ERGA Assembly Report

v24.10.15

Tags: ATLASea[INVALID TAG]

TxID	380120
ToLID	**xbVenCasi1**
Species	Venus casina
Class	Bivalvia
Order	Venerida

Genome Traits	Expected	Observed
Haploid size (bp)	2,056,814,378	2,088,199,216
Haploid Number	19 (source: ancestor)	19
Ploidy	2 (source: ancestor)	2
Sample Sex	Unknown	Unknown

EBP metrics summary and curation notes

Obtained EBP quality metric for collapsed: 7.8.Q50

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

. Kmer completeness value is less than 90 for collapsed

Curator notes

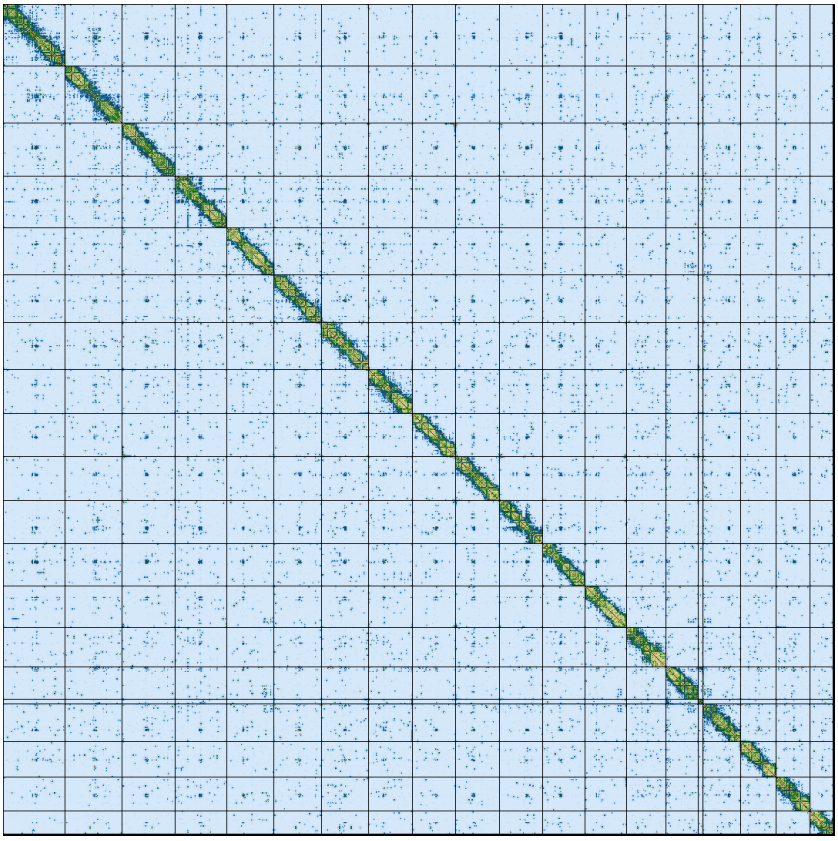
. Interventions/Gb: 2
. Contamination notes: ""
. Other observations: "The assembly of Venus casina (xbVenCasi1) is based on 40X PacBio data and 153X Arima Hi-C data generated as part of the ATLASea programme (https://www.atlasea.fr). The assembly process included the following steps: initial PacBio 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, 32 contigs were identified as contaminants (bacterial, archaeal, or viral), totaling 1.652 Mb (with the largest being 0.139 Mb). Additionally, 214 regions totaling 25.613 Mb (with the largest being 6.346 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, 4 haplotypic regions were removed, totaling 5.832Mb (with the largest being 2.934Mb). 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	2,094,051,565	2,088,199,216
GC %	35.22	35.22
Gaps/Gbp	10.98	11.01
Total gap bp	2,300	2,500
Scaffolds	52	48
Scaffold N50	108,900,207	108,900,207
Scaffold L50	9	9
Scaffold L90	17	17
Contigs	75	71
Contig N50	85,821,000	85,821,000
Contig L50	10	10
Contig L90	26	25
QV	50.6935	50.6912
Kmer compl.	67.4801	67.4121
BUSCO sing.	97.8%	96.6%
BUSCO dupl.	1.7%	1.3%
BUSCO frag.	0.4%	1.6%
BUSCO miss.	0.2%	0.5%

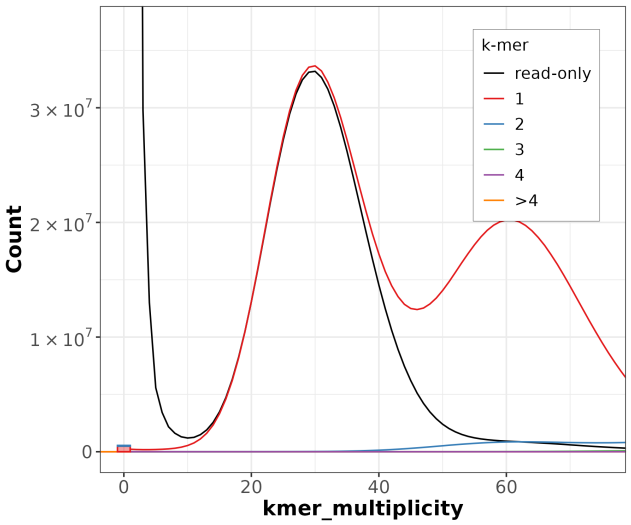
Warning! BUSCO versions or lineage datasets are not the same across results:
BUSCO: 6.0.0 (euk_genome_min, miniprot) / Lineage: mollusca_odb12 (genomes:36, BUSCOs:4421)
BUSCO: 5.8.2 (euk_genome_min, miniprot) / Lineage: mollusca_odb12 (genomes:36, BUSCOs:4421)

HiC contact map of curated assembly

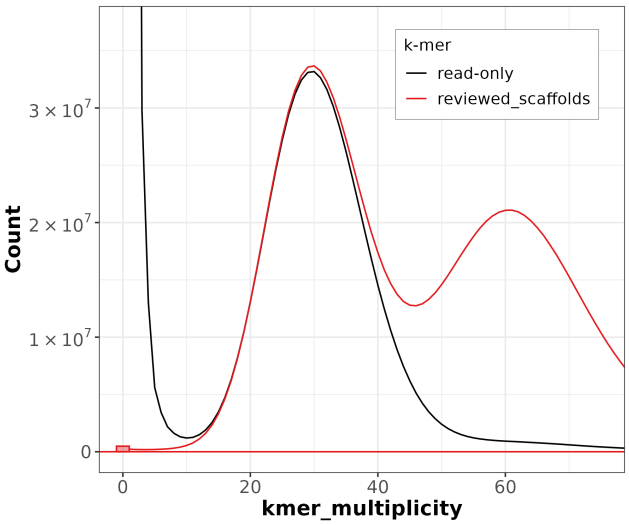


collapsed [LINK]

K-mer spectra of curated assembly

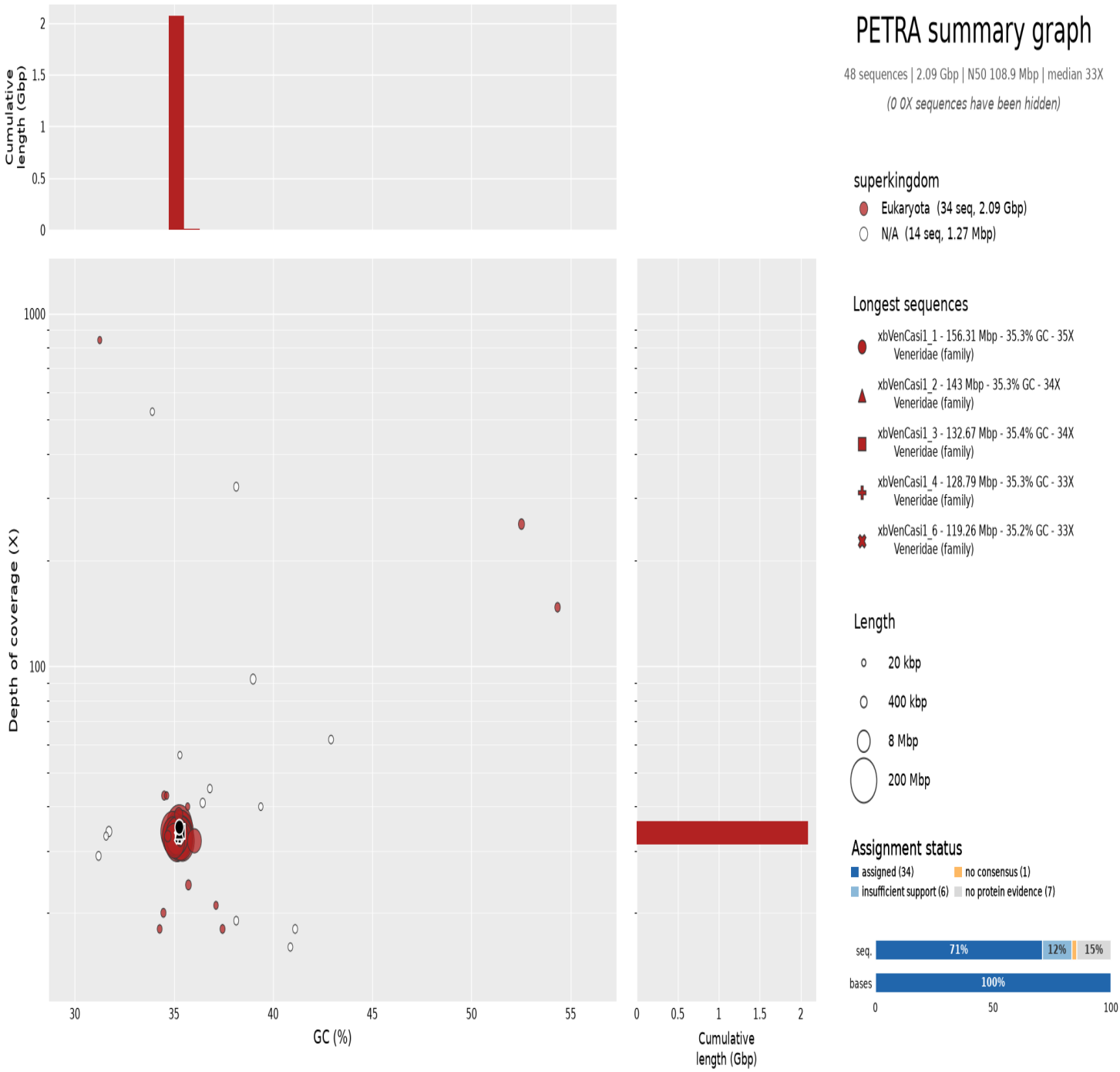


Distribution of k-mer counts per copy numbers found in asm



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	39	153

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: Mohamed Meftah Saoues
Affiliation: Genoscope

Date and time: 2026-09-29 07:25:41 CEST