

ERGA Assembly Report

v24.02.09_beta

Tags: ERGA-BGE

ToLID	xgPhyFlav1
Species	Phyllidia flava
Class	Gastropoda
Order	Nudibranchia

Genome Traits	Expected	Observed
Haploid size (bp)	2,244,081,371	1,900,505,857
Haploid Number	12 (source: ['ancestor'])	13
Ploidy	1 (source: ['ancestor'])	2
Sample Sex	unknown	unknown

EBP metrics summary and curation notes

Obtained EBP quality metric for collapsed: 6.8.Q40

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

- . Observed Haploid Number is different from Expected
- . Observed Ploidy is different from Expected
- . Kmer completeness value is less than 90 for collapsed
- . More than 1000 gaps/Gbp for collapsed

Curator notes

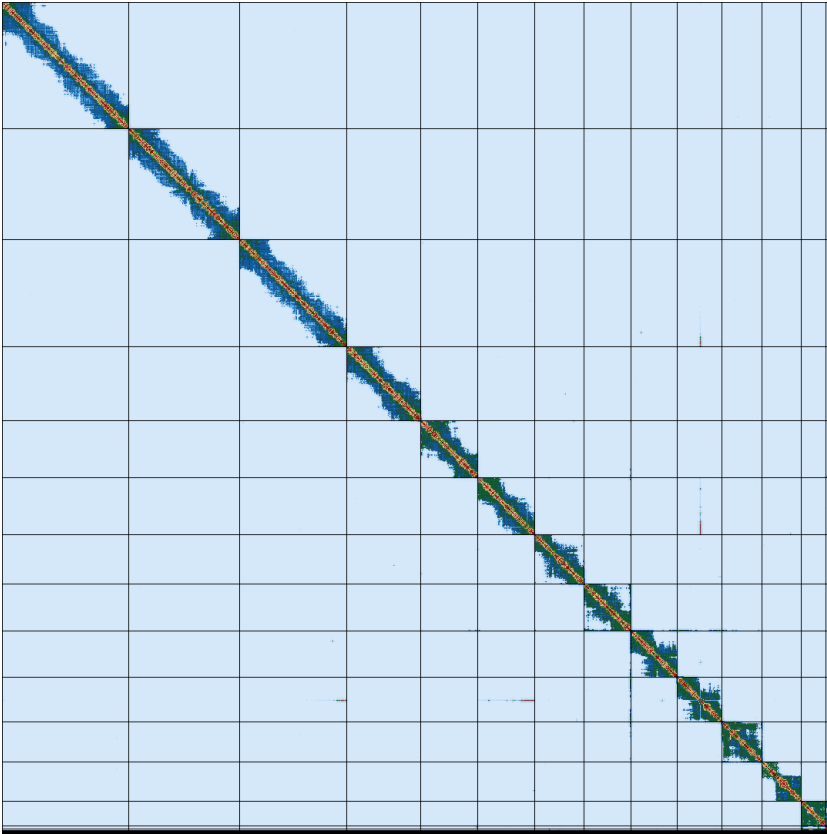
- . Interventions/Gb: 3
- . Contamination notes: "We found a large bacterial sequence (+5Mb) corresponding to a species of proteobacter (pseudomonadota) which was removed (scaffold_15). Furthermore several other scaffolds (75) were removed when they were deemed to be largely made up of contaminants ascribed to different phyla (i.e. arthropoda, chordata, echinodermata, cnidaria, mucoromycota and rotifera)"
- . Other observations: "Curation was a bit difficult due to a fairly low contiguity (2.8 Mb contig N50). We had to break one of the large yahs-assembled scaffolds (scaffold_4) as there was no support for it being a single scaffold but rather two. Some larger unplaced scaffolds/unlocs were not removed even if containing some contaminating sequences since their removal would have resulted in a reduction in the number of busco mollusca & metazoa complete genes"

Quality metrics table

Metrics	Pre-curation collapsed	Curated collapsed
Total bp	1,918,628,656	1,900,505,857
GC %	36.22	36.18
Gaps/Gbp	1,027.82	1,033.41
Total gap bp	394,400	392,800
Scaffolds	141	63
Scaffold N50	242,559,295	169,704,197
Scaffold L50	4	4
Scaffold L90	10	11
Contigs	2,113	2,027
Contig N50	2,386,647	2,398,675
Contig L50	215	212
Contig L90	1,111	1,080
QV	40.4358	40.4285
Kmer compl.	74.9697	74.4559
BUSCO sing.	94.7%	95.0%
BUSCO dupl.	0.7%	0.4%
BUSCO frag.	2.0%	2.0%
BUSCO miss.	2.6%	2.6%

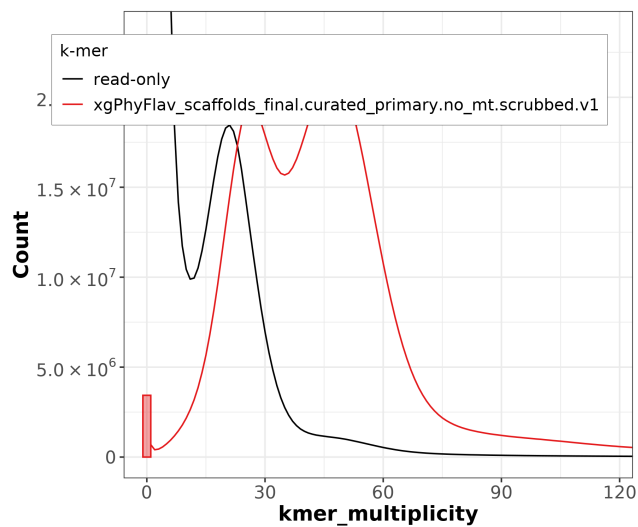
BUSCO 5.4.0 Lineage: metazoa_odb10 (genomes:65, BUSCOs:954)

HiC contact map of curated assembly

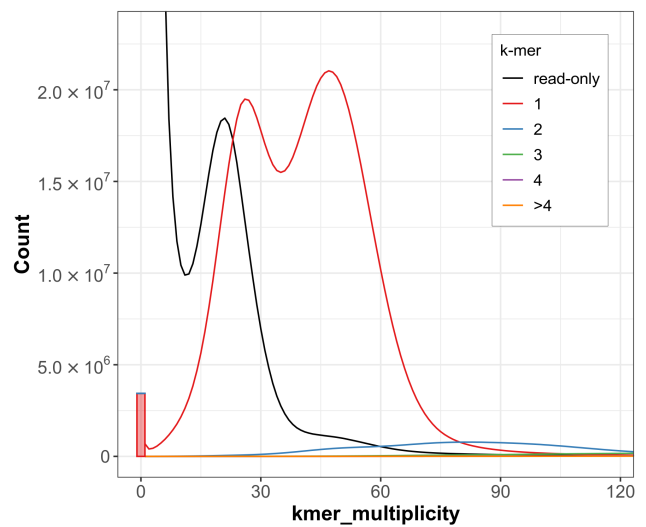


collapsed [\[LINK\]](#)

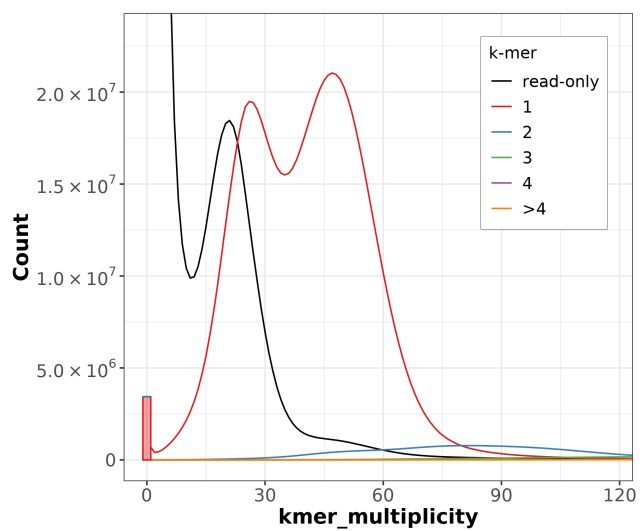
K-mer spectra of curated assembly



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

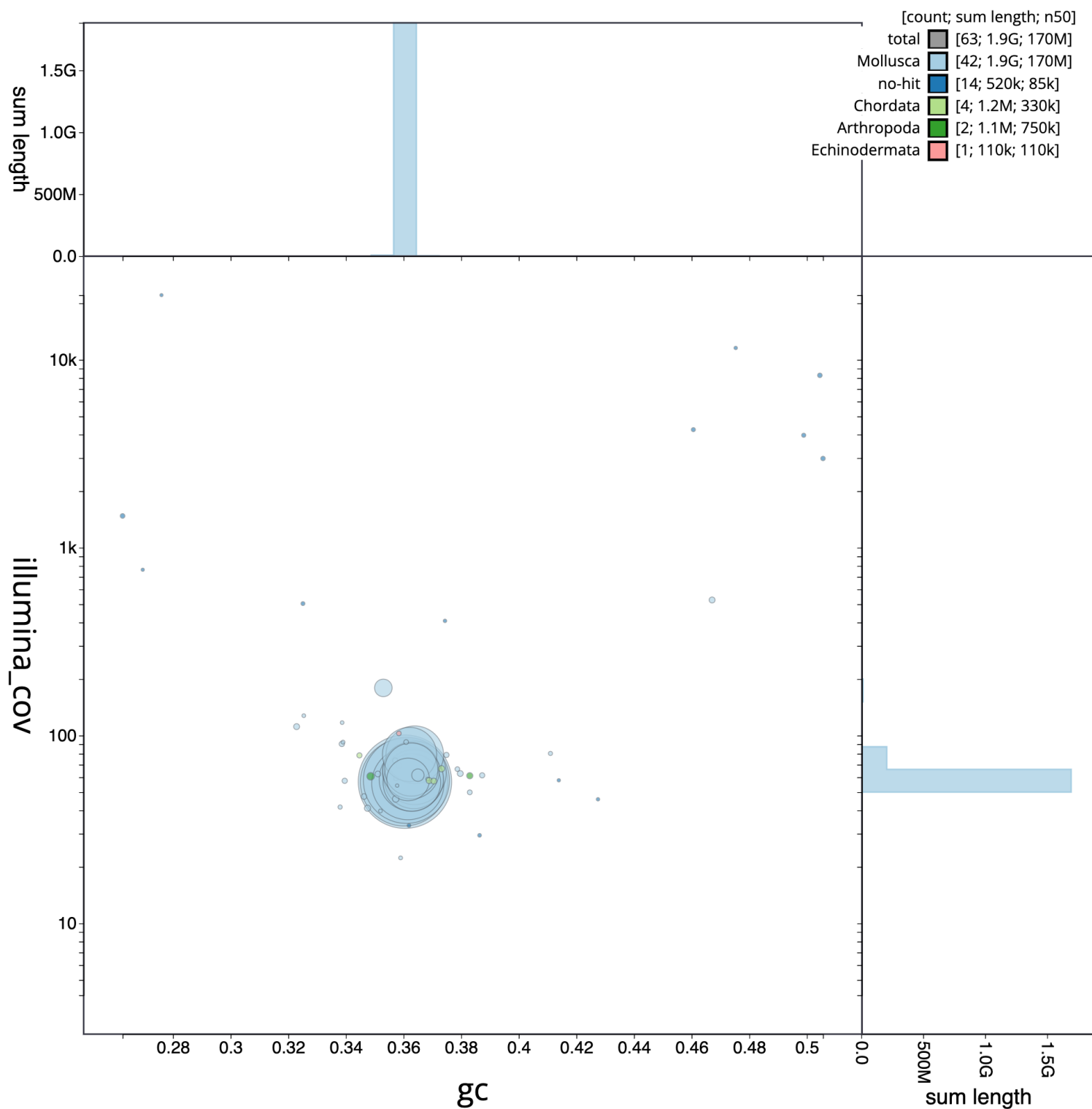


Distribution of k-mer counts per copy numbers found in asm



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

Post-curation contamination screening



collapsed. 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	ONT	Illumina	OmniC
Coverage	72x	74x	64x

Assembly pipeline

- **Trim_Galore**
 - |_ *ver*: 0.6.7
 - |_ *key param*: "--gzip -q 20"
 - |_ *key param*: "--paired"
 - |_ *key param*: "--retain_unpaired"
- **Filtlong**
 - |_ *ver*: 0.2.1
 - |_ *key param*: NA
- **nextdenovo**
 - |_ *ver*: 2.5.0
 - |_ *key param*: NA
- **nextpolish**
 - |_ *ver*: 1.4.1
 - |_ *key param*: NA
- **purge_dups**
 - |_ *ver*: 1.2.6
 - |_ *key param*: NA
- **YaHS**
 - |_ *ver*: 1.2a
 - |_ *key param*: NA

Curation pipeline

- **PretextView**
 - |_ *ver*: 0.2.5
 - |_ *key param*: NA

Submitter: Francisco Camara

Affiliation: CNAG Barcelona

Date and time: 2024-03-07 15:54:47 CET