ERGA Assembly Report

v24.10.15

Tags: ATLASea[INVALID TAG]

TxID	1619589
ToLID	**wpHarFras1**
Species	Harmothoe fraserthomsoni
Class	Polychaeta
Order	Phyllodocida

Genome Traits	Expected	Observed
Haploid size (bp)	1,777,913,986	1,932,930,702
Haploid Number	10 (source: ancestor)	19
Ploidy	2 (source: ancestor)	2
Sample Sex	Unknown	Unknown

EBP metrics summary and curation notes

Obtained EBP quality metric for collapsed: 6.7.Q48

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 collapsed
. Assembly length loss > 3% for collapsed

Curator notes

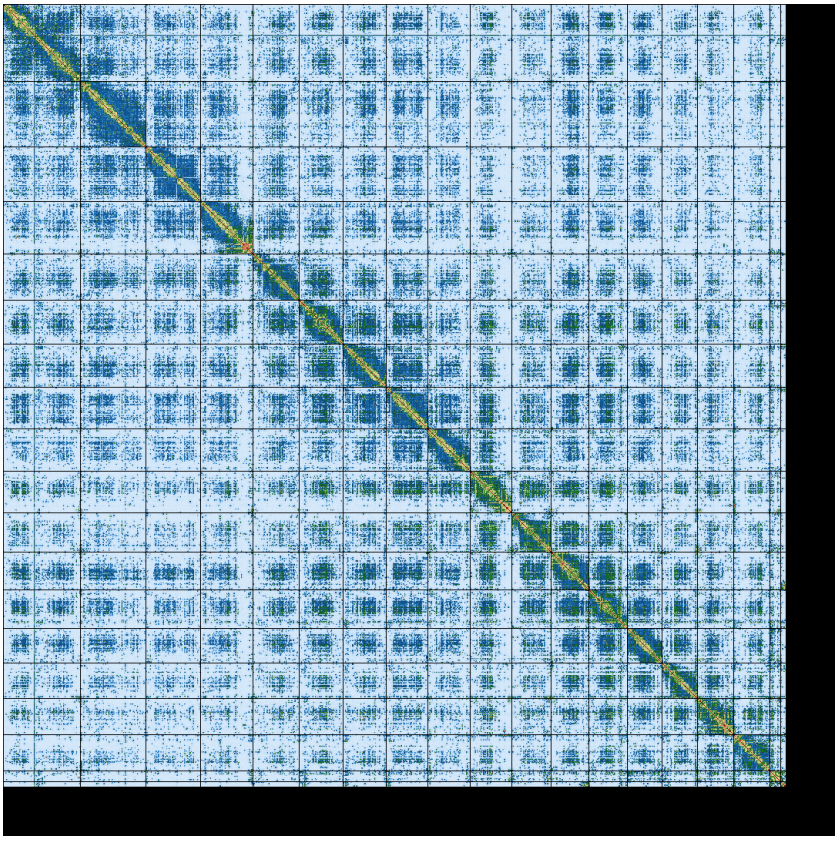
. Interventions/Gb: 16
. Contamination notes: ""
. Other observations: "The assembly of Harmothoe fraserthomsoni (wpHarFras1) is based on 39X PacBio data and 158X Arima Hi-C data generated as part of the ATLASea programme (https://www.atlasea.fr). The assembly process included the following steps: initial PacBio+HiC assembly generation with Hifiasm, removal of contaminant sequences using Context, and Hi-C-based scaffolding with YaHS of each haplotype. In total, 119 contigs were identified as contaminants (bacterial, archaeal, or viral), totaling 4.18 Mb (with the largest being 0.346 Mb). Mitochondrial genome was assembled using OATK. Finally, Both haplotypes were analyzed and manually improved using Pretext. During manual curation, 11 haplotypic regions and 35 contaminant sequences were removed, totaling 11.495Mb and 2.056Mb respectively (with the largest being 2.841Mb and 0.225Mb). The telomeric pattern TTAGGG was identified with TelFinder and used to generate the Pretext Telomeres tracks. 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	2,133,482,091	1,932,930,702
GC %	42.84	42.81
Gaps/Gbp	187.02	368.35
Total gap bp	39,900	84,300
Scaffolds	316	1,512
Scaffold N50	99,211,203	98,078,955
Scaffold L50	9	8
Scaffold L90	20	17
Contigs	715	2,224
Contig N50	9,469,194	5,756,000
Contig L50	58	92
Contig L90	238	416
QV	49.1418	48.4807
Kmer compl.	75.1551	68.7112
BUSCO sing.	87.3%	94.9%
BUSCO dupl.	11.7%	1.8%
BUSCO frag.	0.2%	2.7%
BUSCO miss.	0.7%	0.6%

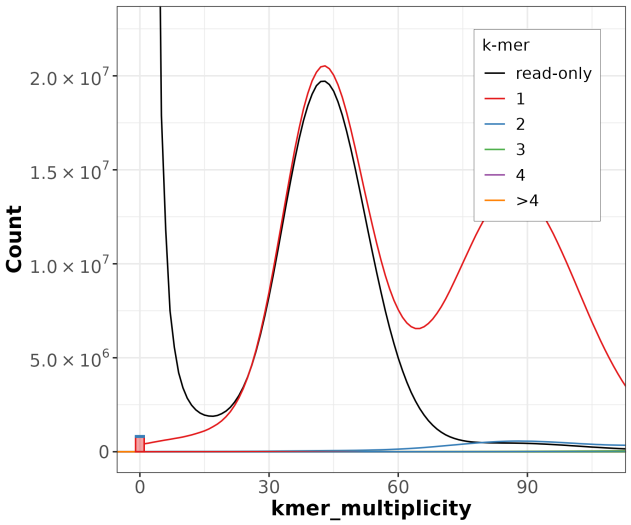
Warning! BUSCO versions or lineage datasets are not the same across results:
BUSCO: 6.0.0 (euk_genome_min, miniprot) / Lineage: lophotrochozoa_odb12 (genomes:75, BUSCOs:1252)
BUSCO: 5.8.2 (euk_genome_min, miniprot) / Lineage: lophotrochozoa_odb12 (genomes:75, BUSCOs:1252)

HiC contact map of curated assembly

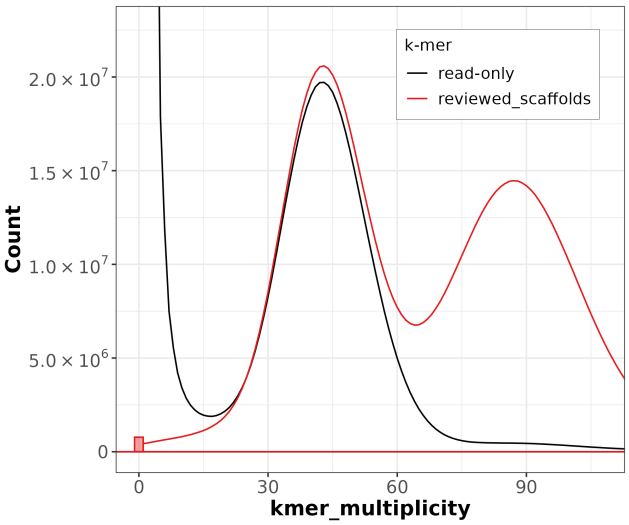


collapsed [LINK]

K-mer spectra of curated assembly

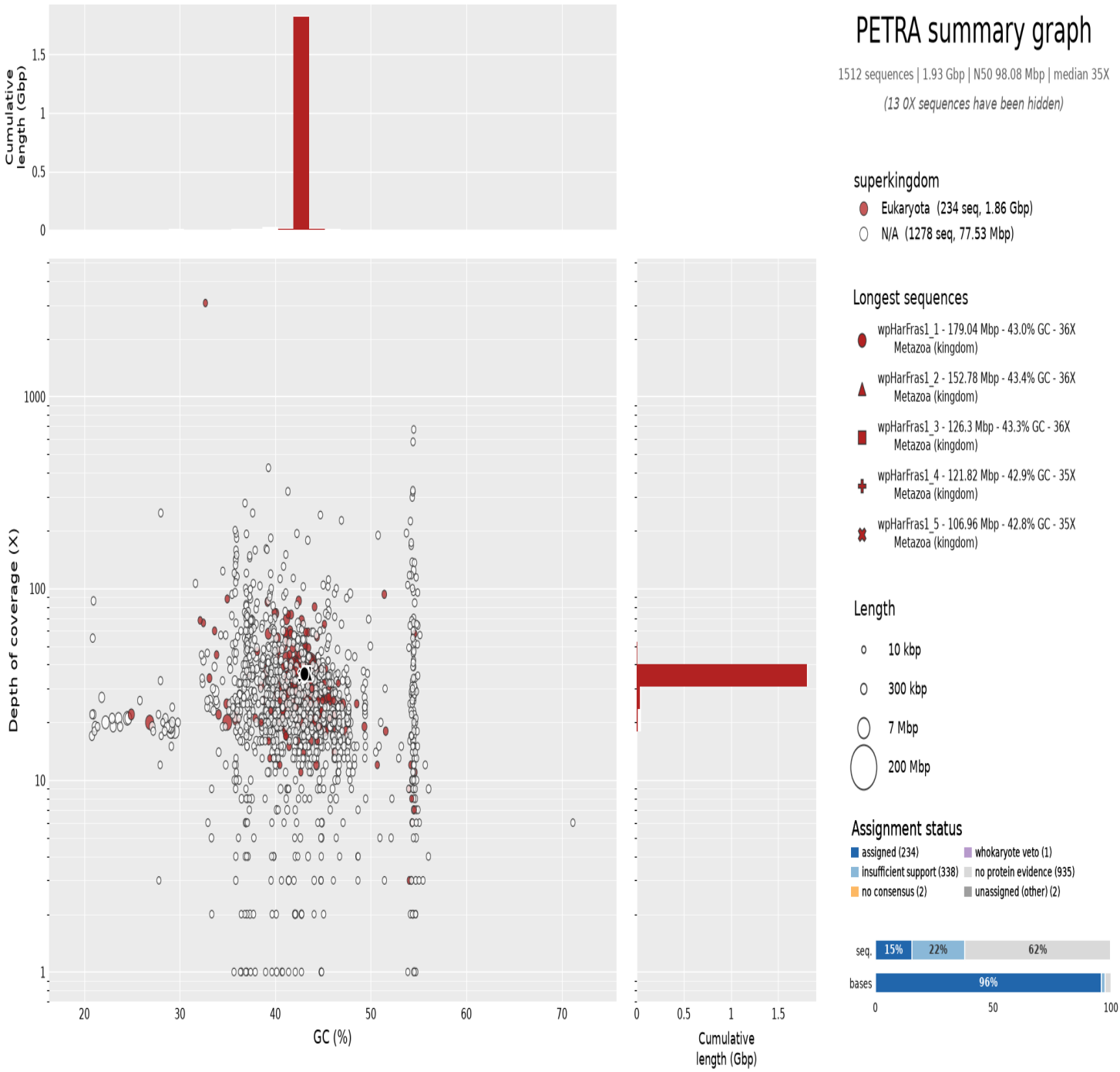


Distribution of k-mer counts per copy numbers found in asm



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	43	174

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: Sophie Layac
Affiliation: Genoscope

Date and time: 2026-09-24 02:23:05 CEST