ERGA Assembly Report

v24.09.10

Tags: ERGA-BGE

TxID	3078426
ToLID	**ibLobSubt1**
Species	Loboptera subterranea
Class	Insecta
Order	Blattodea

Genome Traits	Expected	Observed
Haploid size (bp)	2,449,268,478	2,671,312,452
Haploid Number	17 (source: ancestor)	14
Ploidy	2 (source: ancestor)	2
Sample Sex	XX	XX

EBP metrics summary and curation notes

Obtained EBP quality metric for collapsed: 7.8.Q44

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 collapsed

Curator notes

. Interventions/Gb: 7
. Contamination notes: "Several sequences corresponding to to an intracellular fat body endosymbiont (Blattabacterium) were removed. In a couple of these cases the symbiont sequence were found and remvoved from super scaffolds (SUPER_9 and SUPER_X). In most other cases these sequences span several smaller scaffolds that were fully removed. Moreover, after our internal curation we still found that the following scaffolds, which will be removed from the final version of the assembly also contain symbiont (or other) types of contamination: scaffold_30, scaffold_59, scaffold_96, scaffold_85, scaffold_48,scaffold_39, scaffold_61, scaffold_21, scaffold_55, scaffold_45 and scaffold_49. These have already been removed from the blobtools plot included in the report"
. Other observations: "The original pre-curation assembly was already quite good even though our HiC data contained a higher than desirable proportion of Cis read pairs < 1kb. Interestingly contrary to most other CNAG projects, in this case the ONT data derived from the homomorphic sex (female XX) and the HiC data was derived from the heteromorphic sex (XY). We apoloize but our pretext was not run with the -nosort option so that the few unlocs were placed at the end. However after the review we will build a pretext in which the SUPER scaffolds are kept together. This final EAR report

includes one modification suggested by the reviewers, which consisted of the inverting the contig at the end of SUPER_8. The report is therefore the same except for the busco and gfastats outputs. The -nosort- pretext was shared with the save state that was shared with the reviewers."

Quality metrics table

| Metrics	Pre-curation collapsed	Curated collapsed
Total bp	2,680,084,542	2,671,312,452
GC %	34.38	34.38
Gaps/Gbp	147.38	148.24
Total gap bp	79,000	79,200
Scaffolds	117	90
Scaffold N50	202,725,563	202,725,563
Scaffold L50	6	6
Scaffold L90	12	12
Contigs	512	486
Contig N50	13,138,595	13,138,595
Contig L50	60	60
Contig L90	205	202
QV	44.8731	44.8824
Kmer compl.	87.2601	87.0387
BUSCO sing.	97.3%	97.3%
BUSCO dupl.	1.7%	1.7%
BUSCO frag.	0.4%	0.4%
BUSCO miss.	0.6%	0.6%

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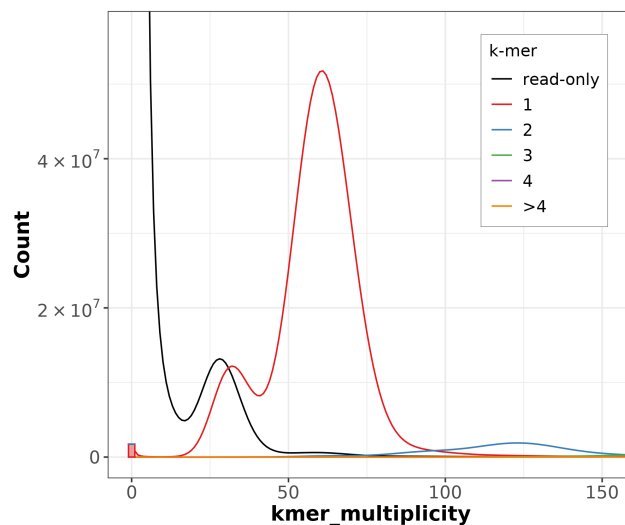
HiC contact map of curated assembly
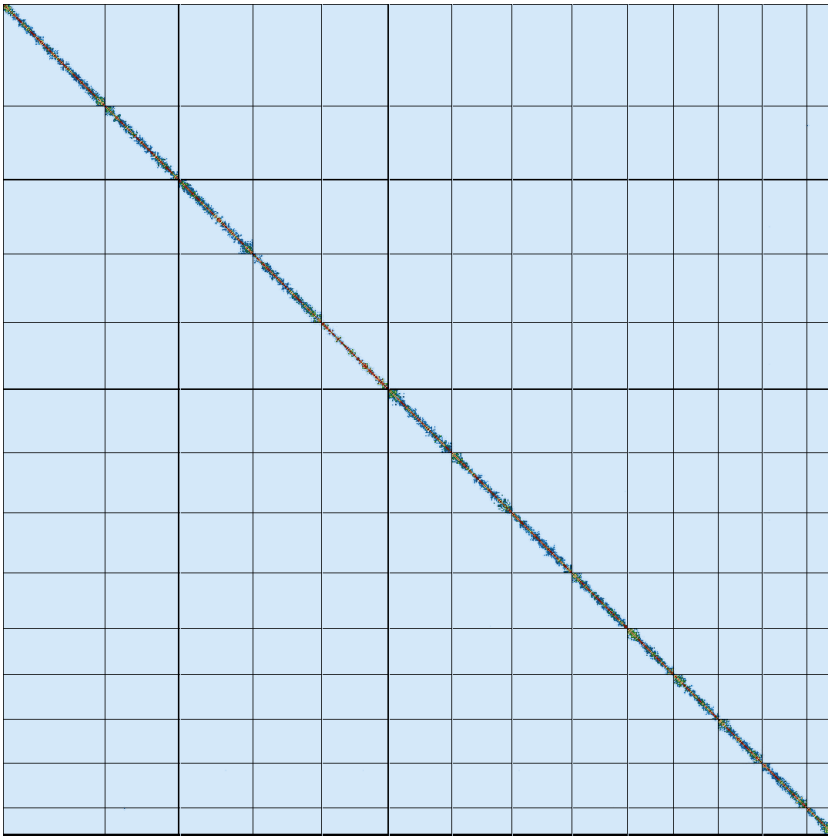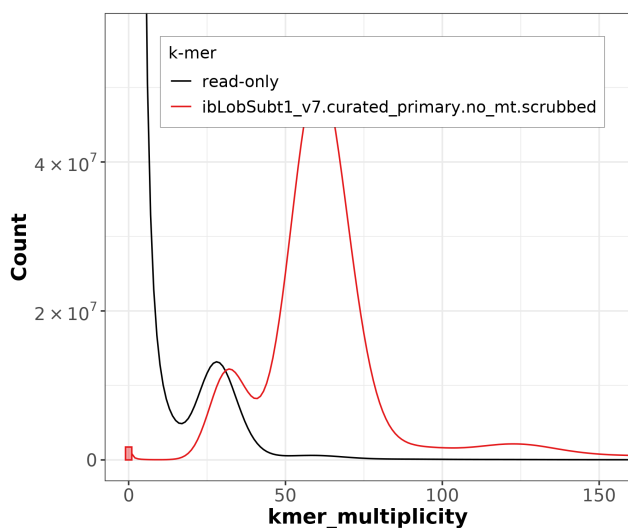

collapsed [LINK]

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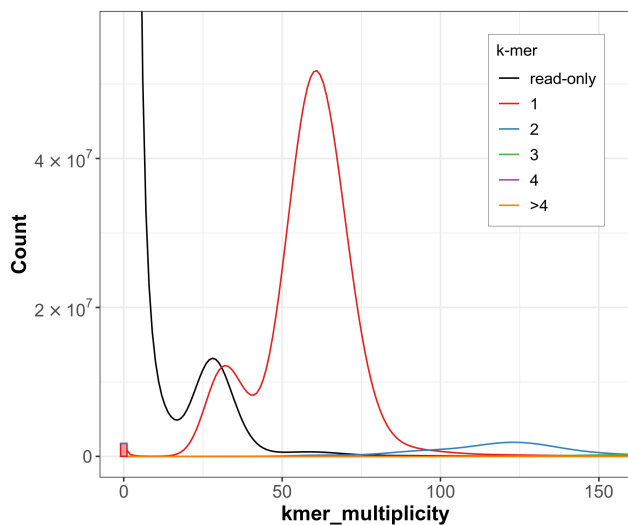
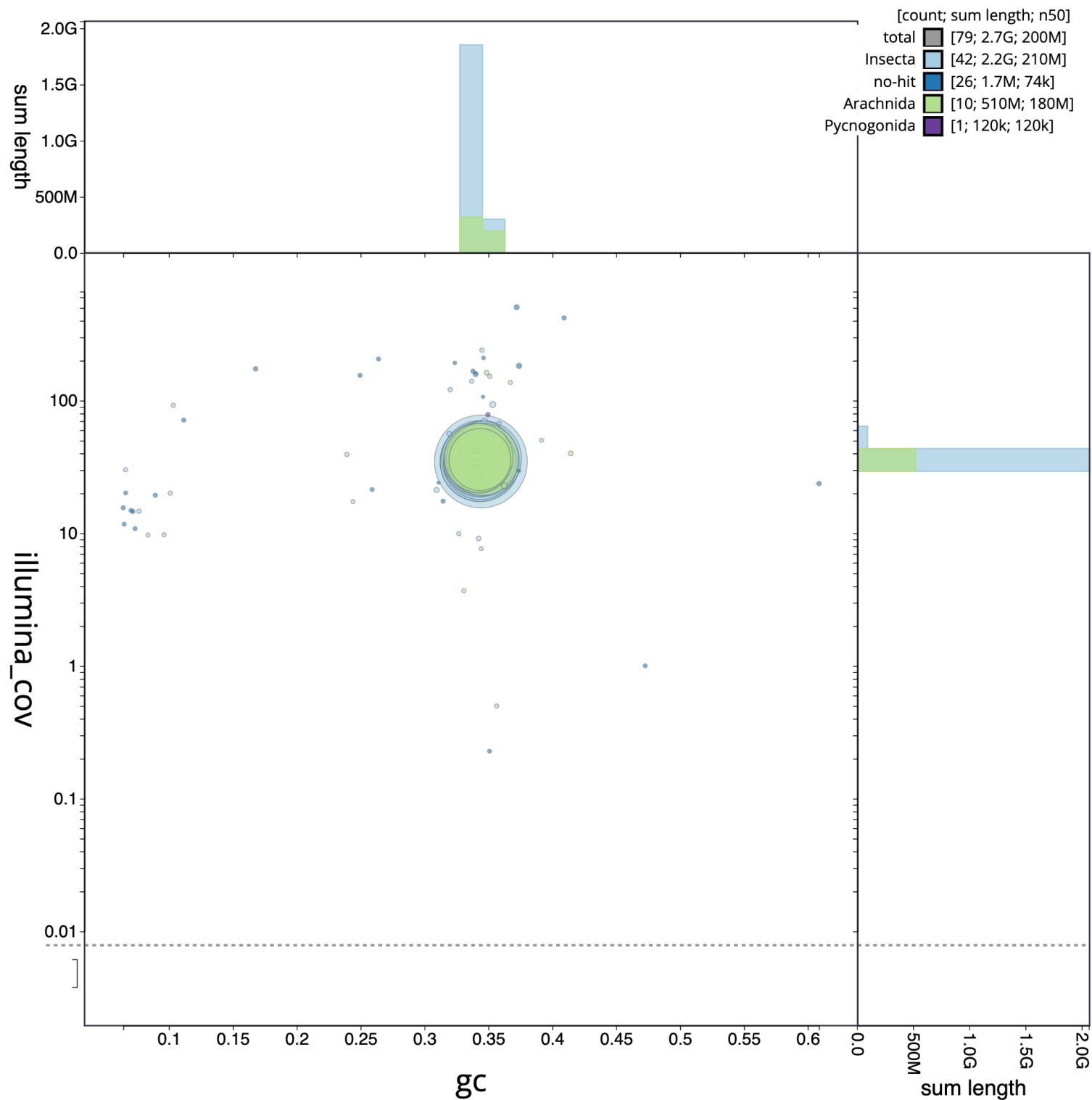Distribution of k-mer counts per copy
numbers found in asm

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	46x	38x	41x

Assembly pipeline

- **Trim_galore**
  - |_ *ver:* 0.6.7
  - |_ *key param:* --gzip
  - |_ *key param:* =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
- **GRIT_Rapid**
  - |_ *ver:* 2.0
  - |_ *key param:* NA

Submitter: Francisco Camara
Affiliation: CNAG Barcelona

Date and time: 2024-10-16 11:49:26 CEST