ERGA Assembly Report

v24.10.15

Tags: ATLASea[INVALID TAG]

TxID	164309
ToLID	**fLepPur2**
Species	Lepadogaster purpurea
Class	Actinopteri
Order	Blenniiformes

Genome Traits	Expected	Observed
Haploid size (bp)	831,451,154	863,662,561
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.Q63

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

. Observed Haploid Number is different from Expected

Curator notes

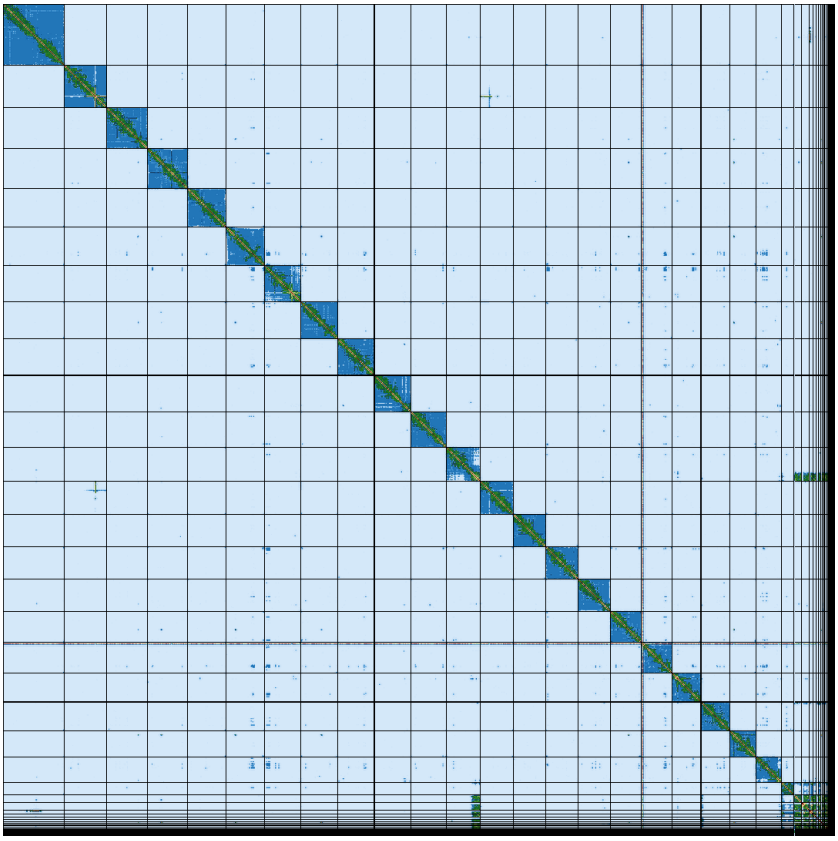
. Interventions/Gb: 23
. Contamination notes: ""
. Other observations: "The assembly of Lepadogaster purpurea (fLepPur2) is based on 40X ONT data and 154X Arima Hi-C data generated as part of the ATLASea programme (https://www.atlasea.fr). The assembly process included the following steps: initial ONT 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, 2 contigs were identified as contaminants (bacterial, archaeal, or viral), totaling 0.034 Mb (with the largest being 0.023 Mb). Additionally, 229 regions totaling 53.596 Mb (with the largest being 3.092 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, 5 haplotypic regions, totaling 0.676Mb, (with the largest being 0.145Mb). 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	864,391,855	863,662,561
GC %	38.83	38.82
Gaps/Gbp	86.77	90.31
Total gap bp	7,500	8,700
Scaffolds	96	92
Scaffold N50	36,980,150	36,980,150
Scaffold L50	11	11
Scaffold L90	21	21
Contigs	171	170
Contig N50	15,794,238	15,794,238
Contig L50	19	19
Contig L90	59	59
QV	49.4522	63.0313
Kmer compl.	93.7897	94.0846
BUSCO sing.	94.8%	94.8%
BUSCO dupl.	0.7%	0.7%
BUSCO frag.	1.1%	1.1%
BUSCO miss.	3.4%	3.4%

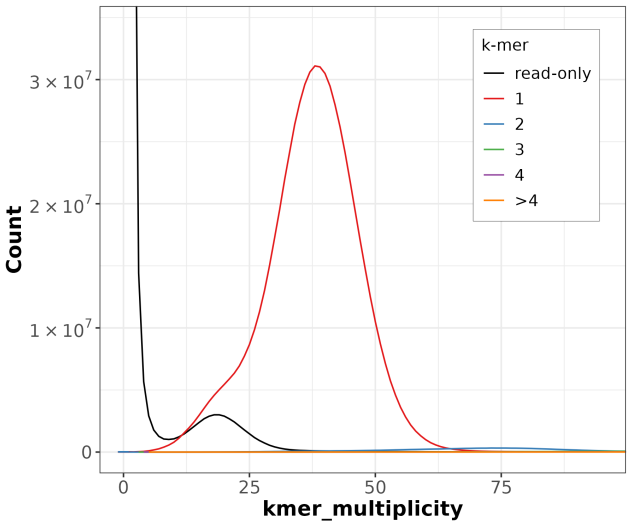
BUSCO: 5.8.2 (euk_genome_met, metaeuk) / Lineage: actinopterygii_odb12 (genomes:75, BUSCOs:7207)

HiC contact map of curated assembly

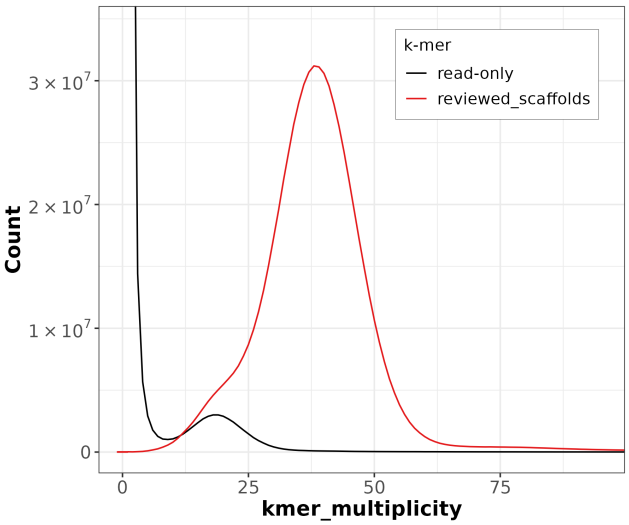


collapsed [LINK]

K-mer spectra of curated assembly

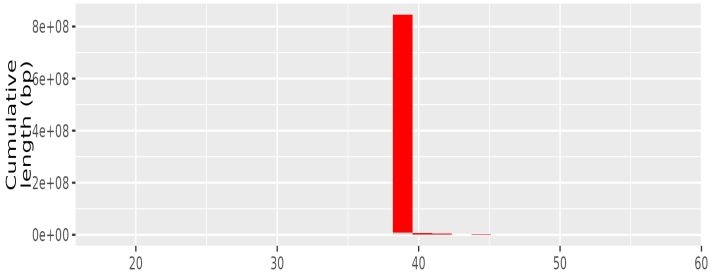


Distribution of k-mer counts per copy numbers found in asm

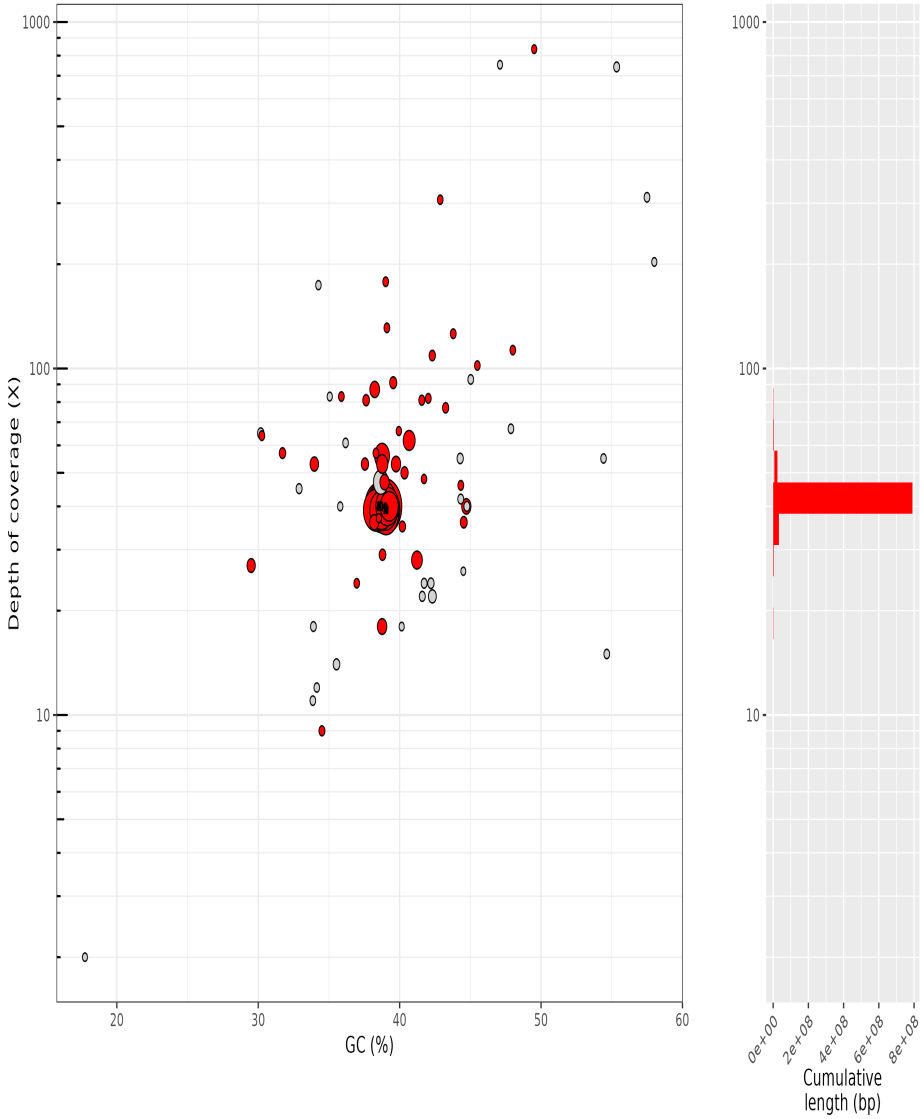


Distribution of k-mer counts coloured by their presence in reads/assemblies

Post-curation contamination screening



TAPAs summary Graph



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	39	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: 2025-04-29 11:18:07 CEST