

ERGA Assembly Report

v24.10.15

Tags: ATLASea[INVALID TAG]

TxID	1764201
ToLID	ucNepSpeb1
Species	Nephroselmis sp. RCC3064
Class	Nephroselmidophyceae
Order	Nephroselmidales

Genome Traits	Expected	Observed
Haploid size (bp)	50,934,642	61,499,123
Haploid Number	7 (source: ancestor)	28
Ploidy	2 (source: ancestor)	2
Sample Sex	Unknown	Unknown

EBP metrics summary and curation notes

Obtained EBP quality metric for collapsed: 6.6.Q37

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
- . QV value is less than 40 for collapsed

Curator notes

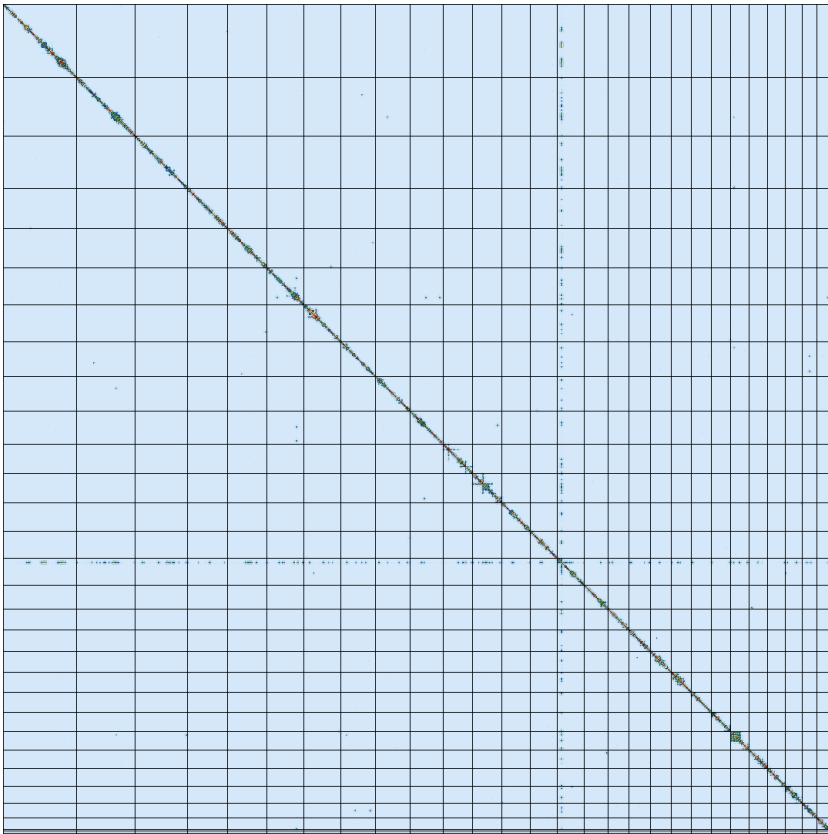
. Interventions/Gb: 0
. Contamination notes: ""
. Other observations: "The assembly of Nephroselmis sp. RCC3064 (ucNepSpeb1) is based on 95X ONT data and 244X Arima Hi-C data generated as part of the ATLASea programme (<https://www.atlasea.fr>). The assembly process included the following steps: initial ONT 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, 3 contigs were identified as contaminants (bacterial, archaeal, or viral), totaling 0.069 Mb (with the largest being 0.047 Mb). Additionally, 190 regions totaling 5.71 Mb (with the largest being 0.13 Mb) were identified as haplotypic duplications and removed. Mitochondrial and chloroplastic genomes were assembled using ptGAUL. Finally, the primary assembly was analyzed and manually improved using Pretext. 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	61,618,922	61,499,123
GC %	73.3	73.39
Gaps/Gbp	16.23	16.26
Total gap bp	100	100
Scaffolds	36	32
Scaffold N50	2,466,477	2,466,477
Scaffold L50	10	10
Scaffold L90	24	24
Contigs	37	33
Contig N50	2,466,477	2,466,477
Contig L50	11	11
Contig L90	25	25
QV	35.3322	37.2628
Kmer compl.	98.7592	98.7597
BUSCO sing.	95.2%	95.1%
BUSCO dupl.	0.3%	0.3%
BUSCO frag.	0.3%	0.3%
BUSCO miss.	4.2%	4.2%

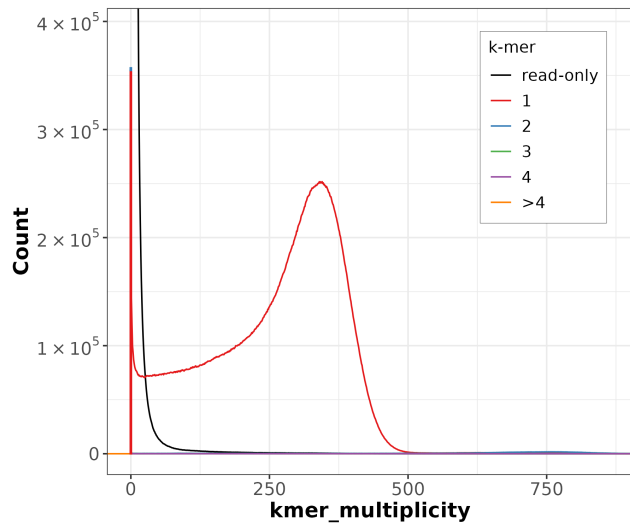
BUSCO: 6.1.0 (euk_genome_min, miniprot) / Lineage: chlorophyta_odb12.2 (genomes:37, BUSCOs:1437)

HiC contact map of curated assembly

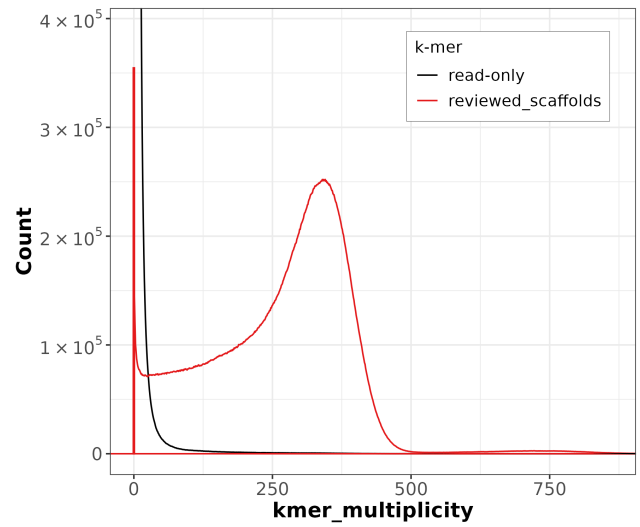


collapsed [\[LINK\]](#)

K-mer spectra of curated assembly

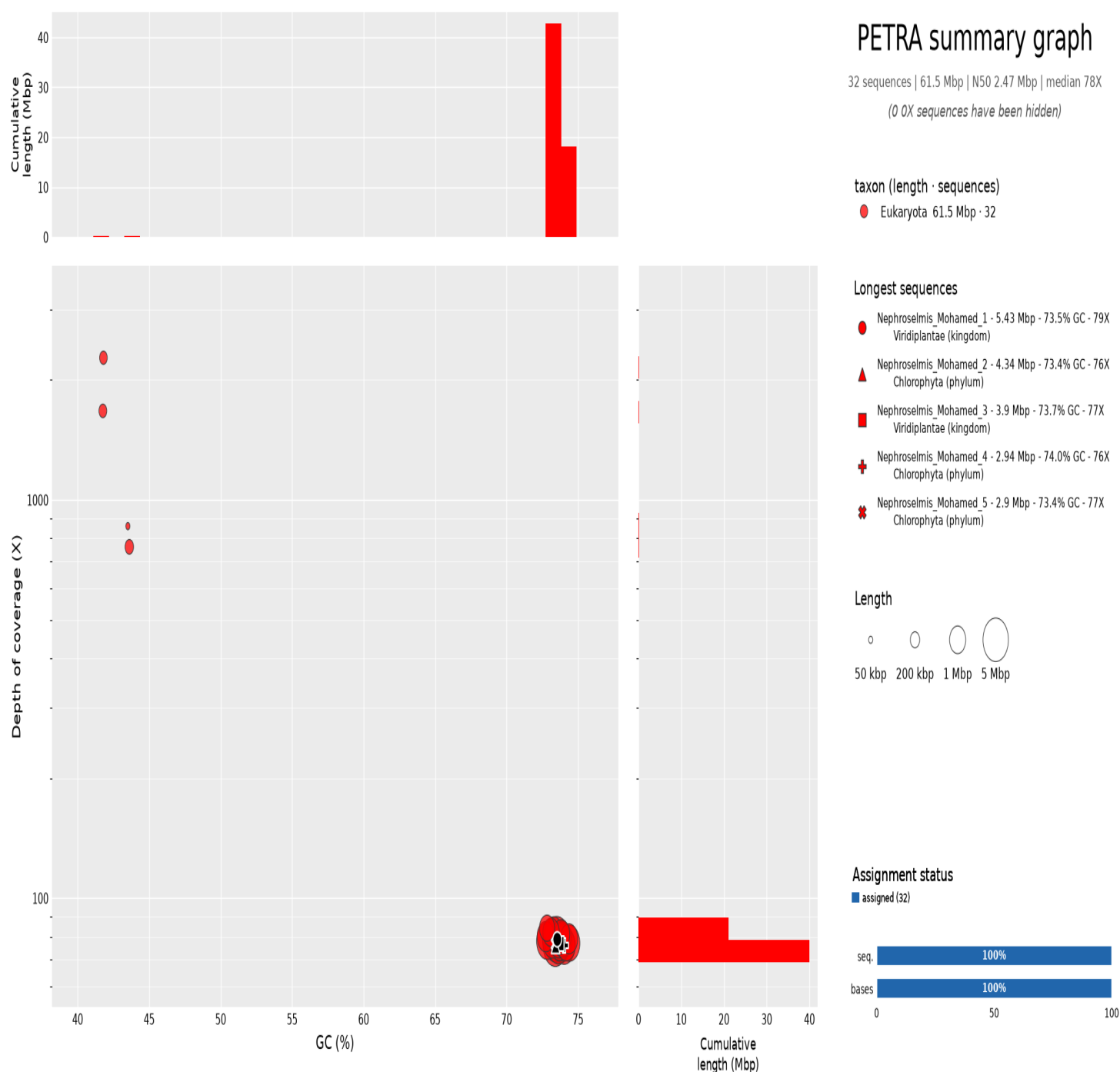


Distribution of k-mer counts per copy numbers found in asm



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	Long reads	Arima
Coverage	95	243

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*: 0.2.4
 - |_ *key param*: NA
- **PretextView**
 - |_ *ver*: 1.0.7
 - |_ *key param*: NA

Submitter: Mohamed Meftah Saoues

Affiliation: Genoscope

Date and time: 2026-10-07 10:58:32 CEST