

ERGA Assembly Report

v24.10.15

Tags: ATLASEa[INVALID TAG]

TxID	260824
ToLID	kaCysDe111
Species	Cystodytes dellechiaiei
Class	Ascidacea
Order	Aplousobranchia

Genome Traits	Expected	Observed
Haploid size (bp)	1,413,758,654	1,333,769,197
Haploid Number	9 (source: ancestor)	21
Ploidy	2 (source: ancestor)	2
Sample Sex	Unknown	Unknown

EBP metrics summary and curation notes

Obtained EBP quality metric for collapsed: 6.7.Q46

The following metrics were automatically flagged as below EBP recommended standards or different from expected:

- . Observed Haploid Number is different from Expected
- . BUSCO single copy value is less than 90% for collapsed

Curator notes

. Interventions/Gb: 65
. Contamination notes: ""
. Other observations: "The assembly of Cystodytes dellechiaiei (kaCysDe111) is based on 32X PacBio data and 180X 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, 189 contigs were identified as contaminants (bacterial, archaeal, or viral), totaling 24.111 Mb (with the largest being 5.22 Mb). Additionally, 470 regions totaling 79.113 Mb (with the largest being 1.904 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, 13 haplotypic regions were removed, totaling 7.147Mb (with the largest being 1.106Mb). Chromosome-scale scaffolds confirmed by Hi-C data were named in order of size. "

Quality metrics table

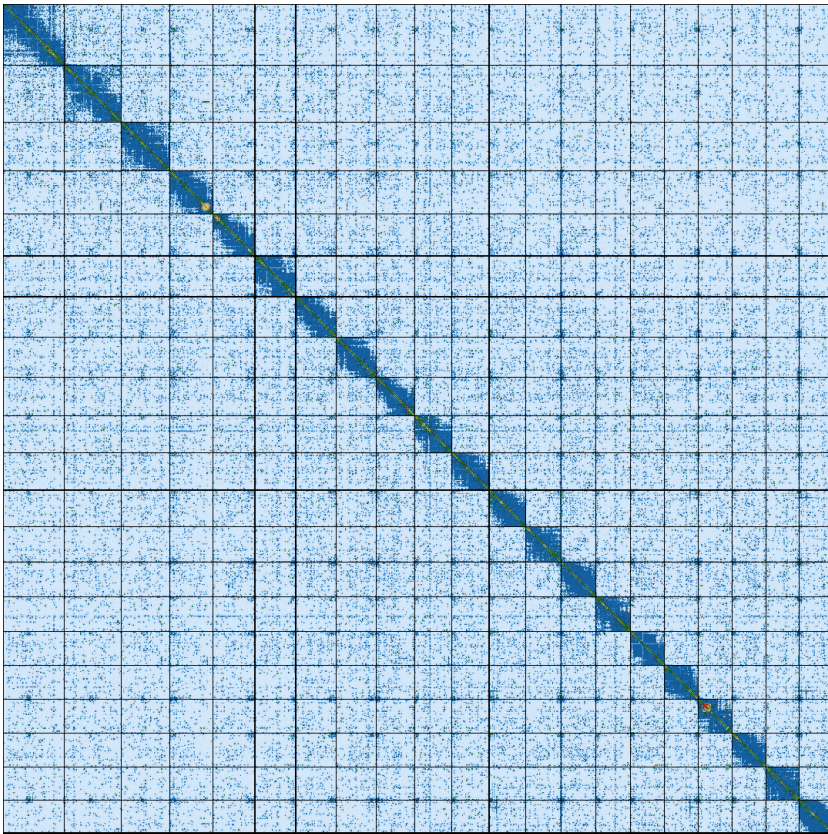
Metrics	Pre-curation collapsed	Curated collapsed
Total bp	1,340,894,754	1,333,769,197
GC %	38.59	38.59
Gaps/Gbp	198.38	217.43
Total gap bp	26,600	34,100
Scaffolds	121	90
Scaffold N50	59,507,780	60,007,437
Scaffold L50	10	10
Scaffold L90	19	19
Contigs	387	380
Contig N50	7,139,953	7,093,973
Contig L50	53	53
Contig L90	177	177
QV	46.7169	46.726
Kmer compl.	90.5889	90.3996
BUSCO sing.	87.6%	87.8%
BUSCO dupl.	4.2%	3.7%
BUSCO frag.	2.5%	2.7%
BUSCO miss.	5.7%	5.8%

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

BUSCO: 6.0.0 (euk_genome_min, miniprot) / Lineage: metazoa_odb12 (genomes:206, BUSCOs:672)

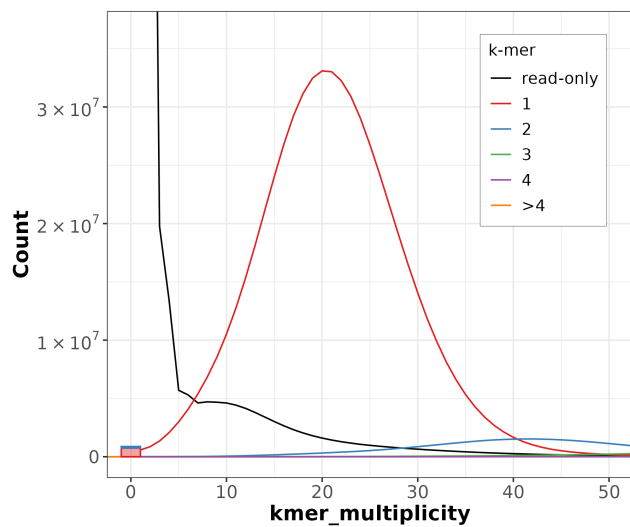
BUSCO: 5.8.2 (euk_genome_min, miniprot) / Lineage: metazoa_odb12 (genomes:206, BUSCOs:672)

HiC contact map of curated assembly

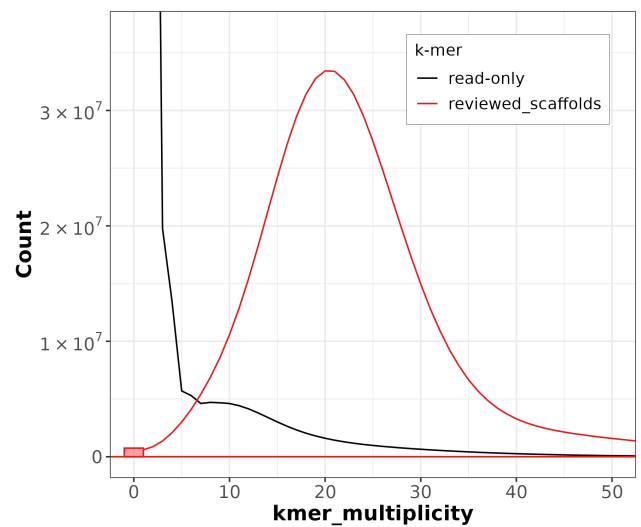


collapsed [\[LINK\]](#)

K-mer spectra of curated assembly

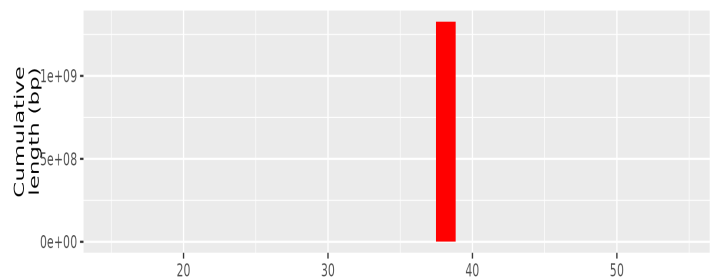


Distribution of k-mer counts per copy numbers found in asm

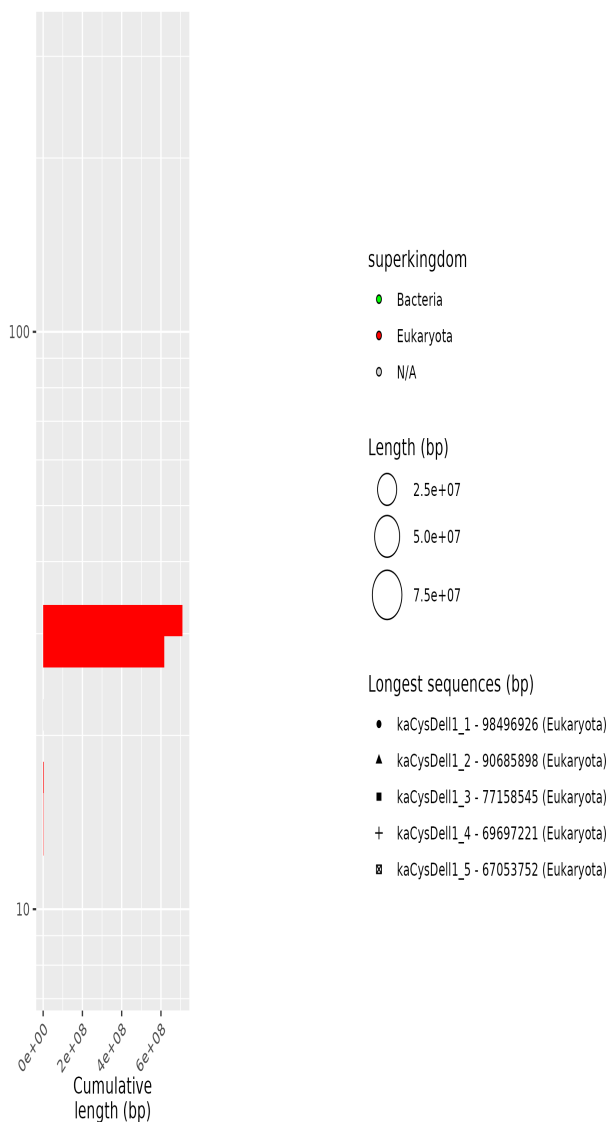
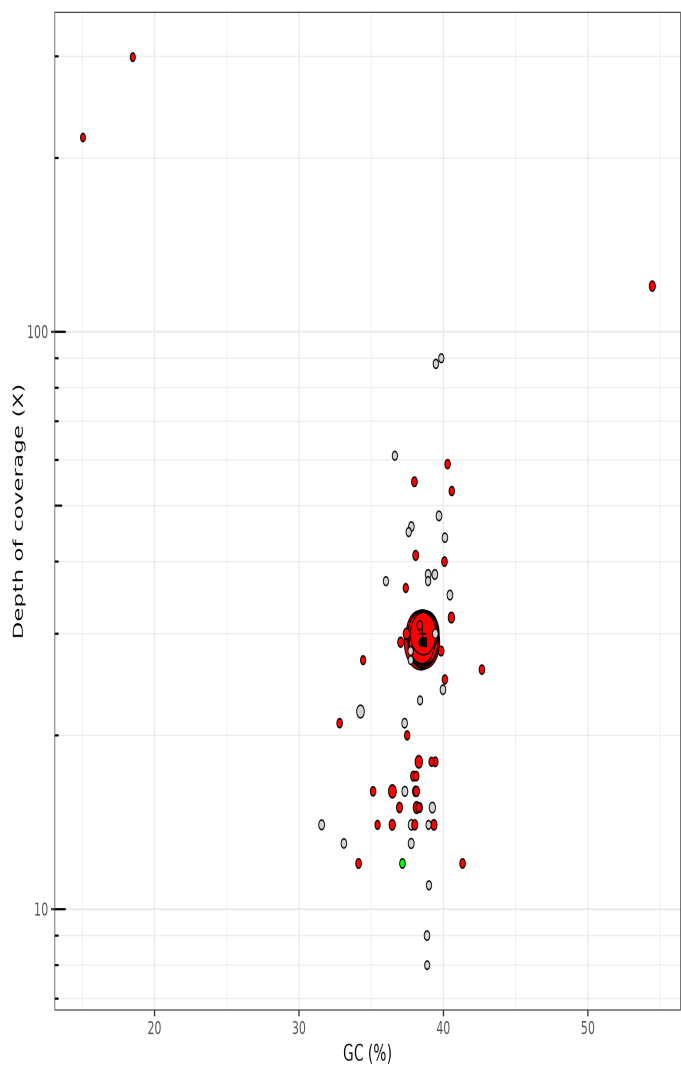


Distribution of k-mer counts coloured by their presence in reads/assemblies

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	32	181

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: 2026-08-29 01:43:58 CEST