

# ERGA Assembly Report

v24.10.15

Tags: ERGA-BGE

TxID	130055
ToLID	<b>jaClaCaes1</b>
Species	Cladocora caespitosa
Class	Anthozoa
Order	Scleractinia

Genome Traits	Expected	Observed
Haploid size (bp)	555,073,251	632,223,877
Haploid Number	6 (source: ancestor)	14
Ploidy	2 (source: ancestor)	2
Sample Sex	unknown	unknown

## EBP metrics summary and curation notes

Obtained EBP quality metric for pri: 6.7.Q55

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 pri
- . Assembly length loss > 3% for pri
- . Not 90% of assembly in chromosomes for pri

### Curator notes

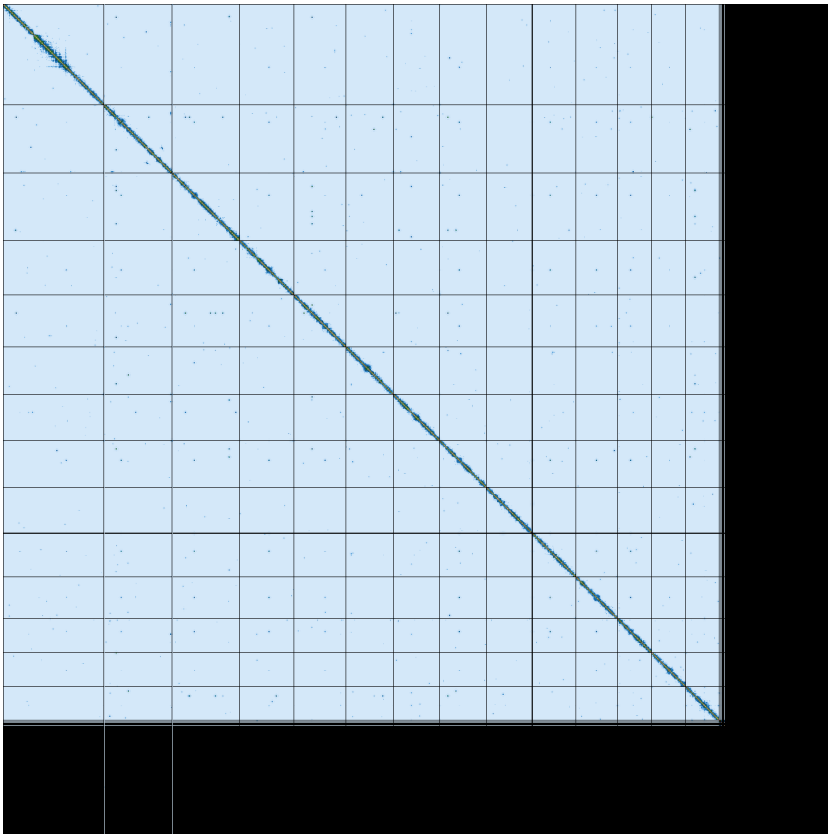
- . Interventions/Gb: 144
- . Contamination notes: "234 contigs were excluded via FCS-GX. Though 900 more contigs are marked with the REVIEW label."
- . Other observations: "The initial PacBio data set shows a very broad range of contamination/cobionts up to 14X. Therefore the PacBio reads were subsampled to 40X, which reduced the "erroneous" kmers to a normal fraction. The hifiasm assembly based on the subsampled reads was quite fragmented (4889 contigs, N50 4.5Mb). FCS-GX decontamination just removed 234 contigs (see contamination notes) and haplotype purging removed 1072 (45.5Mb) additional contigs. The HiC data is of suboptimal quality (40M UU-read pairs with just 3% >40Kb resolution). There is still a large fraction of potential symbionts in the shrapnel contigs. see pretext map and Blobtools plot (low coverage + high GC). But those very short contigs (<30Kb) cannot be reliably identified via FCS-GX nor blobtools. Should we filter by coverage or just mark them as potential cobionts?"

# Quality metrics table

Metrics	Pre-curation pri	Curated pri
Total bp	704,195,618	632,223,877
GC %	41.01	40.71
Gaps/Gbp	0	716.52
Total gap bp	0	51,000
Scaffolds	4,889	3,235
Scaffold N50	4,579,535	35,445,312
Scaffold L50	45	7
Scaffold L90	1,319	271
Contigs	4,889	3,688
Contig N50	4,579,535	4,476,598
Contig L50	45	42
Contig L90	1,319	709
QV	54.9445	55.2565
Kmer compl.	85.3463	82.4133
BUSCO sing.	94.5%	95.6%
BUSCO dupl.	2.4%	1.1%
BUSCO frag.	2.5%	2.5%
BUSCO miss.	0.6%	0.8%

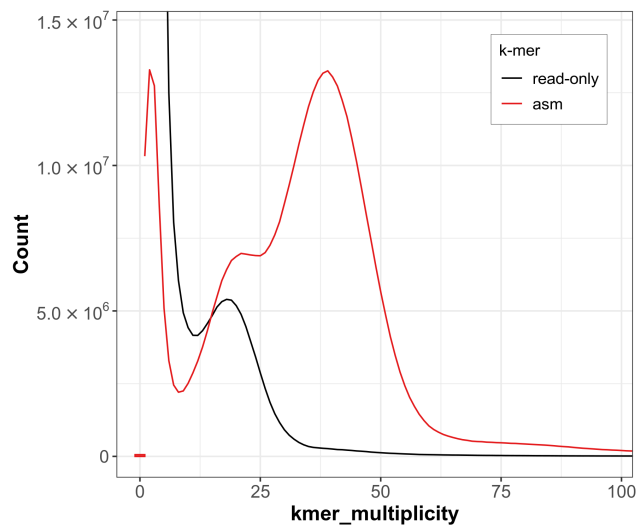
BUSCO: 5.8.3 (euk\_genome\_min, miniprot) / Lineage: anthozoa\_odb12 (genomes:12, BUSCOs:3649)

# HiC contact map of curated assembly

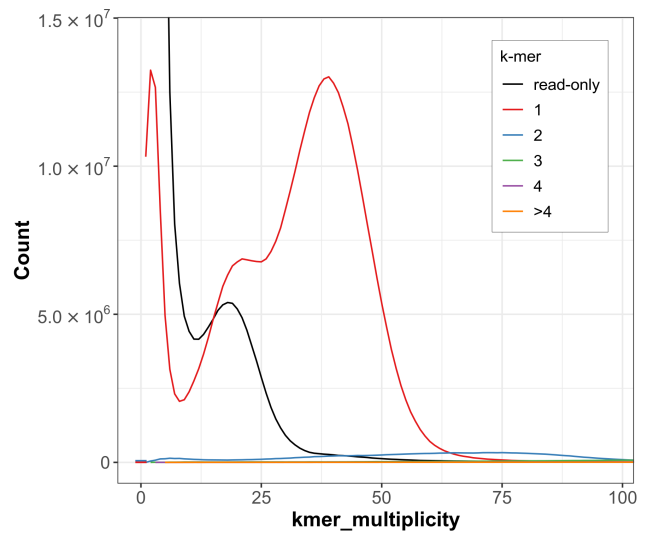


pri [\[LINK\]](#)

# K-mer spectra of curated assembly

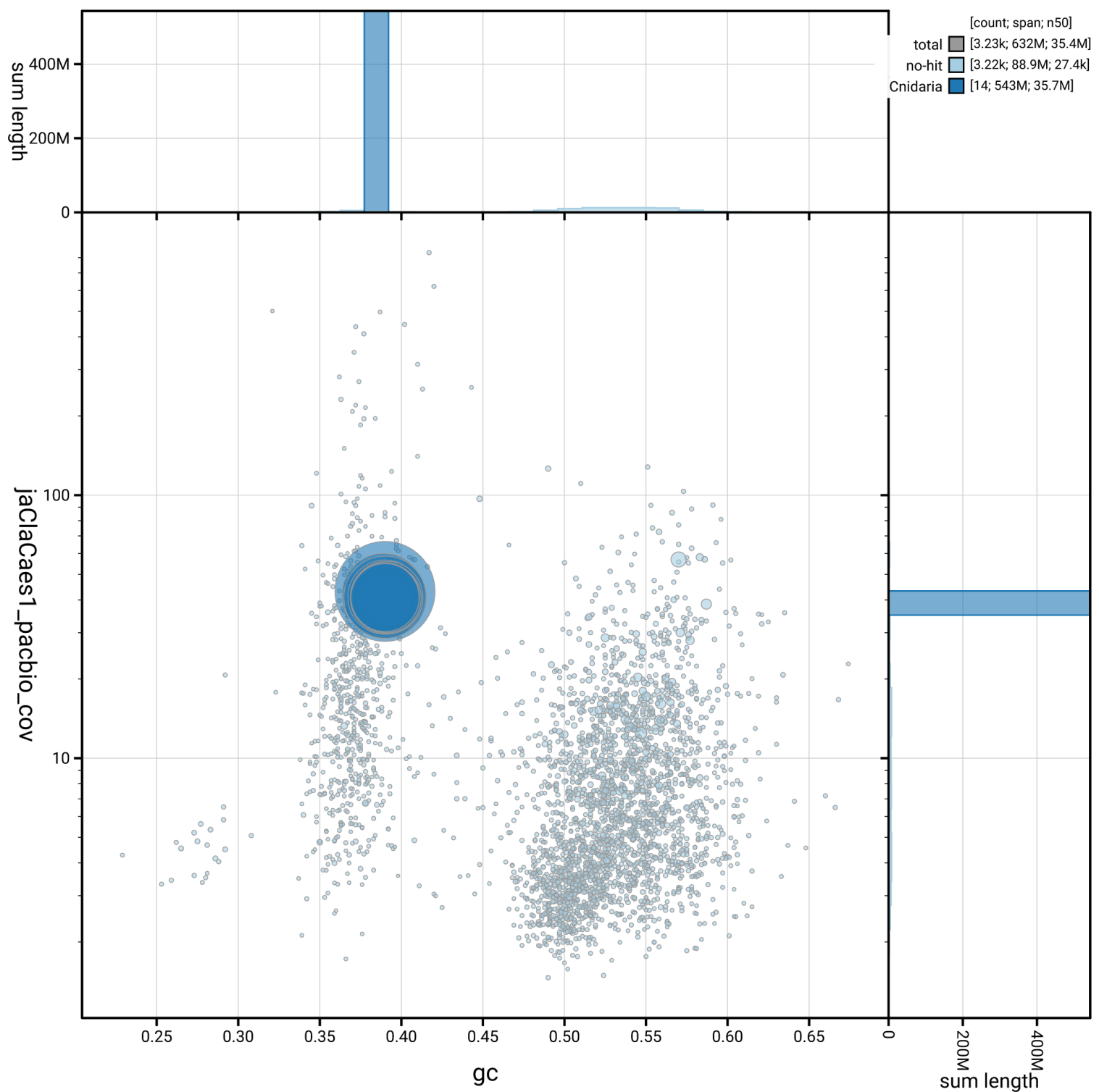


Distribution of k-mer counts coloured by their presence in reads/assemblies



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

# Post-curation contamination screening



**pri.** 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	HiFi	HiC
Coverage	87x	91x

# Assembly pipeline

- **Hifiasm**
  - |\_ *ver*: 0.25.0-r726
  - |\_ *key param*: HiC
  - |\_ *key param*: 13
- **purge\_dups**
  - |\_ *ver*: 1.2.6
  - |\_ *key param*: NA
- **YaHS**
  - |\_ *ver*: 1.2.2
  - |\_ *key param*: NA

# Curation pipeline

- **GRIT\_Rapid**
  - |\_ *ver*: 1a3d79a8
  - |\_ *key param*: NA
- **HiGlass**
  - |\_ *ver*: 0.10.4
  - |\_ *key param*: NA

Submitter: Martin Pippel

Affiliation: SciLifeLab

Date and time: 2025-10-01 11:04:18 CEST