scientific data

Check for updates

OPEN

DATA DESCRIPTOR

De novo assembly at Chromosomal level of the Indigenous Venda chicken

A. H. Molotsi[1]✉, V. Ramothibedi[1], S. Mdyogolo[2], L. T. Nesengani[1], T. S. Tshilate[1], T. M. Masebe[2], N. L Hlongwane[1], R. M. Smith[2], C. A. Gill[1], A. Djikeng[1,3] & N. O. Mapholi[1]✉

Venda chickens are indigenous to South Africa and are known for their robustness traits (ability to grow, reproduce, and survive) in environments with dry, hot, and humid weather and food shortages. Understanding the genetic mechanisms that regulate robustness traits in Venda chicken is important for sustainable breeding and genetic improvement programs in the Southern African poultry industry. Therefore, this study aimed to produce a *de novo* assembly of a female Venda chicken genome at chromosomal level using long-read sequencing technology, complemented by Omni-C data. The assembled genome size is 1.09 Gb with the longest scaffold of 179 Mb. The assembly statistics resulted in a scaffold N50 of 90.4 Mb and contig N50 of 20MB, scaffold and contig L50 were 4, and the GC content 42%. The BUSCO completeness was 97.3% with complete and single-copy (S) genes of 8056. The genome annotation predicted 17891 protein-coding genes, with an average gene length of 4665.8 bp. This is the first chromosomal-level genome of the indigenous Venda chicken and will contribute to conservation and genetic improvement of the breed.

Background & Summary

Venda chickens are indigenous and native to the Limpopo province of South Africa. They are known to be well adapted to hot, subtropical conditions and harsh environments. Venda chickens can survive endemic diseases such as Newcastle disease, coccidiosis and other respiratory diseases without the routine use of vaccinations[1]. Studies also indicated a low prevalence of gastro-intestinal parasites in indigenous chicken in Limpopo and Kwa-Zulu Natal Provinces[1]. Furthermore, Venda chickens are known to exhibit excellent mothering ability and free-range scavenging[2] and are considered a dual-purpose breed utilized for both meat and egg production[3]. They have multi-colored plumage of white, black and red, with green feather tips and reach a mature body weight of about 2.0 kg for males and 1.4 kg for females over 18 weeks[4].

Despite the Venda chicken superior disease resistance traits, when compared to other commercial breeds; Venda chicken reach a slaughter weight of 1.1 kg at 10 weeks[5], while commercial Cobb reaches a slaughter weight of 3.4 kg at 8 weeks[6]. For egg production the Venda chicken produces 153 eggs over a 52-week cycle vs the White Leghorn that produces 279 eggs[7]. This has prompted the need to establish breeding programs to improve production of the well-adapted indigenous breeds, especially in the face of climate change. Notably, there is limited information available on the genetic makeup of the Venda chicken, particularly in relation to their adaptation mechanisms. Some efforts to understand their genetic structure have involved genetic characterization using microsatellites and SNP panels[8–10]. Selection signatures studies also indicated genomic regions linked to disease resistance, growth and reproduction traits in indigenous chickens[4].

However, the abovementioned approaches are limited in explaining the genomic architecture, which includes larger structural variants, that are often missing when using short read sequencing technology. This is because microsatellites cover a small portion of the genome, while SNP panels used in poultry were developed using commercial breeds, which often implies bias when it comes to indigenous breeds. Whole genome sequencing data could assist with the identification of genomic regions that control the expression of adaptation traits, as

[1]Department of Agriculture and Animal Health, College of Agriculture and Environmental Sciences, University of South Africa, Roodepoort, South Africa. [2]Department of Life and Consumer Sciences, College of Agriculture and Environmental Sciences, University of South Africa, Roodepoort, South Africa. [3]International Livestock Research Institute, Nairobi (Kenya) and Addis Ababa (Ethiopia), Nairobi, Kenya. ✉e-mail: molotah@unisa.ac.za; maphon@unisa.ac.za

Fig. 1 A male (left) and female (right) Venda chicken. Photo credits: Annelin Molotsi.

Sequencing platform	Sequencing reads	Total bases (bp)	Read length (mean, bp)	Read Quality
PacBio (HiFi)	5,922,790	63,461,273,932	11910	Q36
Novaseq 6000 (Omni-C)	253,072,230	38,000,000,000	35–151	Q36

Table 1. Summary statistics of the sequencing data, mean read length and phred quality scores.

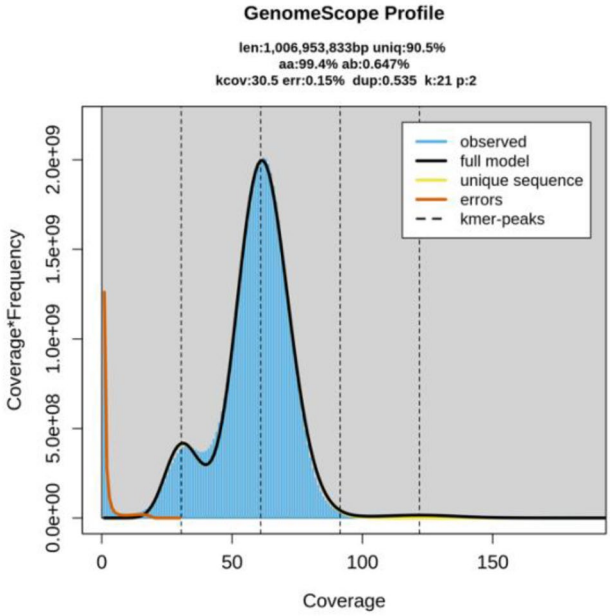

Fig. 2 The Genomescope profile for the indigenous Venda chicken. The figure describes the estimated genome size (len), the heterozygosity (ab), and homozygosity (aa), the number of duplicates (dup), the k-mer coverage (kcov), the user specified K-mer size (k), reads that contained errors (err), and the ploidy (p).

well as SNP markers that are unique for indigenous chickens. This will further assist in the implementation of genomic selection in breeding and conservation programs for Venda chickens.

The current publicly available chicken genomes are skewed towards the commercial exotic and Native Asian chicken breeds, leaving the African chicken breeds underrepresented. The available genome assemblies include the Red Jungle Fowl, which is the ancestor of all domesticated chickens, Galgal6[11] and is known to have 6 missing microchromosomes (MIC) in its assembly. Most of the chicken housekeeping genes and immune pathways are linked to MIC and subtelomeric regions[12,13]. The GRCg7b[14] broiler chicken has been assembled at the chromosomal level, including all the microchromosomes, and is currently used as the reference genome.

However, it is crucial to have the Venda chicken genome assembled to its complete chromosomal composition so that we can have full insight into the genes that play a role in immunity and adaptation. Here, we constructed a high-quality chromosomal-level genome assembly of a female Venda chicken using PacBio Hi-Fi long reads and Omni-C sequencing. This high-quality genome is a valuable genetic resource that can be used to further study the robustness and adaptation traits of Venda chickens.

Statistics without reference	Draft Primary Contig Assembly	Draft Alternate Contig Assembly	Purged Primary Contig Assembly	Purged Alternate Contig Assembly
Number of contigs	1,174	3,112	640	2,592
Largest contig	73,928,295	5,524,578	73,928,295	5,524,578
Total length	1,222,538,838	964,523,839	1,091,172,697	1,025,109,615
Contig N50 (Mb)	19	1.2	20	1.2
L50 contig count	16	229	13	244
GC content (%)	43	42	42	42

Table 2. Statistics of contig assembly before scaffolding.

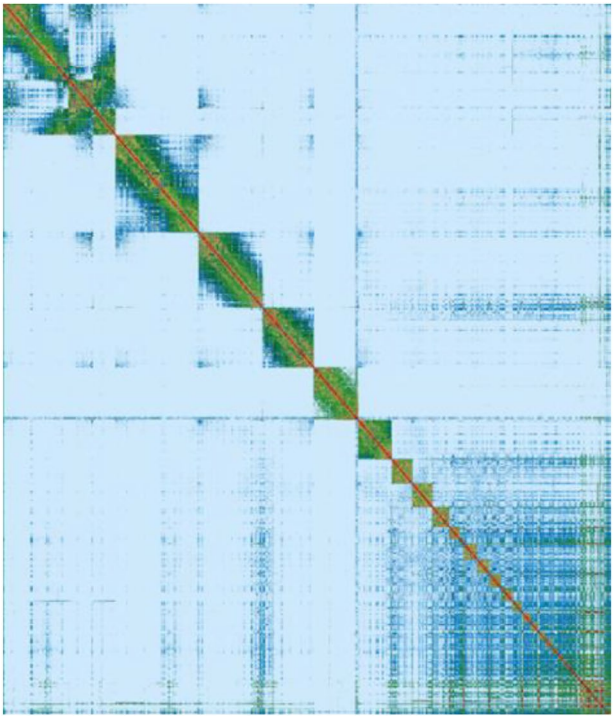

Fig. 3 Pretext map after of the scaffolded genome for Venda chicken A after manual curation.

Statistics	Venda chicken (GCA_055689475.1)[39]	Broiler chicken GRCg7b (GCA_016699485.1)[42]	Red Jungle Fowl -Reference genome Galgal6 (GCA_000002315.5)[43]	Ross chicken (GCA_027557775.1)[44]	Naked neck chicken (GCA_024686465.1)[45]
Genome Size	1.1 Gb	1.1 Gb	1.1 Gb	1.1 Gb	1.0 Gb
Number of chromosomes	41	41	34	41	—
Number of scaffolds	366	213	524	246	1,830
Scaffold N50 (Mb)	90.4	90.9	20.8	91.6	16.4
Scaffold L50	4	4	12	4	19
Number of contigs	562	676	1,402	812	13,837
Contig N50 (Mb)	20	18.8	17.7	5.5	981
GC percent (%)	42	42	42	42	42
Genome coverage	63X	102X	82X	32.3X	53X
Assembly level	Chromosomal	Chromosomal	Chromosomal	Chromosomal	Scaffold

Table 3. Assembly statistics of Venda chicken after manual curation in comparison to other chicken breed assemblies accessed from the NCBI database.

Methods

Ethics statement. The procedures for animal handling, sample collection and all research related activities were approved by the University of South Africa, Animal Research Ethics Committee (AREC-100818-024) and the Department of Agriculture Land Reform and Rural Development (DALRRD) under section 20 of the Animal

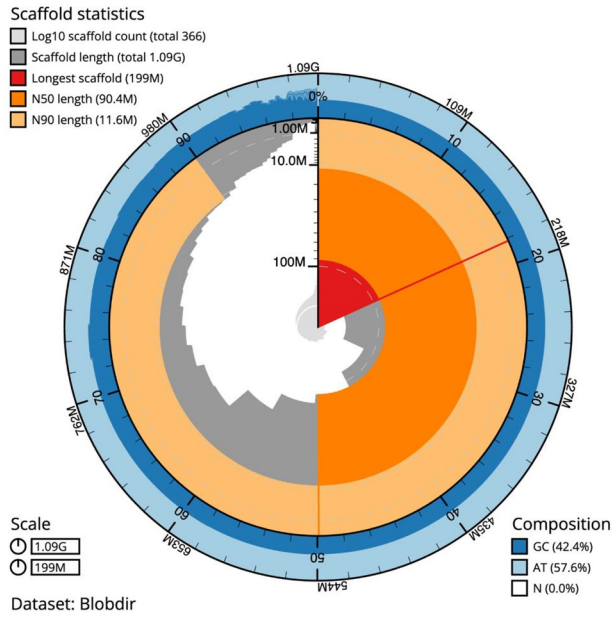

Fig. 4 Snailplot showing the scaffold statistics for the Venda chicken genome.

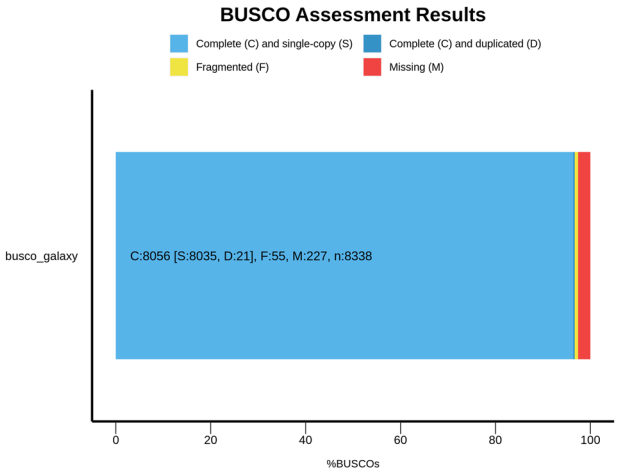

Fig. 5 The BUSCO analysis after scaffolding. The BUSCO analysis of the primary assembly for Venda chicken, where the dark and light blue represent the complete single copy and duplicates, respectively. Yellow represents fragmented genes, and red represents missing genes.

Diseases Act 1984 (Act 35 of 84), reference number: 12/11/1/1/23 (6508 AC).

Sample collection and sequencing. One purebred female Venda chicken (Fig. 1) was selected from a smallholder farmer located in Thohoyandou, Limpopo province (S 22° 57′ 4.788″, E 30° 29′ 7.434″). Blood was collected from the brachial wing vein using an EDTA vacutainer tube with the assistance of a veterinarian. The samples were immediately placed on dry ice, transported to the laboratory, and stored at −80 °C until further processing. Genomic DNA was extracted from 200 µL of whole blood using the Nanobind protocol for high-molecular-weight (HMW) DNA extraction. The SMRTBell® prep kit 3.0 was used to prepare the sequencing library. Sequencing was done using PacBio Sequel IIe with a total HiFi data output of 32.3 Gb and PacBio Revio with a total data output of 31.1 Gb thus providing a total coverage of 63X. To generate Omni-C data for polishing the genome to chromosomal level, we employed Dovetail®Omni-C® kit for Omni-C library preparation. The resulting library was sequenced on NovaSeq 6000 system (Illumina®) which generated an output data of 19.3 Gb for forward and reverse reads, respectively. Initial sequence quality control processing was performed on the SMRTlink Software v11.0 (PacBio) as well as FastQC v0.12.1[15]. The sequencing data from each platform is summarized below in Table 1.

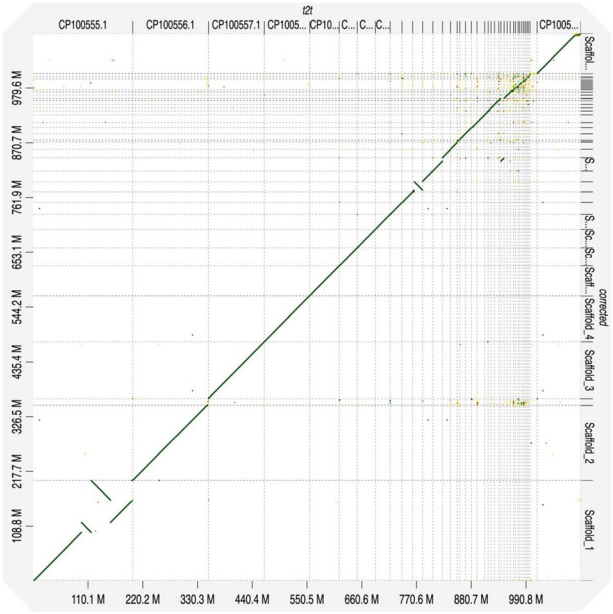

Fig. 6 Co-linearity plot of the curated Venda chicken genome with the T2T Huxu genome GCA_024206055.2[28,46].

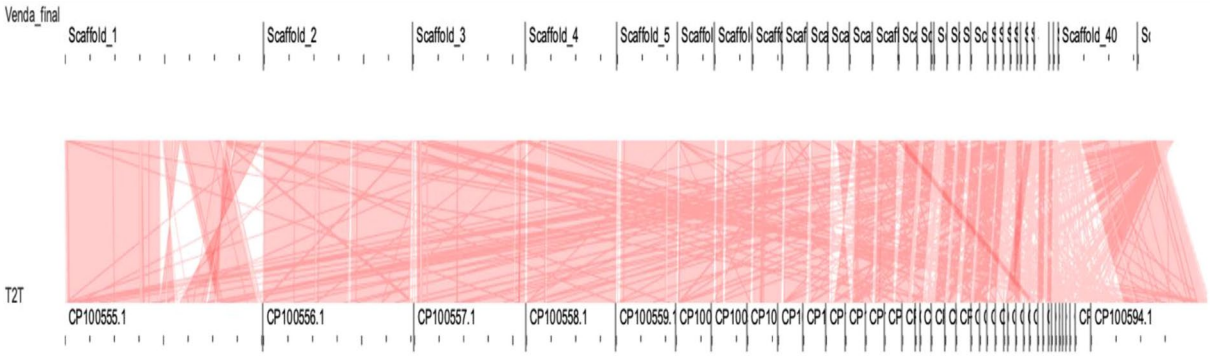

Fig. 7 Synteny plot of the Venda chicken genome against the T2T Huxu genome.

Genome assembly. The genome assembly was carried out through the vertebrate genome project (VGP) workflows on the Galaxy platform[16]. These workflows start with utilising genomeScope2[17] for genome characterization and estimating genome size using HiFi reads, which resulted in 1.0 Gb estimated genome for Venda chicken from k-mer (21) analysis (Fig. 2). Mercury[18] was utilized to estimate the error rate of both the primary and alternate assembly.

This was followed by the genome assembly using HiFiasm[19] on Hi-C mode assembly using both the HiFi reads and Omni-C reads. The assembly was then subjected to purging to remove haplotypic duplication and contig overlaps in the draft assembly using Purge_ dups (v1.2 .6)[20]. Thereafter, the genome was evaluated using the gfastats v 1.3.11 tool[21], which resulted in a genome size of 1.09 GB, with contig and scaffold N50 of 20 Mbp and 48Mbp, respectively (Table 2). The purged primary assembly was further subjected to scaffolding with Hi-C reads, making use of the BWA-MEM2v2.3[22] and YAHS 1.2a.2)[23] to reach the chromosomal level assembly. Following scaffolding, decontamination was done using Kraken2 v2.1.3 databases[24], which was further followed by manual curation using the PretextView v0.2.5 software[25]. The pretext maps before and after manual curation are shown in Fig. 3. After curation, the scaffold N50 was 90.4 Mb, contig N50 was 20 Mb and the number of scaffolds decreased to 366 (Table 3). Blobtoolkit and Blobtoolkit viewer v4.1.0[26] was used to generate a snailplot of scaffold statistics (Fig. 4).

Genome quality assessment. The Benchmarking Universal Single-Copy Orthologs (BUSCO) completeness before scaffolding was 97.3% with 2.7% reported as missing genes (Fig. 3). There is a higher number of conserved single-copy genes observed within the Venda chicken of 7719. After scaffolding with Hi-C data the BUSCO completeness remained at 97.3% with 8056 complete and single-copy (S), and the conserved single-copy genes were 8035 (Fig. 5). The number of duplicates reduced from 340 to 21, which indicates that the erroneous assembly of haplotypes were resolved.

Scaffold name	Chromosome number	Length (bp)
Scaffold 1	Chromosome_1	198,569,007
Scaffold 2	Chromosome_2	148,318,877
Scaffold 3	Chromosome_3	113,453,618
Scaffold 4	Chromosome_4	90,376,613
Scaffold 5	Chromosome_5	59,407,471
Scaffold 6	Chromosome_6	35,531,262
Scaffold 7	Chromosome_7	36,563,567
Scaffold 8	Chromosome_8	29,836,936
Scaffold 9	Chromosome_9	23,776,308
Scaffold 10	Chromosome_10	20,879,195
Scaffold 11	Chromosome_11	20,046,691
Scaffold 12	Chromosome_12	21,445,298
Scaffold 13	Chromosome_13	26,240,519
Scaffold 14	Chromosome_14	17,149,578
Scaffold 15	Chromosome_15	14,119,264
Scaffold 16	Chromosome_16	3,130,056
Scaffold 17	Chromosome_17	2,763,913
Scaffold 18	Chromosome_18	11,588,453
Scaffold 19	Chromosome_19	12,221,156
Scaffold 20	Chromosome_20	10,023,345
Scaffold 21	Chromosome_21	15,384,670
Scaffold 22	Chromosome_22	7,525,834
Scaffold 23	Chromosome_23	6,426,929
Scaffold 24	Chromosome_24	7,259,433
Scaffold 25	Chromosome_25	6,793,231
Scaffold 26	Chromosome_26	3,596,416
Scaffold 27	Chromosome_27	6,526,829
Scaffold 28	Chromosome_28	5,527,825
Scaffold 29	Chromosome_29	3,463,215
Scaffold 30	Chromosome_30	734,165
Scaffold 31	Chromosome_31	1,931,408
Scaffold 32	Chromosome_32	1,764,908
Scaffold 33	Chromosome_33	1,665,007
Scaffold 34	Chromosome_34	2,097,909
Scaffold 35	Chromosome_35	2,098,008
Scaffold 36	Chromosome_36	1,831,508
Scaffold 37	Chromosome_37	2,863,912
Scaffold 38	Chromosome_38	600,450
Scaffold 39	Chromosome_39	4,861,922
Scaffold 40	Chromosome_Z	79,620,663
Scaffold 41	Chromosome_W	12,055,750

Table 4. Scaffold names and corresponding chromosome numbers and lengths of the Venda chicken genome.

Genome alignment. To further assess how the genome aligns compared to the T2T reference genome, we used D-genies software[27] to do an alignment of the Venda chicken genome against a high-quality chicken genome assembly[28] (Fig. 6). We then used JBrowse2 to create a synteny plot between the high-quality Huxu[28] genome and the Venda chicken (Fig. 7). This informed the manual curation process to ensure that we correct mis-assemblies. We were able to assign the Z and W chromosomes to scaffolds 40 and 41, respectively (Table 4).

Telomeric regions. There were 67 telomeric regions identified with an average length of 7966,3 kb with scaffold 32 having the longest telomeric region of 19495 kb. This can be seen in Fig. 8. Telomere length is correlated with cellular stability, with longer lengths being beneficial for cellular longevity[29].

Genome annotation. Repeat elements in the Venda chicken genome were found using RepeatModeler v.0.9[30] and masked using RepeatMaskerv4.1.5[31] as the initial stage of genome annotation. RepeatModeler v0.9[32], which includes tools like RepeatScout v1.0.6, RECON v1.5.0, and TRF v4.09, was used to undertake *de novo* repeat modeling to define Venda chicken-specific repeated elements. The Perl script "one_code_to_find_them_all"[33] was used to further annotate the resultant repeats. Repeat analysis revealed that about 10.9% of the genome consists of repetitive elements. Retroelements are the most abundant among the repeats. Among these, the most

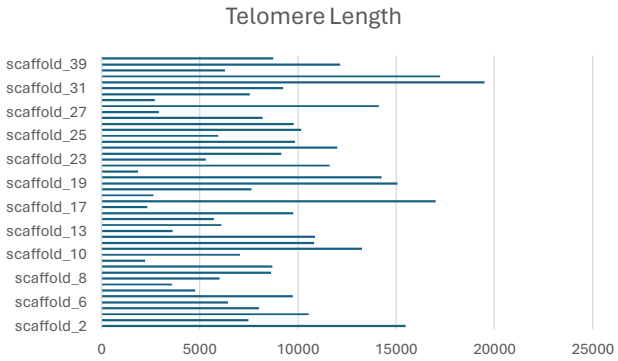

Fig. 8 The number of Telomeric regions present within the Venda chicken genome after scaffolding with Hi-C data.

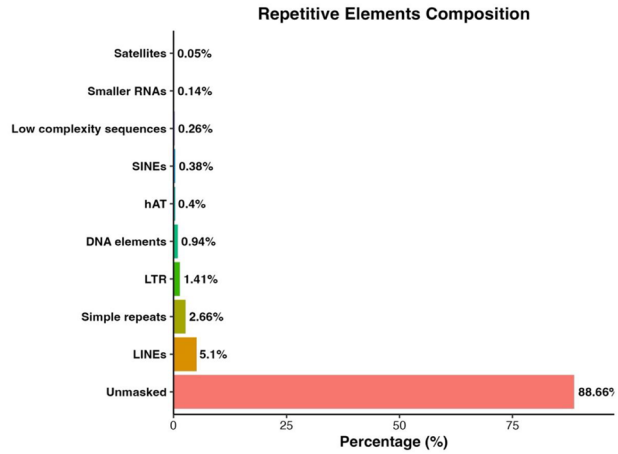

Fig. 9 The repetitive elements in the Venda chicken genome.

Structural Annotation	Value
Number of protein coding genes	17891
Average gene length (bp)	4665.8
Number of exons per gene	20.31
Average exon length (bp)	100.76
Average intron length (bp)	240.8
Number of transcripts	30664
Average transcript length (bp)	959.6

Table 5. Genome annotation of the Venda chicken providing information on the structural annotation.

abundant are Long Interspersed Nuclear Elements (LINEs) 5.01%, followed by Long Terminal Repeats (LTR) elements 1.41% and Short Interspersed Nuclear Elements (SINEs) 0.38%. DNA elements contributed a smaller percentage 0.94%, with contributions of hAT-like elements 0.40% comprising unclassified repeats and may represent species-specific or highly degenerate sequences. Other repeat classes are small RNAs 0.14%, simple repeats 2.66%, satellites 0.05%, and low-complexity sequences (0.26%) (Fig. 9). These results are comparable to the Korean Long tailed chicken[34]. Such a repeat composition represents the structural content of the genome and forms the foundation for exploring its evolutionary history and regulatory complexity.

The masked genome was annotated using Braker3[34,35] hidden Markov chain model to predict the protein-coding gene structures. The results indicated 17891 protein-coding genes, with an average exon length of 100.8 bp and an intron length of 240.8 bp. The number of transcripts was 30664, with an average length of 959.6 as shown in Table 5.

This is the first chromosomal-level assembly of the South African Venda chicken with a high BUSCO completeness and contiguity level. We were able to assign the sex chromosomes and microchromosomes during manual curation. This genome serves as an important genetic resource for the South African poultry industry, as it can be used in further downstream analysis, especially in the identification of genetic variants important for adaptation and production.

Assembly	Unique K-mers	Common k-mers	QV	Error rate
Assembly_01	18644	1222515358	61.3893	7.26221e07
Assembly_02	6981	964461599	64.6259	3.44679e07
Both	25625	2186976957	62.534	5.5796e07

Table 6. Assembly quality as determined by Mercury, including the Unique k-mers, QV and error rate.

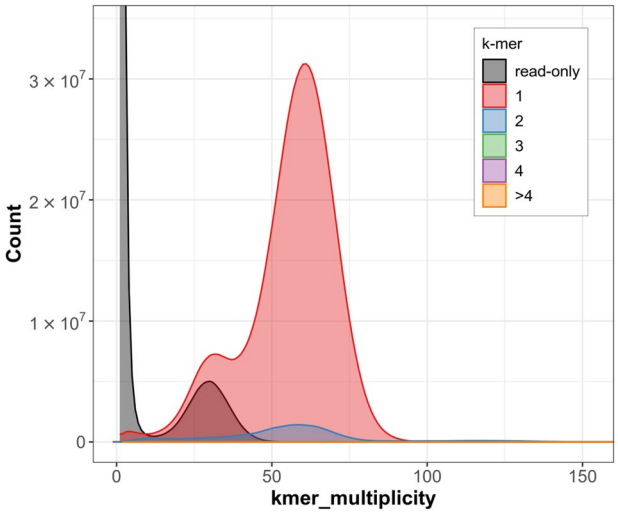

Fig. 10 Kmer multiplicity figure.

Data Record

The raw sequencing data files are available on NCBI under the Bioproject PRJNA1227266, under the SRA section as three data files. The PacBio HiFi reads is in fasta format (SRR35288173[36] -PacBio_SMRT, Revio and SRR35288174[37] -PacBio_SMRT, Sequell II) and the Hi-C reads in fasta format (SRR35288172)[38]. Furthermore, the primary assembly[39] is available on Genbank and NCBI and the alternate assembly[40] files in fasta format are available on Figshare. The genome annotation data[41] is available on Figshare in GTF format.

Technical Validation

DNA sample quality. The High molecular weight DNA molecules were quantified using Qubit 2.0 Fluorometer with the Qubit 1X dsDNA high sensitivity assay kit. DNA fragment length was determined using Agilent Fragment Analyzer to confirm DNA contentration obtained from Qubit. DNA concentration value of 80 ng/ul and fragment lengths on average of 15–20 bp were used for further library preparations. A minimum of 1ug DNA of was used on the flow cell for Sequel IIe and 2 ug DNA for the PacBio Revio sequencing with 24-hour movie time to generate HiFi reads.

Sequencing data assessment. The Phred score for all the sequencing data was above Q30, which is an indication that the base calling rate was more than 90% accurate. This is also provided in Table 1.

Assessment of genome assembly quality. The BUSCO lineage Aves_odb10 was used for the Benchmarking Universal Single-Copy Orthologs (BUSCO) and showed a high completeness of 97.3% for Venda chicken with complete and single-copy (S) of 8056 after scaffolding. The error rate for both the primary and alternate assembly, as calculated through mercury was 5.5796×10^{-7},with a solid k-mer in reads completeness percentage of 99.8% (Table 6 and Fig. 10).

Data availability

The completed genome assembly (GCA_055689475.1)[39] and raw data for Venda Chicken were submitted to the National Centre for Biotechnology Information (NCBI) under the Bioproject PRJNA1227266. The following SRA accession numbers: SSR 35288173, SRR 35288173, SRR 35288172 are for the raw data files.

Code availability

All the analyses done in the current study were processed by employing the VGP pipeline, pretext view for editing the omni-C contact map and RepeatModeler and RepeatMasker for identification of repetitive elements and masking of these elements and Braker3 for annotation. All the commands and pipelines were executed following the manual and protocols of the corresponding bioinformatics software. In all the workflows, unless mentioned and where necessary, we used the default parameters.

Received: 7 October 2025; Accepted: 7 April 2026;
Published online: 20 April 2026

References

1. Malatji, D. P., Tsotetsi, A. M., van Marle-Koster, E. & Muchadeyi, F. C. A description of village chicken production systems and prevalence of gastrointestinal parasites: Case studies in Limpopo and Kwazulu-Natal provinces of south Africa. *Onderstepoort Journal of Veterinary Research*. **83**, https://doi.org/10.4102/ojvr.v83i1.968 (2016).
2. Van Marle-Köster, E. CNH. Phenotypic characterisation of native chicken lines in South Africa. *Animal Genetic Resources Information* **29**, 71–78 (2001).
3. Manyelo, T. G., Selaledi, L., Hassan, Z. M. & Mabelebele, M. Local chicken breeds of Africa: Their description, uses and conservation methods. *Animals* **10**(12), 1–18, https://doi.org/10.3390/ani10122257 (2020).
4. Mogano, R. R., Mpofu, T. J., Mtileni, B. & Hadebe, K. South African indigenous chickens' genetic diversity, and the adoption of ecological niche modelling and landscape genomics as strategic conservation techniques. *Poult Sci. Elsevier Inc*. 104(1), https://doi.org/10.1016/j.psj.2024.104508 (2025).
5. Norris, D. & Ngambi, J. W. Genetic parameter estimates for body weight in local Venda chickens. *Trop Anim Health Prod.* **38**(7-8), 605–609, https://doi.org/10.1007/s11250-006-4420-6 (2006).
6. Sakomura, N. K., Marcato, S. M., Munari, D. P., Freitas, E. C. & Fernandes, J. B. K. *Growth Curves and Body Nutrients Deposition on Two Broiler Chickens Strains* (2006).
7. Grobbelaar, J. A. N., Sutherland, B. & Molalakgotla, N. M. Egg production potentials of certain indigenous chicken breeds from South Africa. *Animal Genetic Resources/Ressources génétiques animales/Recursos genéticos animales* **46**, 25–32, https://doi.org/10.1017/s2078633610000664 (2010).
8. Mtileni, B. J. *et al*. Genetic diversity and conservation of South African indigenous chicken populations. *Journal of Animal Breeding and Genetics* **128**, 209–218, https://doi.org/10.1111/j.1439-0388.2010.00891.x (2011).
9. Khanyile, K. S., Dzomba, E. F. & Muchadeyi, F. C. Population genetic structure, linkage disequilibrium and effective population size of conserved and extensively raised village chicken populations of Southern Africa. *Front Genet*. **5**, https://doi.org/10.3389/fgene.2015.00013 (2015).
10. Mtileni, B., Dzama, K., Nephawe, K. & Rhode, C. Estimates of effective population size and inbreeding in South African indigenous chicken populations: implications for the conservation of unique genetic resources. *Trop Anim Health Prod.* **48**, 943–950, https://doi.org/10.1007/s11250-016-1030-9 (2016).
11. Bellott, D. W. *et al*. Avian W and mammalian Y chromosomes convergently retained dosage-sensitive regulators. *Nat Genet.* **49**(3), 387–394, https://doi.org/10.1038/ng.3778 (2017).
12. Mcqueen, H. A., Siriaco, G. & Bird, A. P. *Chicken Microchromosomes Are Hyperacetylated, Early Replicating, and Gene Rich*. www.genome.org
13. Li, M. *et al*. De Novo Assembly of 20 Chicken Genomes Reveals the Undetectable Phenomenon for Thousands of Core Genes on Microchromosomes and Subtelomeric Regions. *Mol Biol Evol*. 39(4). https://doi.org/10.1093/molbev/msac066 (2022).
14. Smith, J. *et al*. Fourth Report on Chicken Genes and Chromosomes 2022. *Cytogenet Genome Res.* **162**(8-9), 405–528, https://doi.org/10.1159/000529376 (2023).
15. Andrews. FastQC: A Quality Control Tool for High Throughput Sequence Data (2015).
16. Abueg, L. A. L. *et al*. The Galaxy platform for accessible, reproducible, and collaborative data analyses: 2024 update. *Nucleic Acids Res.* **52**(W1), W83–W94, https://doi.org/10.1093/nar/gkae410 (2024).
17. Cheng, H. *et al*. Haplotype-resolved assembly of diploid genomes without parental data. *Nat Biotechnol.* **40**, 1332–1335, https://doi.org/10.1038/s41587-022-01261-x (2022).
18. Rhie, A. Merqury. In GitHub repository. GitHub (2020).
19. Cheng, H., Concepcion, G. T., Feng, X., Zhang, H. & Li, H. Haplotype-resolved de novo assembly using phased assembly graphs with hifiasm. *Nat Methods* **18**, 170–175, https://doi.org/10.1038/s41592-020-01056-5 (2021).
20. Guan, D., McCarthy, S.A., Wood, J., Howe, K., Wang, Y. & Durbin, R. Identifying and removing haplotypic duplication in primary genome assemblies. *Bioinformatics*. 39, 2896-2898, doi: 10.1093/bioinformatics/btaa025. (2020).
21. Formenti, G. *et al*. Gfastats: Conversion, evaluation and manipulation of genome sequences using assembly graphs. *Bioinformatics* **38**, 4214–4216, https://doi.org/10.1093/bioinformatics/btac460 (2022).
22. Vasimuddin, M., Misra, S., Li, H. & Aluru, S. Efficient Architecture-Aware Acceleration of BWA-MEM for Multicore Systems. In: *2019 IEEE International Parallel and Distributed Processing Symposium (IPDPS)*. 314-324. https://doi.org/10.1109/IPDPS.2019.00041 (2019).
23. Zhou, C., McCarthy, S. A. & Durbin, R. YaHS: yet another Hi-C scaffolding tool. *Bioinformatics*. 39(1). https://doi.org/10.1093/bioinformatics/btac808 (2023).
24. Wood, D. E., Lu, J. & Langmead, B. Improved metagenomic analysis with Kraken 2. *Genome Biol*. 20(1). https://doi.org/10.1186/s13059-019-1891-0 (2019).
25. Harry, E. PretextView (Paired REad TEXTure Viewer): A desktop application for viewing pretext contact maps. Published online 2022.
26. Challis, R., Richards, E., Rajan, J., Cochrane, G. & Blaxter, M. BlobToolKit - interactive quality assessment of genome assemblies. *G3: Genes, Genomes, Genetics* **10**(4), 1361–1374, https://doi.org/10.1534/g3.119.400908 (2020).
27. Cabanettes, F. & Klopp, C. D-GENIES: dot plot large genomes in an interactive, efficient and simple way. *PeerJ.* **6**, e4958, https://doi.org/10.7717/peerj.4958 (2018).
28. Huang, Z. *et al*. Evolutionary analysis of a complete chicken genome. *Proc Natl Acad Sci USA*. 120(8), https://doi.org/10.1073/pnas.2216641120 (2023).
29. Yu, H. J., Byun, Y. H. & Park, C. K. Techniques for assessing telomere length: A methodological review. *Comput Struct Biotechnol J. Elsevier B.V.* **23**, 1489–1498, https://doi.org/10.1016/j.csbj.2024.04.011 (2024).
30. Flynn, J. M. *et al*. Repeatmodeler2 for automated genomic discovery of transposable element families. *PNAS* **117**, 9451–9457, https://doi.org/10.1186/s13059-018-1577-z (2020).
31. Smit, A, H. R. & GP. RepeatMasker Open-4.0. (Published online 2015).
32. Bailly-Bechet, M., Haudry, A. & Lerat, E. "One code to find them all": A perl tool to conveniently parse RepeatMasker output files. *Mob DNA*. 5(1), https://doi.org/10.1186/1759-8753-5-13 (2014).
33. Shin, H. D. *et al*. Chromosome-level Genome Assembly of Korean Long-tailed Chicken and Pangenome of 40 Gallus gallus Assemblies. *Scientific Data*. 12(1), https://doi.org/10.1038/s41597-024-04287-9 (2025).
34. Stanke, M., Schöffmann, O., Morgenstern, B. & Waack, S. Gene prediction in eukaryotes with a generalized hidden Markov model that uses hints from external sources. *BMC Bioinformatics*. **7**, https://doi.org/10.1186/1471-2105-7-62 (2006).
35. Stanke, M., Diekhans, M., Baertsch, R. & Haussler, D. Using native and syntenically mapped cDNA alignments to improve de novo gene finding. *Bioinformatics* **24**(5), 637–644, https://doi.org/10.1093/bioinformatics/btn013 (2008).
36. *NCBI Sequence Read Archive* http://identifiers.org/insdc.sra:SRR35288173 (2025).
37. *NCBI Sequence Read Archive* http://identifiers.org/insdc.sra:SRR35288174 (2025).
38. *NCBI Sequence Read Archive* http://identifiers.org/insdc.sra:SRR35288172 (2025).

39. *NCBI Genome assembly Database* http://identifiers.org/insdc.gca:GCA_055689475.1 (2026).
40. Molotsi, A. H. *et al*. Alternate Assembly for Venda chicken genome. *figshare* https://doi.org/10.6084/m9.figshare.31358140 (2026).
41. Molotsi, A. H. *et al*. Annotation for Venda chicken genome. *figshare* https://doi.org/10.6084/m9.figshare.31529176 (2026).
42. *NCBI Genome assembly Database* http://identifiers.org/insdc.gca:GCA_016699485.1 (2021).
43. *NCBI Genome assembly Database* http://identifiers.org/insdc.gca:GCA_000002315.5 (2018).
44. *NCBI Genome assembly Database* http://identifiers.org/insdc.gca:GCA_027557775.1 (2023).
45. *NCBI Genome assemble Database* http://identifiers.org/insdc.gca:GCA_024686465.1 (2022).
46. *NCBI Genome Assembly Database* http://identifiers.org/insdc.gca:GCA_024206055.2 (2024).

Acknowledgements

This project is supported by the University of South Africa, Africa-nuanced sustainable development goals research support programme. The samples for this project were provided by a farmer in Limpopo province of South Africa and the technical support was provided by the Africa BioGenome Project. A special thanks to Mapholi labs for providing the resources and managing the project.

Author contributions

Conceptualization: A.H.M., V.R.; Data Curation: A.H.M., L.N., R.M.S.; Formal Analysis: A.H.M., T.S.T.; Funding Acquisition: N.M.; Investigation: A.H.M.; Methodology: A.H.M., L.N., S.M., R.M.S., T.N.; Project Administration: N.M., T.M.; Resources: N.M. Software: A.H.M.; Validation: L.N., R.M.S., S.M., T.S., N.M.,T.M.; Visualisation: A.H.M., N.L.H.; Writing of the original draft: A.H.M.; Reviewing and editing of the manuscript: A.D., A.H.M., L.N., R.M.S., C.A.G., T.S., S.M., N.L.H., T.M.

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to A.H.M. or N.O.M.