ERGA Assembly Report

v24.10.15

Tags: ATLASea[INVALID TAG]

TxID	2029591
ToLID	**wpLaeHyst1**
Species	Laetmonice hystrix
Class	Polychaeta
Order	Phyllodocida

Genome Traits	Expected	Observed
Haploid size (bp)	584,925,931	627,452,951
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.Q62

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

. Observed Haploid Number is different from Expected

Curator notes

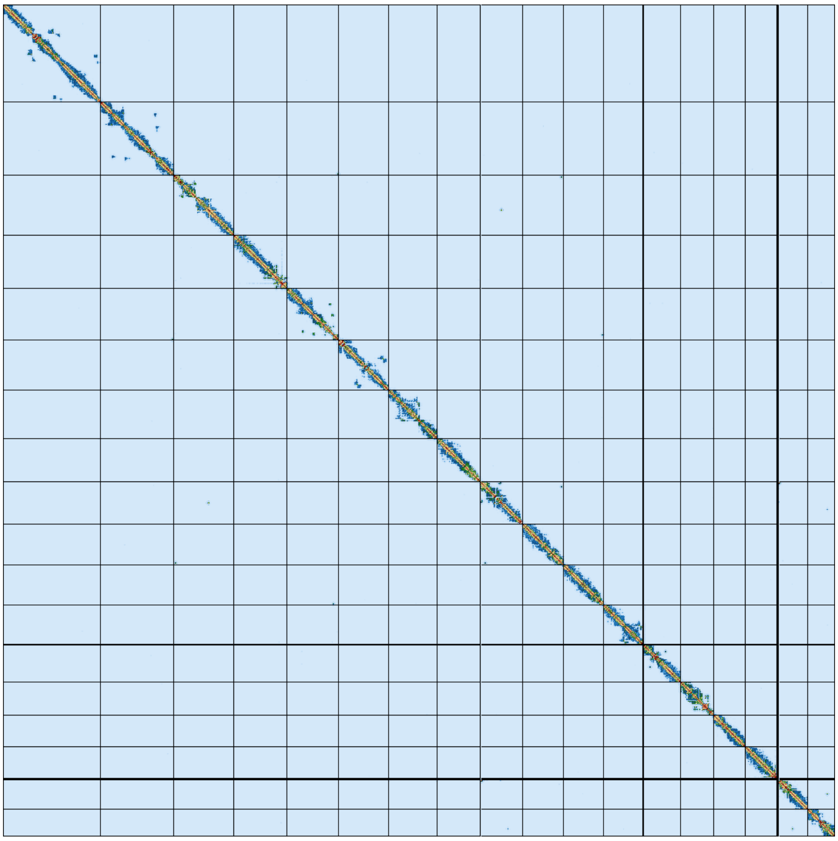
. Interventions/Gb: 98
. Contamination notes: ""
. Other observations: "The assembly of Laetmonice hystrix (wpLaeHyst1) is based on 65X of PacBio data and Arima Hi-C data generated as part of the ATLASea programme (https://www.atlasea.fr). The assembly process included the following steps: initial PacBio 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, 164 contigs were identified as contaminants (bacterial, archaeal, or viral), totaling 11 Mb (with the largest being 0,7 Mb). Additionally, 117 regions totaling 63 Mb were identified as haplotypic duplications and removed (with the largest being 8 Mb). The mitochondrial genome was assembled using OATK. Finally, the primary assembly was analyzed and manually improved using Pretext. During manual curation, 15 haplotypic regions and 2 contaminant sequences were removed, totaling 10 Mb (with the largest being 1.7 Mb). 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	637,599,419	627,452,951
GC %	44.87	44.88
Gaps/Gbp	125.47	162.56
Total gap bp	8,000	13,700
Scaffolds	57	34
Scaffold N50	35,836,971	36,899,702
Scaffold L50	7	7
Scaffold L90	16	16
Contigs	137	136
Contig N50	9,922,540	9,922,540
Contig L50	17	17
Contig L90	62	58
QV	45.7131	62.2514
Kmer compl.	71.5953	71.3931
BUSCO sing.	94.1%	
BUSCO dupl.	2.1%	
BUSCO frag.	2.1%	
BUSCO miss.	1.7%	

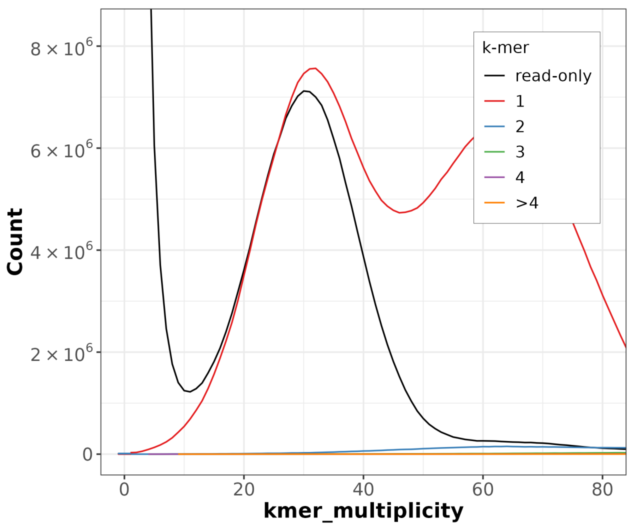
BUSCO: 5.4.3 (euk_genome_met, metaeuk) / Lineage: metazoa_odb10 (genomes:65, BUSCOs:954)

HiC contact map of curated assembly

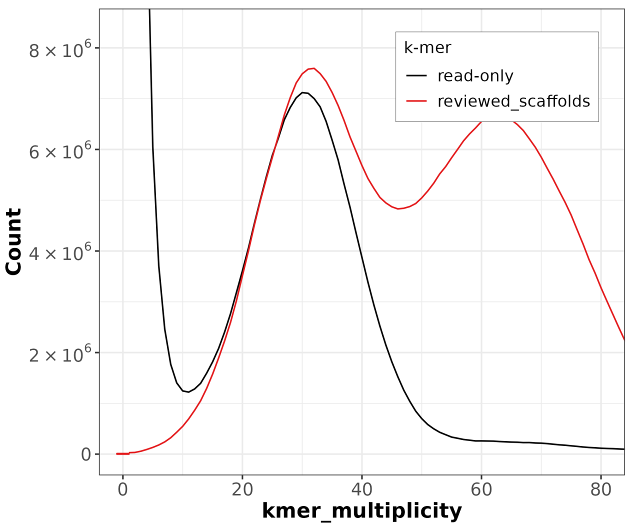


collapsed [LINK]

K-mer spectra of curated assembly

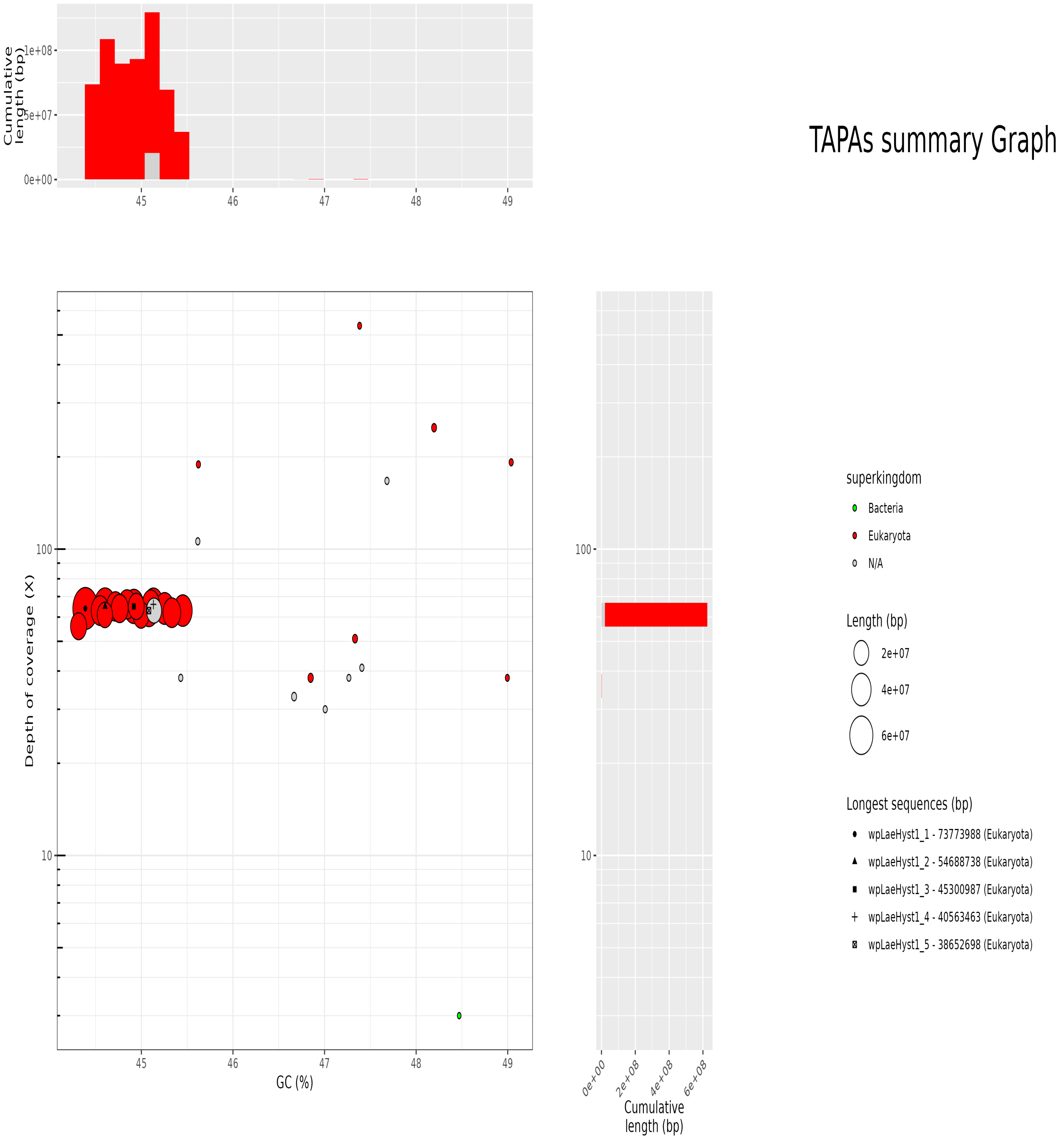


Distribution of k-mer counts per copy numbers found in asm



Distribution of k-mer counts 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	PACBIO Hifi	Arima
Coverage	64	61

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:* 0.1.9
  |_ *key param:* NA
- **PretextView**
  |_ *ver:* 0.2.5
  |_ *key param:* NA

Submitter: Arnaud Couloux
Affiliation: Genoscope

Date and time: 2025-03-25 10:48:17 CET