

# ERGA Assembly Report

v24.04.03\_beta

Tags: ERGA-Pilot

TxID	54556
ToLID	<b>fAlbAlb1</b>
Species	Alburnus alburnus
Class	Actinopteri
Order	Cypriniformes

Genome Traits	Expected	Observed
Haploid size (bp)	1,059,117,179	986,562,608
Haploid Number	25 (source: direct)	25
Ploidy	2 (source: ancestor)	2
Sample Sex	unknown	unknown

## EBP metrics summary and curation notes

Obtained EBP quality metric for hap1: 6.7.Q61

Obtained EBP quality metric for hap2: 6.7.Q60

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

- . Kmer completeness value is less than 90 for hap1
- . Kmer completeness value is less than 90 for hap2

### Curator notes

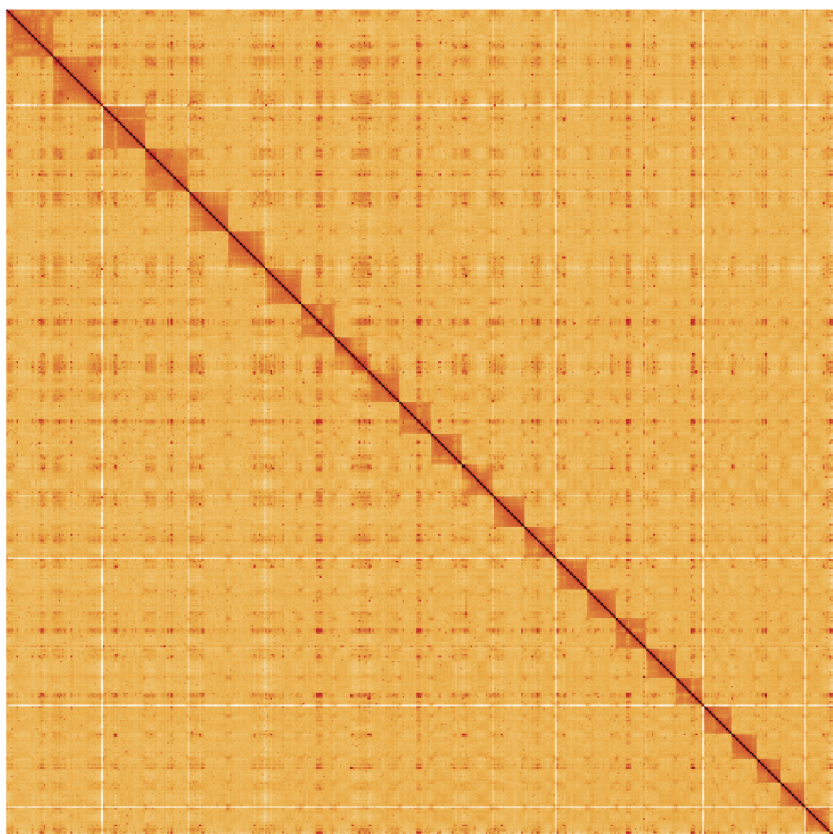
- . Interventions/Gb: None
- . Contamination notes: "No contaminants detected by FCS-GX. Those highlighted by blobtools are very short, repetitive contigs, which FCS either masked or marked as low-coverage hits to other taxa."
- . Other observations: "Curation performed a long time ago (3 years). No information on pre-curation assembly available, so the post-curation statistics have been copied to comply with the EAR."

## Quality metrics table

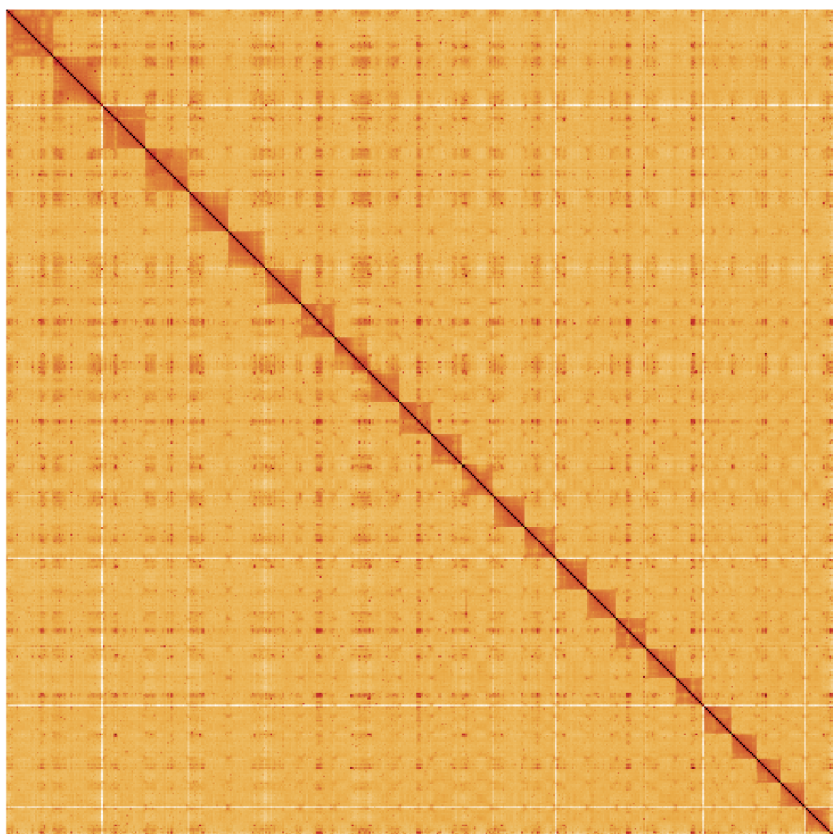
Metrics	Pre-curation hap1	Pre-curation hap2	Curated hap1	Curated hap2
Total bp	1,005,737,902	986,562,608	1,005,737,902	986,562,608
GC %	38.64	38.63	38.64	38.63
Gaps/Gbp	218.74	184.48	218.74	184.48
Total gap bp	87,380	19,460	87,380	19,460
Scaffolds	112	55	112	55
Scaffold N50	38,752,591	38,024,220	38,752,591	38,024,220
Scaffold L50	11	11	11	11
Scaffold L90	22	22	22	22
Contigs	332	237	332	237
Contig N50	7,667,869	8,875,646	7,667,869	8,875,646
Contig L50	39	34	39	34
Contig L90	139	112	139	112
QV	61.895	60.9924	61.895	60.9924
Kmer compl.	76.9516	75.4906	76.9516	75.4906
BUSCO sing.	97.3%	97.1%	97.3%	97.1%
BUSCO dupl.	0.7%	0.6%	0.7%	0.6%
BUSCO frag.	0.5%	0.4%	0.5%	0.4%
BUSCO miss.	1.5%	1.9%	1.5%	1.9%

BUSCO 5.7.1 Lineage: actinopterygii\_odb10 (genomes:26, BUSCOs:3640)

# HiC contact map of curated assembly

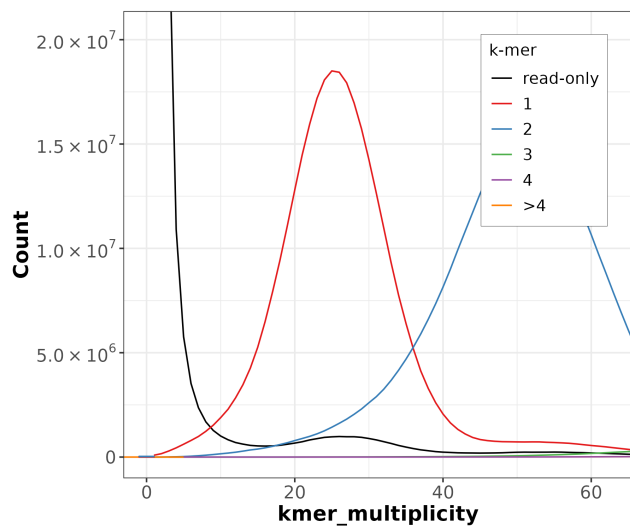


hap1 [\[LINK\]](#)

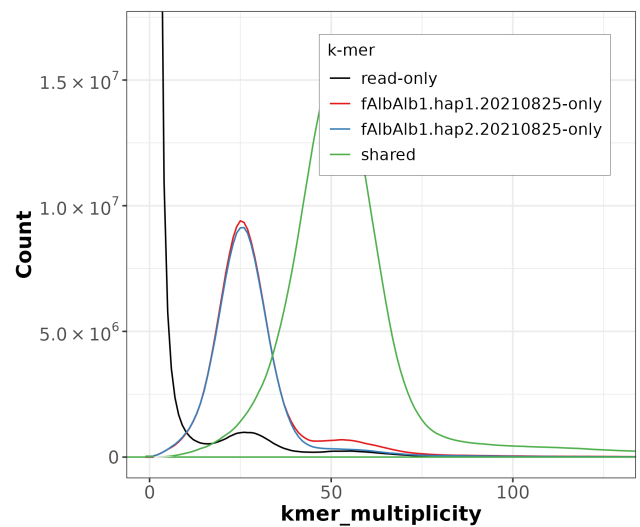


hap2 [\[LINK\]](#)

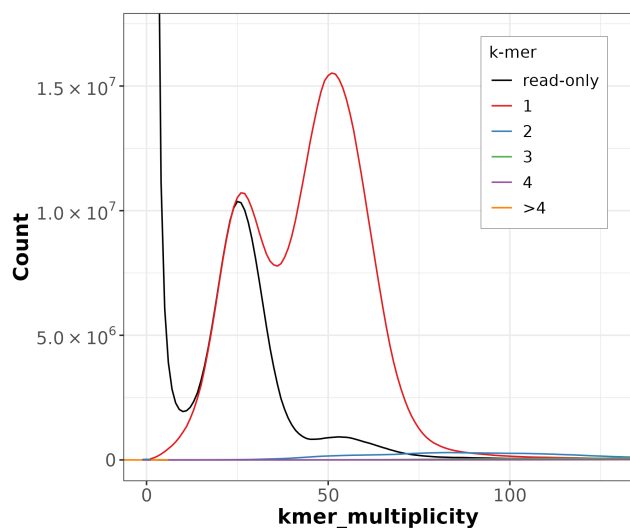
# K-mer spectra of curated assembly



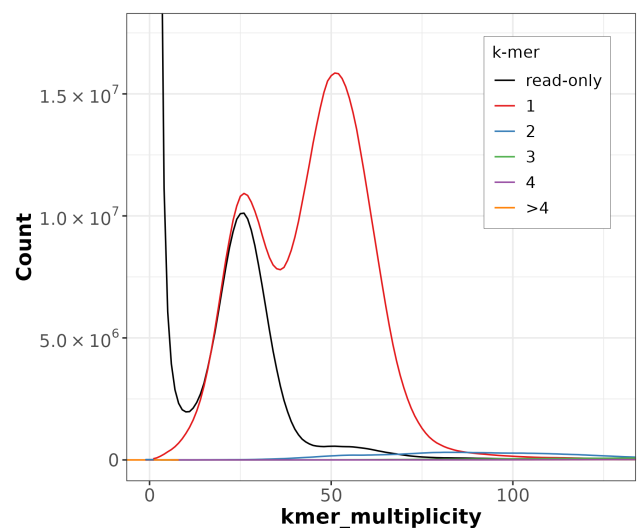
Distribution of k-mer counts per copy numbers found in asm (dipl.)



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

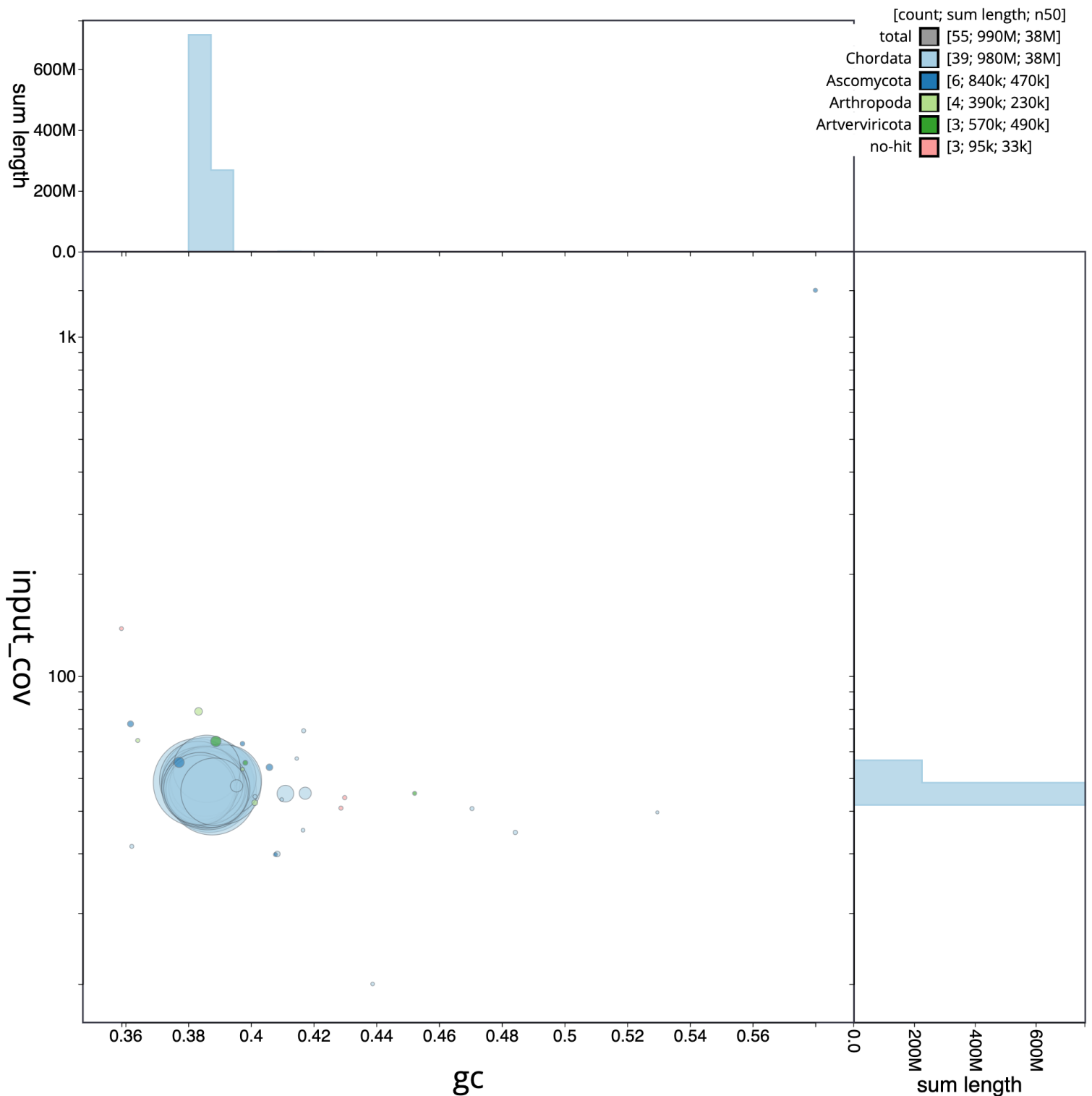


Distribution of k-mer counts per copy numbers found in **hap2** (hapl.)

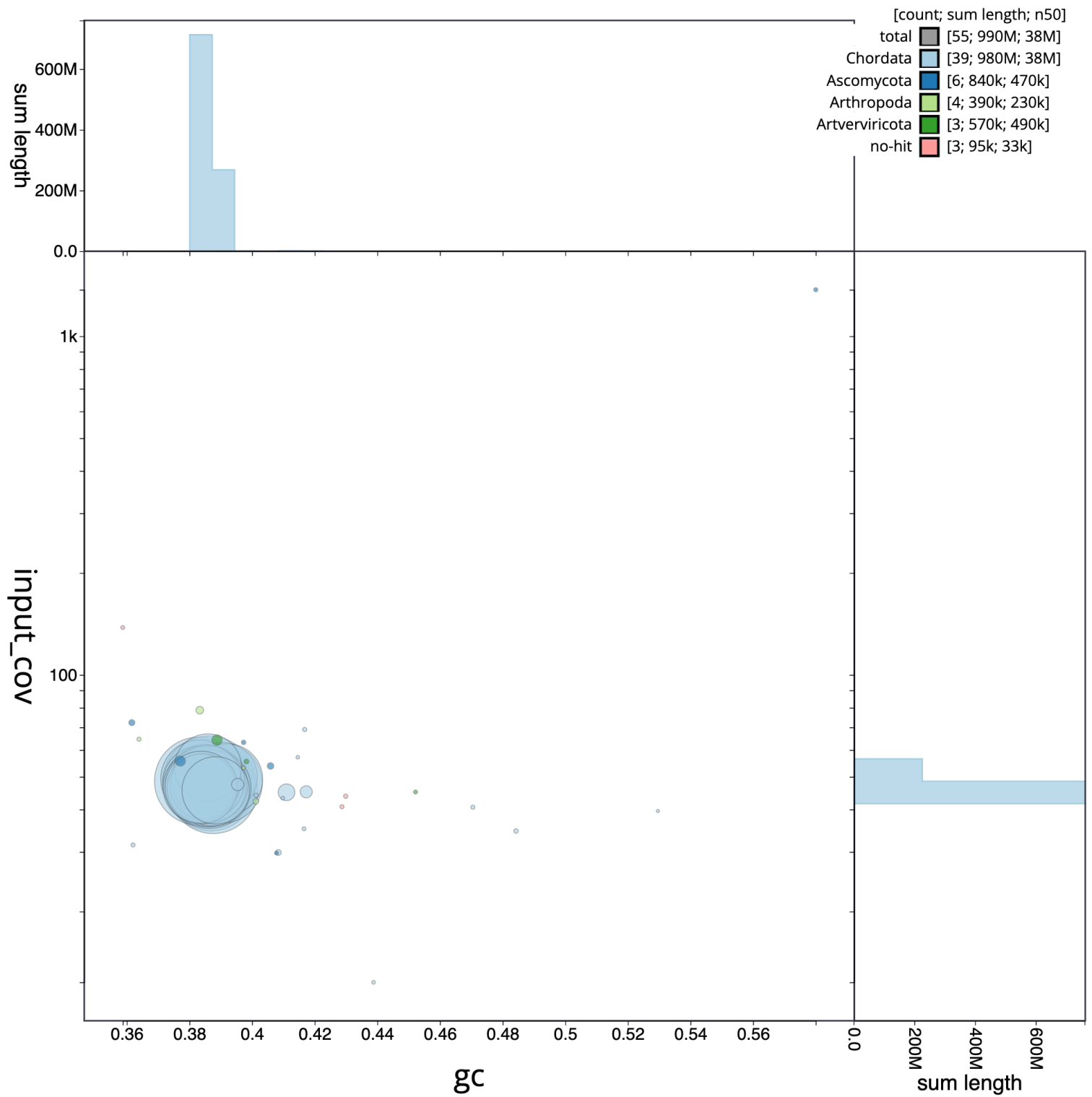


Distribution of k-mer counts per copy numbers found in **hap1** (hapl.)

# 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	Arima HiC
Coverage	50	50

## Assembly pipeline

## Curation pipeline

- **hifiasm**
  - |\_ *ver:*
  - |\_ *key param:* NA
- **salsa2**
  - |\_ *ver:*
  - |\_ *key param:* NA
- **higlass**
  - |\_ *ver:*
  - |\_ *key param:* NA

Submitter: Tom Brown

Affiliation: Max Planck Institute for Molecular Cell Biology and Genetics

Date and time: 2024-05-31 16:30:21 CEST