ERGA Assembly Report

v24.10.15

Tags: ERGA-BGE

TxID	522842
ToLID	**ilMarQuer1**
Species	Marumba quercus
Class	Insecta
Order	Lepidoptera

Genome Traits	Expected	Observed
Haploid size (bp)	596,365,935	595,436,427
Haploid Number	28 (source: ancestor)	28
Ploidy	2 (source: ancestor)	2
Sample Sex	ZZ	ZZ

EBP metrics summary and curation notes

Obtained EBP quality metric for hap1: 7.7.Q70

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

Curator notes

. Interventions/Gb: 7
. Contamination notes: "No contamination was detected by Blobtoolkit which was run on the diploid assembly"
. Other observations: "This assembly was obtained with Hifiasm from ONT data. Given the high quality of both haplotypes, curation was done in diploid mode. We have shared the curated and painted pretext map corresponding to the diploid assembly (after curating each hap separately and combining them). The sex SUPER was initially mistakenly labelled as X, but this is corrected in the round corresponding to diploid curation. The Z was confirmed by alignment to other evolutionarily close species. This EAR includes the modifications suggested by the reviewers and we are sharing a pretext map for each hap generated by our method of diploid curation. Note that we had to remove the hap2 merqury from the yaml since we continually obtained an error when adding it. The code seems to assume both haps are in the same folder. Since we handled both haps separately in this review and there are not mixed haps we obtained an error."

Quality metrics table

Metrics	Pre-curation hap1	Pre-curation hap2	Curated hap1	Curated hap2
Total bp	596,105,788	594,728,036	595,213,122	595,436,427
GC %	39.93	39.91	39.92	39.93
Gaps/Gbp	6.71	6.73	3.36	15.11
Total gap bp	800	800	400	1,800
Scaffolds	32	30	28	28
Scaffold N50	22,445,820	22,081,071	22,445,820	22,081,071
Scaffold L50	12	12	12	12
Scaffold L90	24	24	24	24
Contigs	36	34	30	37
Contig N50	21,511,819	21,256,303	22,445,820	21,036,126
Contig L50	13	13	12	13
Contig L90	26	26	24	28
QV	70.5765		70.9388	
Kmer compl.	96.8579		96.8717	
BUSCO sing.	99.0%	98.9%	98.9%	99.0%
BUSCO dupl.	0.2%	0.2%	0.2%	0.2%
BUSCO frag.	0.5%	0.5%	0.5%	0.5%
BUSCO miss.	0.3%	0.4%	0.4%	0.3%

BUSCO: 5.5.0 (euk_genome_met, metaeuk) / Lineage: arthropoda_odb10 (genomes:90, BUSCOs:1013)
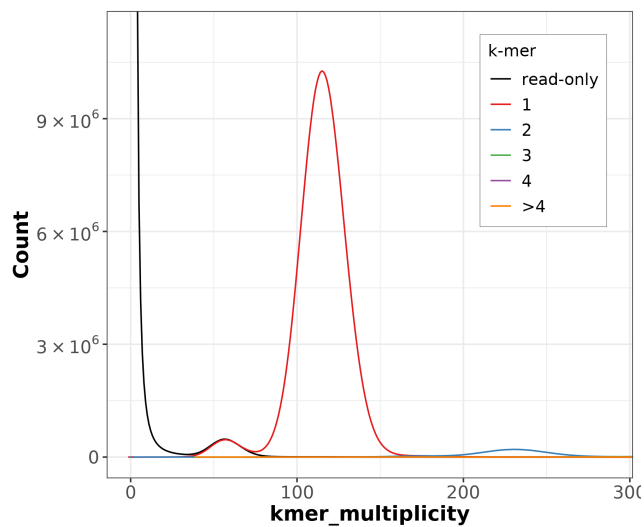
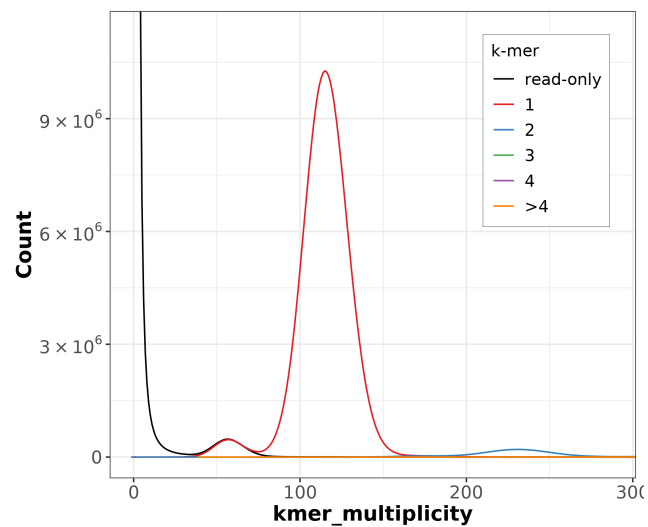# HiC contact map of curated assembly
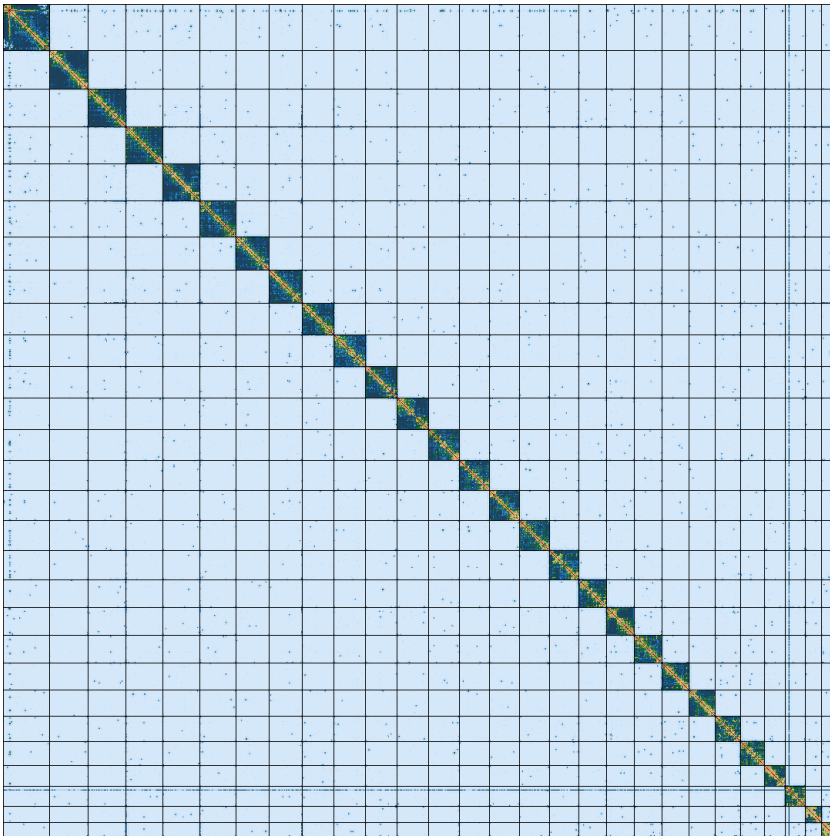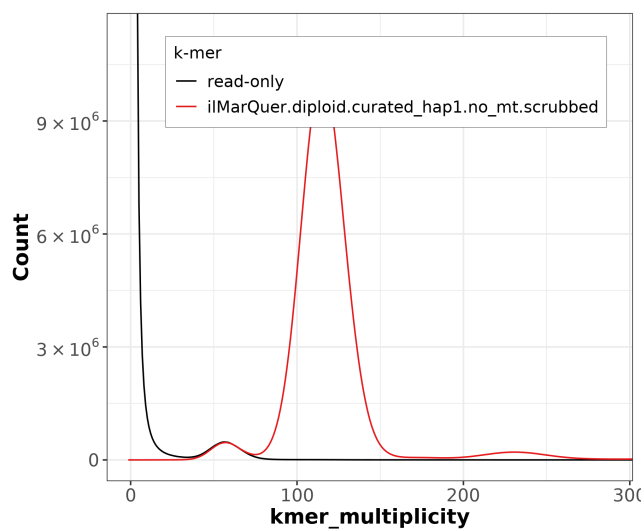

hap1 [LINK]
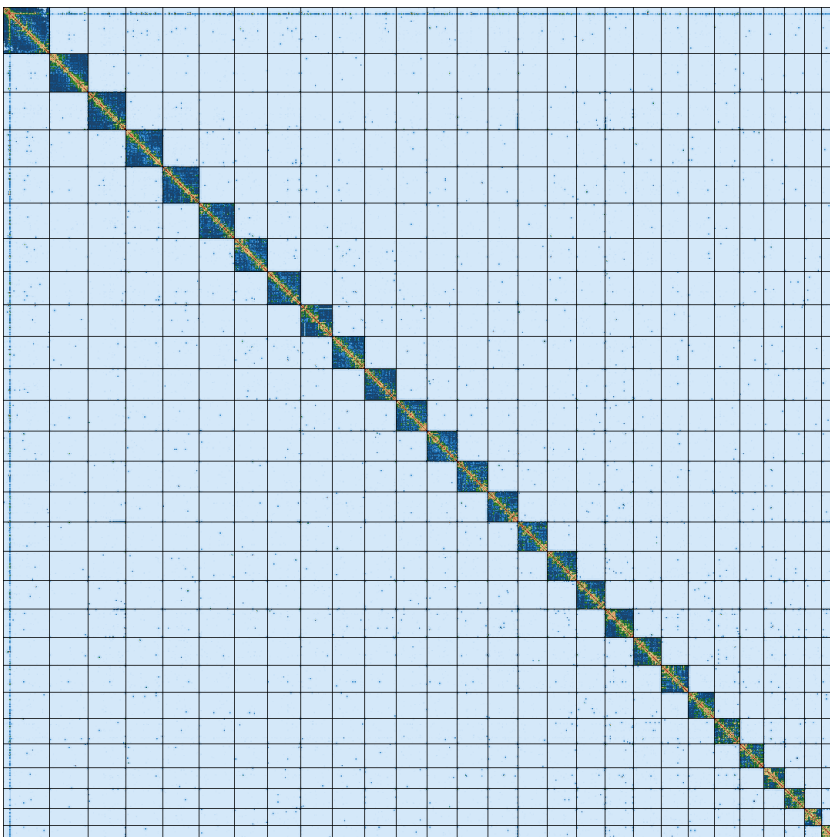
hap2 [LINK]

K-mer spectra of curated assembly

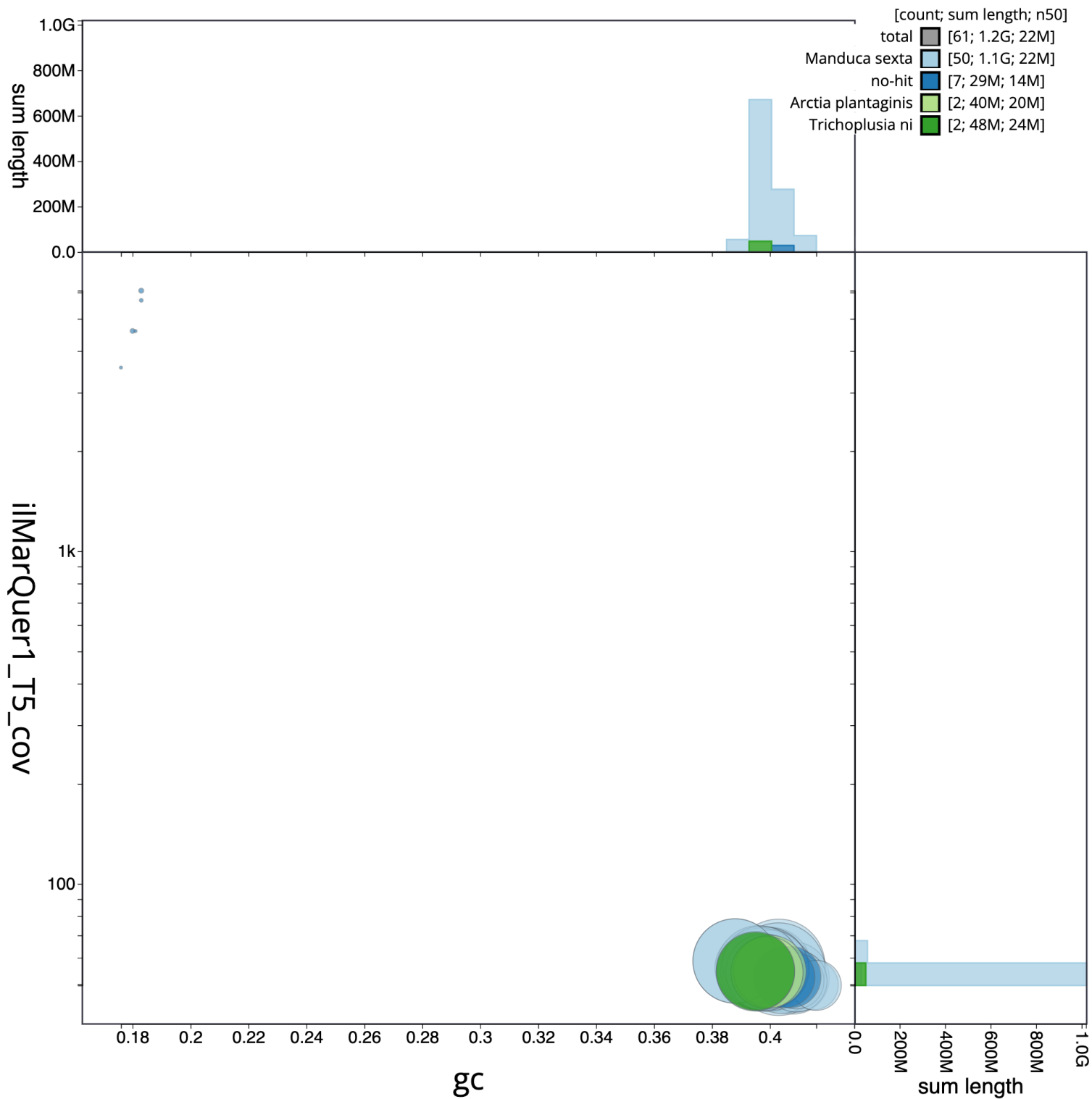Distribution of k-mer counts per copy
numbers found in asm

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

hap1. 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	ONT	Illumina	Omni-C
Coverage	117	72	139

Assembly pipeline

- **CLAWS**
    |_ *ver:* v2.3.0
    |_ *key param:* NA
- **Filtlong**
    |_ *ver:* v0.2.1
    |_ *key param:* -t target_bases 80000000000
    |_ *key param:* --filtlong-min-mean-q 97
- **Hifiasm**
    |_ *ver:* 0.24.0
    |_ *key param:* --ont
- **Yahs**
    |_ *ver:* 1.2a
    |_ *key param:* -mq 0
    |_ *key param:* -no-contig-ec

Curation pipeline

- **Blobtoolkit Nextflow pipeline**
    |_ *ver:* 0.6
    |_ *key param:* NA
- **PretextViewAI**
    |_ *ver:* 1.0.3
    |_ *key param:* NA

Submitter: Francisco Camara Ferreira
Affiliation: CNAG

Date and time: 2025-12-10 14:12:07 CET