ERGA Assembly Report

v24.10.15

Tags: ERGA-BGE

TxID	2034322
ToLID	**iqTroOvul1**
Species	Troglophilus ovuliformis
Class	Insecta
Order	Orthoptera

Genome Traits	Expected	Observed
Haploid size (bp)	6,486,622,086	7,752,424,174
Haploid Number	28 (source: ancestor)	9
Ploidy	2 (source: ancestor)	2
Sample Sex	XY	XY

EBP metrics summary and curation notes

Obtained EBP quality metric for pri: 8.9.Q71

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 pri

Curator notes

. Interventions/Gb: 4.5
. Contamination notes: "Repeated attempts to run BlobTools were unsuccessful. While investigating the cause, we utilized an alternative tool, FCS-GX (Foreign Contamination Screen – Genome eXclusion), which identified several of the smallest chromosomes (all under 3.6 kilobases) as contaminated with human sequence (scaffolds 337, 343, 346, 347, 348, and 351). These scaffolds will be excluded from the final assembly."
. Other observations: "We selected the primary assembly for this species due to its superior contiguity and fewer gaps compared to either haplotype. The sequenced sample was from an XY male. Our analysis identified a putative pseudoautosomal region (PAR), which we positioned at the beginning of chromosome X. We also detected several haplotigs. In the case of the repeat-rich Scaffold_9 (painted), we tentatively flagged certain sequences as unlocalized sequences, even though these could be haplotigs pending reviewer feedback. The Pretext map we are sharing has not yet been remapped, but all putative chromosomes have been painted and identified haplotigs have been tagged. We have also included the mq0 Pretext map for your reference. The statistics (busco,merqury,gfstats) were obtained on the provisionally curated assembly generated from the mq10 pretext which we shared."

Quality metrics table

| Metrics	Pre-curation pri	Curated pri
Total bp	7,800,738,619	7,752,424,174
GC %	40.9	40.9
Gaps/Gbp	4.23	6.19
Total gap bp	6,600	9,600
Scaffolds	353	341
Scaffold N50	1,739,946,135	1,720,636,234
Scaffold L50	3	3
Scaffold L90	11	9
Contigs	386	389
Contig N50	257,295,000	257,295,000
Contig L50	8	8
Contig L90	29	30
QV	71.4565	71.4328
Kmer compl.	89.9503	89.8345
BUSCO sing.	95.4%	96.2%
BUSCO dupl.	3.8%	3.0%
BUSCO frag.	0.5%	0.5%
BUSCO miss.	0.3%	0.3%

BUSCO: 5.5.0 (euk_genome_met, metaeuk) / Lineage: arthropoda_odb10 (genomes:90, BUSCOs:1013)

HiC contact map of curated assembly
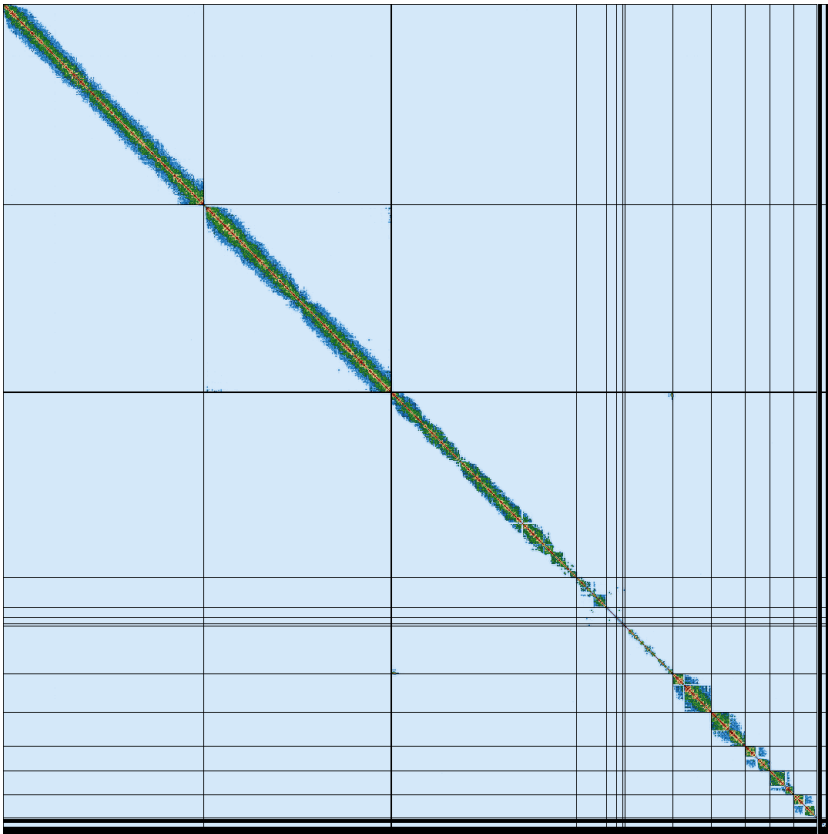
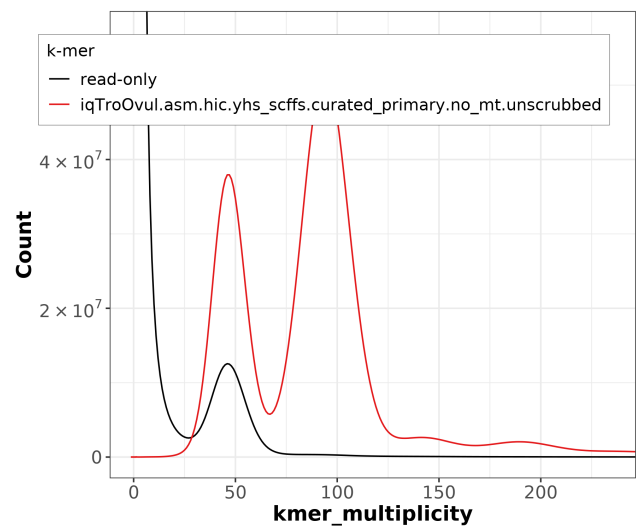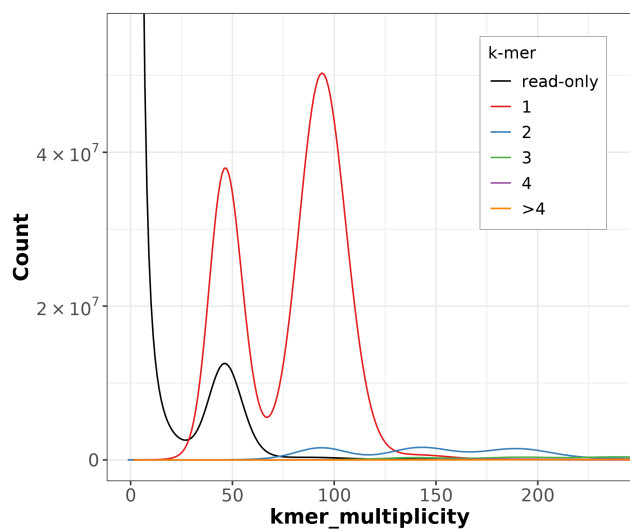**pri** [LINK]

K-mer spectra of curated assembly
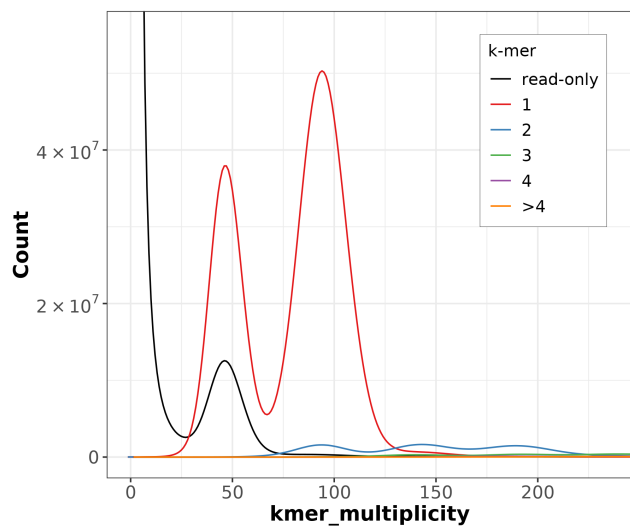
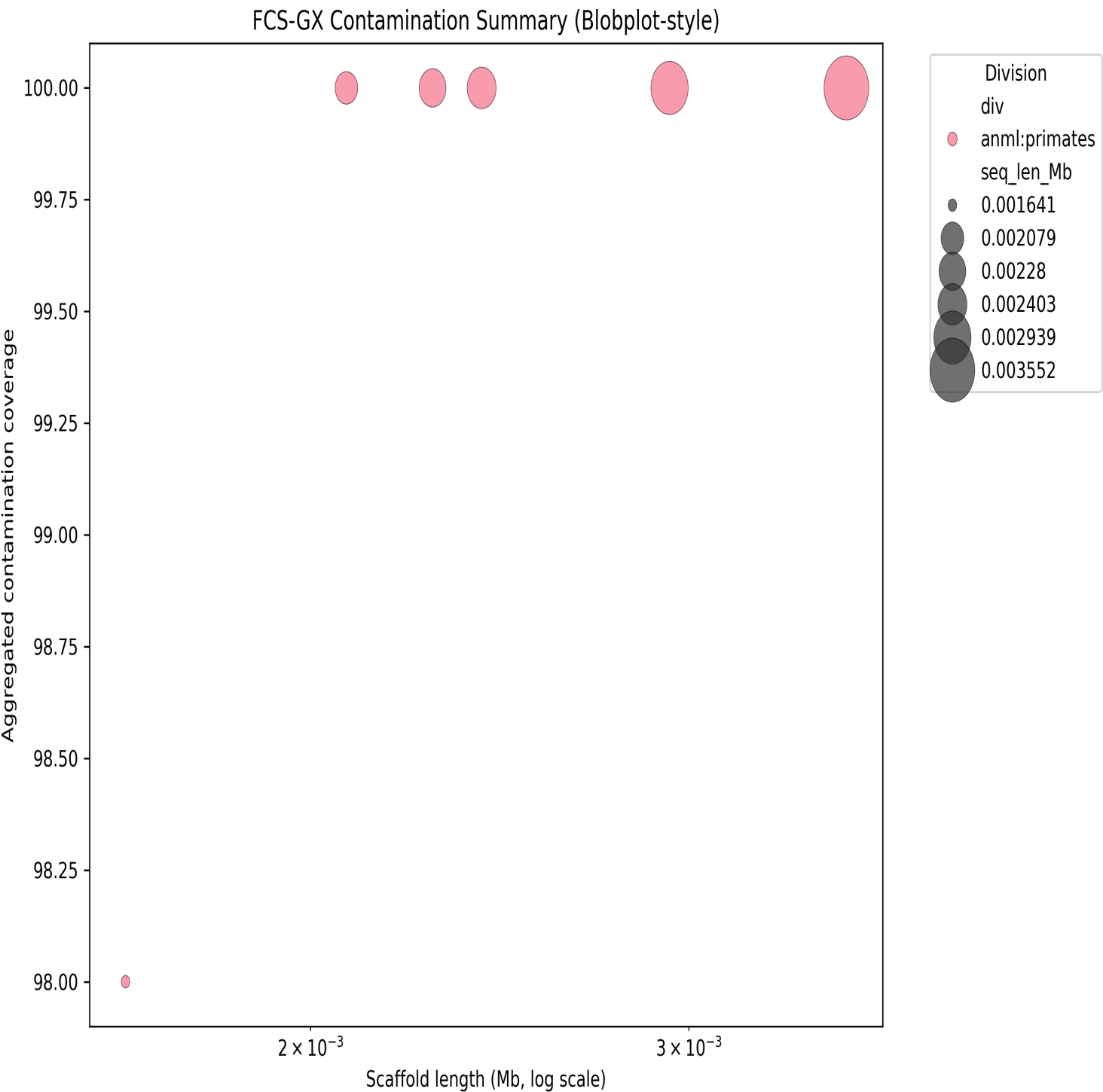Distribution of k-mer counts per copy
numbers found in asm

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

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

Post-curation contamination screening

FCS-GX Contamination Summary (Blobplot-style)

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	ONT	OmniC
Coverage	47x	NA

Assembly pipeline

- **hifiasm**
    |_ *ver:* 0.24.0-r702
    |_ *key param:* NA
- **YaHS**
    |_ *ver:* 1.2a
    |_ *key param:* NA
- **CLAWS**
    |_ *ver:* 2.3
    |_ *key param:* NA

Curation pipeline

- **PretextViewAI**
    |_ *ver:* 1.0.4
    |_ *key param:* NA
- **GRIT_Rapid**
    |_ *ver:* 2.0
    |_ *key param:* NA

Submitter: Francisco Camara
Affiliation: CNAG Barcelona

Date and time: 2025-10-31 14:34:17 CET