ERGA Assembly Report
v24.10.15

Tags: ERGA-BGE

TxID	1960709
ToLID	**wpGesJame1**
Species	Gesiella jameensis
Class	Polychaeta
Order	Phyllodocida

Genome Traits	Expected	Observed
Haploid size (bp)	8,446,252,482	8,852,556,600
Haploid Number	10 (source: ancestor)	9
Ploidy	2 (source: ancestor)	2
Sample Sex	unknown	unknown

EBP metrics summary and curation notes

Obtained EBP quality metric for pri: 6.9.Q66

The following metrics were automatically flagged as below EBP recommended standards or different from expected:

. Observed Haploid Number is different from Expected

. BUSCO single copy value is less than 90% for pri
. BUSCO duplicated value is more than 5% for pri

Curator notes

. Interventions/Gb: 43
. Contamination notes: "FCS-GX identified 59 contaminants (proteobacteria and mycosplasma) which we removed (the larger ones are tagged in pretextview savestate and the rest will be removed with a script)."
. Other observations: "Curation made 25 cuts in contigs, 115 breaks at gaps and 238 joins. Many seemingly collapsed repeats at scaffold gaps. One side with high cov and other with average cov. If both had lower coverage, I removed one and tagged as haptig, but there weren't so many of these. It takes time to run BUSCO and regenerate the maps on the curated assembly, so for the moment I am sharing the final savestate of the curation and showing the BUSCOs prior to curation. "

Quality metrics table

| Metrics	Pre-curation pri	Curated pri
Total bp	8,893,284,574	8,852,556,600
GC %	41.77	41.77
Gaps/Gbp	207.01	221.63
Total gap bp	368,200	392,400
Scaffolds	2,025	1,899
Scaffold N50	1,062,753,756	1,095,131,142
Scaffold L50	4	4
Scaffold L90	8	8
Contigs	3,866	3,861
Contig N50	7,120,079	7,120,079
Contig L50	383	382
Contig L90	1,258	1,255
QV	66.3363	66.3225
Kmer compl.	93.9213	93.7761
BUSCO sing.	83.6%	83.6%
BUSCO dupl.	5.5%	5.5%
BUSCO frag.	7.1%	7.1%
BUSCO miss.	3.8%	3.8%

BUSCO: 5.5.0 (euk_genome_met, metaeuk) / Lineage: metazoa_odb10 (genomes:65, BUSCOs:954)
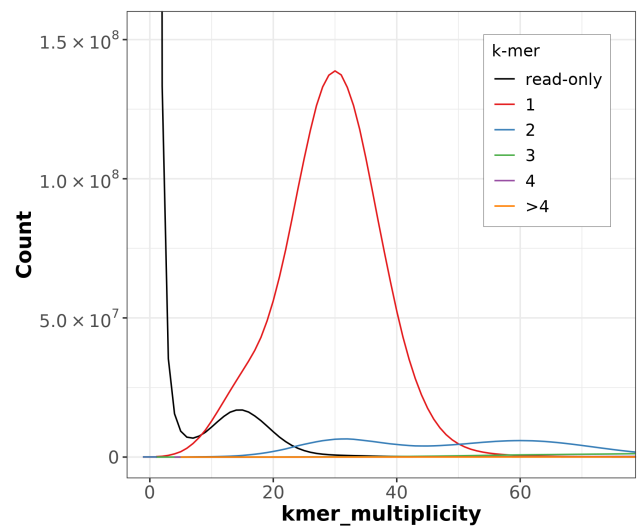
HiC contact map of curated assembly
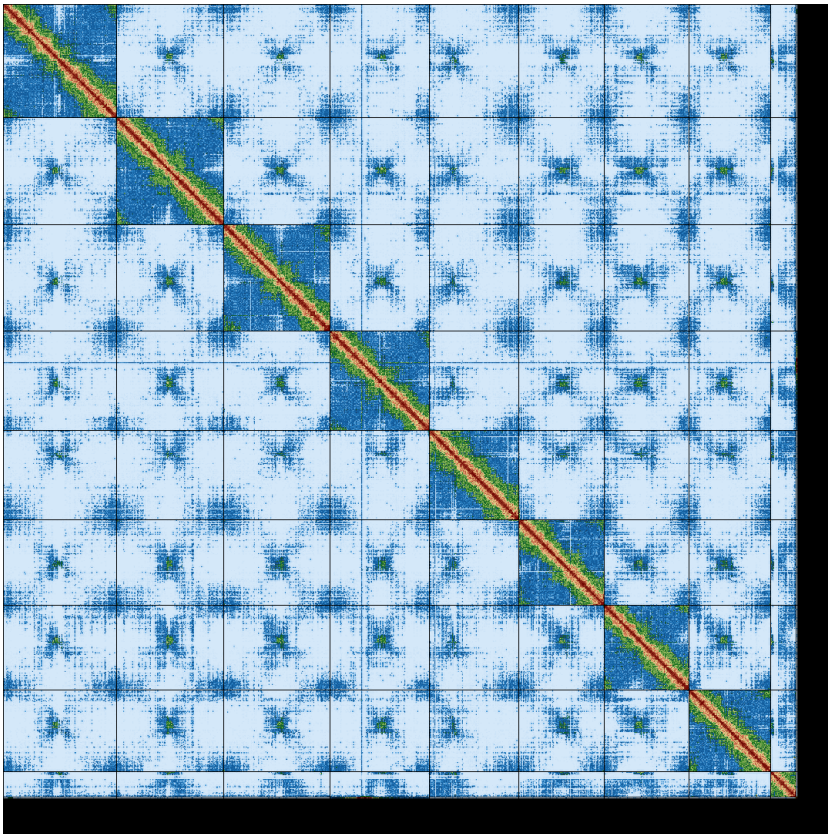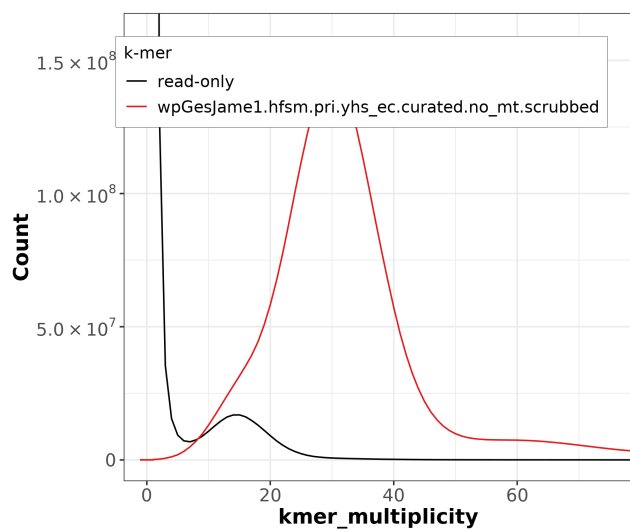

pri [LINK]

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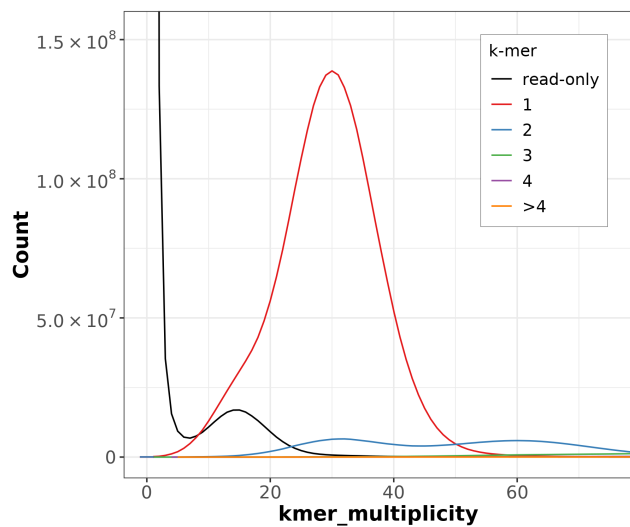
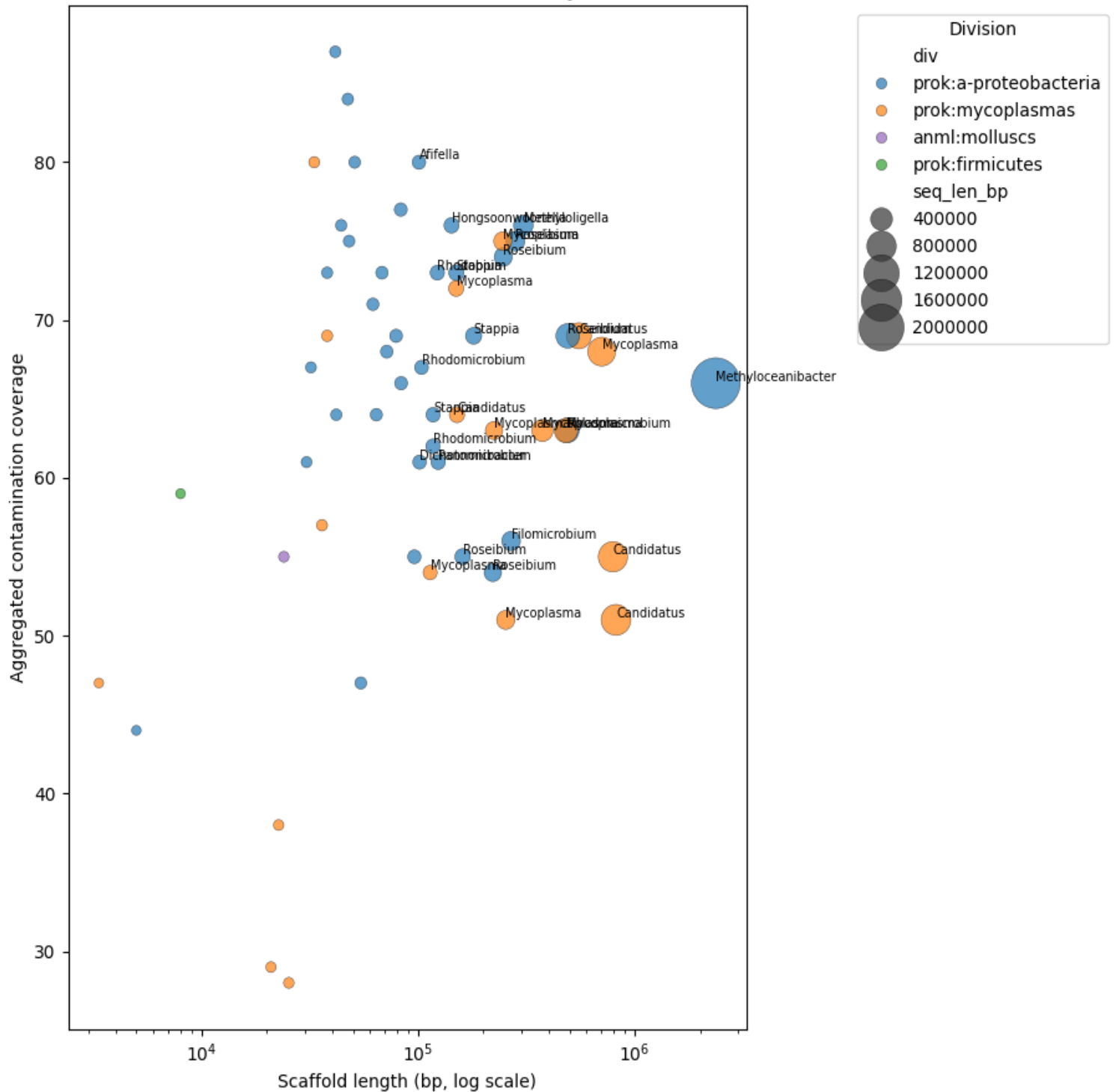Distribution of k-mer counts per copy
numbers found in asm

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

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

Post-curation contamination screening

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	Pacbio HiFi	ONT	Illumina	Omni-C
Coverage	27x	21x	17x	69x

Assembly pipeline

- **Filtlong**
    |_ *ver:* 0.2.1
    |_ *key param:* --min_length 1000 --min_mean_q 80
- **Hifiasm**
    |_ *ver:* 0.24.0
    |_ *key param:* "
- **YaHS**
    |_ *ver:* 1.2a
    |_ *key param:* NA
- **CLAWS pipeline**
    |_ *ver:* 2.3.0
    |_ *key param:* NA
- **Blobtoolkit Nextflow pipeline (latest)**
    |_ *ver:* 0.6
    |_ *key param:* NA
- **FOAM pipeline**
    |_ *ver:* 0.5
    |_ *key param:* NA
- **mitos**
    |_ *ver:* 2.1.3
    |_ *key param:* NA

Curation pipeline

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

Submitter: Tyler Alioto
Affiliation: CNAG

Date and time: 2025-10-30 02:19:49 CET