ERGA Assembly Report

v24.10.15

Tags: ATLASea[INVALID TAG]

TxID	257330
ToLID	**xgConJuda1**
Species	Conus judaeus
Class	Gastropoda
Order	Neogastropoda

Genome Traits	Expected	Observed
Haploid size (bp)	2,268,797,437	2,742,778,799
Haploid Number	14 (source: ancestor)	35
Ploidy	2 (source: ancestor)	2
Sample Sex	Unknown	Unknown

EBP metrics summary and curation notes

Obtained EBP quality metric for collapsed: 6.7.Q40

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

. Observed Haploid size (bp) has >20% difference with Expected
. Observed Haploid Number is different from Expected

. Kmer completeness value is less than 90 for collapsed
. BUSCO single copy value is less than 90% for collapsed
. BUSCO duplicated value is more than 5% for collapsed
. Assembly length loss > 3% for collapsed

Curator notes

. Interventions/Gb: 96
. Contamination notes: ""
. Other observations: "The assembly of Conus judaeus (xgConJuda1) is based on 38X ONT data and 136X Arima Hi-C data generated as part of the ATLASea programme (https://www.atlasea.fr). The assembly process included the following steps: initial ONT 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, 25 contigs were identified as contaminants (bacterial, archaeal, or viral), totaling 1.05 Mb (with the largest being 0.108 Mb). Additionally, 3240 regions totaling 244.759 Mb (with the largest being 0.443 Mb) were identified as haplotypic duplications and removed. Mitochondrial genome was assembled using ptGAUL. Finally, the primary assembly was analyzed and manually improved using Pretext. During manual curation, 61 haplotypic regions and 38 contaminant sequences were removed, totaling 153.541Mb and 2.174Mb respectively (with the largest being 78.256Mb

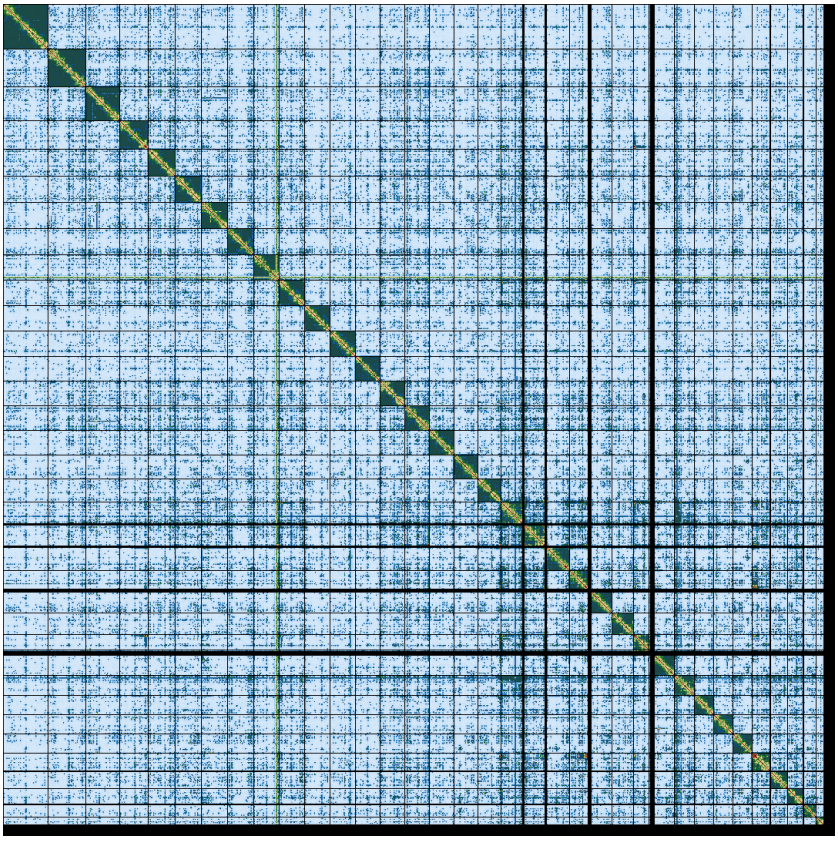
and 0.351Mb). The telomeric pattern TTAGGG was identified with TelFinder and used to generate the Pretext Telomeres tracks. Chromosome-scale scaffolds confirmed by Hi-C data were named in order of size. "

Quality metrics table

Metrics	Pre-curation collapsed	Curated collapsed
Total bp	2,898,484,573	2,742,778,799
GC %	44.17	44.18
Gaps/Gbp	895.3	905.29
Total gap bp	259,500	261,400
Scaffolds	456	346
Scaffold N50	83,354,301	80,472,327
Scaffold L50	14	15
Scaffold L90	32	31
Contigs	3,051	2,829
Contig N50	2,086,645	2,096,137
Contig L50	401	374
Contig L90	1,507	1,400
QV	40.1033	40.0982
Kmer compl.	72.3845	68.9771
BUSCO sing.	78.3%	82.9%
BUSCO dupl.	19.5%	12.1%
BUSCO frag.	0.9%	0.8%
BUSCO miss.	1.3%	4.1%

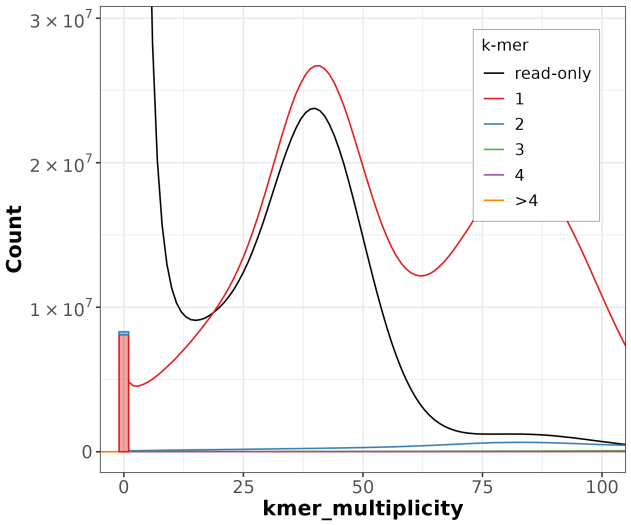
Warning! BUSCO versions or lineage datasets are not the same across results:
BUSCO: 6.0.0 (euk_genome_min, miniprot) / Lineage: mollusca_odb12 (genomes:36, BUSCOs:4421)
BUSCO: 6.1.0 (euk_genome_min, miniprot) / Lineage: mollusca_odb12 (genomes:36, BUSCOs:4421)

HiC contact map of curated assembly

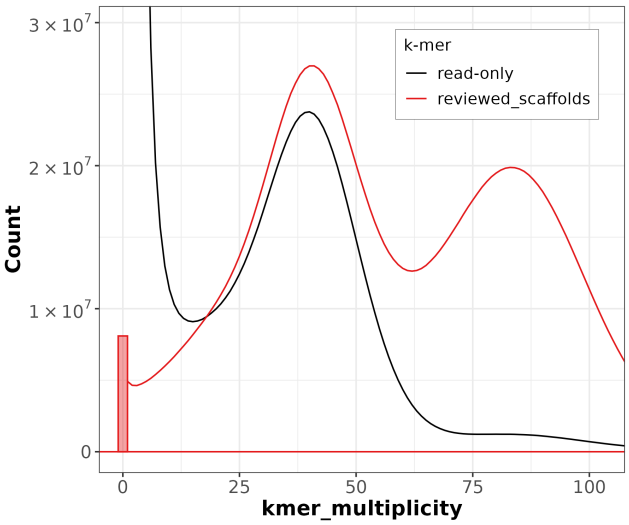


collapsed [LINK]

K-mer spectra of curated assembly

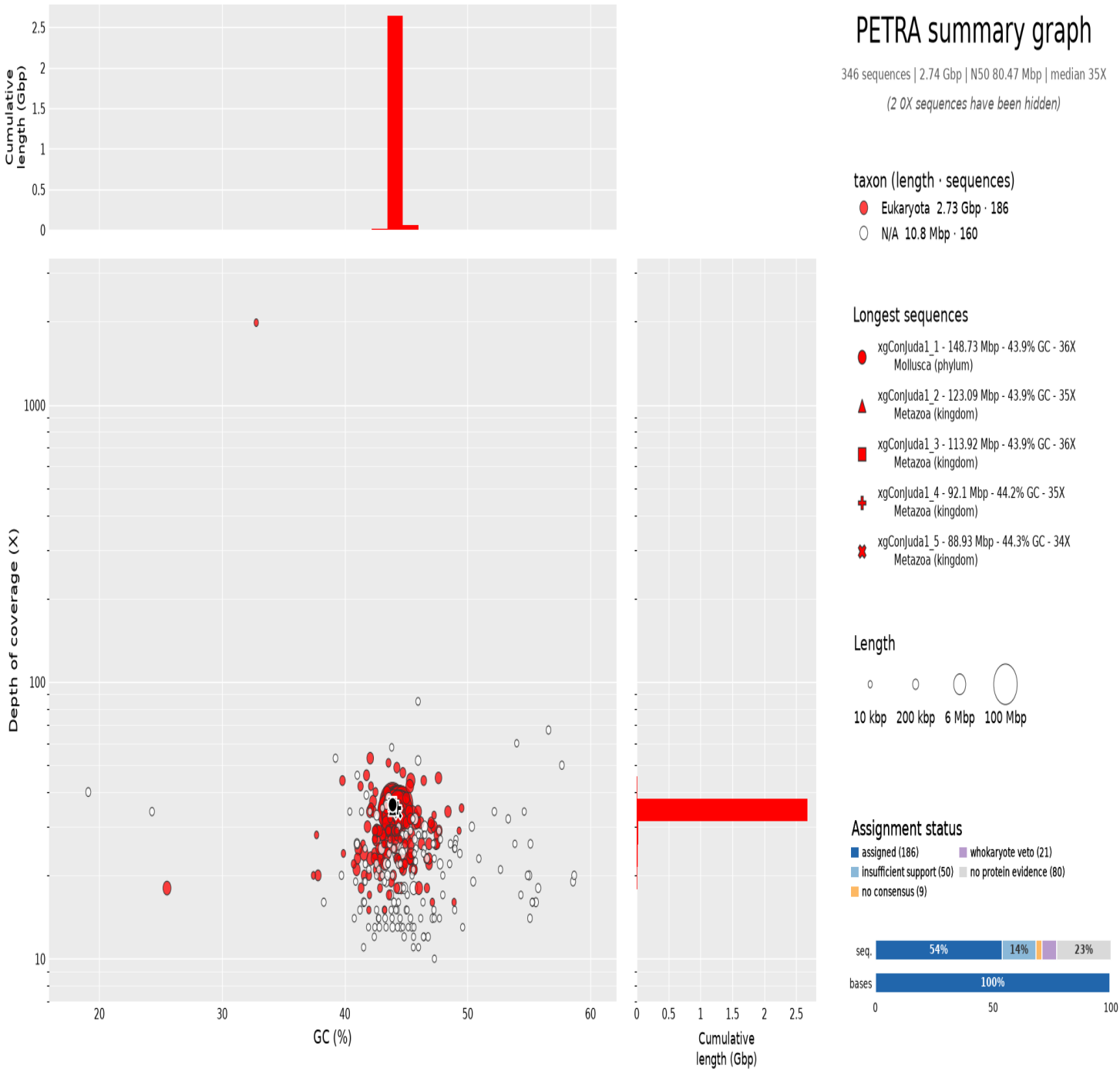


Distribution of k-mer counts per copy numbers found in asm



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	40	143

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: Sophie Layac
Affiliation: Genoscope

Date and time: 2026-10-08 06:28:30 CEST