ERGA Assembly Report

v24.10.15

Tags: ATLASea[INVALID TAG]

TxID	569429
ToLID	**kaPhaNigr1**
Species	Phallusia nigra
Class	Ascidiacea
Order	Phlebobranchia

Genome Traits	Expected	Observed
Haploid size (bp)	44,243,645	162,311,134
Haploid Number	9 (source: ancestor)	10
Ploidy	2 (source: ancestor)	2
Sample Sex	Unknown	Unknown

EBP metrics summary and curation notes

Obtained EBP quality metric for collapsed: 7.7.Q49

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

. Observed Haploid size (bp) has >20% difference with Expected
. Observed Haploid Number is different from Expected
. Kmer completeness value is less than 90 for collapsed

Curator notes

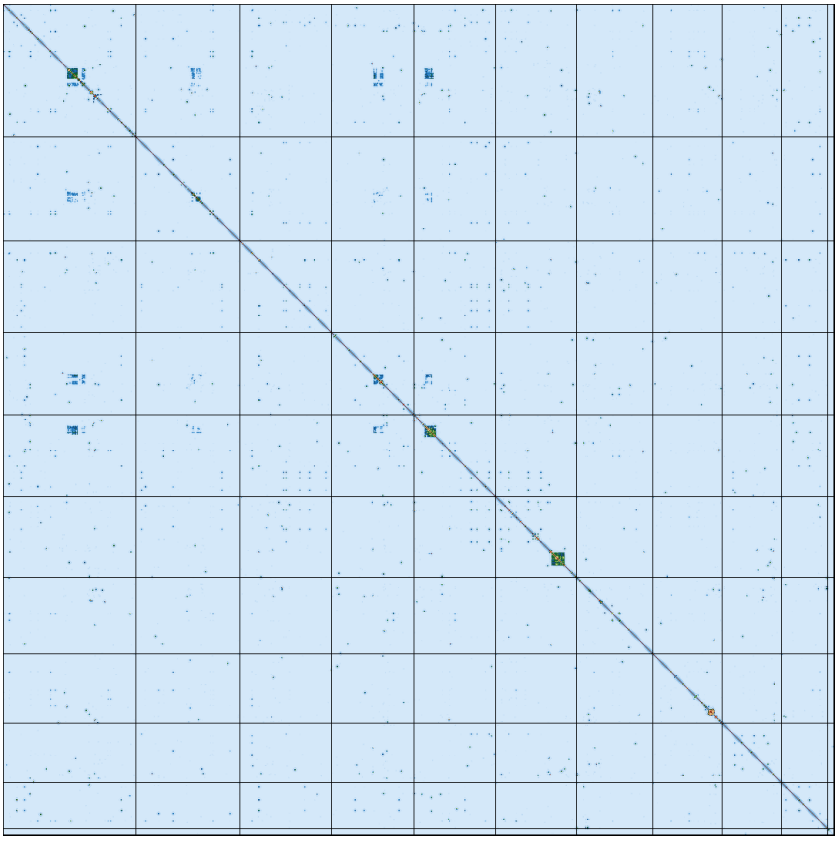
. Interventions/Gb: 137
. Contamination notes: ""
. Other observations: "The assembly of Phallusia nigra (kaPhaNigr1) is based on 210.86X PacBio data and 277.36X Arima Hi-C data generated as part of the ATLASea programme (https://www.atlasea.fr). The assembly process included the following steps: initial PacBio assembly generation with Hifiasm, removal of contaminant sequences using Context, removal of haplotypic duplications using purge_dups, and Hi-C-based scaffolding with YaHS. In total, 703 contigs were identified as contaminants (bacterial, archaeal, or viral), totaling 31.307 Mb (with the largest being 5.159 Mb). Additionally, 262 regions totaling 7.332 Mb (with the largest being 0.699 Mb) were identified as haplotypic duplications and removed. The mitochondrial genome was assembled using Oatk and shows an unusual size compared to currently available reference. Finally, the primary assembly was analyzed and manually improved using Pretext. Chromosome-scale scaffolds confirmed by Hi-C data were named in order of size. "

Quality metrics table

Metrics	Pre-curation collapsed	Curated collapsed
Total bp	161,417,052	162,311,134
GC %	38.53	38.5
Gaps/Gbp	49.56	110.9
Total gap bp	800	3,300
Scaffolds	43	27
Scaffold N50	14,832,700	15,952,774
Scaffold L50	4	5
Scaffold L90	9	9
Contigs	51	45
Contig N50	11,565,000	11,565,000
Contig L50	6	6
Contig L90	14	14
QV	48.2101	49.2301
Kmer compl.	9.41653	9.49282
BUSCO sing.	95.5%	95.5%
BUSCO dupl.	0.6%	0.8%
BUSCO frag.	0.6%	0.6%
BUSCO miss.	3.2%	3.1%

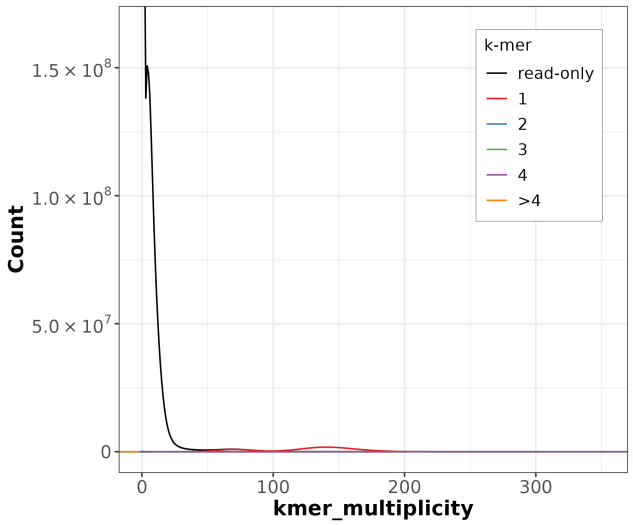
BUSCO: 6.1.0 (euk_genome_min, miniprot) / Lineage: metazoa_odb12.2 (genomes:182, BUSCOs:932)

HiC contact map of curated assembly

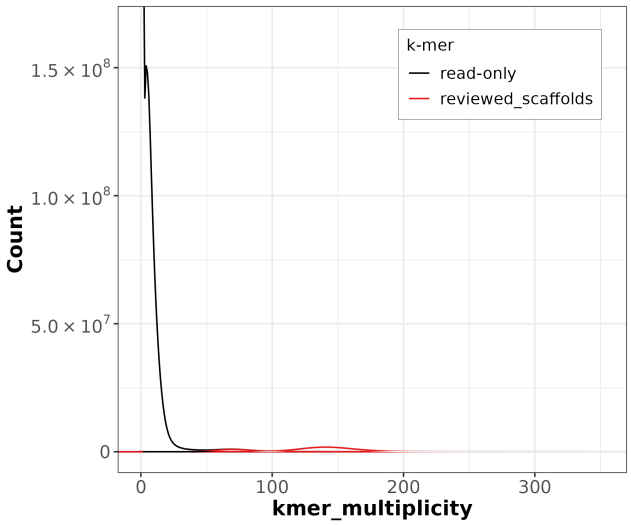


collapsed [LINK]

K-mer spectra of curated assembly

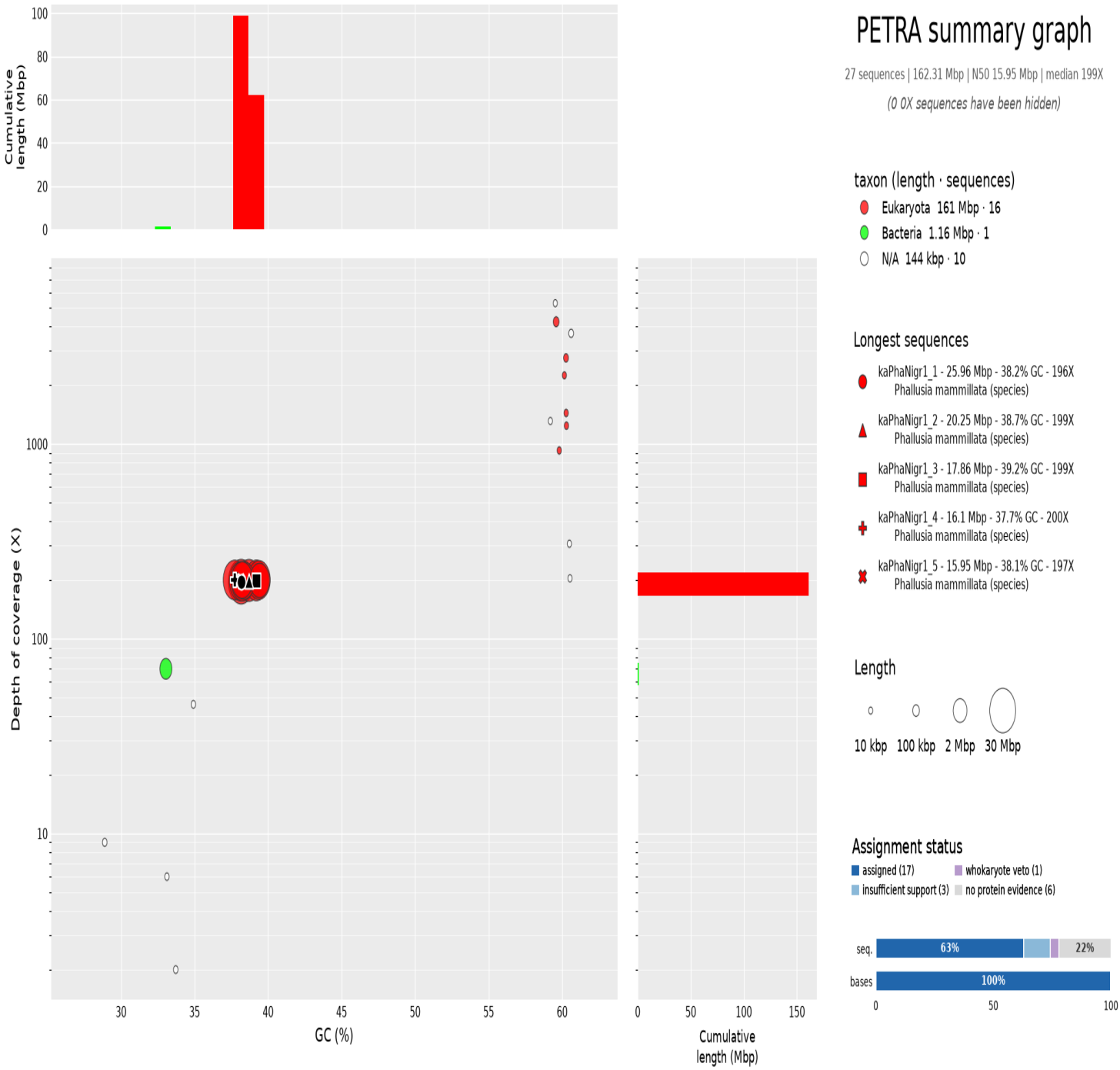


Distribution of k-mer counts per copy numbers found in asm



Distribution of k-mer counts 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	Long reads	Arima
Coverage	209	275

Assembly pipeline

- **Hifiasm**
  |_ *ver:* 0.19.5-r593
  |_ *key param:* NA
- **purge_dups**
  |_ *ver:* 1.2.5
  |_ *key param:* NA
- **YaHS**
  |_ *ver:* 1.2
  |_ *key param:* NA

Curation pipeline

- **PretextMap**
  |_ *ver:* 0.2.4
  |_ *key param:* NA
- **PretextView**
  |_ *ver:* 1.0.7
  |_ *key param:* NA

Submitter: Emilie Teodori
Affiliation: Genoscope

Date and time: 2026-10-19 22:36:10 CEST