

ERGA Assembly Report

v24.10.15

Tags: ATLASEa[INVALID TAG]

TxID	120438
ToLID	xpLepCaje1
Species	Lepidopleurus cajetanus
Class	Polyplacophora
Order	Lepidopleurida

Genome Traits	Expected	Observed
Haploid size (bp)	1,401,909	1,739,191,859
Haploid Number	9 (source: ancestor)	12
Ploidy	2 (source: ancestor)	2
Sample Sex	Unknown	Unknown

EBP metrics summary and curation notes

Obtained EBP quality metric for collapsed: 7.8.Q15

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
- . QV value is less than 40 for collapsed

Curator notes

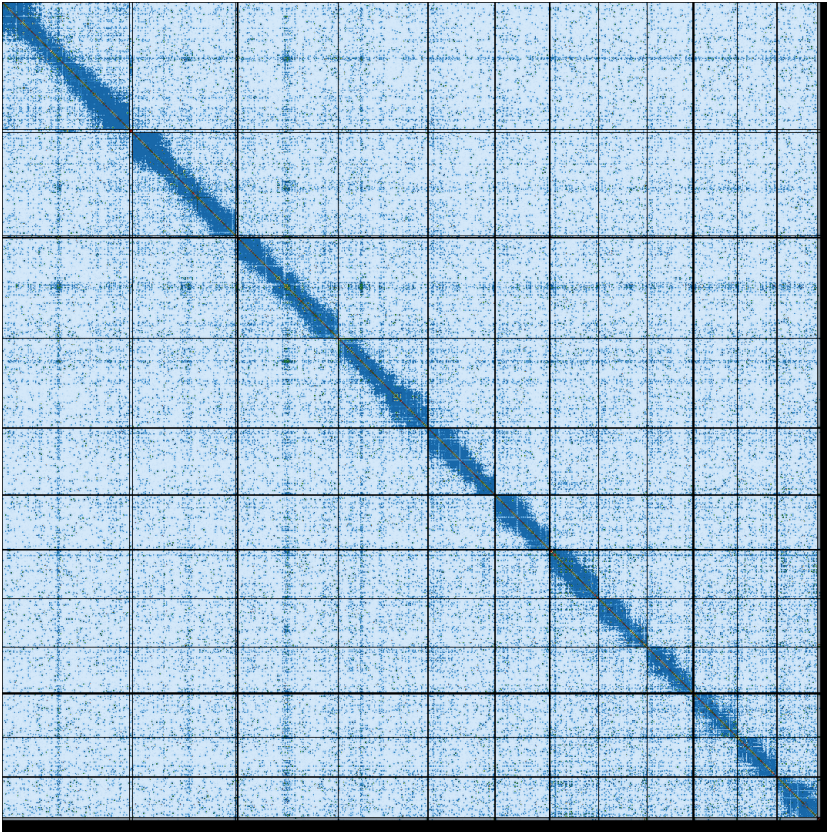
- . Interventions/Gb: 43
- . Contamination notes: ""
- . Other observations: "The assembly of *Lepidopleurus cajetanus* (xpLepCaje1) is based on 34.7X PacBio data and 151.24X Arima Hi-C data generated as part of the ATLASEa programme (<https://www.atlasea.fr>). The assembly process included the following 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, 112 contigs were identified as contaminants (bacterial, archaeal, or viral), totaling 9.627 Mb (with the largest being 1.177 Mb). Additionally, 295 regions totaling 150.579 Mb (with the largest being 22.71 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. During manual curation, 30 haplotypic regions and 1 contaminant sequence were removed, totaling 44.957Mb and 0.097Mb, respectively (with the largest being 6.75Mb and 0.097Mb). 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	1,787,672,871	1,739,191,859
GC %	42.49	42.49
Gaps/Gbp	42.51	45.42
Total gap bp	7,600	11,100
Scaffolds	170	133
Scaffold N50	138,465,912	185,987,349
Scaffold L50	5	4
Scaffold L90	11	11
Contigs	246	212
Contig N50	40,522,000	40,488,162
Contig L50	13	13
Contig L90	44	40
QV	15.8869	15.8697
Kmer compl.	99.6708	99.6392
BUSCO sing.	94.8%	96.7%
BUSCO dupl.	3.1%	1.1%
BUSCO frag.	0.4%	0.4%
BUSCO miss.	1.8%	1.8%

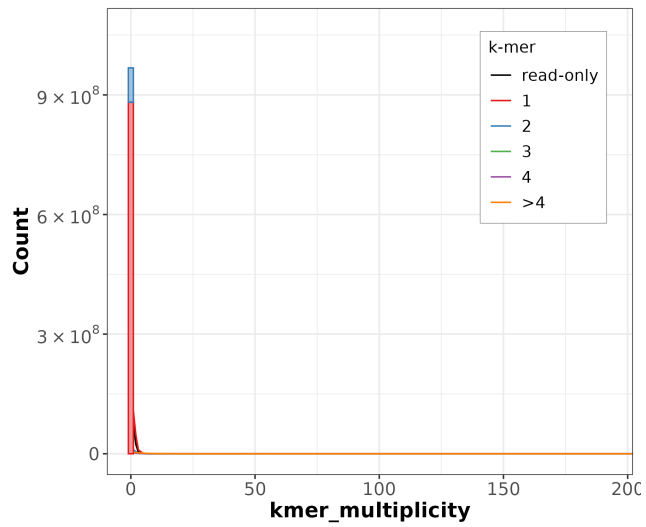
BUSCO: 6.1.0 (euk_genome_min, miniprot) / Lineage: mollusca_odb12.2 (genomes:31, BUSCOs:4236)

HiC contact map of curated assembly

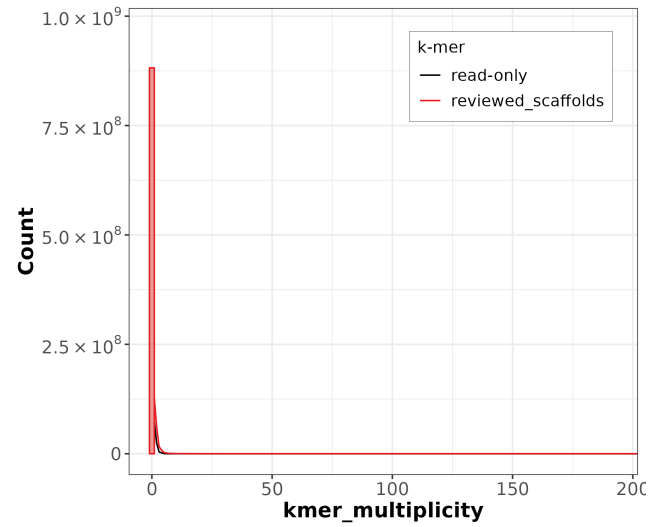


collapsed [\[LINK\]](#)

K-mer spectra of curated assembly

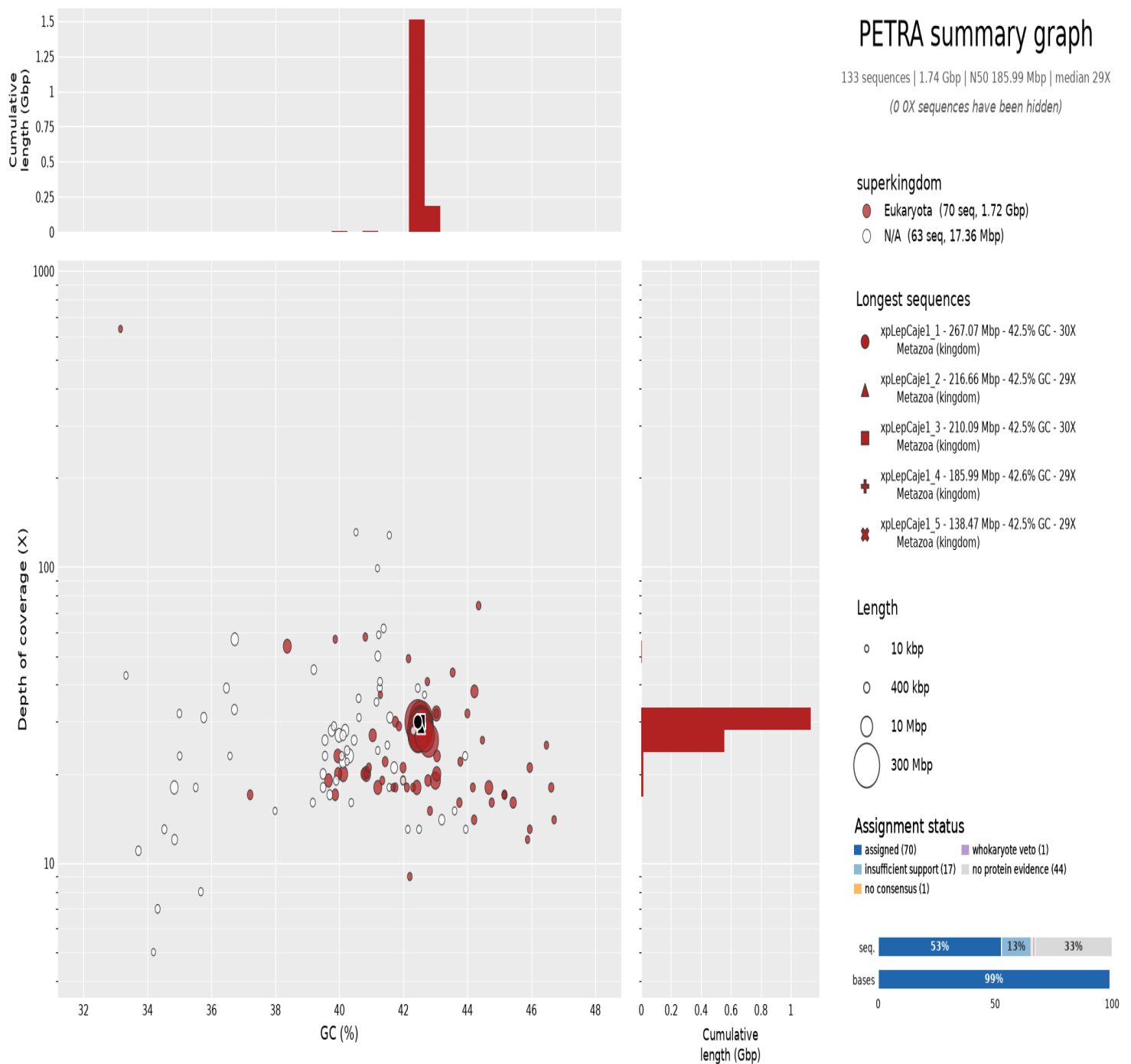


Distribution of k-mer counts per copy numbers found in asm



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	35	155

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: Emilie Teodori

Affiliation: Genoscope

Date and time: 2026-09-10 23:03:34 CEST