ERGA Assembly Report

v24.10.15

Tags: Other[INVALID TAG]

TxID	1182198
ToLID	**fPemSco2_hap1**
Species	Pempheris schomburgkii
Class	Actinopteri
Order	Acropomatiformes

Genome Traits	Expected	Observed
Haploid size (bp)	583,426,511	746,079,414
Haploid Number	24 (source: ancestor)	24
Ploidy	2 (source: ancestor)	2
Sample Sex	Unknown	Unknown

EBP metrics summary and curation notes

Obtained EBP quality metric for collapsed: 6.7.Q60

The following metrics were automatically flagged as below EBP recommended standards or different from expected:

. Observed Haploid size (bp) has >20% difference with Expected

. Kmer completeness value is less than 90 for collapsed

Curator notes

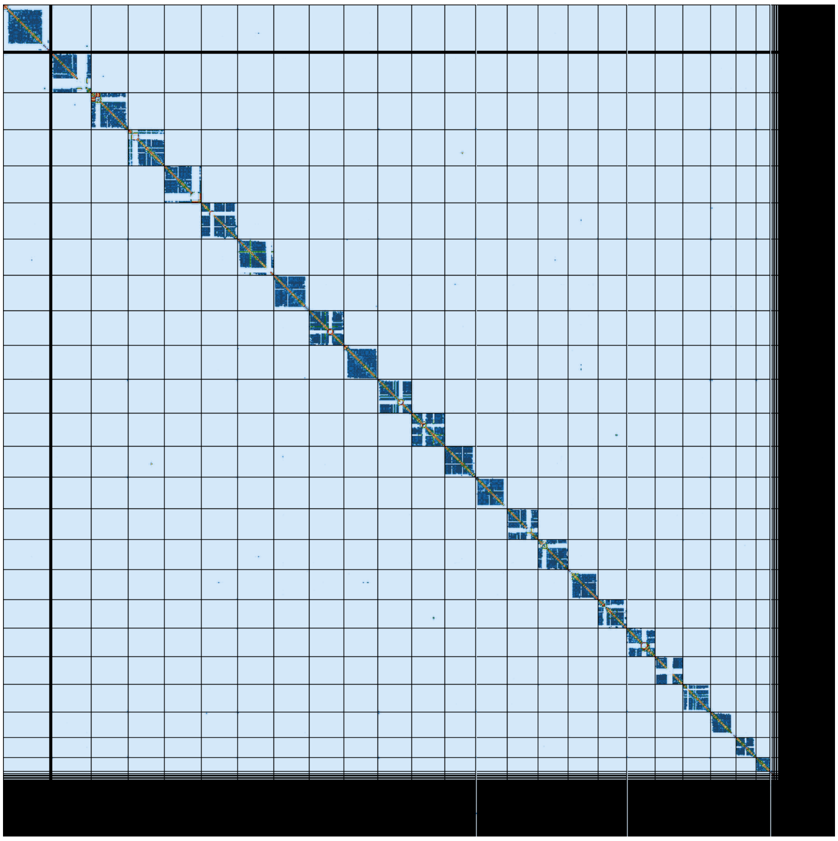
. Interventions/Gb: 173
. Contamination notes: ""
. Other observations: "The assembly of Pempheris schomburgkii (fPemSco2) was generated as part of the Vertebrate Genomes Project (https://vertebrategenomesproject.org/). Both haplotypes were analyzed and manually improved using Pretext "

Quality metrics table

Metrics	Pre-curation collapsed	Curated collapsed
Total bp	689,590,649	746,079,414
GC %	42.48	42.54
Gaps/Gbp	317.58	431.59
Total gap bp	43,800	64,400
Scaffolds	583	538
Scaffold N50	25,620,568	29,112,985
Scaffold L50	12	12
Scaffold L90	39	23
Contigs	802	860
Contig N50	6,532,619	6,822,809
Contig L50	33	36
Contig L90	185	212
QV	60.4581	60.2663
Kmer compl.	89.6667	89.9808
BUSCO sing.	95.6%	98.1%
BUSCO dupl.	0.8%	1.0%
BUSCO frag.	0.9%	0.1%
BUSCO miss.	2.7%	0.8%

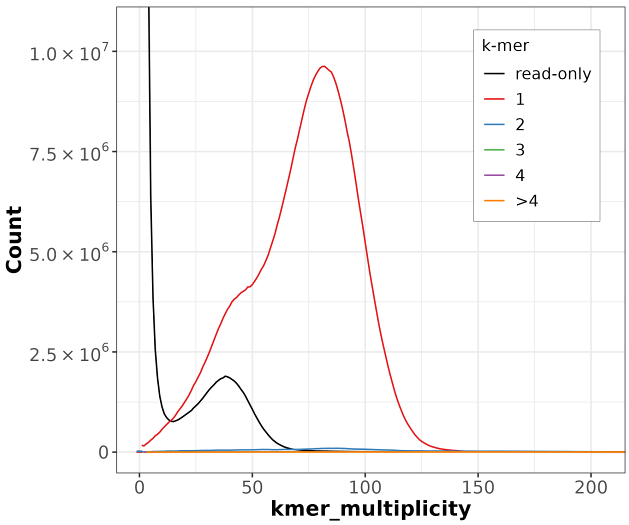
Warning! BUSCO versions or lineage datasets are not the same across results:
BUSCO: 5.8.2 (euk_genome_met, metaeuk) / Lineage: actinopterygii_odb12 (genomes:75, BUSCOs:7207)
BUSCO: 6.0.0 (euk_genome_min, miniprot) / Lineage: actinopterygii_odb12 (genomes:75, BUSCOs:7207)

HiC contact map of curated assembly

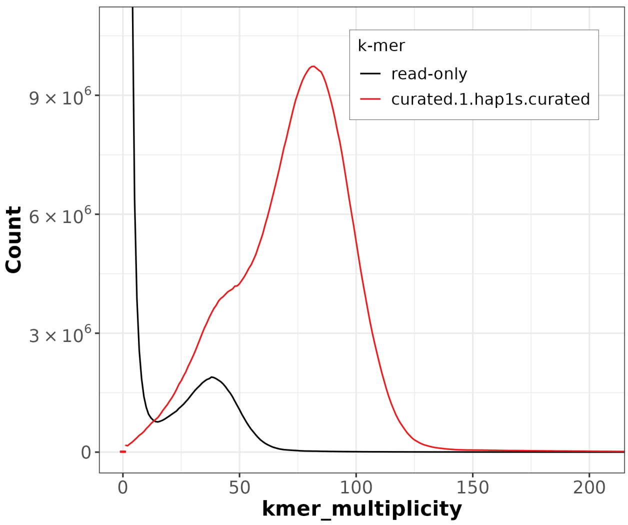


collapsed [LINK]

K-mer spectra of curated assembly

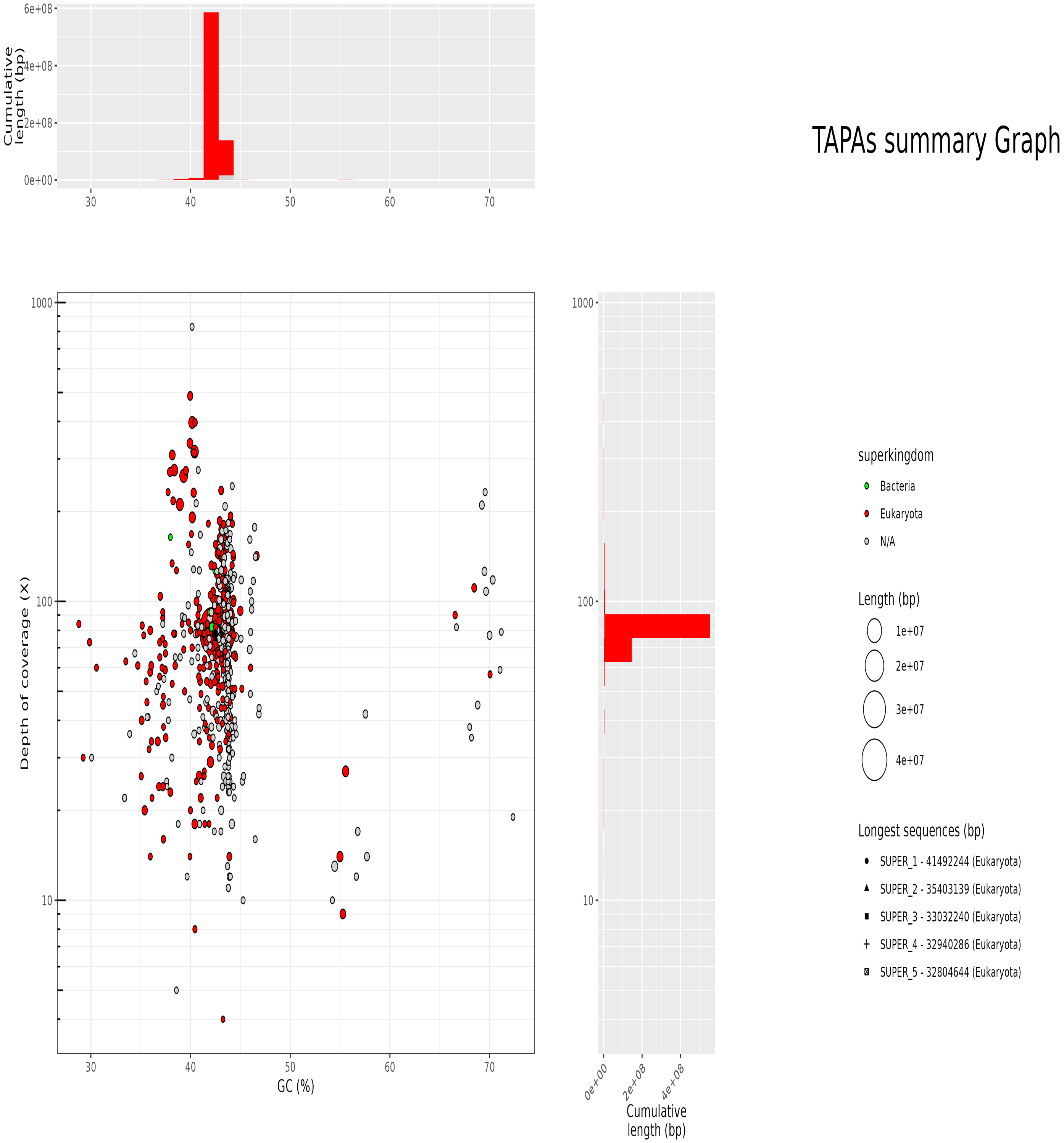


Distribution of k-mer counts per copy numbers found in asm



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	NA	NA

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: Emilie Teodori
Affiliation: Genoscope

Date and time: 2025-11-14 13:12:21 CET