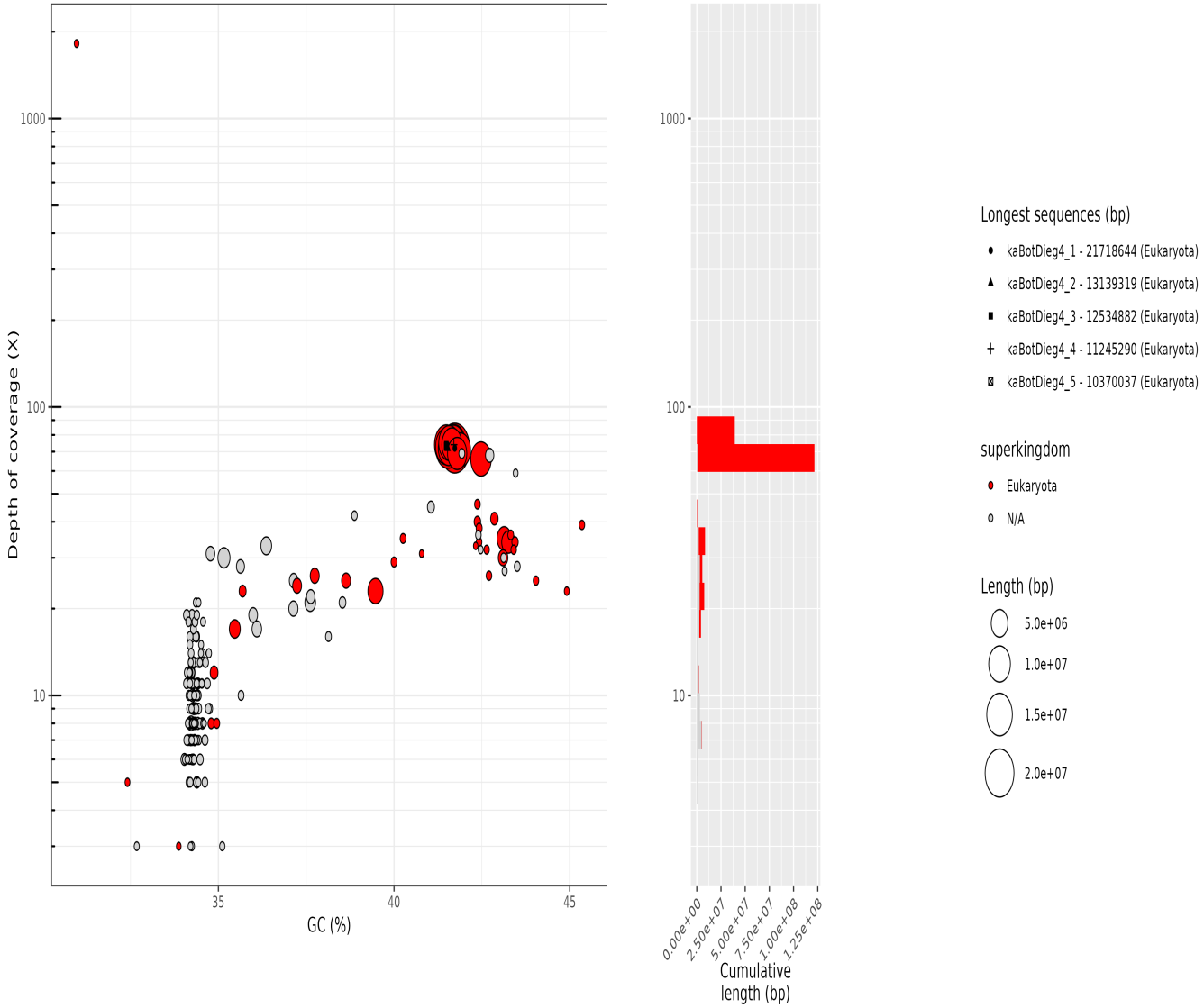


Distribution of k-mer counts coloured by their presence in reads/assemblies

Post-curation contamination screening



TAPAs summary Graph



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	Long reads	Omnic
Coverage	64	334

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: Arnaud Couloux
Affiliation: Genoscope

Date and time: 2026-03-19 19:36:19 CET