

ERGA Assembly Report

v24.04.03_beta

Tags: ERGA-BGE

TxID	60953
ToLID	wrPlaCost1
Species	Placobdella costata
Class	Clitellata
Order	Rhynchobdellida

Genome Traits	Expected	Observed
Haploid size (bp)	486,921,211	488,180,838
Haploid Number	9 (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.Q49

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
- . Assembly length loss > 3% for pri
- . More than 1000 gaps/Gbp for pri

Curator notes

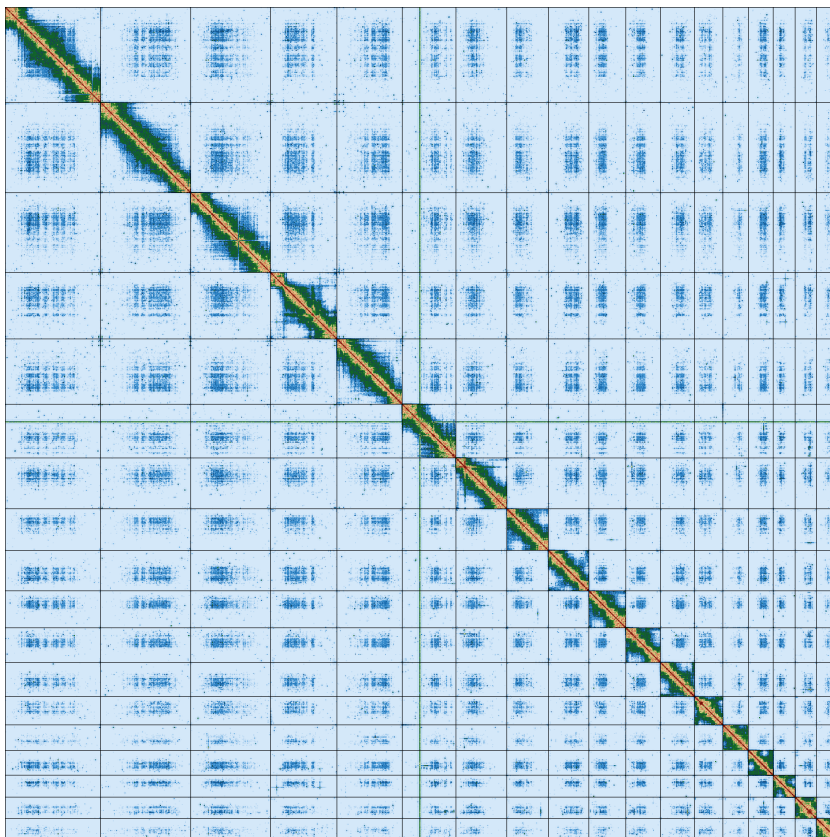
- . Interventions/Gb: 159
- . Contamination notes: "Total length of scaffolds removed: 7,620,786 (1.5 %);Scaffolds removed: 12 (12.1 %);Largest scaffold removed: (3,924,856);FCS-GX contaminant species (number of scaffolds; total length of scaffolds):Yokenella regensburgei, g-proteobacteria (4; 4,053,046);Myxococcus qinghaiensis, d-proteobacteria (1; 3,257,129);Chryseotalea sanaruensis, CFB group bacteria (1; 56,177);Polyangium spumosum, d-proteobacteria (1; 44,272);Citrobacter youngae, g-proteobacteria (1; 43,784);Chryseolinea soli, CFB group bacteria (1; 43,047);Chryseolinea sp., CFB group bacteria (1; 32,751);Chryseotalea sp. WA131a, CFB group bacteria (1; 28,284)"
- . Other observations: "Hi-C from a different individual (wrPlaCost2), so assembly is not Hi-C phased; Expected haploid number from GoaT is an ancestral estimate, likely incorrect in this case"

Quality metrics table

Metrics	Pre-curation pri	Curated pri
Total bp	504,255,463	488,180,838
GC %	32.14	31.77
Gaps/Gbp	1,253.33	1,349.91
Total gap bp	126,400	131,800
Scaffolds	99	28
Scaffold N50	32,061,669	31,539,400
Scaffold L50	6	6
Scaffold L90	15	15
Contigs	731	687
Contig N50	1,889,692	1,764,884
Contig L50	64	64
Contig L90	337	339
QV	49.5	49.5
Kmer compl.	98.85	97.56
BUSCO sing.	89.0%	89.7%
BUSCO dupl.	1.0%	0.7%
BUSCO frag.	4.8%	4.3%
BUSCO miss.	5.2%	5.3%

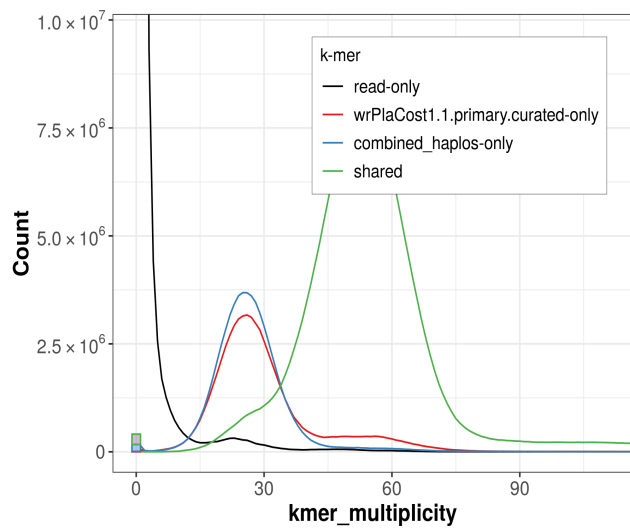
Warning: BUSCO versions or lineage datasets are not the same across results

HiC contact map of curated assembly

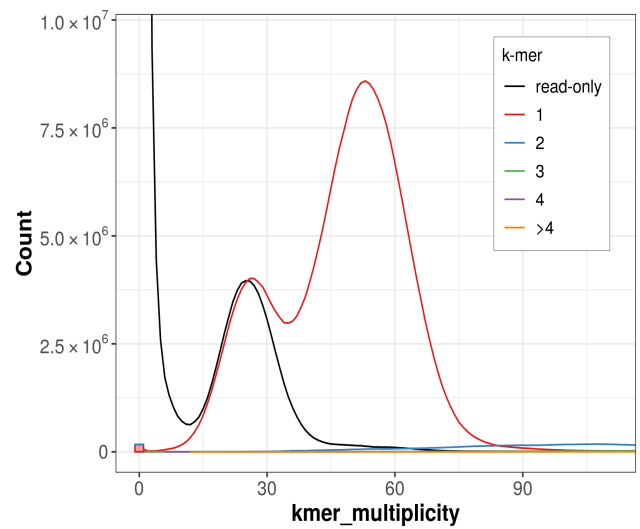


pri [\[LINK\]](#)

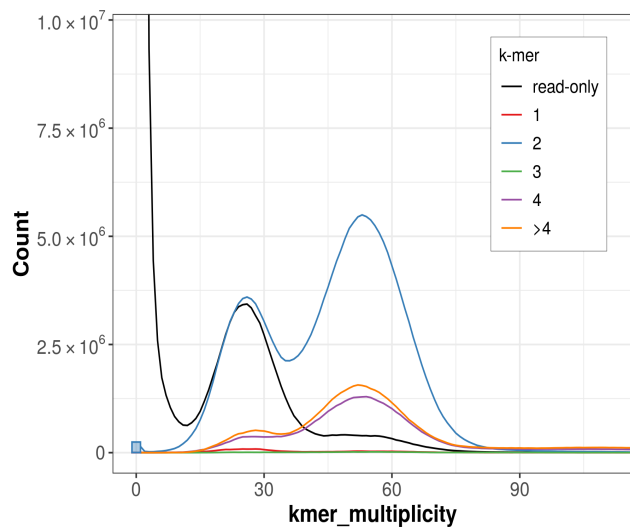
K-mer spectra of curated assembly



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

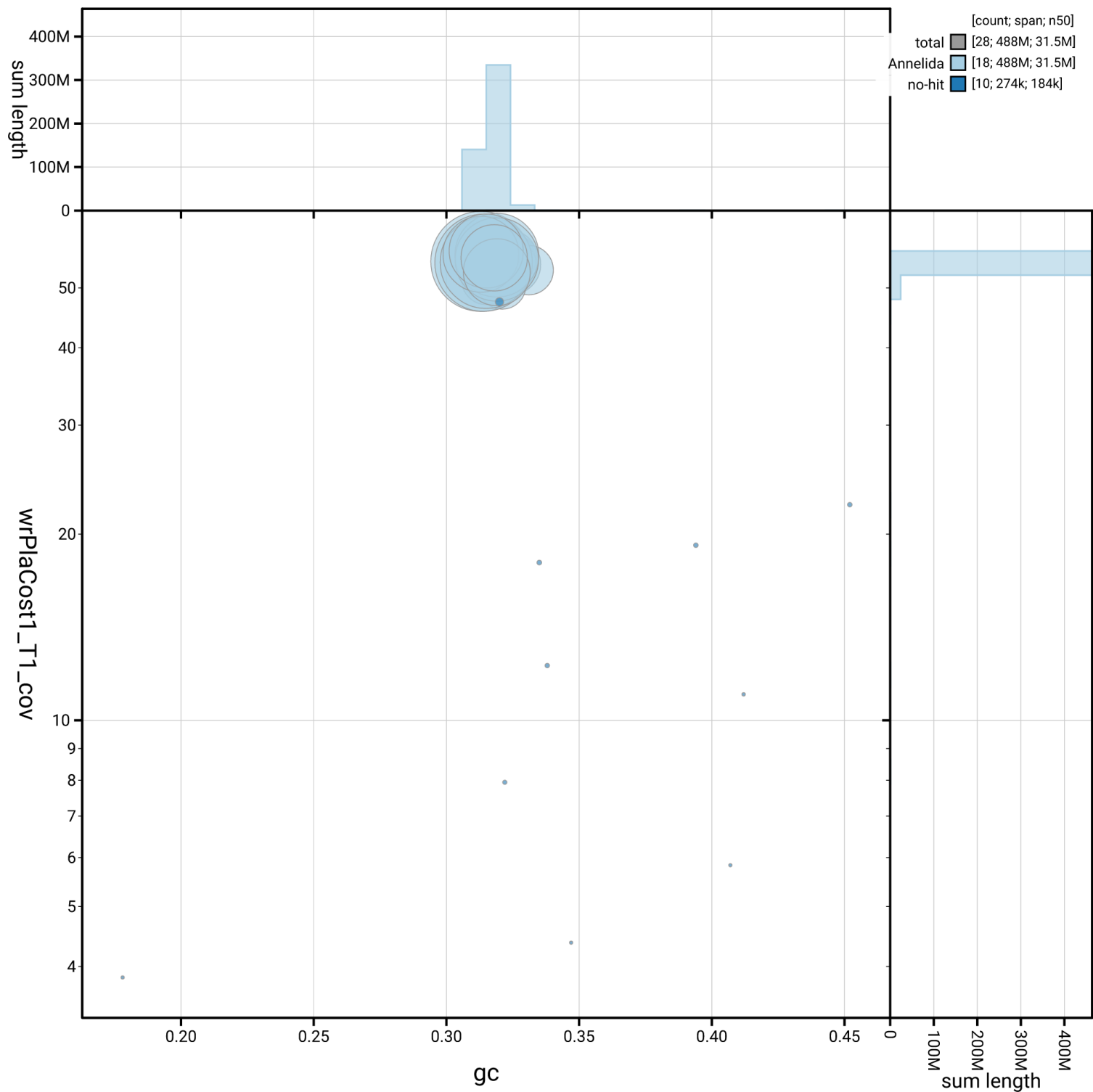


Distribution of k-mer counts per copy numbers found in asm



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

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	54x	204x

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*: 1.1.1
 - |_ *key param*: NA

Submitter: Thomas Mathers

Affiliation: WSI

Date and time: 2024-09-26 11:59:31 CEST