

ERGA Assembly Report

v24.10.15

Tags: ATLASEa[INVALID TAG]

TxID	3465829
ToLID	wpHarSpea1
Species	Harmothoe sp. A-3ZBEU0YED
Class	Polychaeta
Order	Phyllodocida

Genome Traits	Expected	Observed
Haploid size (bp)	2,326,937,269	2,358,279,275
Haploid Number	10 (source: ancestor)	18
Ploidy	2 (source: ancestor)	2
Sample Sex	Unknown	Unknown

EBP metrics summary and curation notes

Obtained EBP quality metric for collapsed: 6.8.Q47

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

Curator notes

. Interventions/Gb: 162
. Contamination notes: ""
. Other observations: "The assembly of Harmothoe (wpHarSpea1) is based on 43X PacBio data and 195X 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, 97 contigs were identified as contaminants (bacterial, archaeal, or viral), totaling 4.16 Mb (with the largest being 0.89 Mb). Additionally, 1546 regions totaling 284.881 Mb (with the largest being 3.654 Mb) were identified as haplotypic duplications and removed. The mitochondrial genome was assembled using OATK. During manual curation, 8 haplotypic regions were removed, totaling 9.3Mb (with the largest being 2.9Mb). Chromosome-scale scaffolds confirmed by Hi-C data were named in order of size. "

Quality metrics table

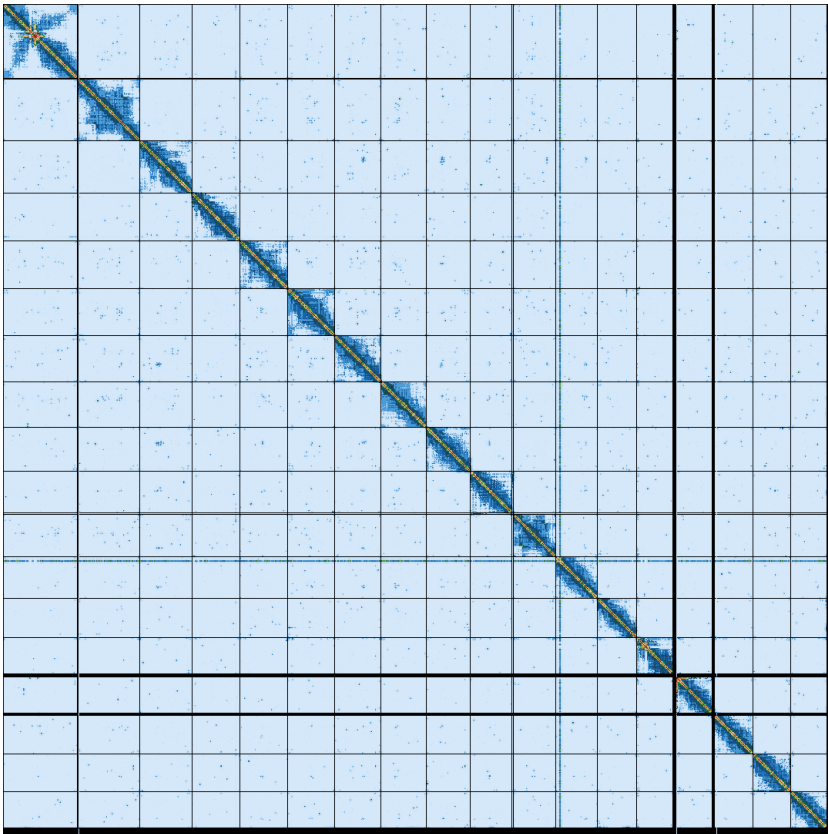
Metrics	Pre-curation collapsed	Curated collapsed
Total bp	2,375,557,729	2,358,279,275
GC %	42.97	42.97
Gaps/Gbp	551.87	549.98
Total gap bp	154,300	162,800
Scaffolds	362	256
Scaffold N50	125,916,576	127,556,776
Scaffold L50	8	8
Scaffold L90	17	17
Contigs	1,614	1,553
Contig N50	3,962,140	3,964,000
Contig L50	165	163
Contig L90	632	626
QV	47.6382	47.657
Kmer compl.	79.354	79.1459
BUSCO sing.	89.5%	97.3%
BUSCO dupl.	1.1%	1.1%
BUSCO frag.	7.3%	0.6%
BUSCO miss.	2.1%	1.0%

Warning! BUSCO versions or lineage datasets are not the same across results:

BUSCO: 5.8.2 (euk_genome_met, metaeuk) / Lineage: lophotrochozoa_odb12 (genomes:75, BUSCOs:1252)

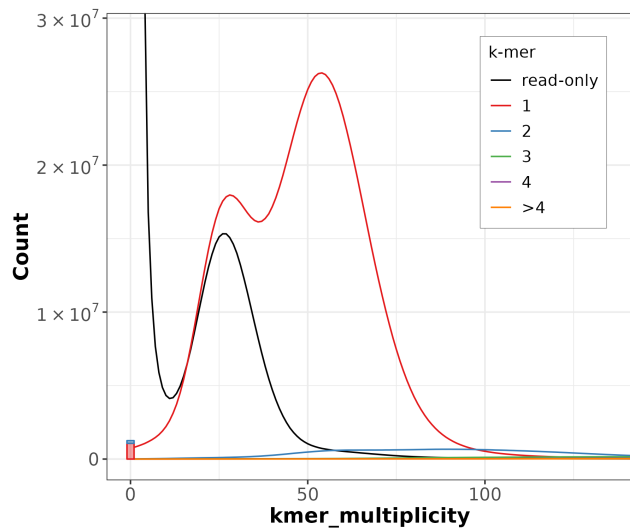
BUSCO: 6.0.0 (euk_genome_min, miniprot) / Lineage: lophotrochozoa_odb12 (genomes:75, BUSCOs:1252)

HiC contact map of curated assembly

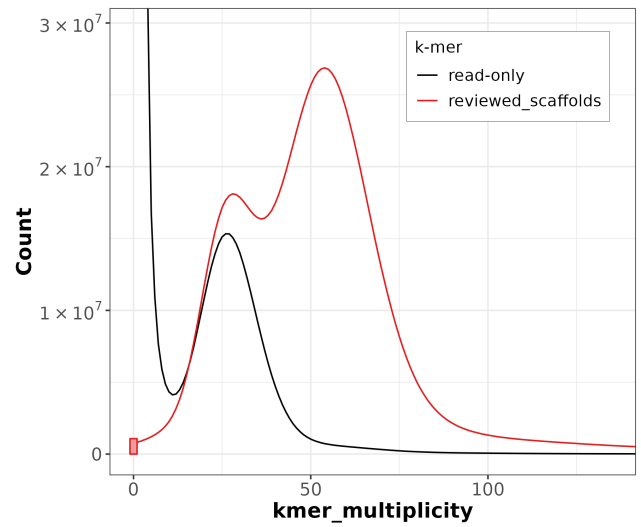


collapsed [\[LINK\]](#)

K-mer spectra of curated assembly

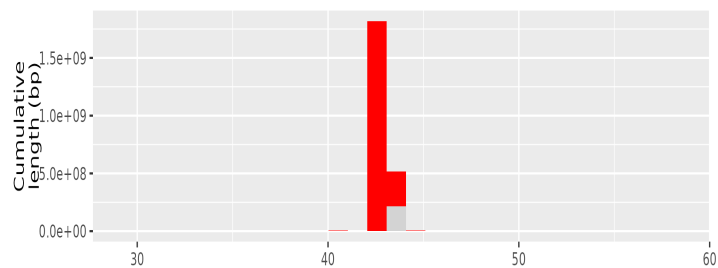


Distribution of k-mer counts per copy numbers found in asm

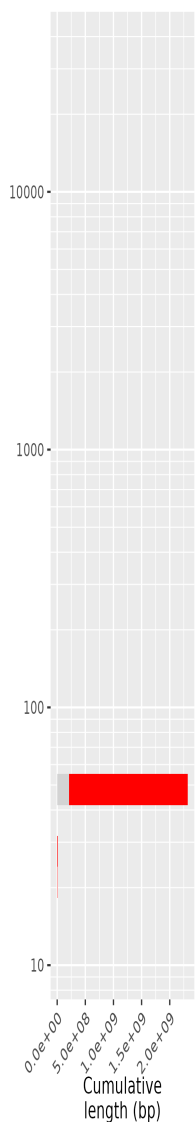
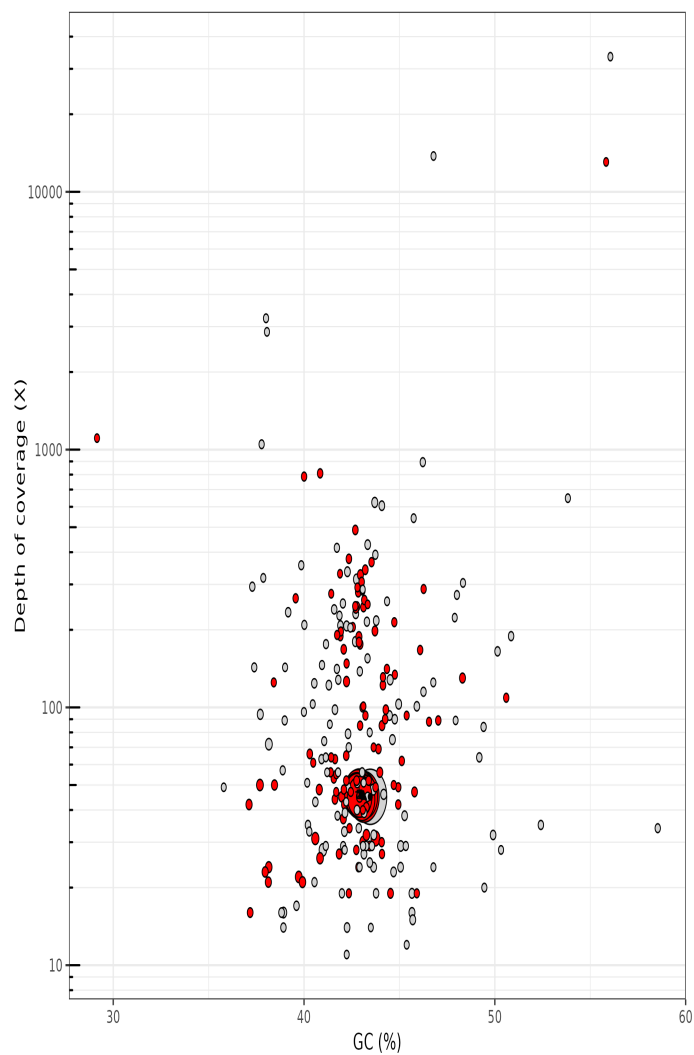


Distribution of k-mer counts coloured by their presence in reads/assemblies

Post-curation contamination screening



TAPAs summary Graph



Longest sequences (bp)

- SUPER_1 - 210973181 (N/A)
- ▲ SUPER_2 - 174251293 (Eukaryota)
- SUPER_3 - 147890429 (Eukaryota)
- + SUPER_4 - 136099631 (Eukaryota)
- ⊠ SUPER_5 - 135570896 (Eukaryota)

superkingdom

- Eukaryota
- N/A

Length (bp)

- 5.0e+07
- 1.0e+08
- 1.5e+08
- 2.0e+08

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	195

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: Benjamin Istace

Affiliation: Genoscope

Date and time: 2025-10-14 00:39:43 CEST