ERGA Assembly Report

v24.10.15

Tags: ATLASea[INVALID TAG]

TxID	464659
ToLID	**urHalSqua1**
Species	Jania squamata
Class	Florideophyceae
Order	Corallinales

Genome Traits	Expected	Observed
Haploid size (bp)	342,719,966	132,186,787
Haploid Number	4 (source: ancestor)	31
Ploidy	4 (source: ancestor)	2
Sample Sex	Unknown	Unknown

EBP metrics summary and curation notes

Obtained EBP quality metric for collapsed: 6.6.Q35

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
. Observed Ploidy is different from Expected

. QV value is less than 40 for collapsed
. 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: 9
. Contamination notes: ""
. Other observations: "The assembly of Jania squamata (urHalSqua1) is based on 62X PacBio data and 150X 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, 5918 contigs were identified as contaminants (bacterial, archaeal, or viral), totaling 234.781 Mb (with the largest being 5.497 Mb). Additionally, 341 regions totaling 16.752 Mb (with the largest being 0.961 Mb) were identified as haplotypic duplications and removed. The mitochondrial genome was assembled using Oatk. The chloroplast genome was assembled using Oatk. Finally, the primary assembly was analyzed and manually improved using Pretext. During manual

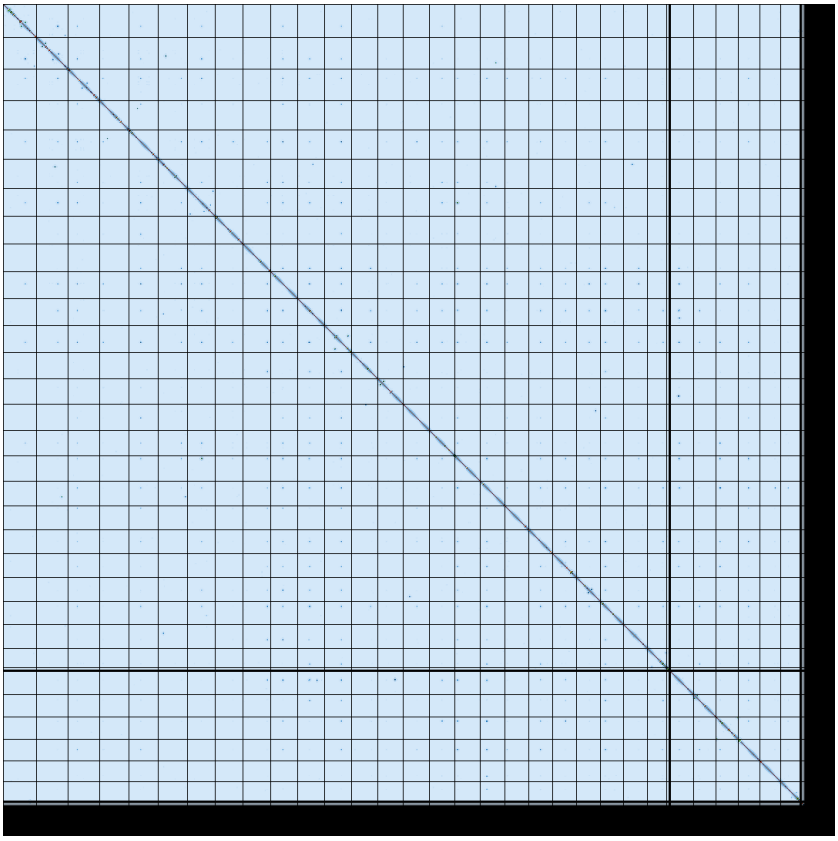
curation, 317 contaminant sequences were removed, totaling 218,734Mb (with the largest being 19.257Mb). Chromosome-scale scaffolds confirmed by Hi-C data were named in order of size. "

Quality metrics table

Metrics	Pre-curation collapsed	Curated collapsed
Total bp	350,764,399	132,186,787
GC %	37.1	44.34
Gaps/Gbp	188.16	83.22
Total gap bp	6,600	1,100
Scaffolds	550	235
Scaffold N50	11,637,298	4,109,289
Scaffold L50	13	15
Scaffold L90	40	29
Contigs	616	246
Contig N50	4,125,058	3,919,172
Contig L50	28	15
Contig L90	73	31
QV	36.9118	35.9073
Kmer compl.	78.4344	1.80918
BUSCO sing.	75.9%	85.3%
BUSCO dupl.	13.2%	2.6%
BUSCO frag.	3.7%	3.5%
BUSCO miss.	7.2%	8.6%

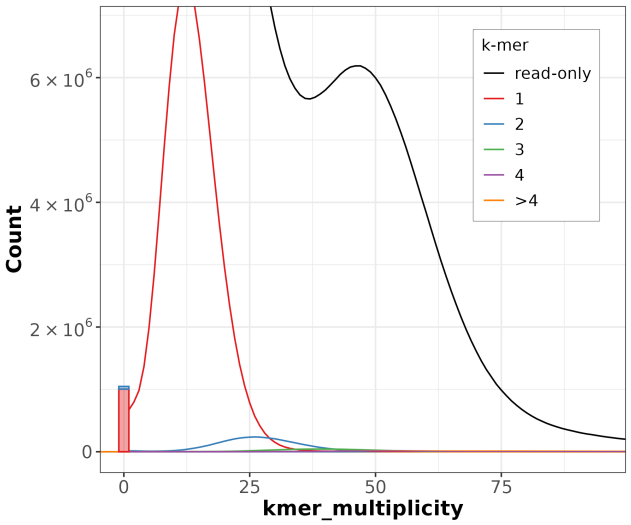
Warning! BUSCO versions or lineage datasets are not the same across results:
BUSCO: 6.0.0 (euk_genome_min, miniprot) / Lineage: rhodophyta_odb12 (genomes:8, BUSCOs:1591)
BUSCO: 6.1.0 (euk_genome_min, miniprot) / Lineage: rhodophyta_odb12 (genomes:8, BUSCOs:1591)

HiC contact map of curated assembly

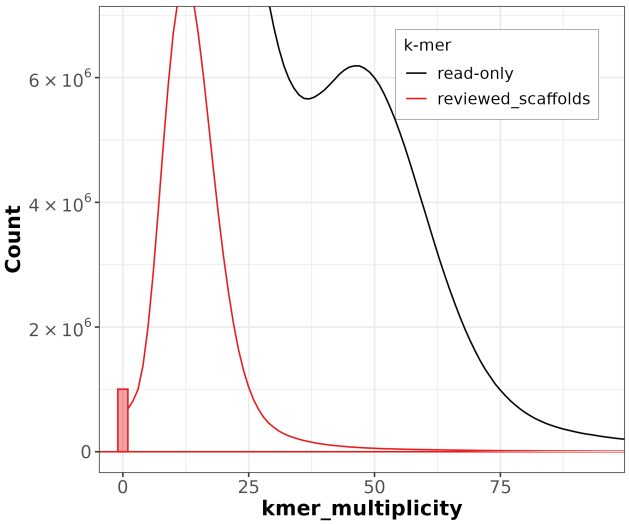


collapsed [LINK]

K-mer spectra of curated assembly

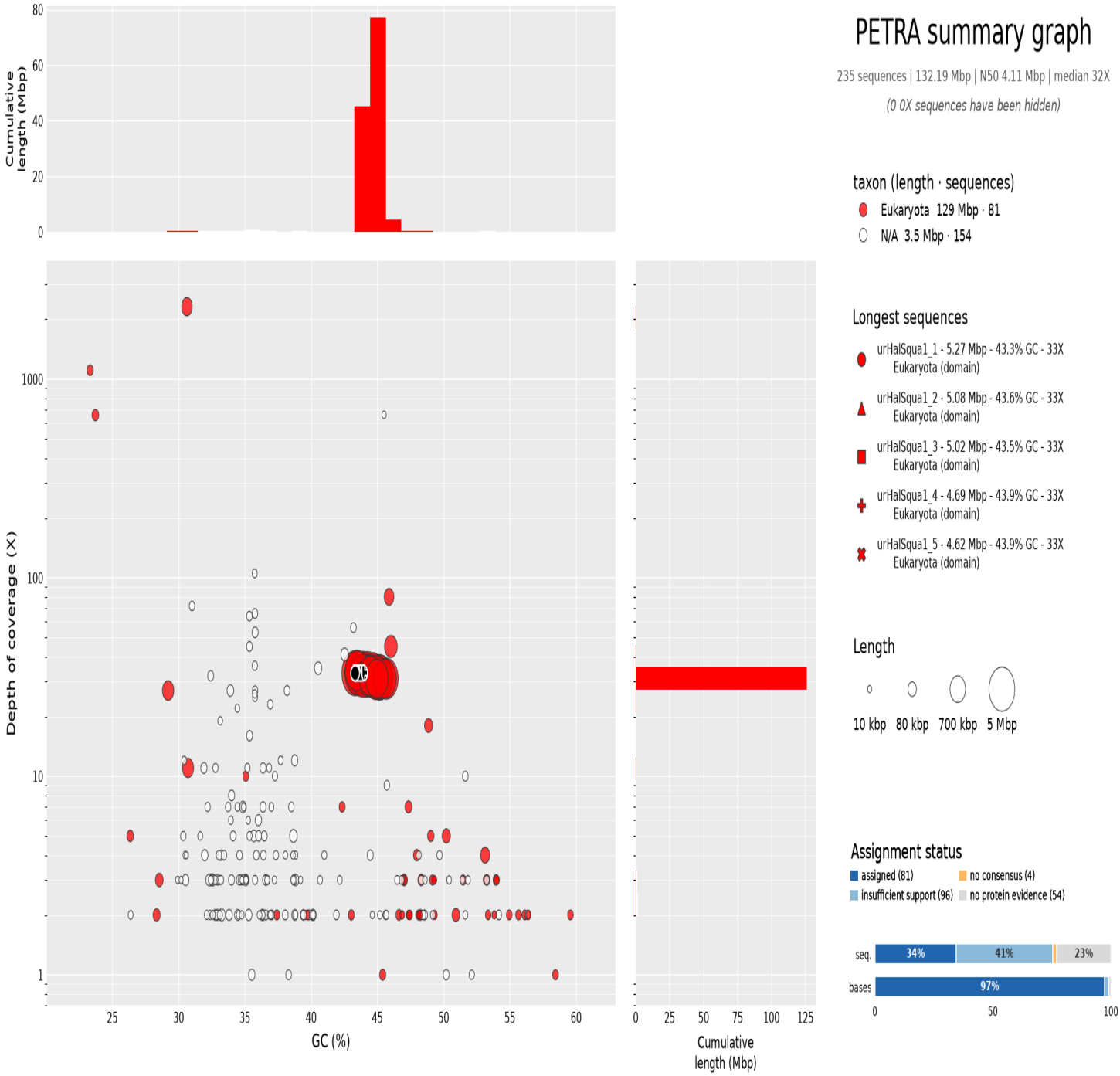


Distribution of k-mer counts per copy numbers found in asm



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	166	399

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: Sophie Layac
Affiliation: Genoscope

Date and time: 2026-10-07 21:44:43 CEST