

# ERGA Assembly Report

v24.10.15

Tags: ERGA-BGE

TxID	317547
ToLID	<b>jaEunCavol1.1</b>
Species	Eunicella cavolini
Class	Anthozoa
Order	Malacalcyonacea

Genome Traits	Expected	Observed
Haploid size (bp)	411,546,873	497,463,504
Haploid Number	6 (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.Q61

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

- . Observed Haploid size (bp) has >20% difference with 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
- . Assembly length loss > 3% for collapsed

### Curator notes

- . Interventions/Gb: 516
- . Contamination notes: ""
- . Other observations: "The assembly of EUNICELLA CAVOLINI (jaEunCavol) is based on 53X PacBio data and 130X Arima Hi-C data generated as part of the European Reference Genome Atlas (ERGA, <https://www.erga-biodiversity.eu/>) via the Biodiversity Genomics Europe project (BGE, <https://biodiversitygenomics.eu/>). 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, 37 contigs were identified as contaminants (bacterial, archaeal, or viral), totaling 1 Mb (with the largest being 88,287 pb). Additionally, 208 regions totaling 168 Mb were identified as haplotypic duplications and removed. However, all sequences larger than 1 Mb were added to the assembly, with purge\_dups removing too many contigs needed for the scaffolding. The mitochondrial genome was assembled using OATK. Finally, the primary assembly was analyzed and manually improved using Pretext. During manual curation, 48

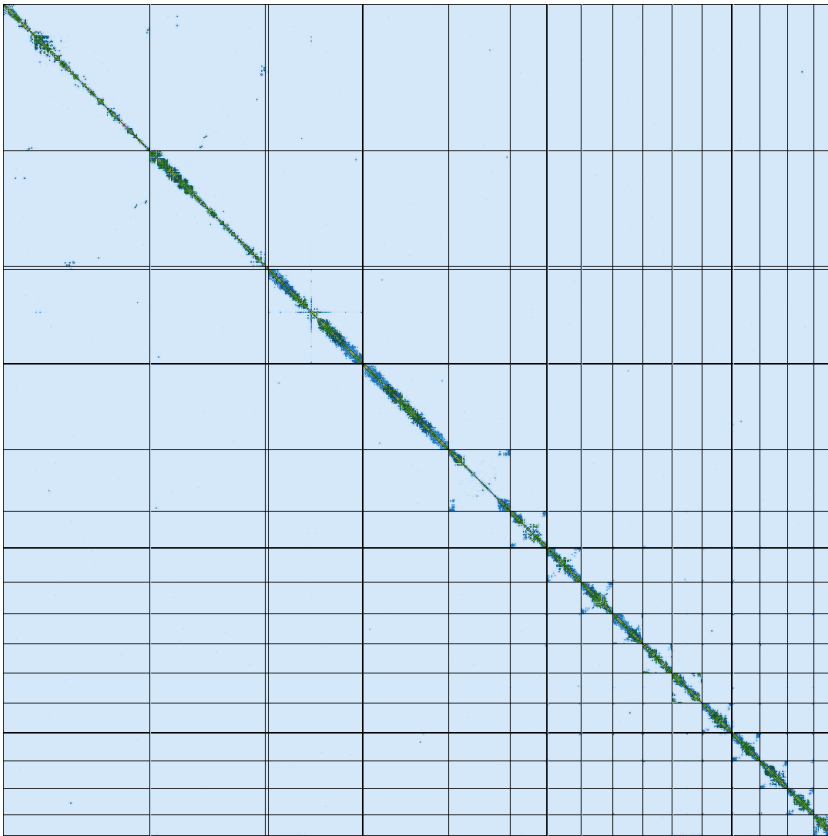
haplotypic regions were removed, totaling 20 Mb (with the largest being 1 Mb).  
Chromosome-scale scaffolds confirmed by Hi-C data were named in order of size. "

## Quality metrics table

Metrics	Pre-curation collapsed	Curated collapsed
Total bp	517,457,888	497,463,504
GC %	37.34	37.35
Gaps/Gbp	0	251.27
Total gap bp	0	17,600
Scaffolds	127	46
Scaffold N50	9,874,030	51,198,936
Scaffold L50	15	4
Scaffold L90	55	13
Contigs	127	171
Contig N50	9,874,030	7,689,143
Contig L50	15	21
Contig L90	55	74
QV	61.6597	61.5375
Kmer compl.	79.7168	78.1301
BUSCO sing.	86.4%	87.7%
BUSCO dupl.	2.4%	1.0%
BUSCO frag.	5.5%	5.5%
BUSCO miss.	5.7%	5.8%

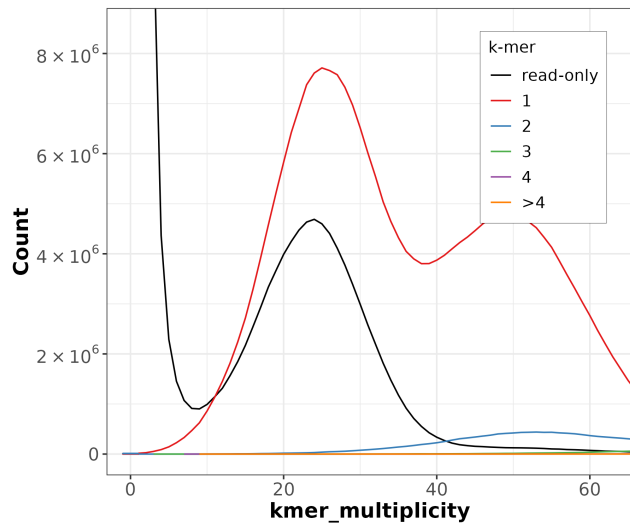
BUSCO: 5.4.3 (euk\_genome\_met, metaeuk) / Lineage: metazoa\_odb10 (genomes:65, BUSCOs:954)

# HiC contact map of curated assembly

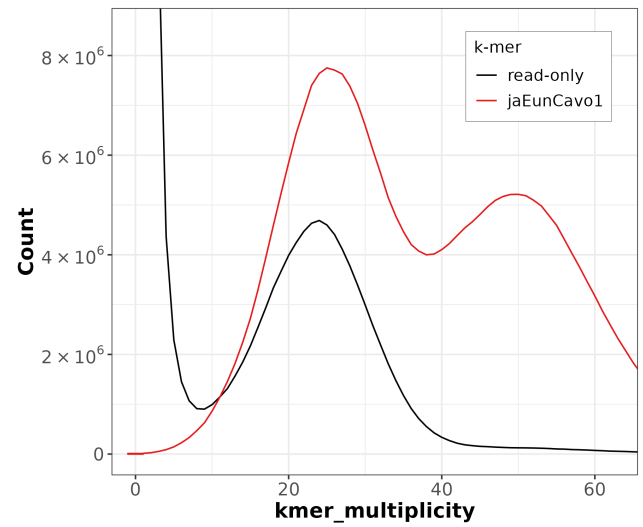


collapsed [\[LINK\]](#)

# K-mer spectra of curated assembly

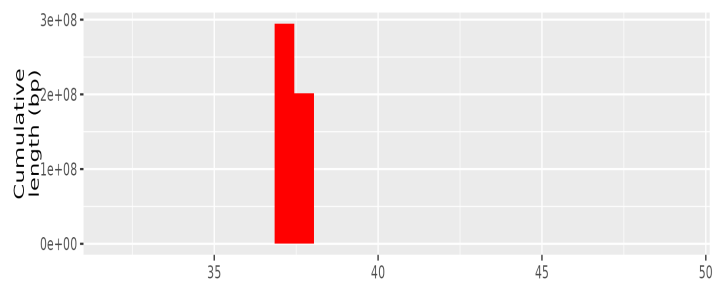


Distribution of k-mer counts per copy numbers found in asm

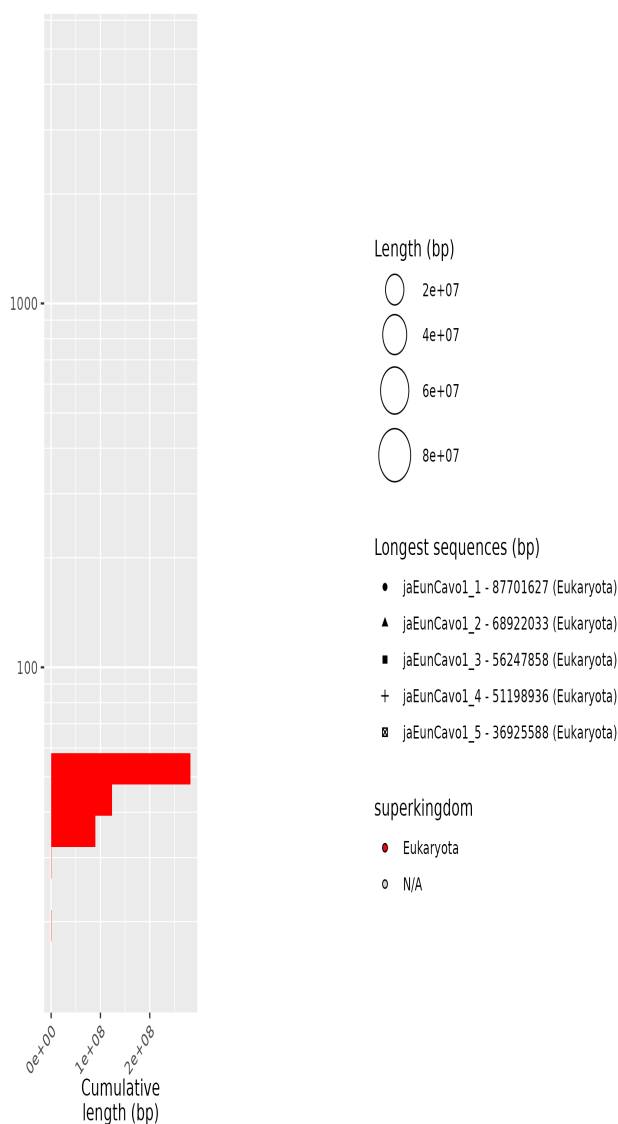
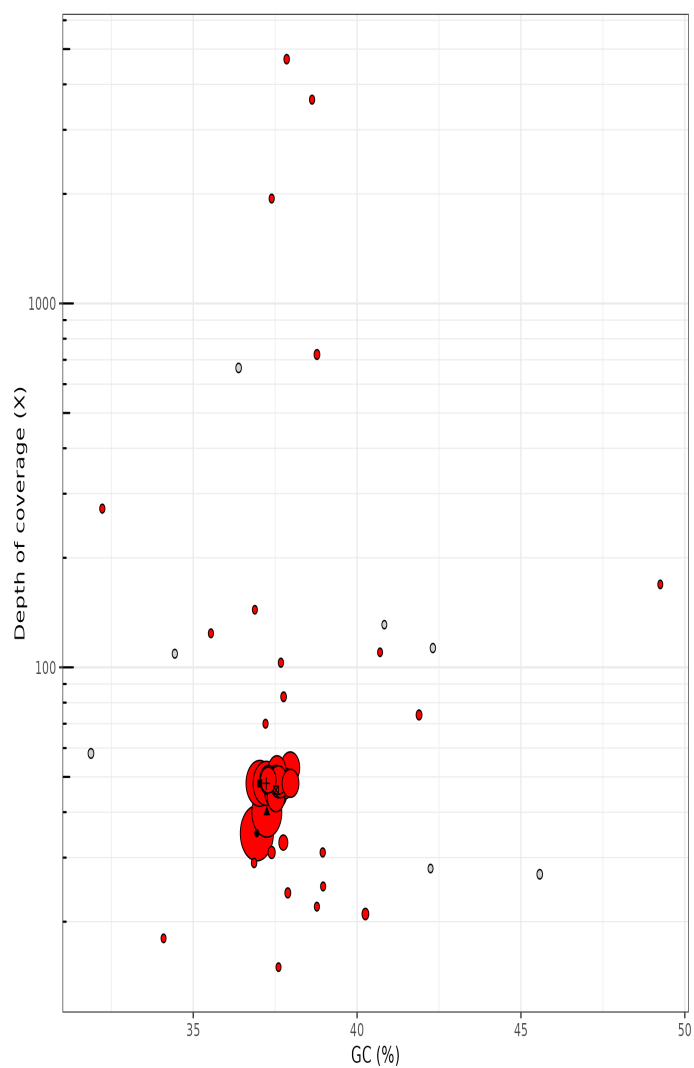


Distribution of k-mer counts coloured by their presence in reads/assemblies

# 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	PACBIO Hifi	Arima
Coverage	53	130

# 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: Lola Demirdjian

Affiliation: Genoscope

Date and time: 2025-02-11 15:39:55 CET