ERGA Assembly Report

v24.05.20_glxy_beta

Tags: ERGA-BGE

TxID	2934182
ToLID	**xgPhyFlav1**
Species	Phyllidia flava
Class	Gastropoda
Order	Nudibranchia

Genome Traits	Expected	Observed
Haploid size (bp)	3,011,504,214	3,343,225,297
Haploid Number	12 (source: ancestor)	28
Ploidy	2 (source: ancestor)	2
Sample Sex	XX	XX

EBP metrics summary and curation notes

Obtained EBP quality metric for hap1: 7.8.Q6
Obtained EBP quality metric for hap2: 7.8.Q69

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

. Observed Haploid Number is different from Expected

. QV value is less than 40 for hap1
. Kmer completeness value is less than 90 for hap1
. Assembly length loss > 3% for hap1
. Assembly length loss > 3% for hap2

Curator notes

. Interventions/Gb: 2
. Contamination notes: "No presence of contaminants. Mitochondrial genome was removed
from the assembly"
. Other observations: "Large collapsed repeat in chr5, haplotypic inversion in chr12"

Quality metrics table

Metrics	Pre-curation hap1	Pre-curation hap2	Curated hap1	Curated hap2
Total bp	3,660,021,265	3,238,925,270	3,343,225,297	3,140,900,549
GC %	41.62	41.25	41.09	41.07
Gaps/Gbp	34.97	23.77	22.13	19.74
Total gap bp	1,874,523	3,509,524	138,686	539,693
Scaffolds	973	555	104	91
Scaffold N50	111,683,922	116,290,662	125,437,426	125,802,197
Scaffold L50	12	10	10	9
Scaffold L90	35	25	24	23
Contigs	1,101	632	178	153
Contig N50	73,275,821	88,190,939	77,716,104	88,213,608
Contig L50	16	12	14	11
Contig L90	85	40	43	36
QV	68.0694	68.4744	6	69.3854
Kmer compl.	95.5615	90.8003	9	90.8758
BUSCO sing.	95.3%	92.9%	95.2%	92.8%
BUSCO dupl.	0.7%	0.8%	0.7%	0.8%
BUSCO frag.	1.1%	1.1%	1.1%	1.1%
BUSCO miss.	2.9%	5.2%	3.0%	5.3%

BUSCO 5.4.7 Lineage: mammalia_odb10 (genomes:24, BUSCOs:9226)
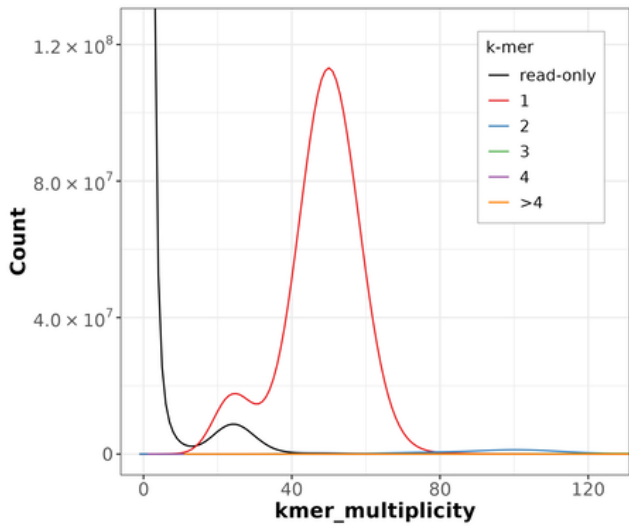
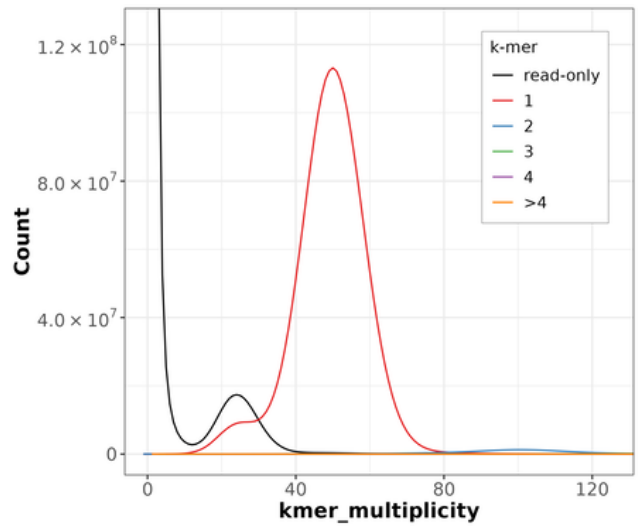# HiC contact map of curated assembly
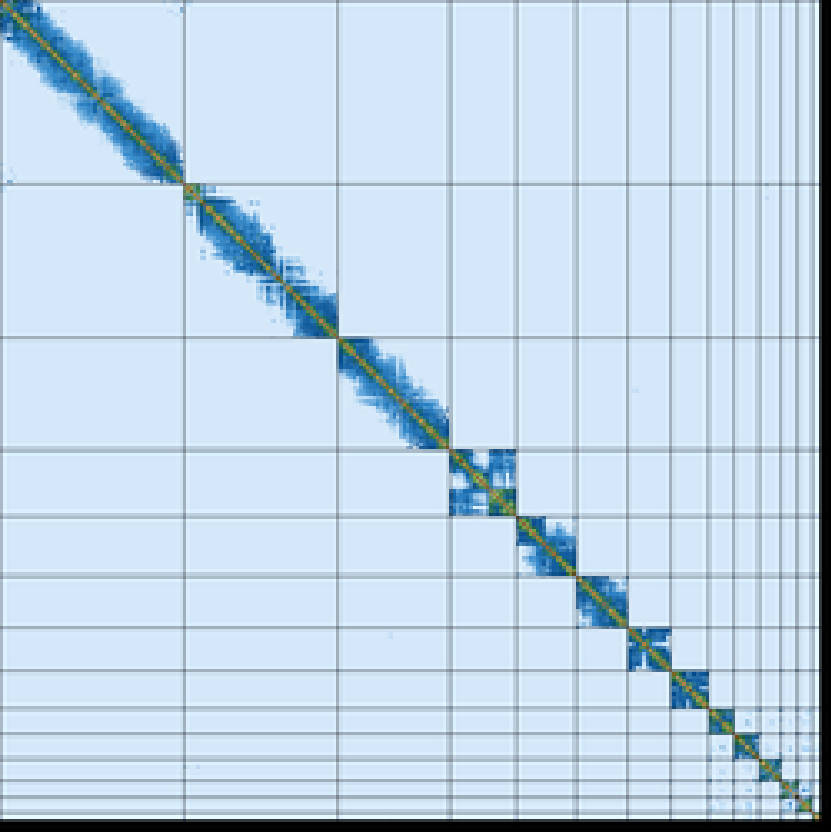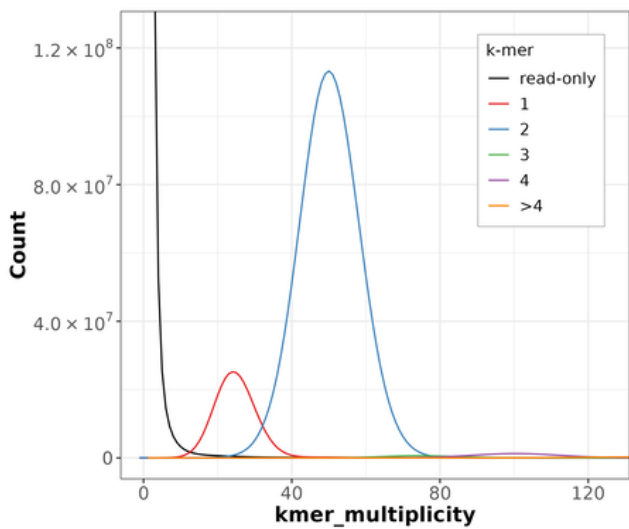

hap1 [LINK]
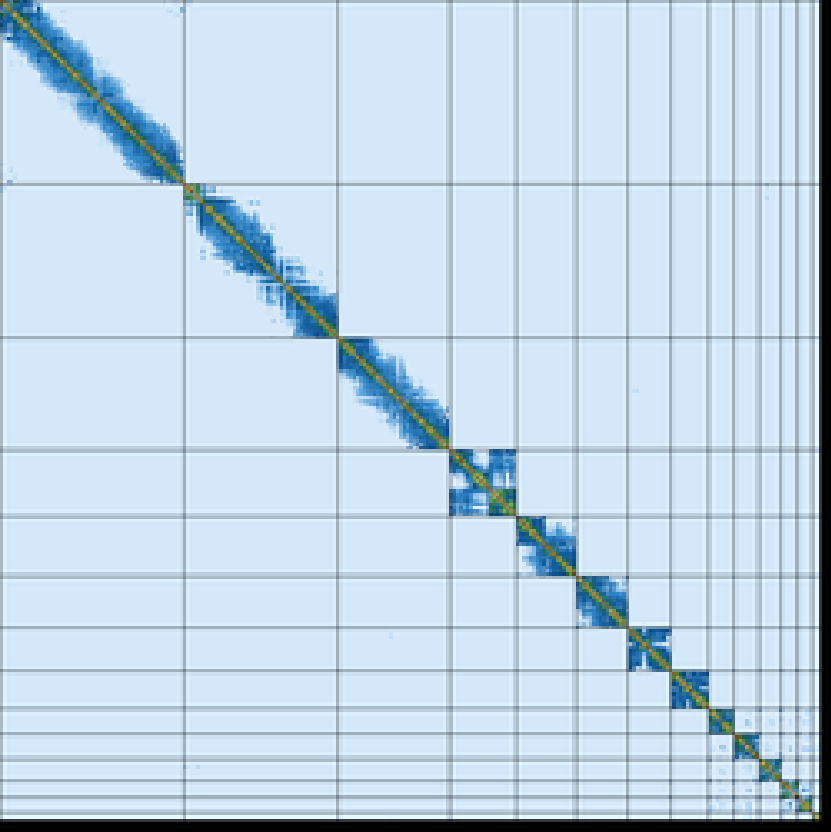
hap2 [LINK]
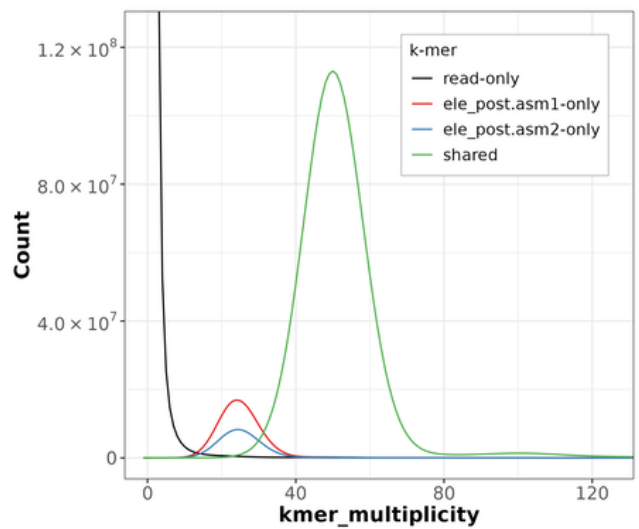
K-mer spectra of curated assembly

dataset_f9122a42-c838-44fc-a8af-8abe91d4351d.dat

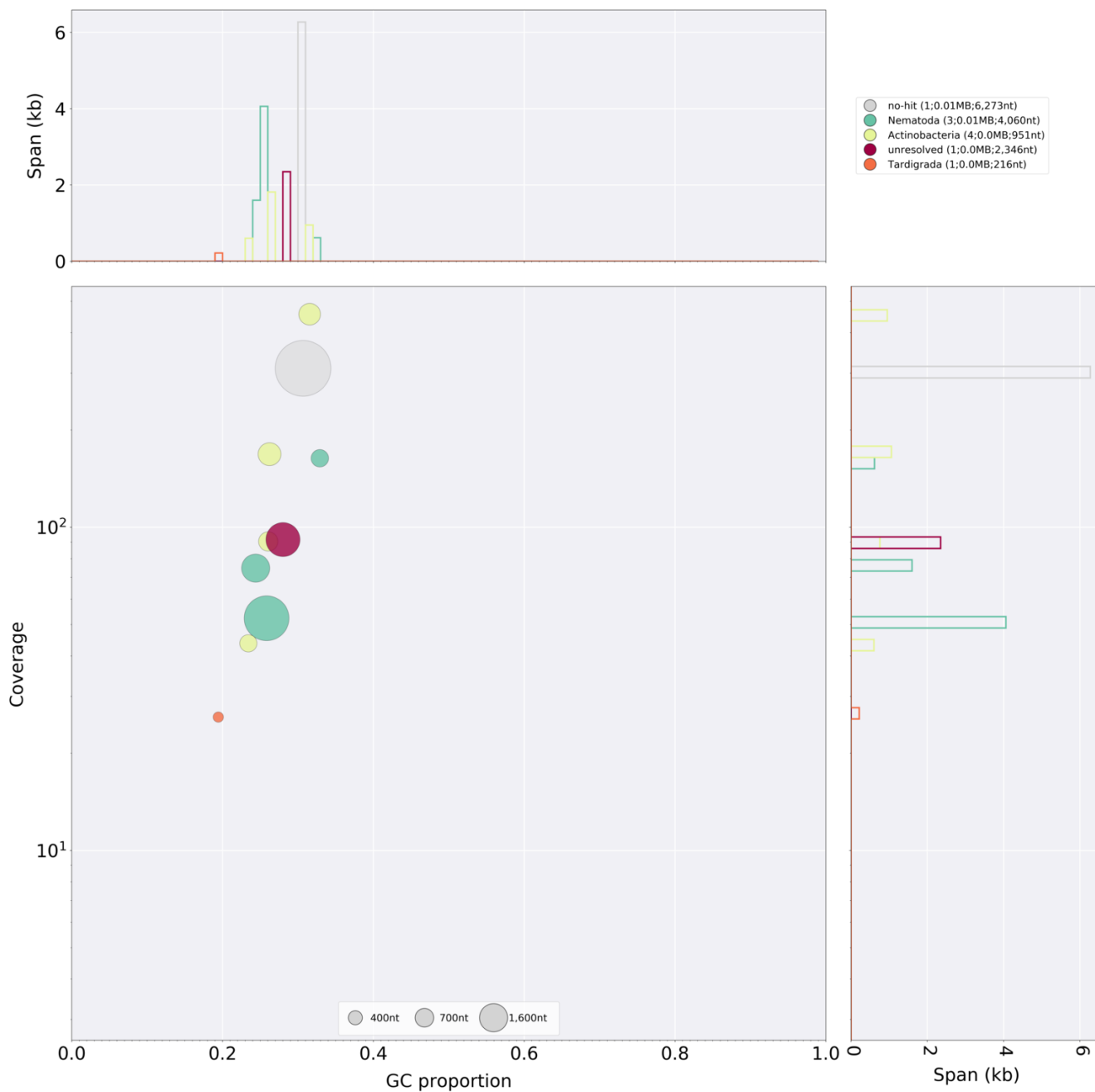dataset_7fa38b17-5b70-4281-9222-9c65709be543.dat

dataset_8d36dd69-969f-4d6c-bfc2-0a91d9cc5022.dat

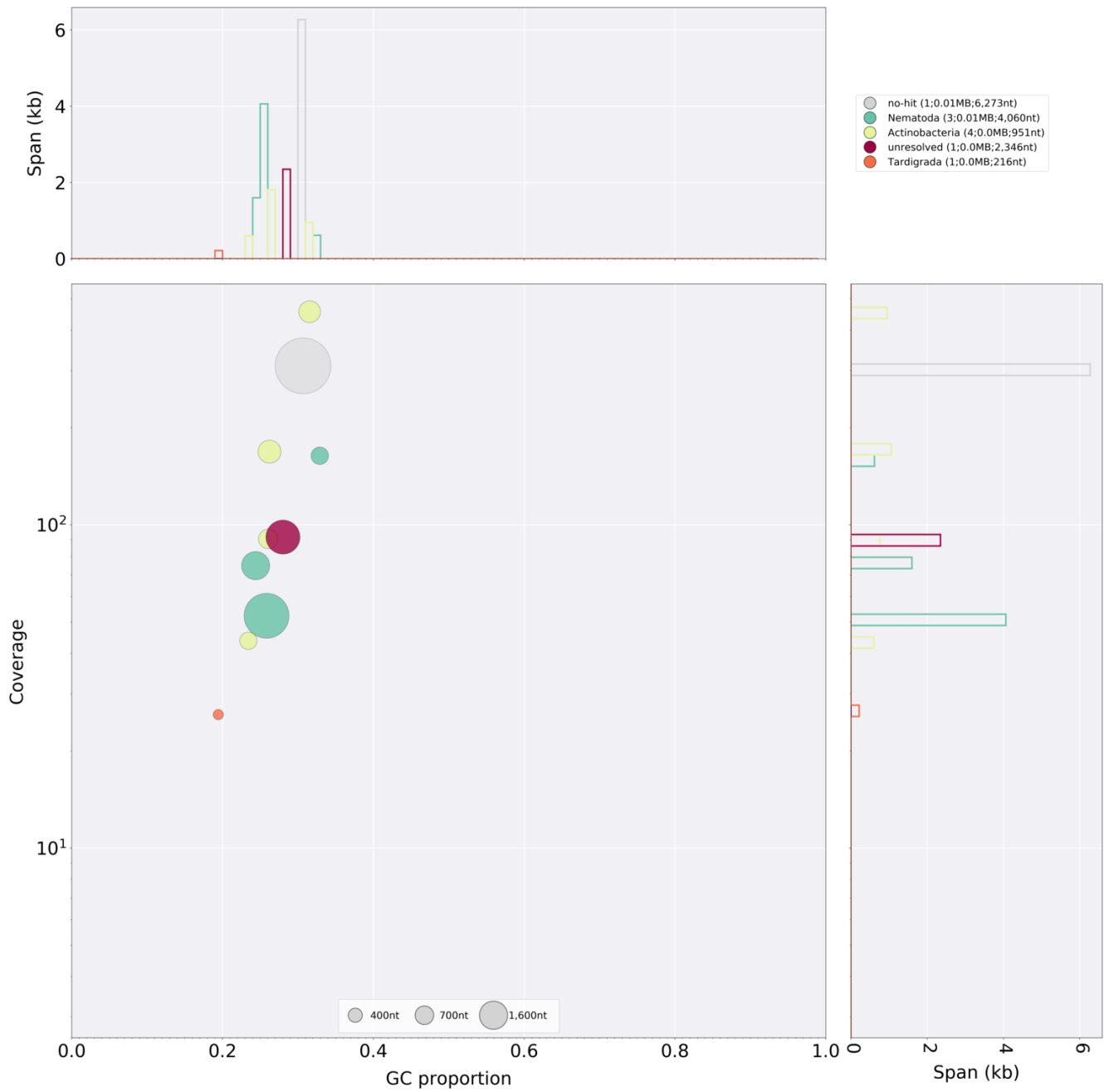dataset_d0f59eeb-c3b6-412c-9610-4b881bb8cdb3.dat

Post-curation contamination screening

example/test.test.blobDB.json.bestsum.phylum.p7.span.100.blobplot.bam0

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.

example/test.test.blobDB.json.bestsum.phylum.p7.span.100.blobplot.bam0

Legend:
- no-hit (1;0.01MB;6,273nt)
- Nematoda (3;0.01MB;4,060nt)
- Actinobacteria (4;0.0MB;951nt)
- unresolved (1;0.0MB;2,346nt)
- Tardigrada (1;0.0MB;216nt)

hap2. 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	HiFi	Bionano	OmniC
Coverage	40x	10x	90x

Assembly pipeline

- **Hifiasm**
  |_ *ver:* 0.19.4
  |_ *key param:* HiC
  |_ *key param:* l0
- **Purge_Dups**
  |_ *ver:* 1.2.6
  |_ *key param:* NA
- **Bionano**
  |_ *ver:* Galaxy_3.7.0
  |_ *key param:* NA
- **YaHS**
  |_ *ver:* 1.1
  |_ *key param:* NA

Curation pipeline

- **GRIT_rapid**
  |_ *ver:* 2.0
  |_ *key param:* NA
- **HiGlass**
  |_ *ver:* 1.0
  |_ *key param:* NA

Submitter: John Doe
Affiliation: Galaxy EU

Date and time: 2024-07-04 14:24:04 CEST