

ERGA Assembly Report

v24.04.03_beta

Tags: ERGA-BGE

TxID	3071549
ToLID	xgAniVori1
Species	Anisus vorticulus
Class	Gastropoda
Order	NA

Genome Traits	Expected	Observed
Haploid size (bp)	1,029,141,755	1,043,023,846
Haploid Number	18 (source: ancestor)	18
Ploidy	2 (source: ancestor)	2
Sample Sex	H	H

EBP metrics summary and curation notes

Obtained EBP quality metric for pri: 6.7.Q54

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

Curator notes

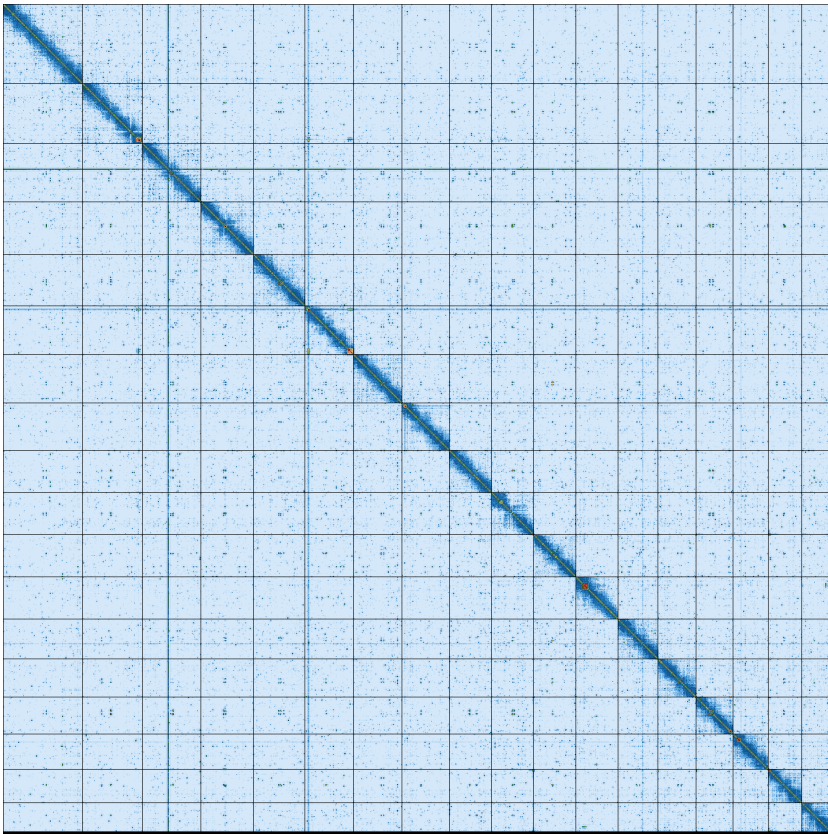
. Interventions/Gb: 123
. Contamination notes: "Total length of scaffolds removed: 1,456,649 (0.1 %);Scaffolds removed: 53 (20.0 %);Largest scaffold removed: (57,806);FCS-GX contaminant species (number of scaffolds; total length of scaffolds):Pararheinheimera texasensis, g-proteobacteria (17; 358,477);Rheinheimera riviphila, g-proteobacteria (11; 369,938);Flavobacterium difficile, CFB group bacteria (5; 193,175);Pseudomonas campi, g-proteobacteria (3; 93,379);Acinetobacter johnsonii, g-proteobacteria (2; 84,971);Hydrogenophaga aromaticivorans, b-proteobacteria (2; 78,989);Flavobacterium amnicola, CFB group bacteria (2; 36,842);Flavobacterium succinicans, CFB group bacteria (1; 41,361);Aquabacterium pictum, b-proteobacteria (1; 37,989);Lacihabitans soyangensis, CFB group bacteria (1; 34,303);Rhodoferax sediminis, b-proteobacteria (1; 32,174);Flavobacterium nackdongense, CFB group bacteria (1; 30,736);Pelomonas sp., b-proteobacteria (1; 29,770);Rubrivivax albus, b-proteobacteria (1; 22,812);Rubrivivax sp., b-proteobacteria (1; 2,000);Flavobacterium bernardetii, CFB group bacteria (1; 2,000);Ideonella sp. WA131b, b-proteobacteria (1; 1,000);Mitochondrion (1; 6,733)"
. Other observations: "This is the second version after feedback from the initial EAR. Chromosomes named by size; Hi-C from a different individual (xgAniVori2), so assembly is not Hi-C phased"

Quality metrics table

Metrics	Pre-curation pri	Curated pri
Total bp	1,049,506,174	1,043,023,846
GC %	35.29	35.28
Gaps/Gbp	662.22	726.73
Total gap bp	139,000	151,600
Scaffolds	265	76
Scaffold N50	55,257,980	59,244,224
Scaffold L50	8	8
Scaffold L90	17	16
Contigs	960	834
Contig N50	2,384,665	2,352,869
Contig L50	129	130
Contig L90	438	442
QV	54.5	54.5
Kmer compl.	98.89	99.27
BUSCO sing.	92.8%	93.2%
BUSCO dupl.	0.9%	0.8%
BUSCO frag.	2.3%	2.0%
BUSCO miss.	4.0%	4.0%

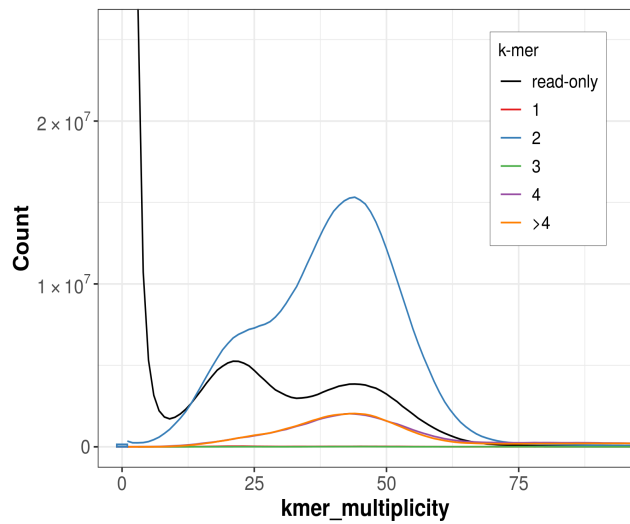
Warning: BUSCO versions or lineage datasets are not the same across results

HiC contact map of curated assembly

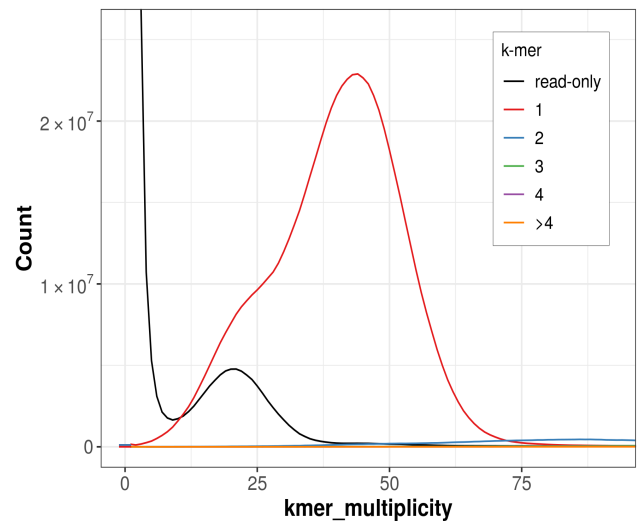


pri [\[LINK\]](#)

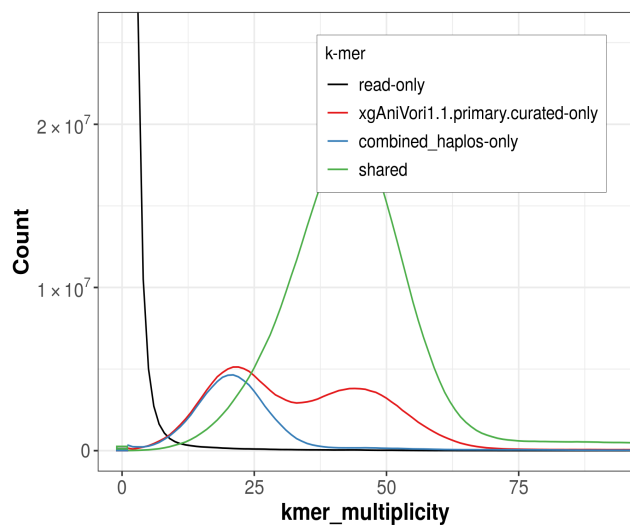
K-mer spectra of curated assembly



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

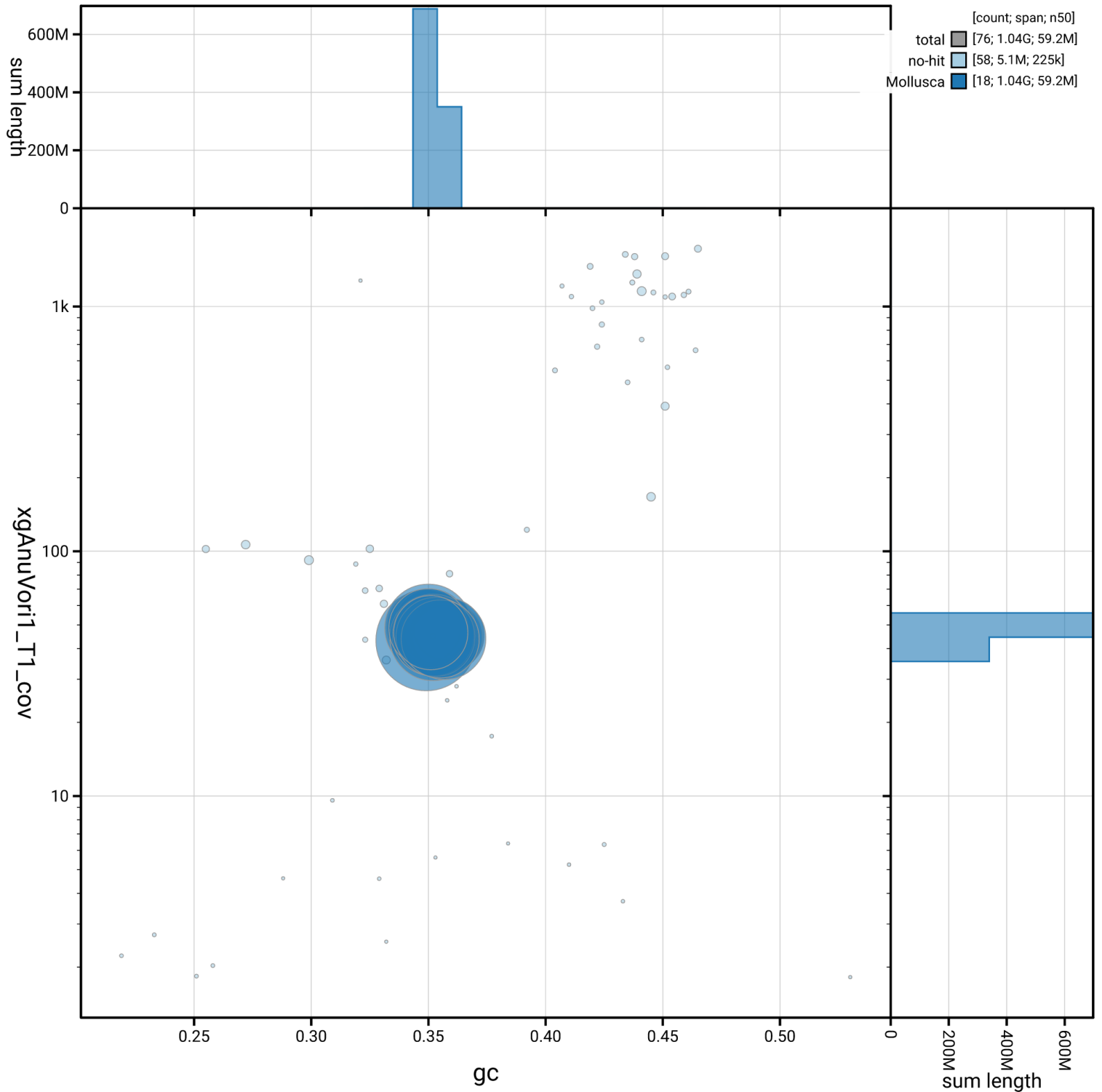


Distribution of k-mer counts per copy numbers found in asm



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	Arima v2
Coverage	45x	142x

Assembly pipeline

- **hifiasm**
 - |_ *ver*: 0.19.8-r603
 - |_ *key param*: --primary
- **purge_dups**
 - |_ *ver*: 1.2.5
 - |_ *key param*: -e
- **yahs**
 - |_ *ver*: 1.2a.2
 - |_ *key param*: NA

Curation pipeline

- **hifiasm**
 - |_ *ver*: 0.19.8-r603
 - |_ *key param*: --primary
- **purge_dups**
 - |_ *ver*: 1.2.5
 - |_ *key param*: -e
- **yahs**
 - |_ *ver*: 1.2a.2
 - |_ *key param*: NA
- **TreeVal**
 - |_ *ver*: 1.1
 - |_ *key param*: NA

Submitter: Michael Paulini

Affiliation: WSI

Date and time: 2024-09-09 12:44:20 CEST