

ERGA Assembly Report

v24.10.15

Tags: ERGA-BGE

TxID	2995250
ToLID	imBolBrac2
Species	<i>Bolivaria brachyptera</i>
Class	Insecta
Order	Mantodea

Genome Traits	Expected	Observed
Haploid size (bp)	3,105,851,667	3,299,746,198
Haploid Number	8 (source: ancestor)	15
Ploidy	2 (source: ancestor)	2
Sample Sex	Unknown	Unknown

EBP metrics summary and curation notes

Obtained EBP quality metric for collapsed: 7.8.Q66

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: 5
- Contamination notes: ""
- Other observations: "The assembly of *Bolivaria brachyptera* (imBolBrac2) is based on 30X PacBio data and Arima Hi-C data generated as part of the European Reference Genome Atlas (ERGA, <https://www.erga-biodiversity.eu/>) via the Biodiversity Genomics Europe project (BGE, <https://biodiversitygenomics.eu/>). The assembly process included the following steps: initial PacBio assembly generation with Hifiasm, removal of contaminant sequences using Context, removal of haplotypic duplications using purge_dups, and Hi-C-based scaffolding with YaHS. In total, no contigs were identified as contaminants (bacterial, archaeal, or viral). Additionally, 679 regions totaling 185 Mb (with the largest being 8.4 Mb) were identified as haplotypic duplications and removed. The mitochondrial genome was assembled using OATK. Finally, the primary assembly was analyzed and manually improved using Pretext. During manual curation, 3 haplotypic regions were removed, totaling 2.3 Mb (with the largest being 1.2 Mb. Chromosome-scale scaffolds confirmed by Hi-C data were named in order of size. The X chromosome was identified thanks to comparison to another insecta genomes. "

Quality metrics table

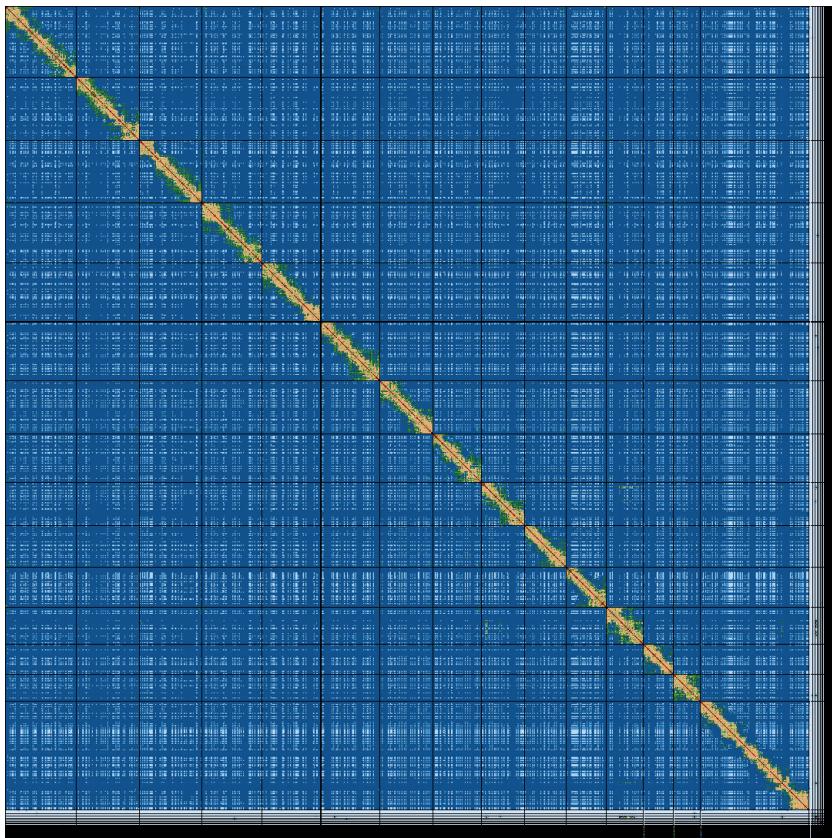
Metrics	Pre-curation collapsed	Curated collapsed
Total bp	3,303,298,582	3,299,746,198
GC %	37.91	37.91
Gaps/Gbp	84.46	82.13
Total gap bp	30,800	31,800
Scaffolds	221	209
Scaffold N50	232,375,610	235,080,318
Scaffold L50	6	6
Scaffold L90	14	14
Contigs	492	480
Contig N50	30,793,000	32,683,822
Contig L50	27	27
Contig L90	118	117
QV	66.8103	66.7904
Kmer compl.	88.7622	88.7313
BUSCO sing.	92.5%	96.8%
BUSCO dupl.	1.7%	1.6%
BUSCO frag.	3.7%	0.3%
BUSCO miss.	2.1%	1.3%

Warning! BUSCO versions or lineage datasets are not the same across results:

BUSCO: 5.8.2 (euk_genome_met, metaeuk) / Lineage: insecta_odb12 (genomes:79, BUSCOs:3114)

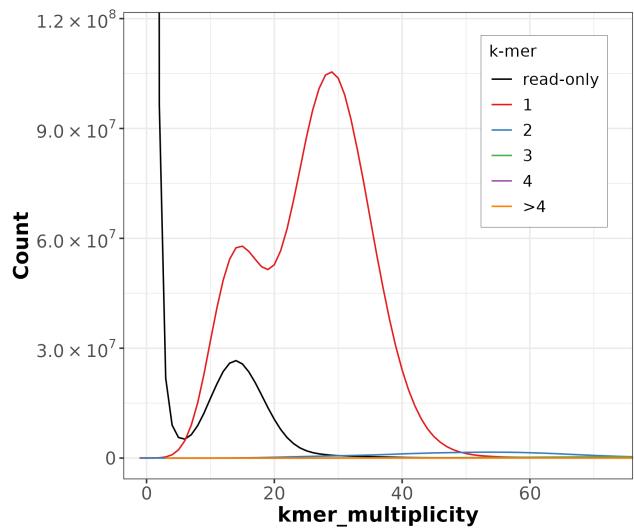
BUSCO: 6.0.0 (euk_genome_min, miniprot) / Lineage: insecta_odb12 (genomes:79, BUSCOs:3114)

HiC contact map of curated assembly

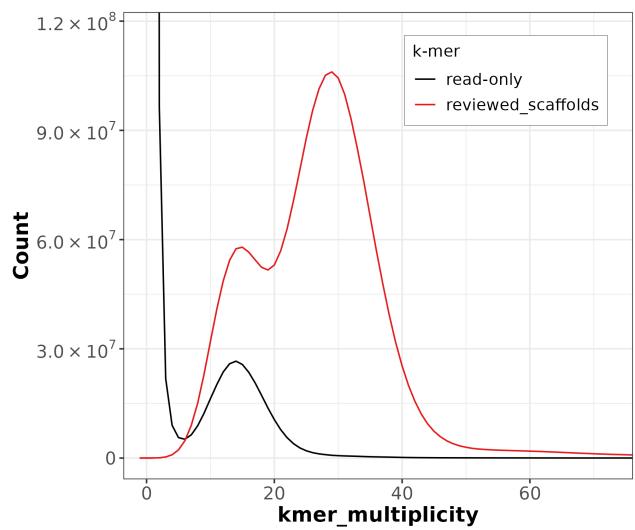


collapsed [\[LINK\]](#)

K-mer spectra of curated assembly

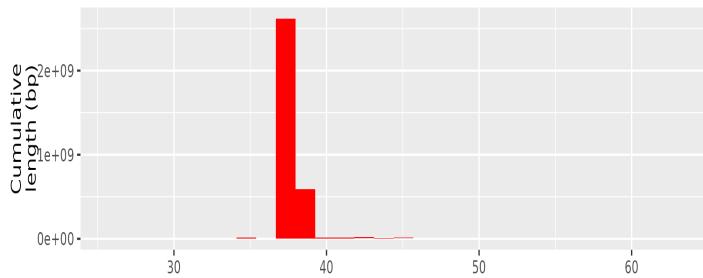


Distribution of k-mer counts per copy numbers found in asm

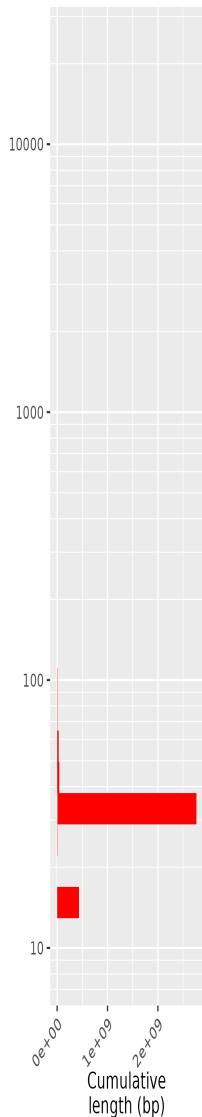
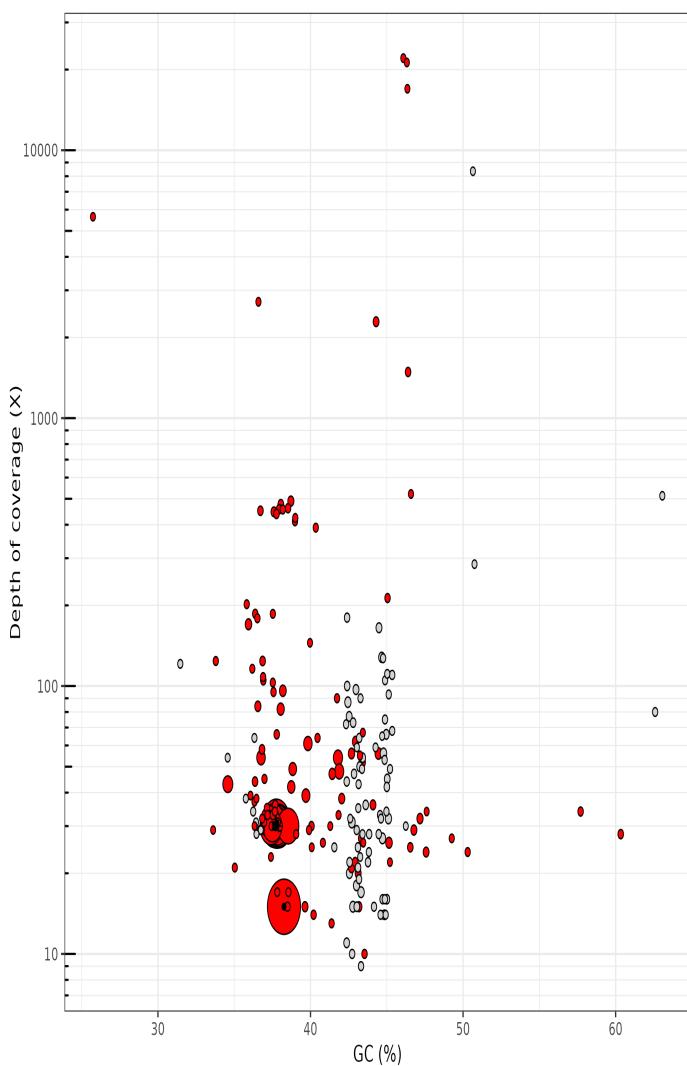


Distribution of k-mer counts coloured by their presence in reads/assemblies

Post-curation contamination screening



TAPAs summary Graph



Length (bp)

- $1e+08$
- $2e+08$
- $3e+08$
- $4e+08$

Longest sequences (bp)

- imBolBrac2_X - 429139322 (Eukaryota)
- ▲ imBolBrac2_1 - 282112534 (Eukaryota)
- imBolBrac2_2 - 252251645 (Eukaryota)
- + imBolBrac2_3 - 246331991 (Eukaryota)
- ▣ imBolBrac2_4 - 235945726 (Eukaryota)

superkingdom

- Eukaryota
- N/A

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	Long reads	Arima
Coverage	30	173

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: Caroline Menguy

Affiliation: Genoscope

Date and time: 2025-10-08 23:45:41 CEST