ERGA Assembly Report
v24.10.15

Tags: ATLASea[INVALID TAG]

TxID	13016
ToLID	**xbCodOrbi2**
Species	Codakia orbicularis
Class	Bivalvia
Order	Lucinida

Genome Traits	Expected	Observed
Haploid size (bp)	1,704,808,314	1,729,544,933
Haploid Number	19 (source: ancestor)	18
Ploidy	2 (source: ancestor)	2
Sample Sex	Unknown	Unknown

EBP metrics summary and curation notes

Obtained EBP quality metric for collapsed: 7.7.Q47

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

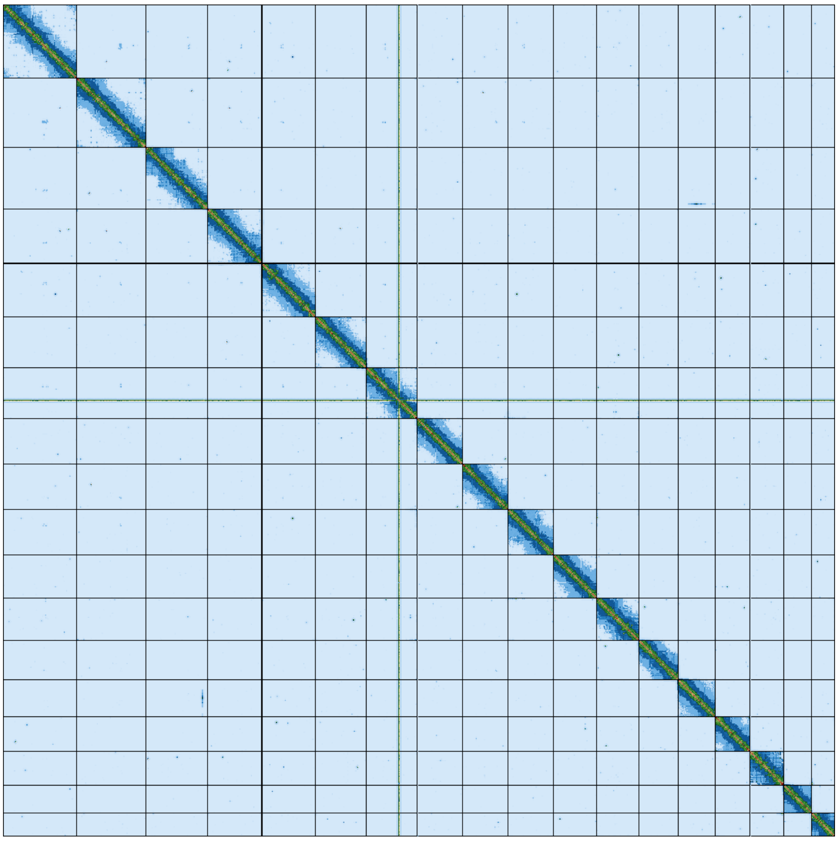
. Interventions/Gb: 5
. Contamination notes: ""
. Other observations: "The assembly of Codakia orbicularis (xbCodOrbi2) is based on 53X PacBio data and 173X Arima Hi-C data generated as part of the ATLASea programme (https://www.atlasea.fr). The assembly process included the following steps: initial 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, 31 contigs were identified as contaminants (bacterial, archaeal, or viral), totaling 5.505 Mb (with the largest being 4.598 Mb). Additionally, 189 regions totaling 53.55 Mb (with the largest being 10.202 Mb) were identified as haplotypic duplications and removed. Mitochondrial genome was assembled using OATK. During manual curation, 6 haplotypic regions were removed, totaling 10,62Mb (with the largest being {args.haplotig_max}Mb).. 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	1,740,211,049	1,729,544,933
GC %	37.12	37.12
Gaps/Gbp	21.26	20.24
Total gap bp	3,700	4,100
Scaffolds	48	44
Scaffold N50	97,884,260	95,407,462
Scaffold L50	8	8
Scaffold L90	16	16
Contigs	85	79
Contig N50	51,581,000	51,581,000
Contig L50	12	12
Contig L90	33	32
QV	47.2717	47.2704
Kmer compl.	69.2739	69.0549
BUSCO sing.	97.2%	98.0%
BUSCO dupl.	1.5%	0.8%
BUSCO frag.	0.6%	0.6%
BUSCO miss.	0.7%	0.7%

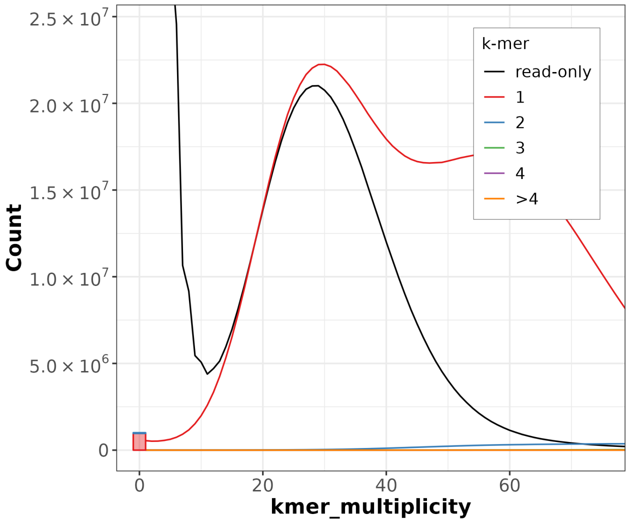
BUSCO: 6.0.0 (euk_genome_min, miniprot) / Lineage: mollusca_odb12 (genomes:36, BUSCOs:4421)

HiC contact map of curated assembly

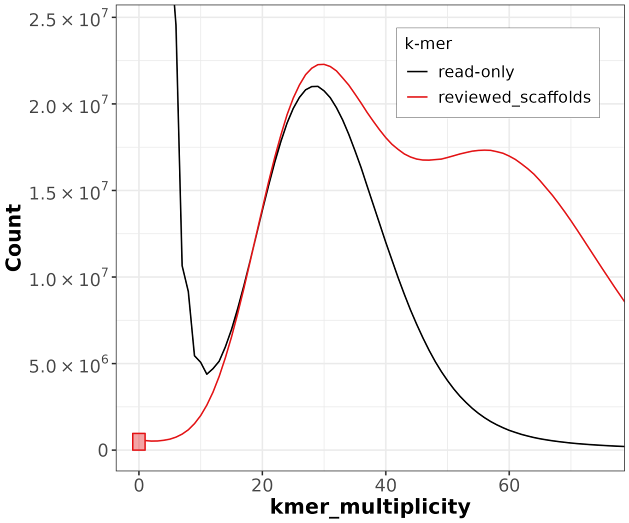


collapsed [LINK]

K-mer spectra of curated assembly

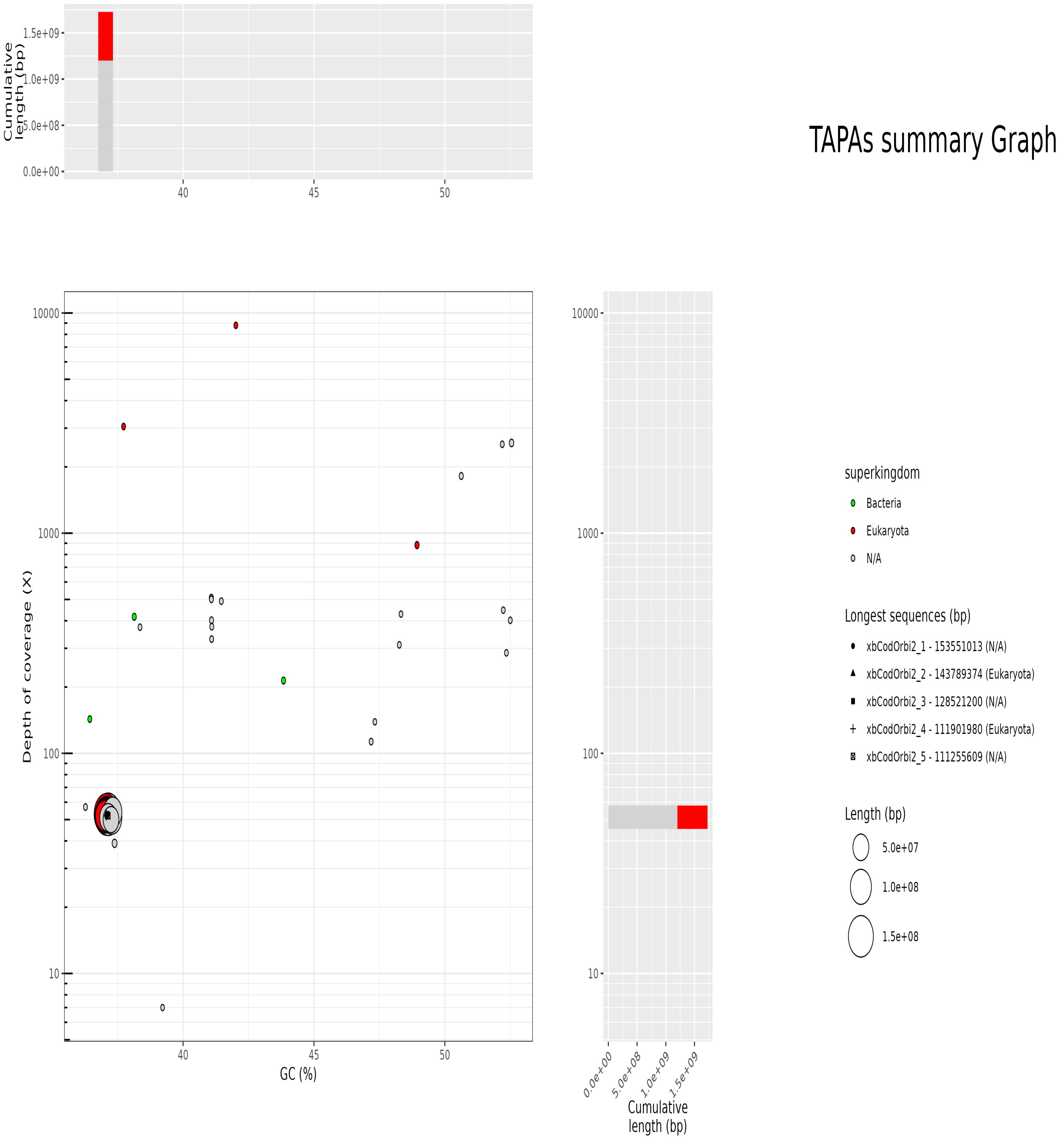


Distribution of k-mer counts per copy numbers found in asm



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

Post-curation contamination screening



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	53	173

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: Arnaud Couloux
Affiliation: Genoscope

Date and time: 2025-12-13 10:05:39 CET