ERGA Assembly Report

v24.10.15

Tags: ATLASea[INVALID TAG]

TxID	3465828
ToLID	**weOchSpea1**
Species	Ochetostoma sp. New Caledonia
Class	Polychaeta
Order	Echiuroidea

Genome Traits	Expected	Observed
Haploid size (bp)	1,345,912,772	1,370,353,031
Haploid Number	18 (source: ancestor)	17
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 Number is different from Expected

. Kmer completeness value is less than 90 for collapsed

Curator notes

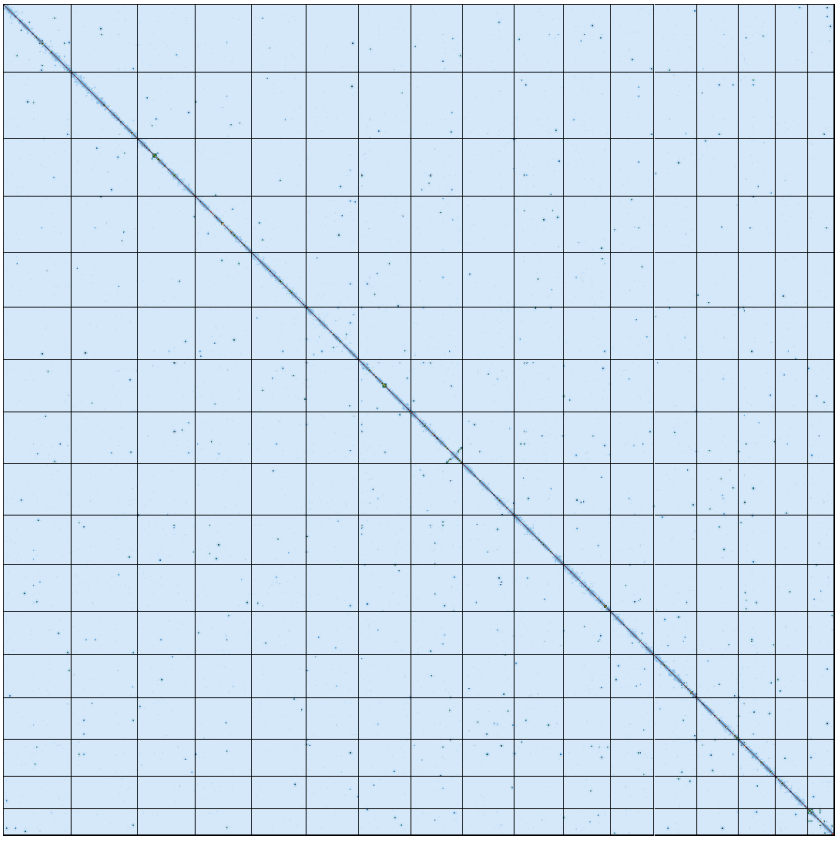
. Interventions/Gb: 6
. Contamination notes: ""
. Other observations: "The assembly of Ochetostoma sp. New Caledonia (weOchSpea1) is based on 56X PacBio data and 295X Arima Hi-C data generated as part of the ATLASea programme (https://www.atlasea.fr).The assembly process included the following steps: 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, 19 contigs were identified as contaminants (bacterial, archaeal, or viral), totaling 0.812 Mb (with the largest being 0.158 Mb). Additionally, 220 regions totaling 61 Mb (with the largest being 7.2 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, 1 supplementary haplotypic region was removed, totaling 1.07 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	1,371,418,013	1,370,353,031
GC %	37.59	37.59
Gaps/Gbp	123.96	47.43
Total gap bp	17,100	7,500
Scaffolds	731	44
Scaffold N50	81,360,609	85,549,285
Scaffold L50	8	8
Scaffold L90	16	15
Contigs	893	109
Contig N50	23,578,183	31,756,802
Contig L50	18	12
Contig L90	62	39
QV	49.3264	49.3282
Kmer compl.	65.8692	65.8463
BUSCO sing.	97.4%	97.5%
BUSCO dupl.	1.5%	1.6%
BUSCO frag.	0.6%	0.4%
BUSCO miss.	0.5%	0.5%

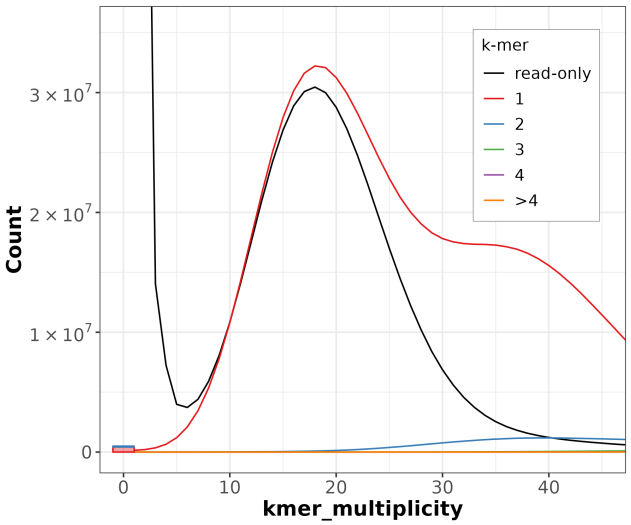
BUSCO: 6.0.0 (euk_genome_min, miniprot) / Lineage: lophotrochozoa_odb12 (genomes:75, BUSCOs:1252)

HiC contact map of curated assembly

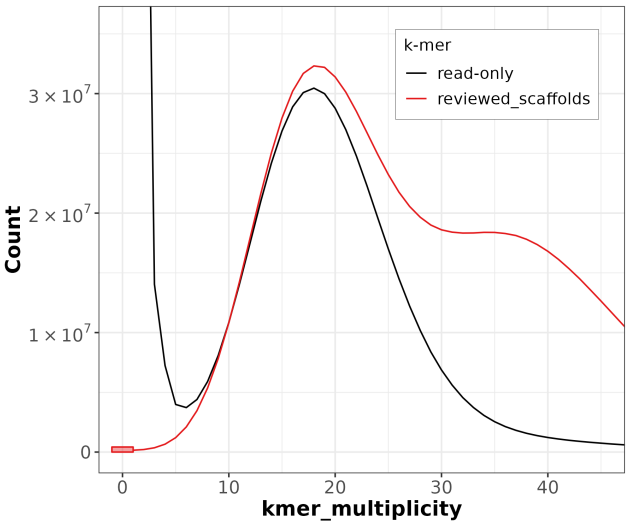


collapsed [LINK]

K-mer spectra of curated assembly

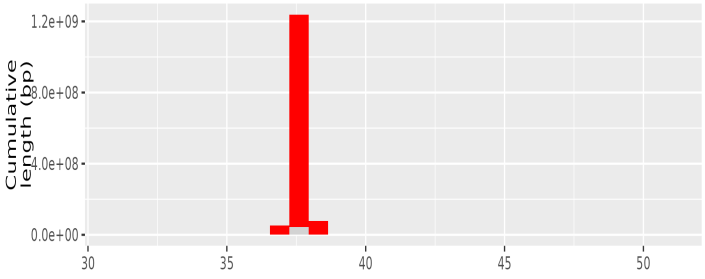


Distribution of k-mer counts per copy numbers found in asm

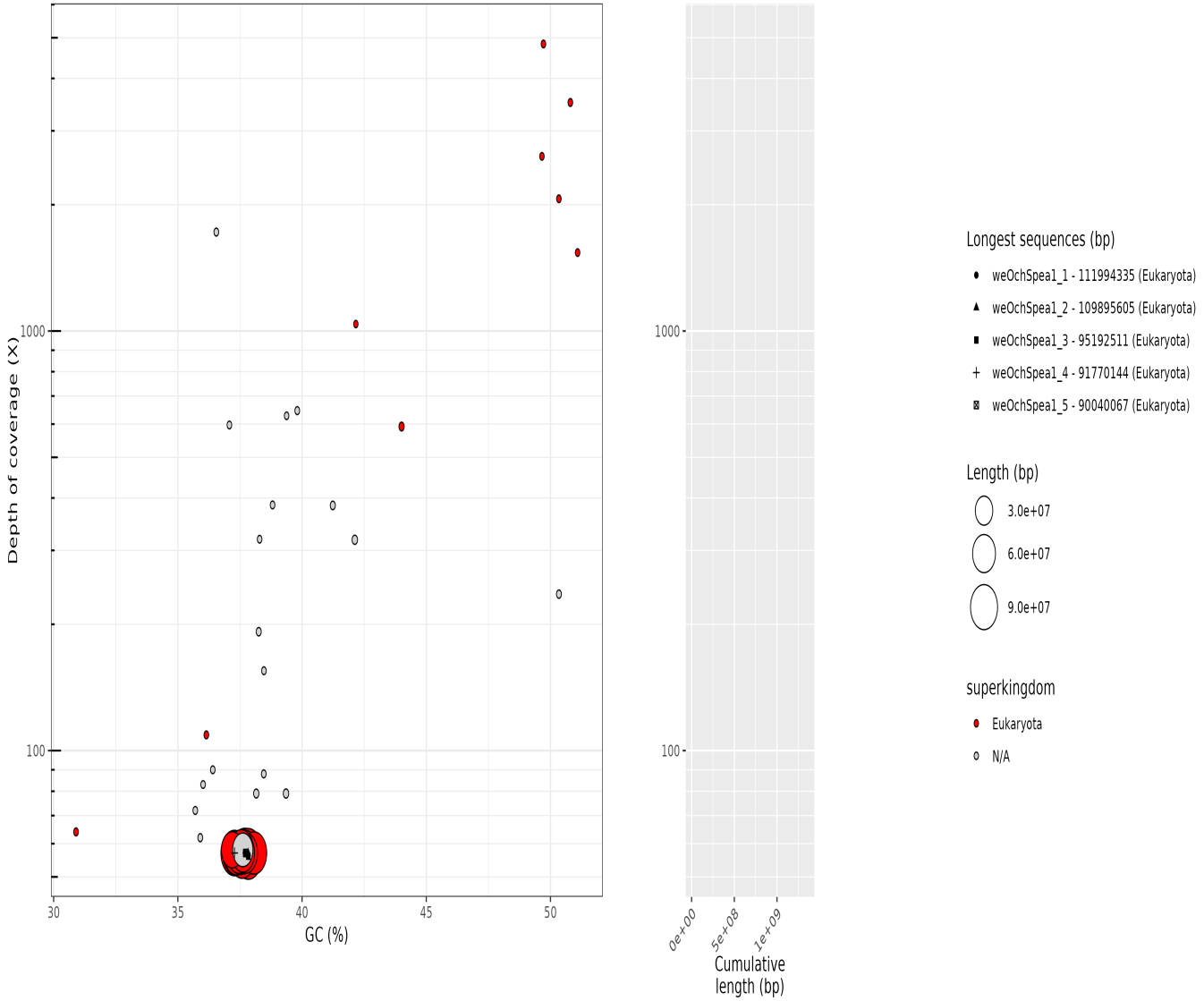


Distribution of k-mer counts coloured by their presence in reads/assemblies

Post-curation contamination screening



TAPAs summary Graph



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	56	295

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: Caroline Menguy
Affiliation: Genoscope

Date and time: 2025-11-04 23:48:29 CET