ERGA Assembly Report

v24.10.15

Tags: ATLASea[INVALID TAG]

TxID	2907997
ToLID	**kaClaSabb1**
Species	Clavelina sabbadini
Class	Ascidiacea
Order	Aplousobranchia

Genome Traits	Expected	Observed
Haploid size (bp)	168,400,425	157,640,194
Haploid Number	9 (source: ancestor)	9
Ploidy	2 (source: ancestor)	2
Sample Sex	Unknown	Unknown

EBP metrics summary and curation notes

Obtained EBP quality metric for collapsed: 6.7.Q61

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

. Kmer completeness value is less than 90 for collapsed
. BUSCO single copy value is less than 90% for collapsed
. Assembly length loss > 3% for collapsed

Curator notes

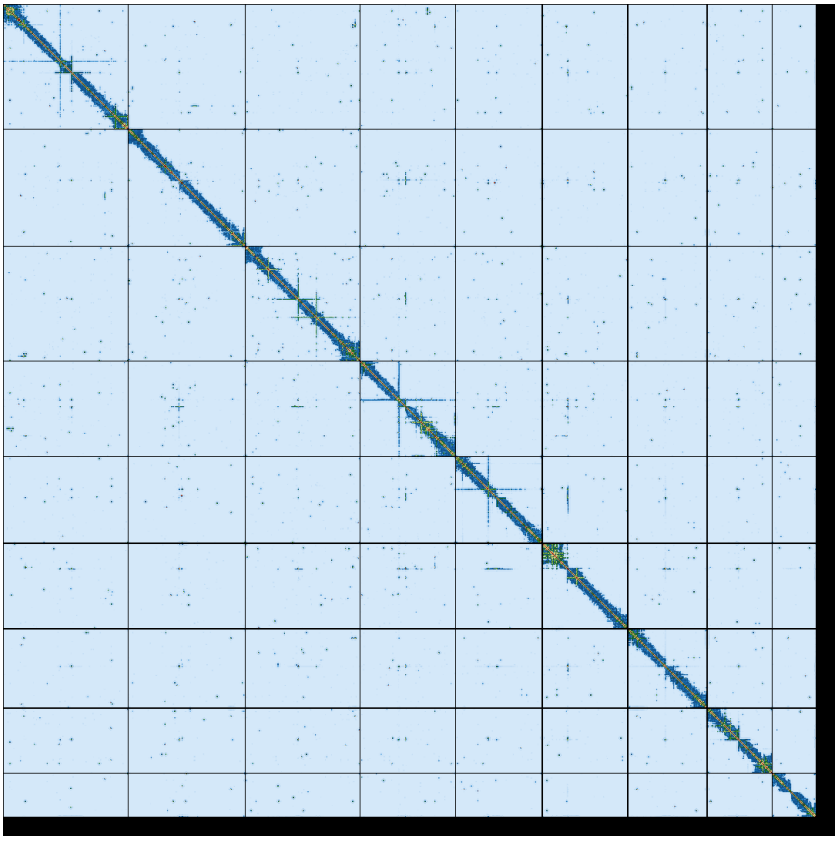
. Interventions/Gb: 170
. Contamination notes: ""
. Other observations: "The assembly of Clavelina sabbadini (kaClaSabb1) is based on 84.51X PacBio HiFi data from an amplified DNA library and 492.01X Arima Hi-C data generated as part of the ATLASea programme (https://www.atlasea.fr). The assembly process 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, 268 contigs were identified as contaminants (bacterial, archaeal, or viral), totaling 9.625 Mb (with the largest being 0.295 Mb). Additionally, 436 regions totaling 9.71 Mb (with the largest being 0.364 Mb) were identified as haplotypic duplications and removed. The mitochondrial genome was assembled using Oatk. Finally, the primary assembly was analyzed and manually improved using Pretext. During manual curation, 6 haplotypic regions were removed, totaling 13.127Mb (with the largest being 12.619Mb). The telomeric pattern AACCCT was identified with TelFinder and used to generate the Pretext Telomeres tracks. 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	170,750,497	157,640,194
GC %	36.58	36.63
Gaps/Gbp	298.68	342.55
Total gap bp	5,100	6,800
Scaffolds	145	143
Scaffold N50	21,066,226	18,095,097
Scaffold L50	4	4
Scaffold L90	8	8
Contigs	196	197
Contig N50	3,693,750	3,755,105
Contig L50	16	15
Contig L90	43	41
QV	62.1813	61.8343
Kmer compl.	85.1534	78.2586
BUSCO sing.	96.1%	88.1%
BUSCO dupl.	0.8%	0.5%
BUSCO frag.	1.1%	1.3%
BUSCO miss.	2.0%	10.1%

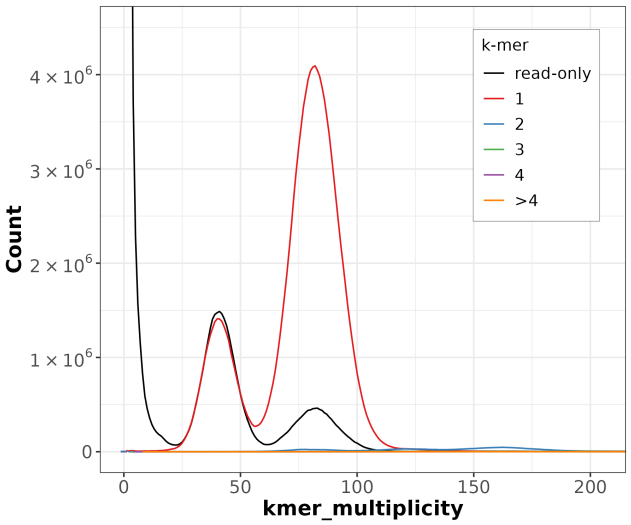
BUSCO: 6.1.0 (euk_genome_min, miniprot) / Lineage: metazoa_odb12.2 (genomes:182, BUSCOs:932)

HiC contact map of curated assembly

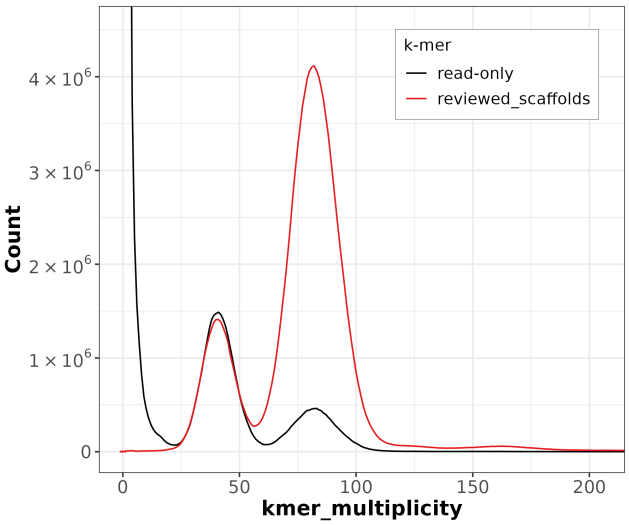


collapsed [LINK]

K-mer spectra of curated assembly

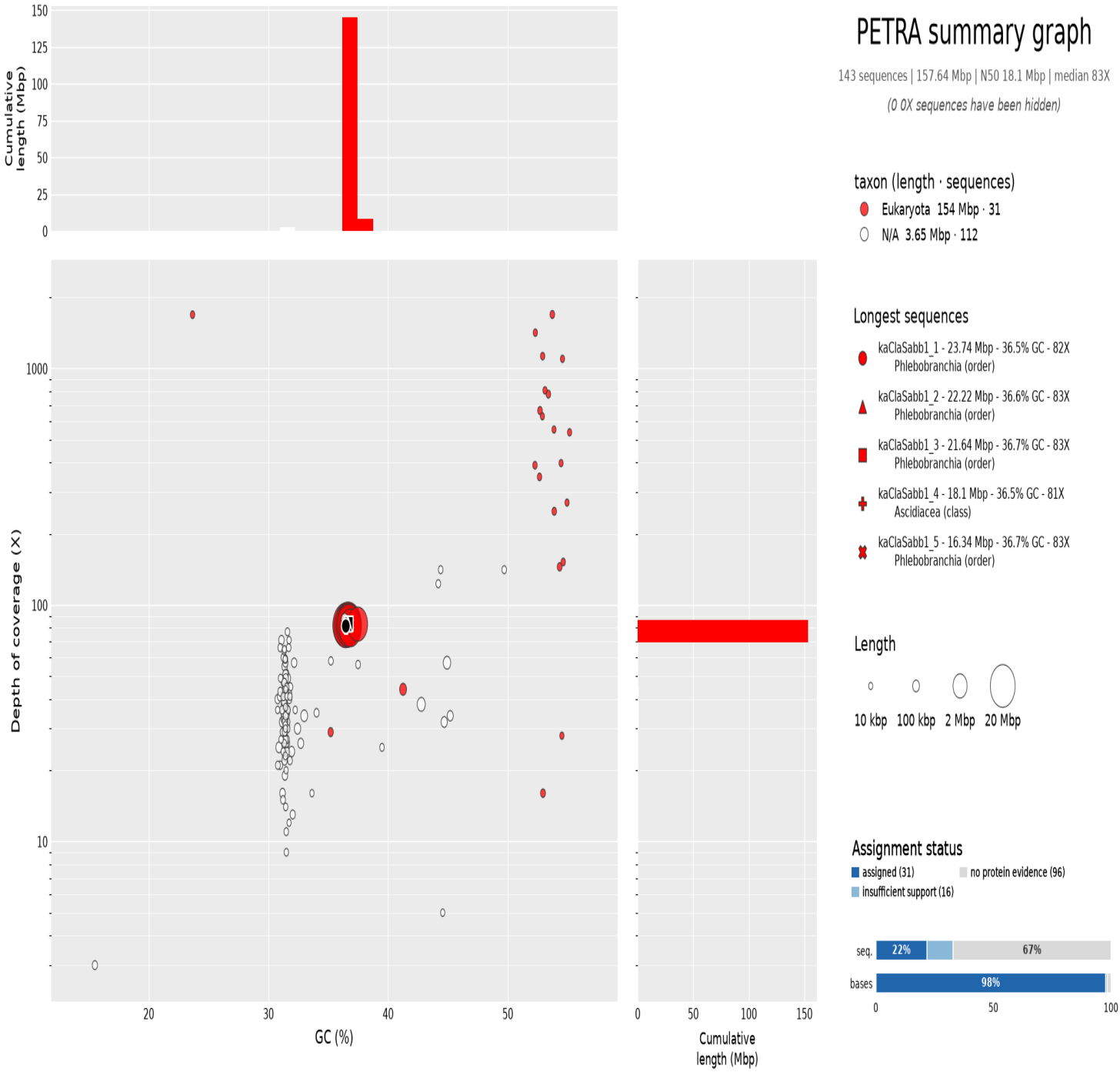


Distribution of k-mer counts per copy numbers found in asm



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

Post-curation contamination screening



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	Arima
Coverage	91	532

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:* 0.2.4
  |_ *key param:* NA
- **PretextView**
  |_ *ver:* 1.0.7
  |_ *key param:* NA

Submitter: Eleanore Lacoste
Affiliation: Genoscope

Date and time: 2026-10-29 18:52:08 CET