

ERGA Assembly Report

v24.10.15

Tags: ATLASEa[INVALID TAG]

TxID	1210411
ToLID	wpLepClav2
Species	Lepidonotus clava
Class	Polychaeta
Order	Phyllodocida

Genome Traits	Expected	Observed
Haploid size (bp)	1,035,948,466	1,015,269,209
Haploid Number	10 (source: ancestor)	18
Ploidy	2 (source: ancestor)	2
Sample Sex	Unknown	Unknown

EBP metrics summary and curation notes

Obtained EBP quality metric for collapsed: 7.7.Q47

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
- . BUSCO single copy value is less than 90% for collapsed
- . Assembly length loss > 3% for collapsed

Curator notes

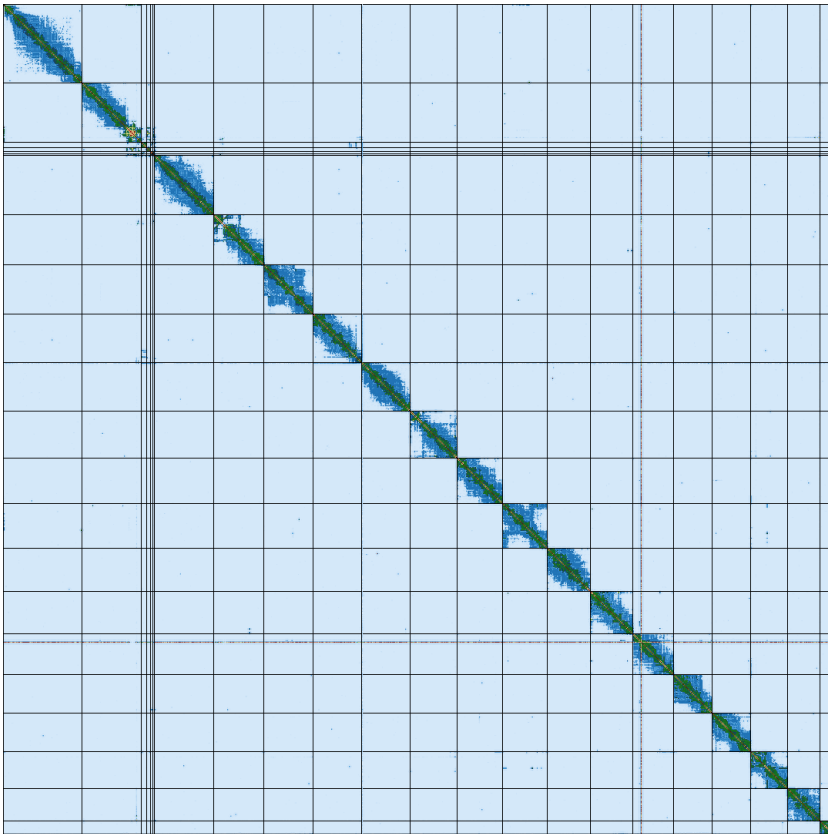
- . Interventions/Gb: 136
- . Contamination notes: ""
- . Other observations: "The assembly of *Lepidonotus clava* (wpLepClav2) is based on 58X PacBio data and 196 X Arima Hi-C data generated as part of the ATLASEa programme (<https://www.atlasea.fr>). The assembly process included the following steps: initial PacBio assembly generation with Hifiiasm, removal of contaminant sequences using Context, removal of haplotypic duplications using purge_dups, and Hi-C-based scaffolding with YaHS. In total, 1138 contigs were identified as contaminants (bacterial and archaeal) , totaling 39 Mb (with the largest being 3 Mb). Additionally, 641 regions totaling 404 Mb (with the largest being 8.9Mb) 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, 46 haplotypic regions and 52 contaminant sequences were removed, totaling 458.5 Mb and 1.2 Mb, respectively (with the largest being 43 Mb and 0.1 Mb).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	1,500,691,440	1,015,269,209
GC %	40.51	40.61
Gaps/Gbp	111.95	149.71
Total gap bp	16,800	21,100
Scaffolds	160	70
Scaffold N50	66,839,741	57,073,074
Scaffold L50	9	8
Scaffold L90	21	16
Contigs	328	222
Contig N50	13,089,000	15,394,861
Contig L50	34	20
Contig L90	119	73
QV	47.1296	47.1411
Kmer compl.	78.1153	56.4469
BUSCO sing.	57.6%	87.6%
BUSCO dupl.	34.0%	0.9%
BUSCO frag.	6.9%	7.7%
BUSCO miss.	1.5%	3.8%

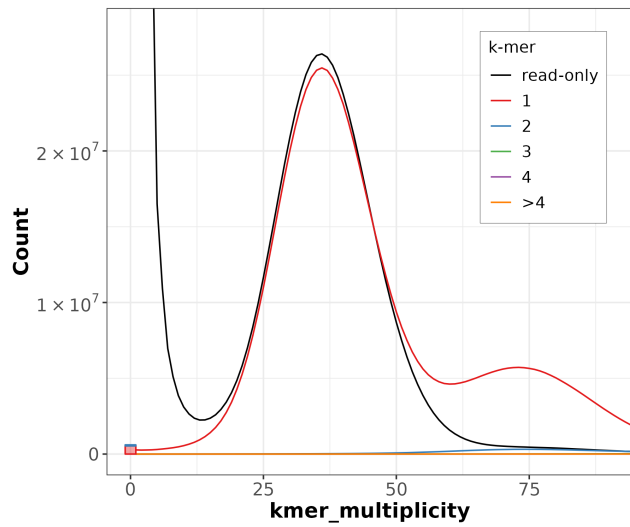
BUSCO: 5.8.2 (euk_genome_met, metaeuk) / Lineage: lophotrochozoa_odb12 (genomes:75, BUSCOs:1252)

HiC contact map of curated assembly

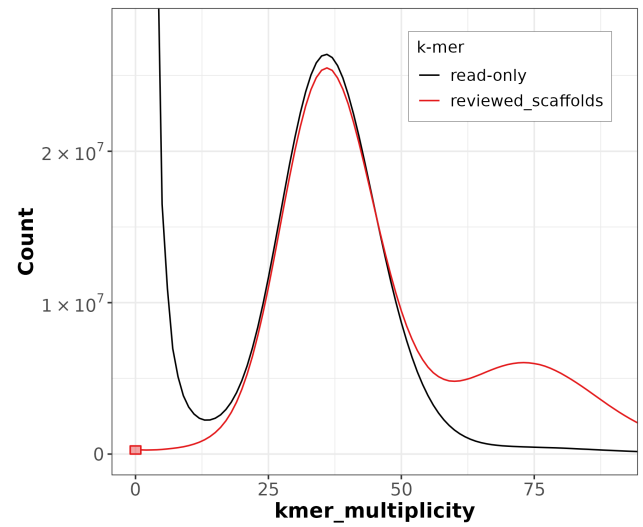


collapsed [\[LINK\]](#)

K-mer spectra of curated assembly

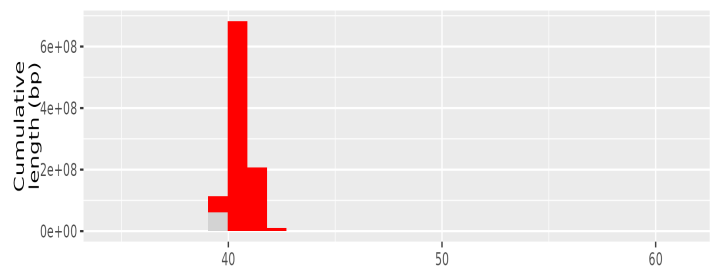


Distribution of k-mer counts per copy numbers found in asm

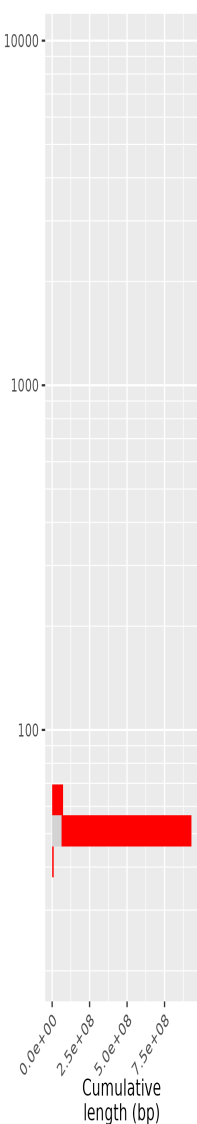
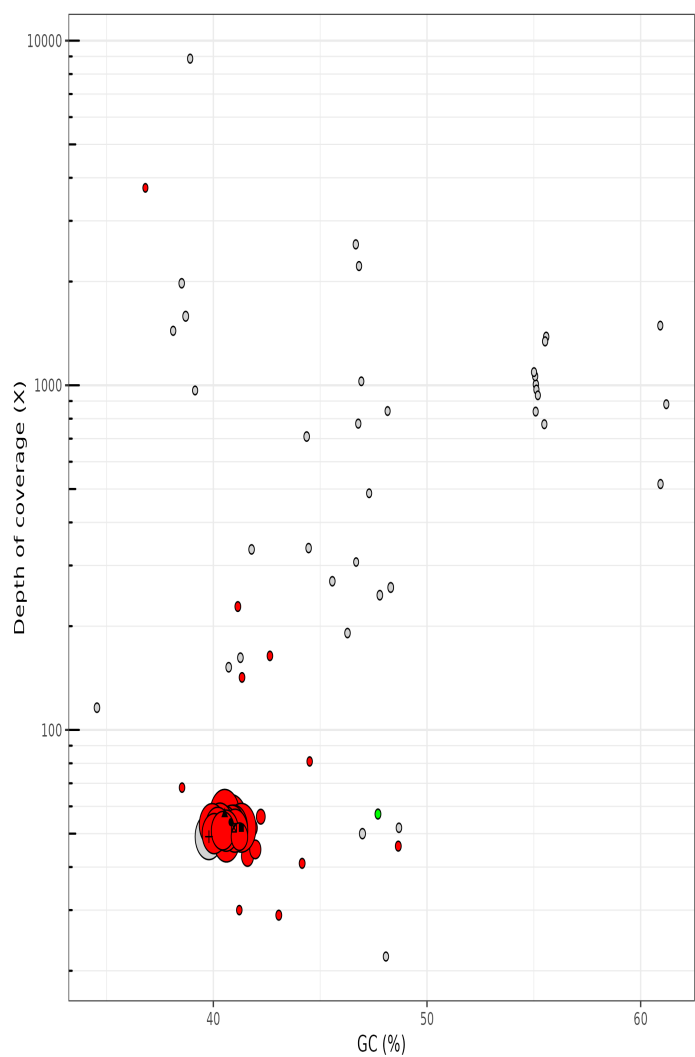


Distribution of k-mer counts coloured by their presence in reads/assemblies

Post-curation contamination screening



TAPAs summary Graph



- superkingdom
- Bacteria
 - Eukaryota
 - N/A
- Length (bp)
- 2.5e+07
 - 5.0e+07
 - 7.5e+07
- Longest sequences (bp)
- SUPER_1 - 95971612 (Eukaryota)
 - ▲ SUPER_2 - 73099444 (Eukaryota)
 - SUPER_3 - 71760907 (Eukaryota)
 - + SUPER_4 - 61813175 (N/A)
 - SUPER_5 - 59904133 (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	58	196

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: Sophie Layac

Affiliation: Genoscope

Date and time: 2025-08-02 09:45:54 CEST