

ERGA Assembly Report

v24.04.03_beta

Tags: ERGA-Pilot

TxID	71183
ToLID	fAnaHis1
Species	Anaecypris hispanica
Class	Actinopteri
Order	Cypriniformes

Genome Traits	Expected	Observed
Haploid size (bp)	968,863,844	984,878,810
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.Q60

Obtained EBP quality metric for hap2: 6.7.Q59

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

Curator notes

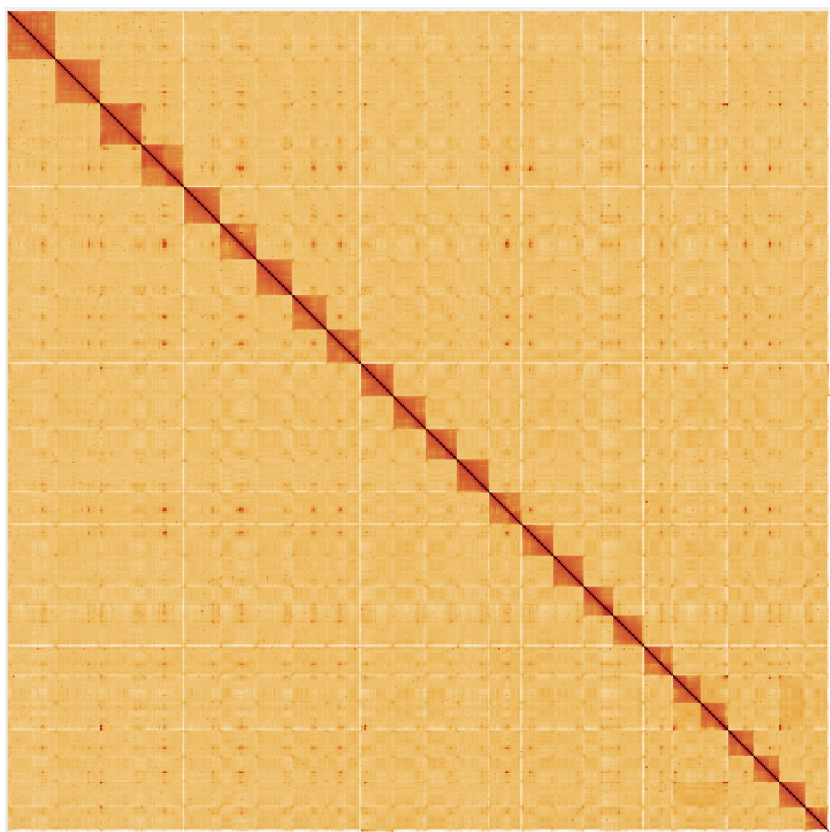
- . Interventions/Gb: None
- . Contamination notes: "No contaminants detected by FCS-GX. Those highlighted by blobtools are very short, repetitive contigs, which FCS either masked or marked as low-coverage hits to other taxa."
- . Other observations: "Curation performed a long time ago (3 years). No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

Quality metrics table

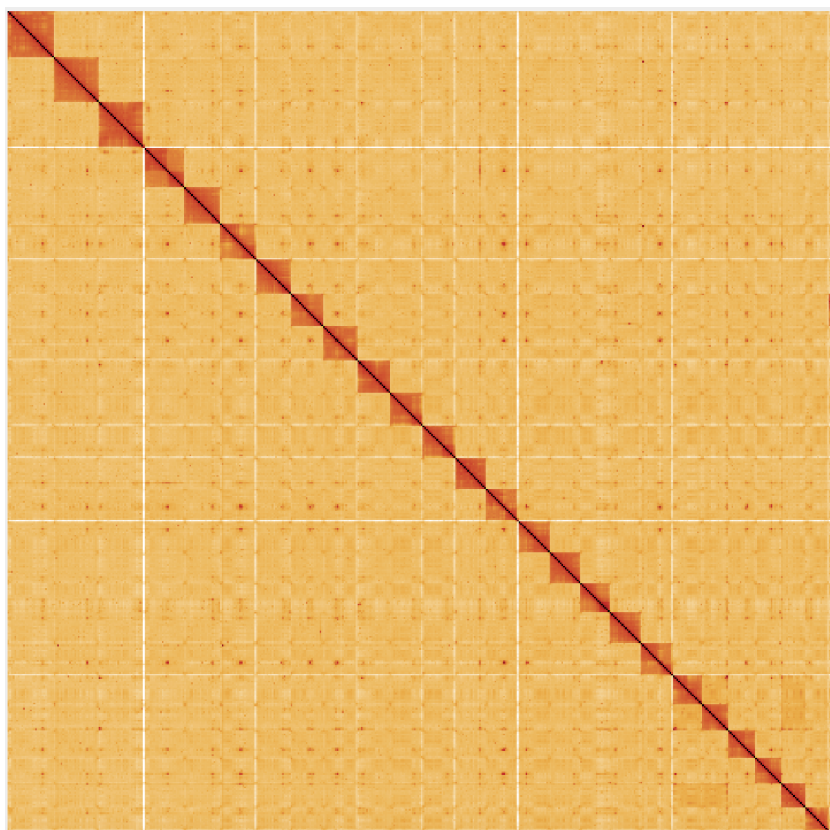
Metrics	Pre-curation hap1	Pre-curation hap2	Curated hap1	Curated hap2
Total bp	983,738,696	984,878,810	983,738,696	984,878,810
GC %	38.78	38.78	38.78	38.78
Gaps/Gbp	471.67	349.28	471.67	349.28
Total gap bp	191,962	36,080	191,962	36,080
Scaffolds	52	41	52	41
Scaffold N50	38,642,748	38,563,620	38,642,748	38,563,620
Scaffold L50	11	11	11	11
Scaffold L90	22	22	22	22
Contigs	516	385	516	385
Contig N50	4,199,513	5,445,377	4,199,513	5,445,377
Contig L50	75	57	75	57
Contig L90	262	189	262	189
QV	60.7852	59.6049	60.7852	59.6049
Kmer compl.	91.6665	91.5819	91.6665	91.5819
BUSCO sing.	97.4%	97.5%	97.4%	97.5%
BUSCO dupl.	0.7%	0.6%	0.7%	0.6%
BUSCO frag.	0.7%	0.6%	0.7%	0.6%
BUSCO miss.	1.2%	1.3%	1.2%	1.3%

BUSCO 5.7.1 Lineage: actinopterygii_odb10 (genomes:26, BUSCOs:3640)

HiC contact map of curated assembly

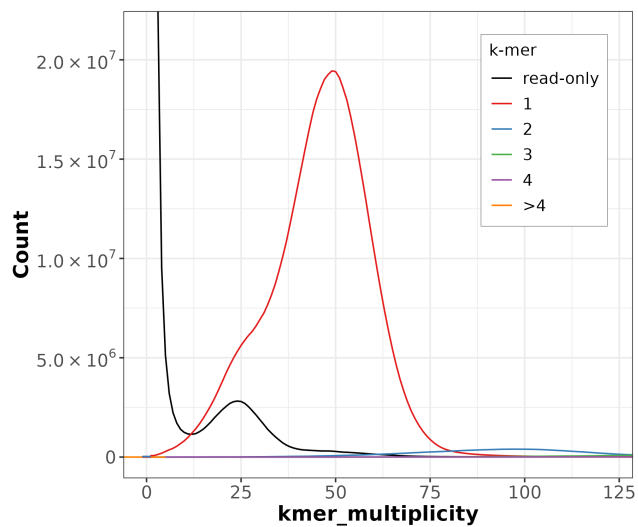


hap1 [\[LINK\]](#)

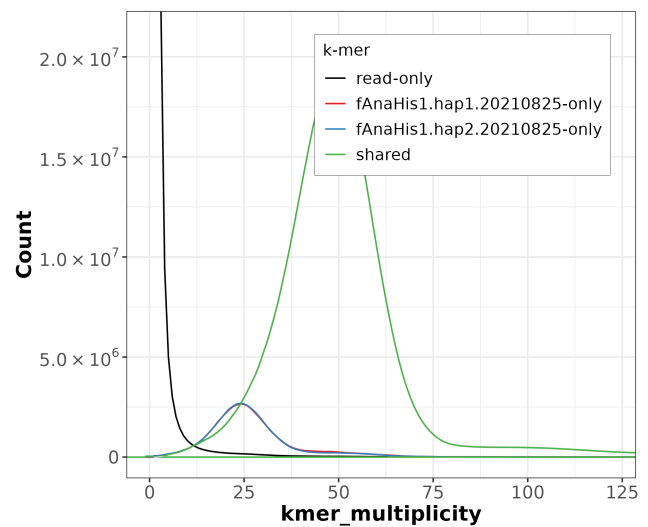


hap2 [\[LINK\]](#)

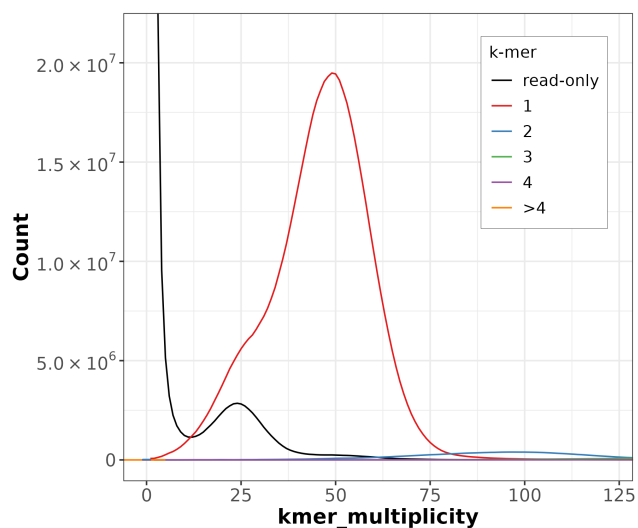
K-mer spectra of curated assembly



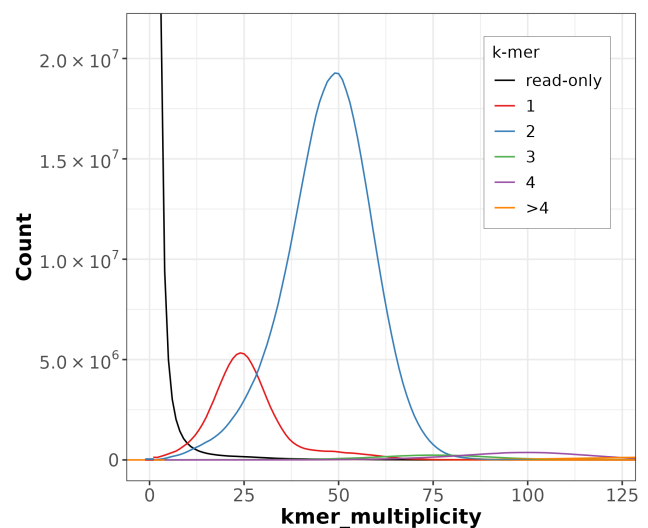
Distribution of k-mer counts per copy numbers found in **hap2** (hap1.)



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

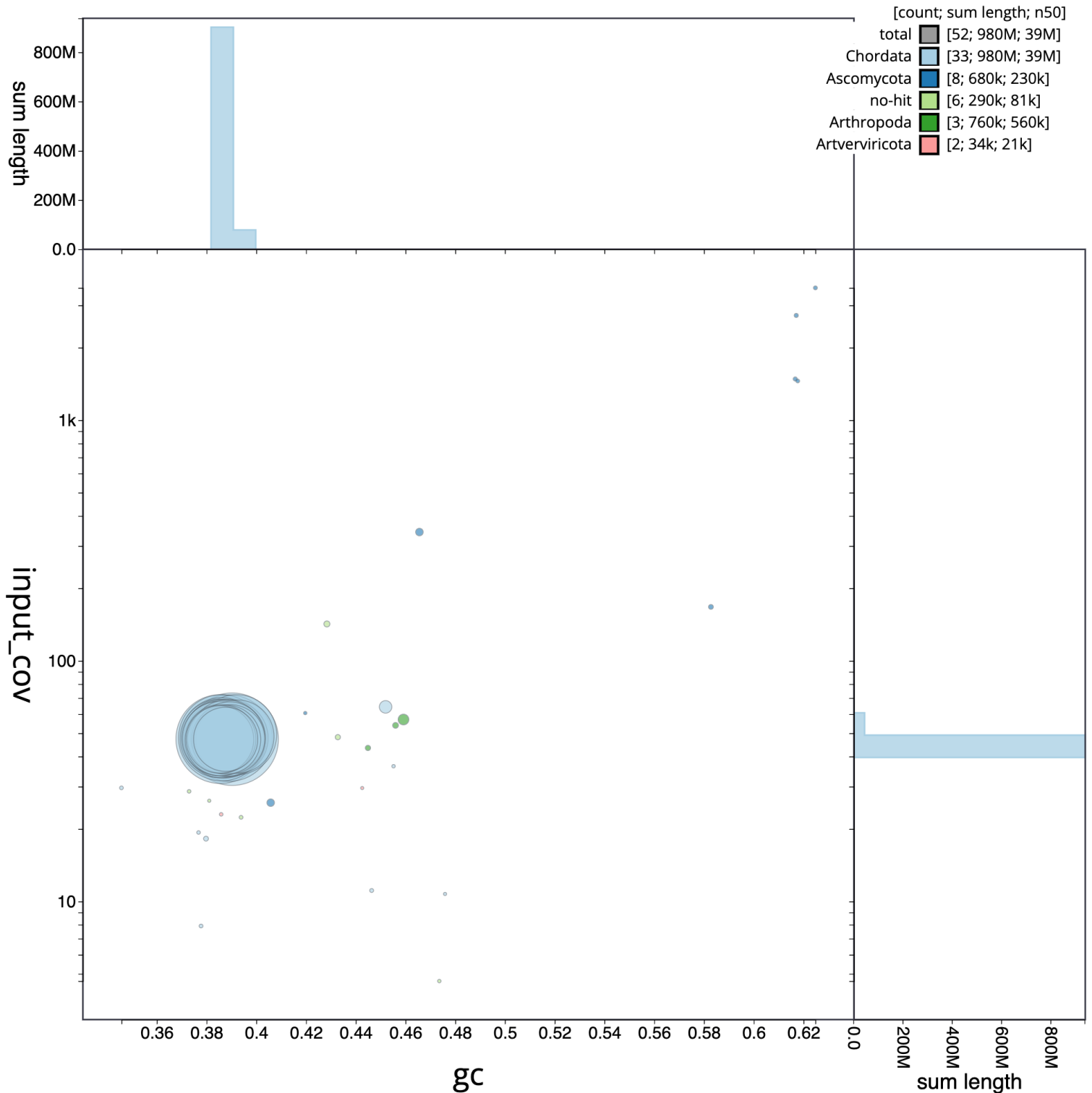


Distribution of k-mer counts per copy numbers found in **hap1** (hap1.)

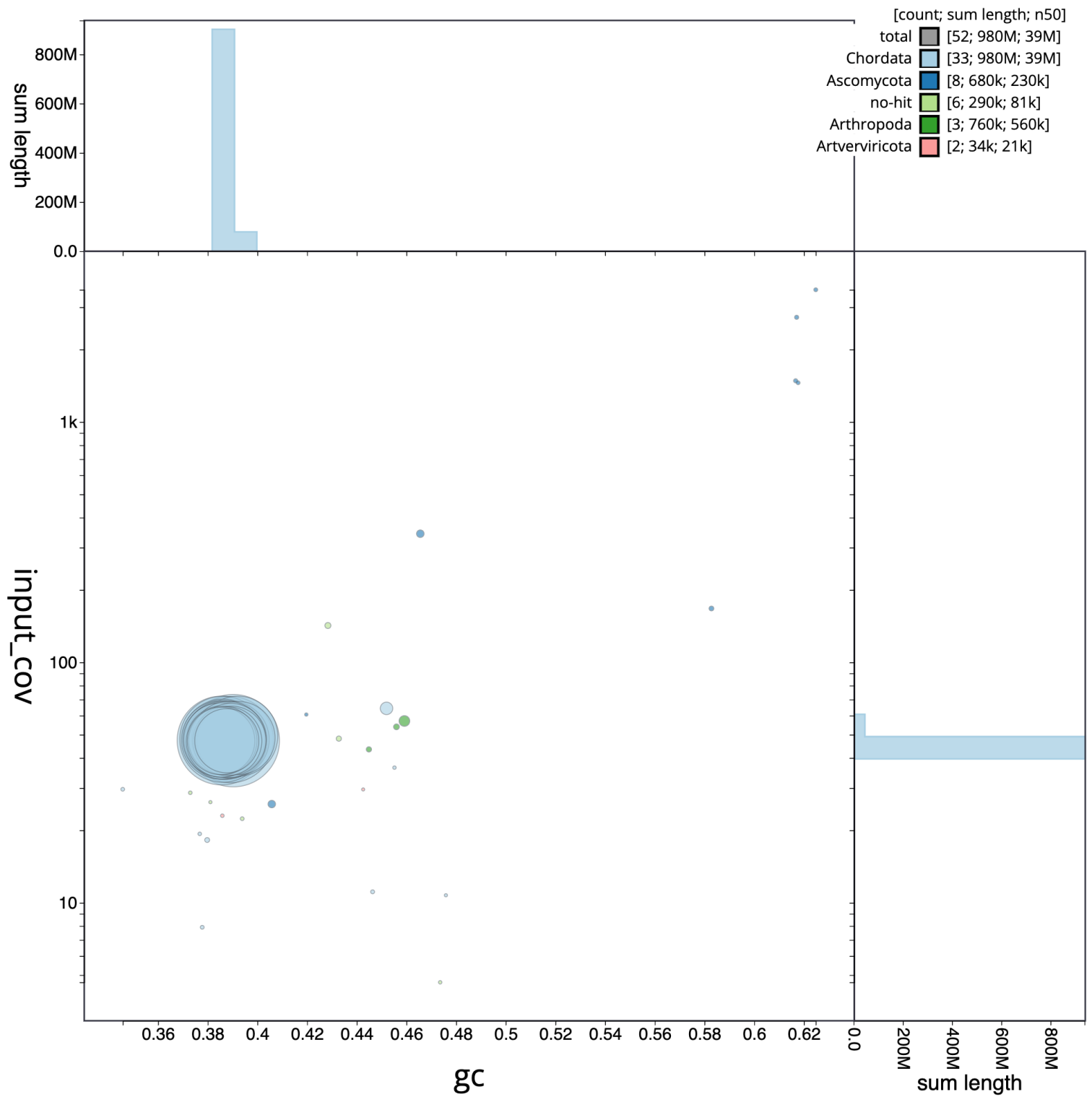


Distribution of k-mer counts per copy numbers found in **asm** (dipl.)

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: 2024-05-28 10:32:04 CEST