

ERGA Assembly Report

v24.10.15

Tags: ERGA-BGE

TxID	2967106
ToLID	ilMusMuscl
Species	Muscula muscula
Class	Insecta
Order	Lepidoptera

Genome Traits	Expected	Observed
Haploid size (bp)	751,486,236	815,388,997
Haploid Number	31 (source: ancestor)	30
Ploidy	2 (source: ancestor)	2
Sample Sex	ZW	ZW

EBP metrics summary and curation notes

Obtained EBP quality metric for hap1: 7.7.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 hap1

Curator notes

. Interventions/Gb: 4

. Contamination notes: "Repeated attempts to run BlobTools were unsuccessful. We instead utilized an alternative tool, FCS-GX (Foreign Contamination Screen - Genome eXclusion), which identified three scaffolds as contaminated with bacterial sequence. One scaffold appears to correspond to the full genome of the endosymbiont Serratia marcescens. These contaminated scaffolds were removed prior to diploid curation"

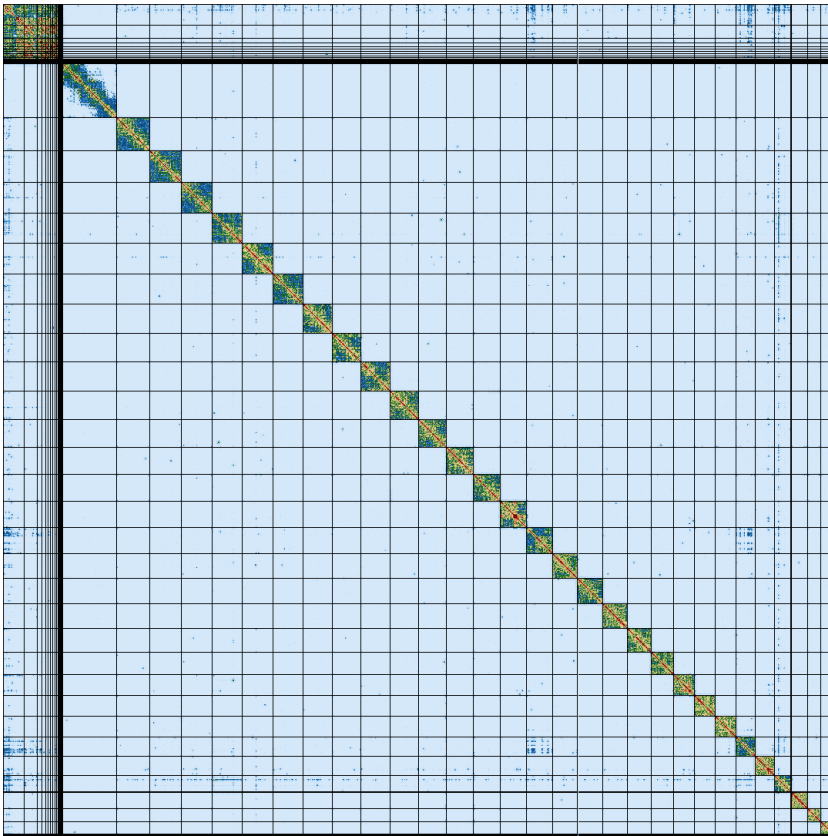
. Other observations: "This assembly was generated with Hifiiasm using ONT data. Given the high quality of both haplotypes, curation was performed in diploid mode. In this report, only hap1 is shared. This haplotype is more complete than hap2, as it includes the full Z chromosome (originally in hap1) and the fragmented W chromosome, which was present in both hap1 and hap2 prior to diploid curation. After the reviewing process we agreed that SUPER_W_unloc_12 and SUPER_W_unloc_13 did not appear to be Unlocs of SUPER_W. They were moved to the Chaff and renamed scaffold_3 and scaffold_4. We also found two other segments of SUPER_15 and SUPER_21 respectively that did not seem to match specifically to these SUPERS and we also moved these to the chaff while renaming them scaffold_1 and scaffold_2"

Quality metrics table

Metrics	Pre-curation hap1	Curated hap1
Total bp	786,754,004	815,388,997
GC %	39.36	39.27
Gaps/Gbp	0	3.68
Total gap bp	0	600
Scaffolds	75	63
Scaffold N50	26,643,511	26,102,625
Scaffold L50	13	14
Scaffold L90	27	28
Contigs	75	66
Contig N50	26,643,511	26,102,625
Contig L50	13	14
Contig L90	27	29
QV	71.9384	71.7977
Kmer compl.	86.5872	86.0378
BUSCO sing.	98.7%	98.8%
BUSCO dupl.	0.9%	0.8%
BUSCO frag.	0.2%	0.2%
BUSCO miss.	0.2%	0.2%

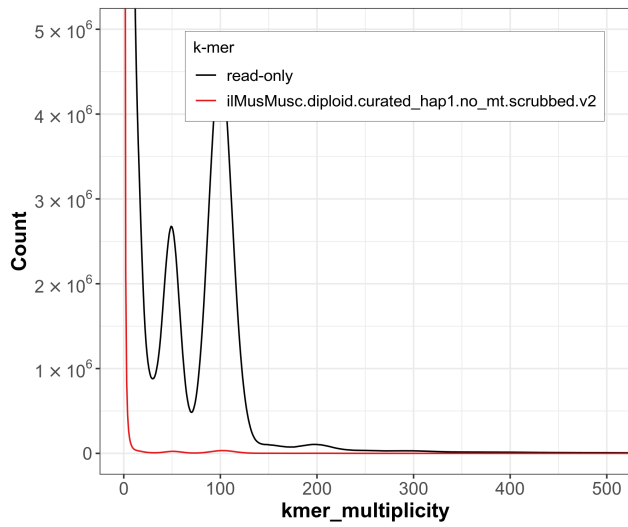
BUSCO: 6.0.0 (euk_genome_min, miniprot) / Lineage: lepidoptera_odb10 (genomes:16, BUSCOs:5286)

HiC contact map of curated assembly

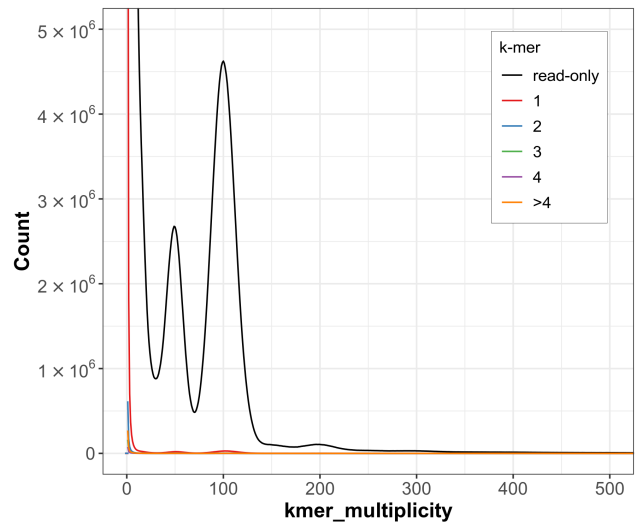


hap1 [\[LINK\]](#)

K-mer spectra of curated assembly

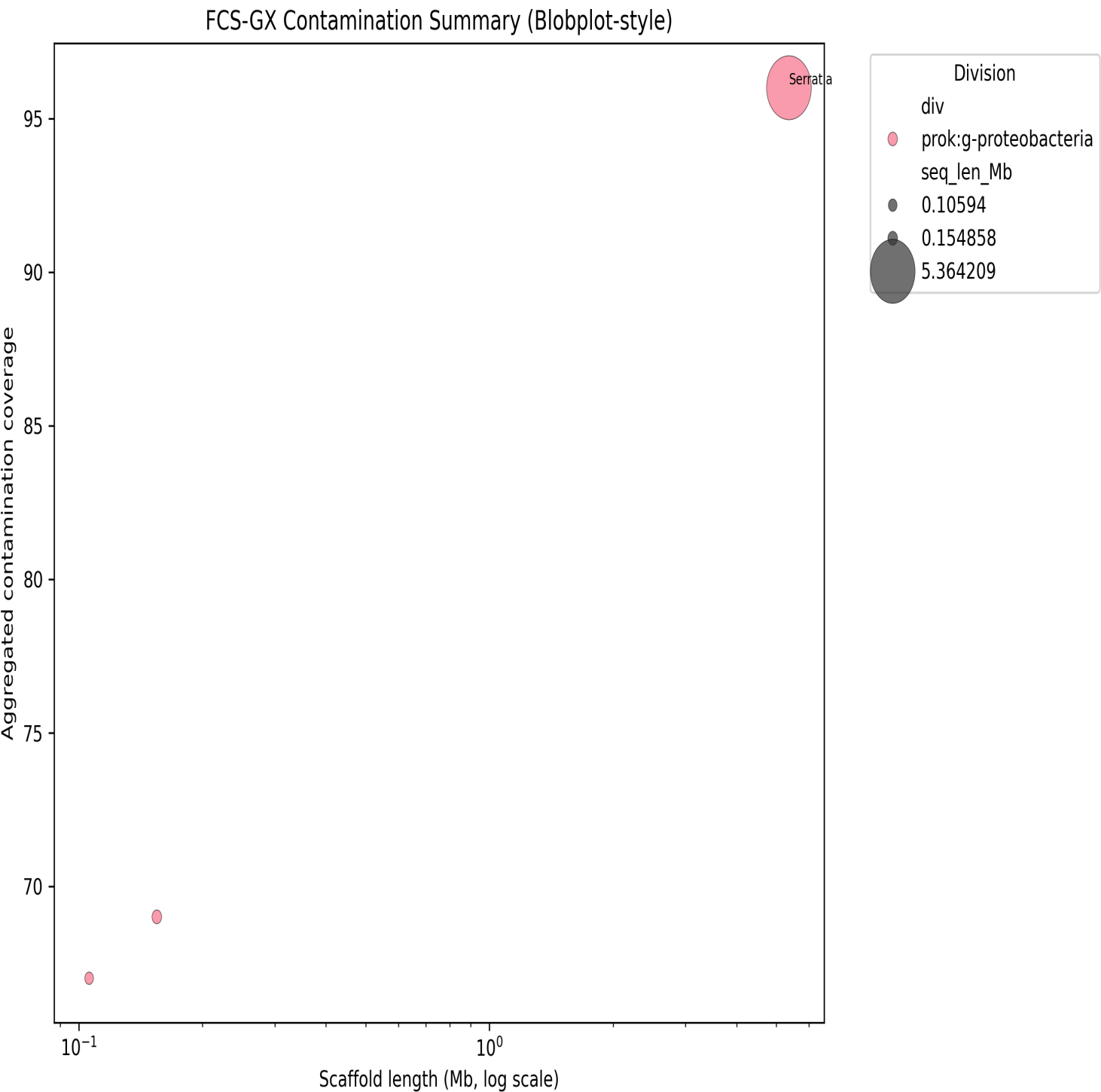


Distribution of k-mer counts coloured by their presence in reads/assemblies



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	Illumina	Omni-C
Coverage	100	100	268

Assembly pipeline

- **CLAWS**
 - |_ *ver*: v2.3.0
 - |_ *key param*: NA
- **Filtlong**
 - |_ *ver*: v0.2.1
 - |_ *key param*: -t target_bases 80000000000
 - |_ *key param*: --filtlong-min-mean-q 97
- **Hifiasm**
 - |_ *ver*: 0.24.0
 - |_ *key param*: --ont
- **Yahs**
 - |_ *ver*: 1.2a
 - |_ *key param*: -mq 0
 - |_ *key param*: -no-contig-ec

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: 2026-02-03 16:09:42 CET