

ERGA Assembly Report

v24.10.15

Tags: ERGA-BGE

TxID	1608717
ToLID	ilManCypr1
Species	<i>Maniola cypricola</i>
Class	Insecta
Order	Lepidoptera

Genome Traits	Expected	Observed
Haploid size (bp)	341,210,855	394,264,579
Haploid Number	28 (source: ancestor)	29
Ploidy	2 (source: ancestor)	2
Sample Sex	ZW	ZW

EBP metrics summary and curation notes

Obtained EBP quality metric for pri: 6.7.Q60

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 pri
- . Assembly length loss > 3% for pri
- . More than 1000 gaps/Gbp for pri

Curator notes

- . Interventions/Gb: 680
- . Contamination notes: "We found four small scaffolds with bacterial contaminants, two of which were labelled and removed using pretextview and the rapid curation scripts. The other two will be removed after the reviewing process."
- . Other observations: "We selected the primary assembly for this genome because neither haplotype achieved a cN50 of 1 Mb. Although purgeDups was used to manage duplicated regions, we subsequently reintroduced a substantial portion of the REPEAT labelled purged sequences in an attempt to recover the W chromosome, but this effort was unsuccessful. The curation process was relatively complex, requiring approximately 680 edits per gigabase. Across two rounds of manual curation, we removed 63 haplotigs. We also identified and labelled an additional haplotig in the PretextView map (mq0) that we have shared for review. This haplotig had been intentionally removed in an earlier iteration but was reintroduced in the final curated assembly because it had been contigged into a region containing a BUSCO gene. The haplotig was positioned in a region showing no clear HiC contacts to SUPER_24 and was therefore moved to the chaff. It also lay adjacent to a smaller segment aligning

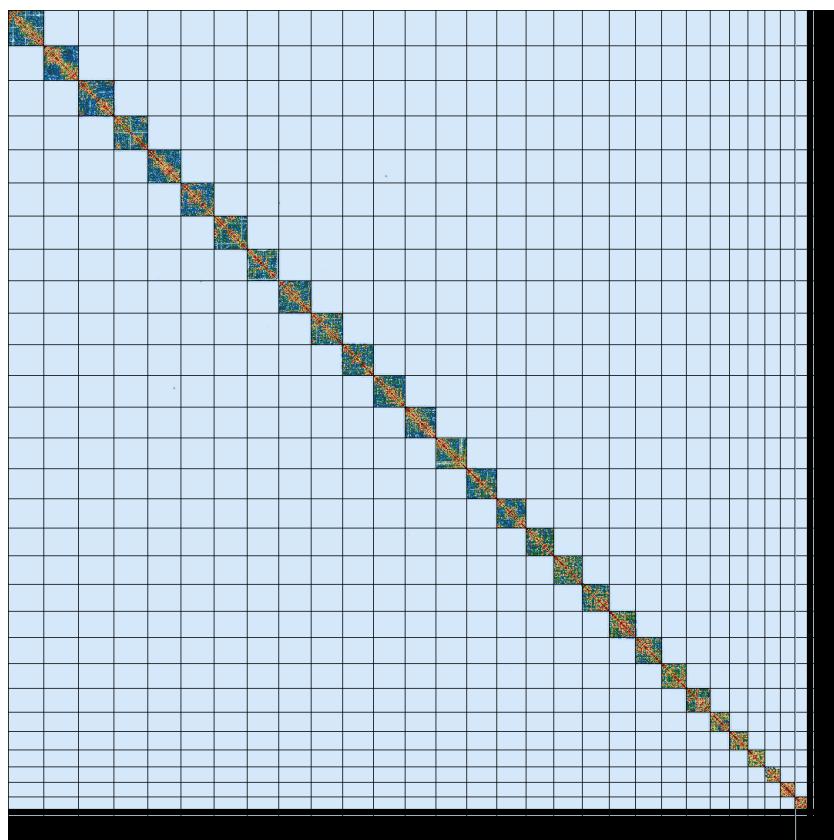
to SUPER_6, which contains the aforementioned BUSCO gene that had been lost during haplotig removal. We subsequently reassigned this short segment to SUPER_6. Note that we are sharing to save_states of our mq0 pretext map curation effort"

Quality metrics table

Metrics	Pre-curation pri	Curated pri
Total bp	409,130,404	394,264,579
GC %	36.75	36.76
Gaps/Gbp	1,290.54	1,496.46
Total gap bp	105,600	118,000
Scaffolds	395	342
Scaffold N50	14,195,737	14,681,771
Scaffold L50	14	13
Scaffold L90	28	26
Contigs	923	932
Contig N50	1,166,861	1,109,810
Contig L50	111	112
Contig L90	377	383
QV	60.7991	60.723
Kmer compl.	77.5562	76.3837
BUSCO sing.	93.5%	96.0%
BUSCO dupl.	4.1%	1.6%
BUSCO frag.	0.9%	0.9%
BUSCO miss.	1.5%	1.5%

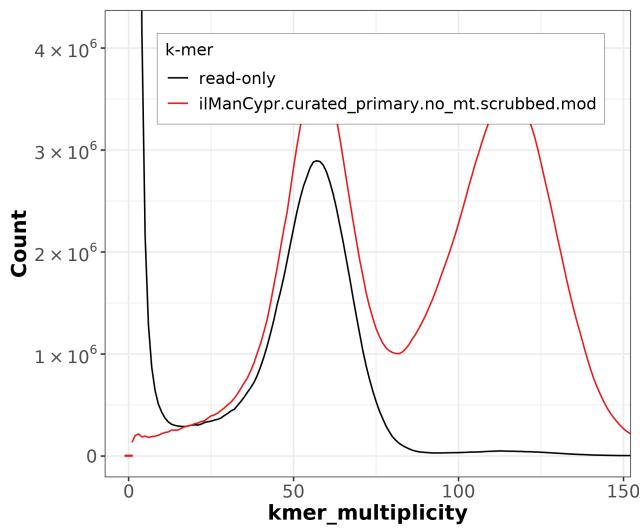
BUSCO: 5.5.0 (euk_genome_met, metaeuk) / Lineage: arthropoda_odb10 (genomes:90, BUSCOs:1013)

HiC contact map of curated assembly

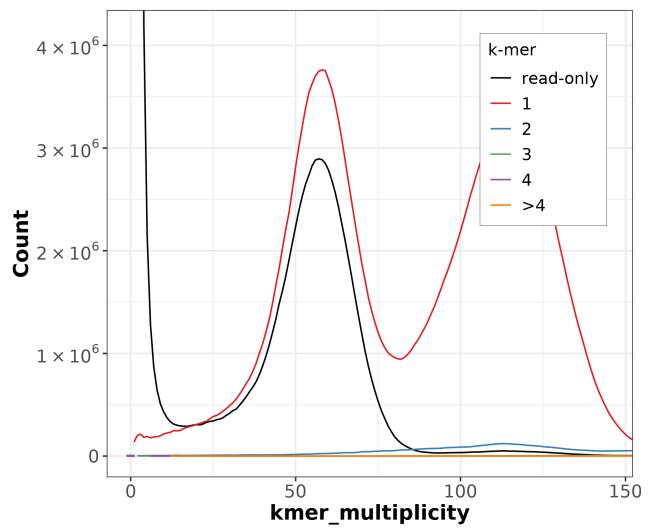


pri [LINK]

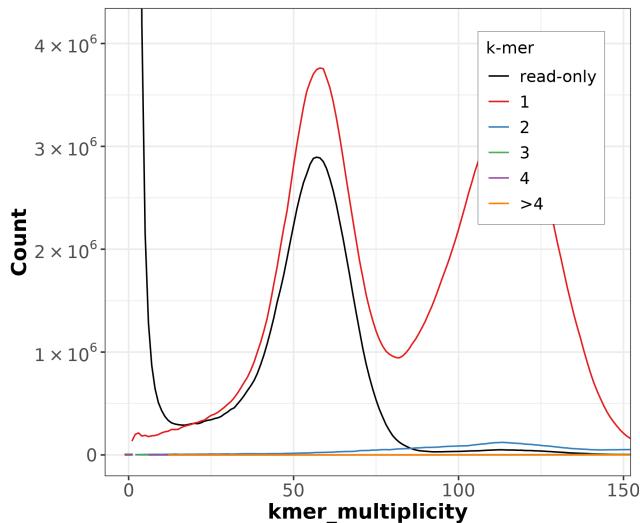
K-mer spectra of curated assembly



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

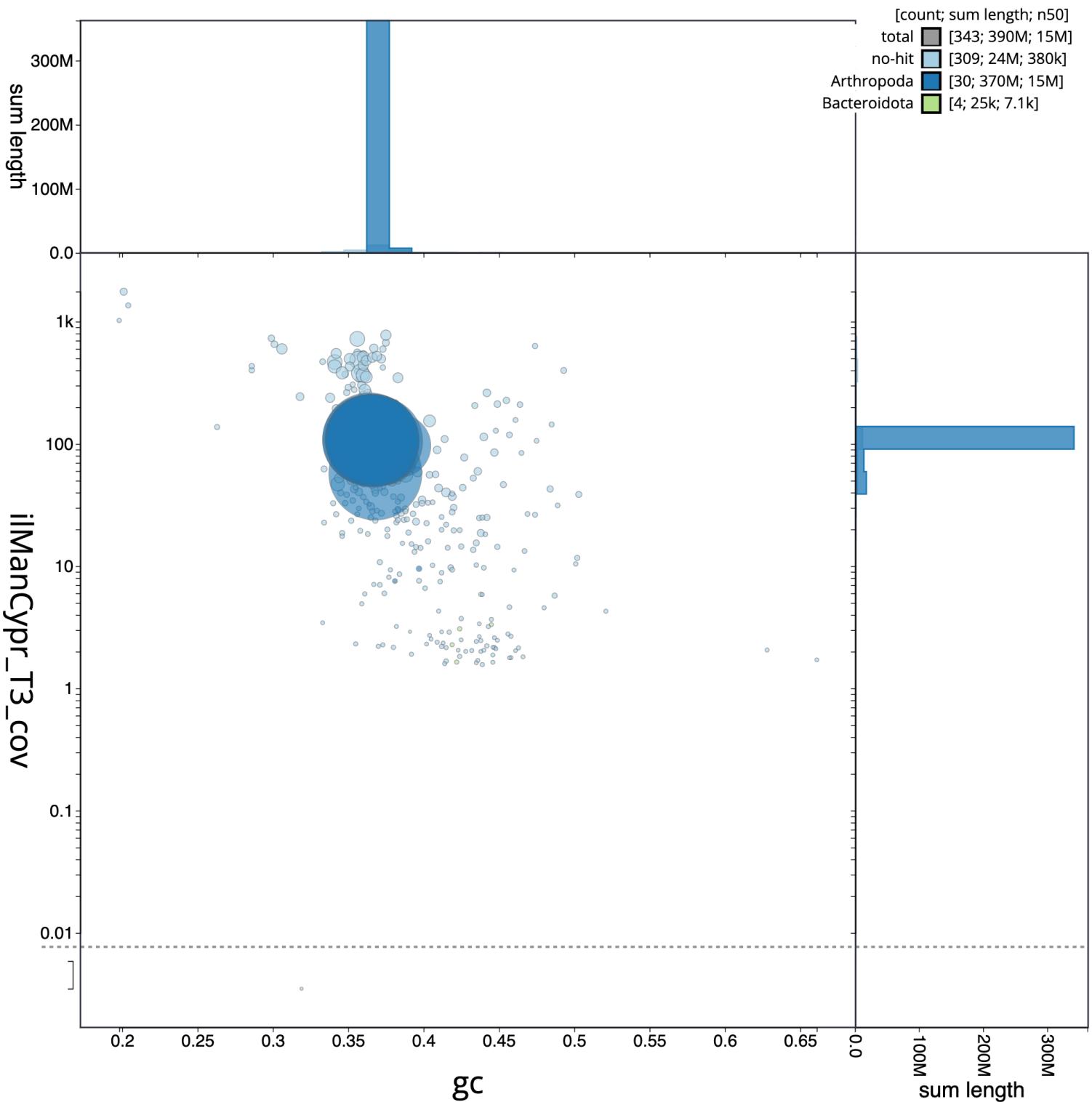


Distribution of k-mer counts per copy numbers found in asm



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	PACBIO HIFI	OmniC
Coverage	98x	90x

Assembly pipeline

```
- hifiasm
  |_ ver: 0.24.0-r702
  |_ key param: --primary
- purge_dups
  |_ ver: 1.2.6
  |_ key param: NA
- YaHS
  |_ ver: 1.2a
  |_ key param: NA
- CLAWS
  |_ ver: 2.3
  |_ key param: NA
```

Curation pipeline

```
- PretextViewAI
  |_ ver: 1.0.5
  |_ key param: NA
- GRIT_Rapid
  |_ ver: 2.0
  |_ key param: NA
```

Submitter: Francisco Camara

Affiliation: CNAG Barcelona

Date and time: 2026-01-26 14:56:25 CET