ERGA Assembly Report

v24.10.15

Tags: ERGA-BGE

TxID	3229267
ToLID	**iyHalNico6**
Species	Halictus nicosiae
Class	Insecta
Order	Hymenoptera

Genome Traits	Expected	Observed
Haploid size (bp)	362,073,405	367,819,907
Haploid Number	14 (source: ancestor)	7
Ploidy	1 (source: ancestor)	1
Sample Sex	NA	NA

EBP metrics summary and curation notes

Obtained EBP quality metric for collapsed: 7.7.Q75

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:
. Contamination notes: "FCS-GX identified some small bacterial fragments, as well as two large circular contigs matching Spiroplasma and Wolbachia. The identity of these was confirmed though BLAST. The Spiroplasma genome is 2.48 Mb long with 25.16% GC, and it has highest similarity to other Spiroplasma cobionts of bees such as Seladonia, Ammophila, and Lassioglossum. The Wolbachia genome is 1.34 Mb long with 35.22% GC, and it has high similarity with other Wolbachia group A genomes from a wide range of insects."
. Other observations: "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. The genome seems to have only metacentric or sub-metacentric chromosomes and the gaps are usually near where the centromeres would be. Some putative peri-centromeric scaffolds could be assigned to chromosomes abeit with the unlocalized tag. Also, because it is common for non-homologous chromosomes to share repetitive DNA, some repetitive scaffolds were left unassigned to chromosomes because of this uncertainty. Karyotype is unkown but another genus from the same tribe (genus Lasioglossum, tribe Halictini) displays a wide range of karyotypes (n=6-20), so the GOAT ancestor inference is unlikely to be useful here."

Quality metrics table

| Metrics	Pre-curation collapsed	Curated collapsed
Total bp	367,818,707	367,819,907
GC %	40.99	40.99
Gaps/Gbp	27.19	43.5
Total gap bp	2,000	3,200
Scaffolds	58	52
Scaffold N50	38,644,030	49,072,324
Scaffold L50	4	4
Scaffold L90	10	7
Contigs	68	68
Contig N50	24,371,020	24,371,020
Contig L50	7	7
Contig L90	14	14
QV	75.6357	75.6357
Kmer compl.	93.1568	93.1568
BUSCO sing.	98.3%	98.3%
BUSCO dupl.	0.4%	0.4%
BUSCO frag.	0.2%	0.2%
BUSCO miss.	1.1%	1.1%

BUSCO: 6.0.0 (euk_genome_min, miniprot) / Lineage: hymenoptera_odb10 (genomes:40, BUSCOs:5991)
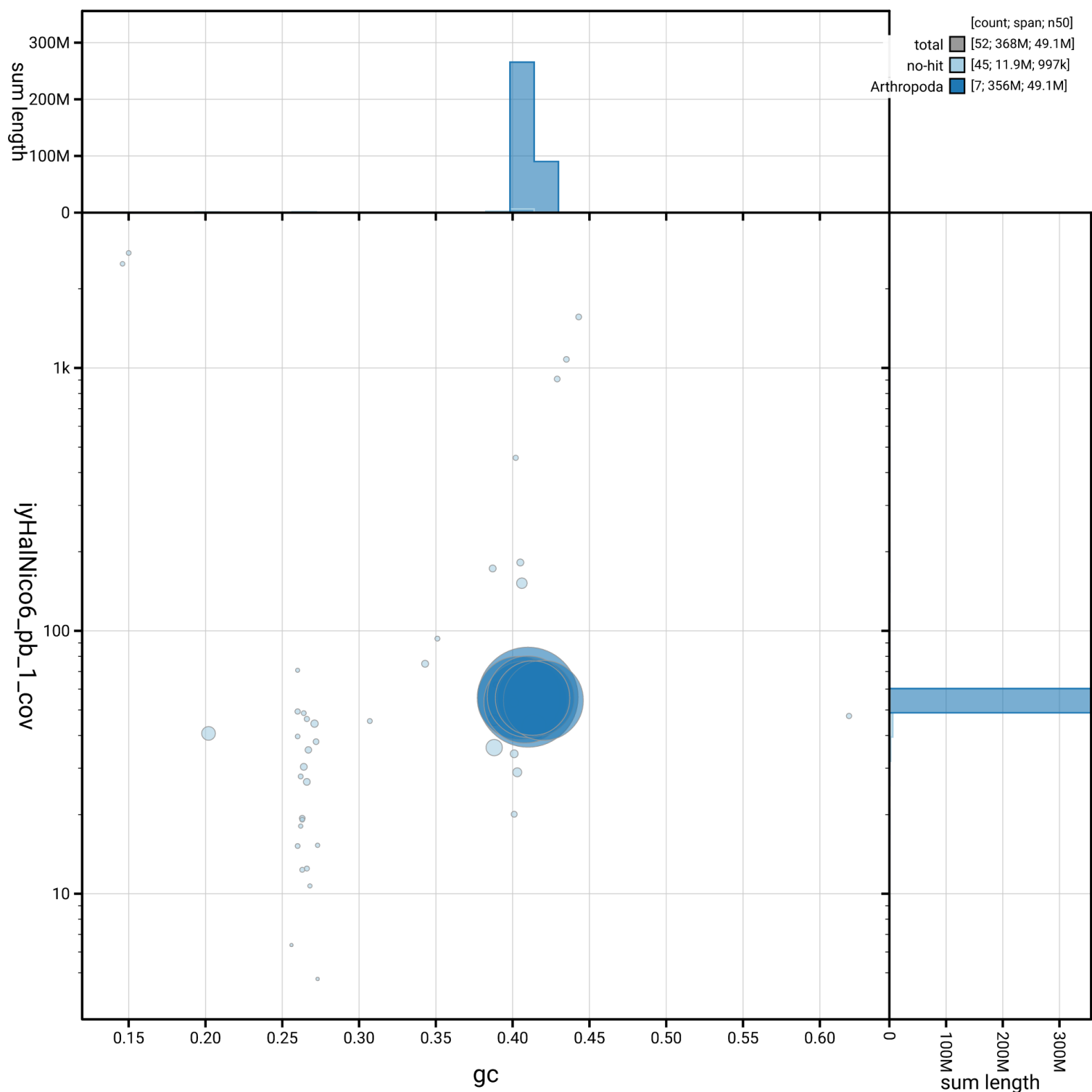
HiC contact map of curated assembly
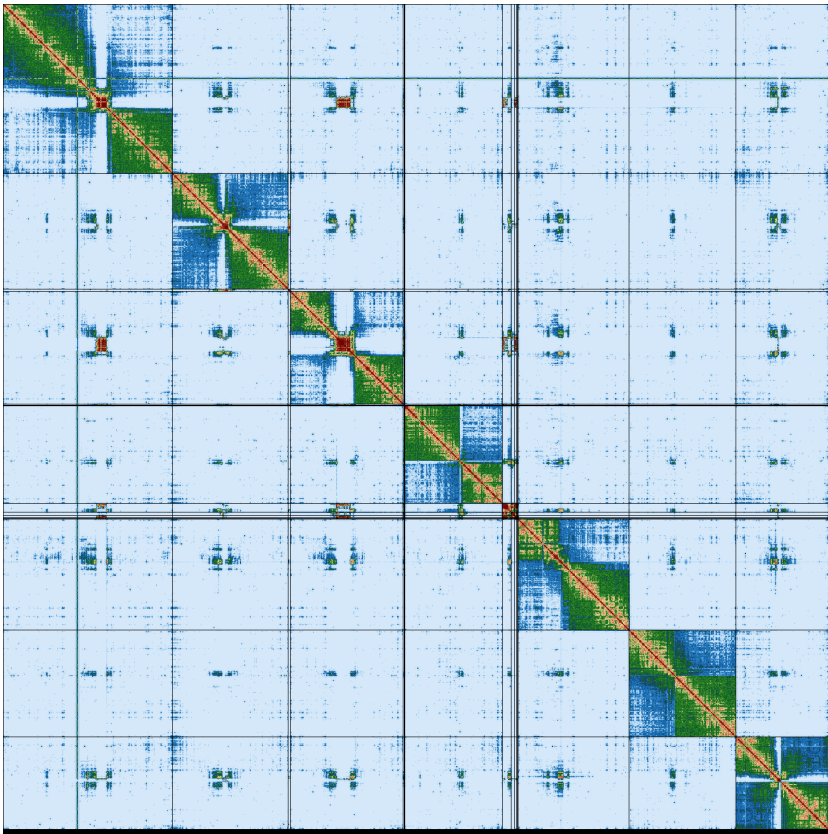
collapsed [LINK]

K-mer spectra of curated assembly
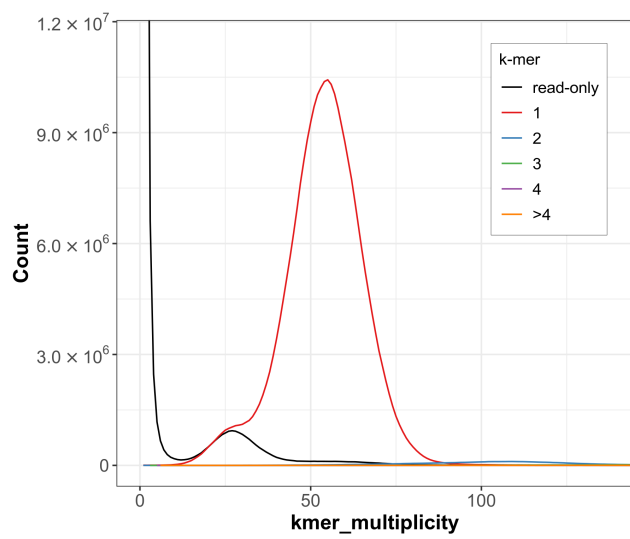
Distribution of k-mer counts per copy
numbers found in asm

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	55x	204x

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
- **sanger-tol/blobtoolkit**
    |_ *ver:* 0.8.0
    |_ *key param:* NA
- **YaHS**
    |_ *ver:* 1.2a.2
    |_ *key param:* NA
- **NBIS/Earth-Biogenome-Project-pilot**
    |_ *ver:* 99c8937
    |_ *key param:* NA

Curation pipeline

- **sanger-tol/curationpretext**
    |_ *ver:* 1.4.2
    |_ *key param:* NA
- **sanger-tol/agp-tpf-utils**
    |_ *ver:* 35ea45d
    |_ *key param:* NA

Submitter: Guilherme B. Dias
Affiliation: SciLifeLab

Date and time: 2025-09-12 10:56:27 CEST