

ERGA Assembly Report

v24.10.15

Tags: ATLASEa[INVALID TAG]

TxID	3465858
ToLID	odHexSpea1
Species	Hexadella sp. A-XEH1ECDDD
Class	Demospongiae
Order	Verongiida

Genome Traits	Expected	Observed
Haploid size (bp)	248,983,316	162,810,289
Haploid Number	5 (source: ancestor)	22
Ploidy	2 (source: ancestor)	2
Sample Sex	Unknown	Unknown

EBP metrics summary and curation notes

Obtained EBP quality metric for collapsed: 6.6.Q43

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
- . BUSCO single copy value is less than 90% for collapsed
- . Assembly length loss > 3% for collapsed

Curator notes

- . Interventions/Gb: 123
- . Contamination notes: ""
- . Other observations: "The assembly of Hexadella sp. (odHexSpea1) is based on 78X PacBio data and 168X 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 Hifiiasm, removal of contaminant sequences using Context, removal of haplotypic duplications using purge_dups, and Hi-C-based scaffolding with YaHS. In total, 665 contigs were identified as contaminants (bacterial, archaeal, or viral), totaling 54.316 Mb (with the largest being 2.528 Mb). Additionally, 242 regions totaling 10.145 Mb (with the largest being 0.248 Mb) were identified as haplotypic duplications and removed. The mitochondrial genome was assembled using OATK. Finally, the primary assembly was analyzed and manually improved using Pretext. During manual curation, 2 haplotypic regions and 113 contaminant sequences were removed, totaling 1.13Mb and 4.38Mb, respectively (with the largest being 0.91Mb and 0.20Mb). Chromosome-scale scaffolds confirmed by Hi-C

data were named in order of size. "

Quality metrics table

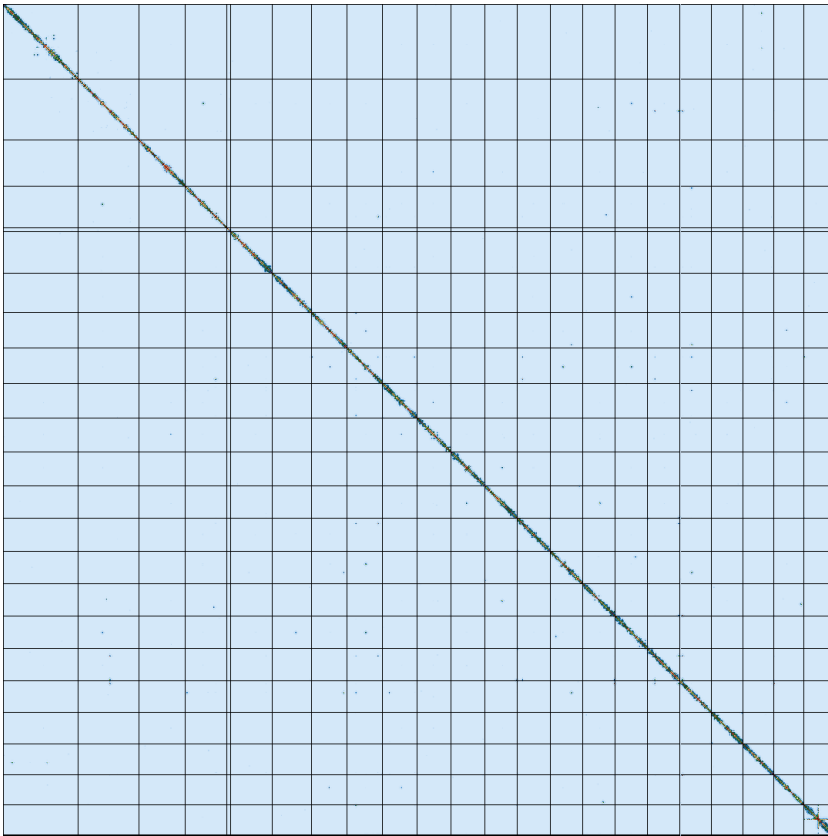
Metrics	Pre-curation collapsed	Curated collapsed
Total bp	168,373,564	162,810,289
GC %	46.77	46.65
Gaps/Gbp	605.8	595.79
Total gap bp	11,000	11,600
Scaffolds	162	29
Scaffold N50	6,850,060	6,585,021
Scaffold L50	10	10
Scaffold L90	21	20
Contigs	263	126
Contig N50	3,598,763	3,923,007
Contig L50	15	14
Contig L90	50	44
QV	37.2528	43.0412
Kmer compl.	70.0719	69.5691
BUSCO sing.	52.5%	82.0%
BUSCO dupl.	0.9%	0.7%
BUSCO frag.	28.0%	5.7%
BUSCO miss.	18.6%	11.6%

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

BUSCO: 5.8.2 (euk_genome_met, metaeuk) / Lineage: metazoa_odb12 (genomes:206, BUSCOs:672)

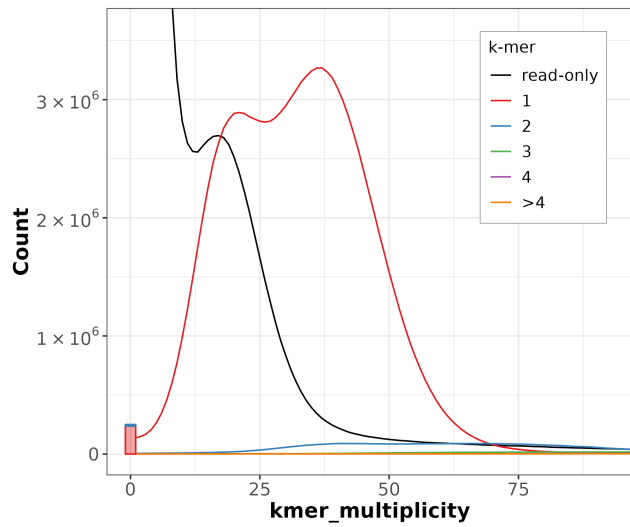
BUSCO: 6.0.0 (euk_genome_min, miniprot) / Lineage: metazoa_odb12 (genomes:206, BUSCOs:672)

HiC contact map of curated assembly

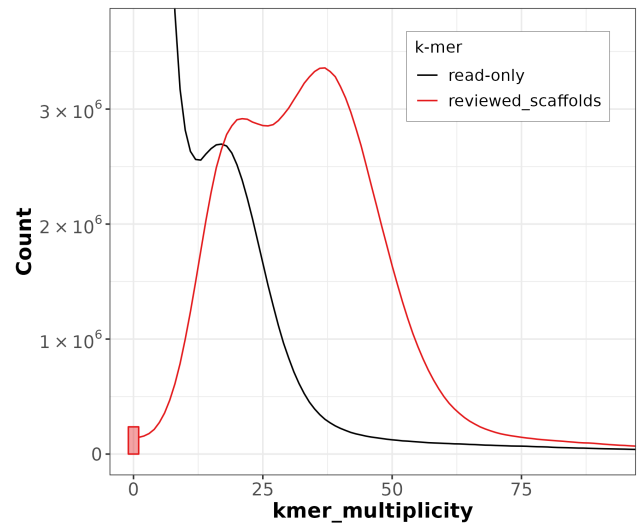


collapsed [\[LINK\]](#)

K-mer spectra of curated assembly

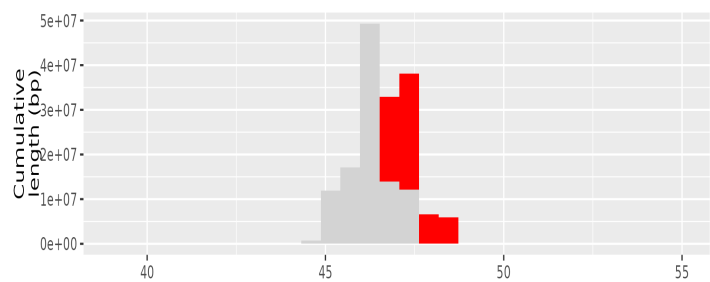


Distribution of k-mer counts per copy numbers found in asm

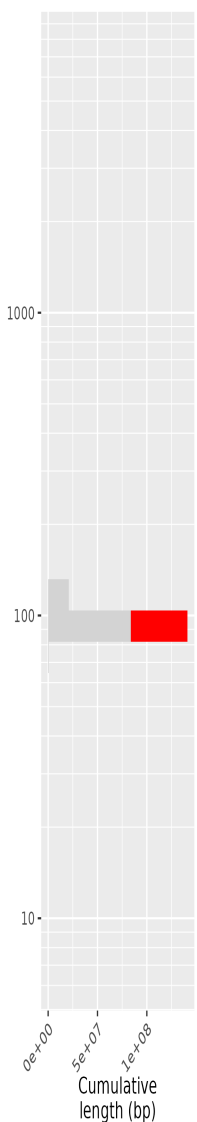
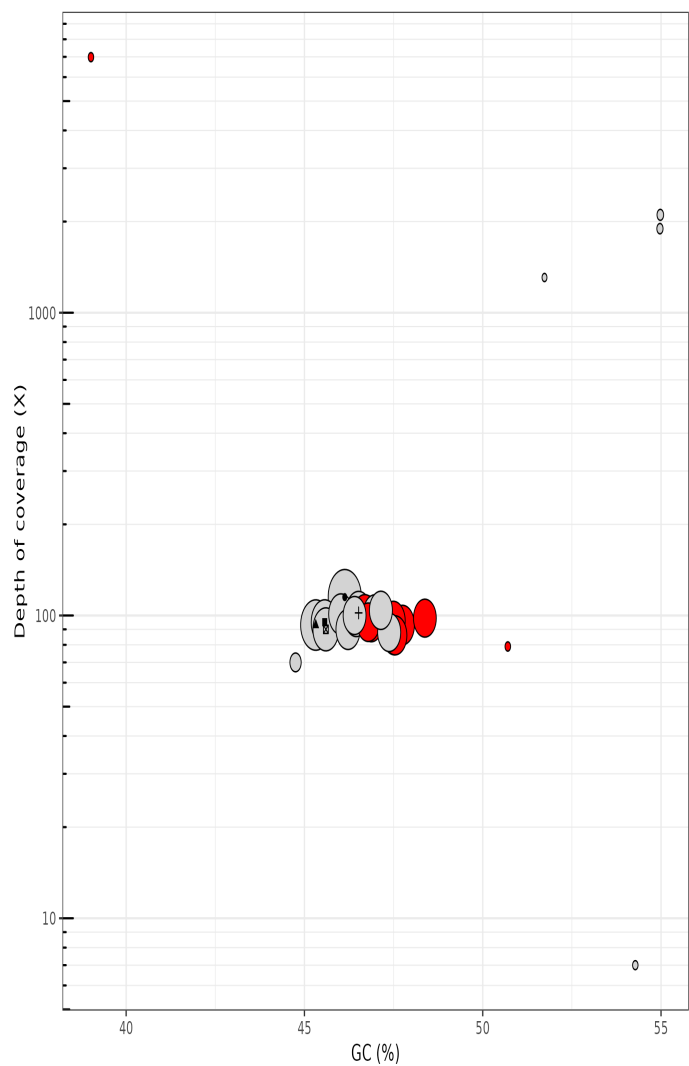


Distribution of k-mer counts coloured by their presence in reads/assemblies

Post-curation contamination screening



TAPAs summary Graph



- Longest sequences (bp)
- odHexSpea1_1 - 14750572 (N/A)
 - ▲ odHexSpea1_2 - 11893337 (N/A)
 - odHexSpea1_3 - 9009239 (N/A)
 - + odHexSpea1_5 - 8121852 (N/A)
 - ▣ odHexSpea1_4 - 8105055 (N/A)
- Length (bp)
- 5.0e+06
 - 1.0e+07
- superkingdom
- Eukaryota
 - N/A

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	77	167

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: Jean-Marc Aury

Affiliation: Genoscope

Date and time: 2025-10-13 06:08:38 CEST