

ERGA Assembly Report

v24.10.15

Tags: ATLASEa[INVALID TAG]

TxID	2991936
ToLID	odCelCioc1
Species	Celtodoryx ciocalyptoides
Class	Demospongiae
Order	Poecilosclerida

Genome Traits	Expected	Observed
Haploid size (bp)	166,822,604	161,126,432
Haploid Number	5 (source: ancestor)	20
Ploidy	2 (source: ancestor)	2
Sample Sex	Unknown	Unknown

EBP metrics summary and curation notes

Obtained EBP quality metric for collapsed: 6.6.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

Curator notes

. Interventions/Gb: 208
. Contamination notes: ""
. Other observations: "The assembly of *Celtodoryx ciocalyptoides* (odCelCioc1) is based on 44X PacBio data and 146X 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, 541 contigs were identified as contaminants (bacterial, archaeal, or viral), totaling 39.942 Mb (with the largest being 4.782 Mb). Additionally, 114 regions totaling 9.835 Mb (with the largest being 0.357 Mb) were identified as haplotypic duplications and removed. Mitochondrial genome was assembled using OATK. Finally, the primary assembly was analyzed and manually improved using Pretext. During manual curation, 5 haplotypic regions and 34 contaminant sequences were removed, totaling 1.868Mb and 0.999Mb, respectively (with the largest being 0.689Mb and 0.105Mb). The Hi-C contact map contained relatively few long-range interactions, making chromosome validation and orientation more challenging. However, the assembly was highly contiguous, with very few gaps, which

facilitated the reconstruction of chromosome-scale scaffolds. Chromosome-scale scaffolds confirmed by Hi-C data were named in order of size. "

Quality metrics table

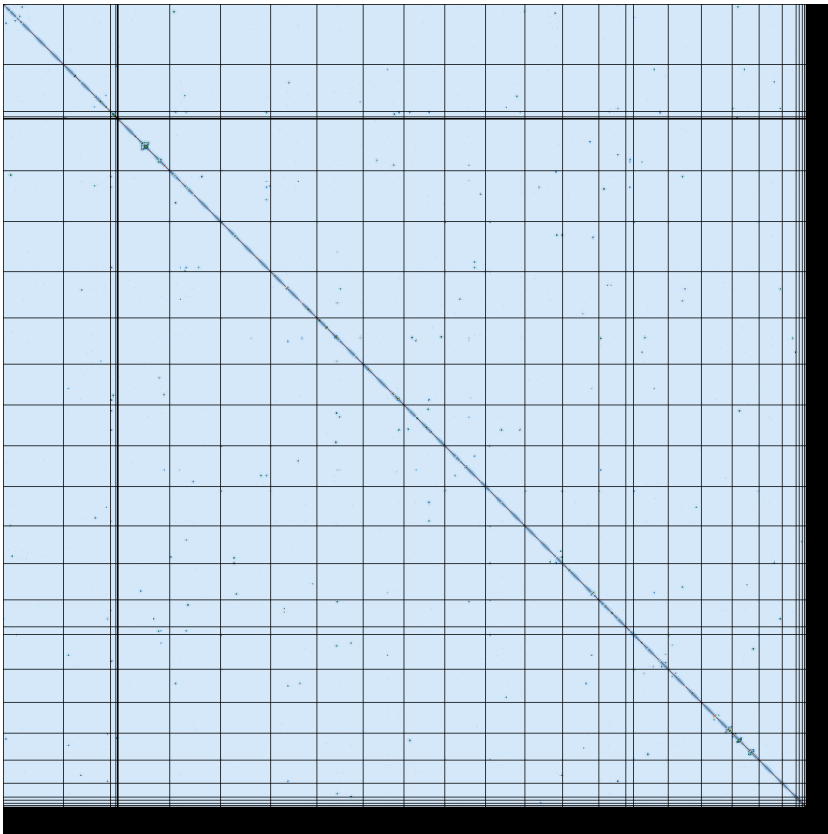
Metrics	Pre-curation collapsed	Curated collapsed
Total bp	164,060,488	161,126,432
GC %	42.61	42.66
Gaps/Gbp	902.11	906.12
Total gap bp	16,200	17,900
Scaffolds	197	120
Scaffold N50	6,708,599	7,911,763
Scaffold L50	10	9
Scaffold L90	25	19
Contigs	330	266
Contig N50	1,492,303	1,492,303
Contig L50	35	35
Contig L90	120	116
QV	43.0057	44.3676
Kmer compl.	68.9589	68.433
BUSCO sing.	53.9%	81.2%
BUSCO dupl.	0.7%	1.2%
BUSCO frag.	27.7%	5.4%
BUSCO miss.	17.7%	12.2%

Warning! BUSCO versions or lineage datasets are not the same across results:

BUSCO: 5.8.2 (euk_genome_met, metaeuk) / Lineage: metazoa_odb12 (genomes:206, BUSCOs:672)

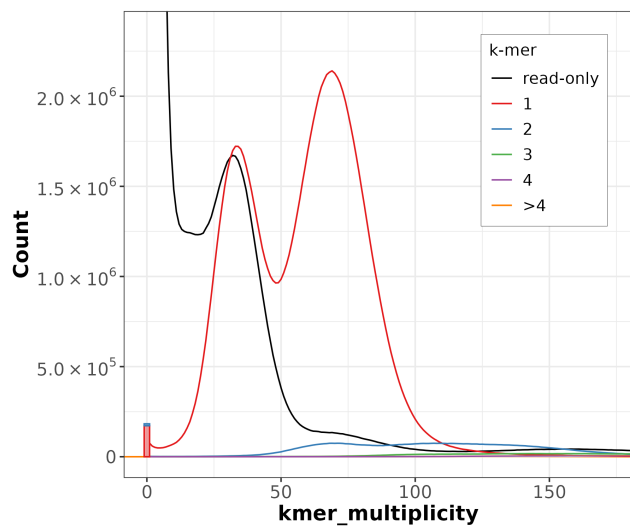
BUSCO: 5.8.2 (euk_genome_min, miniprot) / Lineage: metazoa_odb12 (genomes:206, BUSCOs:672)

HiC contact map of curated assembly

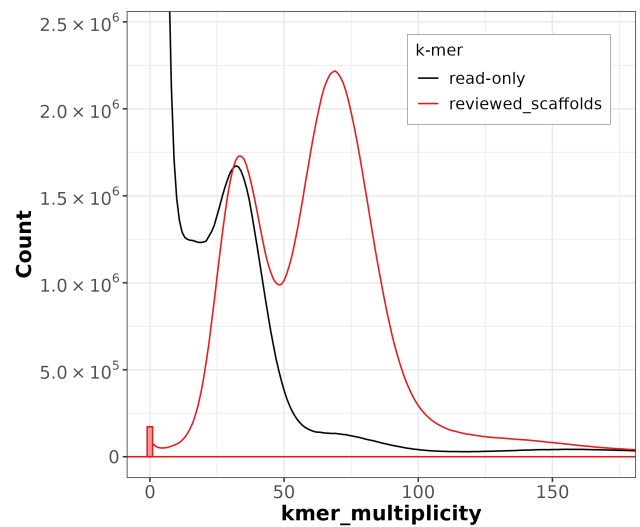


collapsed [\[LINK\]](#)

K-mer spectra of curated assembly

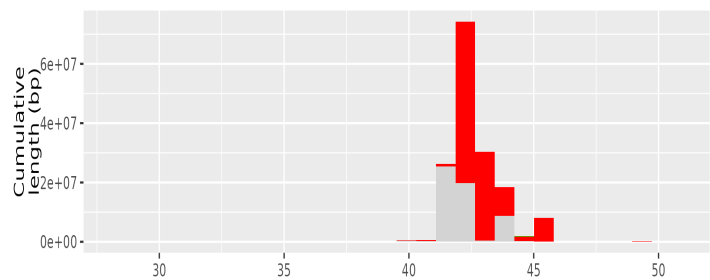


Distribution of k-mer counts per copy numbers found in asm

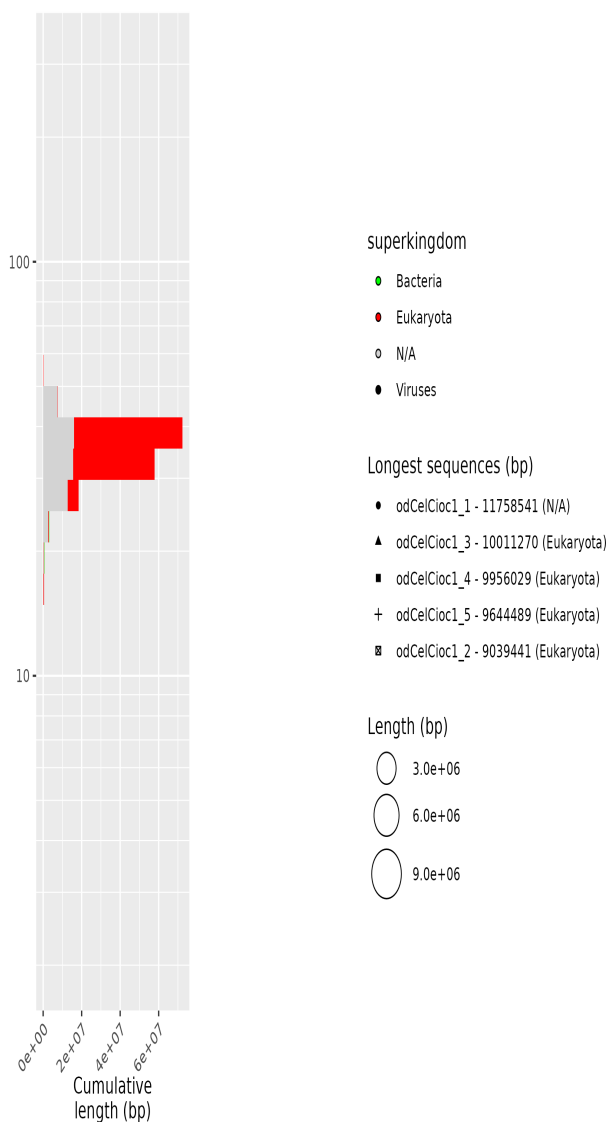
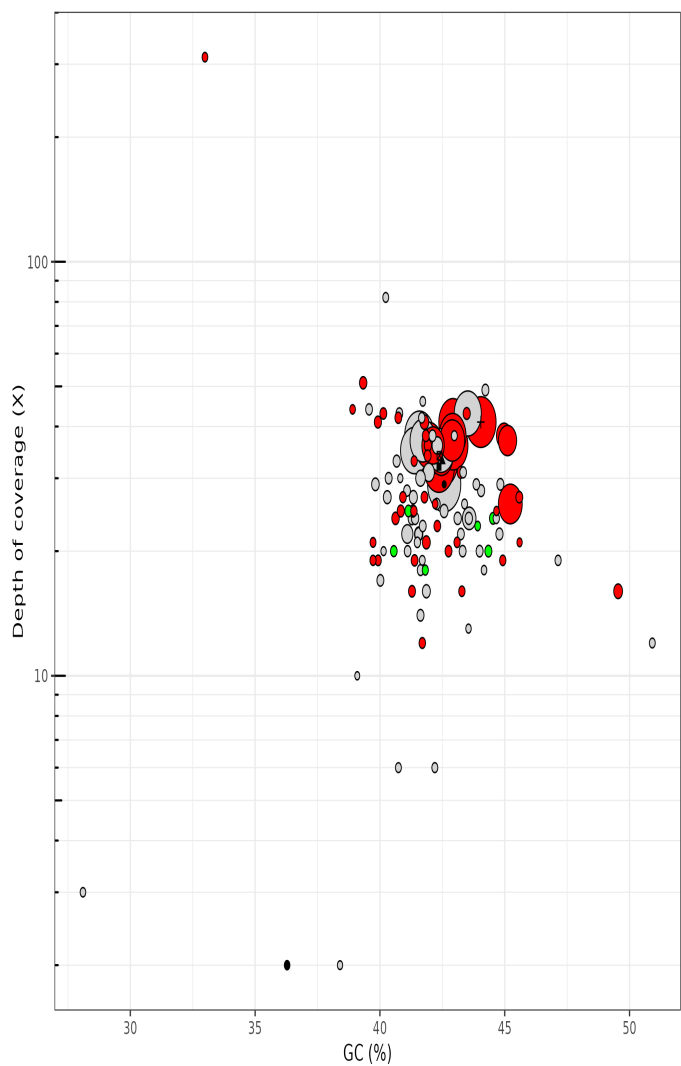


Distribution of k-mer counts coloured by their presence in reads/assemblies

Post-curation contamination screening



TAPAs summary Graph



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	46	151

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: 2026-08-08 02:54:47 CEST