

ERGA Assembly Report

v24.10.15

Tags: ATLASEa[INVALID TAG]

TxID	286208
ToLID	kaPolLaca1
Species	Polysyncraton lacazei
Class	Ascidacea
Order	Aplousobranchia

Genome Traits	Expected	Observed
Haploid size (bp)	403,657,564	772,749,557
Haploid Number	9 (source: ancestor)	23
Ploidy	2 (source: ancestor)	2
Sample Sex	Unknown	Unknown

EBP metrics summary and curation notes

Obtained EBP quality metric for collapsed: 5.7.Q46

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

- . Observed Haploid size (bp) has >20% difference with Expected
- . Observed Haploid Number is different from Expected
- . Kmer completeness value is less than 90 for collapsed
- . More than 1000 gaps/Gbp for collapsed

Curator notes

- . Interventions/Gb: 342
- . Contamination notes: ""
- . Other observations: "The assembly of Polysyncraton lacazei (kaPolLaca1) is based on 35X PacBio data generated from an amplified DNA library and 103X 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, 698 contigs were identified as contaminants (bacterial, archaeal, or viral), totaling 23.776 Mb (with the largest being 1.524 Mb). Additionally, 1074 regions totaling 37.857 Mb (with the largest being 0.29 Mb) were identified as haplotypic duplications and removed. Finally, the primary assembly was analyzed and manually improved using Pretext. During manual curation, 8 haplotypic regions and 5 contaminant sequences were removed, totaling 1.371Mb and 0.256Mb, respectively (with the largest being 0.395Mb and 0.176Mb). Chromosome-scale scaffolds confirmed by Hi-C data were named in order of size. "

Quality metrics table

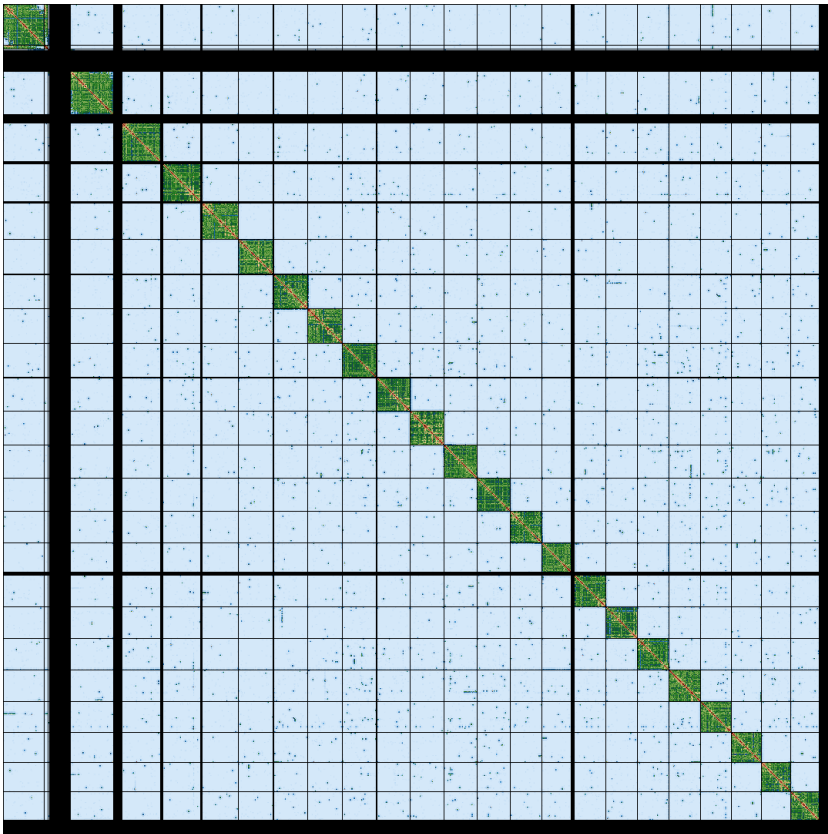
Metrics	Pre-curation collapsed	Curated collapsed
Total bp	774,356,821	772,749,557
GC %	38.26	38.26
Gaps/Gbp	2,081.73	2,148.17
Total gap bp	161,200	180,100
Scaffolds	489	461
Scaffold N50	30,498,439	30,732,984
Scaffold L50	11	12
Scaffold L90	22	23
Contigs	2,101	2,121
Contig N50	651,866	642,306
Contig L50	354	360
Contig L90	1,174	1,192
QV	46.4665	46.4909
Kmer compl.	84.8584	84.7621
BUSCO sing.	91.4%	91.1%
BUSCO dupl.	3.6%	3.4%
BUSCO frag.	2.5%	3.0%
BUSCO miss.	2.5%	2.5%

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

BUSCO: 6.0.0 (euk_genome_min, miniprot) / Lineage: metazoa_odb12 (genomes:206, BUSCOs:672)

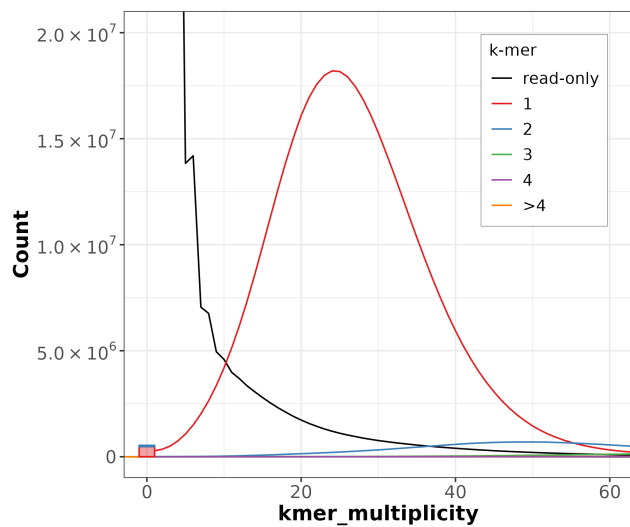
BUSCO: 5.8.2 (euk_genome_min, miniprot) / Lineage: metazoa_odb12 (genomes:206, BUSCOs:672)

HiC contact map of curated assembly

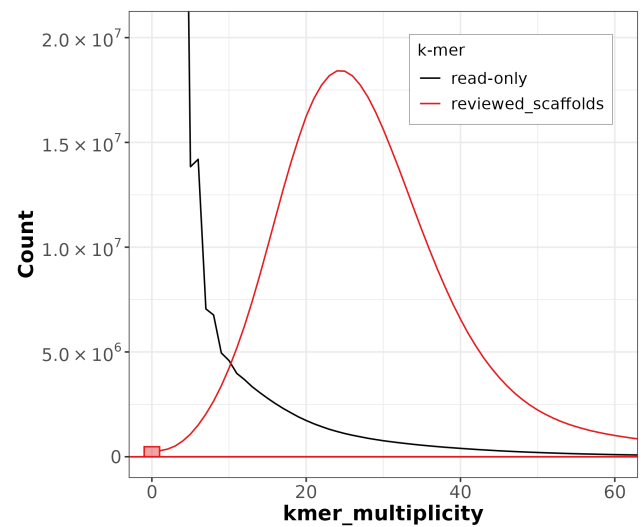


collapsed [\[LINK\]](#)

K-mer spectra of curated assembly

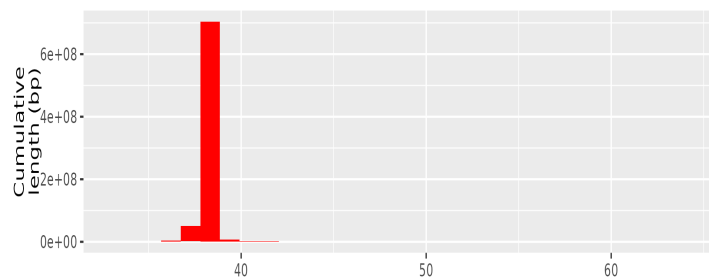


Distribution of k-mer counts per copy numbers found in asm

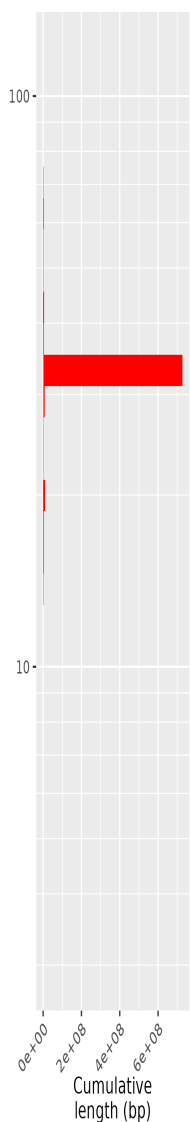
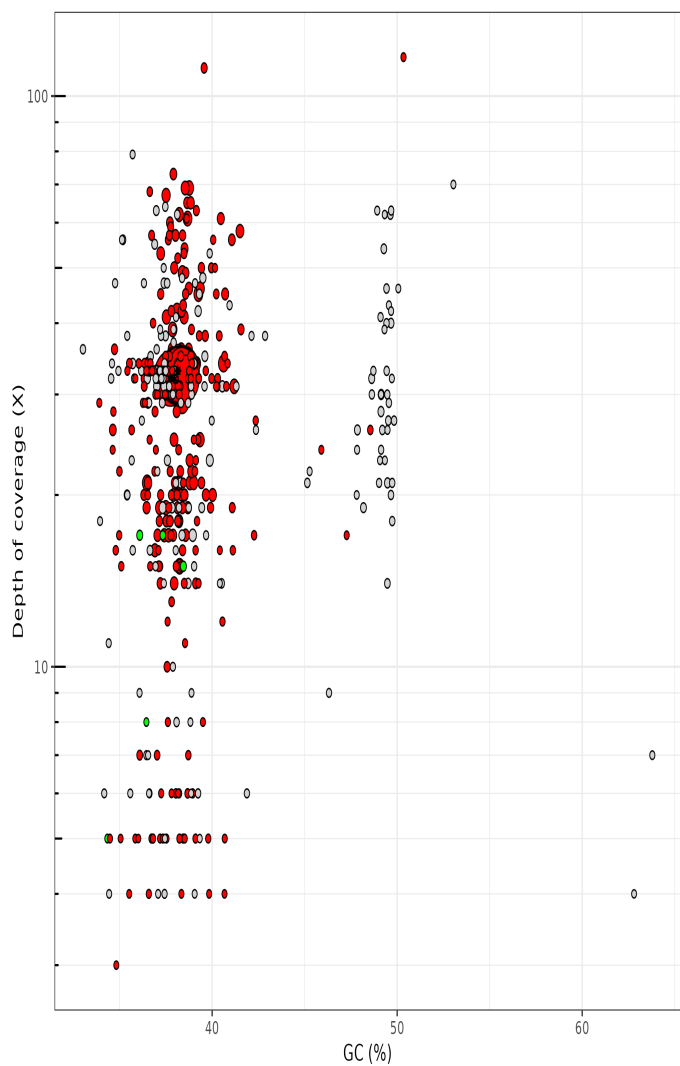


Distribution of k-mer counts coloured by their presence in reads/assemblies

Post-curation contamination screening



TAPAs summary Graph



- superkingdom
- Bacteria
 - Eukaryota
 - N/A
- Length (bp)
- 1e+07
 - 2e+07
 - 3e+07
- Longest sequences (bp)
- kaPollaca1_2 - 39974356 (Eukaryota)
 - ▲ kaPollaca1_1 - 38631024 (Eukaryota)
 - kaPollaca1_3 - 35494623 (Eukaryota)
 - + kaPollaca1_4 - 35206372 (Eukaryota)
 - ⊠ kaPollaca1_5 - 33740367 (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	Long reads	Arima
Coverage	35	102

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: 2026-07-28 00:33:50 CEST