

# ERGA Assembly Report

v24.10.15

Tags: ERGA-BGE

TxID	3229155
ToLID	<b>ilEumAren4</b>
Species	Eumannia arenbergeri
Class	Insecta
Order	Lepidoptera

Genome Traits	Expected	Observed
Haploid size (bp)	401,697,388	420,681,263
Haploid Number	30 (source: ancestor)	33
Ploidy	2 (source: ancestor)	2
Sample Sex	NA	NA

## EBP metrics summary and curation notes

Obtained EBP quality metric for hap1: 7.7.Q68

Obtained EBP quality metric for hap2: 7.7.Q68

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 hap1
- . Kmer completeness value is less than 90 for hap2

### Curator notes

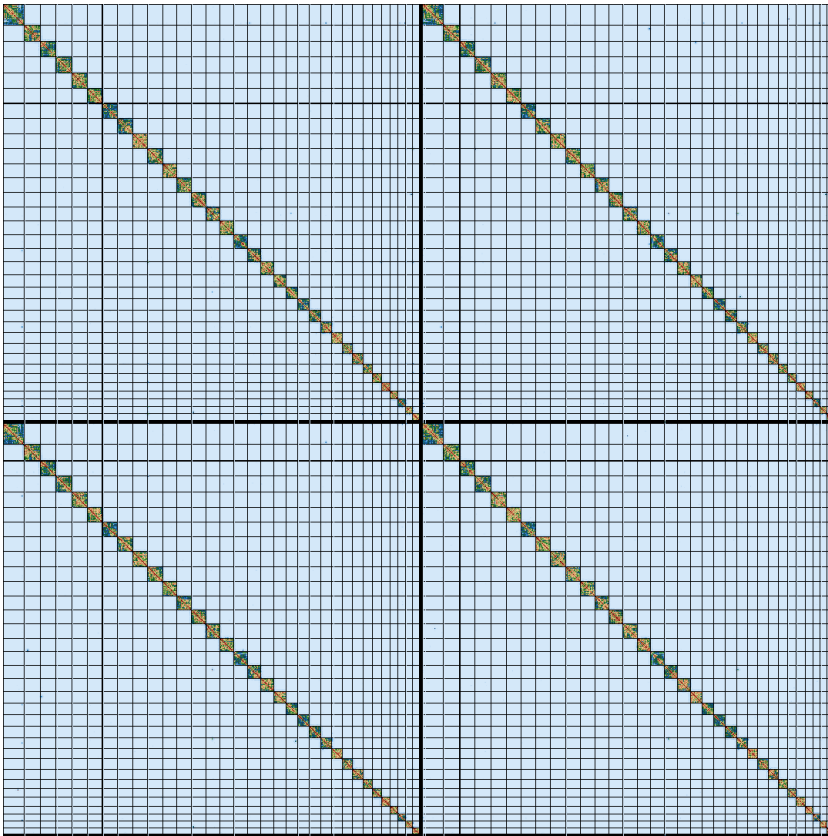
- . Interventions/Gb: None
- . Contamination notes: "One scaffold of 1,311,256 bp identified as Wolbachia was removed from hap1. Additionally, two scaffolds of 1,311,256 bp and 59,890 bp also identified as Wolbachia were removed from hap2."
- . Other observations: "This genome assembly was produced by assembling PacBio HiFi data with Hifiiasm and then scaffolding with Yahs. Given the good quality of both haplotypes, curation was made in diploid mode and we are planning to submit both haplotypes to the ENA. Curation made 1 cut in contigs, 3 breaks at gaps and 4 joins. Mitogenome was successfully assembled with FOAM."

# Quality metrics table

Metrics	Pre-curation hap1	Pre-curation hap2	Curated hap1	Curated hap2
Total bp	420,474,646	416,273,480	420,681,263	413,384,861
GC %	36.24	36.18	36.24	36.2
Gaps/Gbp	11.89	33.63	14.26	36.29
Total gap bp	1,000	2,800	1,200	3,000
Scaffolds	133	82	130	81
Scaffold N50	13,950,272	13,940,839	14,082,768	13,940,839
Scaffold L50	14	14	14	14
Scaffold L90	29	28	29	28
Contigs	138	96	136	96
Contig N50	13,775,434	12,863,022	14,082,768	12,247,657
Contig L50	14	15	14	15
Contig L90	30	32	29	33
QV	67.5081	69.8644	68.3577	68.6677
Kmer compl.	73.9321	73.7533	73.5634	73.3054
BUSCO sing.	98.0%	98.0%	98.0%	98.0%
BUSCO dupl.	0.7%	0.7%	0.8%	0.7%
BUSCO frag.	0.1%	0.2%	0.1%	0.1%
BUSCO miss.	1.1%	1.1%	1.1%	1.2%

BUSCO: 6.0.0 (euk\_genome\_min, miniprot) / Lineage: lepidoptera\_odb12 (genomes:79, BUSCOs:5760)

# HiC contact map of curated assembly

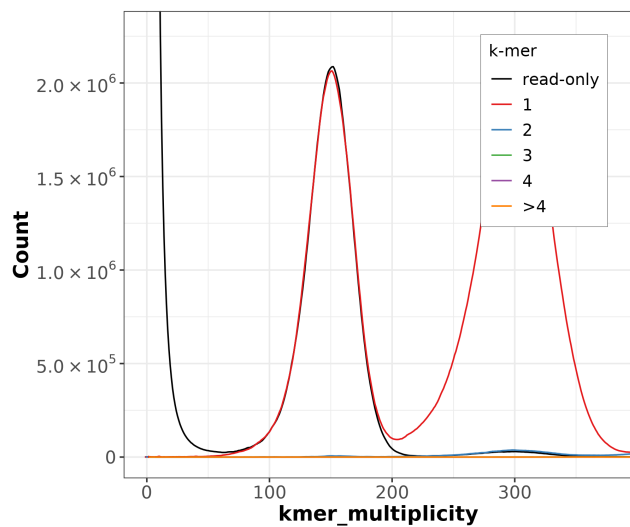


hap1 [\[LINK\]](#)

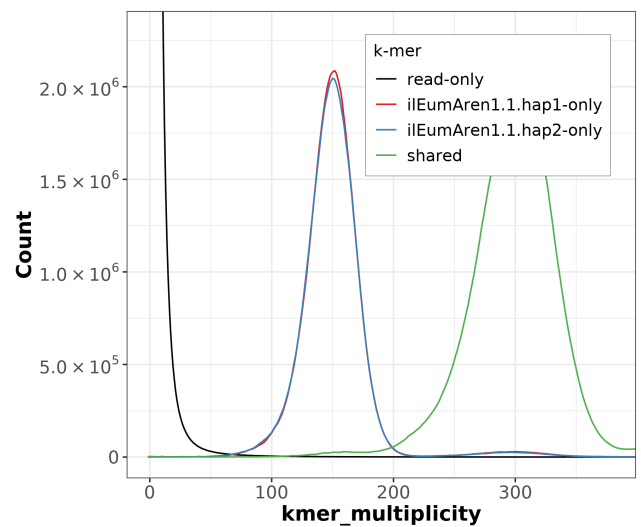
hap2 HiC PNG is missing!

hap2 File link is missing!

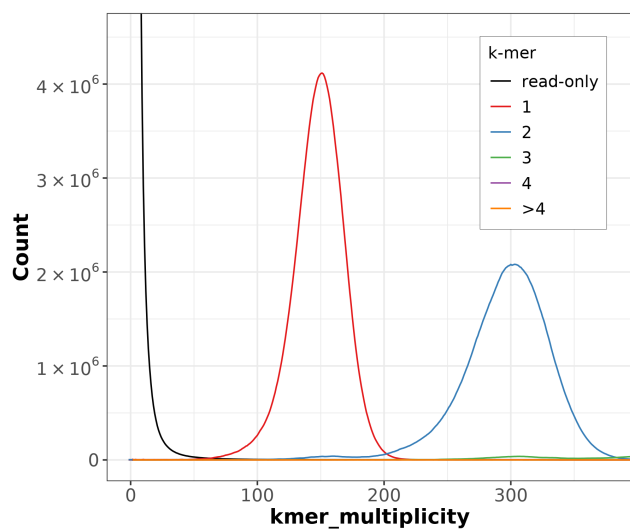
# K-mer spectra of curated assembly



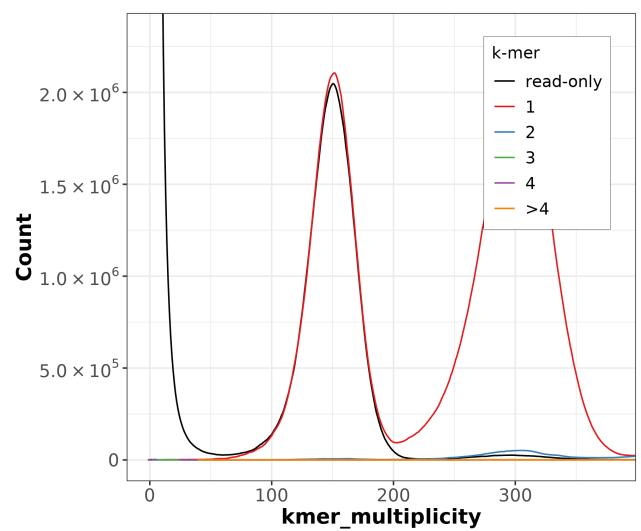
Distribution of k-mer counts per copy numbers found in **hap2** (hap1.)



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

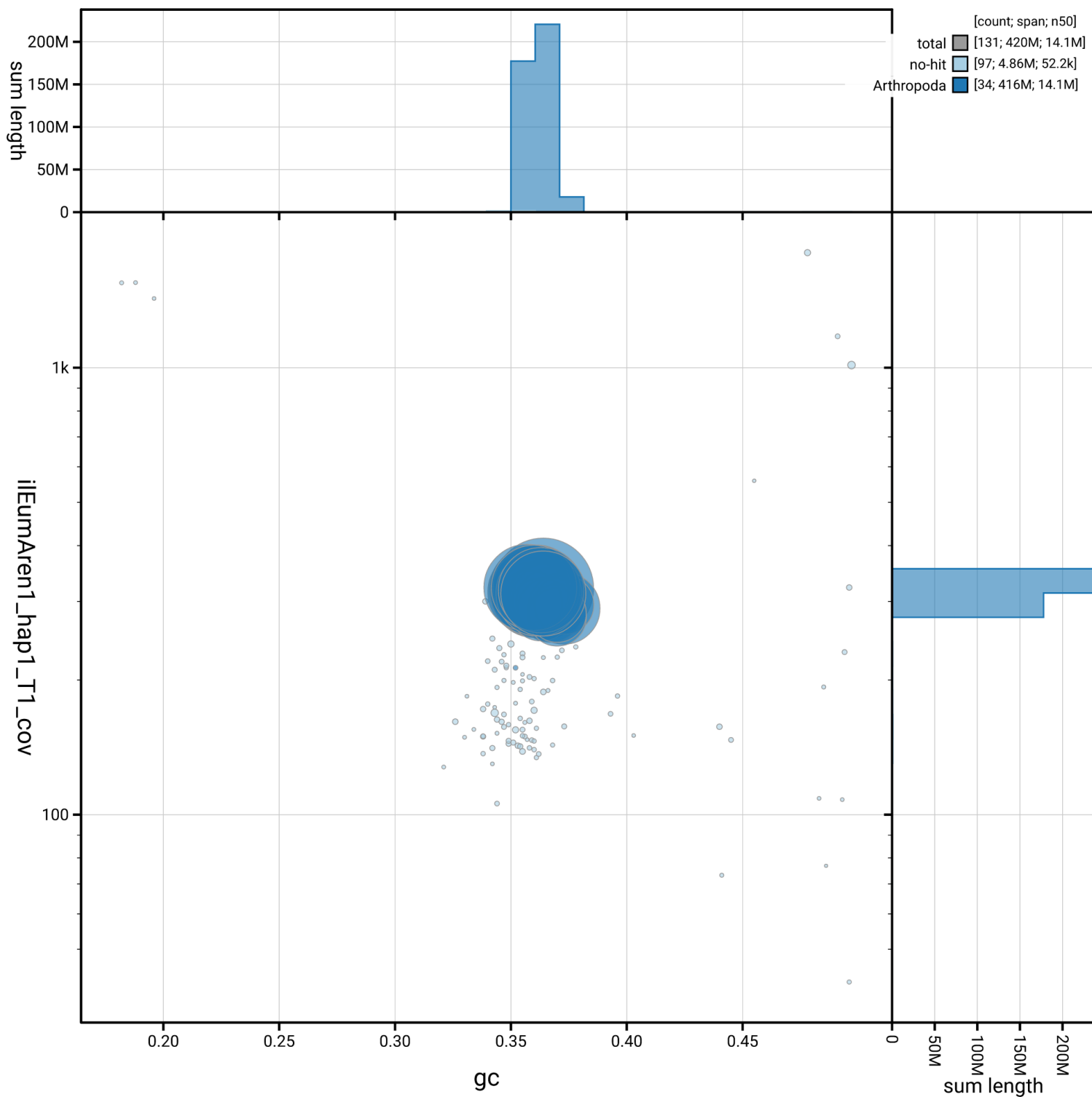


Distribution of k-mer counts per copy numbers found in **asm** (dipl.)

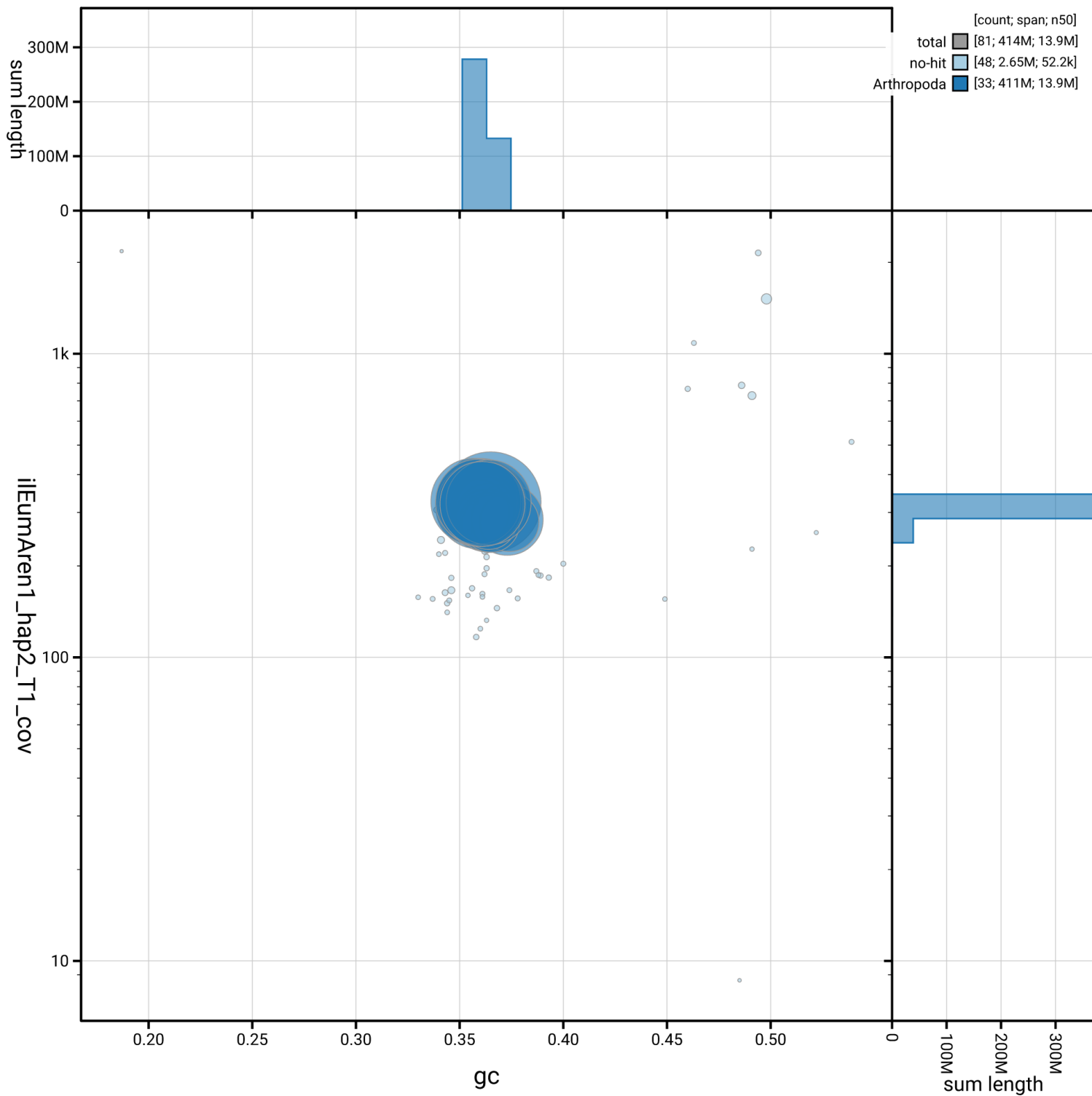


Distribution of k-mer counts per copy numbers found in **hap1** (hap1.)

# Post-curation contamination screening



**hap1.** 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.



**hap2.** 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	296	144

# Assembly pipeline

- **CLAWS**
  - |\_ *ver*: v3.0
  - |\_ *key param*: NA
- **Hifiasm**
  - |\_ *ver*: 0.24.0
  - |\_ *key param*: --telo-m TTAGG
- **Yahs**
  - |\_ *ver*: 1.2a
  - |\_ *key param*: -mq 10
  - |\_ *key param*: --no-contig-ec
- **FOAM**
  - |\_ *ver*: 0.5
  - |\_ *key param*: NA

# Curation pipeline

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

Submitter: Jessica Gomez-Garrido

Affiliation: CNAG

Date and time: 2026-01-30 11:42:48 CET