

ERGA Assembly Report

v24.04.03_beta

Tags: ERGA-BGE

TxID	37273
ToLID	fGamHol14
Species	Gambusia holbrooki
Class	Actinopteri
Order	Cyprinodontiformes

Genome Traits	Expected	Observed
Haploid size (bp)	652,850,389	674,605,490
Haploid Number	18 (source: direct)	24
Ploidy	2 (source: ancestor)	2
Sample Sex	M	M

EBP metrics summary and curation notes

Obtained EBP quality metric for pri: 7.7.Q59

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

- . Observed Haploid Number is different from Expected

Curator notes

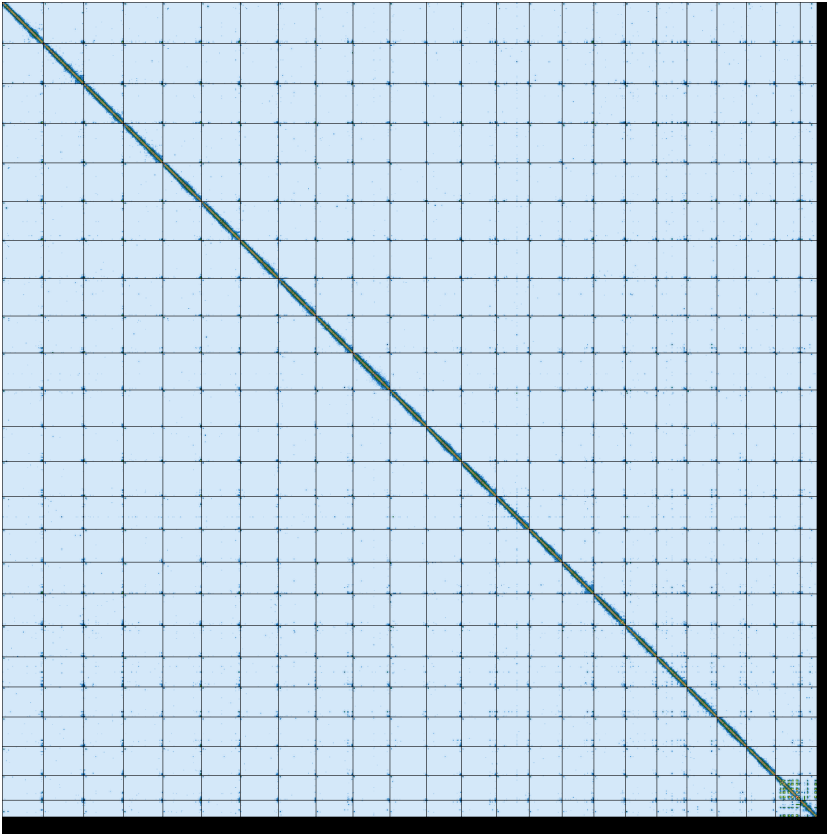
- . Interventions/Gb: 25
- . Contamination notes: "Contamination report for assembly labelled hap1. Total length of scaffolds removed: 1,576,701 (0.2 %); Scaffolds removed: 23 (6.5 %); Largest scaffold removed: (247,497); FCS-GX contaminant species (number of scaffolds; total length of scaffolds): Candidatus Ichthyocystis hellenicum, b-proteobacteria (20; 1,443,046); Crotalus tigris, reptiles (1; 17,401); Mitochondrion (2; 116,254)"
- . Other observations: "Assembly was Hi-C phased; There is a haplotypic inversion on Chromosome 9 at ~8.31-10.04Mb; "

Quality metrics table

Metrics	Pre-curation pri	Curated pri
Total bp	676,976,343	674,605,490
GC %	39.01	39.01
Gaps/Gbp	147.72	152.68
Total gap bp	10,000	11,100
Scaffolds	353	318
Scaffold N50	29,628,722	29,628,722
Scaffold L50	11	11
Scaffold L90	22	22
Contigs	453	421
Contig N50	15,918,108	15,918,108
Contig L50	16	16
Contig L90	63	61
QV	59.8	59.8
Kmer compl.	99.79	99.78
BUSCO sing.	97.7%	97.7%
BUSCO dupl.	0.8%	0.7%
BUSCO frag.	0.5%	0.5%
BUSCO miss.	1.0%	1.1%

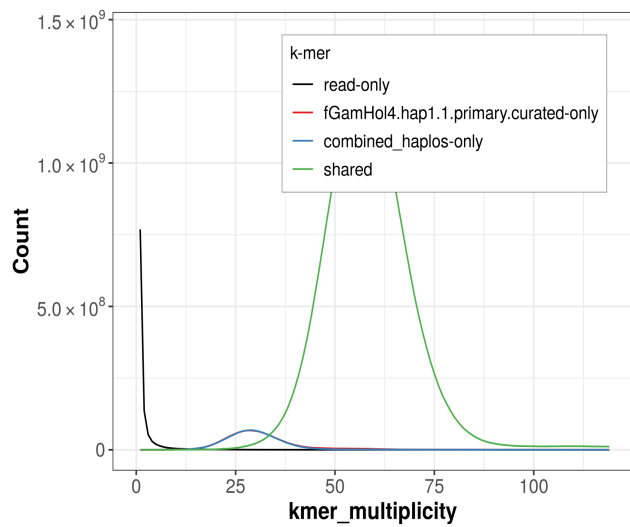
BUSCO 5.3.0 Lineage: actinopterygii_odb10 (genomes:26, BUSCOs:3640)

HiC contact map of curated assembly

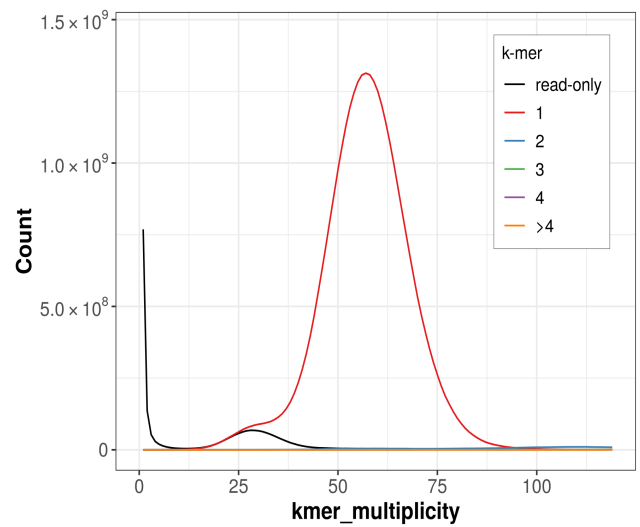


pri [\[LINK\]](#)

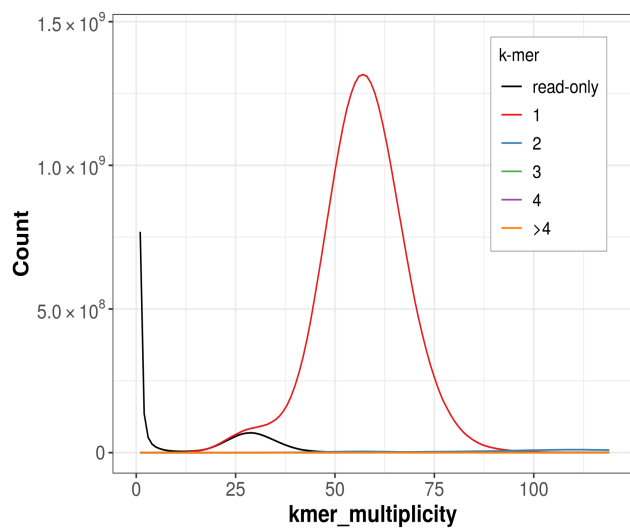
K-mer spectra of curated assembly



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

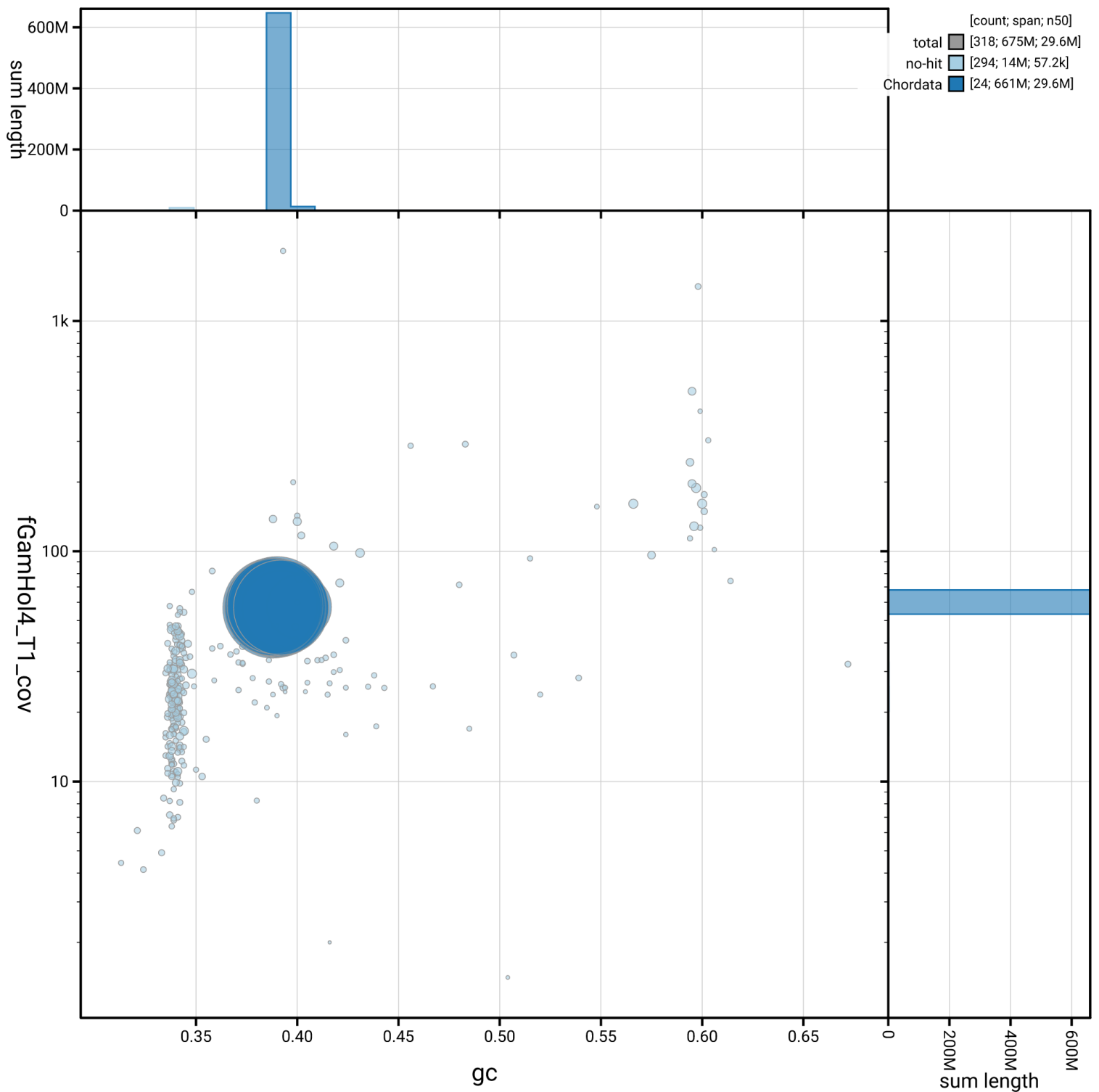


Distribution of k-mer counts per copy numbers found in asm



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

Post-curation contamination screening



pri. 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 v2
Coverage	56x	147x

Assembly pipeline

- **hifiasm-hic**
 - |_ *ver*: 0.19.8-r603
 - |_ *key param*: --primary
- **yahs**
 - |_ *ver*: 1.2.2
 - |_ *key param*: NA

Curation pipeline

- **hifiasm-hic**
 - |_ *ver*: 0.19.8-r603
 - |_ *key param*: --primary
- **yahs**
 - |_ *ver*: 1.2.2
 - |_ *key param*: NA
- **TreeVal**
 - |_ *ver*: 1.2.1
 - |_ *key param*: NA

Submitter: Joanna Collins

Affiliation: WSI

Date and time: 2025-04-23 13:52:07 CEST