

ERGA Assembly Report

v24.10.15

Tags: ERGA-BGE

TxID	87964
ToLID	xgPatRust1
Species	Patella rustica
Class	Gastropoda
Order	NA

Genome Traits	Expected	Observed
Haploid size (bp)	725,915,189	719,386,390
Haploid Number	9 (source: ancestor)	9
Ploidy	2 (source: ancestor)	2
Sample Sex	female	female

EBP metrics summary and curation notes

Obtained EBP quality metric for collapsed: 6.7.Q44

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

- . Kmer completeness value is less than 90 for collapsed

Curator notes

- . Interventions/Gb: 40
- . Contamination notes: "During the curation process we pre-labeled eight small scaffolds as contaminants based on their much lower ONT coverage than observed across the rest of the assembly. We subsequently analyzed these scaffolds using blobtools and confirmed that these contaminants all matched Spirochaetota. Spirocheta have previously been found in bivalves but little is known about the characteristics of these associations (commensalism or parasitism). These eight scaffolds were not re-mapped with HiC data and are absent from the for-review pretext file. They were also removed from the final assembly fasta. Moreover, running blobtools (using species buscoregions) on the curated assembly allowed us to identify six additional small scaffolds (scaffold_39, scaffold_42, scaffold_43, scaffold_48, scaffold_31 & scaffold_50) that contain fish sequence-contaminants. One of the unlocs and another scaffold also contained short matches to fish sequences these seemed to be false positive."
- . Other observations: "The curation was generally not too difficult except that several of the differet centromere regions required some level of modification. We also found two haplotypic duplications that were removed from the pretext map sent for reviewing. The organism was determined to be a female but we are not sure of the sex determination system of this species. The curated assembly stats (gfastas, busco

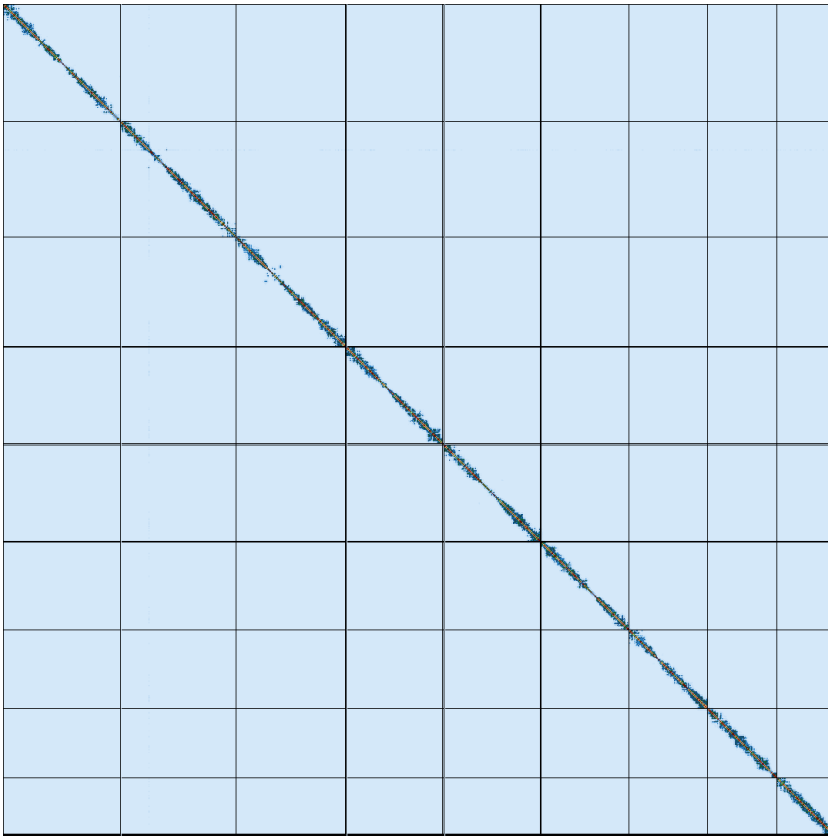
and merqury) and the blobplot were generated from the 24-sequence assembly after filtering out for the scaffold-containing contaminants. The pretext does contain the six scaffolds with fish DNA."

Quality metrics table

Metrics	Pre-curation collapsed	Curated collapsed
Total bp	723,798,427	719,386,390
GC %	36.42	36.4
Gaps/Gbp	165.79	184.88
Total gap bp	24,000	26,600
Scaffolds	52	24
Scaffold N50	83,441,781	83,582,575
Scaffold L50	4	4
Scaffold L90	8	8
Contigs	172	157
Contig N50	8,438,000	8,519,583
Contig L50	23	22
Contig L90	79	79
QV	44.057	44.0838
Kmer compl.	82.3326	81.9078
BUSCO sing.	97.0%	96.9%
BUSCO dupl.	0.3%	0.3%
BUSCO frag.	1.3%	1.3%
BUSCO miss.	1.4%	1.5%

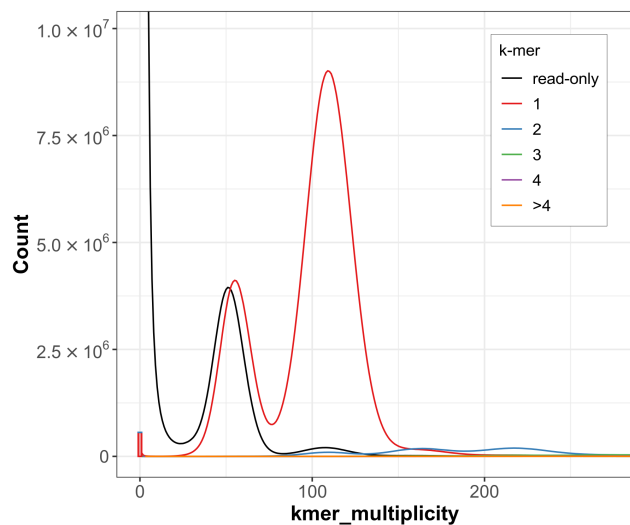
BUSCO: 5.4.0 (euk_genome_met, metaeuk) / Lineage: metazoa_odb10 (genomes:65, BUSCOs:954)

HiC contact map of curated assembly

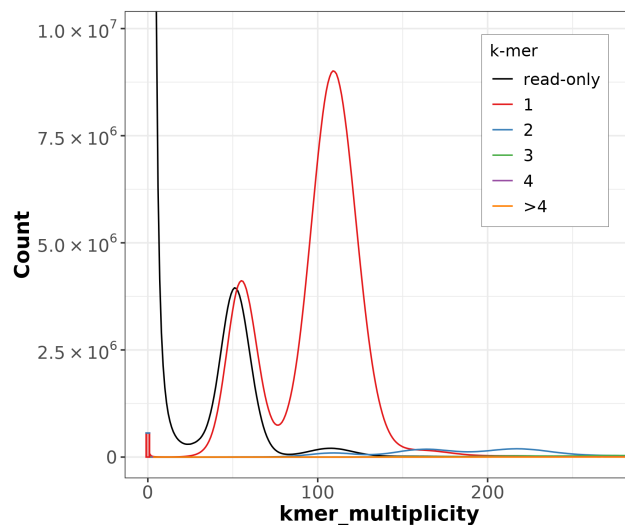


collapsed [\[LINK\]](#)

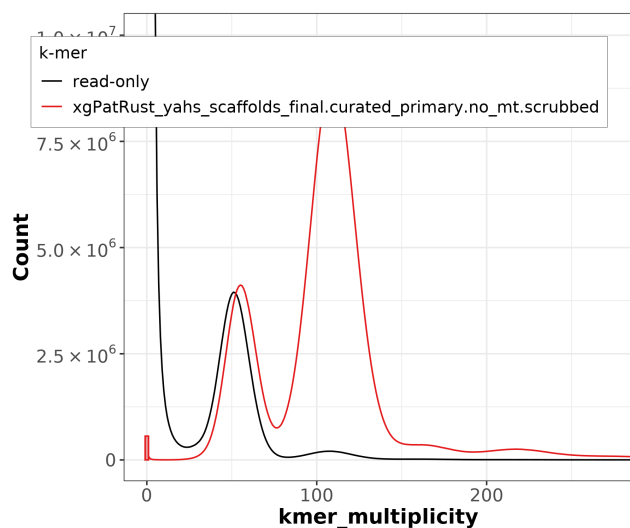
K-mer spectra of curated assembly



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

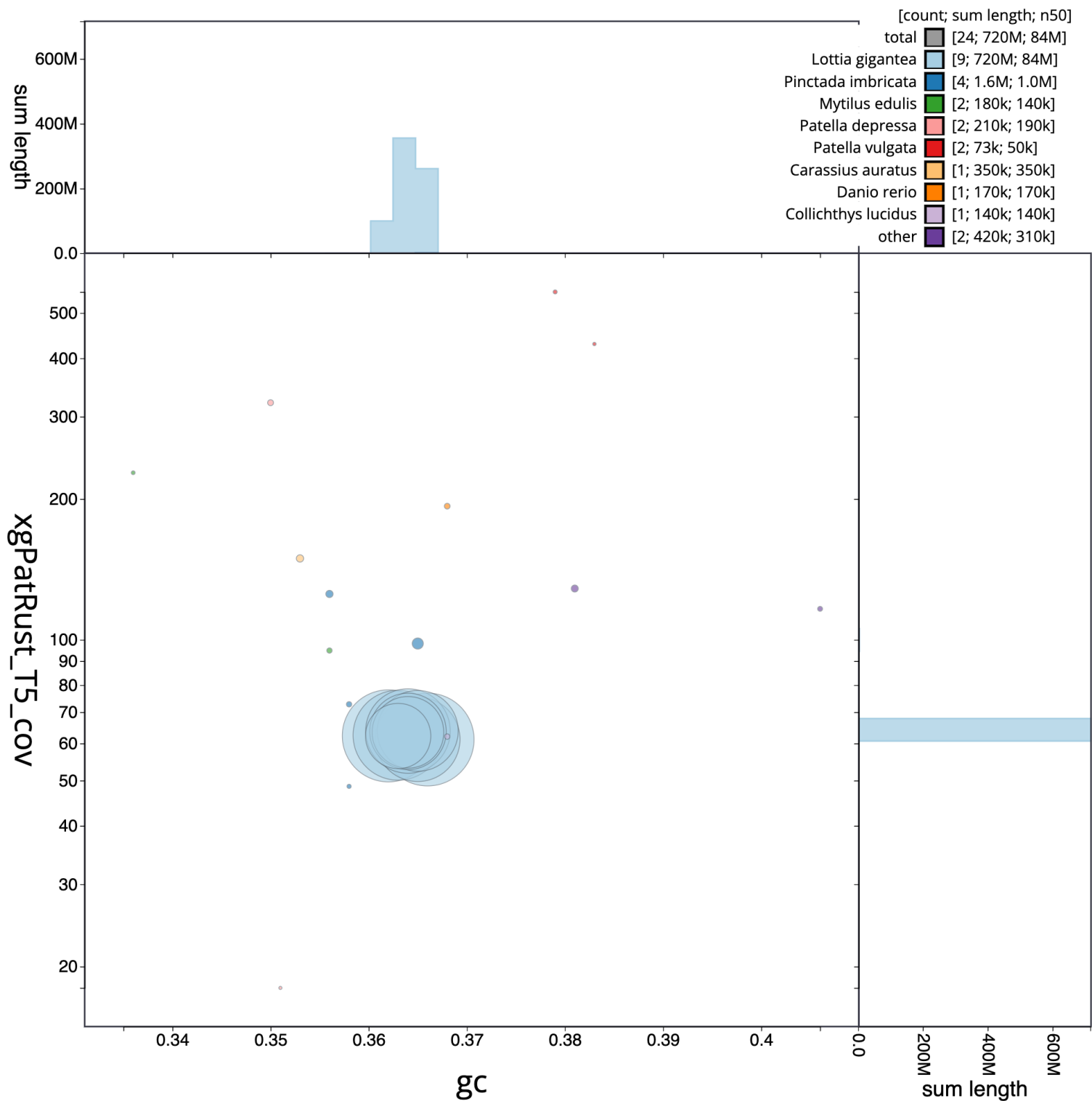


Distribution of k-mer counts per copy numbers found in asm



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

Post-curation contamination screening



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	59x	65x	147x

Assembly pipeline

- **Trim_galore**
 - |_ *ver*: 0.6.7
 - |_ *key param*: --gzip
 - |_ *key param*: =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
- **hypo**
 - |_ *ver*: 1.0.3
 - |_ *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
- **GRIT_Rapid**
 - |_ *ver*: 2.0
 - |_ *key param*: NA

Submitter: Francisco Camara

Affiliation: CNAG Barcelona

Date and time: 2024-12-12 15:03:31 CET