ERGA Assembly Report

v24.02.09_beta

Tags: ERGA-BGE

ToLID	**ilErePala4**
Species	**Erebia palarica**
Class	Insecta
Order	Lepidoptera

Genome Traits	Expected	Observed
Haploid size (bp)	483,621,817	493,162,700
Haploid Number	22 (source: ['ancestor'])	14
Ploidy	1 (source: ['ancestor'])	2
Sample Sex	Z0	Z0

EBP metrics summary and curation notes

Obtained EBP quality metric for collapsed: 7.7.Q44

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

. Observed Haploid Number is different from Expected
. Observed Ploidy is different from Expected
. Kmer completeness value is less than 90 for collapsed

Curator notes

. Interventions/Gb: 2
. Contamination notes: "what after curation we assumed to be SUPER_14 turned out to be the sequence of the insect endosymbiont Wolbachia & highly similar to Wolbachia pipiens (1.48 Mbases). Our Wolbachia sequence spans 1.55 Mbases"
. Other observations: "Yahs did an excellent initial job. Manual curation was minimal. Stats, merqury file and busco were obtained from the curated assembly -without- the wolbachia sequence. The post-curation contamination screening plot corresponds to the curated assembly before removal of the wolbachia sequence."

Quality metrics table

| Metrics	Pre-curation collapsed	Curated collapsed
Total bp	494,716,229	493,162,700
GC %	36.86	36.87
Gaps/Gbp	42.45	42.58
Total gap bp	4,200	4,200
Scaffolds	18	17
Scaffold N50	38,441,827	38,441,827
Scaffold L50	6	6
Scaffold L90	12	12
Contigs	39	38
Contig N50	34,188,440	34,188,440
Contig L50	6	6
Contig L90	16	16
QV	44.6729	44.6769
Kmer compl.	89.1241	88.7451
BUSCO sing.	96.6%	96.6%
BUSCO dupl.	1.3%	1.3%
BUSCO frag.	0.8%	0.8%
BUSCO miss.	1.3%	1.3%

BUSCO 5.4.0 Lineage: arthropoda_odb10 (genomes:90, BUSCOs:1013)
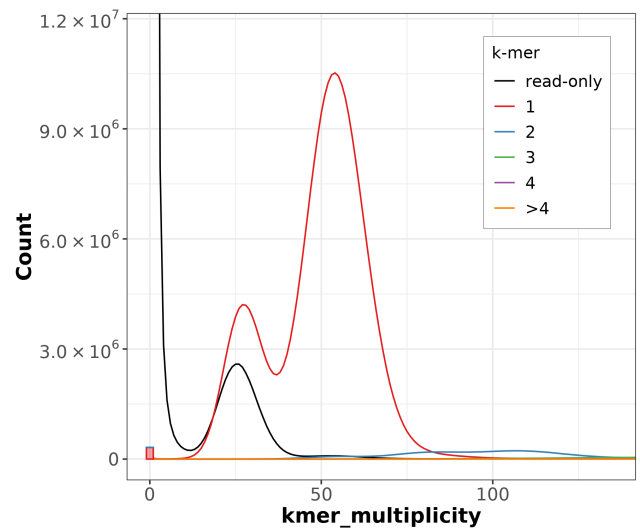
HiC contact map of curated assembly
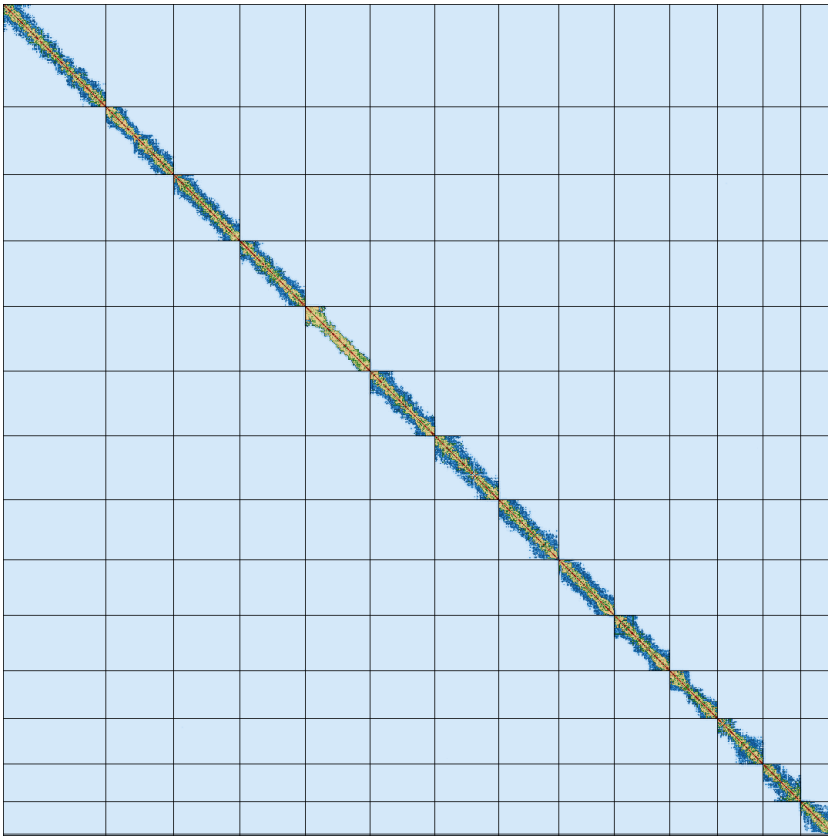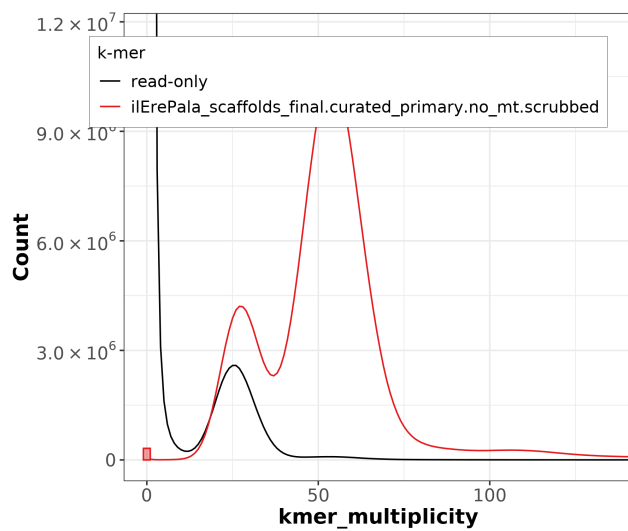

collapsed [LINK]

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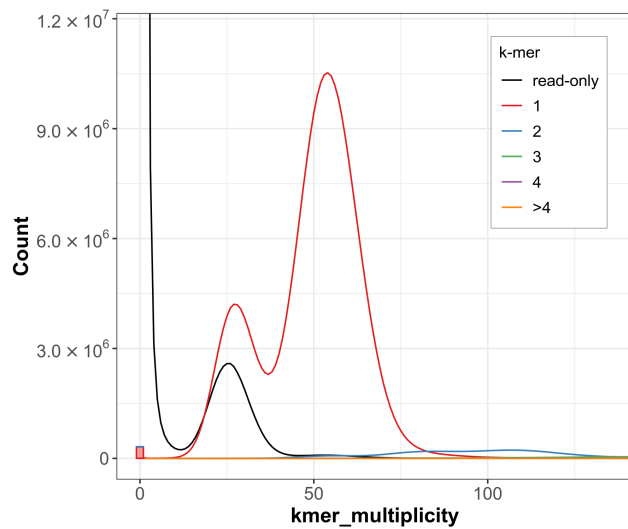
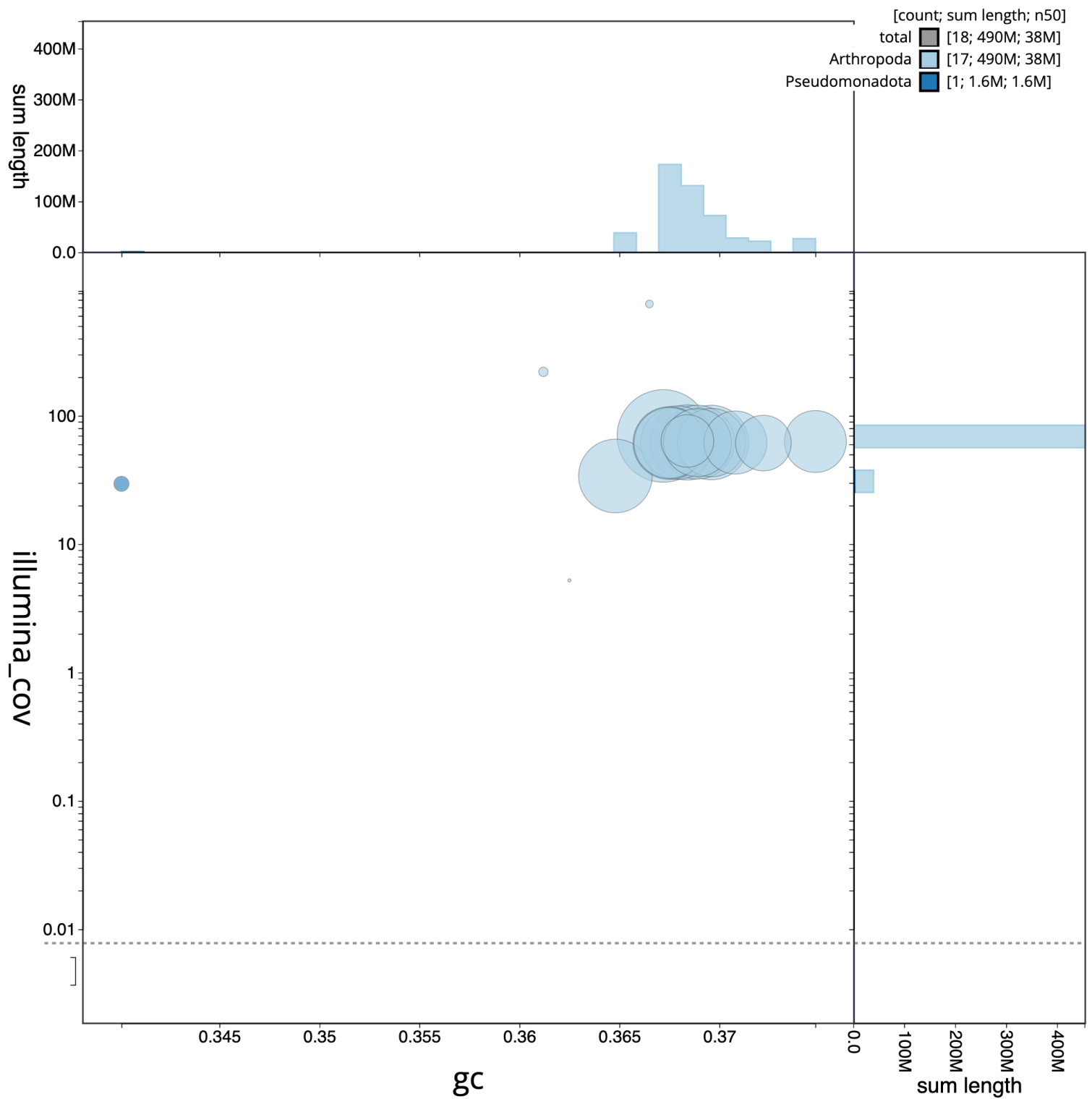Distribution of k-mer counts per copy
numbers found in asm

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

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	104x	64x	81x

Assembly pipeline

- **Trim_Galore**
  |_ *ver:* 0.6.7
  |_ *key param:* "--gzip -q 20"
  |_ *key param:* "--paired"
  |_ *key param:* "--retain_unpaired"
- **Filtlong**
  |_ *ver:* 0.2.1
  |_ *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

Submitter: Francisco Camara
Affiliation: CNAG Barcelona

Date and time: 2024-02-19 11:38:26 CET