

ERGA Assembly Report

v24.10.15

Tags: ATLASEa[INVALID TAG]

TxID	1076719
ToLID	xg0xyMacu1
Species	Oxymeris maculata
Class	Gastropoda
Order	Neogastropoda

Genome Traits	Expected	Observed
Haploid size (bp)	1,971,393,780	2,122,731,883
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.Q44

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
- . BUSCO single copy value is less than 90% for collapsed
- . BUSCO duplicated value is more than 5% for collapsed

Curator notes

- . Interventions/Gb: 147
- . Contamination notes: ""
- . Other observations: "The assembly of Oxymeris maculata (xg0xyMacu1) is based on 45X ONT data and 166X 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 Hifiiasm, removal of contaminant sequences using Context, removal of haplotypic duplications using purge_dups, and Hi-C-based scaffolding with YaHS. In total, 24 contigs were identified as contaminants (bacterial, archaeal, or viral), totaling 4.416 Mb (with the largest being 1.912 Mb). Additionally, 696 regions totaling 118.011 Mb (with the largest being 1.39 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, 83 haplotypic regions were removed, totaling 32.662Mb (with the largest being 2.861Mb). Note that chromosome 6 contains substantially more assembly gaps than the other chromosomes and exhibits an unusual coverage pattern, with approximately half coverage across the 0-25 Mb and 50-76 Mb regions, separated

by a region of normal coverage between 25 and 50 Mb. This pattern is consistent with a possible fusion of differentiated sex chromosomes. In the absence of supporting evidence from related species or a well-defined breakpoint, the scaffold was retained as assembled. Users should therefore interpret the structure of this chromosome with caution. Chromosome-scale scaffolds confirmed by Hi-C data were named in order of size. "

Quality metrics table

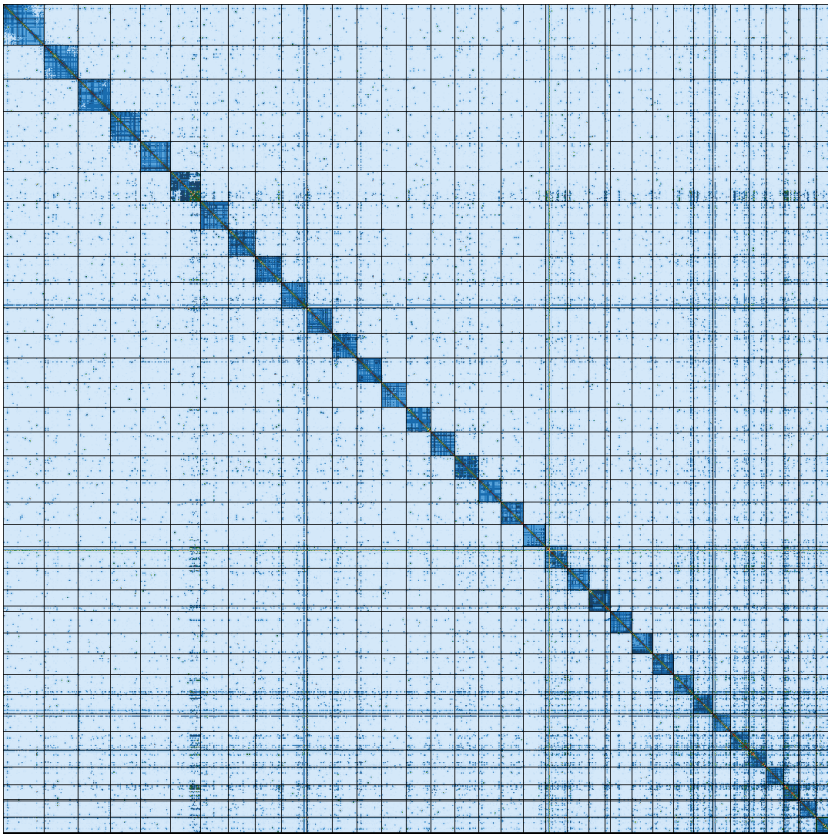
Metrics	Pre-curation collapsed	Curated collapsed
Total bp	2,155,363,006	2,122,731,883
GC %	40.67	40.67
Gaps/Gbp	219.45	253.92
Total gap bp	47,300	69,500
Scaffolds	214	148
Scaffold N50	62,672,996	61,915,275
Scaffold L50	15	15
Scaffold L90	31	31
Contigs	687	687
Contig N50	7,565,517	7,256,292
Contig L50	81	82
Contig L90	308	309
QV	44.0896	44.0927
Kmer compl.	72.4683	71.9697
BUSCO sing.	79.1%	78.9%
BUSCO dupl.	18.6%	16.7%
BUSCO frag.	0.8%	2.8%
BUSCO miss.	1.5%	1.7%

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)

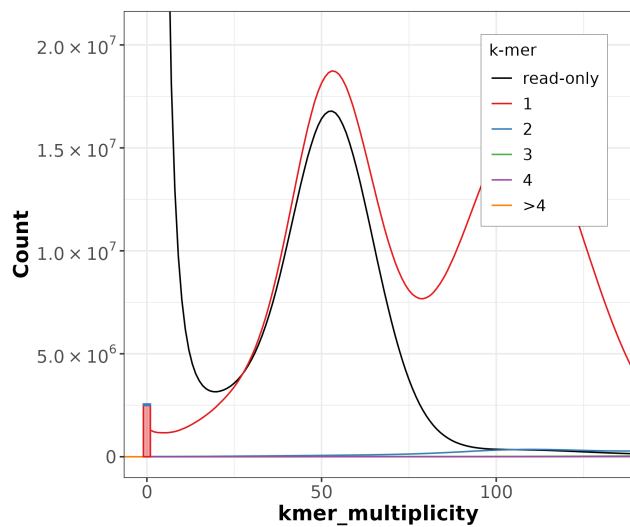
BUSCO: 5.8.2 (euk_genome_min, miniprot) / Lineage: mollusca_odb12 (genomes:36, BUSCOs:4421)

HiC contact map of curated assembly

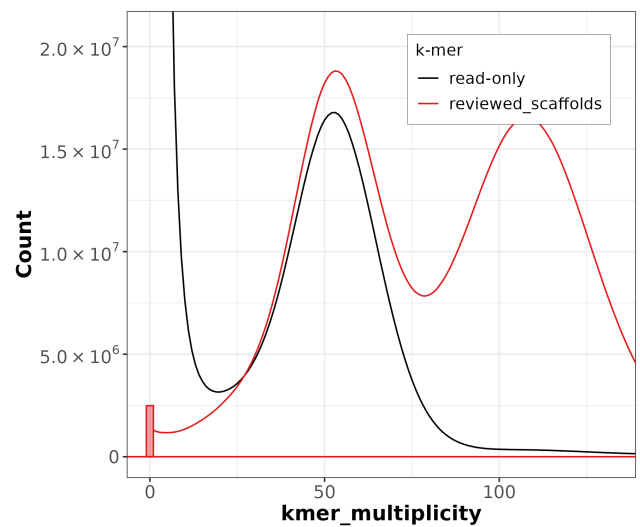


collapsed [\[LINK\]](#)

K-mer spectra of curated assembly

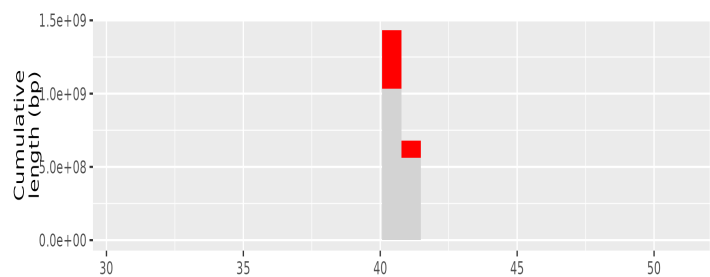


Distribution of k-mer counts per copy numbers found in asm

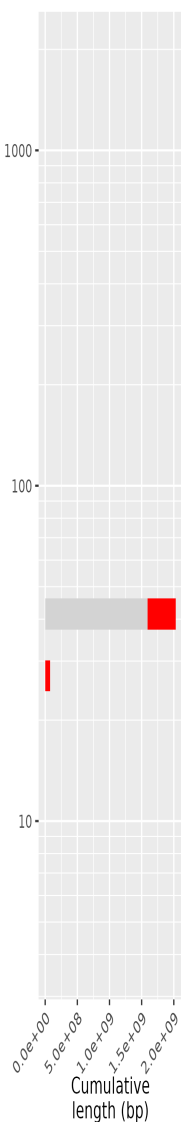
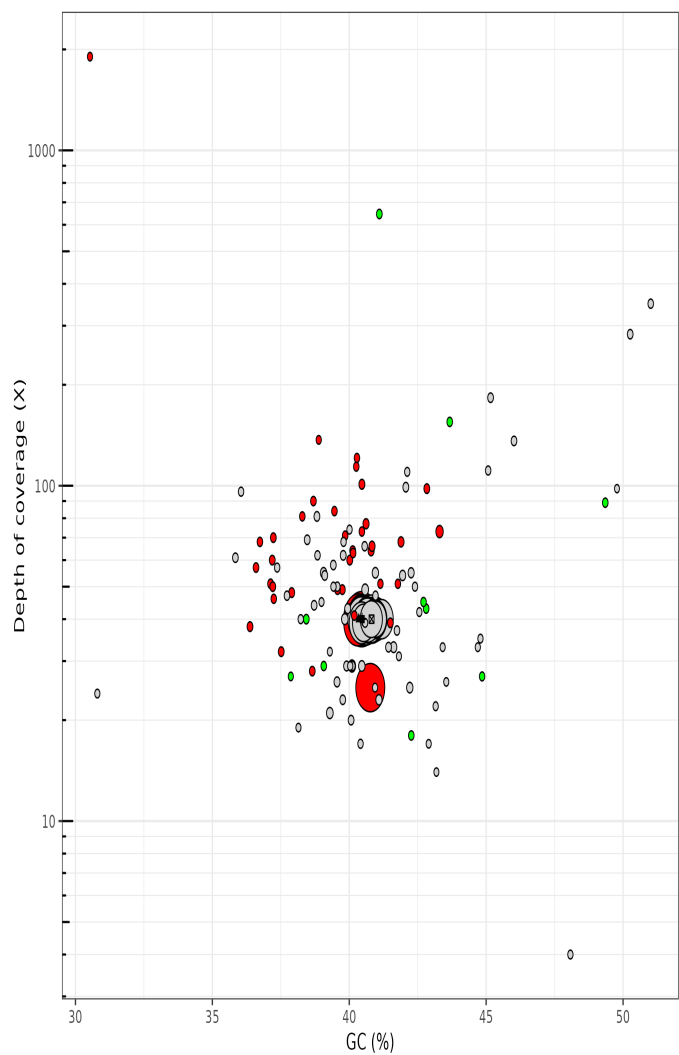


Distribution of k-mer counts coloured by their presence in reads/assemblies

Post-curation contamination screening



TAPAs summary Graph



- superkingdom
- Bacteria
 - Eukaryota
 - N/A
- Length (bp)
- 2.5e+07
 - 5.0e+07
 - 7.5e+07
 - 1.0e+08
- Longest sequences (bp)
- xgOxyMacu1_1 - 104839367 (Eukaryota)
 - ▲ xgOxyMacu1_2 - 85957703 (Eukaryota)
 - xgOxyMacu1_3 - 83038319 (N/A)
 - + xgOxyMacu1_5 - 76667732 (N/A)
 - xgOxyMacu1_4 - 76473808 (N/A)

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	45	168

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: Jean-Marc Aury

Affiliation: Genoscope

Date and time: 2026-08-06 21:59:34 CEST