

# ERGA Assembly Report

v24.09.10

Tags: ERGA-BGE

TxID	554486
ToLID	<b>ilCoeOedi1</b>
Species	Coenonympha oedippus
Class	Insecta
Order	Lepidoptera

Genome Traits	Expected	Observed
Haploid size (bp)	400,913,152	397,581,223
Haploid Number	29 (source: direct)	29
Ploidy	2 (source: ancestor)	2
Sample Sex	ZW	ZW

## EBP metrics summary and curation notes

Obtained EBP quality metric for collapsed: 6.7.Q56

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

- . Kmer completeness value is less than 90 for collapsed

### Curator notes

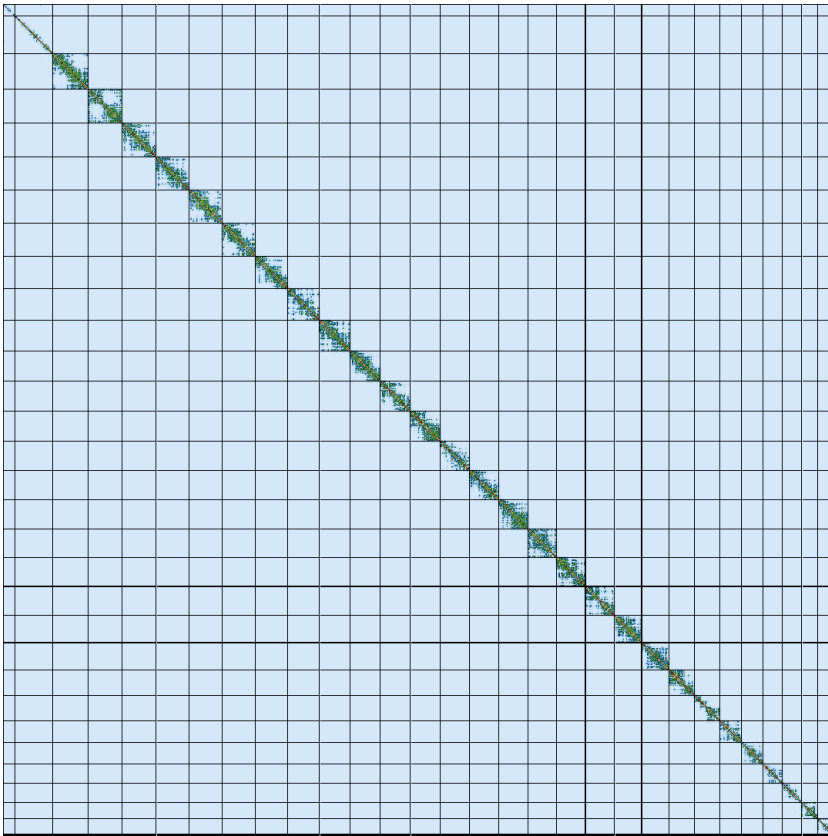
- . Interventions/Gb: 100
- . Contamination notes: "1 scaffold matching Wolbachia was removed from the Flye assembly"
- . Other observations: "We curated the Flye assembly with exception of the W scaffold that was better reconstructed by Nextdenovo. We replaced the Flye W scaffold for the nextdenovo W scaffold. In addition, 83 small scaffolds with remaining W fragments or just too short were removed from the curated Flye assembly. During the review, we made 2 additional cuts in contigs, 6 breaks at gaps and 12 joins localizing 5 contigs into superscaffolds"

# Quality metrics table

Metrics	Pre-curation collapsed	Curated collapsed
Total bp	398,249,265	397,581,223
GC %	37.32	37.32
Gaps/Gbp	778.41	827.5
Total gap bp	61,900	65,700
Scaffolds	165	58
Scaffold N50	14,195,870	14,242,255
Scaffold L50	13	13
Scaffold L90	26	26
Contigs	475	387
Contig N50	2,683,349	2,771,813
Contig L50	46	45
Contig L90	156	154
QV	55.0646	56.0202
Kmer compl.	71.2387	70.8353
BUSCO sing.	97.5%	97.6%
BUSCO dupl.	1.3%	1.2%
BUSCO frag.	0.2%	0.2%
BUSCO miss.	1.0%	1.0%

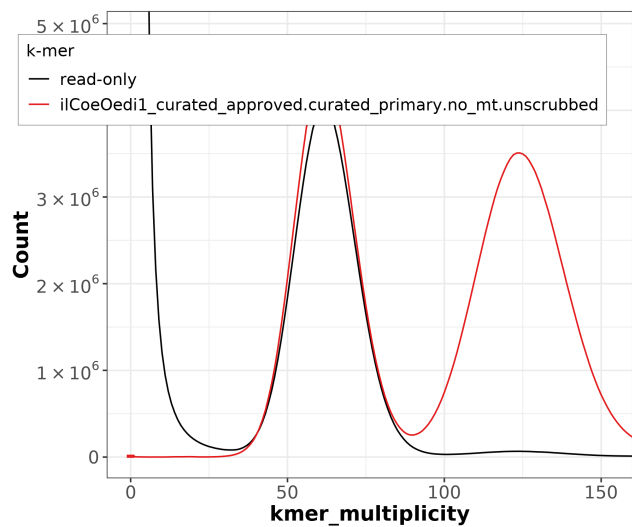
BUSCO 5.4.0 Lineage: insecta\_odb10 (genomes:75, BUSCOs:1367)

# HiC contact map of curated assembly

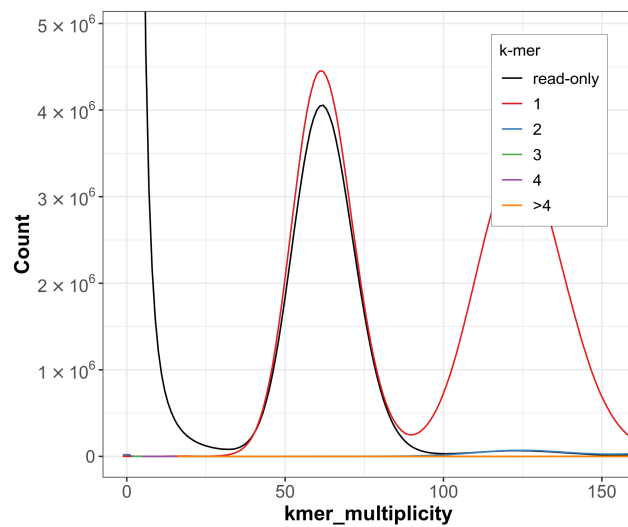


collapsed [\[LINK\]](#)

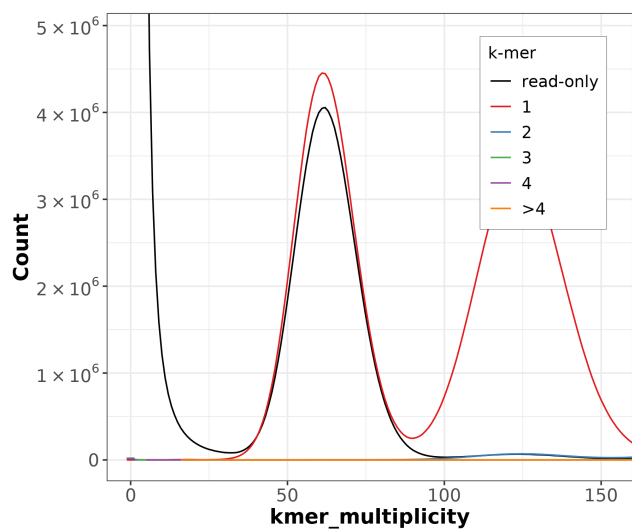
# K-mer spectra of curated assembly



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

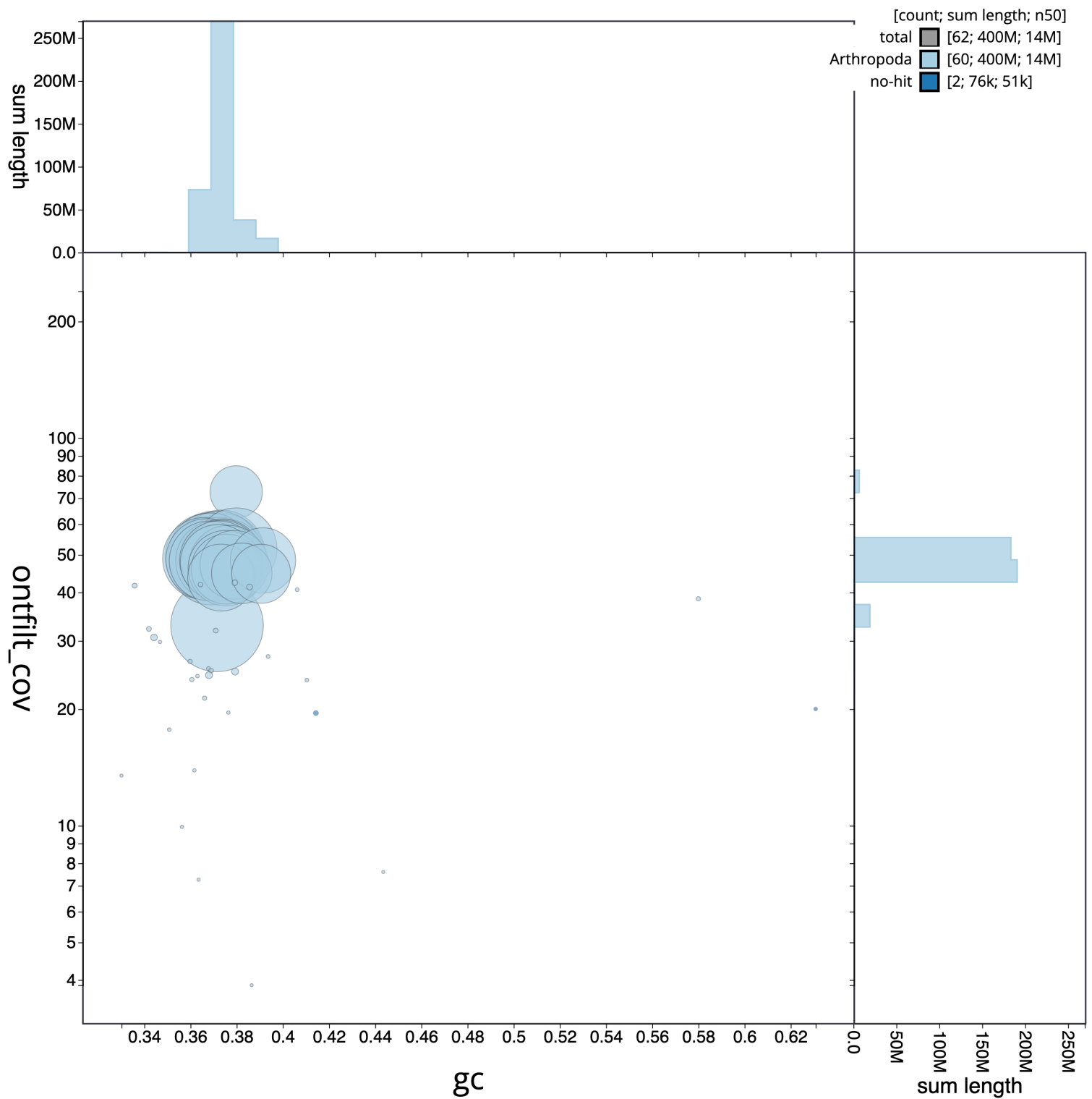


Distribution of k-mer counts per copy numbers found in asm



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

# Post-curation contamination screening



**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	ONT	Illumina	OmniC
Coverage	226x	75x	183x

# Assembly pipeline

- **Trim\_galore**
  - |\_ ver: 0.6.7
  - |\_ key param: --gzip
  - |\_ key param: =q 20
  - |\_ key param: --paired
  - |\_ key param: retain\_unpaired
  - |\_ key param: --max\_n
  - |\_ key param: =0
- **Filtlong**
  - |\_ ver: 0.2.1
  - |\_ key param: --minlen
  - |\_ key param: =1000
  - |\_ key param: --min\_mean\_q
  - |\_ key param: =80
  - |\_ key param: --target\_bases
  - |\_ key param: =27000000000
- **flye**
  - |\_ ver: 2.9.1
  - |\_ key param: NA
- **nextdenovo**
  - |\_ ver: 2.5.0
  - |\_ key param: NA
- **hypo**
  - |\_ ver: 1.0.3
  - |\_ key param: NA
- **purge\_dups**
  - |\_ ver: 1.2.6
  - |\_ key param: NA
- **YaHS**
  - |\_ ver: 1.2a
  - |\_ key param: NA

# Curation pipeline

- **PretextView**
  - |\_ ver: 0.2.5
  - |\_ key param: NA
- **GRIT\_Rapid**
  - |\_ ver: 2.0
  - |\_ key param: NA

Submitter: Fernando Cruz  
Affiliation: CNAG Barcelona

Date and time: 2024-11-04 11:58:04 CET