



ERGA Community Genome Review Certificate

Species: *Citrullus colocynthis*

ToLID: ddCitColo1

Taxonomy ID: 252529

Submitted by: Anestis Gkanogiannis

Affiliation: International Center for Biosaline Agriculture
(ICBA)

Metadata

Sample metadata have been reviewed and approved by Christian de Guttery (Swiss Institute of Bioinformatics - SIB) as meeting the mandatory requirements of the ERGA manifest v2.5.

ENA: <https://www.ebi.ac.uk/ena/browser/view/SAMEA114392512>

Genome Assembly

The genome assembly was reviewed and approved as meeting Earth BioGenome Project (EBP) criteria by Diego De Panis (Leibniz-Institut für Zoo- und Wildtierforschung), and the process supervised by Tom Brown (Leibniz-Institut für Zoo- und Wildtierforschung).

EAR: <https://github.com/ERGA-consortium/EARs/pull/46>

ERGA Assembly Report

v24.04.03_beta

Tags: ERGA-Satellite

TxID	252529
ToLID	ddCitColo1
Species	Citrullus colocynthis
Class	Magnoliopsida
Order	Cucurbitales

Genome Traits	Expected	Observed
Haploid size (bp)	346,730,326	366,112,299
Haploid Number	12 (source: direct)	11
Ploidy	2 (source: ancestor)	2
Sample Sex	NA	NA

EBP metrics summary and curation notes

Obtained EBP quality metric for hap1: 7.7.Q67

Obtained EBP quality metric for hap2: 7.7.Q66

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

- . Observed Haploid Number is different from Expected

Curator notes

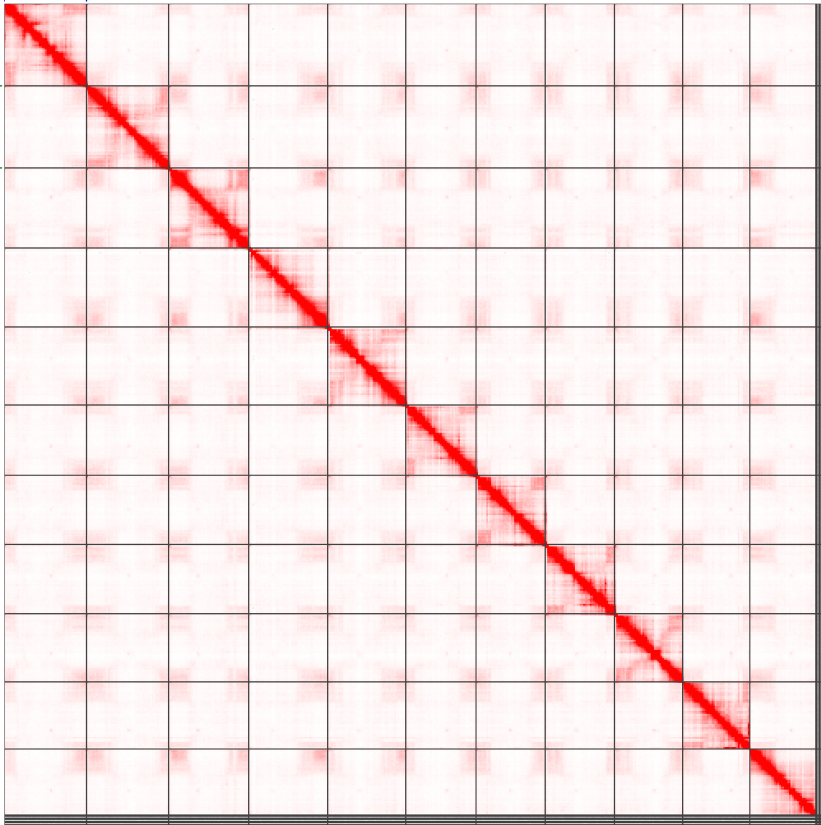
- . Interventions/Gb: None
- . Contamination notes: "Removed a 43kbp contaminant contig from hap1 assembly."
- . Other observations: "True haploid number is 11."

Quality metrics table

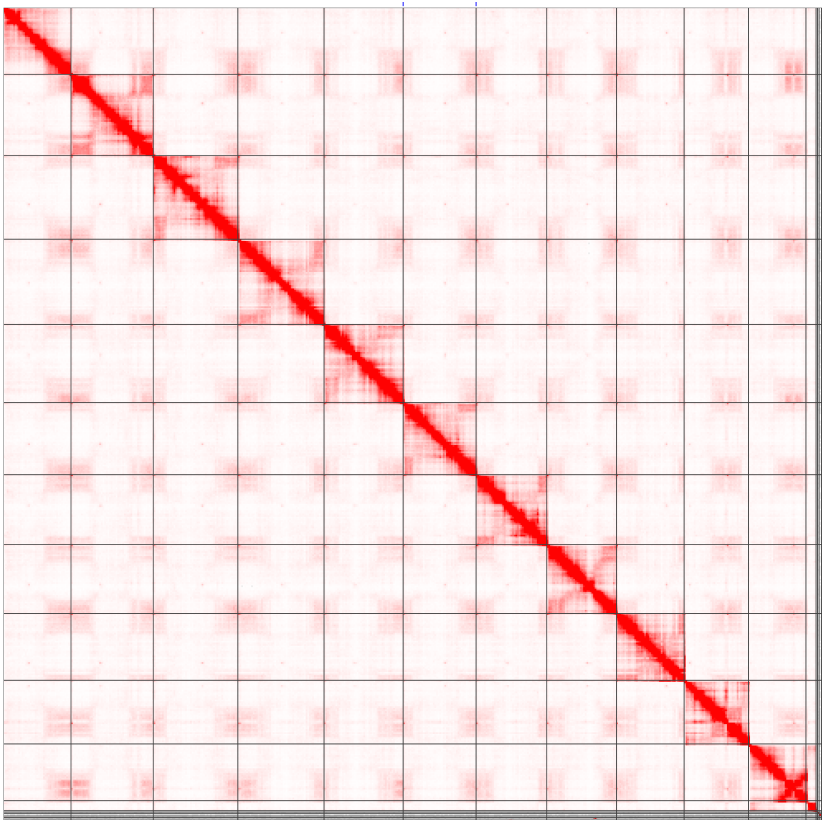
Metrics	Pre-curation hap1	Pre-curation hap2	Curated hap1	Curated hap2
Total bp	366,111,899	356,939,597	366,112,299	356,935,950
GC %	33.6	33.64	33.6	33.64
Gaps/Gbp	0	0	5.46	25.21
Total gap bp	0	0	400	1,800
Scaffolds	88	95	86	88
Scaffold N50	30,853,444	29,357,918	31,341,397	30,674,282
Scaffold L50	6	6	6	6
Scaffold L90	11	14	10	10
Contigs	88	95	88	97
Contig N50	30,853,444	29,357,918	30,853,444	29,345,706
Contig L50	6	6	6	6
Contig L90	11	14	11	14
QV	67.0945	66.9325	67.0945	66.9325
Kmer compl.	99.7458	97.7577	99.7458	97.7577
BUSCO sing.	95.7%	94.5%	95.7%	94.5%
BUSCO dupl.	1.0%	1.2%	1.0%	1.2%
BUSCO frag.	0.5%	0.6%	0.6%	0.6%
BUSCO miss.	2.8%	3.7%	2.7%	3.7%

BUSCO 3.0.2 Lineage: eudicotyledons_odb10 (genomes:31, BUSCOs:2326)

HiC contact map of curated assembly

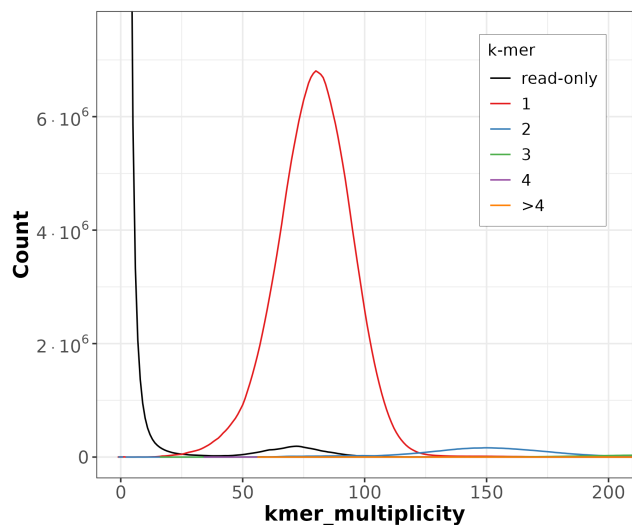


hap1 [\[LINK\]](#)

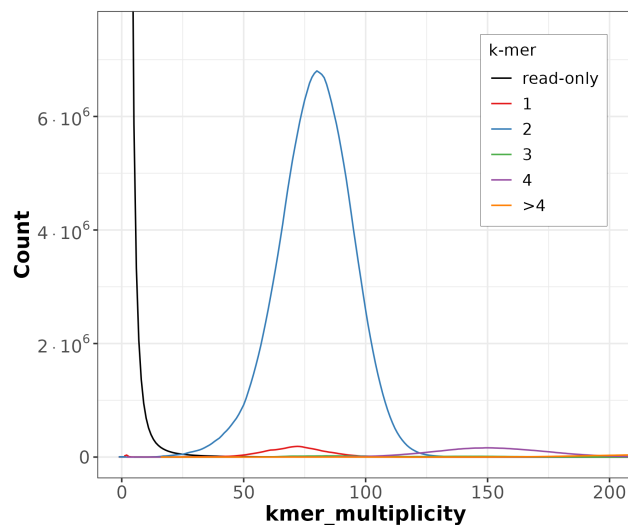


hap2 [\[LINK\]](#)

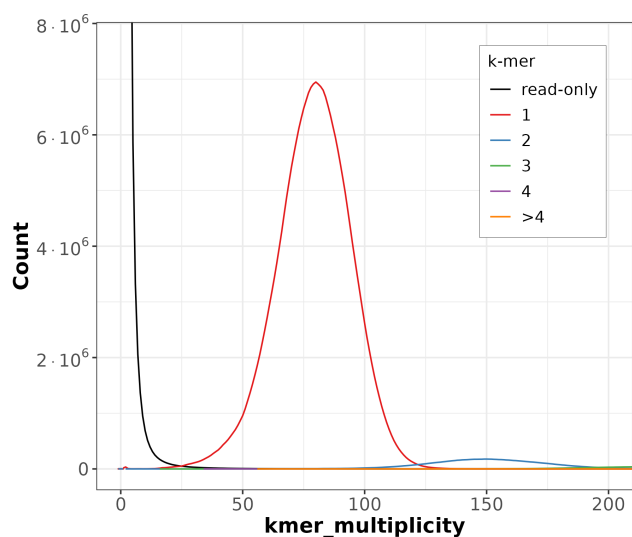
K-mer spectra of curated assembly



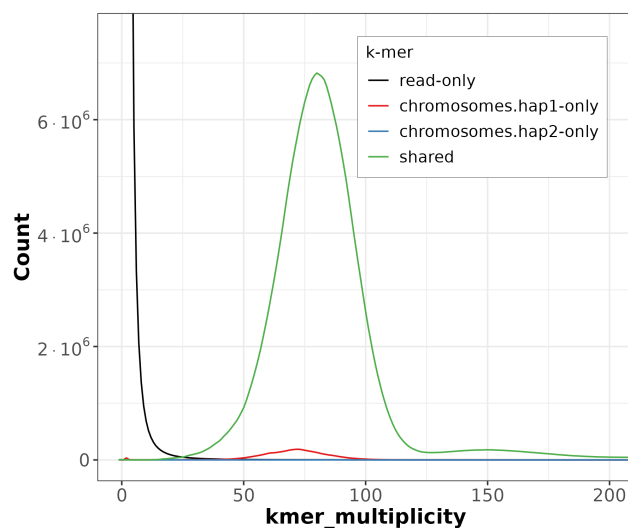
Distribution of k-mer counts per copy numbers found in **hap2** (hapl.)



Distribution of k-mer counts per copy numbers found in **asm** (dipl.)

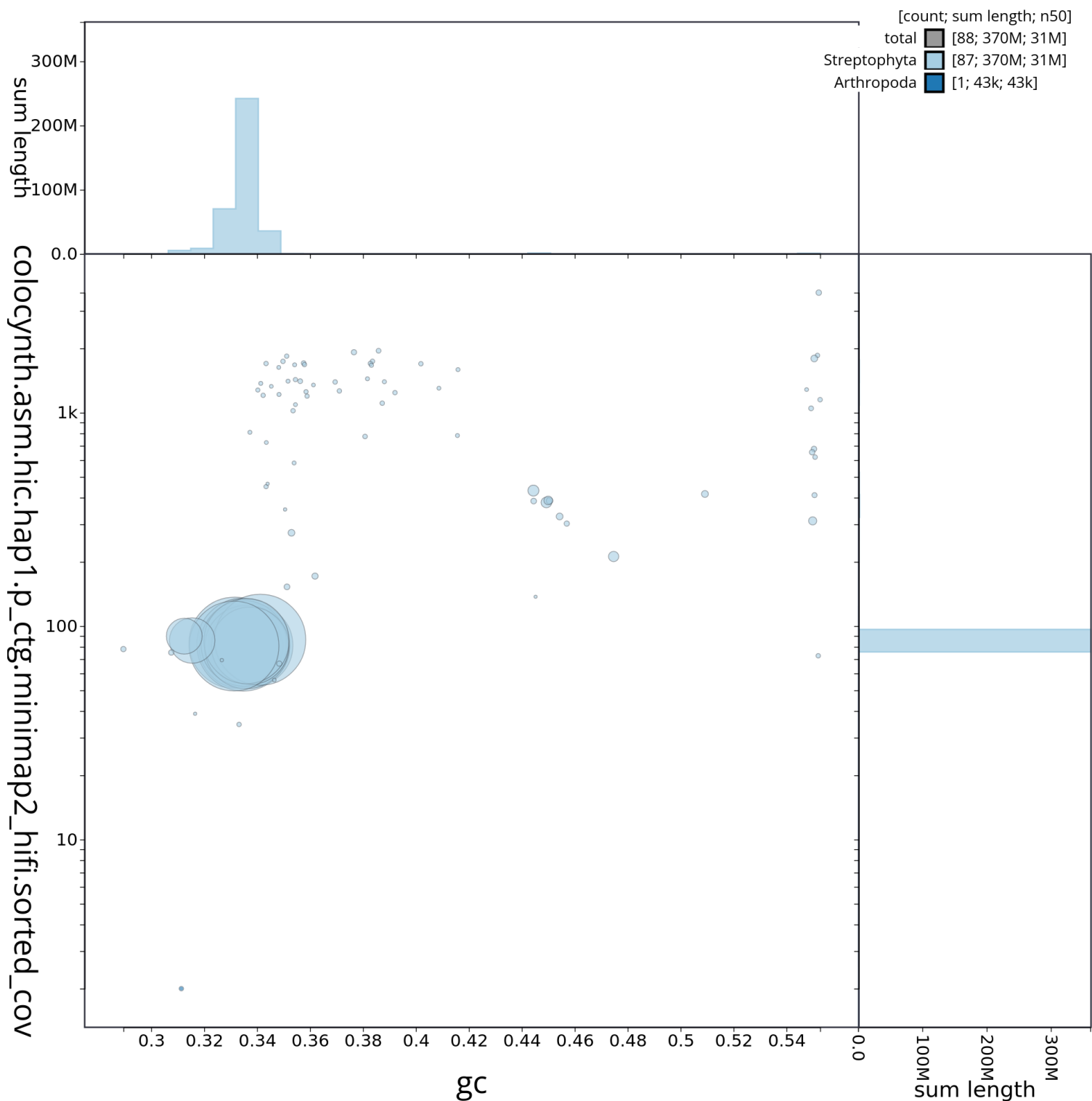


Distribution of k-mer counts per copy numbers found in **hap1** (hapl.)

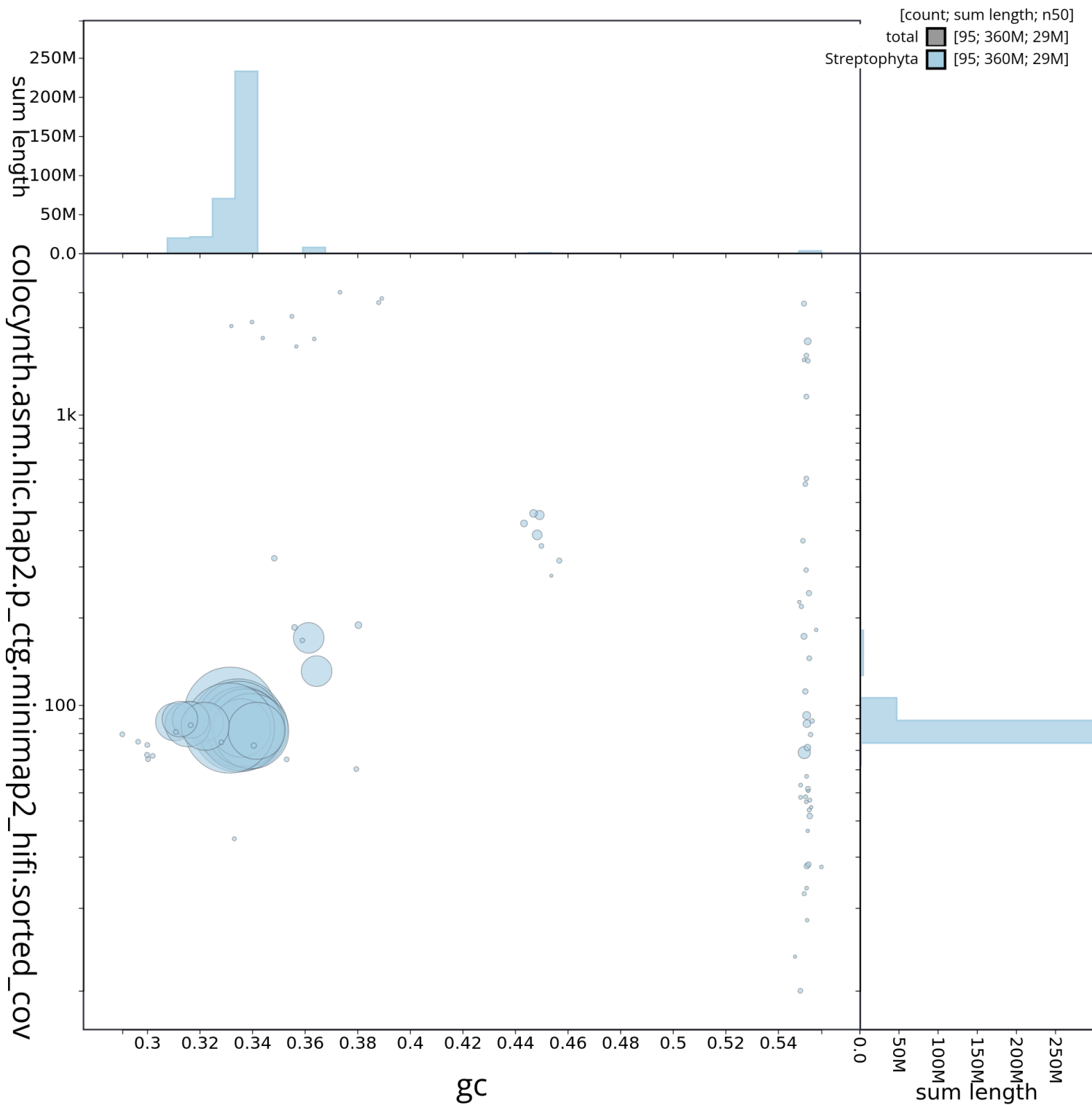


Distribution of k-mer counts coloured by their presence in reads/assemblies

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-HiFi	Illumina-HiC	DNBSEQ-G400-WGS
Coverage	85x	120x	260x

Assembly pipeline

- **Hifiasm**
 - |_ *ver*: 0.18.9-r527
 - |_ *key param*: NA

Curation pipeline

- **YaHS**
 - |_ *ver*: 1.2a.1
 - |_ *key param*: NA
- **JBAT**
 - |_ *ver*: 2.20.00
 - |_ *key param*: NA

Submitter: Anestis Gkanogiannis
Affiliation: International Center for Biosaline Agriculture (ICBA)
Date and time: 2024-07-04 14:03:42 CEST