ERGA Assembly Report

v24.10.15

Tags: ATLASea[INVALID TAG]

TxID	2079803
ToLID	**kaCioEdwa1**
Species	Ciona edwardsi
Class	Ascidiacea
Order	Phlebobranchia

Genome Traits	Expected	Observed
Haploid size (bp)	160,656,425	157,111,833
Haploid Number	9 (source: ancestor)	14
Ploidy	2 (source: ancestor)	2
Sample Sex	Unknown	Unknown

EBP metrics summary and curation notes

Obtained EBP quality metric for collapsed: 6.6.Q48

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

. Observed Haploid Number is different from Expected

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

Curator notes

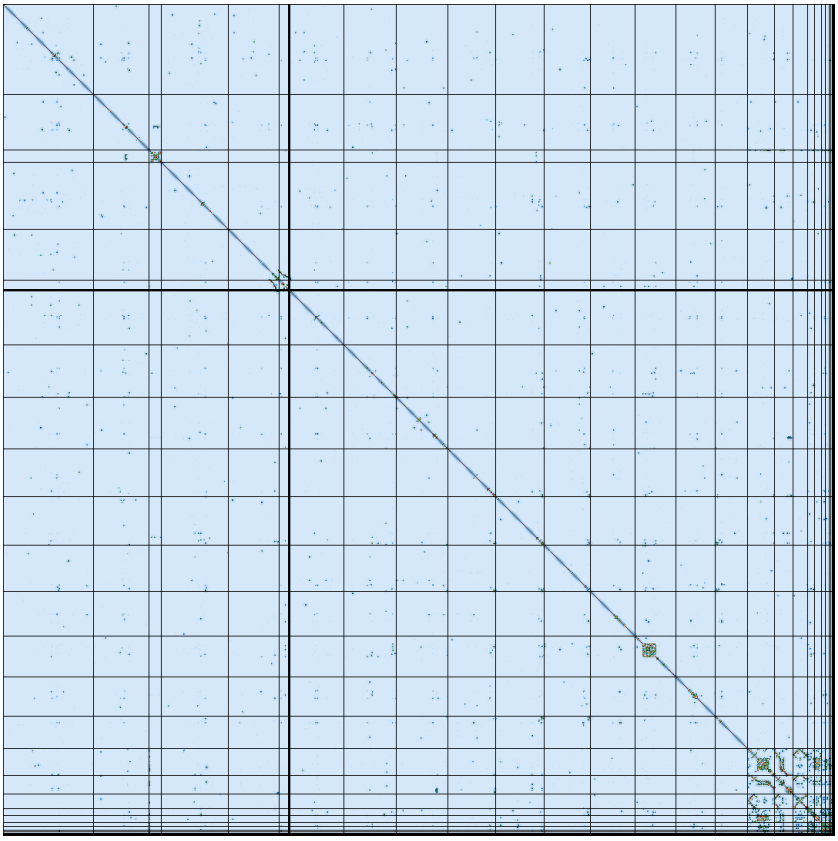
. Interventions/Gb: 6
. Contamination notes: ""
. Other observations: "The assembly of Ciona edwardsi (kaCioEdwa1) is based on 194.79X PacBio data and 171.56X Arima Hi-C data generated as part of the ATLASea programme (https://www.atlasea.fr). The assembly process included the following steps: initial 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, 99 contigs were identified as contaminants (bacterial, archaeal, or viral), totaling 6.857 Mb (with the largest being 2.935 Mb). Additionally, 260 regions totaling 14.529 Mb (with the largest being 1.177 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. The telomeric pattern AACCCCT 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	157,122,325	157,111,833
GC %	35.8	35.8
Gaps/Gbp	89.1	19.09
Total gap bp	1,400	300
Scaffolds	126	42
Scaffold N50	8,190,160	9,576,136
Scaffold L50	8	7
Scaffold L90	21	16
Contigs	140	45
Contig N50	6,554,000	9,105,670
Contig L50	10	7
Contig L90	30	17
QV	48.0885	48.0809
Kmer compl.	80.9117	80.9127
BUSCO sing.	95.5%	95.7%
BUSCO dupl.	1.2%	0.9%
BUSCO frag.	1.6%	1.6%
BUSCO miss.	1.6%	1.8%

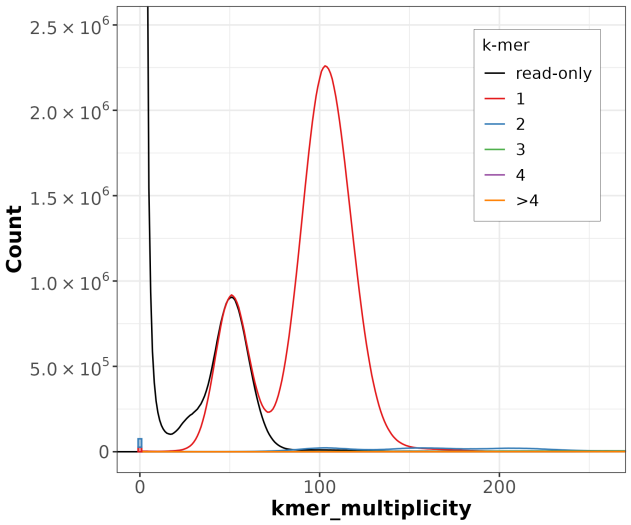
Warning! BUSCO versions or lineage datasets are not the same across results:
BUSCO: 6.0.0 (euk_genome_min, miniprot) / Lineage: metazoa_odb12 (genomes:206, BUSCOs:672)
BUSCO: 6.1.0 (euk_genome_min, miniprot) / Lineage: metazoa_odb12 (genomes:206, BUSCOs:672)

HiC contact map of curated assembly

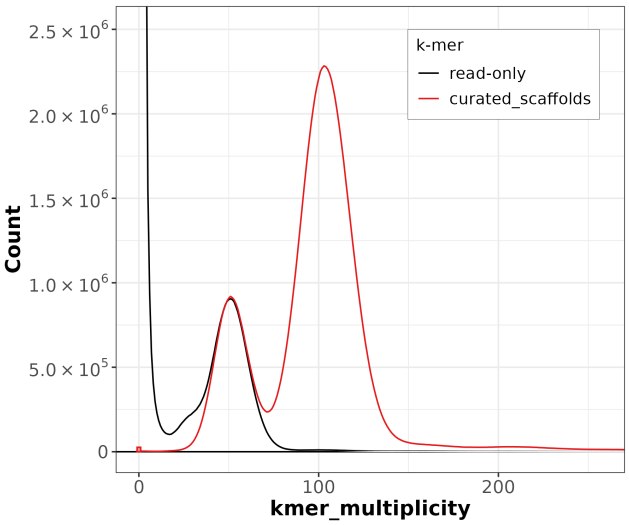


collapsed [LINK]

K-mer spectra of curated assembly

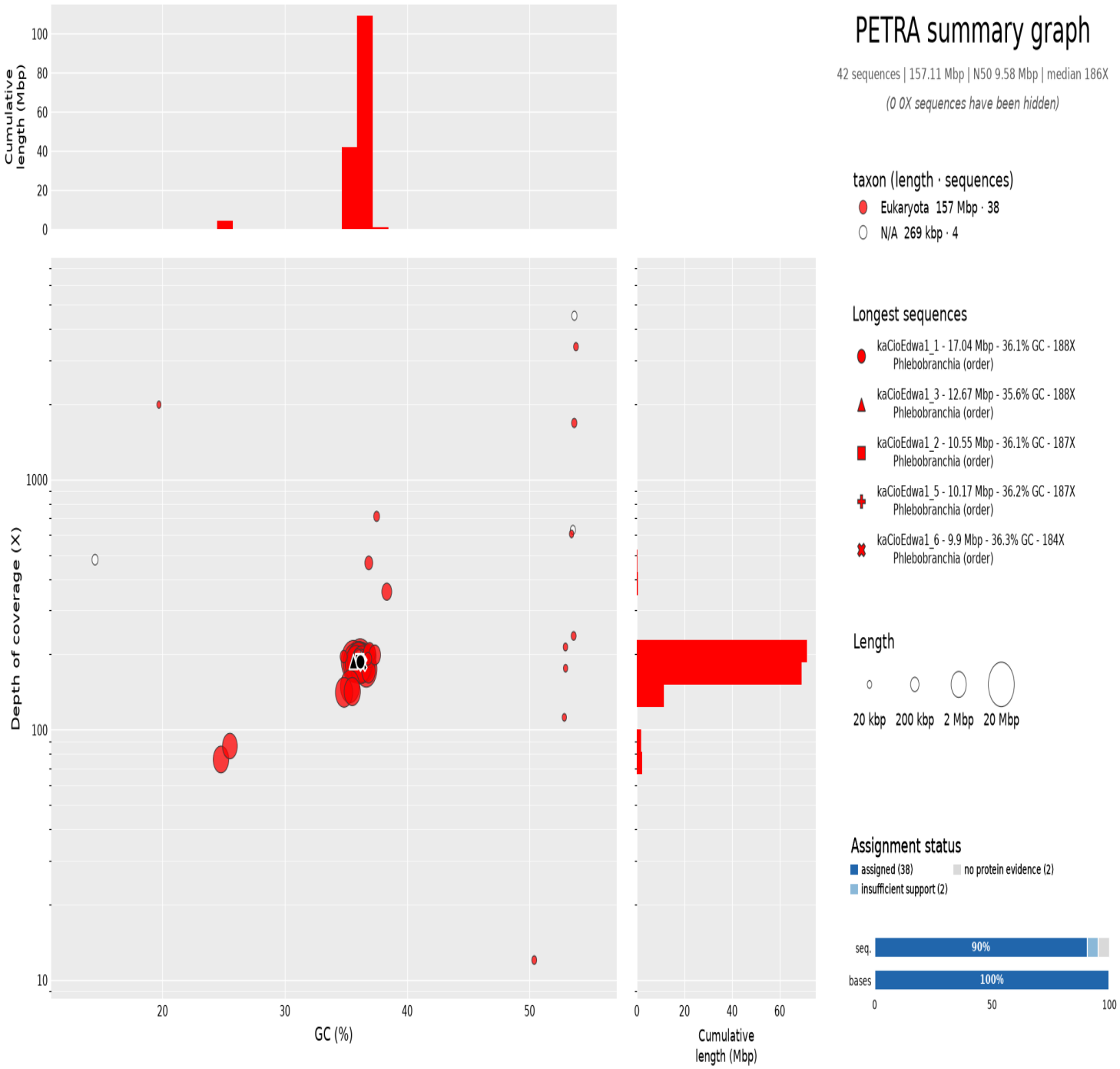


Distribution of k-mer counts per copy numbers found in asm



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

Post-curation contamination screening



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	194	171

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: Benjamin Istace
Affiliation: Genoscope

Date and time: 2026-10-07 13:02:33 CEST