ERGA Assembly Report

v24.10.15

Tags: ERGA-BGE

TxID	3350012
ToLID	**ihAphHill11**
Species	Aphis hillerislambersi
Class	Insecta
Order	Hemiptera

Genome Traits	Expected	Observed
Haploid size (bp)	489,340,424	427,291,510
Haploid Number	4 (source: ancestor)	4
Ploidy	2 (source: ancestor)	2
Sample Sex	XX	XX

EBP metrics summary and curation notes

Obtained EBP quality metric for pri: 7.8.Q69

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

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

Curator notes

. Interventions/Gb: 20
. Contamination notes: "Contaminations were identified and removed with FCS-GX (removed: 760 contigs, 24.3M). Mitochondrial genome was removed from the assembly."
. Other observations: "PacBio reads were assembled with hifiasm (yield: 509M, N50: 37.7Mb, 850 contigs). FCS-GX detected several contaminations which where removed (yield: 485M, N50: 50.4Mb, 90 contigs). Purge_dups was run to remove haplotypic duplications (purged: 62 sequences, 38.8Mb remaining assembly: yield: 446M, N50: 48.4Mb, 28 contigs). The HiC quality is really suboptimal (just 130K cis read pairs >40Kb) and 4 individuals were used for the HiC sequencing (see comments in assembly_curation slack channel from 16th October 2025) Due to the low number of contigs and the low HiC quality the scaffolding was done manually in the manual curation step. Obvious contig misjoins (see old slack message) where removed based on off-diagonal HiC signals, on coverage, and by investigating PacBio read alignments in IGV. The newer Hifiasm version (0.25.0-r726) created far less contig errors compared to the older version back in October 2025. Synteny analysis with 5 chromosome-scale Aphis assemblies were used to confirm the overall chromosome structure and the X-chromosome."
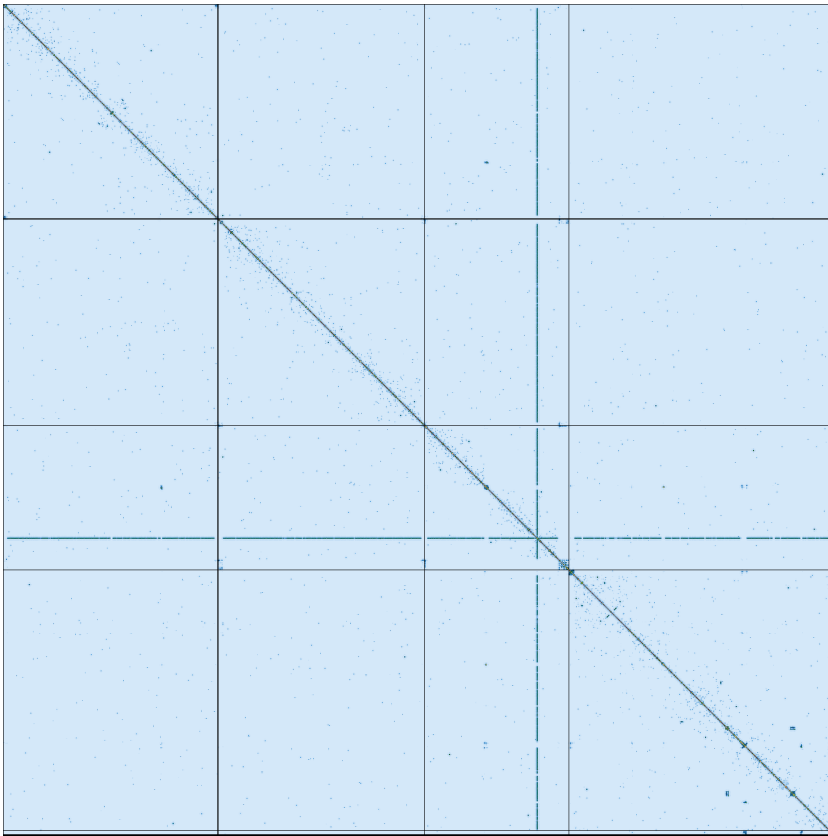
Quality metrics table

Metrics	Pre-curation pri	Curated pri
Total bp	446,144,015	427,291,510
GC %	28.81	28.79
Gaps/Gbp	0	23.4
Total gap bp	0	2,000
Scaffolds	25	13
Scaffold N50	48,412,940	110,330,946
Scaffold L50	4	2
Scaffold L90	10	4
Contigs	25	23
Contig N50	48,412,940	48,412,940
Contig L50	4	4
Contig L90	10	10
QV	69.419	69.8166
Kmer compl.	89.5035	89.2276
BUSCO sing.	93.7%	97.7%
BUSCO dupl.	4.2%	1.3%
BUSCO frag.	0.9%	0.0%
BUSCO miss.	1.2%	1.0%

Warning! BUSCO versions or lineage datasets are not the same across results:
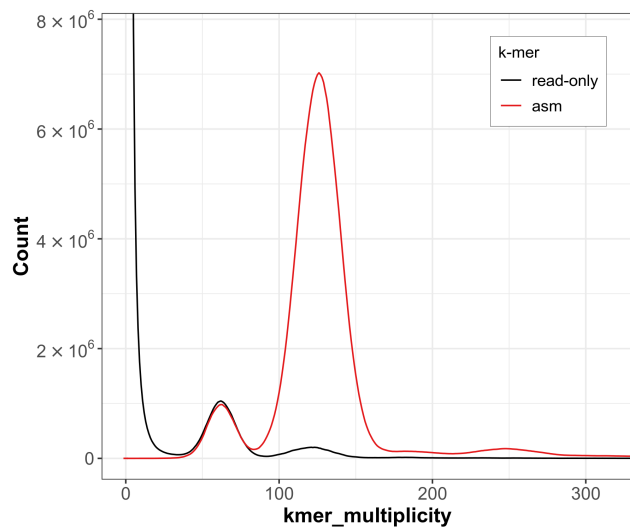BUSCO: 5.8.3 (euk_genome_min, miniprot) / Lineage: hemiptera_odb12 (genomes:32, BUSCOs:3396)
BUSCO: 6.0.0 (euk_genome_min, miniprot) / Lineage: hemiptera_odb12 (genomes:32, BUSCOs:3396)
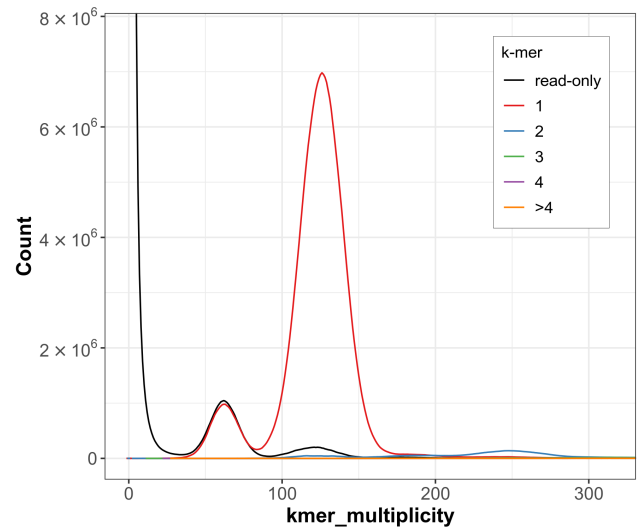
HiC contact map of curated assembly

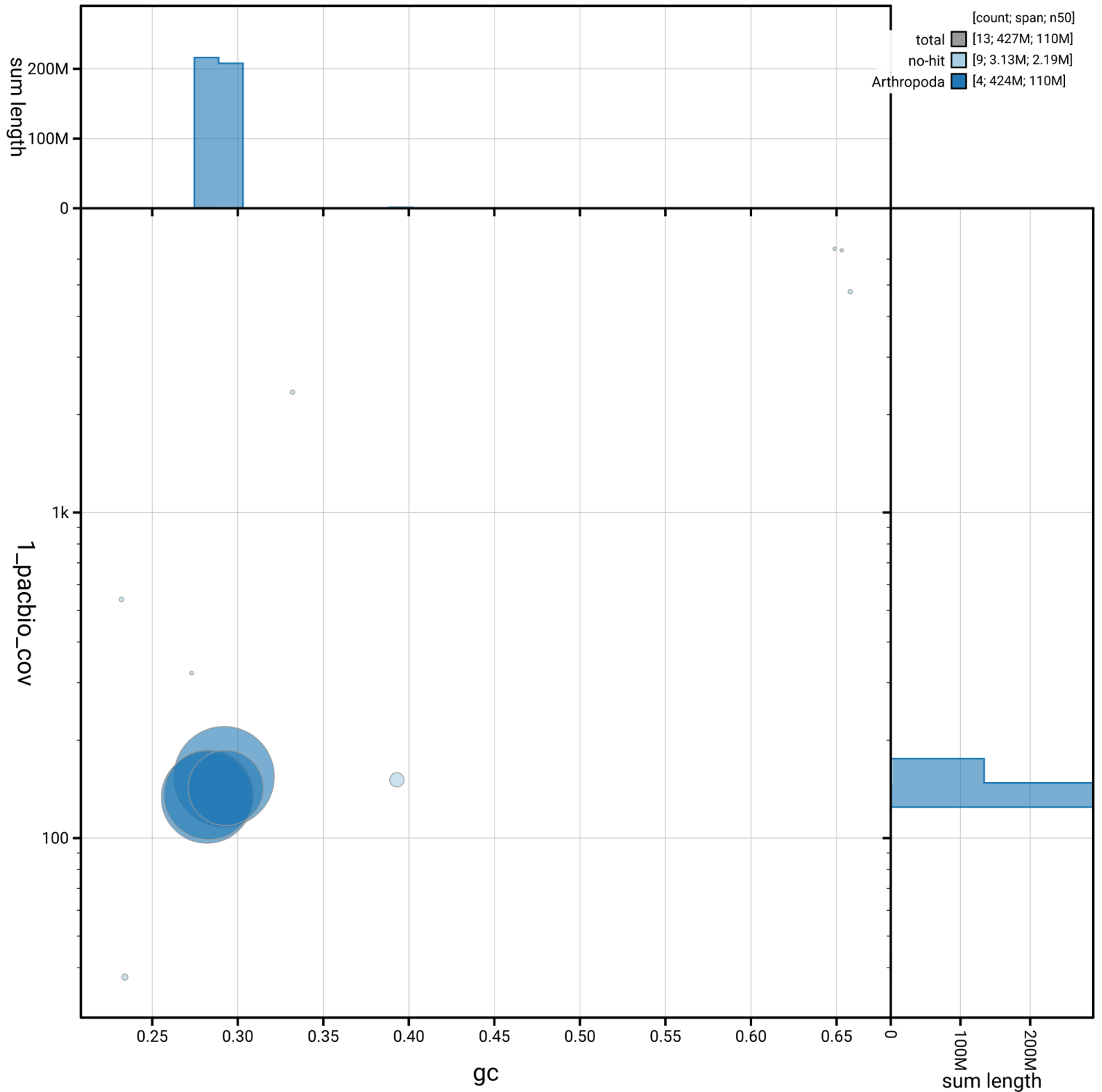**pri** [LINK]

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

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	70x	92x

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: 2026-02-13 22:14:52 CET