

ERGA Assembly Report

v24.10.15

Tags: ATLASEa[INVALID TAG]

TxID	473301
ToLID	fCerMad1
Species	Ceratoscopelus maderensis
Class	Actinopteri
Order	Myctophiformes

Genome Traits	Expected	Observed
Haploid size (bp)	1,351,347,209	1,452,682,932
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.Q46

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

- . Kmer completeness value is less than 90 for collapsed
- . BUSCO duplicated value is more than 5% for collapsed

Curator notes

. Interventions/Gb: 77
. Contamination notes: ""
. Other observations: "The assembly of *Ceratoscopelus maderensis* (fCerMad1) is based on 51X PacBio data and 185X 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, 1 contigs was identified as contaminants (bacterial, archaeal, or viral), totaling 0.024 Mb (with the largest being 0.024 Mb). Additionally, 1291 regions totaling 150.476 Mb (with the largest being 1.201 Mb) were identified as haplotypic duplications and removed. Mitochondrial genome was assembled using OATK. Finally, the primary assembly was analyzed and manually improved using Pretext. During manual curation, 32 haplotypic regions were removed, totaling 26.966Mb (with the largest being 3.213Mb). Chromosome-scale scaffolds confirmed by Hi-C data were named in order of size. "

Quality metrics table

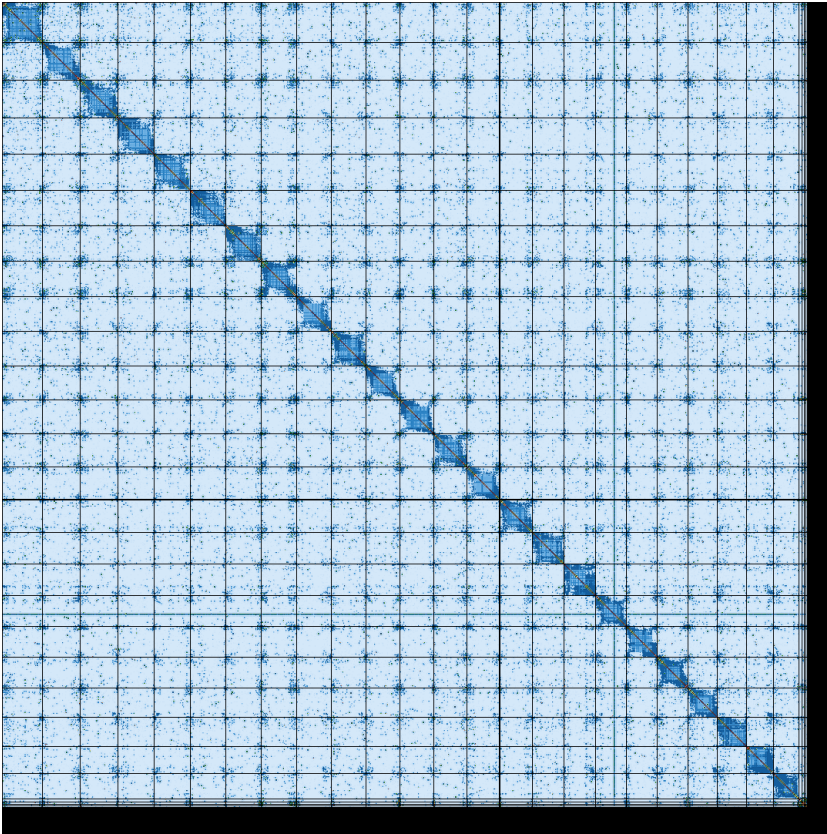
Metrics	Pre-curation collapsed	Curated collapsed
Total bp	1,479,667,175	1,452,682,932
GC %	42.65	42.66
Gaps/Gbp	633.93	640.19
Total gap bp	93,800	97,200
Scaffolds	441	429
Scaffold N50	59,886,613	59,189,287
Scaffold L50	11	12
Scaffold L90	22	23
Contigs	1,379	1,359
Contig N50	4,721,031	4,775,208
Contig L50	69	66
Contig L90	388	377
QV	46.2695	46.2668
Kmer compl.	73.3574	72.7152
BUSCO sing.	91.3%	91.2%
BUSCO dupl.	7.3%	6.1%
BUSCO frag.	0.2%	1.1%
BUSCO miss.	1.2%	1.6%

Warning! BUSCO versions or lineage datasets are not the same across results:

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

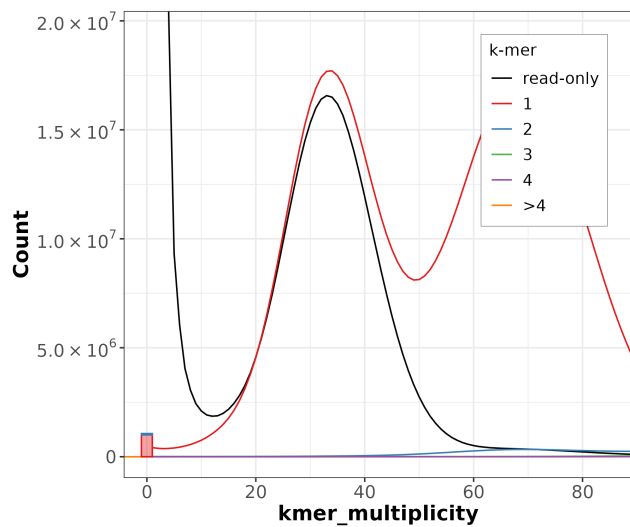
BUSCO: 5.8.2 (euk_genome_min, miniprot) / Lineage: actinopterygii_odb12 (genomes:75, BUSCOs:7207)

HiC contact map of curated assembly

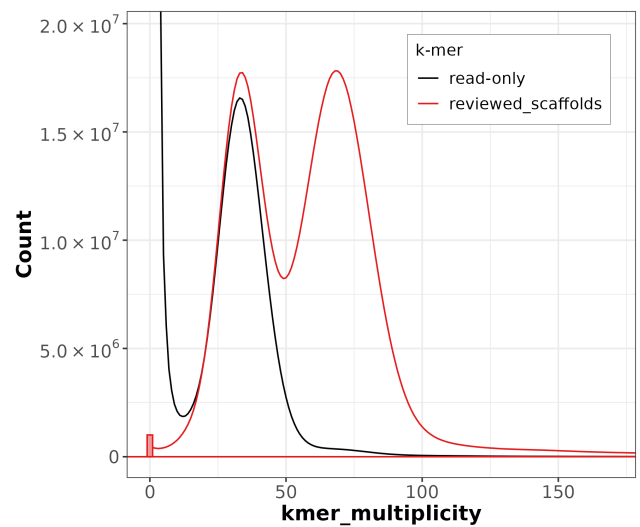


collapsed [\[LINK\]](#)

K-mer spectra of curated assembly

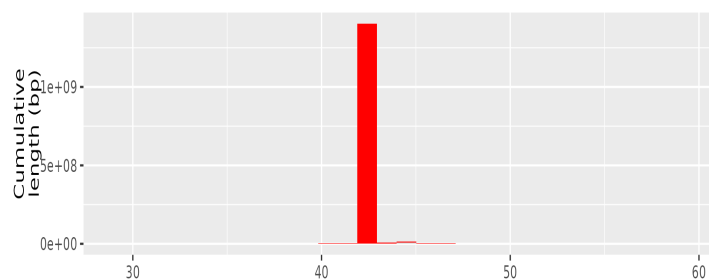


Distribution of k-mer counts per copy numbers found in asm

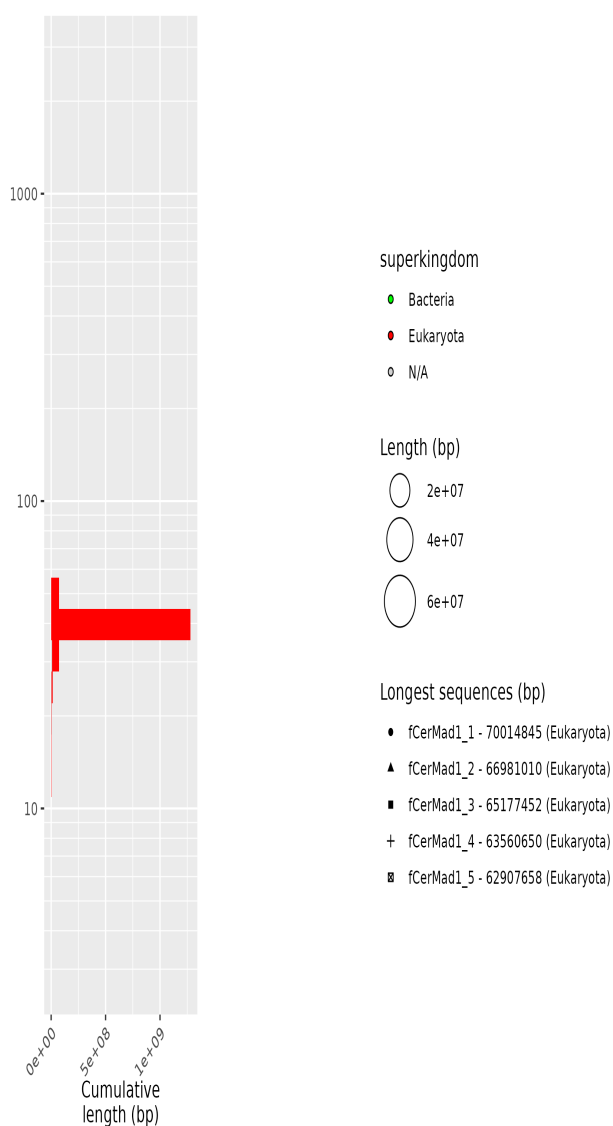
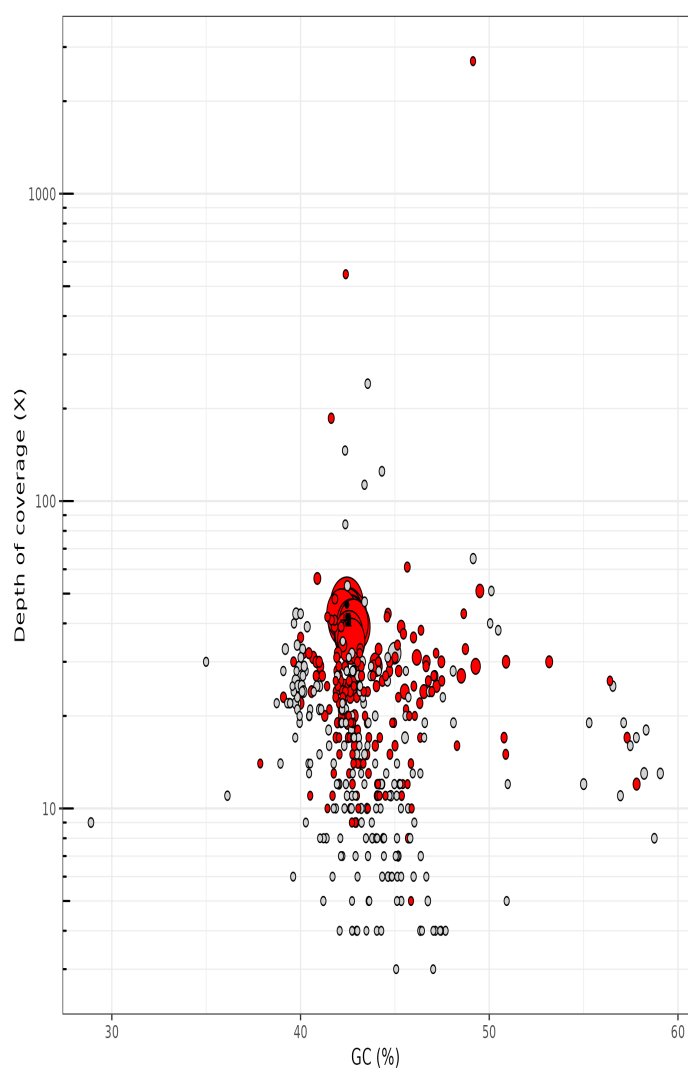


Distribution of k-mer counts coloured by their presence in reads/assemblies

Post-curation contamination screening



TAPAs summary Graph



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	52	188

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: Jean-Marc Aury

Affiliation: Genoscope

Date and time: 2026-08-08 16:55:57 CEST