ERGA Assembly Report

v24.10.15

Tags: ERGA-BGE

TxID	928672
ToLID	**bPyrMur1**
Species	Pyrrhula murina
Class	Aves
Order	Passeriformes

Genome Traits	Expected	Observed
Haploid size (bp)	1,085,438,021	1,163,867,538
Haploid Number	39 (source: ancestor)	41
Ploidy	2 (source: ancestor)	2
Sample Sex	ZW	ZW

EBP metrics summary and curation notes

Obtained EBP quality metric for hap1: 7.7.Q71

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: 4
. Contamination notes: "No contaminants found."
. Other observations: "Haploid number within normal range. No reliable signal to join possible chromosome arms. Will produce an MQ0 map if necessary, but we don't think it is needed. Made three breaks (yahs joined four contigs that were T2T into one scaffold) and two joins (had to invert one contig in SUPER_32 based on contact map of primary assembly, which had no gap in this chromosome). Placed a few unlocs. bPyrMur.hfsm.hic.hap1.yhs_scffs.curated 1,163,866,538 bp sequence (minus gaps) Autosomes: n = 40 154,884,556 RL_1 ... ... 3,442,651 RL_42 1,043,826,223 bp total Named: n = 2 31,746,173 W 86,793,513 Z 118,539,686 bp total Unplaced: n = 11 296,000 scaffold_43 ... ... 65,000 scaffold_59 1,500,629 bp total Curation made 0 cuts in contigs, 3 breaks at gaps and 2 joins "

Quality metrics table

| Metrics	Pre-curation hap1	Curated hap1
Total bp	1,163,867,738	1,163,867,538
GC %	43.29	43.29
Gaps/Gbp	5.16	4.3
Total gap bp	1,200	1,000
Scaffolds	59	60
Scaffold N50	76,036,967	76,036,967
Scaffold L50	6	6
Scaffold L90	23	23
Contigs	65	65
Contig N50	64,819,875	64,819,875
Contig L50	6	6
Contig L90	26	26
QV	72.9629	71.4214
Kmer compl.	98.6215	98.6211
BUSCO sing.	96.2%	96.3%
BUSCO dupl.	0.7%	0.6%
BUSCO frag.	0.5%	0.5%
BUSCO miss.	2.6%	2.6%

BUSCO: 5.5.0 (euk_genome_met, metaeuk) / Lineage: aves_odb10 (genomes:62, BUSCOs:8338)

HiC contact map of curated assembly
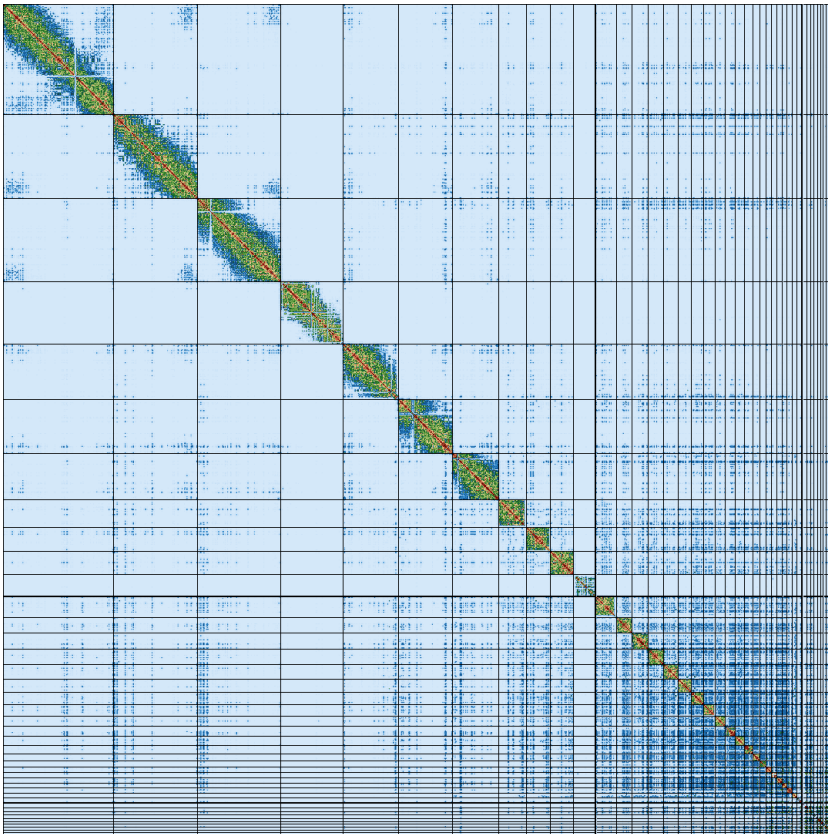
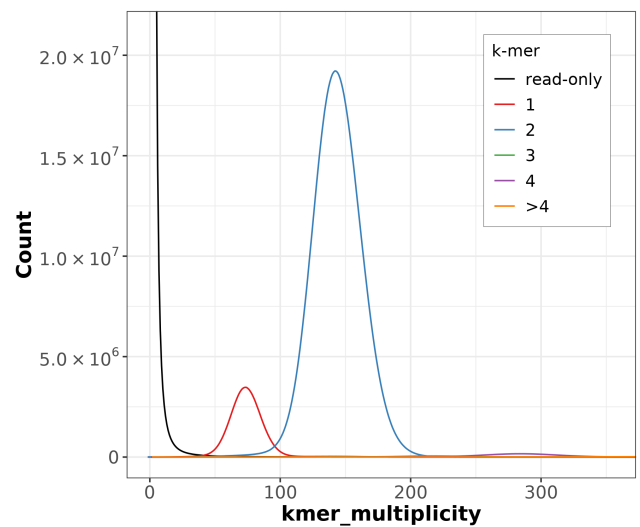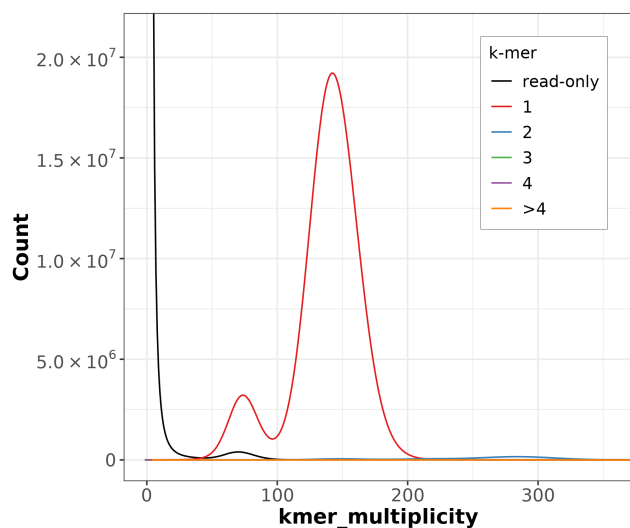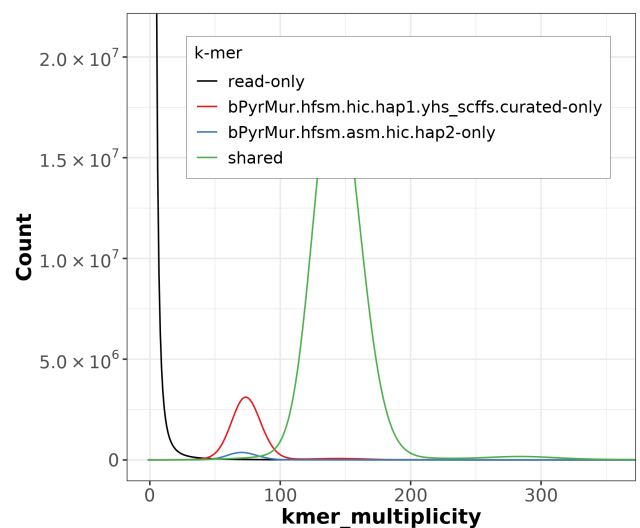**hap1** [LINK]

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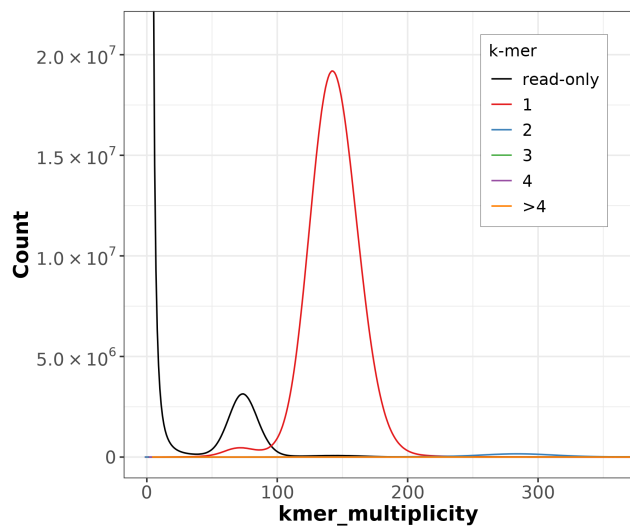
Distribution of k-mer counts per copy
numbers found in **hap2** (hapl.)

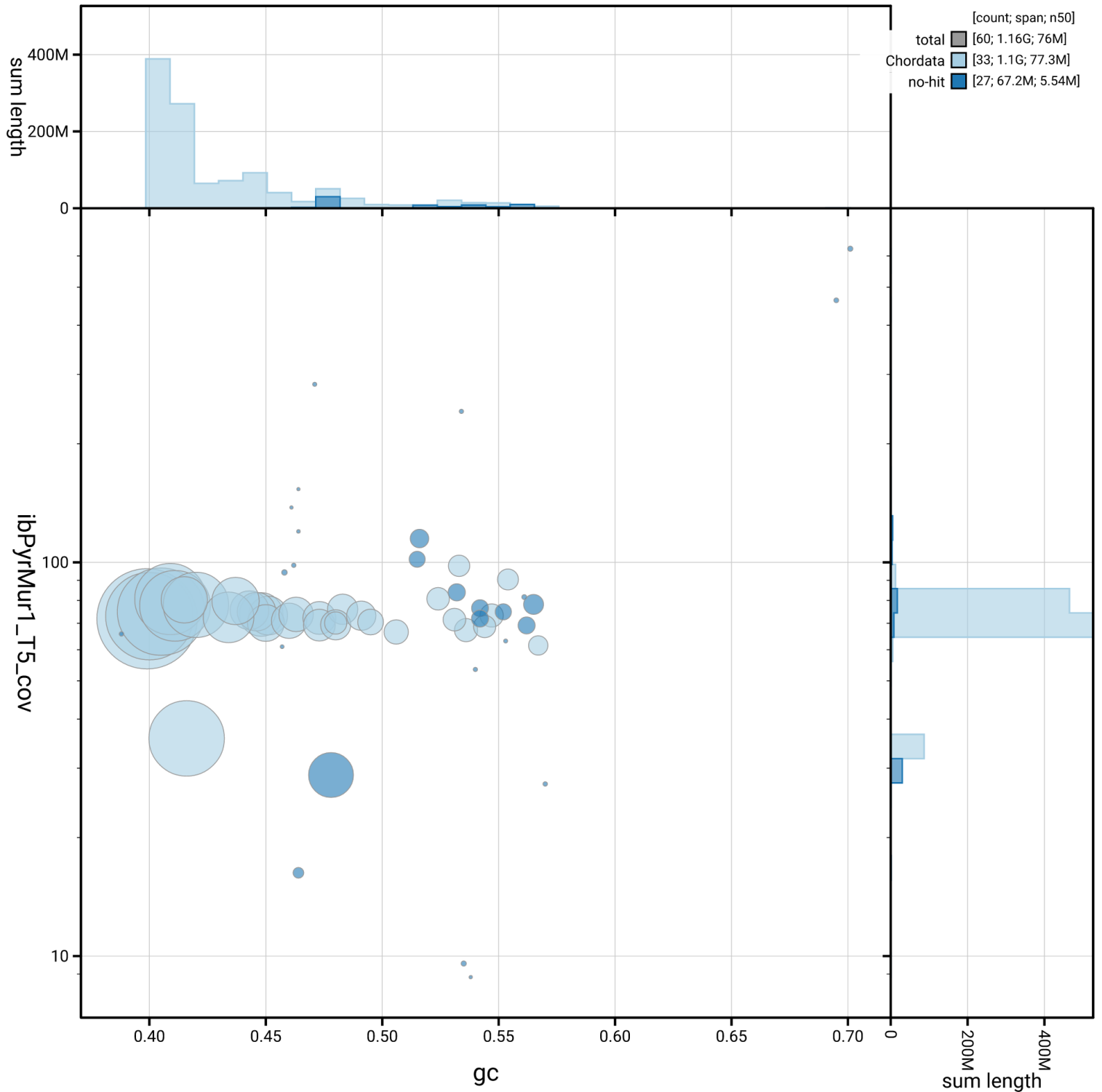Distribution of k-mer counts per copy
numbers found in asm (dipl.)

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

Data profile

Data	ONT	Omni-C	Illumina WGS
Coverage	100x	79x	99x

Assembly pipeline

- **Filtlong**
    |_ *ver:* 0.2.1
    |_ *key param:* --min_length 1000 --min_mean_q 97 -t 80000000000
- **Hifiasm**
    |_ *ver:* 0.24.0
    |_ *key param:* --ont --h1 hic.1.fastq.gz --h2 hic.2.fastq.gz
- **YaHS**
    |_ *ver:* 1.2a
    |_ *key param:* NA
- **CLAWS pipeline**
    |_ *ver:* 2.2.0
    |_ *key param:* NA
- **Blobtoolkit Nextflow pipeline (latest)**
    |_ *ver:* 0.6
    |_ *key param:* NA
- **FOAM pipeline**
    |_ *ver:* 0.5
    |_ *key param:* NA
- **mitos**
    |_ *ver:* 2.1.3
    |_ *key param:* NA

Curation pipeline

- **PretextView**
    |_ *ver:* 0.2.5
    |_ *key param:* NA
- **GRIT_Rapid**
    |_ *ver:* 2.0
    |_ *key param:* NA

Submitter: Tyler Alioto
Affiliation: CNAG

Date and time: 2025-02-27 01:29:17 CET