ERGA Assembly Report

v24.10.15

Tags: ATLASea[INVALID TAG]

TxID	3449502
ToLID	**wpIphFimb1**
Species	Iphione sp. A-MEVVRFRX5
Class	Polychaeta
Order	Phyllodocida

Genome Traits	Expected	Observed
Haploid size (bp)	1,140,716,906	1,179,141,958
Haploid Number	10 (source: ancestor)	11
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
. BUSCO single copy value is less than 90% for collapsed
. Assembly length loss > 3% for collapsed

Curator notes

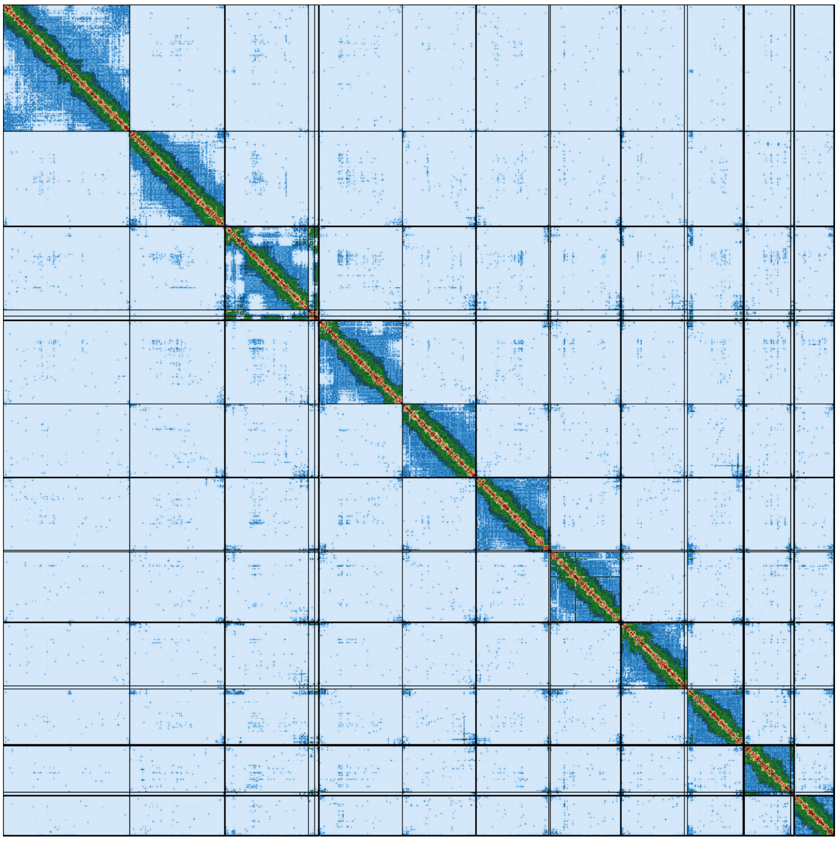
. Interventions/Gb: 97
. Contamination notes: ""
. Other observations: "The assembly of Iphione fimbriata (wpIphFimb1) is based on 58X PacBio data and 266X Arima Hi-C data generated as part of the ATLASea programme (https://www.atlasea.fr). The assembly process included the following steps: initial PacBio 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, 1382 contigs were identified as contaminants (bacterial, archaeal, or viral), totaling 27.985 Mb (with the largest being 0.248 Mb). Additionally, 686 regions totaling 78.182 Mb (with the largest being 2.503 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, 45 haplotypic regions were removed, totaling 41Mb, (with the largest being 4,5Mb). 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	1,221,004,372	1,179,141,958
GC %	43.37	43.32
Gaps/Gbp	262.08	267.99
Total gap bp	32,000	36,400
Scaffolds	139	89
Scaffold N50	106,525,323	104,334,169
Scaffold L50	5	5
Scaffold L90	10	10
Contigs	459	405
Contig N50	6,416,059	6,589,000
Contig L50	47	44
Contig L90	181	169
QV	47.7588	47.9278
Kmer compl.	70.1074	68.9335
BUSCO sing.	86.9%	88.3%
BUSCO dupl.	3.0%	1.3%
BUSCO frag.	8.0%	8.1%
BUSCO miss.	2.2%	2.3%

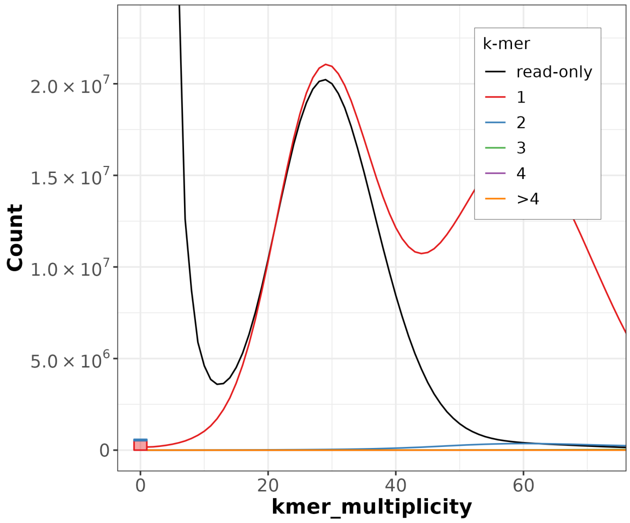
BUSCO: 5.8.2 (euk_genome_met, metaeuk) / Lineage: lophotrochozoa_odb12 (genomes:75, BUSCOs:1252)

HiC contact map of curated assembly

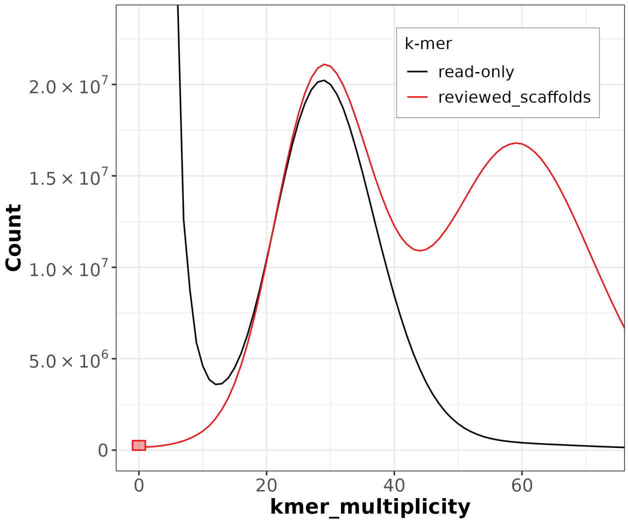


collapsed [LINK]

K-mer spectra of curated assembly

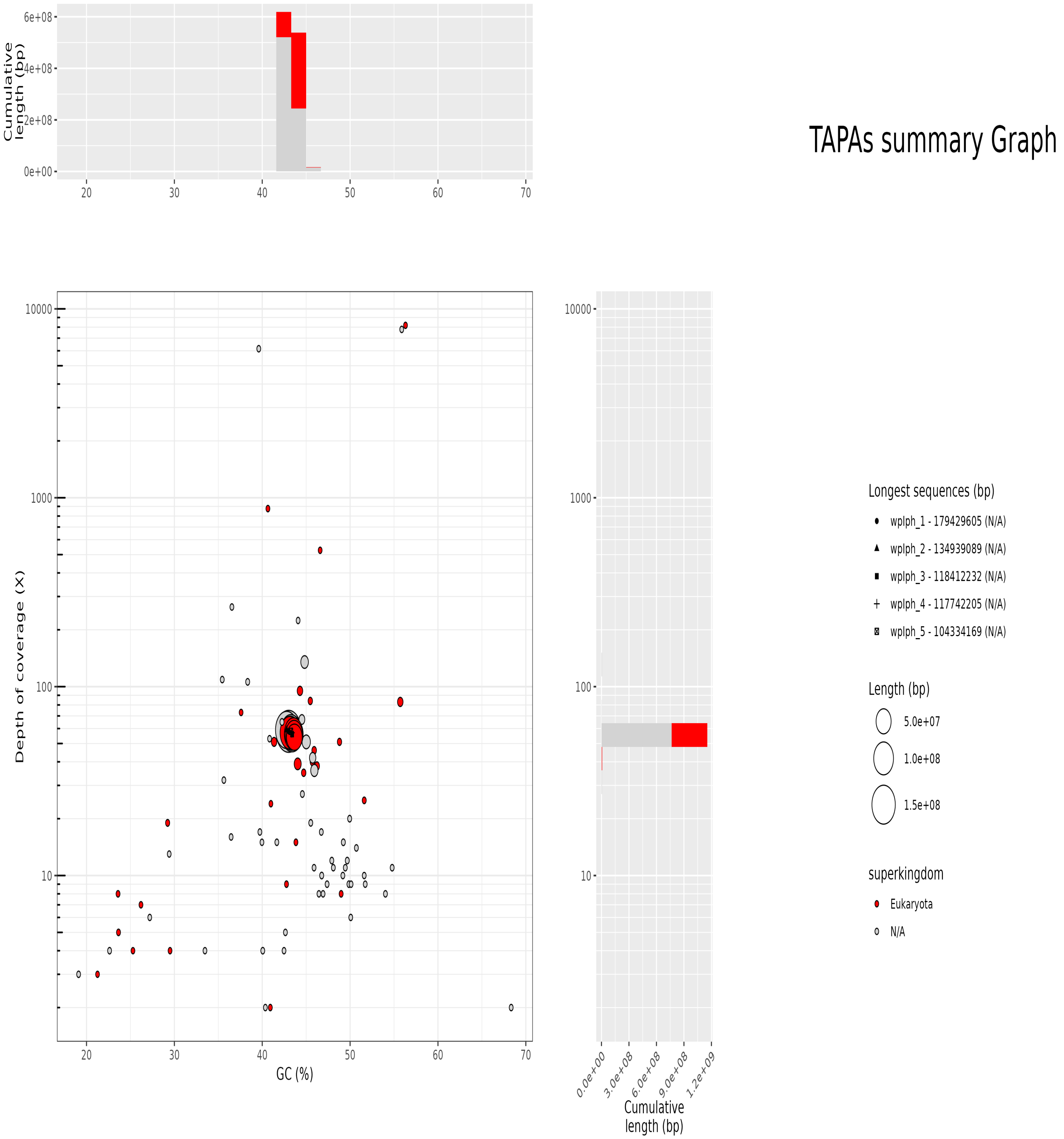


Distribution of k-mer counts per copy numbers found in asm



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	PACBIO Hifi	Arima
Coverage	58	266

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: 2025-09-02 21:59:49 CEST