ERGA Assembly Report
v24.10.15

Tags: ERGA-BGE

TxID	470658
ToLID	**iqPlaFalx1**
Species	Platycleis falx
Class	Insecta
Order	Orthoptera

Genome Traits	Expected	Observed
Haploid size (bp)	4,863,021,292	5,216,350,721
Haploid Number	15 (source: ancestor)	16
Ploidy	2 (source: ancestor)	2
Sample Sex	X0	X0

EBP metrics summary and curation notes

Obtained EBP quality metric for collapsed: 7.8.Q69

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

. Observed Haploid Number is different from Expected

. Kmer completeness value is less than 90 for collapsed
. Assembly length loss > 3% for collapsed

Curator notes

. Interventions/Gb:
. Contamination notes: "FCS-GX found a single bacterial fragment, tiara found a couple of bacteria and one archaea contigs. Additional short contigs removed based on coverage."
. Other observations: "Sample is recorded as male and the observed karyotype (n=16, X0) is consistent with the literature. Assembly was performed with hifiasm using HiC data with the --h1/--h2 parameters. The collapsed p_ctg was selected based on N50/N90 and L50/L90 stats. purge_dups removed too many large contigs despite the cutoffs being very clear, causing 30 loss in complete BUSCOS. So, instead, I used the purge_dups classification together with BUSCO and tiara/FCS-GX reports to manually remove any contigs flagged by purge_dups that did not contain any BUSCOs (eukaryota, metazoa, arthropoda, or insecta from odb10), as well as flagged contaminants."

Quality metrics table

| Metrics	Pre-curation collapsed	Curated collapsed
Total bp	5,584,225,582	5,216,350,721
GC %	39.3	39.28
Gaps/Gbp	0	25.69
Total gap bp	0	26,800
Scaffolds	846	21
Scaffold N50	57,970,750	399,530,575
Scaffold L50	30	4
Scaffold L90	104	12
Contigs	846	155
Contig N50	57,970,750	62,235,952
Contig L50	30	29
Contig L90	104	88
QV	67.5191	69.6433
Kmer compl.	86.9597	84.9089
BUSCO sing.	93.5%	95.1%
BUSCO dupl.	6.1%	4.7%
BUSCO frag.	0.1%	0.0%
BUSCO miss.	0.3%	0.2%

BUSCO: 6.0.0 (euk_genome_min, miniprot) / Lineage: arthropoda_odb10 (genomes:90, BUSCOs:1013)
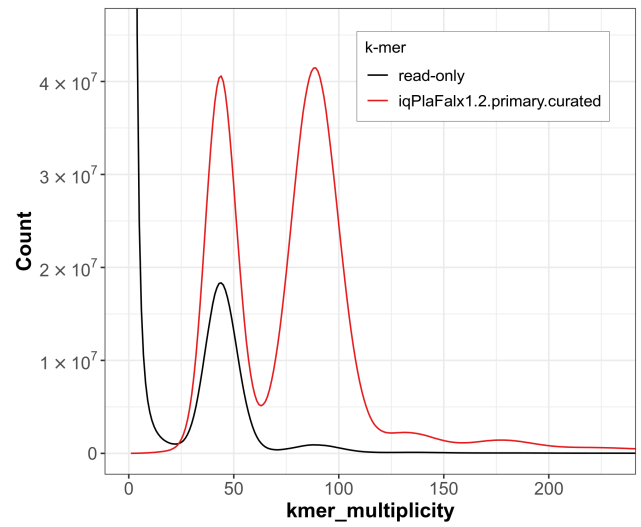
HiC contact map of curated assembly
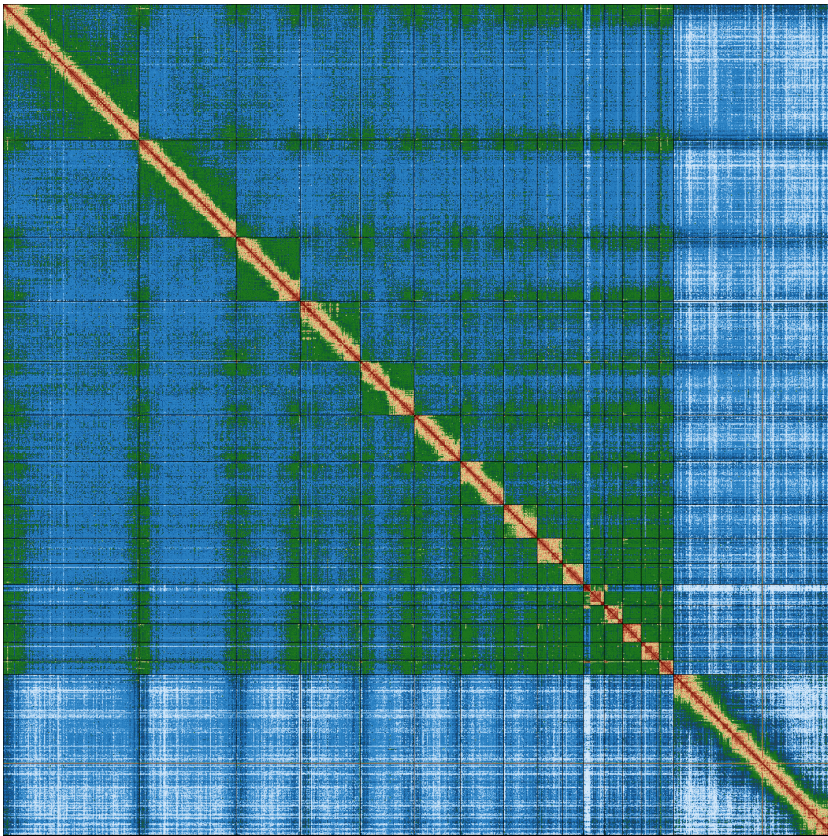
collapsed [LINK]
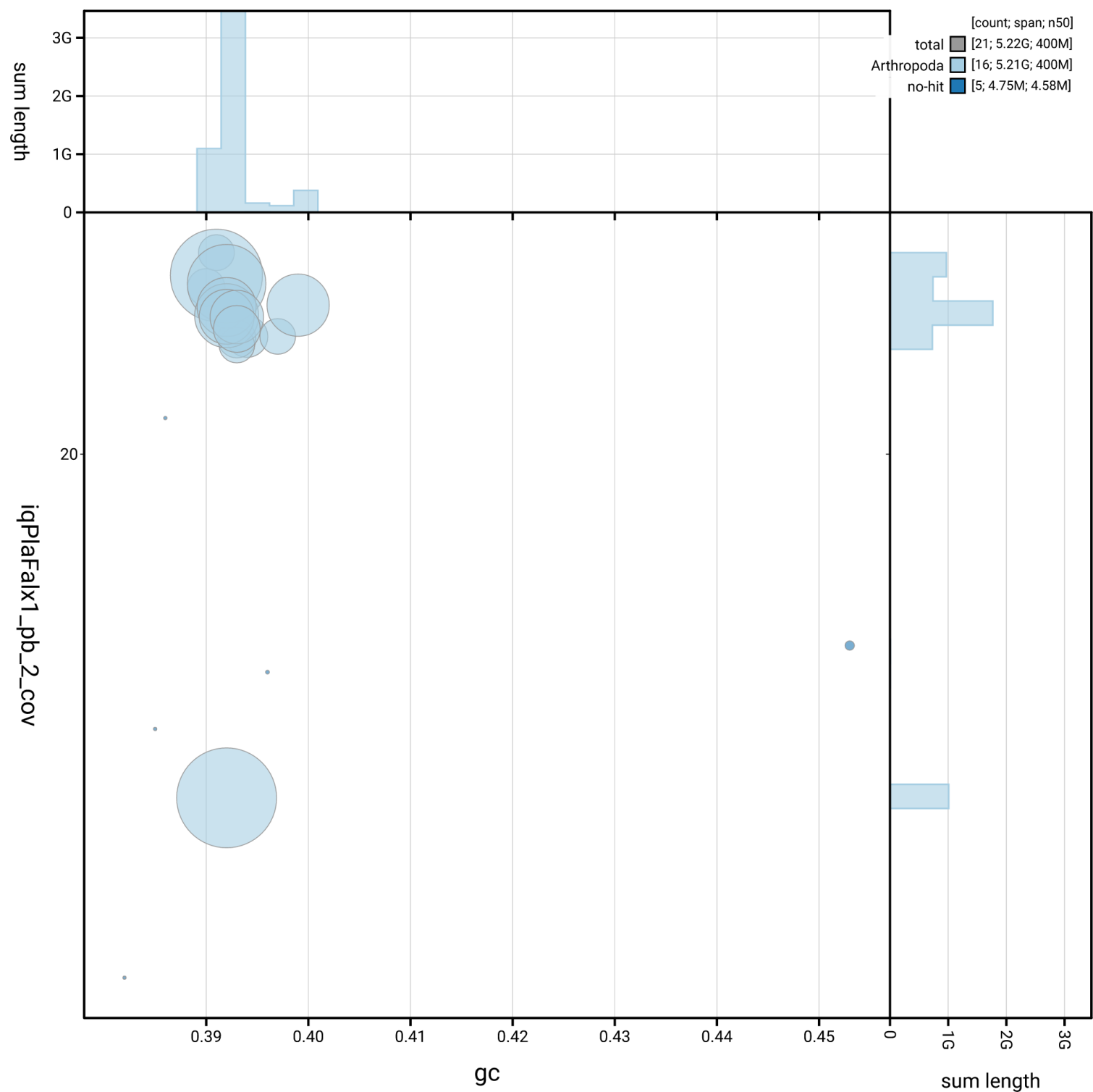
K-mer spectra of curated assembly
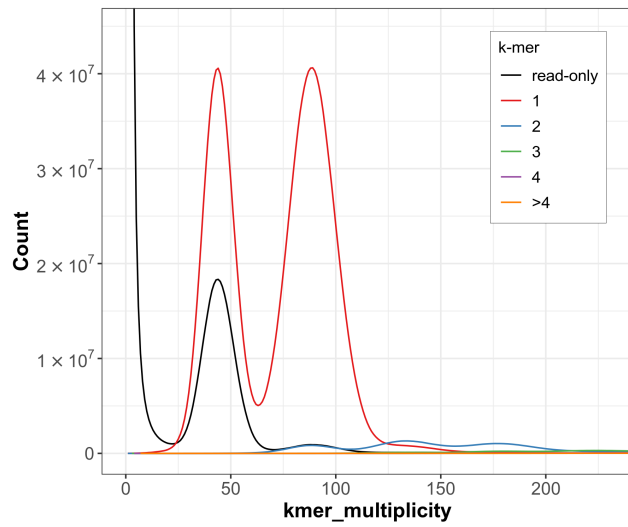
Distribution of k-mer counts per copy
numbers found in asm

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

Post-curation contamination screening

collapsed. 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 v2
Coverage	89x	66x

Assembly pipeline

- **hifiasm-hic**
    |_ *ver:* 0.25.0-r726
    |_ *key param:* --h1
    |_ *key param:* --h2
- **purge_dups**
    |_ *ver:* 1.2.5
    |_ *key param:* NA
- **tiara**
    |_ *ver:* 1.0.3
    |_ *key param:* NA
- **YaHS**
    |_ *ver:* 1.2a.2
    |_ *key param:* --no-contig-ec
- **NBIS/Earth-Biogenome-Project-pilot**
    |_ *ver:* 1d2a1e4
    |_ *key param:* NA

Curation pipeline

- **sanger-tol/curationpretext**
    |_ *ver:* 1.4.2
    |_ *key param:* NA
- **GRIT_Rapid**
    |_ *ver:* 1a3d79a8
    |_ *key param:* NA

Submitter: Guilherme B. Dias
Affiliation: SciLifeLab

Date and time: 2025-09-03 09:48:22 CEST