

ERGA Assembly Report

v24.04.03_beta

Tags: ERGA-BGE

TxID	62148
ToLID	qqGluDors1.3
Species	<i>Gluvia dorsalis</i>
Class	Arachnida
Order	Solifugae

Genome Traits	Expected	Observed
Haploid size (bp)	790,185,518	787,019,465
Haploid Number	11 (source: ancestor)	5
Ploidy	2 (source: ancestor)	2
Sample Sex	unknown	unknown

EBP metrics summary and curation notes

Obtained EBP quality metric for collapsed: 7.8.Q48

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

- . Observed Haploid Number is different from Expected

Curator notes

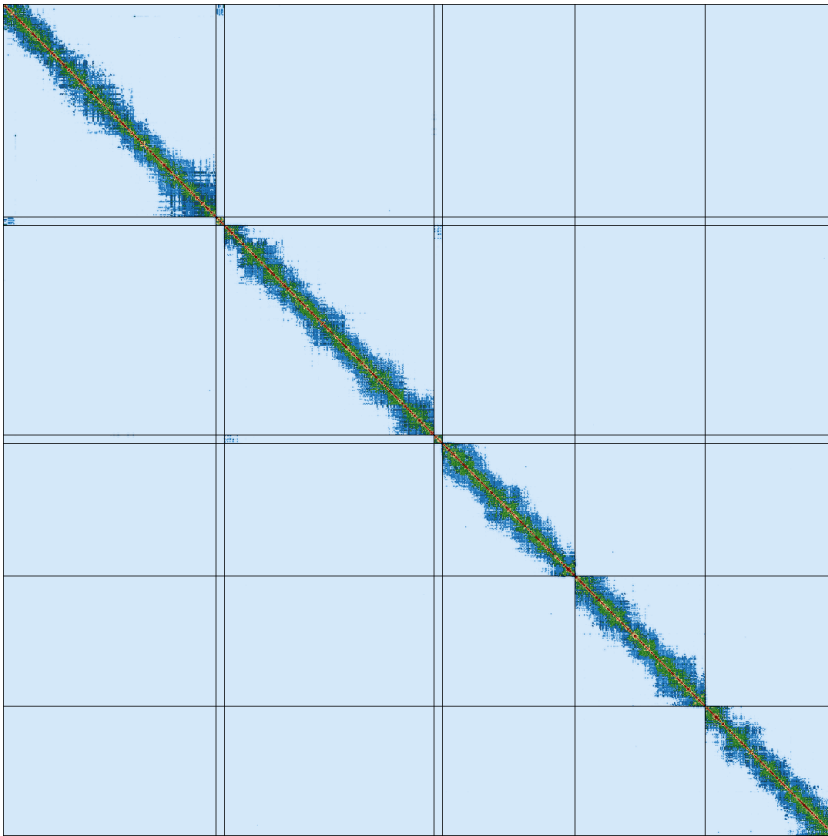
. Interventions/Gb: 7.62
. Contamination notes: "BTK detected 3 contaminant scaffolds matching bacteria (nt database) and they were discarded. Scaffold_12 was considered a false positive contaminant and retained (has no matches in nt and a short hit to Rotifera in uniprot). All other sequences were assigned to phylum Arthropoda by BTK. The complete mitochondrial genome was assembled into a single circular contig of 14,734 base pairs with excellent base accuracy using the FOAM pipeline (<https://github.com/cnag-aat/FOAM>). Following current SAC decision the MT is not included in this assembly version but it will be uploaded to ENA."
. Other observations: "Genome Assembly obtained with CLAWS v2.1 pipeline (<https://workflowhub.eu/workflows/567>). Input assembly (nextdenovo.hypol.purged) for HiC scaffolding was already highly contiguous (N50=47.35Mb), YaHS placed all contigs into 14 scaffolds. Manual curation involved localizing two scaffolds into SUPER_5 and SUPER_6 (made 2 cuts in contigs, 0 breaks at gaps and 4 joins). After review, two SUPER (qqGluDors1.2) of approximately 7Mb were placed at SUPER_1 and SUPER_2 and tagged as Unloc. We could not tag the sex chromosome(s) as all SUPER show diploid coverage. Two scaffolds remain unplaced, they are repetitive and show multiple contacts to several SUPER. The hic_FullMap_link points to a folder containing two pretext maps with default and high resolution, respectively."

Quality metrics table

Metrics	Pre-curation collapsed	Curated collapsed
Total bp	788,347,192	787,019,465
GC %	39.71	39.73
Gaps/Gbp	46.93	52.1
Total gap bp	7,400	8,200
Scaffolds	14	9
Scaffold N50	198,509,873	198,509,873
Scaffold L50	2	2
Scaffold L90	5	5
Contigs	51	50
Contig N50	37,604,012	37,604,012
Contig L50	8	8
Contig L90	26	26
QV	48.0932	48.093
Kmer compl.	95.6018	95.3951
BUSCO sing.	97.3%	97.3%
BUSCO dupl.	0.8%	0.8%
BUSCO frag.	1.4%	1.4%
BUSCO miss.	0.5%	0.5%

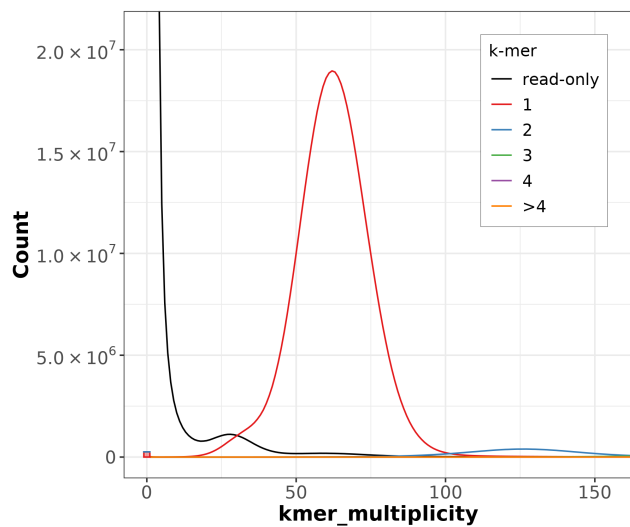
BUSCO 5.4.0 Lineage: arthropoda_odb10 (genomes:90, BUSCOs:1013)

HiC contact map of curated assembly

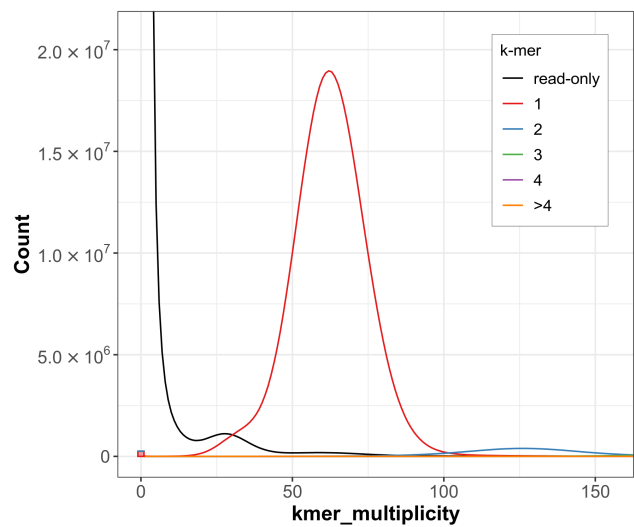


collapsed [\[LINK\]](#)

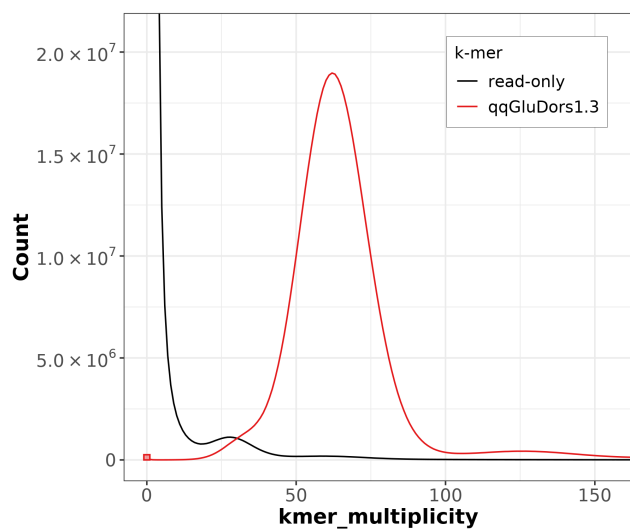
K-mer spectra of curated assembly



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

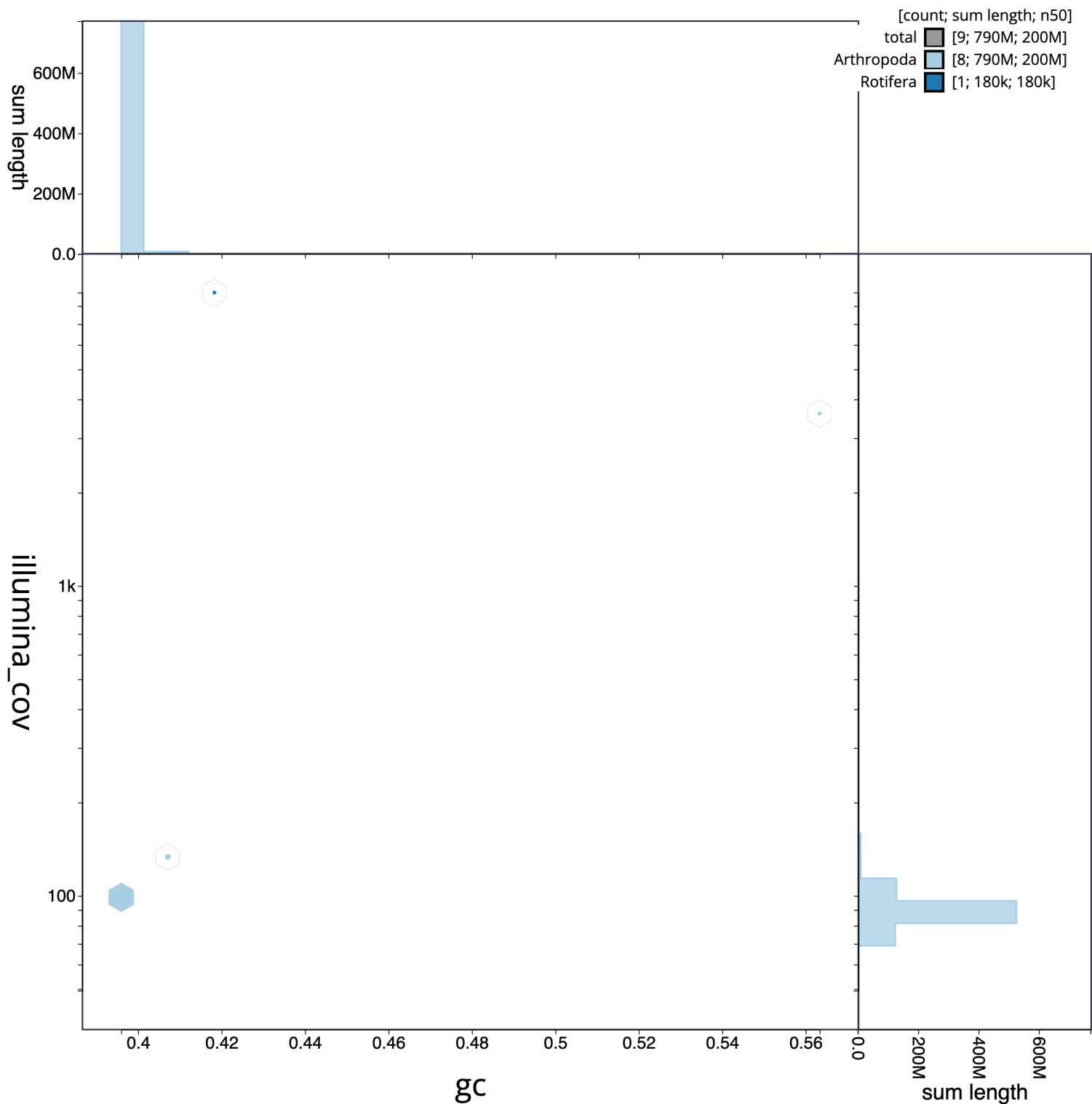


Distribution of k-mer counts per copy numbers found in asm



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	116x	102x	97x

Assembly pipeline

- **Trim_Galore**
 - |_ *ver*: 0.6.7
 - |_ *key param*: "--gzip -q 20"
 - |_ *key param*: "--paired"
 - |_ *key param*: "--retain_unpaired"
 - |_ *key param*: "--max_n 0"
- **Filtlong**
 - |_ *ver*: 0.2.1
 - |_ *key param*: "--target_bases 700000000000"
- **nextdenovo**
 - |_ *ver*: 2.5.0
 - |_ *key param*: NA
- **hypo**
 - |_ *ver*: 1.0.3
 - |_ *key param*: "1 round polishing"
- **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: Fernando Cruz

Affiliation: CNAG Barcelona

Date and time: 2024-04-17 14:21:35 CEST