

ERGA Assembly Report

v24.10.15

Tags: ATLASEa[INVALID TAG]

TxID	934979
ToLID	xgPruPsel1
Species	<i>Pruvotfolia pselliotes</i>
Class	Gastropoda
Order	Nudibranchia

Genome Traits	Expected	Observed
Haploid size (bp)	1,093,452,623	1,103,564,245
Haploid Number	11 (source: ancestor)	14
Ploidy	2 (source: ancestor)	2
Sample Sex	Unknown	Unknown

EBP metrics summary and curation notes

Obtained EBP quality metric for collapsed: 7.7.Q45

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

Curator notes

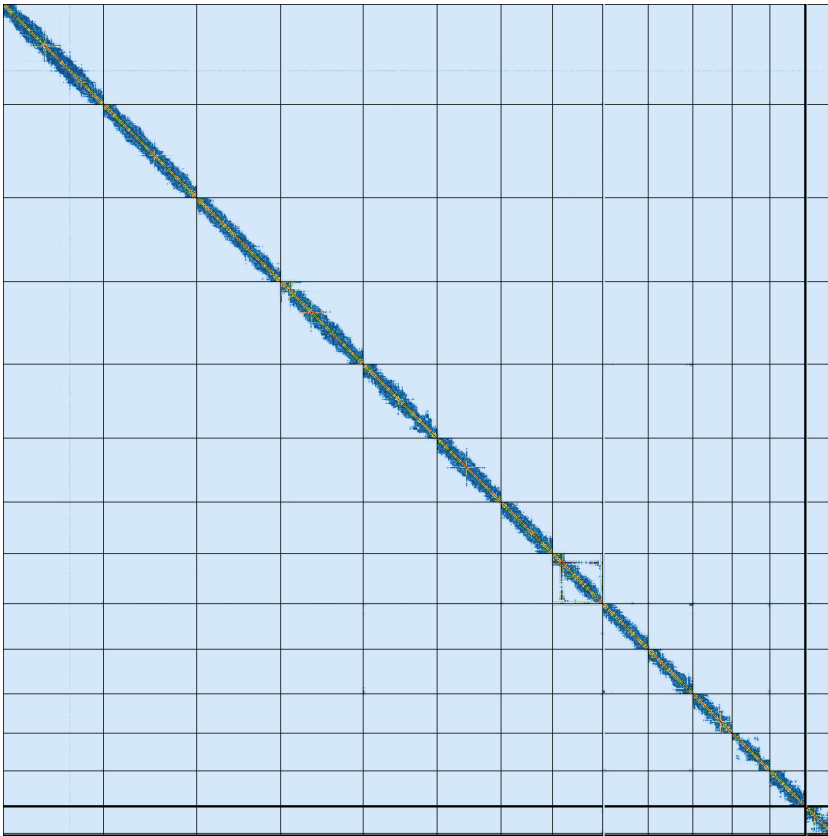
. Interventions/Gb: 30
. Contamination notes: ""
. Other observations: "The assembly of '*Pruvotfolia pselliotes*' (xgPruPsel1) is based on 81X ONT data and 137X Arima Hi-C data generated as part of the ATLASEa programme (<https://www.atlasea.fr>). The assembly process included the following steps: initial 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, 57 contigs were identified as contaminants (bacterial, archaeal, or viral), totaling 6.116 Mb (with the largest being 3.124 Mb). Additionally, 141 regions totaling 83.798 Mb (with the largest being 12.588 Mb) were identified as haplotypic duplications and removed. Mitochondrial genome was assembled using ptGAUL. During manual curation, 12 haplotypic regions were removed, totaling 31.008Mb (with the largest being 12.754Mb). 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	1,134,569,379	1,103,564,245
GC %	37.87	37.88
Gaps/Gbp	37.9	45.31
Total gap bp	4,300	6,700
Scaffolds	73	60
Scaffold N50	95,813,200	98,649,140
Scaffold L50	5	5
Scaffold L90	12	12
Contigs	116	110
Contig N50	30,154,000	31,238,996
Contig L50	14	13
Contig L90	38	37
QV	45.3621	45.3534
Kmer compl.	71.7604	70.7364
BUSCO sing.	94.6%	95.7%
BUSCO dupl.	2.6%	1.2%
BUSCO frag.	0.5%	0.5%
BUSCO miss.	2.4%	2.6%

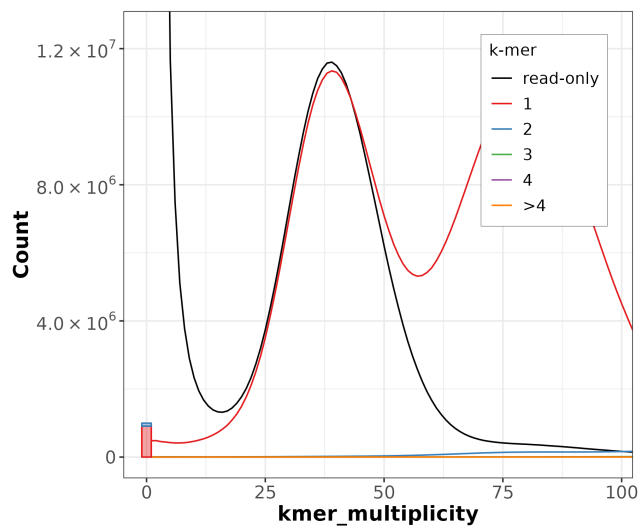
BUSCO: 6.0.0 (euk_genome_min, miniprot) / Lineage: mollusca_odb12 (genomes:36, BUSCOs:4421)

HiC contact map of curated assembly

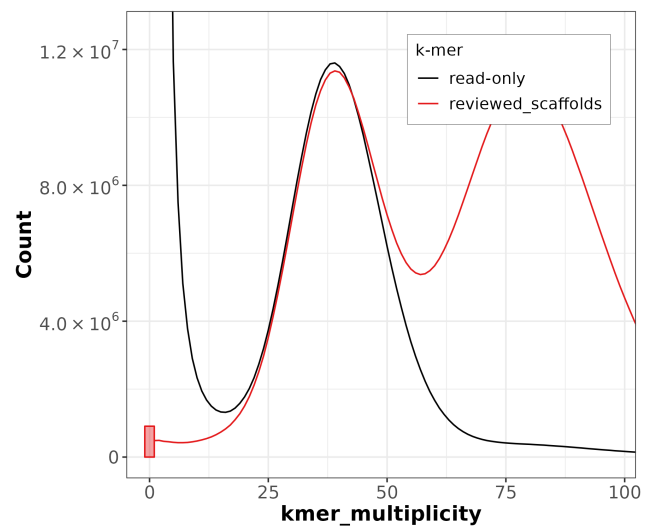


collapsed [\[LINK\]](#)

K-mer spectra of curated assembly

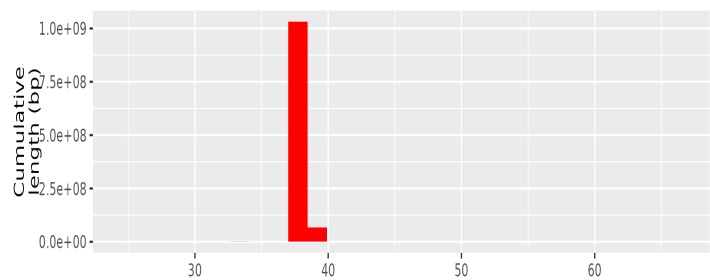


Distribution of k-mer counts per copy numbers found in asm



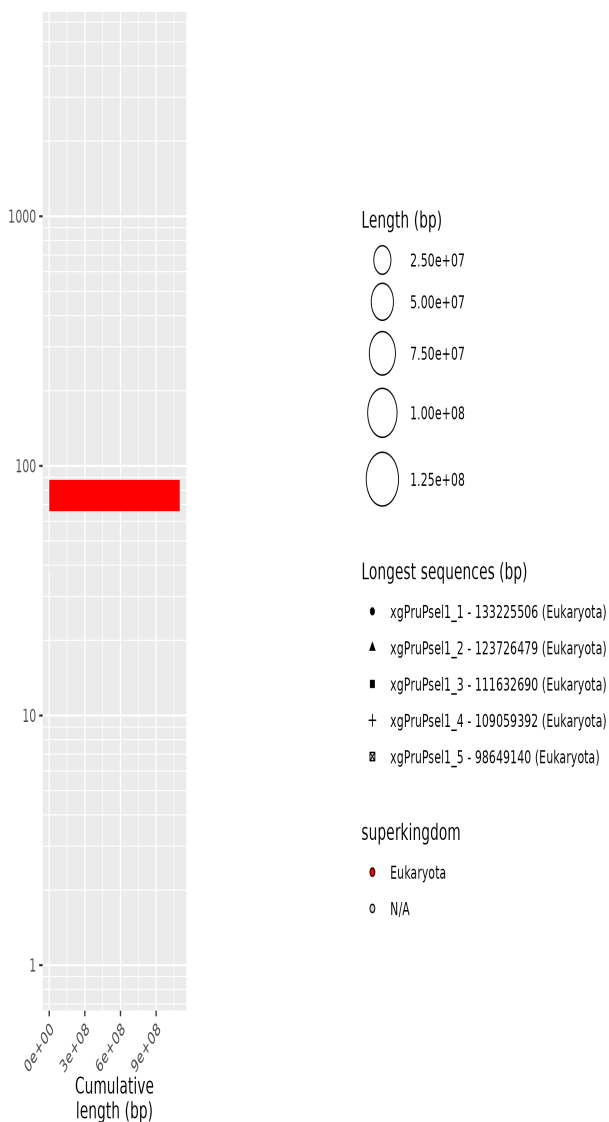
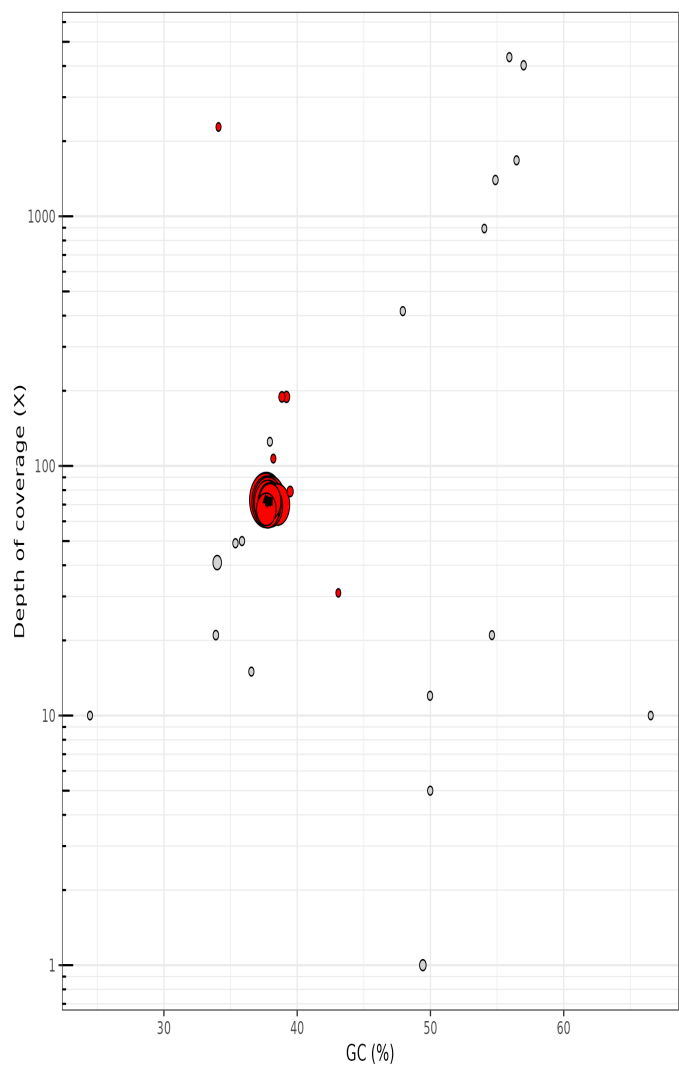
Distribution of k-mer counts coloured by their presence in reads/assemblies

Post-curation contamination screening



TAPAs summary Graph

(22 0X contigs have been hidden)



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	83	142

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*: 0.1.9
 - |_ *key param*: NA
- **PretextView**
 - |_ *ver*: 0.2.5
 - |_ *key param*: NA

Submitter: Simone Duprat

Affiliation: Genoscope

Date and time: 2026-02-04 05:03:57 CET