

ERGA Assembly Report

v24.10.15

Tags: ATLASEa[INVALID TAG]

TxID	622441
ToLID	xgChiRamo1
Species	Chicoreus ramosus
Class	Gastropoda
Order	Neogastropoda

Genome Traits	Expected	Observed
Haploid size (bp)	1,660,931,552	1,827,947,987
Haploid Number	12 (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.Q41

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
- . More than 1000 gaps/Gbp for collapsed

Curator notes

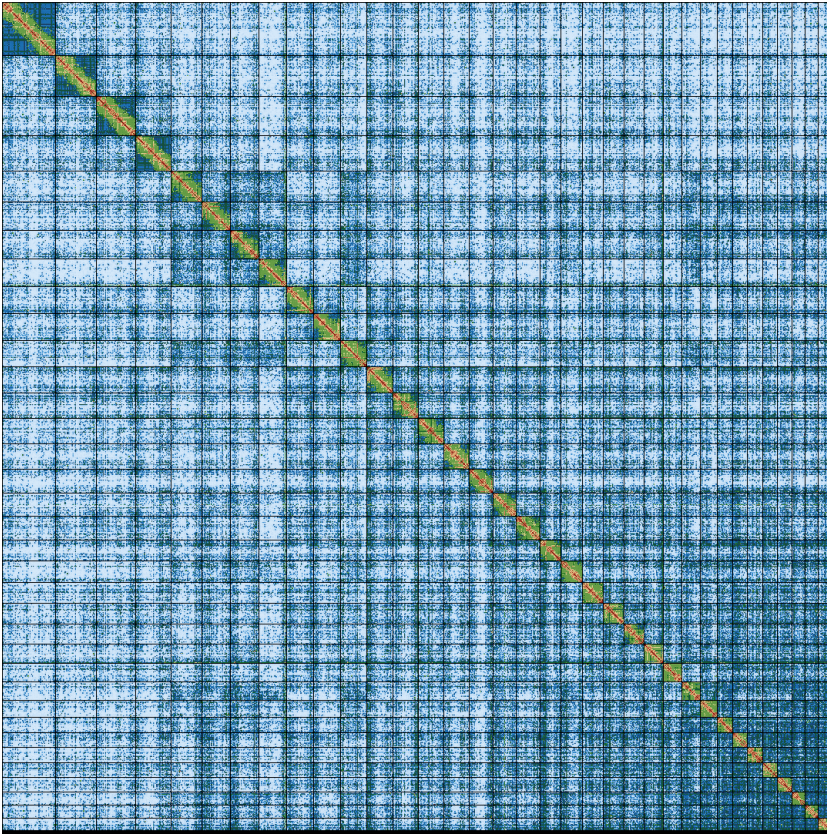
- . Interventions/Gb: 163
- . Contamination notes: ""
- . Other observations: "The assembly of *Chicoreus ramosus* (xgChiRamo1) is based on 27X ONT data and 138X Arima Hi-C data generated as part of the ATLASEa programme (<https://www.atlasea.fr>). The assembly process included the following steps: initial 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, 52 contigs were identified as contaminants (bacterial, archaeal, or viral), totaling 1.847 Mb (with the largest being 0.207 Mb). Additionally, 3021 regions totaling 151.207 Mb (with the largest being 0.293 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 14.346Mb (with the largest being 0.393Mb). 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	1,842,292,782	1,827,947,987
GC %	41.85	41.85
Gaps/Gbp	1,523.64	1,546.54
Total gap bp	280,700	295,800
Scaffolds	266	231
Scaffold N50	56,664,674	56,830,550
Scaffold L50	13	13
Scaffold L90	30	30
Contigs	3,073	3,058
Contig N50	1,058,000	1,060,653
Contig L50	532	529
Contig L90	1,735	1,728
QV	41.5882	41.5934
Kmer compl.	87.0697	86.8734
BUSCO sing.	79.4%	79.8%
BUSCO dupl.	18.1%	17.6%
BUSCO frag.	0.9%	0.9%
BUSCO miss.	1.7%	1.8%

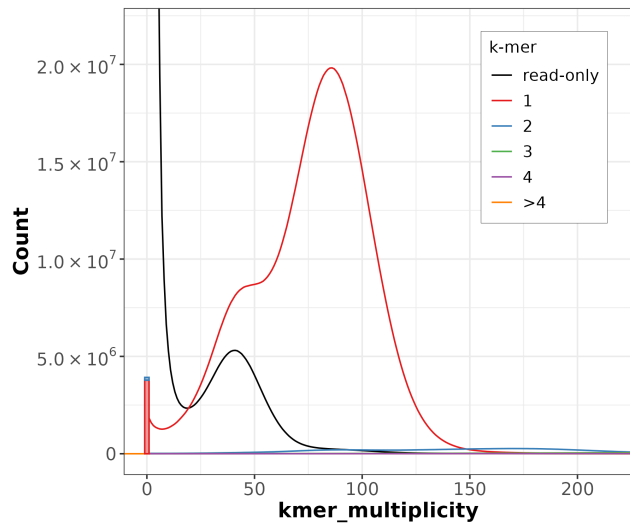
BUSCO: 6.0.0 (euk_genome_min, miniprot) / Lineage: mollusca_odb12 (genomes:36, BUSCOs:4421)

HiC contact map of curated assembly

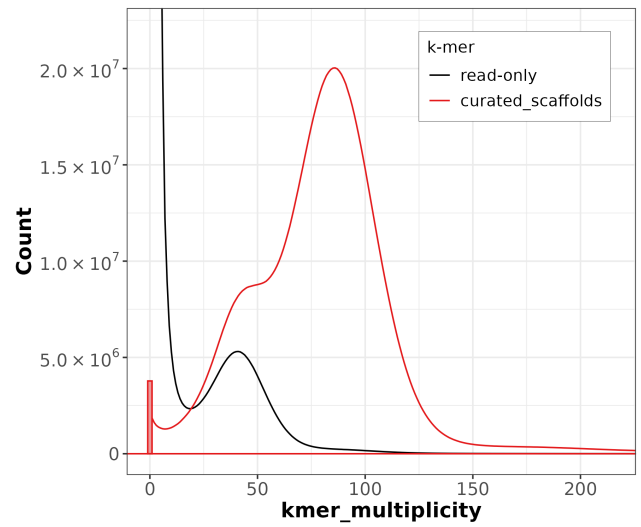


collapsed [\[LINK\]](#)

K-mer spectra of curated assembly

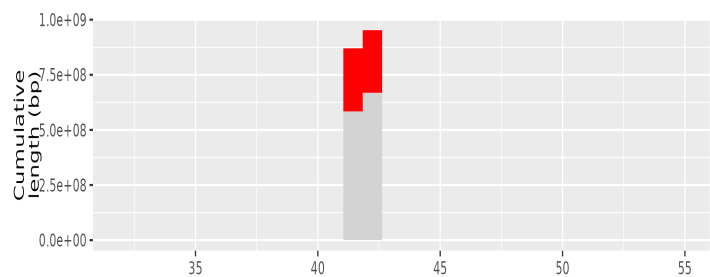


Distribution of k-mer counts per copy numbers found in asm

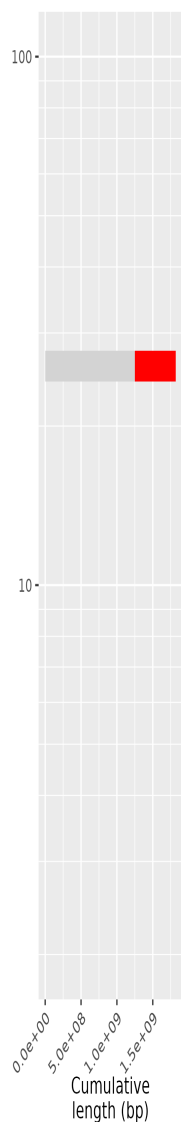
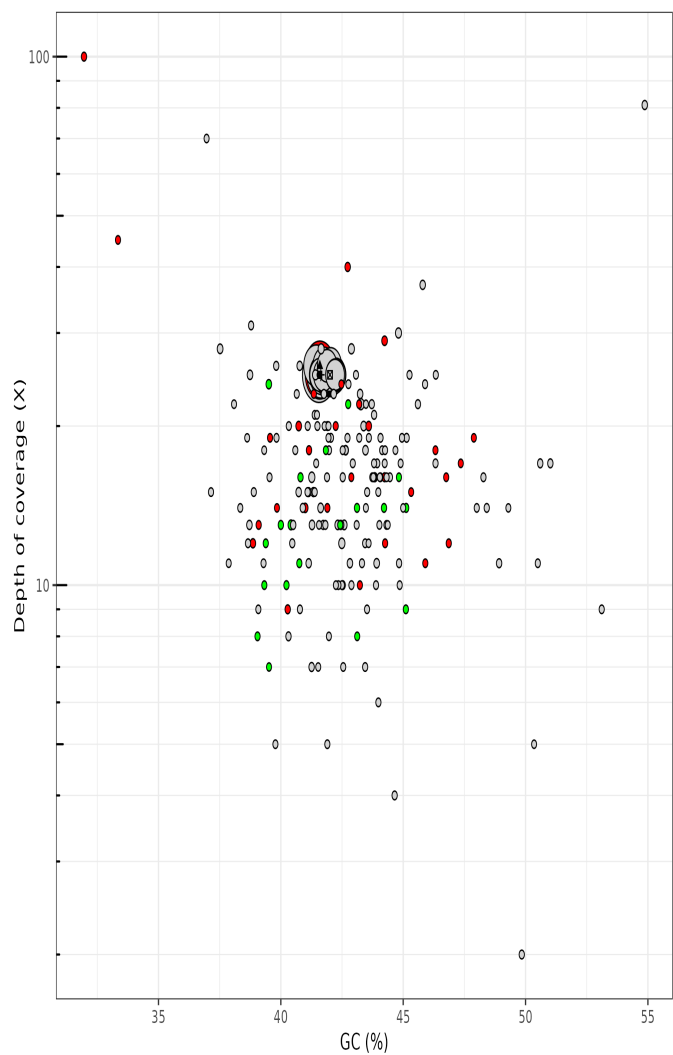


Distribution of k-mer counts coloured by their presence in reads/assemblies

Post-curation contamination screening



TAPAs summary Graph



- superkingdom
- Bacteria
 - Eukaryota
 - N/A
- Length (bp)
- 3.0e+07
 - 6.0e+07
 - 9.0e+07
- Longest sequences (bp)
- xgChiRamo1_1 - 117289264 (N/A)
 - ▲ xgChiRamo1_2 - 90355118 (Eukaryota)
 - xgChiRamo1_3 - 85897122 (Eukaryota)
 - + xgChiRamo1_4 - 77863671 (N/A)
 - ▣ xgChiRamo1_5 - 66933393 (Eukaryota)

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	27	138

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 18:25:43 CEST