

ERGA Assembly Report

v24.10.15

Tags: ERGA-BGE

TxID	3139642
ToLID	qdBraStyg1
Species	Brachydesmus stygivagus
Class	Diplopoda
Order	Polydesmida

Genome Traits	Expected	Observed
Haploid size (bp)	291,521,673	340,699,620
Haploid Number	12 (source: ancestor)	10
Ploidy	2 (source: ancestor)	2
Sample Sex	unknown	unknown

EBP metrics summary and curation notes

Obtained EBP quality metric for pri: 6.7.Q64

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
- . Assembly length loss > 3% for pri
- . Not 90% of assembly in chromosomes for pri

Curator notes

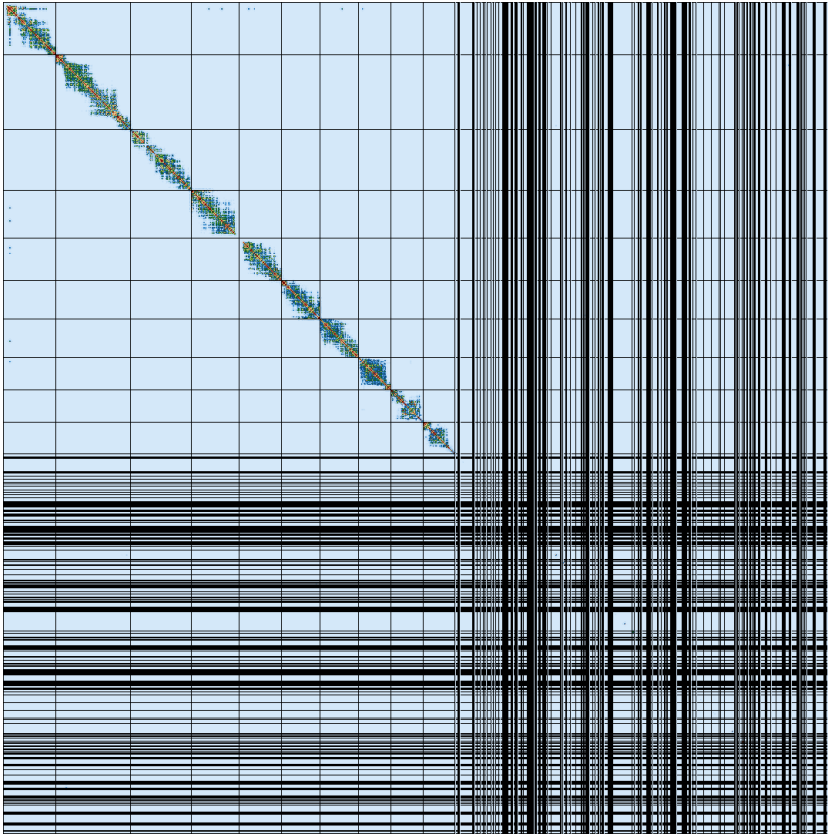
- . Interventions/Gb: 300
- . Contamination notes: "BlobTools identified more than 300 small scaffolds as potential contaminants, primarily corresponding to bacteria, with additional matches to fungi and chordates. With the exception of a couple, these contaminated scaffolds were removed from the assembly provided for review. The elimination of these putative contaminants did not seem to affect the BUSCO scores"
- . Other observations: "B. stygivagus has a repeat rich genome. For this species, we selected the primary assembly due to its superior contiguity, as none of the alternative haplotypes reached a cN50 of 1 Mb. The PacBio assembly remained fragmented, and the HiC library quality was suboptimal because the long read and HiC datasets originated from different individuals. As a result, the YaHS scaffolding required more extensive curation than usual. We were unable to identify a sex chromosome, as no scaffold showed halved coverage and no closely related species with annotated sex chromosomes was available for comparison. Both MQ_0 and MQ_10 Pretext maps were used to guide the curation process. Each has a save state in the directory I have shared"

Quality metrics table

Metrics	Pre-curation pri	Curated pri
Total bp	357,806,435	340,699,620
GC %	32.79	32.4
Gaps/Gbp	433.2	548.87
Total gap bp	31,000	37,400
Scaffolds	685	350
Scaffold N50	3,203,691	13,201,337
Scaffold L50	23	9
Scaffold L90	176	91
Contigs	840	537
Contig N50	1,566,716	1,599,458
Contig L50	60	56
Contig L90	279	233
QV	63.7115	64.5638
Kmer compl.	86.2459	86.0617
BUSCO sing.	92.4%	92.7%
BUSCO dupl.	1.4%	1.0%
BUSCO frag.	2.8%	2.9%
BUSCO miss.	3.4%	3.4%

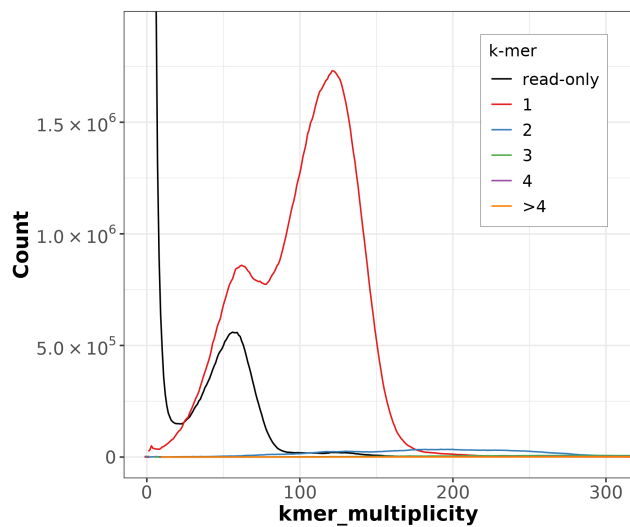
BUSCO: 5.5.0 (euk_genome_met, metaeuk) / Lineage: arthropoda_odb10 (genomes:90, BUSCOs:1013)

HiC contact map of curated assembly

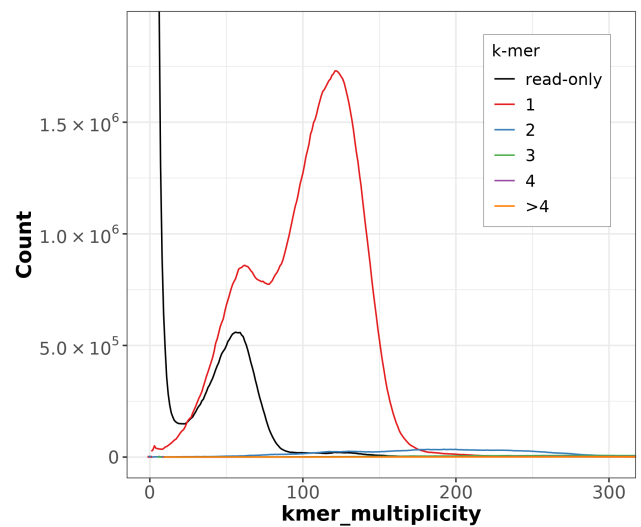


pri [\[LINK\]](#)

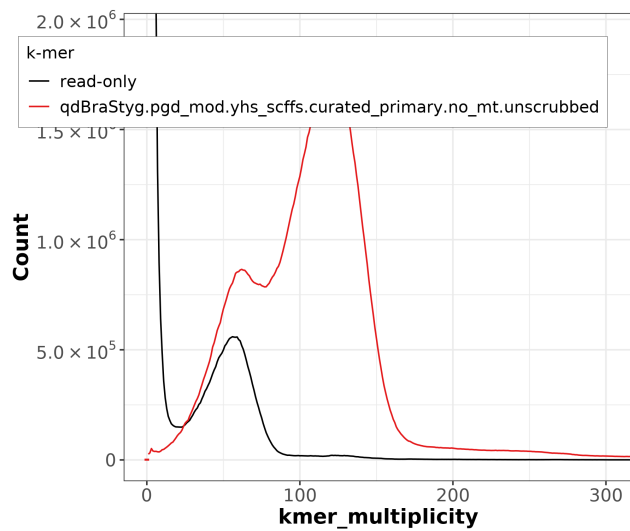
K-mer spectra of curated assembly



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

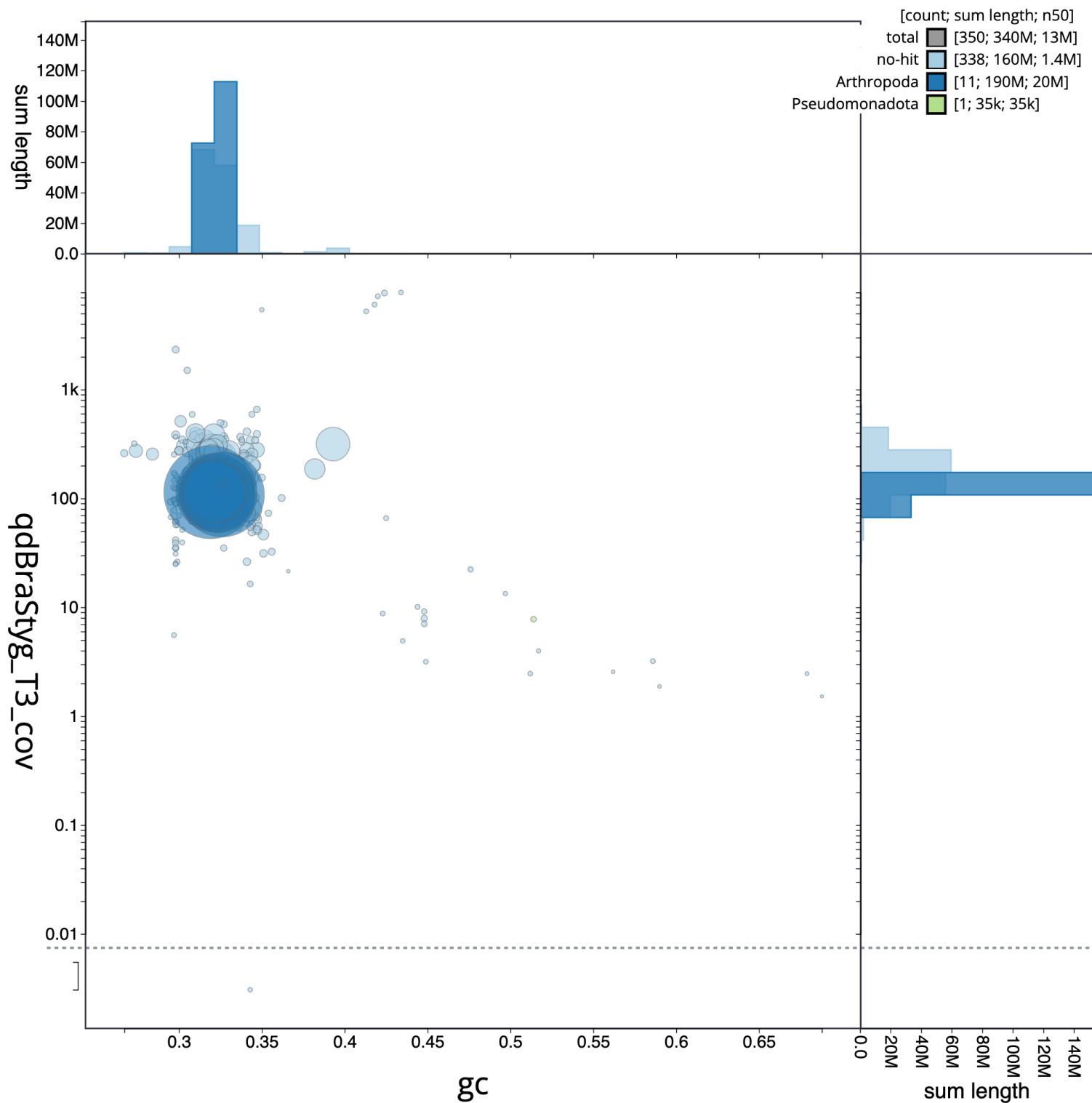


Distribution of k-mer counts per copy numbers found in asm



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	OmniC
Coverage	100x	105x

Assembly pipeline

- **hifiasm**
 - |_ *ver*: 0.24.0-r702
 - |_ *key param*: NA
- **purge_dups**
 - |_ *ver*: 1.2.6
 - |_ *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: 2026-01-09 16:42:45 CET