

# ERGA Assembly Report

v24.10.15

Tags: ERGA-BGE

TxID	2799294
ToLID	<b>qdGloPulc1</b>
Species	Glomeris pulchra
Class	Diplopoda
Order	Glomerida

Genome Traits	Expected	Observed
Haploid size (bp)	218,806,032	553,799,360
Haploid Number	12 (source: ancestor)	17
Ploidy	2 (source: ancestor)	4
Sample Sex	Unknown	unknown

## EBP metrics summary and curation notes

Obtained EBP quality metric for hap1: 6.7.Q49

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
- . Observed Ploidy is different from Expected
- . Kmer completeness value is less than 90 for hap1
- . Assembly length loss > 3% for hap1
- . Not 90% of assembly in chromosomes for hap1

### Curator notes

- . Interventions/Gb: 29
- . Contamination notes: "Repeated attempts to run BlobTools were unsuccessful, so we used an alternative tool called FCS-GX (Foreign Contamination Screen - Genome eXclusion), to identify contaminated scaffolds. In this species, hap1 contained 26 contaminated scaffolds with diverse bacterial sequences. Of these, four scaffolds exceeded 1 Mbp. All contaminated scaffolds from both haps were flagged as contaminants during a preliminary curation step using pretextview, and were subsequently removed from the analysis in the pretextview file shared for review."
- . Other observations: "We are sharing hap1 of the qdGloPulc hifi asm assembly. BUSCO results suggest a partial, and likely recent, whole genome duplication. Purge\_dups was unable to remove the duplicated regions, which we have subsequently labeled as haplotigs. We selected hap1 because hap2 had lower cN50 and sN50 values, although BUSCO scores and other assembly statistics were otherwise similar between the two haplotypes. In fact we are currently remapping the HiC data to a curated hap2 as well

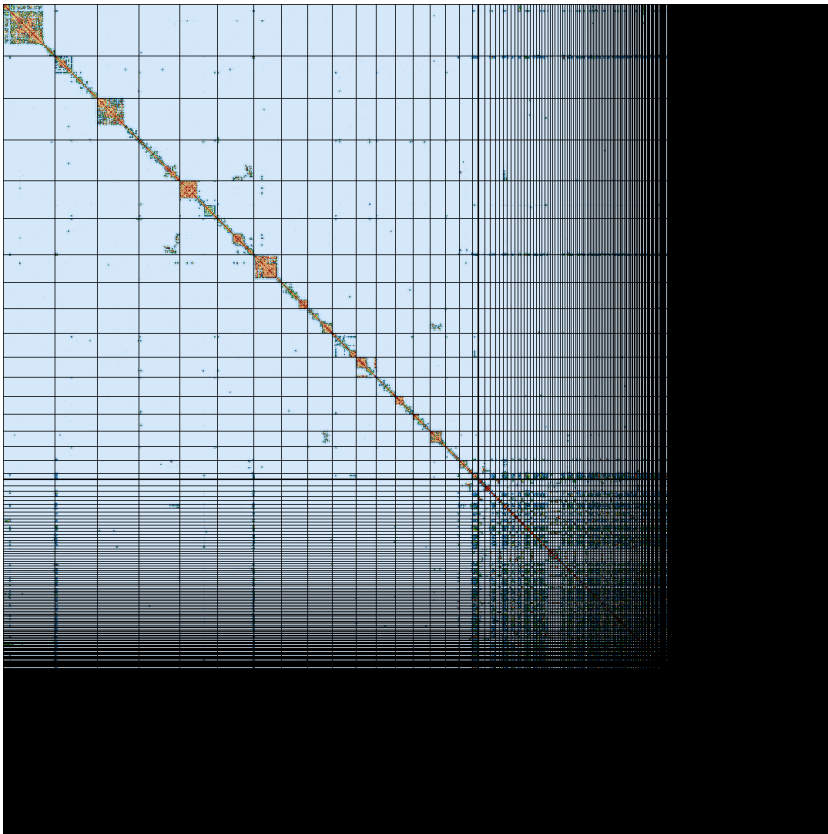
in order to compare both haps. The duplicated hap1 assembly was approximately 830 Mb in size. Following the first round of curation and removal of a large putative duplicated region, the assembly size was reduced to approximately 550 Mb. This curation led to a drop in cN50 of about half. We are currently considering curating hap2 as well to assess whether a more contiguous assembly can be achieved and are open to suggestions from the reviewers. Interestingly, GenomeScope predicts a haploid assembly size of 200 Mb assuming tetraploidy, and 440 Mb assuming the genome is diploid. We are currently running Smudgeplot (a rather slow process) to try to get a clearer idea of the ploidy of this species. Incidentally we were not able to successfully run hifiasm with its ploidy option set to four. After the first round of curation, we remapped the Hi-C data to the assembly and generated an updated Pretext map, which we are sharing for review. We have subsequently curated the map further. Please refer to save state three for the most up-to-date Pretext map."

# Quality metrics table

Metrics	Pre-curation hap1	Curated hap1
Total bp	831,031,655	553,799,360
GC %	41.11	41.8
Gaps/Gbp	577.6	733.12
Total gap bp	96,000	81,200
Scaffolds	1,004	956
Scaffold N50	31,684,699	11,699,835
Scaffold L50	9	14
Scaffold L90	228	317
Contigs	1,484	1,362
Contig N50	7,768,657	3,977,923
Contig L50	31	23
Contig L90	562	649
QV	49.8564	49.1682
Kmer compl.	96.5047	81.3073
BUSCO sing.	3.8%	95.1%
BUSCO dupl.	94.2%	1.4%
BUSCO frag.	0.4%	0.4%
BUSCO miss.	1.7%	3.2%

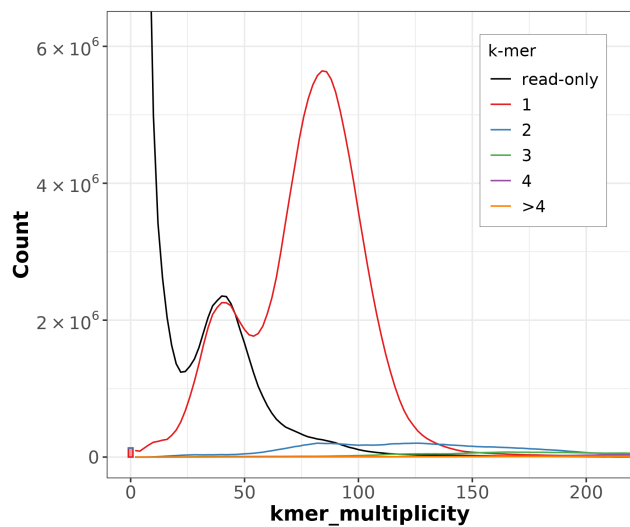
BUSCO: 6.0.0 (euk\_genome\_min, miniprot) / Lineage: arthropoda\_odb10 (genomes:90, BUSCOs:1013)

# HiC contact map of curated assembly

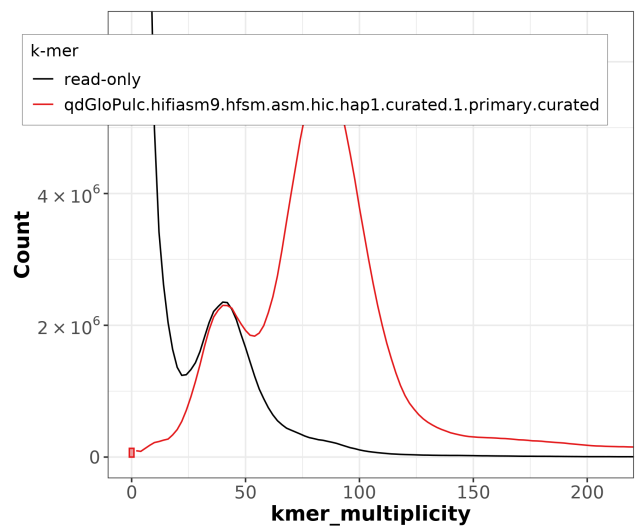


hap1 [\[LINK\]](#)

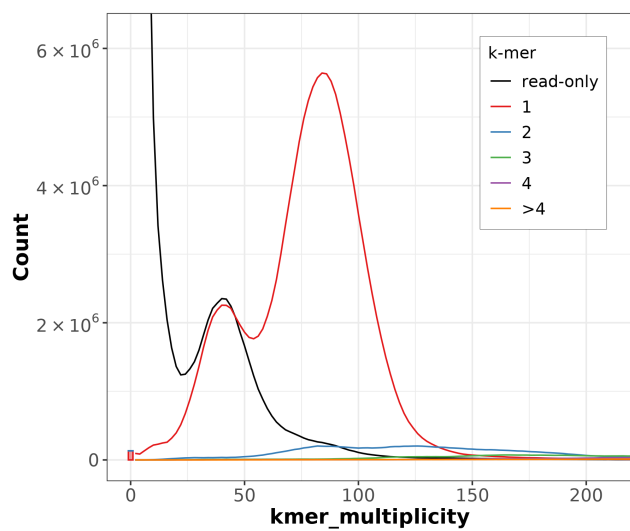
# K-mer spectra of curated assembly



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

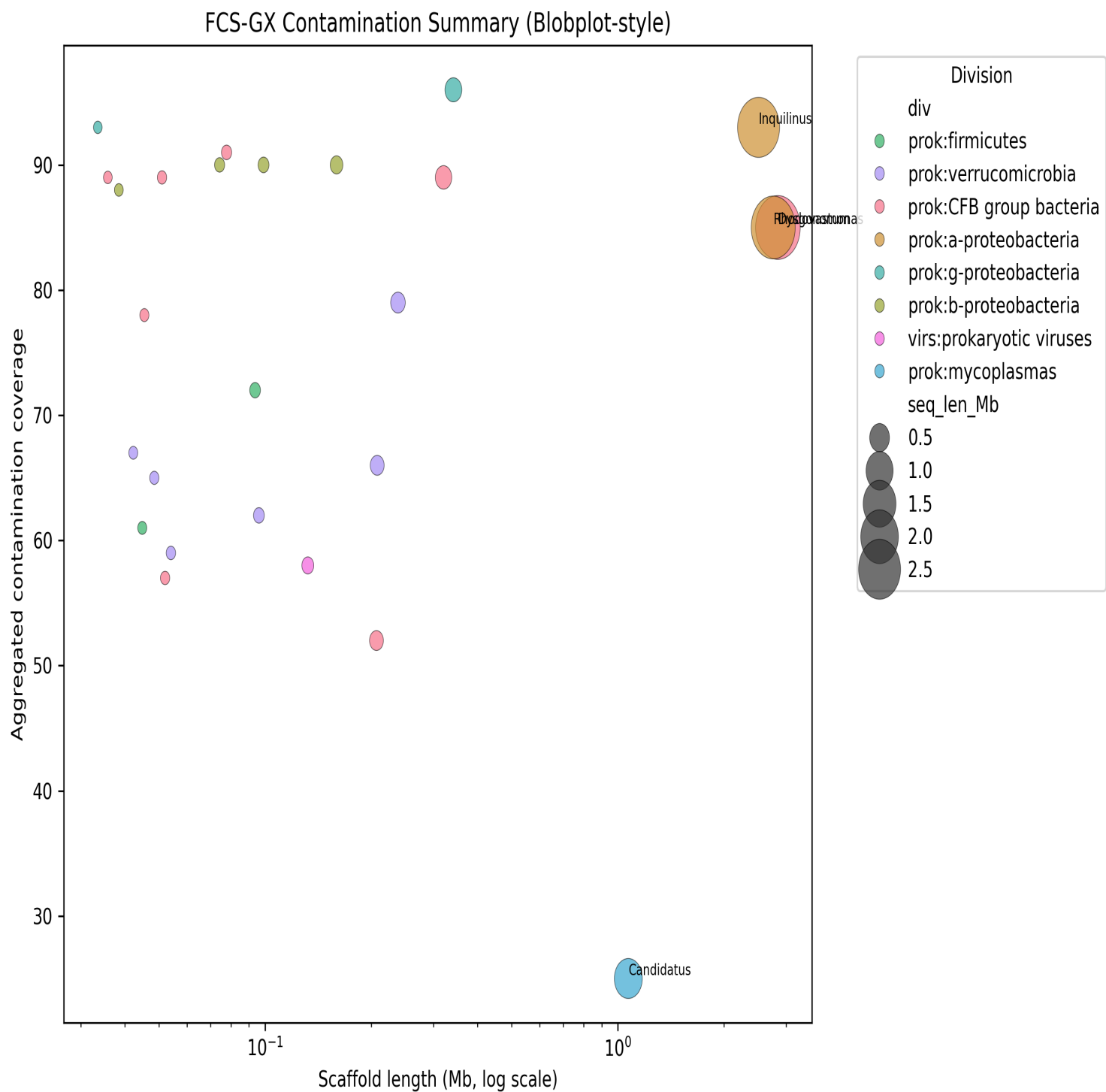


Distribution of k-mer counts coloured by their presence in reads/assemblies



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

# 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.

# Data profile

Data	ONT	Omni-C
Coverage	72	127

# Assembly pipeline

- **CLAWS**
  - |\_ ver: v3.1.0
  - |\_ key param: NA
- **Filtlong**
  - |\_ ver: v0.2.1
  - |\_ key param: -t target\_bases 40000000000
  - |\_ key param: --filtlong-min-mean-q 99
- **Hifiasm**
  - |\_ ver: 0.24.0
  - |\_ key param: --ont
- **Purge\_Dups**
  - |\_ ver: 1.2.6
  - |\_ key param: NA
- **Yahs**
  - |\_ ver: 1.2a
  - |\_ key param: NA

# Curation pipeline

- **FCS-GX (ForeignContamination Screen-Genome eXclusion) pipeline**
  - |\_ ver: v0.5.5-0
  - |\_ key param: NA
- **PretextViewAI**
  - |\_ ver: 1.0.5
  - |\_ key param: NA

Submitter: Francisco Camara Ferreira

Affiliation: CNAG

Date and time: 2026-02-24 12:07:52 CET