

ERGA Assembly Report

v24.09.10

Tags: ERGA-BGE

TxID	3023605
ToLID	icCarBano7
Species	Carabus banonii
Class	Insecta
Order	Coleoptera

Genome Traits	Expected	Observed
Haploid size (bp)	71,013,296	207,825,851
Haploid Number	14 (source: ancestor)	15
Ploidy	2 (source: ancestor)	2
Sample Sex	Unknown	Unknown

EBP metrics summary and curation notes

Obtained EBP quality metric for collapsed: 6.7.Q57

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

- . Observed Haploid size (bp) has >20% difference with Expected
- . Observed Haploid Number is different from Expected
- . Kmer completeness value is less than 90 for collapsed

Curator notes

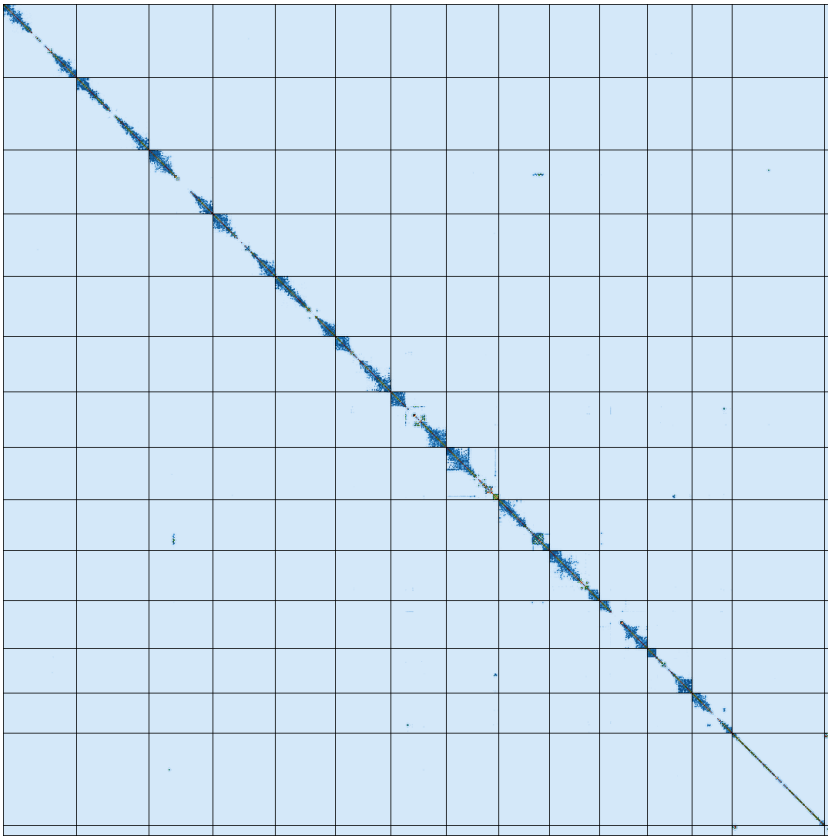
- . Interventions/Gb: 426
- . Contamination notes: "None bacterial contamination. But some small scaffolds for which no HiC links was found were removed "
- . Other observations: "Presence of larges repeated regions (telomeric or centromeric?) at the ends of the scaffolds,therefore need to invert the arms of chromosomes "

Quality metrics table

Metrics	Pre-curation collapsed	Curated collapsed
Total bp	208,441,780	207,825,851
GC %	35.43	35.39
Gaps/Gbp	62.37	336.82
Total gap bp	1,300	13,800
Scaffolds	83	19
Scaffold N50	12,327,626	14,906,334
Scaffold L50	8	6
Scaffold L90	15	13
Contigs	96	89
Contig N50	7,793,566	7,491,500
Contig L50	12	12
Contig L90	27	28
QV	40.112	57.4508
Kmer compl.	80.0782	81.1154
BUSCO sing.	97.3%	99.2%
BUSCO dupl.	2.4%	0.4%
BUSCO frag.	0.4%	0.4%
BUSCO miss.	0.1%	0.0%

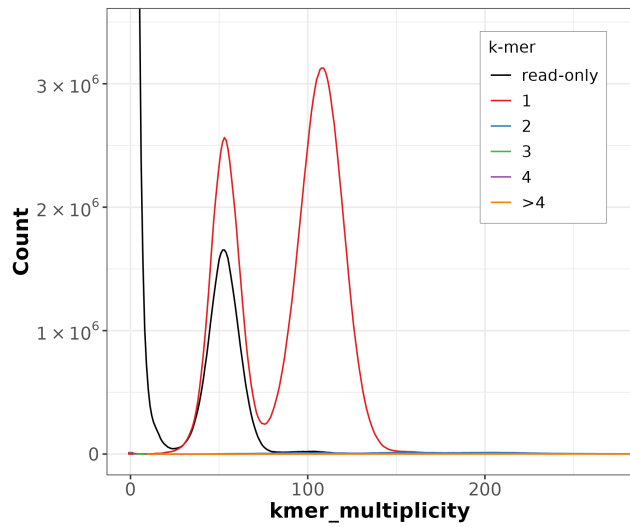
BUSCO 5.4.3 Lineage: eukaryota_odb10 (genomes:70, BUSCOs:255)

HiC contact map of curated assembly

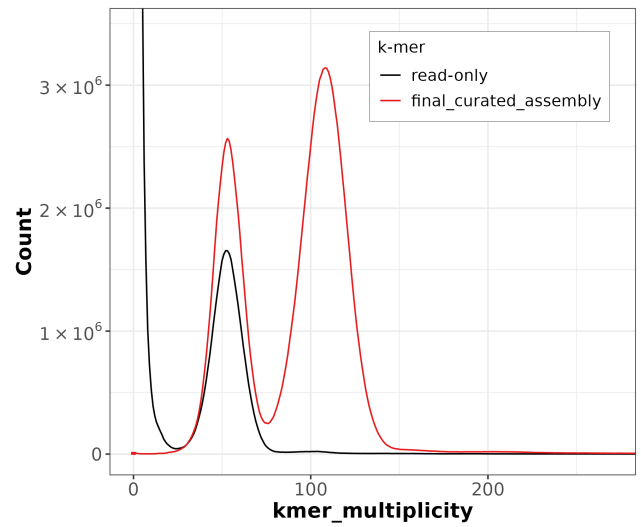


collapsed [\[LINK\]](#)

K-mer spectra of curated assembly

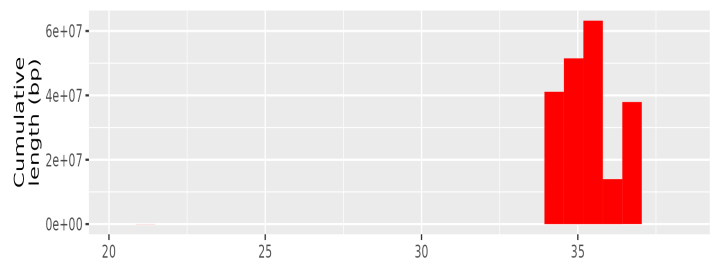


Distribution of k-mer counts per copy numbers found in asm

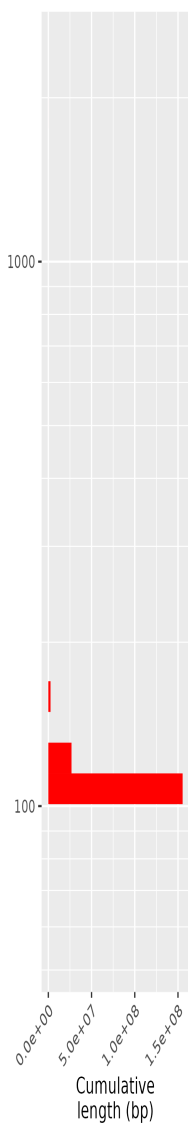
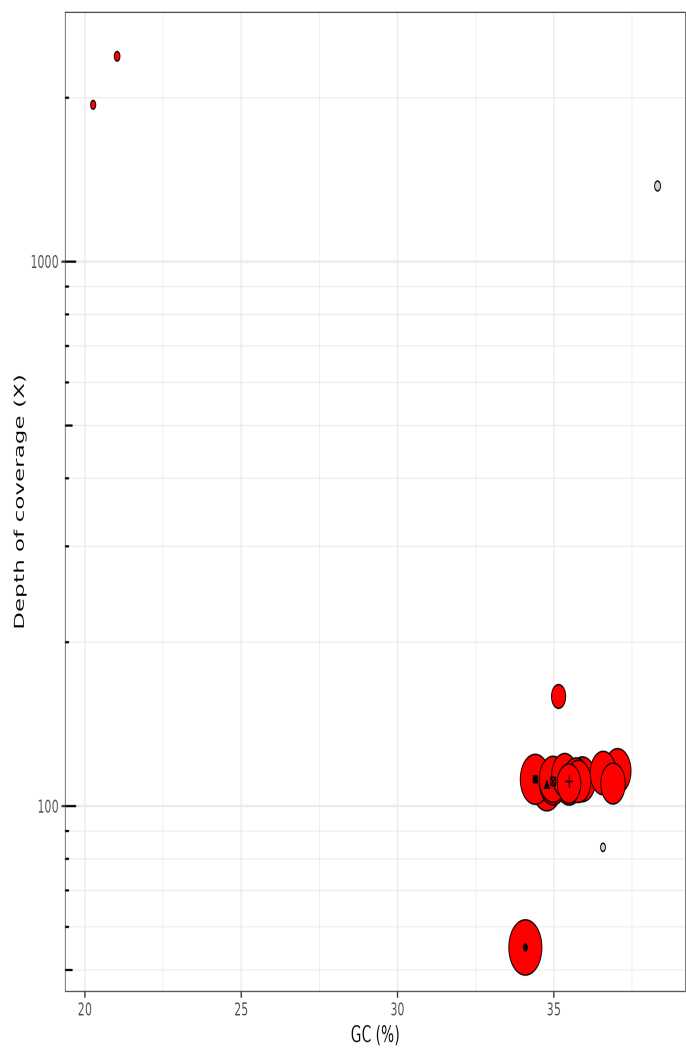


Distribution of k-mer counts coloured by their presence in reads/assemblies

Post-curation contamination screening



TAPAs summary Graph



- Longest sequences (bp)
- X - 22990679 (Eukaryota)
 - ▲ SUPER_1 - 18430360 (Eukaryota)
 - SUPER_2 - 18126551 (Eukaryota)
 - + SUPER_3 - 15914161 (Eukaryota)
 - ▣ SUPER_4 - 15633705 (Eukaryota)

- superkingdom
- Eukaryota
 - N/A

- Length (bp)
- 5.0e+06
 - 1.0e+07
 - 1.5e+07
 - 2.0e+07

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	PACBIO Hifi	Arima (4-enz)
Coverage	30	50

Assembly pipeline

- **Hifiasm**
 - |_ *ver*: 0.19.5-r593
 - |_ *key param*: NA
- **purge_dups**
 - |_ *ver*: 1.2.5
 - |_ *key param*: NA
- **YaHS**
 - |_ *ver*: 1.2
 - |_ *key param*: NA

Curation pipeline

- **PretextMap**
 - |_ *ver*: 0.1.9
 - |_ *key param*: NA
- **PretextView**
 - |_ *ver*: 0.2.5
 - |_ *key param*: NA

Submitter: Arnaud Couloux

Affiliation: Genoscope

Date and time: 2024-10-22 09:28:26 CEST