

ERGA Assembly Report

v24.10.15

Tags: ATLASEa[INVALID TAG]

TxID	55702
ToLID	wpSthBoax2
Species	Sthenelais boa
Class	Polychaeta
Order	Phyllodocida

Genome Traits	Expected	Observed
Haploid size (bp)	514,312,242	514,930,757
Haploid Number	10 (source: ancestor)	11
Ploidy	2 (source: ancestor)	2
Sample Sex	Unknown	Unknown

EBP metrics summary and curation notes

Obtained EBP quality metric for collapsed: 6.7.Q41

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
- . Assembly length loss > 3% for collapsed

Curator notes

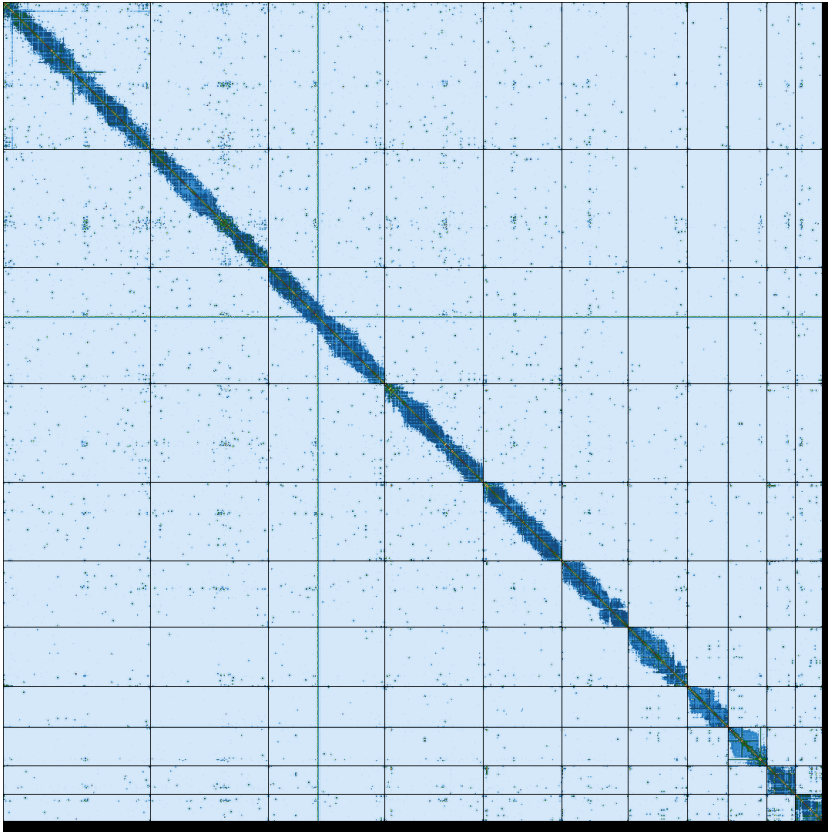
. Interventions/Gb: 166
. Contamination notes: ""
. Other observations: "The assembly of \'Sthenelais boa\' (wpSthBoax2) is based on 34X PacBio data and 173X Arima Hi-C data generated as part of the ATLASEa programme (<https://www.atlasea.fr>). The assembly process included the following steps: initial 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, 205 contigs were identified as contaminants (bacterial, archaeal, or viral), totaling 7.455 Mb (with the largest being 1.783 Mb). Additionally, 457 regions totaling 22.975 Mb (with the largest being 0.73 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, 34 haplotypic regions were removed, totaling 22.952Mb (with the largest being 6.359Mb). 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	538,426,045	514,930,757
GC %	41.02	41.02
Gaps/Gbp	523.75	557.36
Total gap bp	28,200	32,500
Scaffolds	313	270
Scaffold N50	61,854,699	61,135,444
Scaffold L50	4	4
Scaffold L90	9	9
Contigs	595	557
Contig N50	5,592,298	5,496,845
Contig L50	26	25
Contig L90	123	116
QV	41.9477	41.8634
Kmer compl.	73.678	72.0205
BUSCO sing.	97.7%	98.5%
BUSCO dupl.	1.8%	0.4%
BUSCO frag.	0.0%	0.1%
BUSCO miss.	0.5%	1.1%

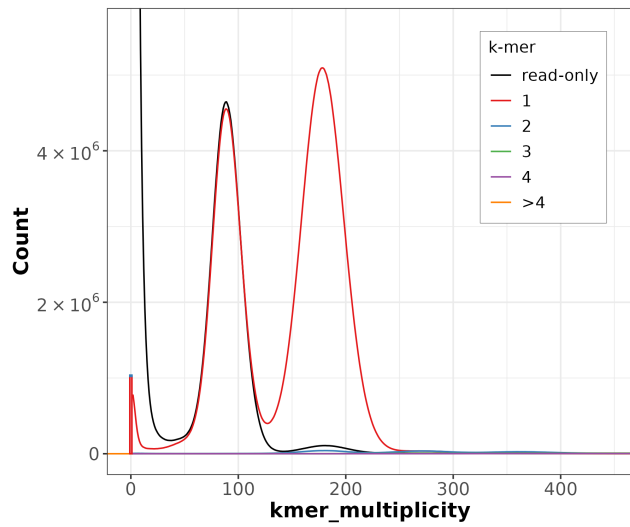
BUSCO: 6.1.0 (euk_genome_min, miniprot) / Lineage: lophotrochozoa_odb12.2 (genomes:63, BUSCOs:1329)

HiC contact map of curated assembly

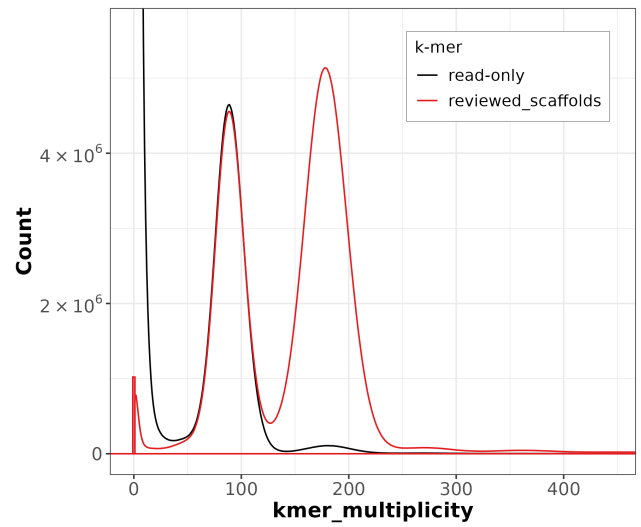


collapsed [\[LINK\]](#)

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



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	181

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*: 0.2.4
 - |_ *key param*: NA
- **PretextView**
 - |_ *ver*: 1.0.7
 - |_ *key param*: NA

Submitter: Simone Duprat

Affiliation: Genoscope

Date and time: 2026-09-28 16:32:39 CEST