ERGA Assembly Report

v24.10.15

Tags: Other[INVALID TAG]

TxID	100859
ToLID	**bVanVan3**
Species	Vanellus vanellus
Class	Aves
Order	Charadriiformes

Genome Traits	Expected	Observed
Haploid size (bp)	1,228,149,300	1,393,446,359
Haploid Number	38 (source: direct)	41
Ploidy	3 (source: ancestor)	2
Sample Sex	Unknown	Unknown

EBP metrics summary and curation notes

Obtained EBP quality metric for collapsed: 7.7.Q61

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

. Observed Haploid Number is different from Expected
. Observed Ploidy is different from Expected

. Kmer completeness value is less than 90 for collapsed

Curator notes

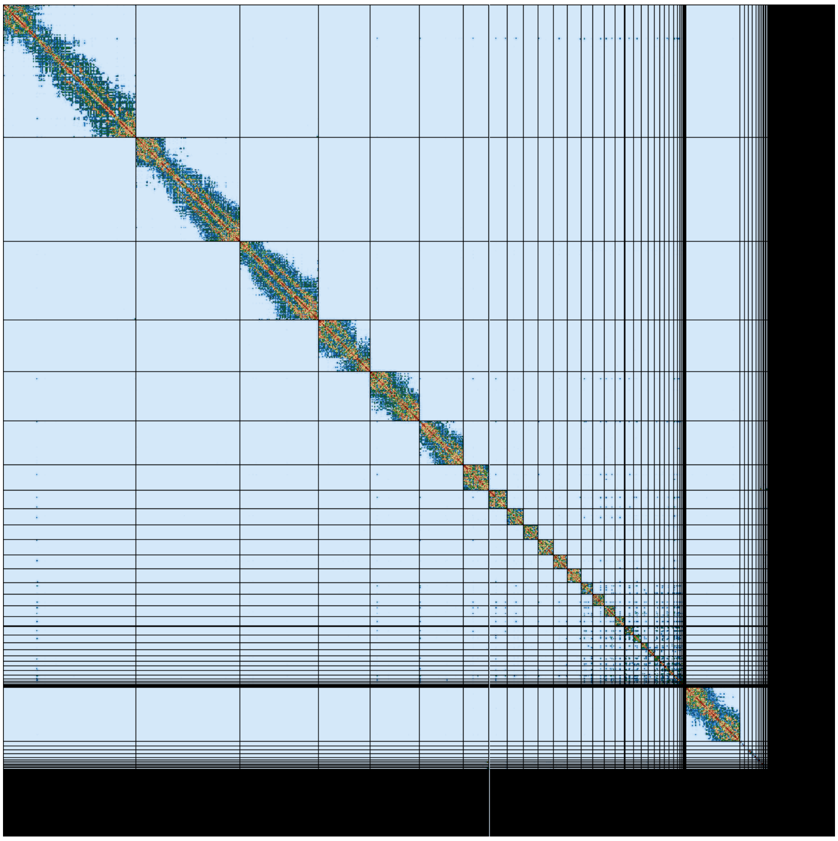
. Interventions/Gb: 12
. Contamination notes: ""
. Other observations: "The assembly of Vanellus vanellus (bVanVan3) is based on 70X of PacBio data and Arima Hi-Cdata generated as part of the ATLASea programme (https://www.atlasea.fr). The assembly process included the following steps: initial PacBio 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, 19 contigs were identified as contaminants (bacterial, archaeal, or viral), totaling 0.75 Mb (with the largest being 0.28 Mb). Additionally, 353 regions totaling 29 Mb were identified as haplotypic duplications and removed (with the largest being 4 Mb). The mitochondrial genome was assembled using OATK. Finally, the primary assembly was analyzed and manually improved using Pretext. During manual curation, 1 haplotypic regions and 0 contaminant sequences were removed, totaling 0.74 Mb (with the largest being 0.74 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	1,394,275,708	1,393,446,359
GC %	44.05	44.05
Gaps/Gbp	120.49	129.89
Total gap bp	16,800	19,800
Scaffolds	1,014	986
Scaffold N50	87,105,396	87,105,696
Scaffold L50	5	5
Scaffold L90	35	34
Contigs	1,182	1,167
Contig N50	22,717,496	22,783,031
Contig L50	18	17
Contig L90	92	92
QV	61.6665	61.6601
Kmer compl.	89.1362	89.1327
BUSCO sing.	97.4%	97.4%
BUSCO dupl.	0.3%	0.3%
BUSCO frag.	0.3%	0.3%
BUSCO miss.	2.0%	2.0%

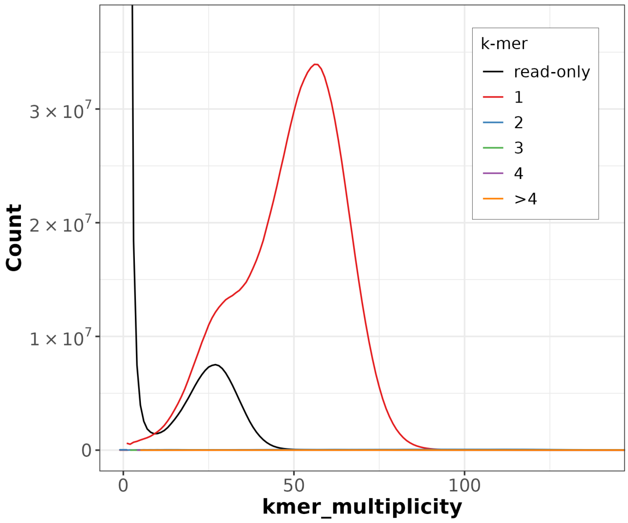
BUSCO: 5.4.3 (euk_genome_met, metaeuk) / Lineage: aves_odb10 (genomes:62, BUSCOs:8338)

HiC contact map of curated assembly

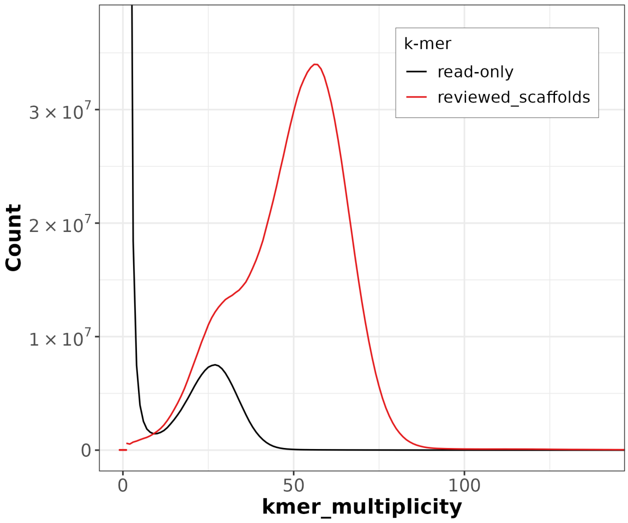


collapsed [LINK]

K-mer spectra of curated assembly

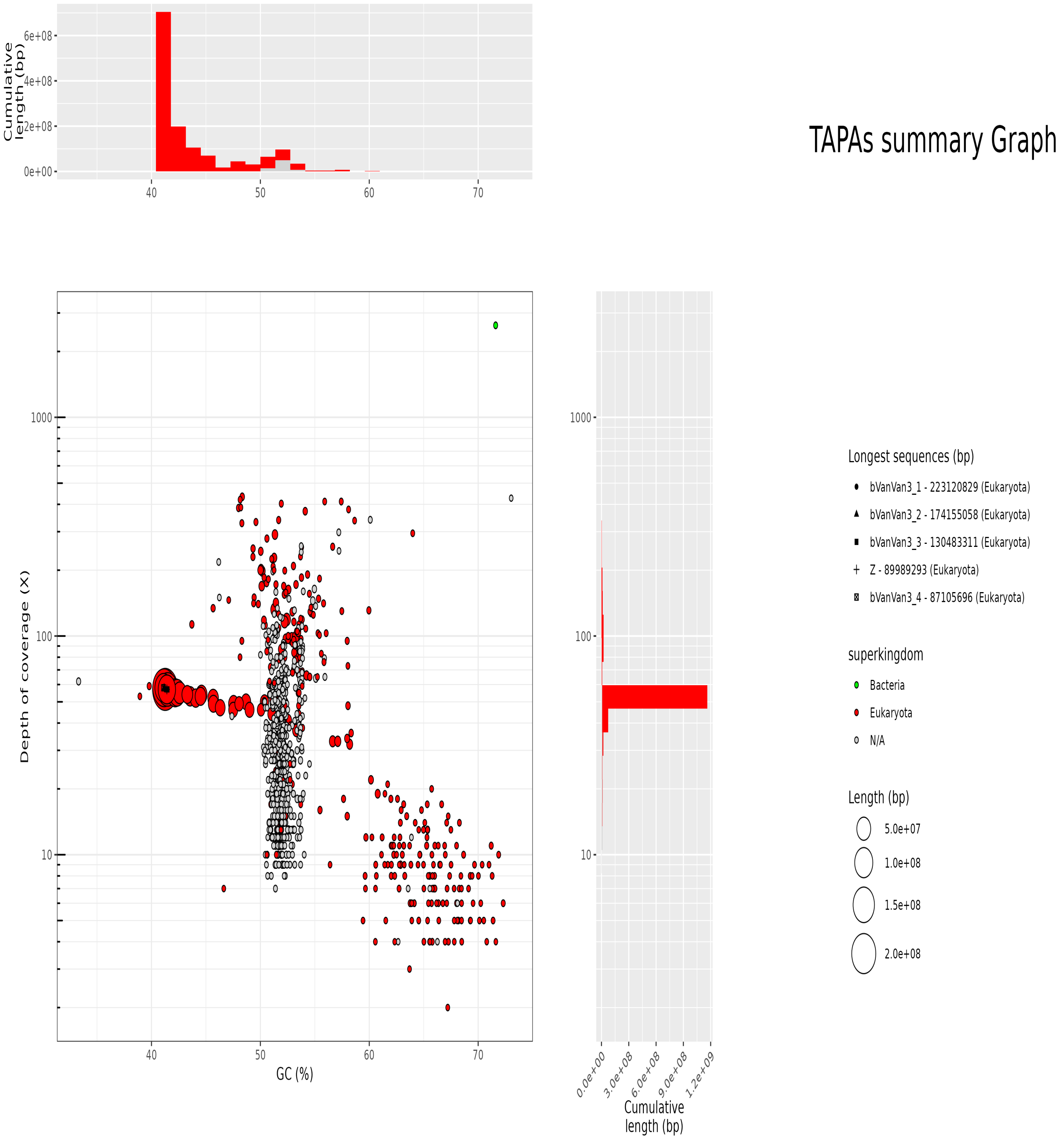


Distribution of k-mer counts per copy numbers found in asm



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	PACBIO Hifi	Omnic
Coverage	57	36

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: 2025-04-05 18:28:36 CEST