

ERGA Assembly Report

v24.10.15

Tags: ATLASEa[INVALID TAG]

TxID	3094989
ToLID	xgCan0dhn4
Species	Candiella odhneri
Class	Gastropoda
Order	Nudibranchia

Genome Traits	Expected	Observed
Haploid size (bp)	1,019,708,199	1,111,637,671
Haploid Number	11 (source: ancestor)	16
Ploidy	2 (source: ancestor)	2
Sample Sex	Unknown	Unknown

EBP metrics summary and curation notes

Obtained EBP quality metric for collapsed: 6.7.Q40

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
- . Assembly length loss > 3% for collapsed

Curator notes

. Interventions/Gb: 376
. Contamination notes: ""
. Other observations: "The assembly of Candiella odhneri (xgCan0dhn4) is based on 44X ONT data and 127X 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, 12 contigs were identified as contaminants (bacterial, archaeal, or viral), totaling 1.089 Mb (with the largest being 0.594 Mb). Additionally, 1149 regions totaling 154.937 Mb (with the largest being 0.984 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, 194 haplotypic regions and 3 contaminant sequences were removed, totaling 113.349Mb and 0.11Mb, respectively (with the largest being 2.266Mb and 0.061Mb). Chromosome-scale scaffolds confirmed by Hi-C data were named in order of size. "

Quality metrics table

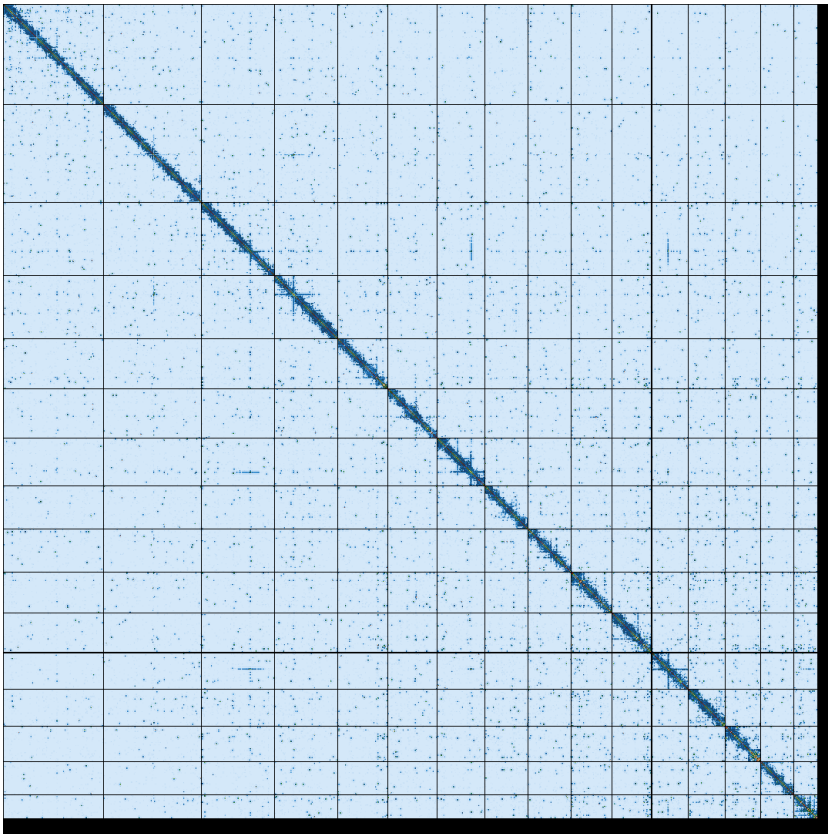
Metrics	Pre-curation collapsed	Curated collapsed
Total bp	1,225,074,667	1,111,637,671
GC %	38.42	38.37
Gaps/Gbp	644.86	630.6
Total gap bp	79,000	86,000
Scaffolds	302	248
Scaffold N50	68,872,221	65,506,304
Scaffold L50	7	6
Scaffold L90	15	14
Contigs	1,092	949
Contig N50	2,747,000	2,928,369
Contig L50	122	107
Contig L90	498	400
QV	40.541	40.6263
Kmer compl.	69.7984	65.5337
BUSCO sing.	90.5%	95.1%
BUSCO dupl.	7.8%	1.1%
BUSCO frag.	0.4%	2.2%
BUSCO miss.	1.3%	1.6%

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)

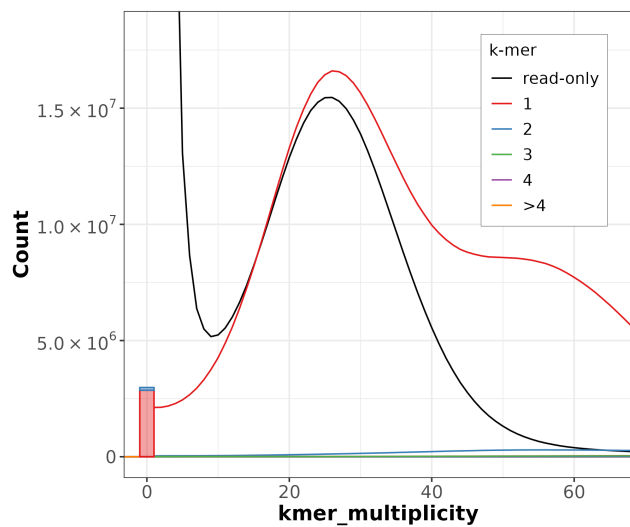
BUSCO: 5.8.2 (euk_genome_min, miniprot) / Lineage: mollusca_odb12 (genomes:36, BUSCOs:4421)

HiC contact map of curated assembly

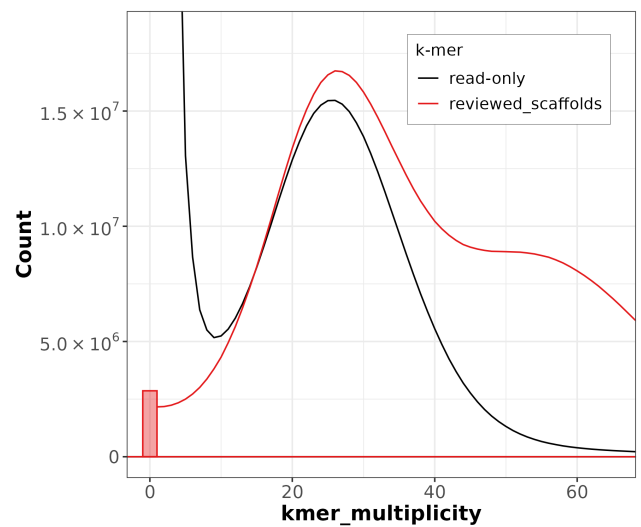


collapsed [\[LINK\]](#)

K-mer spectra of curated assembly

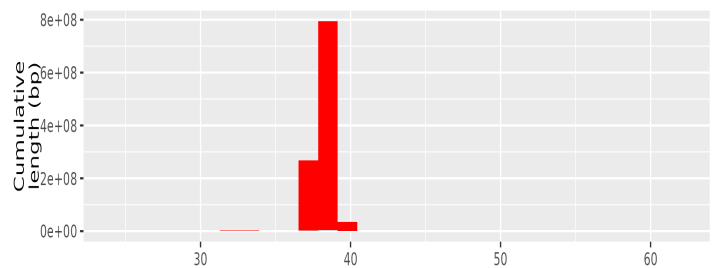


Distribution of k-mer counts per copy numbers found in asm



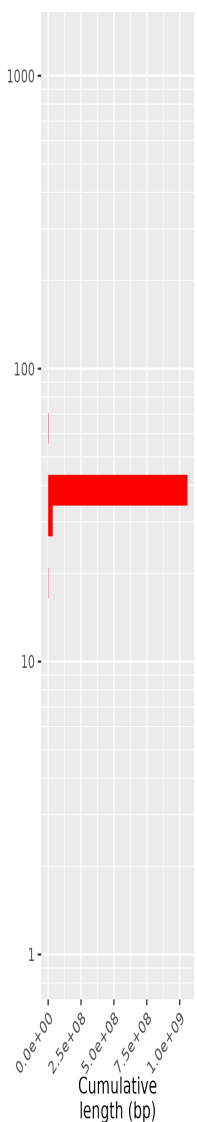
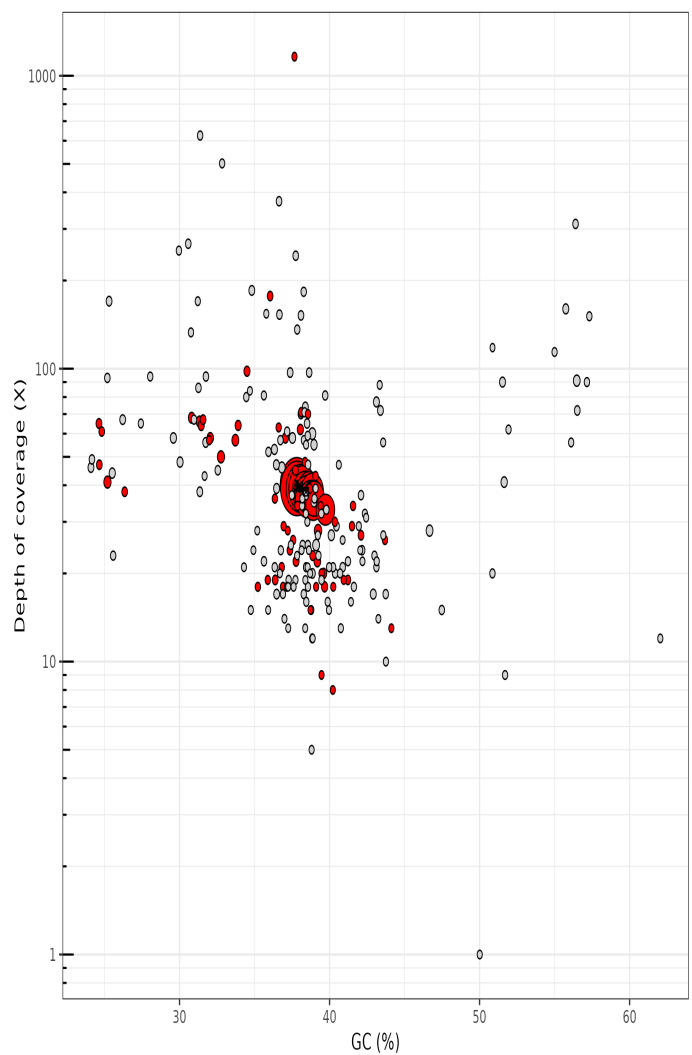
Distribution of k-mer counts coloured by their presence in reads/assemblies

Post-curation contamination screening

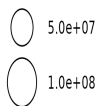


TAPAs summary Graph

(10 0X contigs have been hidden)



Length (bp)



Longest sequences (bp)

- xgCanOdhn4_1 - 134769338 (Eukaryota)
- ▲ xgCanOdhn4_2 - 130533061 (Eukaryota)
- xgCanOdhn4_3 - 97240718 (Eukaryota)
- + xgCanOdhn4_4 - 85009605 (Eukaryota)
- ▣ xgCanOdhn4_5 - 66809807 (Eukaryota)

superkingdom

- Eukaryota
- 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	48	139

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: 2026-08-02 18:09:49 CEST