ERGA Assembly Report

v24.10.15

Tags: ERGA-BGE

TxID	765876
ToLID	**ihDacCocc15**
Species	Dactylopius coccus
Class	Insecta
Order	Hemiptera

Genome Traits	Expected	Observed
Haploid size (bp)	365,644,368	360,319,176
Haploid Number	8 (source: direct)	8
Ploidy	2 (source: ancestor)	2
Sample Sex	unknown	unknown

EBP metrics summary and curation notes

Obtained EBP quality metric for pri: 7.7.Q73

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

. BUSCO single copy value is less than 90% for pri

Curator notes

. Interventions/Gb: 0
. Contamination notes: "FCS-GX detected 17 contigs represetning the following species: Spiroplasma endosymbiont of Danaus chrysippus, Candidatus Dactylopiibacterium carminicum, Spiroplasma endosymbiont of Lariophagus distinguendus and Wolbachia endosymbiont (group A) of Rhinocyllus conicus"
. Other observations: "PacBio HiFi reads were subsampled to 60x coverage. HiFiasm (yield: 383Mb, N50: 27.9M) + FCS-GX (removed 17 contigs, 17.5Mb) + purge_dups (yield: 368Mb, N50: 37.8M) + yahs. The HiC data is of a very bad quality, i.e. only 258K read pairs span a region of +1Kb. Therefore yahs was run without contig error correction. In the curation the pretext map was barely usable. Three larger repetitive contigs were removed due to very low PacBio covarage. Three more joins were done based on "slightly" increased pixel density and the orientation is based on the telomere motifs. I tried to map against other chromosome-scale assemblies but I could not find close relatives ones. These ones: GCA_009761765.1_ASM976176v1, GCA_032883995.1_ASM3288399v1, GCA_039619475.1_ASM3961947v1, GCA_039619475.1_ASM3961947v1 and BGE:Scarites_abbreviatus show barely any hit. I attached a telomere plot to show that we potentially achieved chromosomes-scale, as almost all of the 8 longest scaffolds have telomere motifs at both ends."

Quality metrics table

Metrics	Pre-curation pri	Curated pri
Total bp	368,570,170	360,319,176
GC %	31.17	31.03
Gaps/Gbp	0	22.2
Total gap bp	0	1,100
Scaffolds	24	13
Scaffold N50	37,830,695	46,180,480
Scaffold L50	4	4
Scaffold L90	12	8
Contigs	24	21
Contig N50	37,830,695	37,830,695
Contig L50	4	4
Contig L90	12	12
QV	73.5344	73.5344
Kmer compl.	99.172	99.172
BUSCO sing.	89.1%	89.2%
BUSCO dupl.	1.4%	1.4%
BUSCO frag.	3.4%	3.5%
BUSCO miss.	6.0%	5.9%

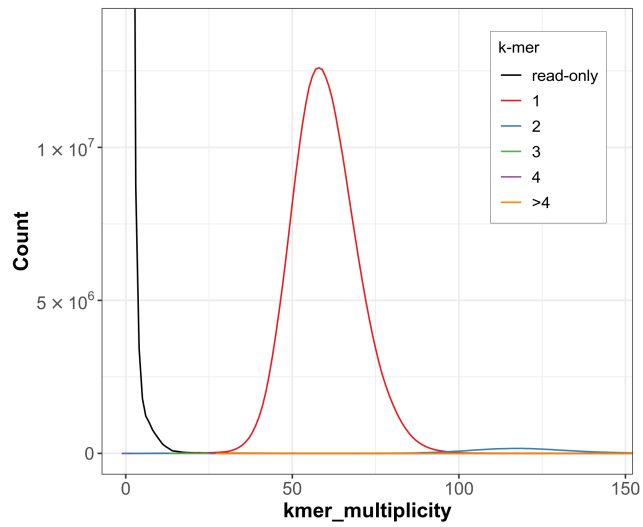BUSCO: 5.8.3 (euk_genome_min, miniprot) / Lineage: hemiptera_odb12 (genomes:32, BUSCOs:3396)

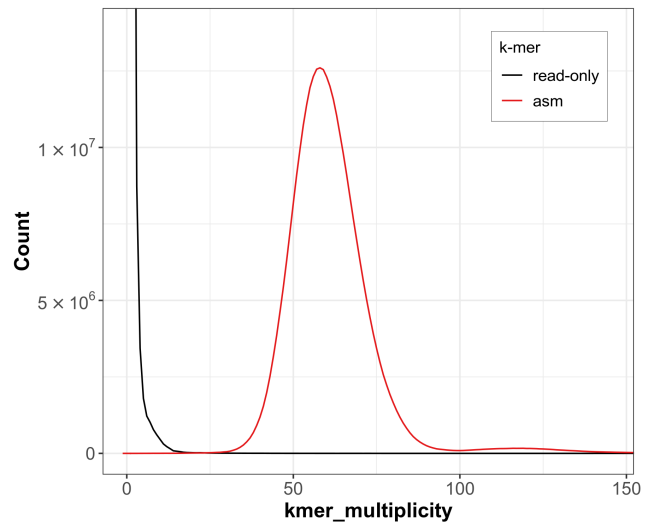# HiC contact map of curated assembly

Genomic scaffolds

SUPER_1

SUPER_2

SUPER_3

SUPER_4

SUPER_5

SUPER_6

SUPER_7

SUPER_8
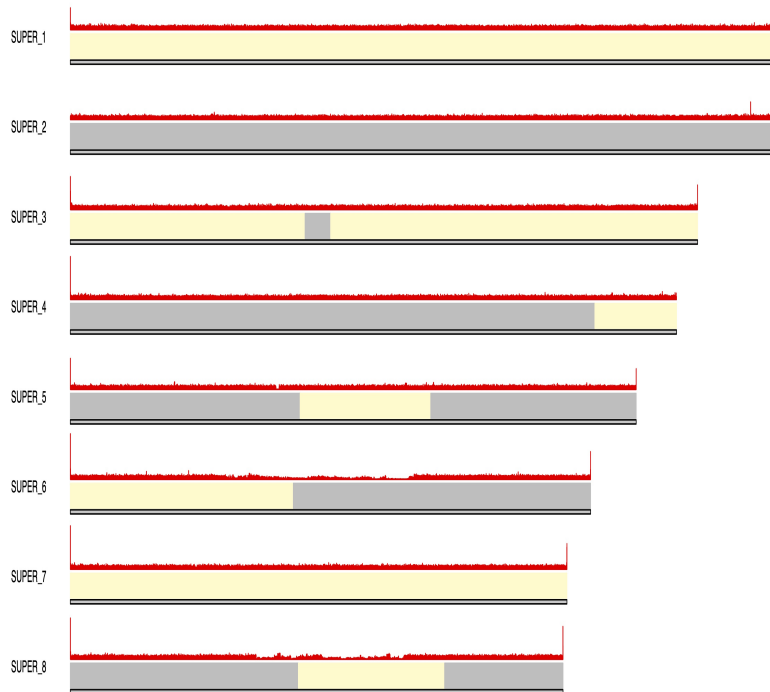
pri [LINK]
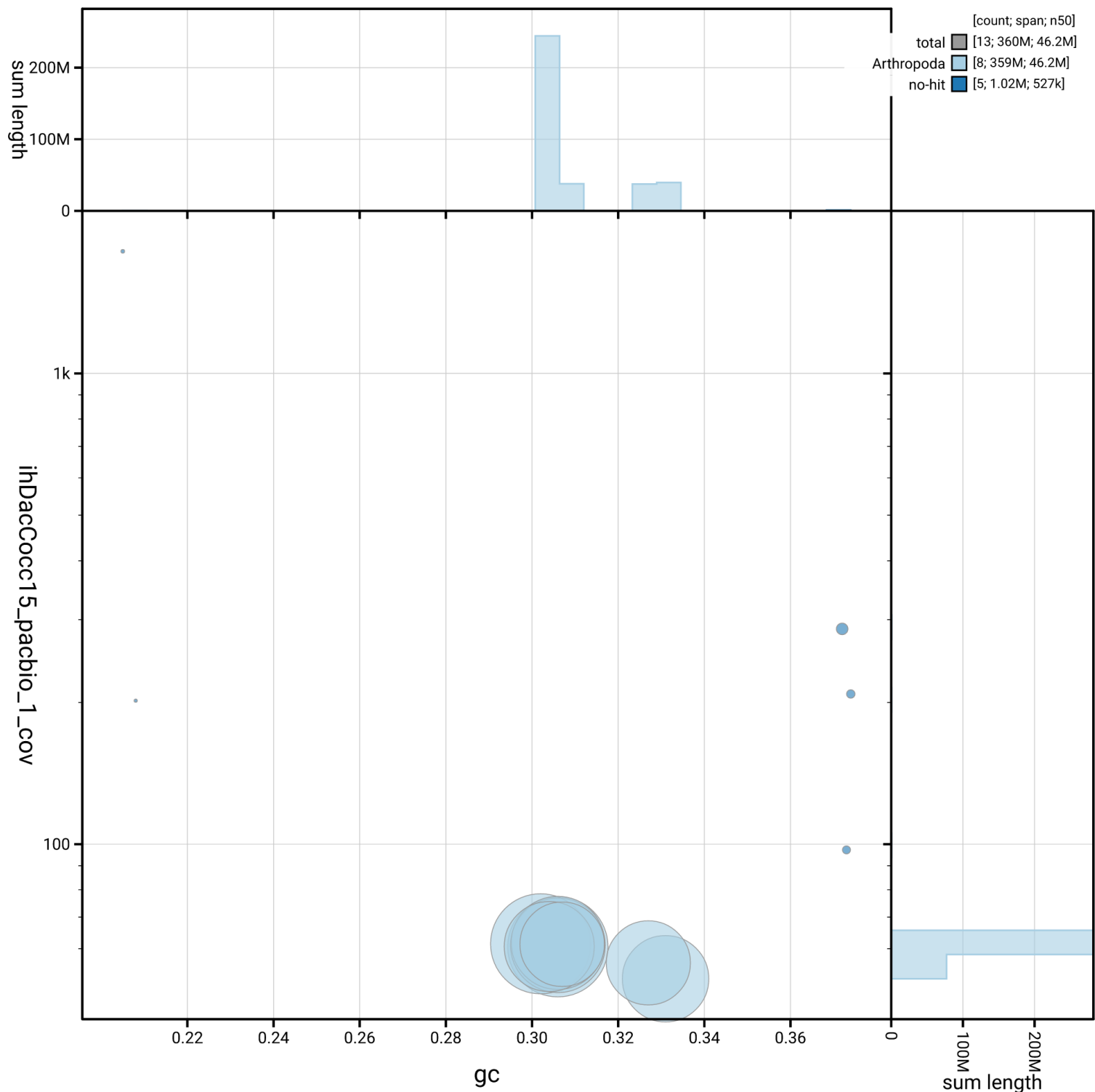
K-mer spectra of curated assembly

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

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	HiFi	HiC
Coverage	264x	131x

Assembly pipeline

- **Hifiasm**
   |_ *ver:* 0.25.0-r726
   |_ *key param:* l3
- **purge_dups**
   |_ *ver:* 1.2.6
   |_ *key param:* NA

Curation pipeline

- **GRIT_Rapid**
   |_ *ver:* 1a3d79a8
   |_ *key param:* NA
- **HiGlass**
   |_ *ver:* 0.10.4
   |_ *key param:* NA

Submitter: Martin Pippel
Affiliation: SciLifeLab

Date and time: 2025-10-09 09:01:54 CEST