

ERGA Assembly Report

v24.10.15

Tags: ERGA-BGE

TxID	119148
ToLID	qqDysUngu2
Species	Dysdera unguimmanis
Class	Arachnida
Order	Araneae

Genome Traits	Expected	Observed
Haploid size (bp)	1,553,664,201	1,824,778,328
Haploid Number	11 (source: ancestor)	8
Ploidy	2 (source: ancestor)	3
Sample Sex	XX	NA

EBP metrics summary and curation notes

Obtained EBP quality metric for Pri: 6.8.Q59

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
- . Observed sex is different from Sample sex
- . BUSCO single copy value is less than 90% for Pri
- . BUSCO duplicated value is more than 5% for Pri
- . Assembly length loss > 3% for Pri
- . More than 1000 gaps/Gbp for Pri

Curator notes

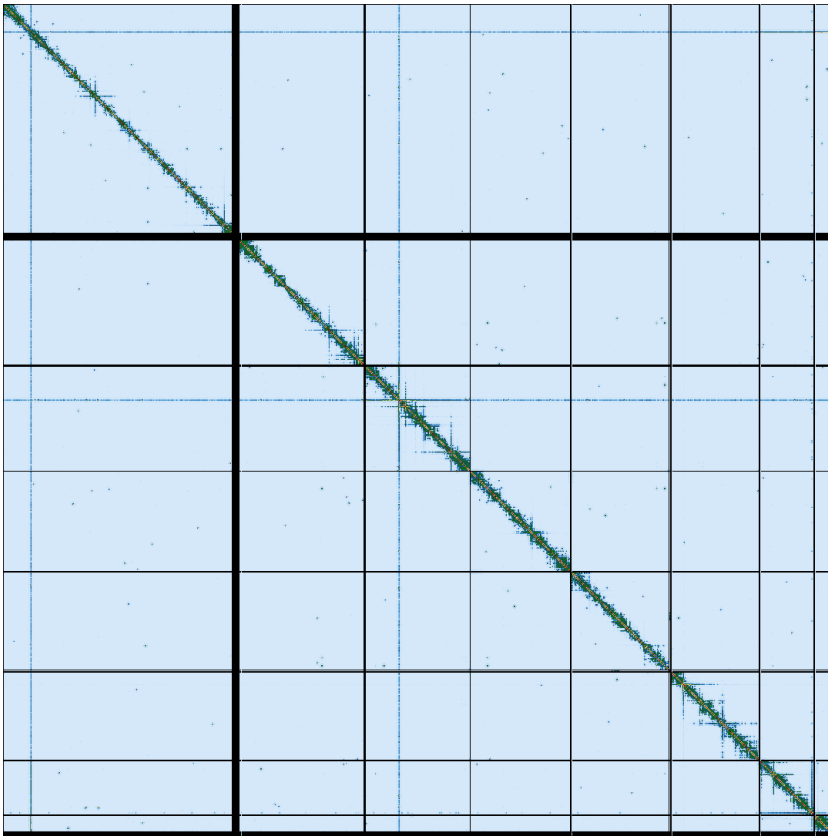
- . Interventions/Gb: None
- . Contamination notes: "FCS-GX identified 32 bacterial sequences that were removed from the assembly, the total amount of these contaminants added up 3,812,705 bp of sequence. "
- . Other observations: "This assembly was obtained with Hifiasm from HiFi data. After scaffolding with Yaho the genome was still highly fragmented and duplicated but we could make several joins and edits during curation with PretextMap, resulting in a substantial increase on scaffold N50 from (~2 Mb to 228Mb) and reducing duplication. Curation made 24 cuts in contigs, 15 breaks at gaps and 37 joins. 178 haplotig sequences, adding up a total of 121 Mb of sequence, were removed during curaton. Before curation, purge_dups had removed 7341 sequences, adding up a total of 1.5 Gb of sequence and reducing busco dups from 81% to 21% while keeping completeness at 97%. "

Quality metrics table

Metrics	Pre-curation Pri	Curated Pri
Total bp	1,949,241,826	1,824,778,328
GC %	35.2	35.21
Gaps/Gbp	981.92	1,246.18
Total gap bp	382,800	454,800
Scaffolds	722	181
Scaffold N50	9,020,784	228,685,931
Scaffold L50	61	3
Scaffold L90	268	7
Contigs	2,636	2,455
Contig N50	1,067,851	1,060,824
Contig L50	627	589
Contig L90	1,726	1,622
QV	59.5959	59.6
Kmer compl.	93.8064	92.0246
BUSCO sing.	74.8%	79.3%
BUSCO dupl.	21.1%	15.8%
BUSCO frag.	1.2%	1.2%
BUSCO miss.	2.8%	3.7%

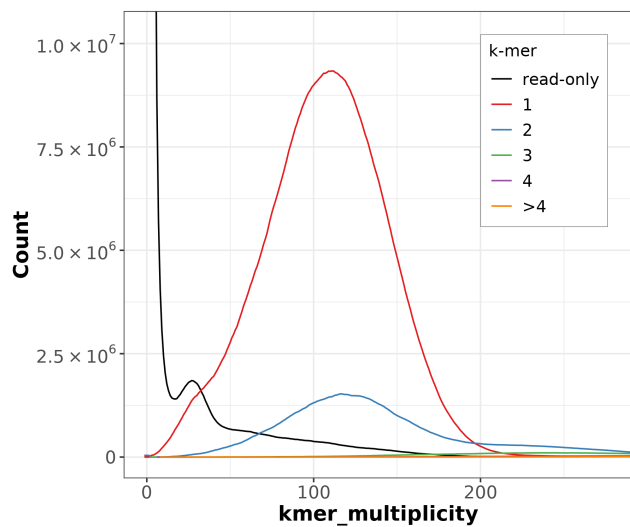
BUSCO: 6.0.0 (euk_genome_min, miniprot) / Lineage: arachnida_odb12 (genomes:34, BUSCOs:1123)

HiC contact map of curated assembly

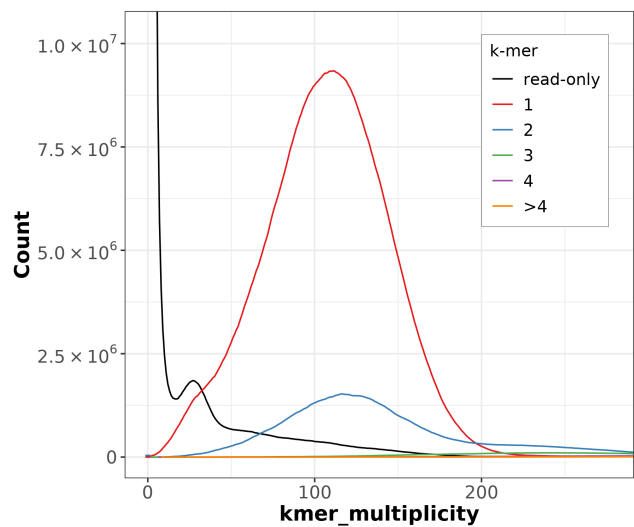


Pri [\[LINK\]](#)

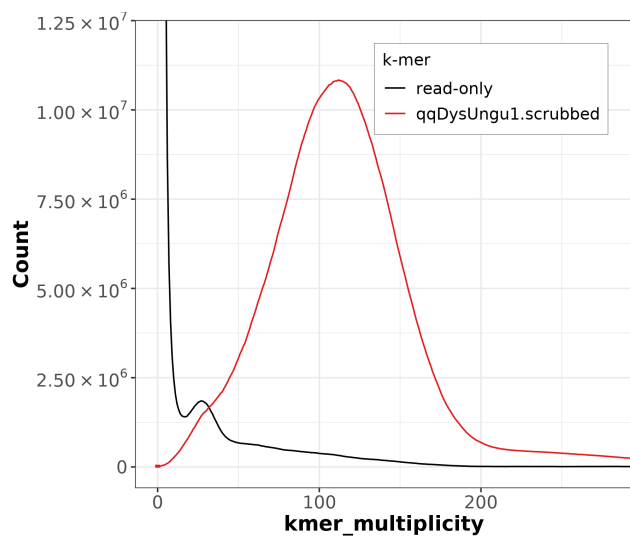
K-mer spectra of curated assembly



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

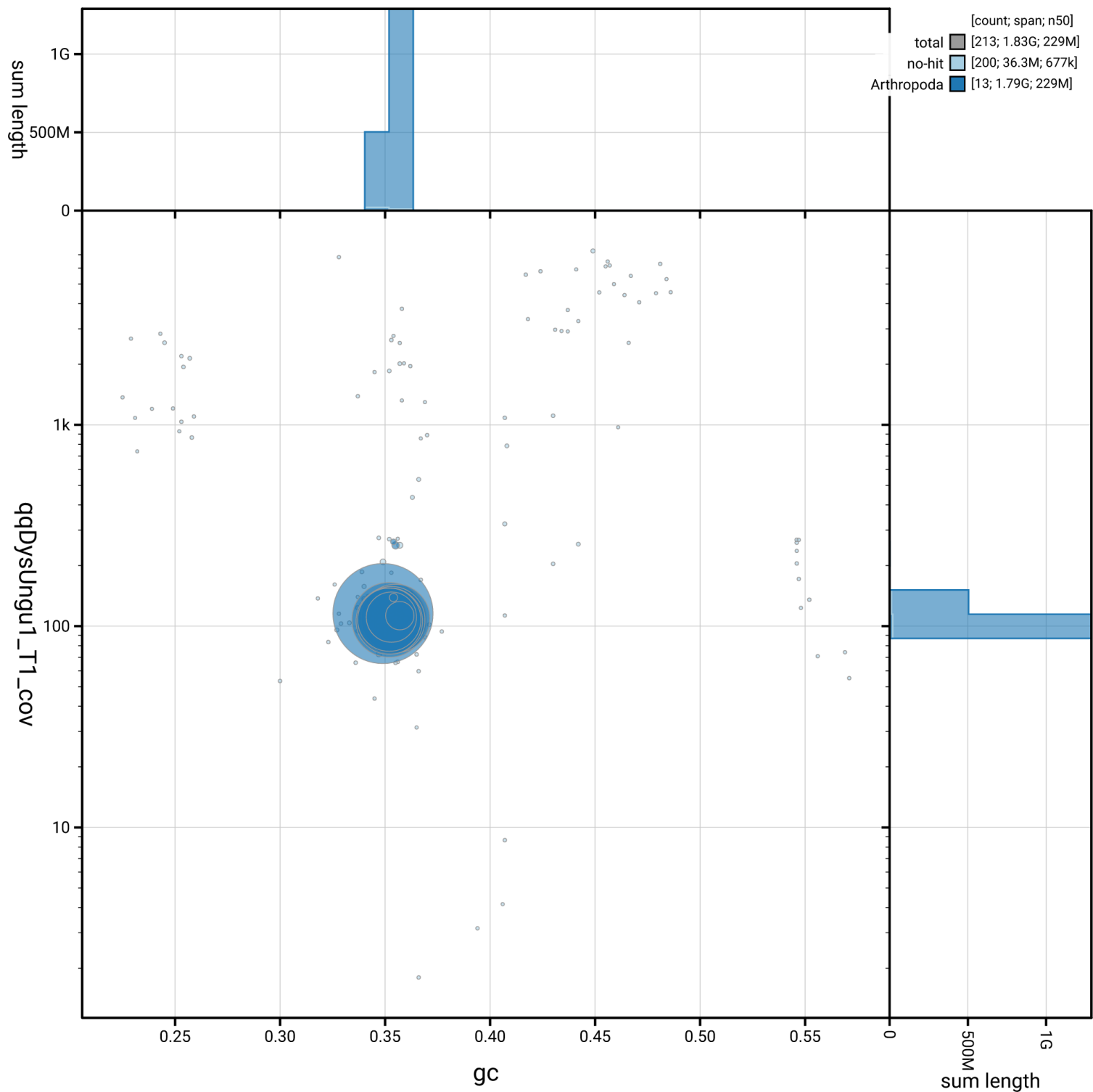


Distribution of k-mer counts per copy numbers found in asm



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

Post-curation contamination screening



Pri. 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 Hi-C
Coverage	95	115

Assembly pipeline

- **CLAWS**
 - |_ *ver*: v3.1
 - |_ *key param*: NA
- **Filtlong**
 - |_ *ver*: v0.2.1
 - |_ *key param*: --minlen 1000
 - |_ *key param*: --min_mean_q 80
- **Hifiasm**
 - |_ *ver*: 0.24.0
 - |_ *key param*: --telo-m TTAGG
- **Purge_dups**
 - |_ *ver*: 1.2.6
 - |_ *key param*: NA
- **Yahs**
 - |_ *ver*: 1.2a
 - |_ *key param*: -mq 10
 - |_ *key param*: --no-contig-ec

Curation pipeline

- **Blobtoolkit Nextflow pipeline**
 - |_ *ver*: 0.6
 - |_ *key param*: NA
- **PretextViewAI**
 - |_ *ver*: 1.0.5
 - |_ *key param*: NA
- **FCS-GX**
 - |_ *ver*: None
 - |_ *key param*: NA

Submitter: Jessica Gomez-Garrido
Affiliation: CNAG

Date and time: 2026-02-24 10:08:55 CET