

ERGA Assembly Report

v24.10.15

Tags: ERGA-BGE

TxID	130055
ToLID	jaClaCaes1
Species	Cladocora caespitosa
Class	Anthozoa
Order	Scleractinia

Genome Traits	Expected	Observed
Haploid size (bp)	555,073,251	554,556,895
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.Q59

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

Curator notes

. Interventions/Gb: 144

. Contamination notes: "234 contigs were excluded via FCS-GX. Though 900 more contigs are marked with the REVIEW label."

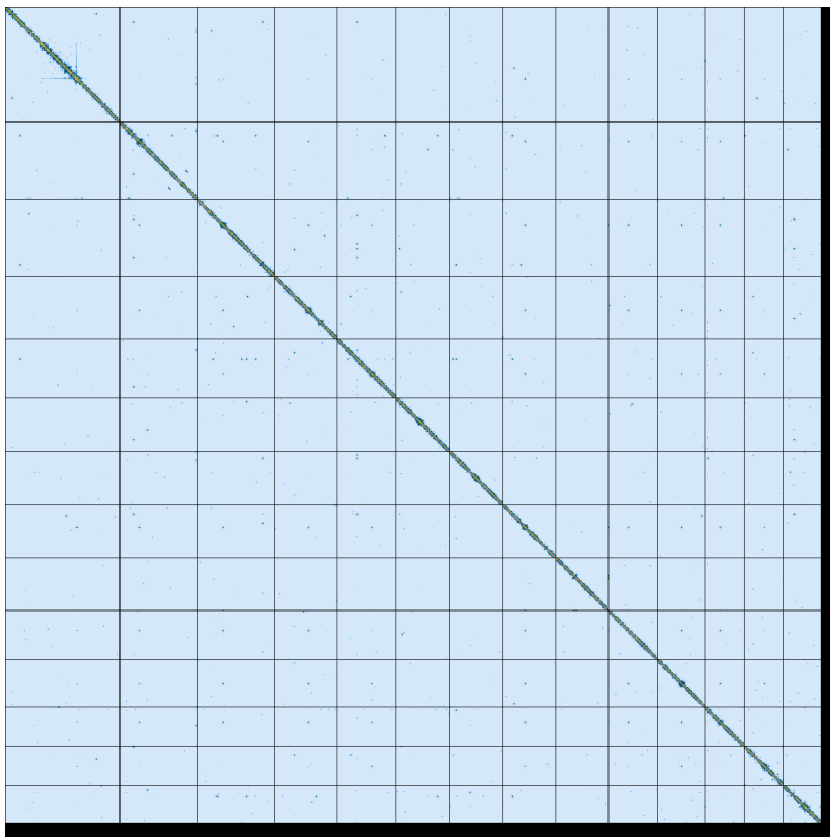
. Other observations: "The initial PacBio data set shows a very broad range of contamination/cobionts up to 14X. Therefore the PacBio reads were subsampled to 40X, which reduced the "erroneous" kmers to a normal fraction. The hifiasm assembly based on the subsampled reads was quite fragmented (4889 contigs, 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). Further contamination were filtered out by running tiara, purge_dups again, investigating the from FCS-GX review-tagged contigs, blobtoolKit, and by applying a GC filtering (>0.45) and PacBio coverage <5."

Quality metrics table

Metrics	Pre-curation pri	Curated pri
Total bp	704,195,618	554,556,895
GC %	41.01	38.98
Gaps/Gbp	0	290.32
Total gap bp	0	21,800
Scaffolds	4,889	630
Scaffold N50	4,579,535	35,719,839
Scaffold L50	45	6
Scaffold L90	1,319	13
Contigs	4,889	791
Contig N50	4,579,535	5,274,245
Contig L50	45	34
Contig L90	1,319	112
QV	54.9445	59.368
Kmer compl.	85.3463	80.5198
BUSCO sing.	94.5%	95.6%
BUSCO dupl.	2.4%	1.1%
BUSCO frag.	2.5%	2.5%
BUSCO miss.	0.6%	0.8%

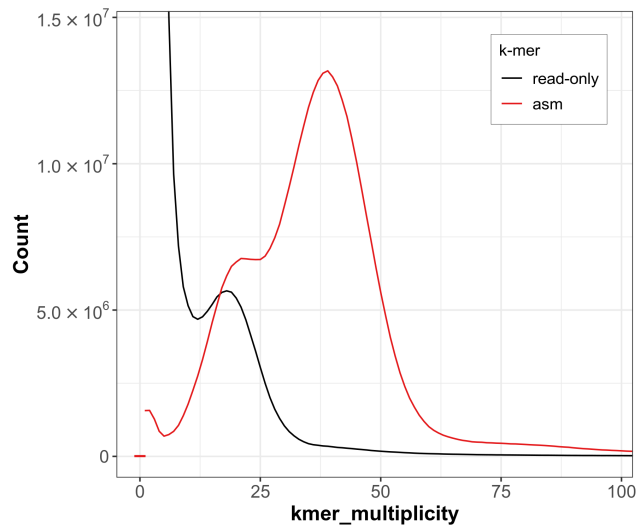
BUSCO: 5.8.3 (euk_genome_min, miniprot) / Lineage: anthozoa_odb12 (genomes:12, BUSCOs:3649)

HiC contact map of curated assembly

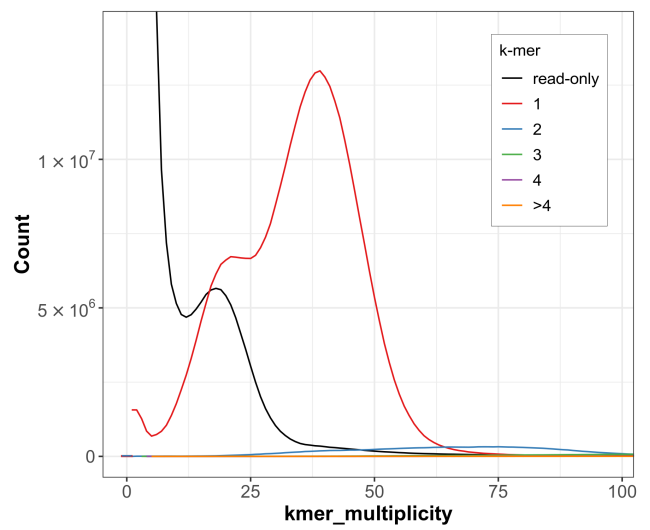


pri [\[LINK\]](#)

K-mer spectra of curated assembly

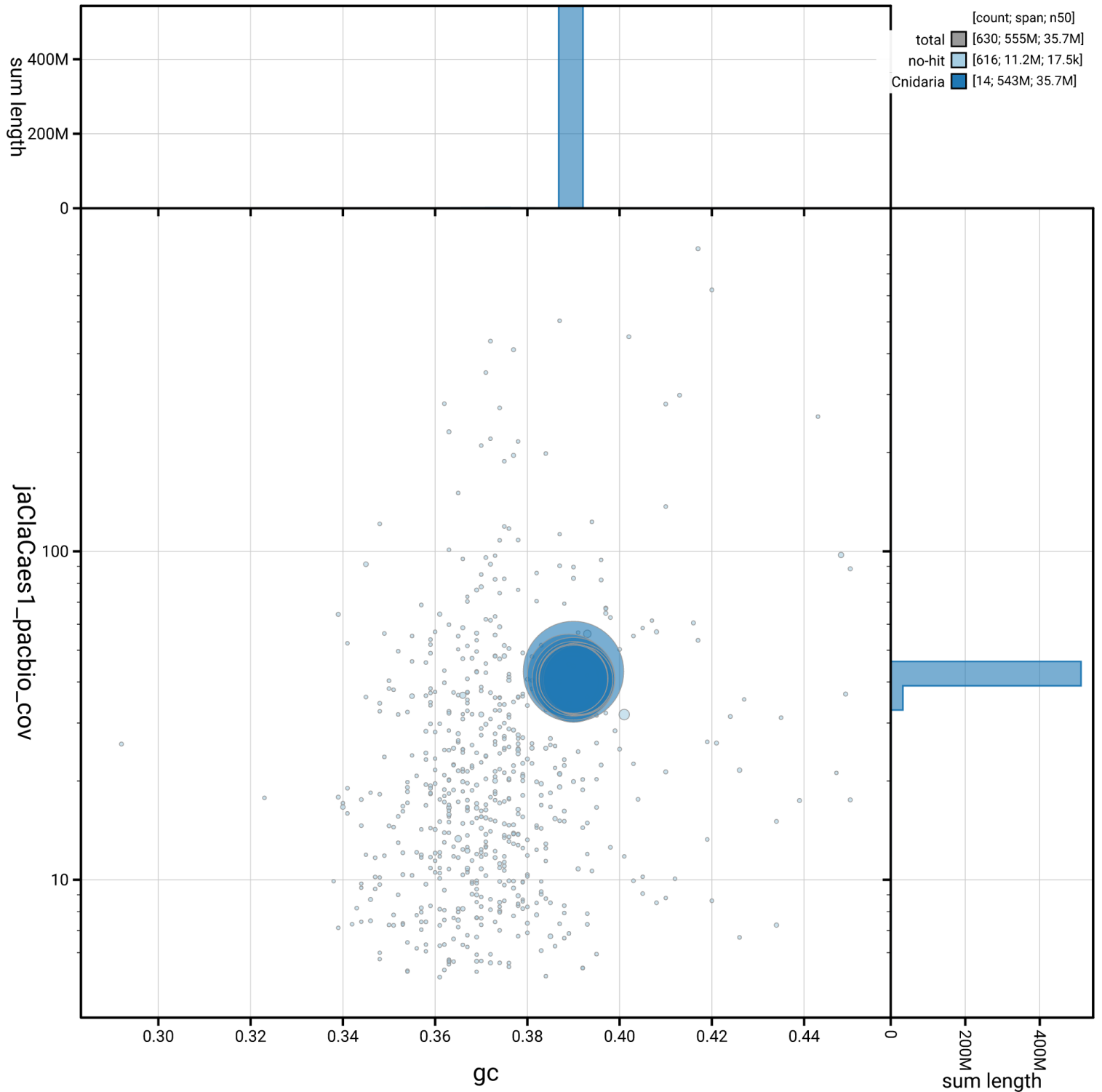


Distribution of k-mer counts coloured by their presence in reads/assemblies



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: 2025-10-09 15:10:19 CEST