ERGA Assembly Report
v24.10.15

Tags: ERGA-BGE

TxID	43341
ToLID	**xgPhoTurb1**
Species	Phorcus turbinatus
Class	Gastropoda
Order	Trochida

Genome Traits	Expected	Observed
Haploid size (bp)	947,398,423	952,421,223
Haploid Number	9 (source: ancestor)	18
Ploidy	2 (source: ancestor)	2
Sample Sex	unknown	unknown

EBP metrics summary and curation notes

Obtained EBP quality metric for pri: 7.7.Q65

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

Curator notes

. Interventions/Gb: 1
. Contamination notes: "BTK detected 817 putative contaminant scaffolds using nt, reference proteomes and 15 BUSCO databases. As the The mitogenome hit to Anopheles gambiae and a hit to Mytilus coruscus were considered false positives, we removed 815 scaffolds shorter than 445Kb and accounting for 36 Mb of sequence."
. Other observations: "The hifiasm primary assembly was highly contiguous and complete. However, after performing a whole-genome alignment (WGA) against Phorcus lineatus (xgPhoLine1.1, with 15 MYA divergence) we observed high colinearity between the larger contigs in our assembly and the P. lineatus' chromosomes, except for chromosome 17. Thanks to the WGA, we indentified a 5Mb carrying the telomere end of the chromosome 17 homologue and joined the two contigs (i.e. SUPER_17), ending up with 18 super-scaffolds. Given the little amount of sample received for the sequenced specimen, and the fact that a high proportion of snail shell was shattered and mixed with the remaining tissue, the cell lysate was not enough to obtain a HiC library. In absence of HiC data for this species, we mapped the available HiC reads for xgPhoLine1.1 using a MQ threshold of 10 and obtaining more than 153.8 million valid pairs. The contact map confirms the chromosome territories and the likely-real

chromosome inversions observed in the WGA between both species. Noticeably, 14 out of 18 superscaffolds present telomere sequence peaks at both ends while the other 4 have a peak at just one chromosome end. The shared folder contains these three evidences: 1. synteny:WGA ,2. telomeres, 3. hic of sister species. These are supporting the chromosome level assembly (xgPhoTurb1.1) presented here, that has identical observed haploid number to Phorcus lineatus. Regarding the organelles, we assembled the mitogenome using FOAM v0.5 into a single circular contig of 17,345 bp staring with the trnF gene."

Quality metrics table

| Metrics	Pre-curation pri	Curated pri
Total bp	987,726,549	952,421,223
GC %	35.71	35.49
Gaps/Gbp	1.01	1.05
Total gap bp	500	500
Scaffolds	918	104
Scaffold N50	50,589,882	50,589,882
Scaffold L50	8	8
Scaffold L90	17	16
Contigs	919	105
Contig N50	50,589,882	50,589,882
Contig L50	8	8
Contig L90	17	16
QV	52.9659	65.5484
Kmer compl.	80.2551	79.9082
BUSCO sing.	97.2%	97.5%
BUSCO dupl.	0.8%	0.5%
BUSCO frag.	1.2%	1.2%
BUSCO miss.	0.8%	0.8%

BUSCO: 5.5.0 (euk_genome_met, metaeuk) / Lineage: metazoa_odb10 (genomes:65, BUSCOs:954)
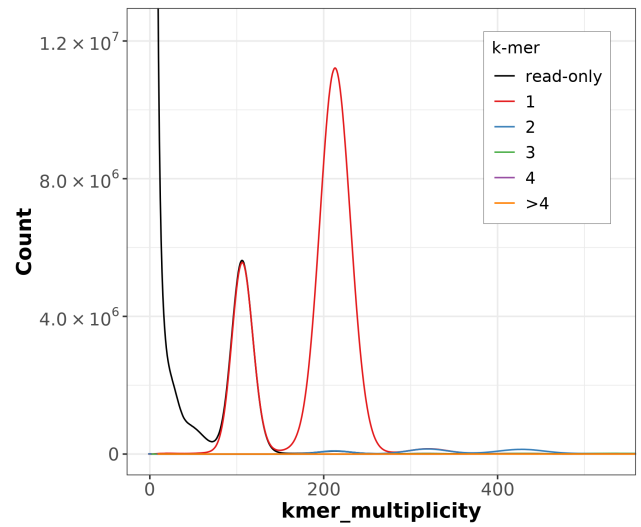
HiC contact map of curated assembly
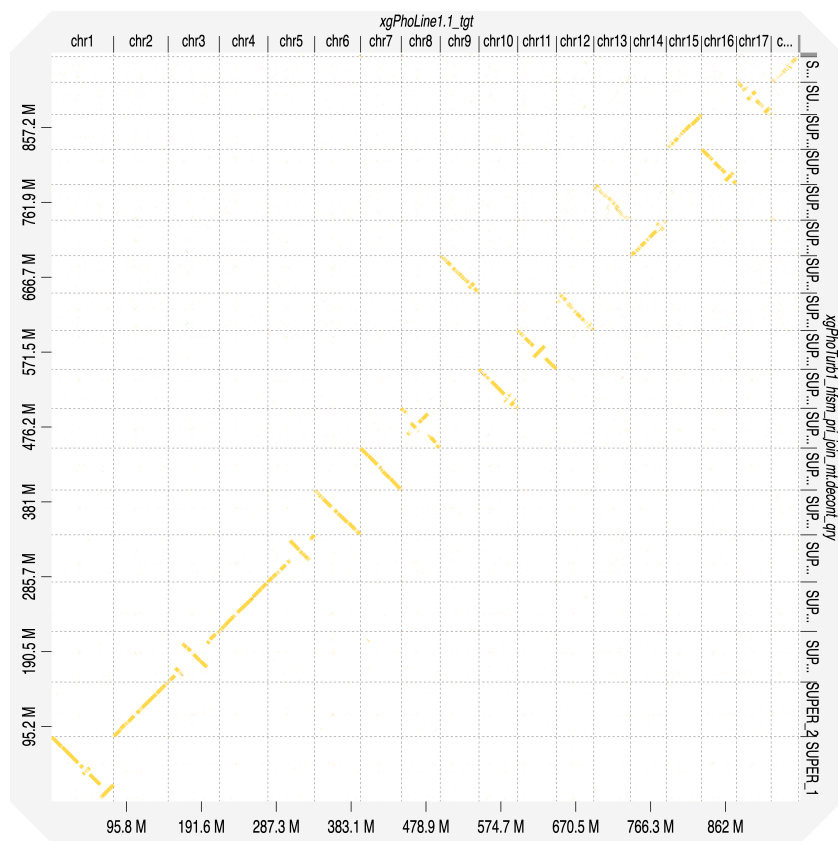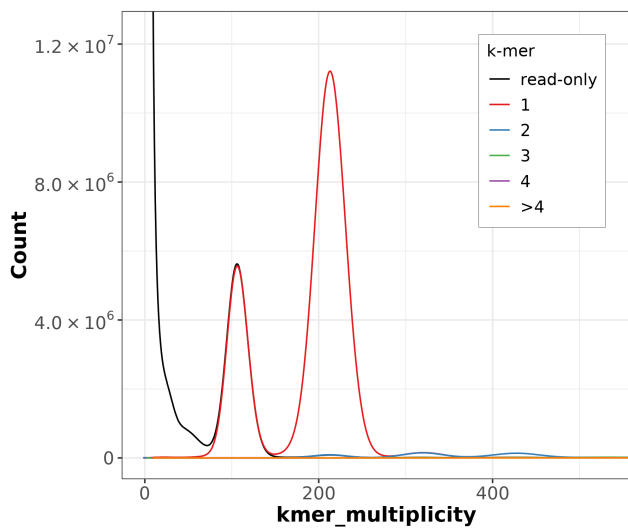

pri [LINK]

K-mer spectra of curated assembly
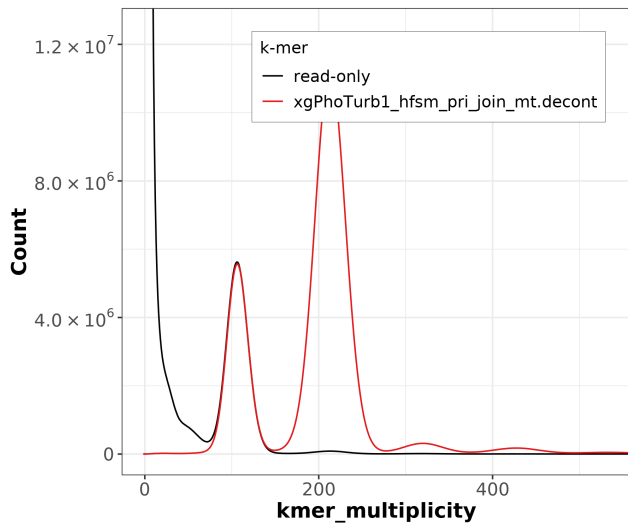
Distribution of k-mer counts coloured by
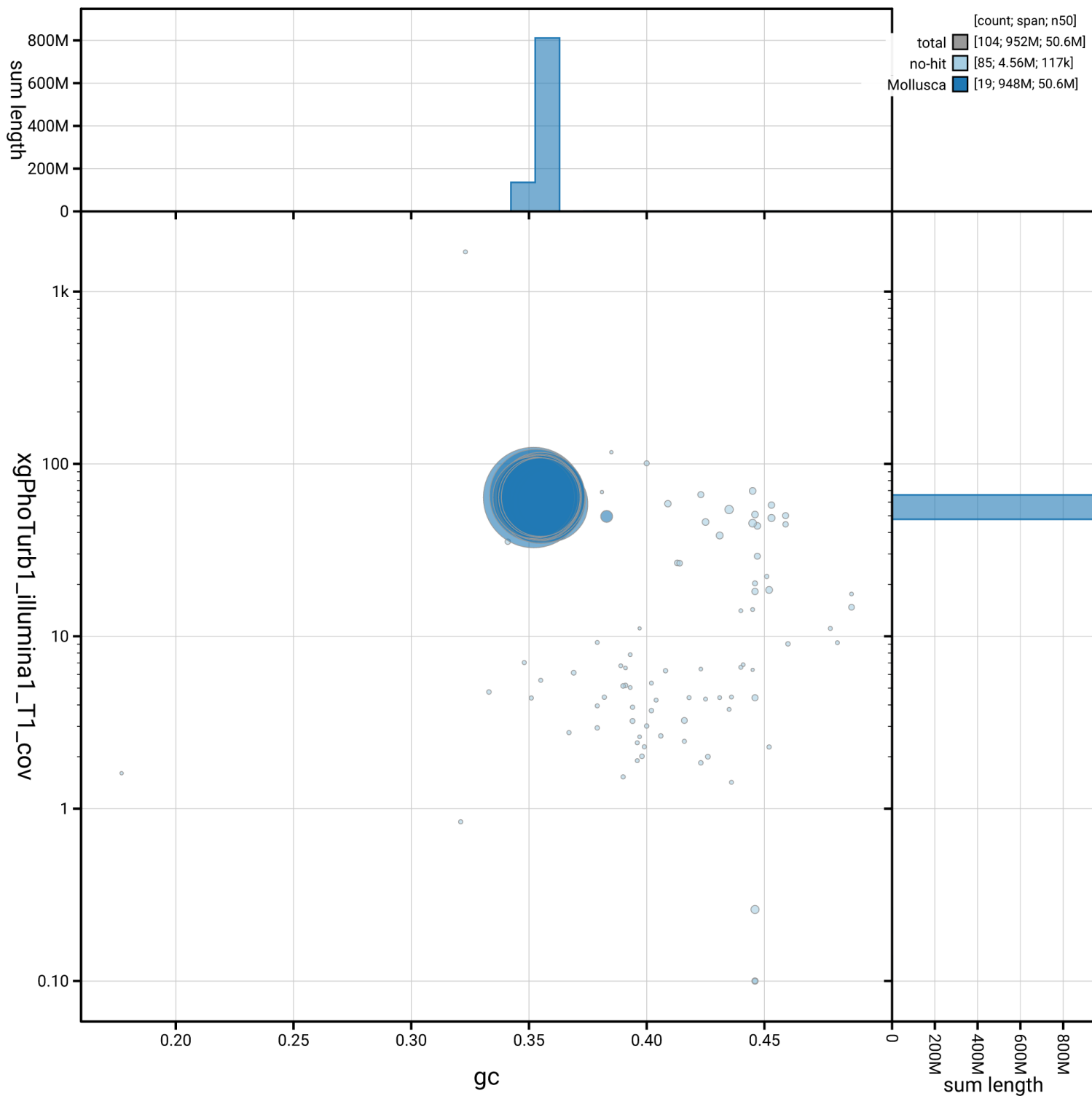their presence in reads/assemblies

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	ONT	Illumina	OmniC
Coverage	234x	135x	0x

Assembly pipeline

- **CLAWS pipeline**
  |_ *ver:* 2.2.0 and 2.3.0
  |_ *key param:* NA
- **Trim_galore**
  |_ *ver:* 0.6.7
  |_ *key param:* NA
- **Filtlong**
  |_ *ver:* 0.2.1
  |_ *key param:* NA
- **flye**
  |_ *ver:* 2.9.1
  |_ *key param:* NA
- **nextdenovo**
  |_ *ver:* 2.5.0
  |_ *key param:* NA
- **hifiasm**
  |_ *ver:* 0.24.0
  |_ *key param:* NA
- **hypo**
  |_ *ver:* 1.0.3
  |_ *key param:* NA
- **purge_dups**
  |_ *ver:* 1.2.6
  |_ *key param:* NA
- **YaHS**
  |_ *ver:* 1.2a
  |_ *key param:* NA
- **Blobtoolkit Nextflow pipeline (latest)**
  |_ *ver:* 0.6
  |_ *key param:* NA
- **FOAM pipeline**
  |_ *ver:* 0.5
  |_ *key param:* NA
- **mitos**
  |_ *ver:* 2.1.3
  |_ *key param:* NA

Curation pipeline

- **PretextViewAI**
  |_ *ver:* 1.0.0
  |_ *key param:* NA

Submitter: Fernando Cruz

Affiliation: CNAG

Date and time: 2025-03-27 11:41:04 CET