

# ERGA Assembly Report

v24.10.15

Tags: ERGA-BGE

TxID	159417
ToLID	<b>kaClaLepa1</b>
Species	Clavelina lepadiformis
Class	Ascidiacea
Order	Aplousobranchia

Genome Traits	Expected	Observed
Haploid size (bp)	191,230,229	218,899,632
Haploid Number	9 (source: direct)	9
Ploidy	2 (source: direct)	2
Sample Sex	Hermaphrodite	Hermaphrodite

## EBP metrics summary and curation notes

Obtained EBP quality metric for hap1: 6.7.Q65

Obtained EBP quality metric for hap2: 6.7.Q65

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

- . Assembly length loss > 3% for hap2
- . Not 90% of assembly in chromosomes for hap2

### Curator notes

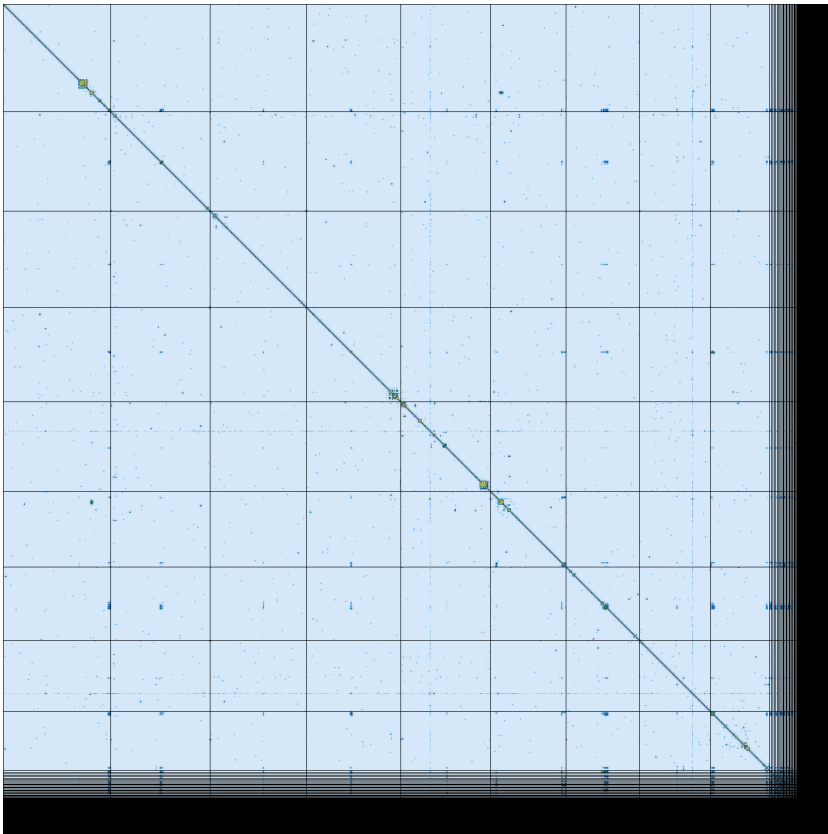
- . Interventions/Gb: 98
- . Contamination notes: "The organism was moderately contaminated. Initial screening was performed with Tiara and BlobToolKit. Those contaminants were identified in the Hi-C contact map and clustered with other contigs showing high contact frequency. These blocks, along with other contigs showing minimal contact frequency to the primary scaffolds, were removed from the assembly."
- . Other observations: "The Hifi data was subsampled to 130x, which improved contiguity and reduced contamination. The Sanger Hi-C library (ERR10501018) was used for scaffolding and manual curation after the SciLifeLab library proved insufficient. SciLifeLab data was subsequently remapped for final validation. Because the Sanger library used a different individual, signal breaks within contigs were ignored to prioritize HiFi data accuracy. The SciLifeLab library was used primarily for contamination checks. The pretext file of Sanger Hi-C against the Hifi is also in <https://export.uppmass.uu.se/snuc2021-6-194/ERGA/kaClaLepa1>"

## Quality metrics table

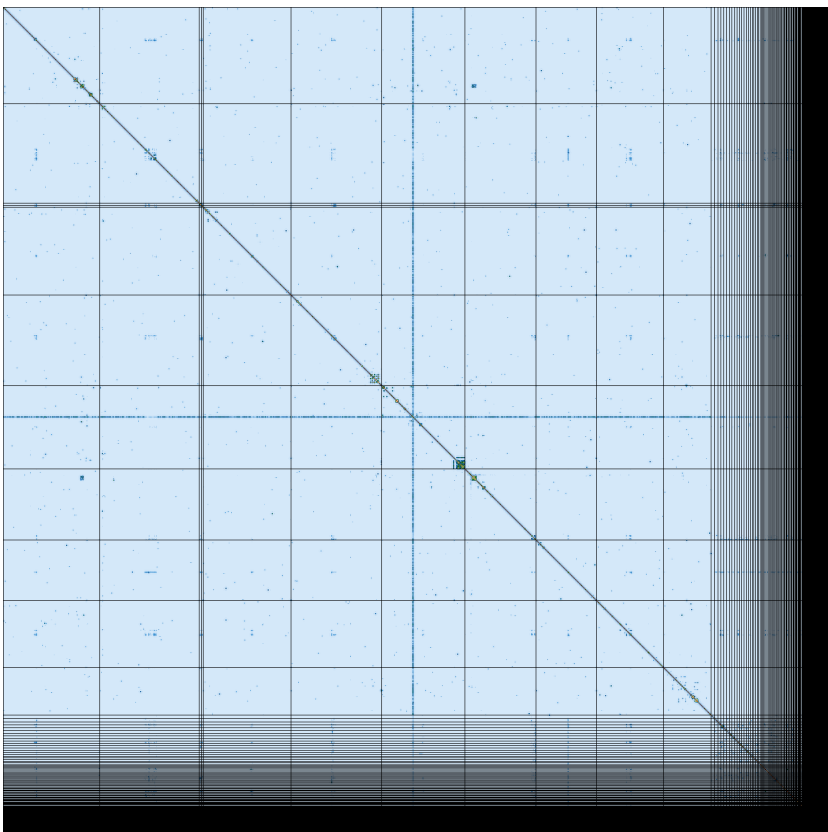
Metrics	Pre-curation hap1	Pre-curation hap2	Curated hap1	Curated hap2
Total bp	198,159,941	228,295,693	193,722,113	218,899,632
GC %	35.88	35.13	35.8	35.12
Gaps/Gbp	282.6	232.16	454.26	315.21
Total gap bp	11,200	10,600	17,600	13,800
Scaffolds	353	282	176	189
Scaffold N50	22,014,189	21,931,320	20,804,929	21,998,938
Scaffold L50	4	5	5	5
Scaffold L90	15	46	9	27
Contigs	409	335	264	258
Contig N50	3,344,678	4,310,601	3,016,126	4,930,265
Contig L50	17	17	19	16
Contig L90	70	95	65	80
QV	24.7	24.7	65.6	65.6
Kmer compl.	98.93	98.93	98.11	98.11
BUSCO sing.	93.7%	93.6%	94.2%	94.0%
BUSCO dupl.	1.5%	1.9%	0.8%	1.4%
BUSCO frag.	0.3%	0.4%	0.4%	0.5%
BUSCO miss.	4.5%	4.1%	4.5%	4.1%

BUSCO: 6.0.0 (euk\_genome\_min, miniprot) / Lineage: metazoa\_odb10 (genomes:65, BUSCOs:954)

# HiC contact map of curated assembly

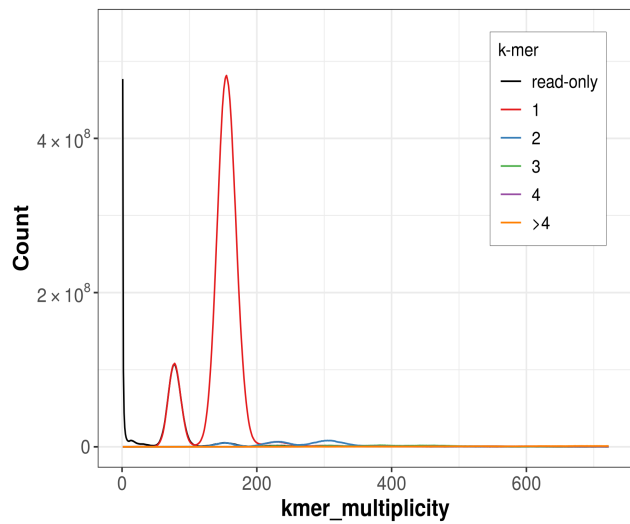


hap1 [\[LINK\]](#)

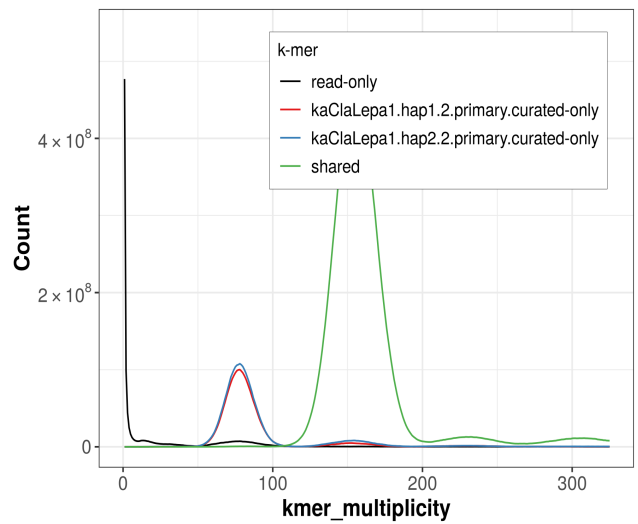


hap2 [\[LINK\]](#)

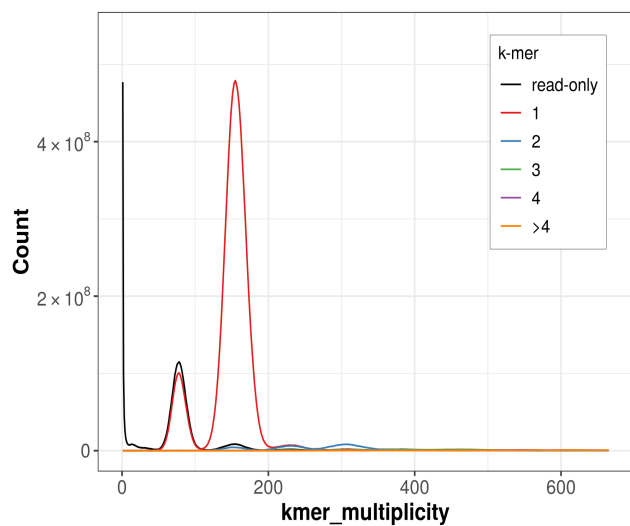
# K-mer spectra of curated assembly



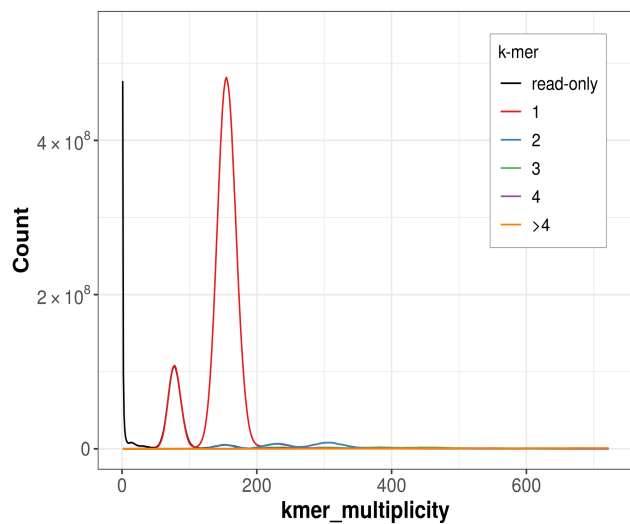
Distribution of k-mer counts per copy numbers found in asm



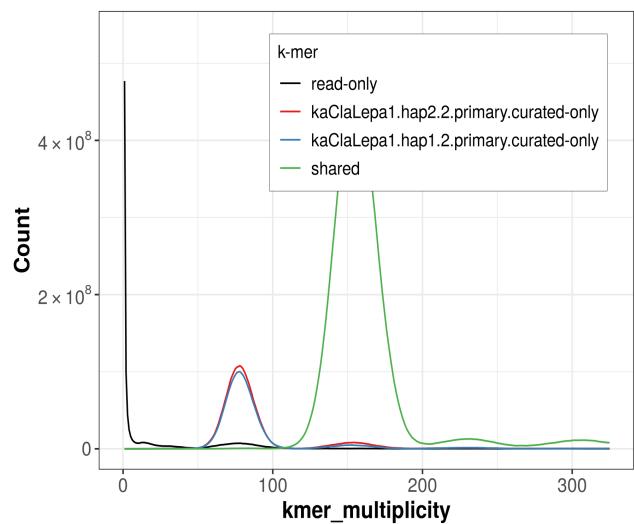
Distribution of k-mer counts coloured by their presence in reads/assemblies



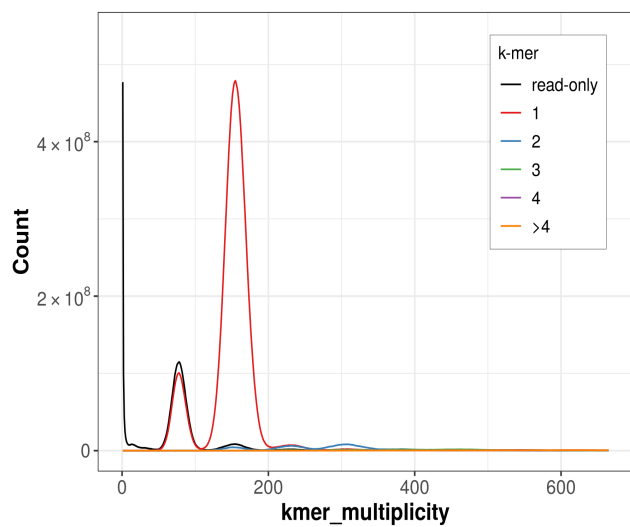
Distribution of k-mer counts per copy numbers found in asm



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

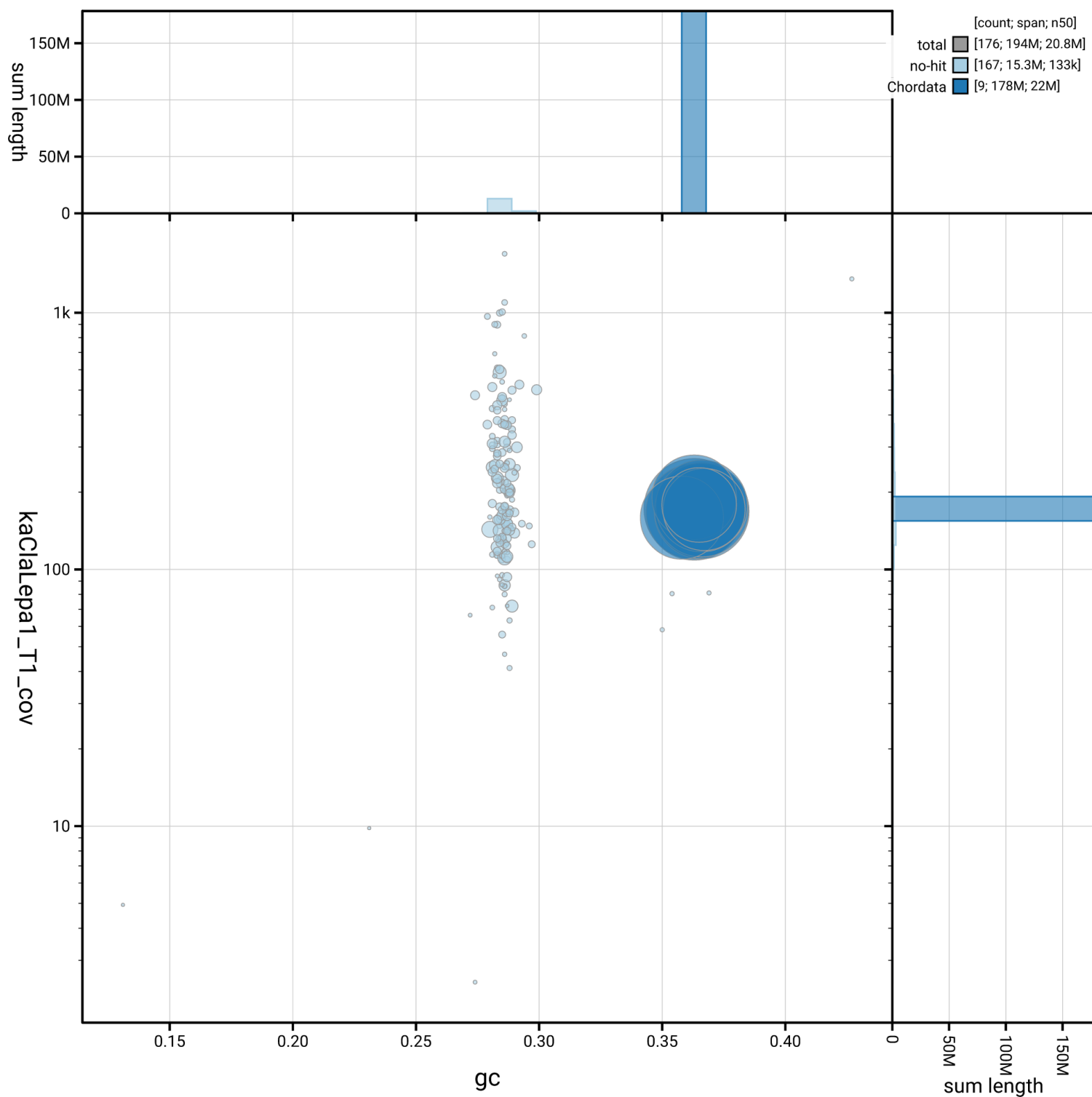


Distribution of k-mer counts coloured by their presence in reads/assemblies

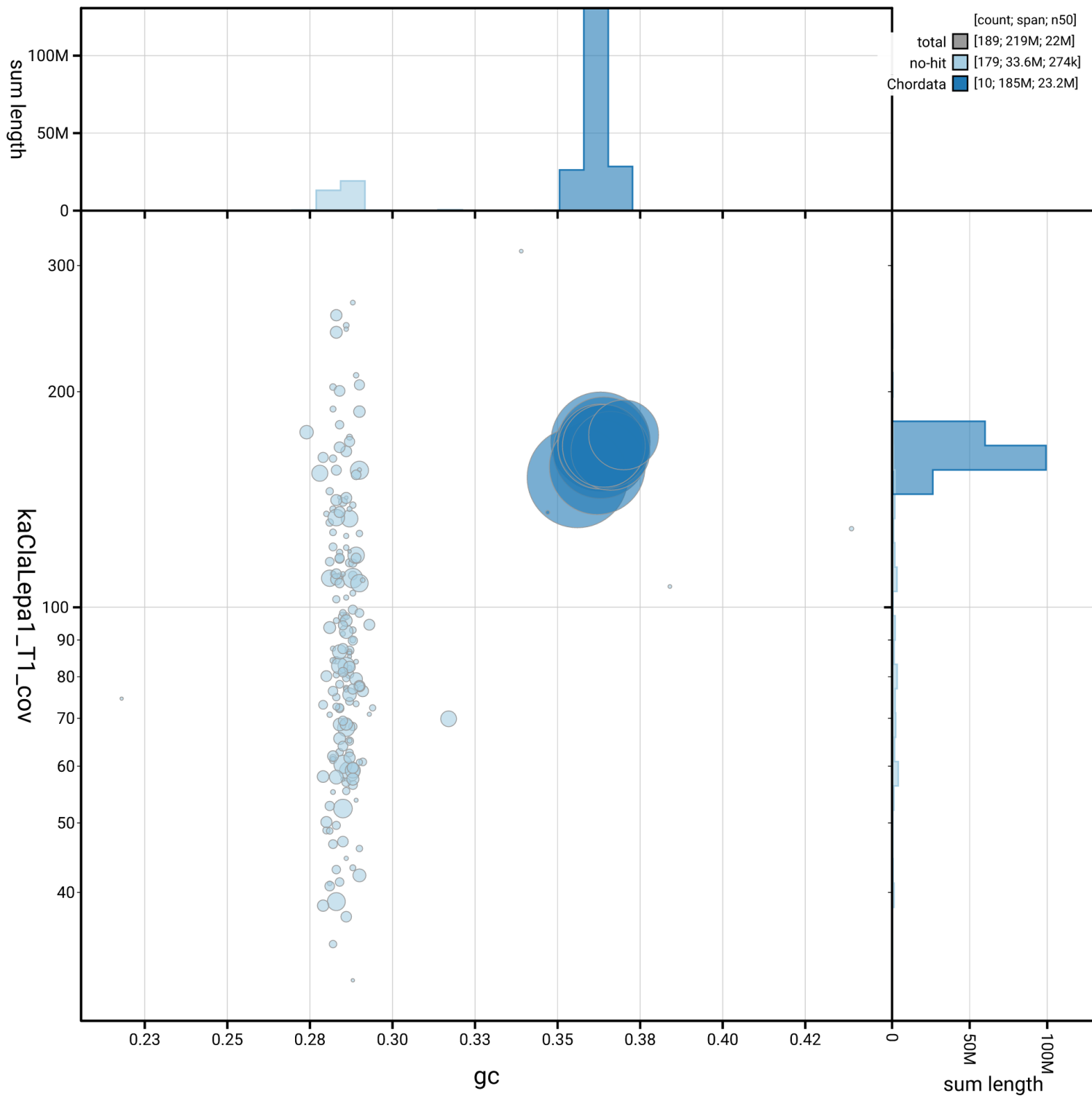


Distribution of k-mer counts per copy numbers found in asm

# Post-curation contamination screening



**hap1.** 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.



**hap2.** 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	Hi-C	Hi-C_ERR10501018
Coverage	259	226	869

## Assembly pipeline

```
- Rasusa
  | _ ver: 2.2.2
  | _ key param: --coverage 130
- NBISweden/Earth-Biogenome-Project-pilot
  | _ ver: commit:85e04ded1d
  | _ key param: --steps "inspect,preprocess,assemble,screen,purge,scaffold"
  | _ key param: --use_phased true
  | _ key param: --hifiasm_with_hic true
  | _ key param: --oatkdb "path-to-OatkDB-v20230921"
  | _ key param: --busco.lineages_db_path "path-to-busco-v5_lineage_sets"
- Hifiasm
  | _ ver: 0.25.0-r726
  | _ key param: --h1
  | _ key param: --h2
- purge_dups
  | _ ver: 1.2.5
  | _ key param: NA
- YaHS
  | _ ver: 1.2a.1
  | _ key param: --no-contig-ec
- FCSGX
  | _ ver: 0.5.4
  | _ key param: NA
- Mitohifi
  | _ ver: 3.0.0
  | _ key param: -o 13
```

## Curation pipeline

```
- sanger-tol/curationpretext
  | _ ver: 1.5.1
  | _ key param: NA
- sanger-tol/blobtoolkit
  | _ ver: 0.9.0
  | _ key param: NA
- PretextView
  | _ ver: 1.0.4
  | _ key param: NA
- sanger-tol/agp-tpf-utils
  | _ ver: commit: 35ea45dbcd8f3ece7a91104e5813541b55beb2ff
  | _ key param: NA
- sanger-tol/sanger-tol-ear
  | _ ver: commit: 2b15915ec8
  | _ key param: NA
```



Submitter: Mahesh Binzer-Panchal

Affiliation: SciLifeLab

Date and time: 2026-02-09 09:39:05 CET