

ERGA Assembly Report

v24.10.15

Tags: ATLASEa[INVALID TAG]

TxID	2740763
ToLID	fLepCan2
Species	Lepadogaster candolii
Class	Actinopteri
Order	Blenniiformes

Genome Traits	Expected	Observed
Haploid size (bp)	1,012,449,432	1,100,949,132
Haploid Number	19 (source: ancestor)	23
Ploidy	2 (source: ancestor)	2
Sample Sex	Unknown	Unknown

EBP metrics summary and curation notes

Obtained EBP quality metric for collapsed: 7.7.Q48

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

- . Observed Haploid Number is different from Expected
- . Kmer completeness value is less than 90 for collapsed

Curator notes

. Interventions/Gb: 109
. Contamination notes: ""
. Other observations: "The assembly of Lepadogaster candolii (fLepCan2.1) is based on 81X PacBio data and 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, 22 contigs were identified as contaminants (bacterial, archaeal, or viral), totaling 1.5 Mb (with the largest being 0.33 Mb). Additionally, 191 regions totaling 74.6 Mb (with the largest being 3.53 Mb) were identified as haplotypic duplications and removed. The mitochondrial genome was assembled using OATK. Finally, the primary assembly was analyzed and manually improved using Pretext. During manual curation, 8 haplotypic regions and 19 contaminant sequences were removed, totaling 8.17 Mb and 0.54 Mb (with the largest being 3.53 Mb and 0.04 Mb). Centromeric regions were more fragmented, and the organization of contigs in these regions has lower confidence. Centromeric regions were more fragmented, and the organization of contigs in these regions has lower confidence. Chromosome 5 appears to be a sex chromosome (half the coverage of other

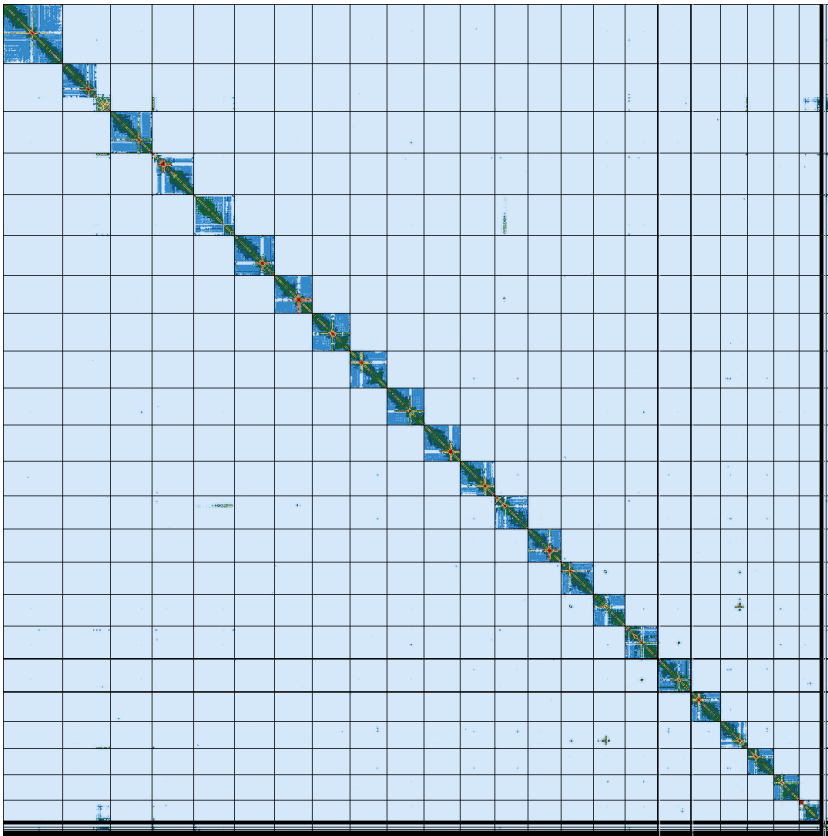
chromosomes), suggesting that it might correspond to two sex chromosomes that have been fused. However, there is no clear evidence of this on the Hi-C map. Chromosome-scale scaffolds confirmed by Hi-C data were named in order of size. "

Quality metrics table

Metrics	Pre-curation collapsed	Curated collapsed
Total bp	1,109,624,145	1,100,949,132
GC %	39.16	39.15
Gaps/Gbp	60.38	129.89
Total gap bp	6,700	23,700
Scaffolds	210	114
Scaffold N50	43,667,422	49,201,026
Scaffold L50	11	10
Scaffold L90	25	21
Contigs	277	257
Contig N50	18,941,000	18,941,000
Contig L50	21	21
Contig L90	66	64
QV	47.2322	48.91
Kmer compl.	80.8236	80.7562
BUSCO sing.	93.8%	94.1%
BUSCO dupl.	1.5%	1.1%
BUSCO frag.	1.2%	1.1%
BUSCO miss.	3.5%	3.6%

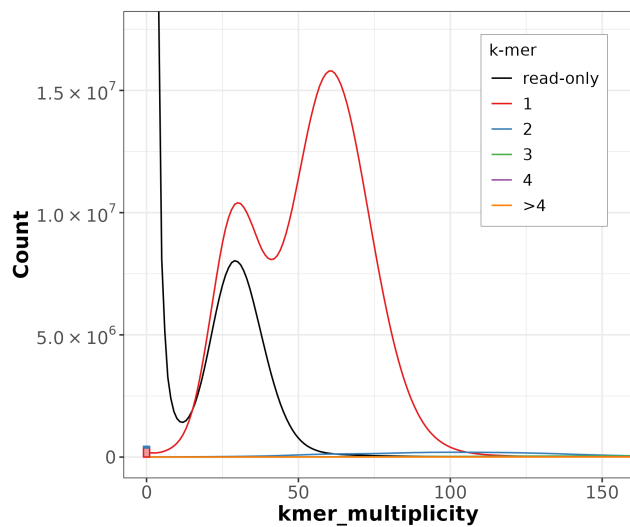
BUSCO: 5.8.2 (euk_genome_met, metaeuk) / Lineage: actinopterygii_odb12 (genomes:75, BUSCOs:7207)

HiC contact map of curated assembly

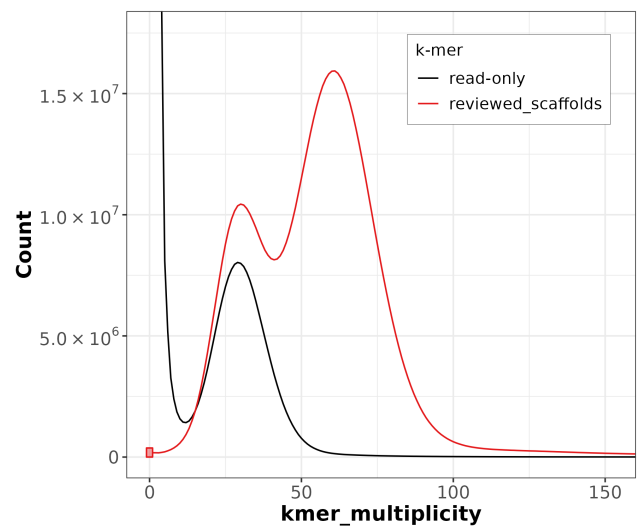


collapsed [\[LINK\]](#)

K-mer spectra of curated assembly

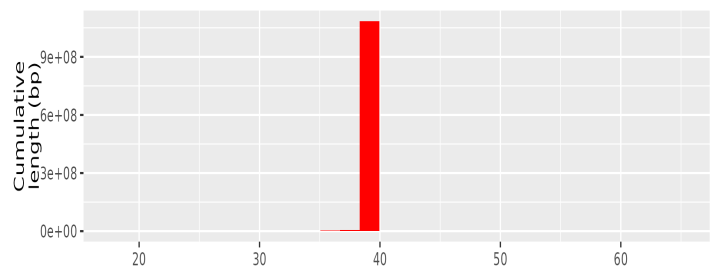


Distribution of k-mer counts per copy numbers found in asm

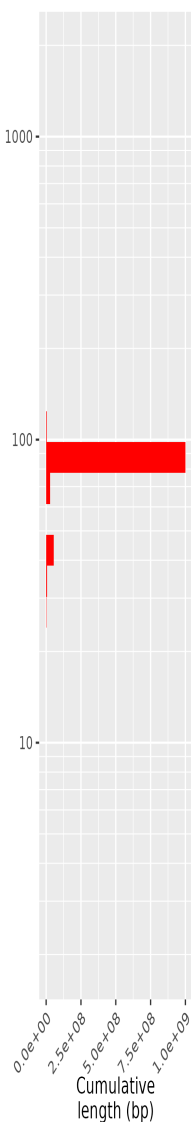
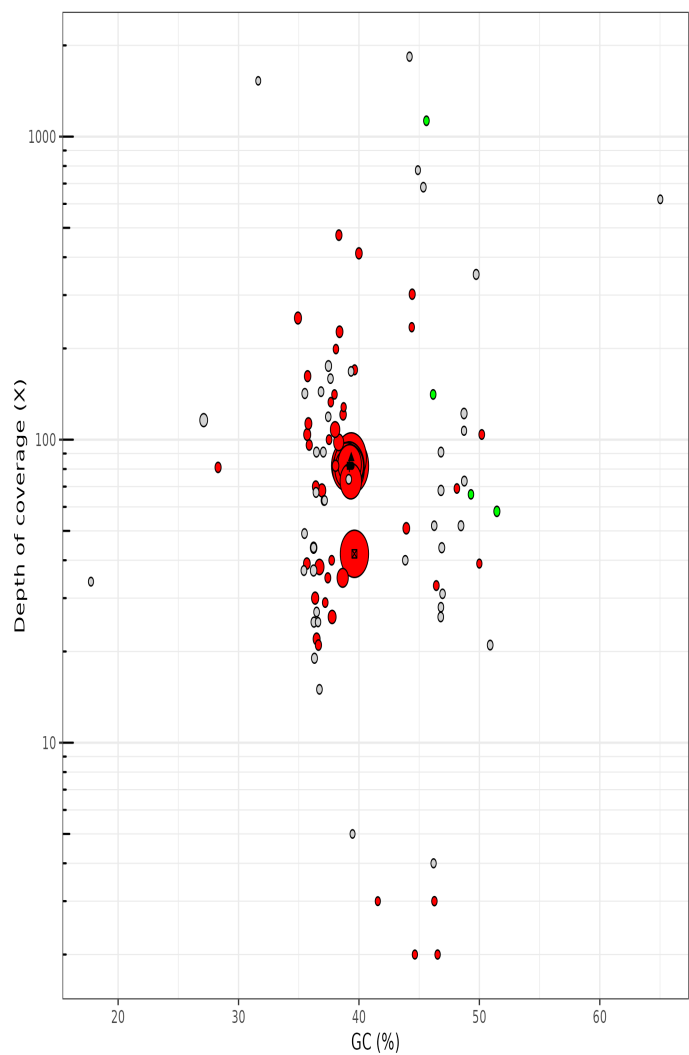


Distribution of k-mer counts coloured by their presence in reads/assemblies

Post-curation contamination screening



TAPAs summary Graph



- superkingdom
- Bacteria
 - Eukaryota
 - N/A
- Length (bp)
- 2e+07
 - 4e+07
 - 6e+07
- Longest sequences (bp)
- fLepCan1_1 - 78554009 (Eukaryota)
 - ▲ fLepCan1_2 - 63349531 (Eukaryota)
 - fLepCan1_3 - 55765364 (Eukaryota)
 - + fLepCan1_4 - 54671200 (Eukaryota)
 - ⊠ fLepCan1_5 - 54159018 (Eukaryota)

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	PACBIO Hifi	Arima
Coverage	81	138

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: 2025-09-10 10:50:45 CEST