ERGA Assembly Report

v24.10.15

Tags: ATLASea[INVALID TAG]

TxID	40499
ToLID	**fOcyChr1**
Species	Ocyurus chrysurus
Class	Actinopteri
Order	Lutjaniformes

Genome Traits	Expected	Observed
Haploid size (bp)	1,023,446,102	1,030,891,425
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: 7.7.Q50

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

. Kmer completeness value is less than 90 for collapsed

Curator notes

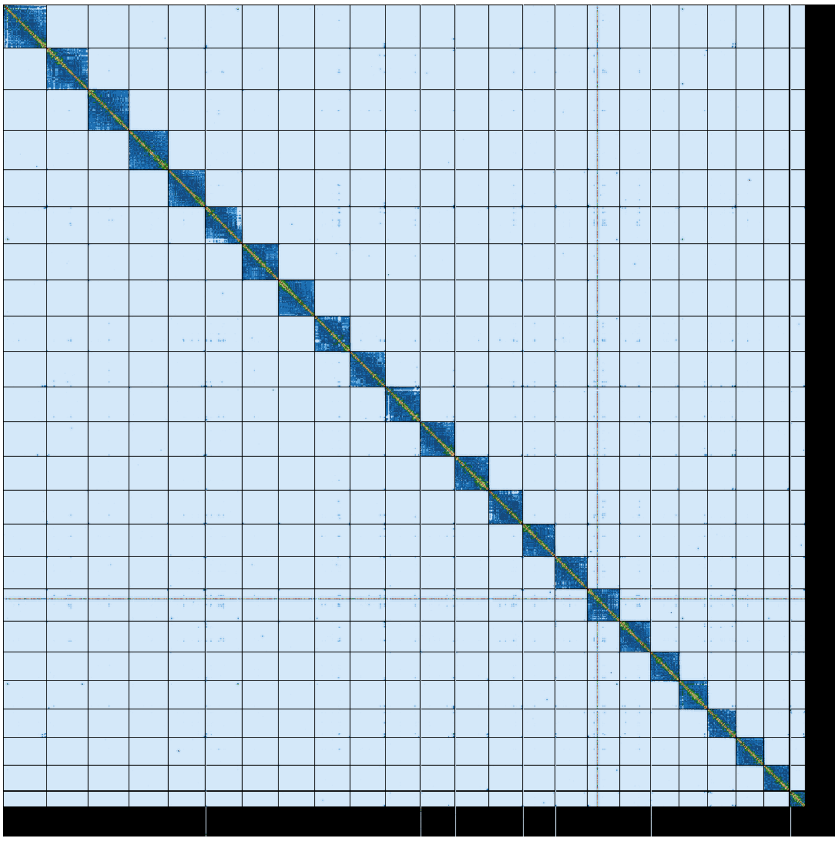
. Interventions/Gb: 73
. Contamination notes: ""
. Other observations: "The assembly of Ocyurus chrysurus (fOcyChr1) is based on 29X PacBio data and 163X Arima Hi-C data generated as part of the ATLASea programme (https://www.atlasea.fr). The assembly process included the following steps: initial PacBio 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, 31 contigs were identified as contaminants (bacterial, archaeal, or viral), totaling 3.007 Mb (with the largest being 0.337 Mb). Additionally, 351 regions totaling 30.304 Mb (with the largest being 2.298 Mb) were identified as haplotypic duplications and removed. The mitochondrial genome was assembled using OATK. During manual curation, 1 haplotypic regions was removed, totaling 0.570Mb. 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	1,031,543,081	1,030,891,425
GC %	40.08	40.08
Gaps/Gbp	170.62	198.86
Total gap bp	19,400	27,900
Scaffolds	520	475
Scaffold N50	42,157,256	43,064,661
Scaffold L50	12	11
Scaffold L90	22	22
Contigs	687	680
Contig N50	16,904,131	16,904,131
Contig L50	19	19
Contig L90	98	98
QV	50.0799	50.1186
Kmer compl.	87.4065	87.3845
BUSCO sing.	99.6%	99.6%
BUSCO dupl.	0.1%	0.1%
BUSCO frag.	0.0%	0.0%
BUSCO miss.	0.3%	0.3%

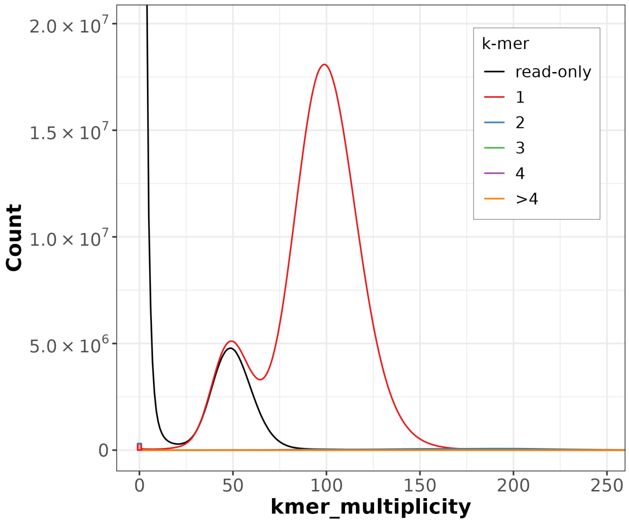
BUSCO: 6.0.0 (euk_genome_min, miniprot) / Lineage: actinopterygii_odb12 (genomes:75, BUSCOs:7207)

HiC contact map of curated assembly

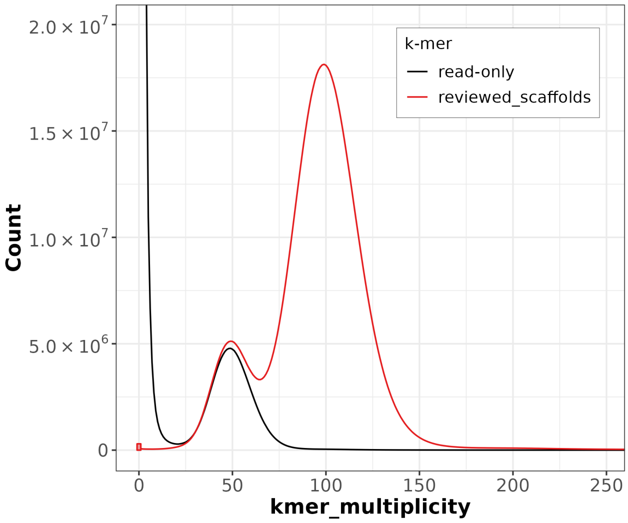


collapsed [LINK]

K-mer spectra of curated assembly

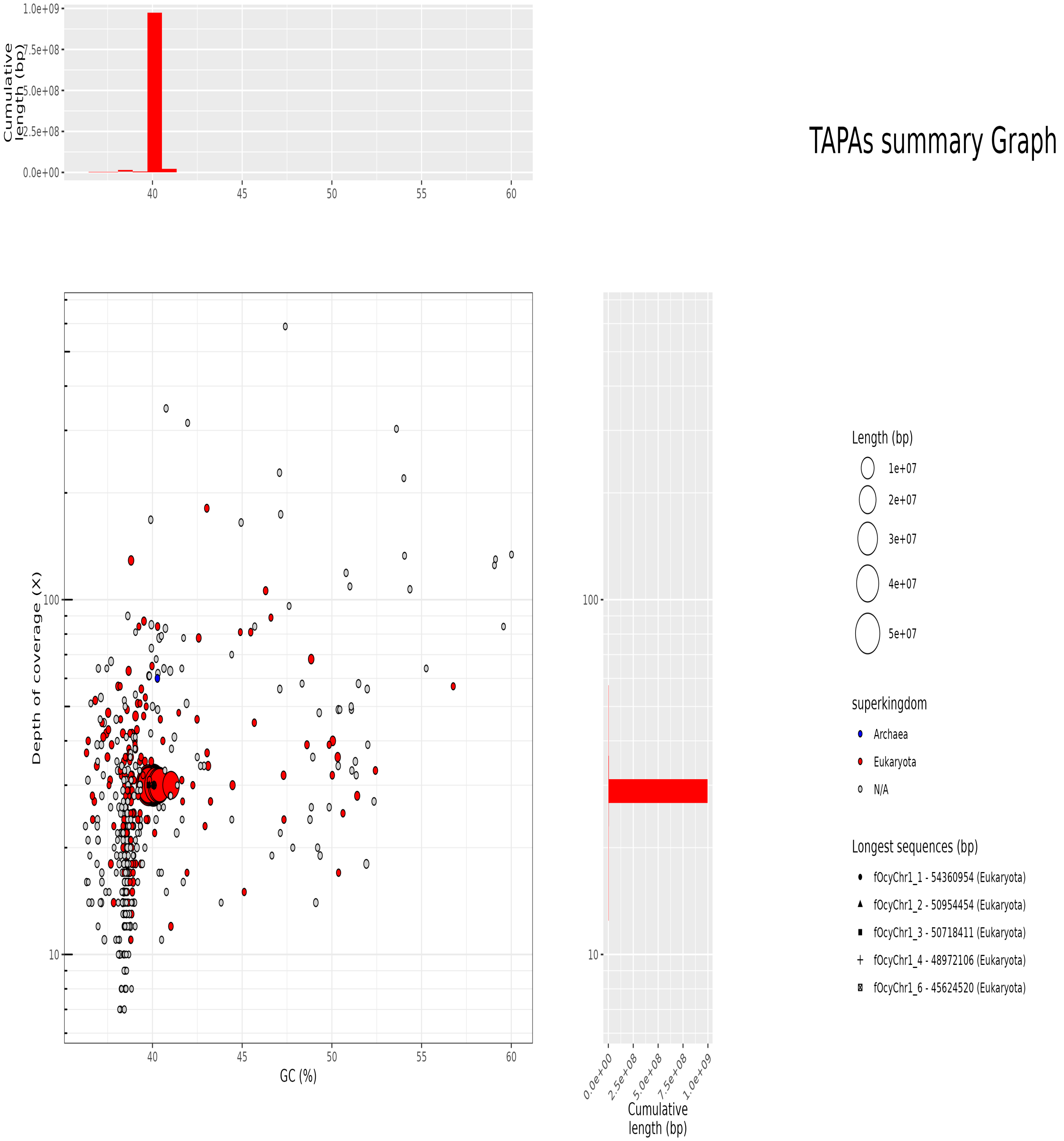


Distribution of k-mer counts per copy numbers found in asm



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	29	163

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: Benjamin Istace
Affiliation: Genoscope

Date and time: 2025-11-04 10:15:54 CET