ERGA Assembly Report

v24.10.15

Tags: ATLASea[INVALID TAG]

TxID	50731
ToLID	**fAbuSax1**
Species	Abudefduf saxatilis
Class	Actinopteri
Order	NA

Genome Traits	Expected	Observed
Haploid size (bp)	784,073,943	793,691,323
Haploid Number	19 (source: ancestor)	24
Ploidy	2 (source: ancestor)	2
Sample Sex	Unknown	Unknown

EBP metrics summary and curation notes

Obtained EBP quality metric for collapsed: 7.7.Q50

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

. Observed Haploid Number is different from Expected

Curator notes

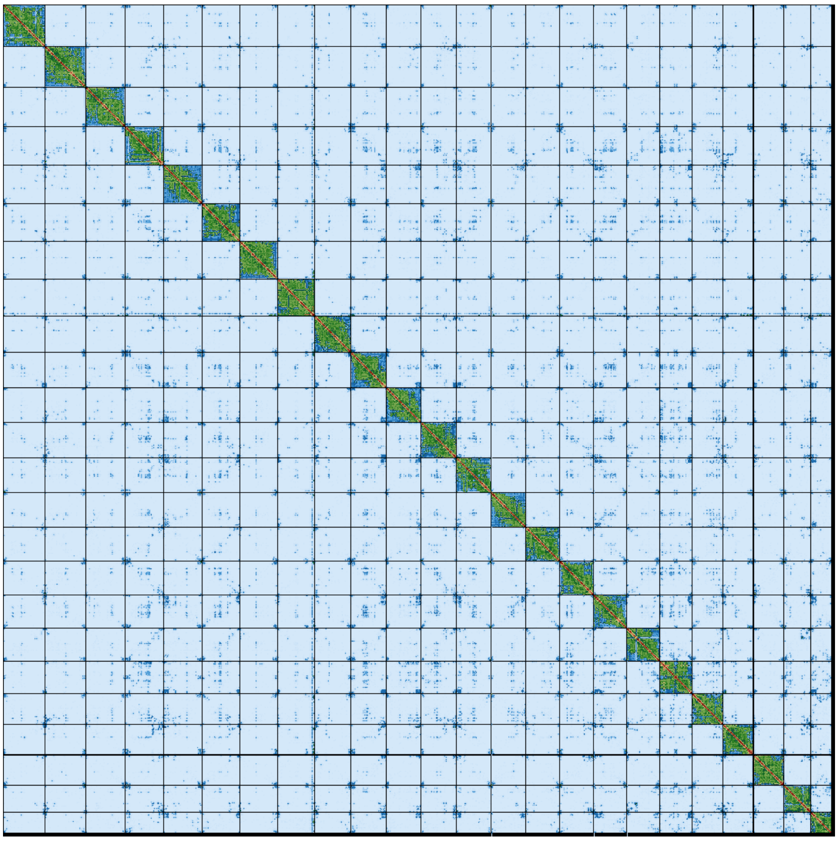
. Interventions/Gb: 81
. Contamination notes: ""
. Other observations: "The assembly of Abudefduf saxatilis (fAbuSax1) is based on 69X PacBio data and 154X Arima Hi-C data generated as part of the ATLASea programme (https://www.atlasea.fr). The assembly process included the following steps: initial PacBio 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, 1 contigs were identified as contaminants (bacterial, archaeal, or viral), totaling 0.032 Mb (with the largest being 0.032 Mb). Additionally, 83 regions totaling 12.06 Mb (with the largest being 3.954 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. 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	793,667,214	793,691,323
GC %	40.93	40.93
Gaps/Gbp	55.44	89.46
Total gap bp	4,600	12,000
Scaffolds	95	66
Scaffold N50	33,204,765	33,453,840
Scaffold L50	12	11
Scaffold L90	22	21
Contigs	136	137
Contig N50	21,421,956	21,421,956
Contig L50	16	16
Contig L90	41	41
QV	50.0225	50.0226
Kmer compl.	90.0967	90.0988
BUSCO sing.	97.4%	99.4%
BUSCO dupl.	0.2%	0.2%
BUSCO frag.	0.5%	0.0%
BUSCO miss.	1.9%	0.4%

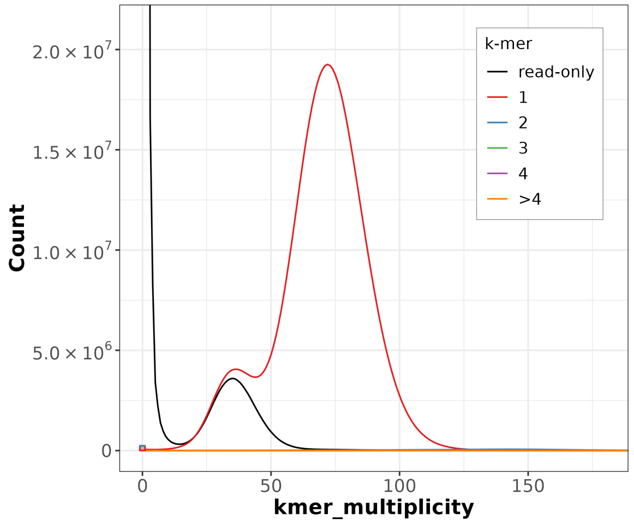
Warning! BUSCO versions or lineage datasets are not the same across results:
BUSCO: 5.8.2 (euk_genome_met, metaeuk) / Lineage: actinopterygii_odb12 (genomes:75, BUSCOs:7207)
BUSCO: 6.0.0 (euk_genome_min, miniprot) / Lineage: actinopterygii_odb12 (genomes:75, BUSCOs:7207)

HiC contact map of curated assembly

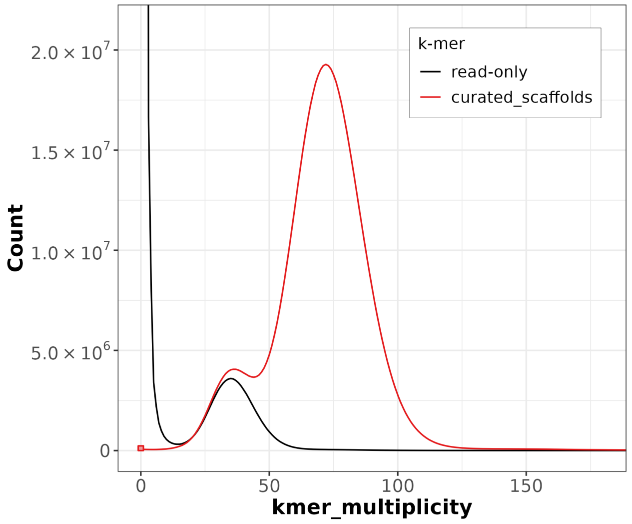


collapsed [LINK]

K-mer spectra of curated assembly

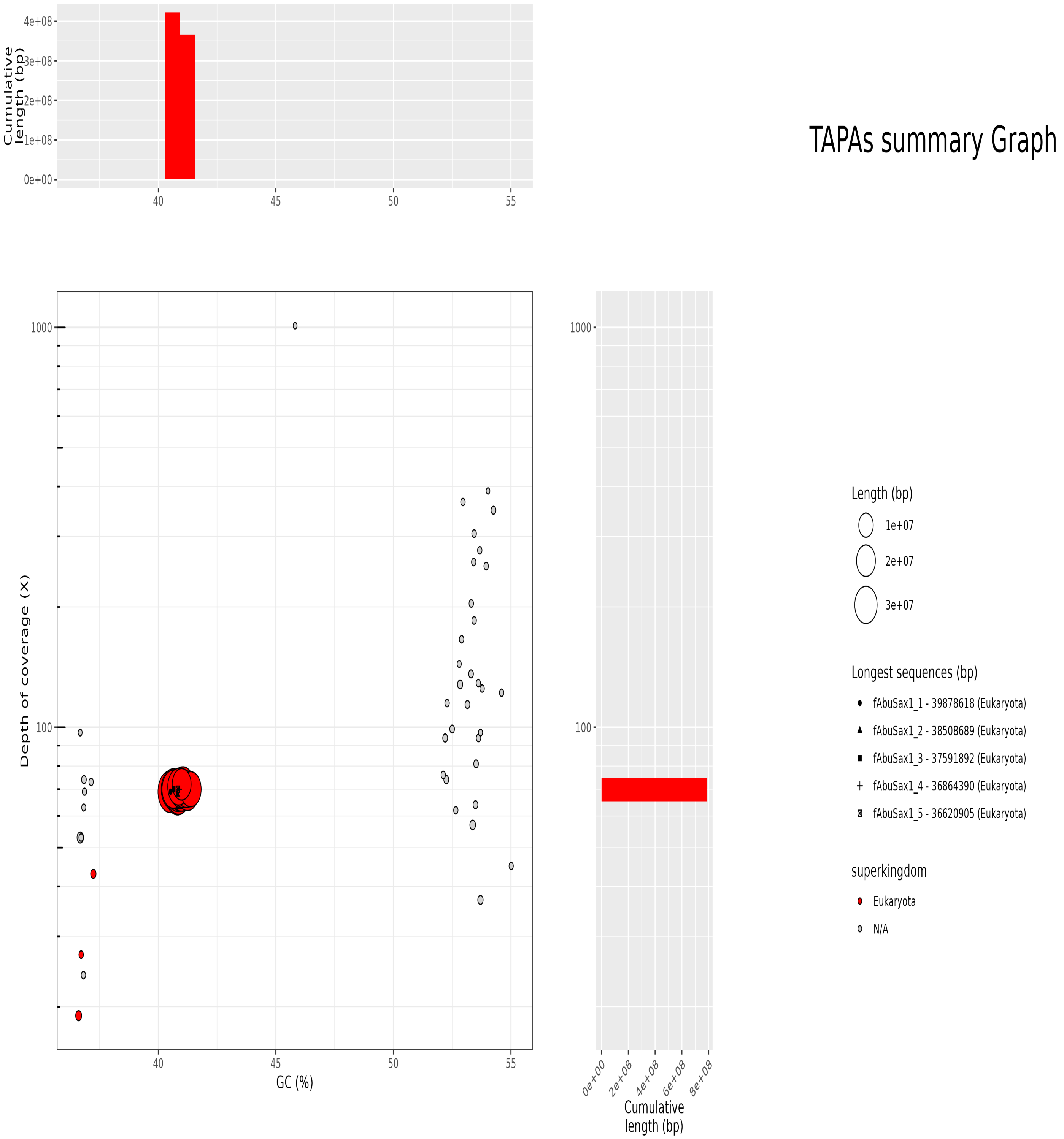


Distribution of k-mer counts per copy numbers found in asm



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

Post-curation contamination screening



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	69	154

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: Arnaud Couloux
Affiliation: Genoscope

Date and time: 2025-12-11 22:31:16 CET