

# ERGA Assembly Report

v24.10.15

Tags: ERGA-BGE

TxID	297749
ToLID	<b>qqCypTeyr2</b>
Species	Cyphophthalmus teyrovskyi
Class	Arachnida
Order	Opiliones

Genome Traits	Expected	Observed
Haploid size (bp)	332,466,274	400,051,399
Haploid Number	8 (source: ancestor)	69
Ploidy	2 (source: ancestor)	2
Sample Sex	unknown	unknown

## EBP metrics summary and curation notes

Obtained EBP quality metric for collapsed: 6.6.Q50

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
- . Kmer completeness value is less than 90 for collapsed
- . BUSCO single copy value is less than 90% for collapsed
- . Not 90% of assembly in chromosomes for collapsed

### Curator notes

- . Interventions/Gb: None
- . Contamination notes: "FCS-GX and Blobtoolkit identified 90 contigs as contaminants and they were removed from the assembly."
- . Other observations: "Sequencing was done at UNIBA for two different specimens: qqCypTeyr2 (HiFi) and qqCypTeyr8 (Hi-C). Assembly was performed by the CNAG. We obtained a contig level assembly as detailed. We first assembled the HiFi reads with Hifiasm and then ran purge\_dups on it to collapse duplicated haplotigs and repeats. The Hi-C barely mapped to the purged assembly so scaffolding did not work. We did not assemble the mitogenome. We did not find evidences of contamination in the Hi-C or HiFi reads, and our QC of the HiC library shows <1% mapped reads and the few Cis Read Pairs mapping at distances shorter than 1kb"

# Quality metrics table

Metrics	Pre-curation collapsed	Curated collapsed
Total bp	404,302,988	400,051,399
GC %	37.06	37.03
Gaps/Gbp	0	0
Total gap bp	0	0
Scaffolds	612	522
Scaffold N50	1,407,397	1,421,856
Scaffold L50	70	69
Scaffold L90	281	271
Contigs	612	522
Contig N50	1,407,397	1,421,856
Contig L50	70	69
Contig L90	281	271
QV	49.5001	50.1812
Kmer compl.	86.5403	86.5213
BUSCO sing.	75.3%	75.5%
BUSCO dupl.	3.8%	3.6%
BUSCO frag.	5.2%	5.1%
BUSCO miss.	15.7%	15.9%

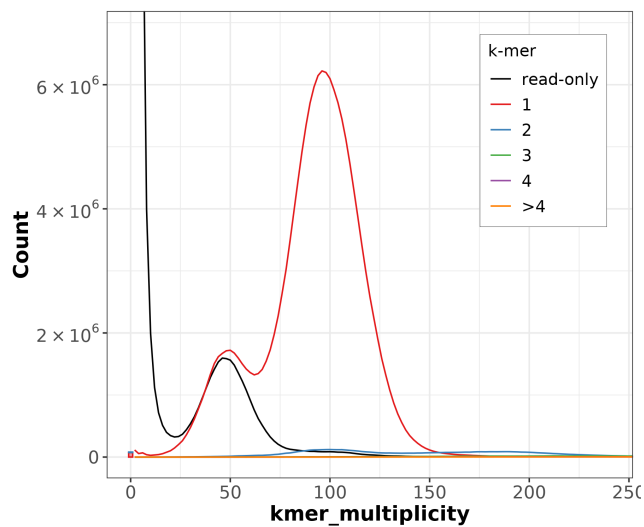
BUSCO: 6.0.0 (euk\_genome\_min, miniprot) / Lineage: arachnida\_odb12 (genomes:34, BUSCOs:1123)

# HiC contact map of curated assembly

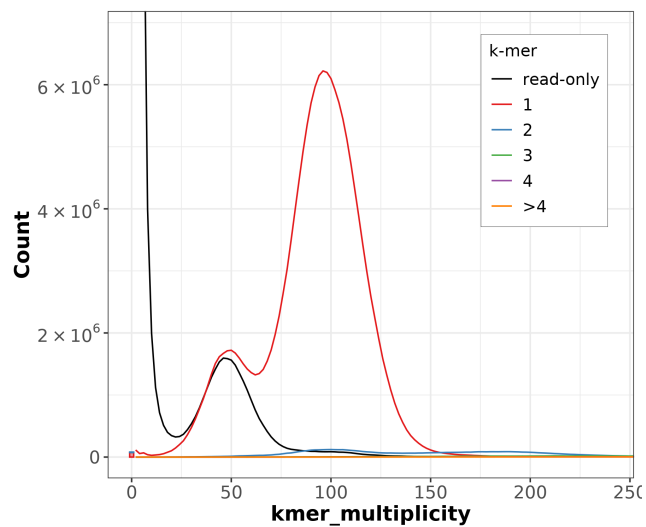
**collapsed** HiC PNG is missing!

**collapsed** File link is missing!

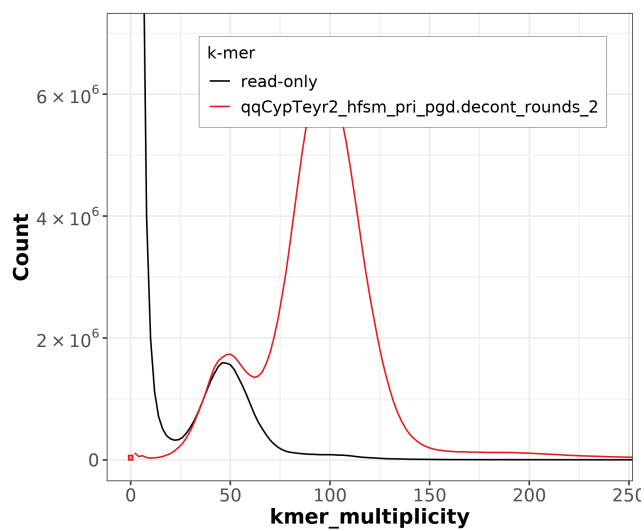
# K-mer spectra of curated assembly



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



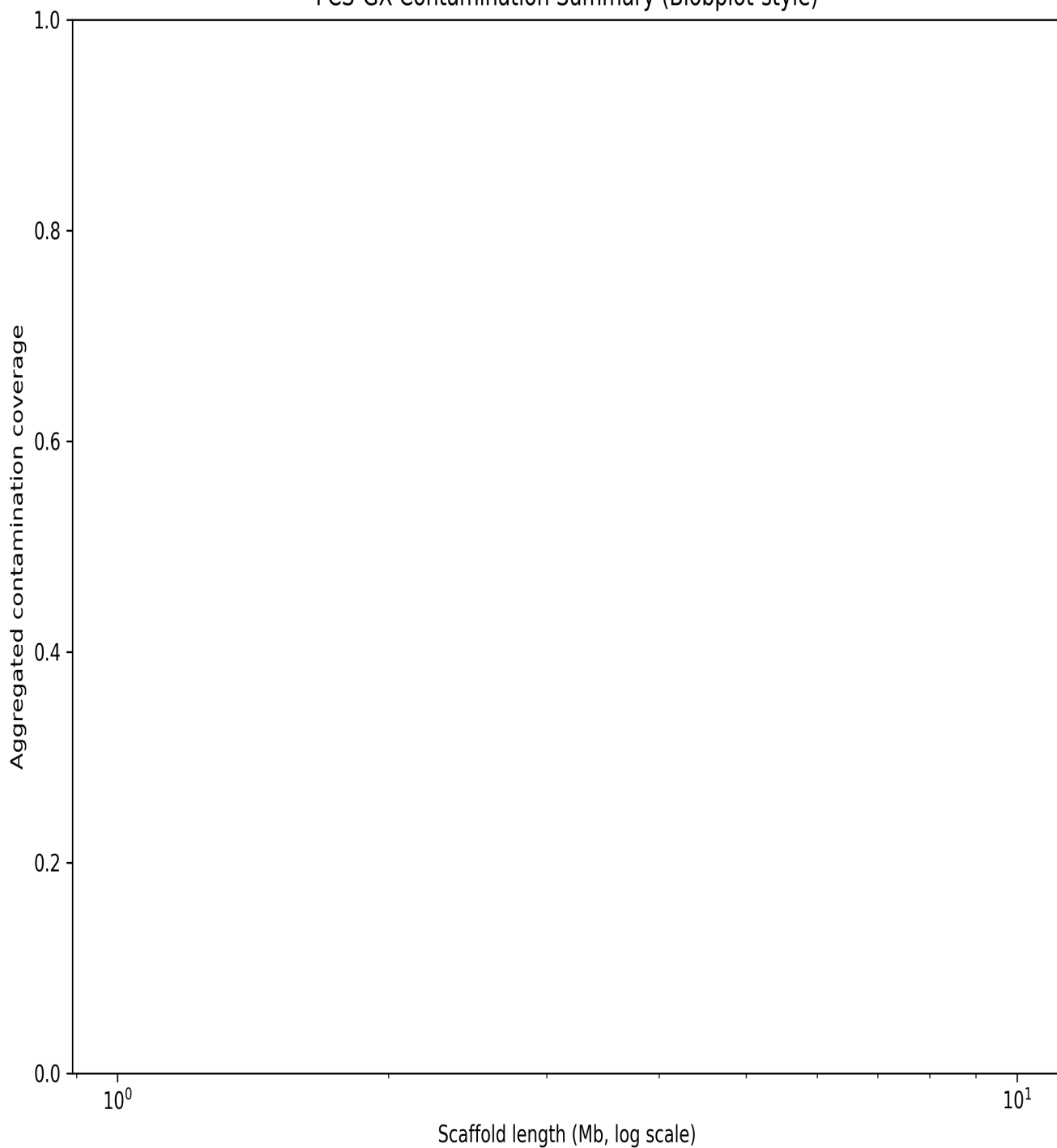
Distribution of k-mer counts per copy numbers found in asm



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

# Post-curation contamination screening

FCS-GX Contamination Summary (Blobplot-style)



**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	Pacbio HiFi	Hi-C
Coverage	85x	166x

## Assembly pipeline

- **CLAWS pipeline**
  - |\_ *ver*: 3.1
  - |\_ *key param*: NA
- **Filtlong**
  - |\_ *ver*: 0.2.1
  - |\_ *key param*: --min\_length 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
- **Merqury**
  - |\_ *ver*: 1.3
  - |\_ *key param*: NA
- **BUSCO**
  - |\_ *ver*: 6.0.0
  - |\_ *key param*: NA
- **tidk**
  - |\_ *ver*: 0.2.65
  - |\_ *key param*: NA
- **chromap**
  - |\_ *ver*: 0.3.2
  - |\_ *key param*: NA

## Curation pipeline

- **Blobtoolkit DToL Nextflow pipeline**
  - |\_ *ver*: 0.6.0
  - |\_ *key param*: NA
- **FCS-GX**
  - |\_ *ver*: 0.5.5
  - |\_ *key param*: NA

Submitter: Fernando Cruz  
Affiliation: CNAG

Date and time: 2026-02-20 12:48:20 CET