

A Phylogeographic Assessment of The Greater Kudu (*Tragelaphus Strepsiceros*) Across South Africa

Ruan Jacobs

University of the Free State - Bloemfontein Campus: University of the Free State

Willem Gabriel Coetzer (✉ coetzerwg@outlook.com)

University of the Free State <https://orcid.org/0000-0003-2189-5539>

Paul Grobler

University of the Free State - Bloemfontein Campus: University of the Free State

Research Article

Keywords: ungulate, mitochondrial DNA, divergence dating, adaptive genetic variation, control region, HSPA5, BMP4

Posted Date: February 23rd, 2022

DOI: <https://doi.org/10.21203/rs.3.rs-1299029/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License. [Read Full License](#)

Abstract

The greater kudu (*Tragelaphus strepsiceros*) is widespread throughout South Africa and occurs in higher densities in the north-eastern and southern regions, with isolated populations in the Eastern Cape Province and in the Kimberley area in the Northern Cape Province. This study aimed to quantify the levels of genetic diversity and differentiation within and between individual greater kudu populations, based on neutral mitochondrial DNA regions as well as adaptive variation. A total of 116 kudu were sampled from various biomes, separated by geographic features and distance, across six South African provinces. Sampling was based on skin samples collected from taxidermists. Genetic diversity and differentiation were quantified using sequence data from the mitochondrial control region (mtDNA CR) and the Heat Shock Protein Subunit 5 (*HSPA5*) gene. A short tandem repeat (STR) sequences was identified at the 3'-UTR of the Bone Morphogenetic Protein 4 (*BMP4*) gene and used for downstream analyses. Twenty-six haplotypes were identified the CR dataset, three for the *HSPA5* region, and 14 alleles were identified for the *BMP4* STR. The CR phylogenetic analyses identified two distinct genetic clades representing an Eastern and Western group respectively. Molecular divergence dating identified the most recent common ancestor of the Eastern and Western South African kudu clades as older (2.237 Mya) than some well-known African antelope species. This data was further supported by the *HSPA5* and *BMP4* results. Environmental selective pressures, such as rainfall and ambient temperature were also identified as possible driving forces for evolution at the *HSPA5* gene region. These results can provide support for future conservation management decisions to ensure the conservation of natural patterns of diversity in this majestic antelope species in South Africa.

Introduction

Greater kudus are large, spiral-horned antelope, rufous to grey-brown in colour, with bulls being greyer than cows. Horns are long and are found in bulls only (Frost 2014; Stuart and Stuart 2015). The greater kudu is a popular game species due to its horn characteristics and large physical size, prized by tourists and hunters (Taylor et al. 2015; Furstenburg 2016). The earliest appearance of greater kudu fossils date from the lower Pleistocene, approximately 2 million years ago (MYA) (Gentry 1978). Other early recorded scientific reports of kudu in South Africa noted the presence of kudu in the Eastern Cape Province of South Africa, specifically in the Albany or Grahamstown district. The kudu used to be found over a wide area stretching from the Limpopo Province down to Cape Town in the Western Cape Province. Some reports state that the populations in the Eastern and Western Cape Provinces became isolated from the rest of the species' range north of the Orange and Vaal rivers at various periods in the 1700's and 1800's, although no conclusive causative event has been identified (Simpson 1972; Sakwa 2005; Frost 2014; Stuart and Stuart 2015).

The popularity of the greater kudu as a trophy species as well as in the venison industry, coupled with the recent expansion in the South African wildlife trade (van der Merwe and Saayman 2003; Taylor et al. 2015), has led to an increased interest in this charismatic bovid as a commercial wildlife species (Furstenburg 2016). Previous reports of a relatively recent bottlenecks, coupled with potential fragmentation, have led to speculations about the current genetic structure of the South African kudu populations (Nersting and Arctander 2001; Sakwa 2005; Stuart and Stuart 2015; Furstenburg 2016).

Both anthropogenic and natural selection have likely contributed in shaping the genetic structure of current kudu populations (Furstenburg 2016). The greater kudu shares a historical geographic distribution with other southern African bovids that have shown definite signs of fragmentation and genetic differentiation (e.g. the common impala (Nersting and Arctander 2001; Lorenzen et al. 2006), eland (Lorenzen et al. 2010), plains zebra (Lorenzen et al. 2008), Cape mountain zebra (Kotzé et al. 2019), roan antelope (Alpers et al. 2004), gemsbok (Osmers et al. 2012), and wildebeest (Codron and Brink 2007)). These studies all found some extent of population fragmentation in the various species, with evidence of Pleistocene refugia and historic population declines and rescue. It is therefore highly likely that populations of the greater kudu have also become fragmented to some degree, due to both anthropogenic and natural barriers to migration (Nersting and Arctander 2001; Furstenburg 2016). In terms of anthropogenic factors, the greater kudu suffered significant habitat loss and consequent population decline in the past (Nersting and Arctander 2001; Sakwa 2005; Furstenburg 2016). This was mostly the result of habitat loss due to agricultural growth and development, and over-hunting (Nersting and Arctander 2001; Furstenburg 2016). Furthermore, breeding practices to produce trophy-quality wildlife has become a common practice amongst commercial wildlife ranchers in South Africa (Taylor et al. 2015). Taylor et al. (2015) stated that the kudu was amongst the most popular trophy-hunting species in South Africa. It was ranked 4th, behind springbok (*Antidorcas marsupialis*), warthog (*Phacochoerus africanus*) and impala (*Aepyceros melampus*).

The aforementioned recent and historical events may have contributed to current patterns of genetic differentiation and may have resulted in genetic drift within populations. Little has however been published on the population structure of the greater kudu (Nersting and Arctander 2001), although many unpublished reports have suggested the presence of fragmentation and phenotypic differentiation throughout the total South African population. Various anecdotal reports by farmers and state veterinarians have reported that kudu populations differ morphologically (body size; horn shape) between regions, although nothing has been published to confirm this. This is especially true when comparing individuals, for example, from the northern and southern regions in South Africa (personal communication, W.G. Coetzer). The only recognized greater kudu sub-population in South Africa is that of the valley bushveld areas of the Sundays River and Great Fish River valleys in the Eastern Cape Province (Furstenburg 2016). However, this author reported that the man-made barriers, such as agricultural lands, that isolate the Eastern Cape Province kudu from populations in the Northern Cape Province are shrinking significantly. This could in part be due to the expanding game farming industry as well as *Acacia (Vachellia) karroo* encroachment due to the climatic changes (Furstenburg 2016). On a sub-continental scale, previous investigations identified genetic structure with an east-west split in the current greater kudu populations across southern Africa (Nersting and Arctander 2001; Sakwa 2005). Gaining a better understanding of the genetic structure of these populations will hopefully aid in the long-term conservation of natural patterns of genetic diversity in the greater kudu. In this regard, current observed patterns of differentiation are likely to be a combination of the consequences of anthropogenic actions through selective removal or translocation (in terms of selecting certain individuals for hunting or breeding purposes) in recent times, and the consequences of earlier natural selection due to differing environmental pressures.

For a complete view on genetic structure in greater kudu, it is useful to include both neutral markers and markers with adaptive significance. The mitochondrial control region (mtDNA CR) is a widely used neutral genetic marker. The mitochondrial genome is relatively conserved across animal species, with no introns, few duplications, and short intergenic regions. The region is known for its high mutation rate that leads to high variability in natural populations. It is also relatively easy to amplify and is present in numerous copies in the cell, making it a relatively simple and cost-effective marker for structural genetic analyses (Gissi et al. 2008; Galtier et al. 2009).

The Heat Shock Protein Family A (Hsp70) Member 5 (*HSPA5*) gene can be used to study variation linked to fitness. This gene encodes the HSP70-5 protein, which is located in the endoplasmic reticulum (ER) of all cell types (Gething 1999; Daugaard et al. 2007). Its primary function is to facilitate the transport of newly synthesized proteins into the ER lumen and aid in their subsequent folding (Henle and Dethlefsen 1978). The strongest link with the HSP70 family of genes has been established with temperature-tolerance in various animals (Henle and Dethlefsen 1978; Daugaard et al. 2007; Galtier et al. 2009). This gene is thus potentially suitable to study the adaptive significance of genetic diversity in kudu populations. This family of genes facilitate a wide array of diverse functions and their ability to do so is a consequence of an ability to cooperate with upstream and downstream chaperone proteins (Clerico et al. 2015). The peptide binding domain (PBD) plays a crucial role in enabling the HSP70 family of proteins to bind to these chaperones, in turn enabling the HSPs to facilitate numerous diverse functions (Strub et al. 2002; Clerico et al. 2015).

The bone morphogenic protein 4 (*BMP4*) gene also codes for characteristics that can elucidate the fitness effects of genetic variation. It is one of 20 genes in the transforming growth factor- β family of proteins (Chen et al. 2004; Albertson and Kocher 2006; McCord and Westneat 2016) and is located on chromosome 10 in cattle (*Bos taurus*, NCBI Reference Sequence: XP_024853077.1). The role of *BMP4* in the development and diversification of cranial morphologies in a wide range of organisms has made it a popular molecular tool for functional morphologists, developmental biologists, and geneticists (Albertson and Kocher 2006; Zhong et al. 2010; McCord and Westneat 2016). One such report focussed on investigating the link between *BMP4* variation and population structure in a southern African ungulate species and was reported by van Aswegen et al. (2012). These authors investigated the population structure of the southern African springbok (*Antidorcas marsupialis*) populations by sequencing the cytochrome b (*cytb*) and *BMP4* genes. It was reported that the *BMP4* gene showed variation based on short tandem repeats (STRs) at the 3'-end of the fragment. The STR displayed a variable number of repeats of the cytosine and adenine bases (CA_n), thus forming several alleles (van Aswegen et al. 2012). Van Aswegen et al. (2012) also found that this STR could potentially be linked with morphological variation in springbok. Variation in the numbers of tandem repeats in introns, regulatory regions, and coding sequences have previously been linked to variations in gene expression and phenotypes (Myers 2007; Gemayel et al. 2012).

Our current study aimed to investigate the distribution of genetic diversity of the South African greater kudu population using a neutral marker (CR) and two markers with adaptive significance (*HSPA5* and *BMP4*). This was achieved by quantifying the levels of genetic diversity within South African kudu populations sampled from across the distribution range, and to identify possible signatures of isolation and fragmentation. Furthermore, it was aimed to identify the genetic structure of the greater kudu across specific ecological regions (biomes) in South Africa. The results from this study will further our understanding of this well-known African ungulate, thereby expanding conservation authorities' ability to properly conserve the species.

Materials And Methods

Ethical statement

All sampling methods were approved by the General/Human Research Ethics Committee (GHREC) of the University of the Free State (Ethics approval number UFS-HSD2018/0265). Representative samples for the Northern Cape, Free State, Limpopo and North West Province were sampled at taxidermists in and around Bloemfontein, Free State, and carried out under Permit number JM 3363/2018 (Free State Department of Economic, Small Business Development, Tourism and Environmental Affairs). The samples from KwaZulu-Natal were sampled at taxidermists situated in KwaZulu-Natal and sampling was performed under Permit number OP 1265/2019 (Ezemvelo KZN Wildlife). The collection of taxidermy samples were exempted from Section 20 clearance by the Department of Agriculture, Forestry and Fisheries. The Eastern Cape Province samples were sourced from previous projects under rules prescribed by that province in 2014. The collection, transport and storage of these samples were approved by the Interfaculty Animal Ethics Committee of the University of the Free State in 2014 (Ethics approval number 13/2014).

Sample collection

A total of 116 kudu samples were sourced from various localities across six provinces in South Africa spanning several biomes (Fig. 1). Most samples (n = 90) were collected from taxidermists. The origin of the samples ranged from animals on private farms to culled individuals from nature reserves. Some samples were obtained from private ranches that preferred to remain anonymous and not disclose the exact location of the ranches, other than the broader region. The sampling strategy covered most of the current natural distributional range of kudu in South Africa.

A 1x1 cm sample was taken from each salted skin sourced from taxidermists. Hides that have been chemically treated and dyed were not sampled as the DNA from these tissues were potentially degraded. Samples were immediately stored in 96% alcohol and stored on ice for transportation. All samples were stored at -20° C in the Department of Genetics laboratory at the University of The Free State (UFS) on arrival. The DNA samples originating from the Eastern Cape Province were obtained from carcasses at private game ranches that were not explicitly hunted for the purpose of this study. These samples also consisted of 1x1 cm skin tissue samples. The same protocol for storage and transportation, as detailed for taxidermy samples, was followed for these samples. No live animals were sampled during this study.

DNA extraction

The Roche High Pure PCR Template Preparation Kit (Roche Diagnostics) was used for DNA extraction throughout the study. All extractions were performed according to the manufacturer's instructions. DNA quantification was performed with a Nanodrop® ND-1000 Spectrophotometer v3.7 (Thermo Fisher Scientific, Waltham, MA, USA) to evaluate the success of the extraction procedures, by measuring both the quantity and quality of isolated DNA.

PCR amplification and sequencing

Primers were specifically designed for the CR in Geneious Prime 2020 software (Kearse et al. 2012), using *Tragelaphus* sp. mitochondrial genome sequences (*Tragelaphus strepsiceros*- JN632708.1; *Tragelaphus scriptus*- JN632707.1; *Tragelaphus angasii*- JN632702.1) downloaded from the Genbank database. The CR primers covers approximately 800 base-pairs and overlaps the hypervariable region 1 (HVR-1), the central conserved region (CCR), and hypervariable region 2 (HVR-2). A ~900 bp region of the nuclear *HSPA5* gene was targeted that includes partial segments of intron 7 and exon 8. Exon 8 contains the peptide binding domain (PBD) of the gene, which plays a critical role in cell regulation by mediating protein interactions involved in

signalling (Strub et al. 2002). Four bovid *HSPA5* sequences were downloaded from Genbank (*Ovis aries* - XM_004005637.2; *Capra hircus* - XM_005687138.3; *Pantholops hodgsonii* - XM_005953827.1 and *Bos taurus* - BT030726.1), aligned, and used for *HSPA5* primer design in Geneious.

For the *BMP4* region, primers originally designed by van Aswegen et al. (2012) were used to obtain a small number of good quality kudu *BMP4* sequences of ~800 bp each. These sequences were then used to design a new forward-primer targeting the CA STR in the 3'- untranslated region (UTR) of the gene, located 20 bp after the stop codon, for fragment analysis. This STR was used for the downstream analysis as the majority of the targeted 800 bp gene region showed little diversity. All primer sequences are provided in Supplementary Table S1.

CR and HSPA5 genetic diversity

The haplotype frequency distributions for both the CR and *HSPA5* samples were calculated using the program DnaSP v. 6.0 (Rozas et al. 2017). The diploid *HSPA5* sequence dataset were first resolved into haplotypes using a Bayesian method in Phase v2.1 (Stephens et al. 2001; Stephens and Scheet 2005) as implemented in DnaSP. Haplotype frequencies were estimated by counting and presented as haplotype counts. A search for shared haplotypes was conducted to identify shared haplotypes within and among regional samples. Intra-population nucleotide diversity (π) was assessed in DnaSP for each sampled population. The software GenAEx v. 6.5 (Peakall and Smouse 2006) was used to determine the inbreeding coefficients for all three gene regions. Nucleotides of the sequence data were numerically coded as follows: A=1, C=2, G=3, T=4, with missing data coded as 0, following the suggested methods in the GenAEx user manual. Geneious software was used to assess any potential amino acid changes caused by non-synonymous single nucleotide polymorphisms (SNPs) at the *HSPA5* region. GenAEx was used to perform an analysis of molecular variance (AMOVA) and pairwise F_{ST} based on the CR and *HSPA5* sequence results.

Phylogenetic analysis

Thirty-two control region sequences were downloaded from GeneBank, representing the three major clades observed by Nersting and Arctander (2001), and combined with the CR sequences from the current study. This combined dataset was then trimmed to 439 bp, following DNA alignment, to allow equal length of sequences. Four *Tragelaphus* sp. sequences were used as outgroups for the phylogenetic analysis and downloaded from GenBank (*Tragelaphus oryx*, NC_020750.1; *Tragelaphus angasii*, NC_020748.1; *Tragelaphus spekii*, NC_020620.1; *Tragelaphus scriptus*, NC_020751.1). A phylogenetic tree was constructed for the combined CR alignment using maximum likelihood (ML) procedures on the online platform PhyML 3.0 (Guindon et al. 2010). The Hasegawa-Kishino-Yano, 85 substitution model (Lanave et al. 1984) with gamma distribution and invariable sites (HKY85 +G +I) was used. The substitution model was determined by the Smart Model Selection (SMS) system (Lefort et al. 2017), implemented in PhyML 3.0 based on 1 000 bootstrap iterations. The software FigTree v.1.4.4 (Rambaut 2018) was used to visualize the phylogenetic tree generated with the PhyML software. Median joining haplotype networks were constructed for both the CR and *HSPA5* gene regions using the software Population Analysis with Reticulate Trees (PopART; Leigh and Bryant 2015).

A molecular clock analysis was additionally performed to better understand the history of the South African *T. strepsiceros*. The haplotype data from our CR sequences were combined with 21 outgroups species representing Bovidae and Cervidae families (See Supplementary Table S2 for species names and GenBank accession numbers). The analysis was performed in the program BEAST v.2.5 (Bouckaert et al. 2019). The HKY85 +G +I nucleotide substitution model was used as identified by jModelTest v.2.1 (Darriba et al. 2012), and a lognormal relaxed-clock model with the Yule process tree prior was selected. Four calibration dates were used: Crown Bovidae, mean age = 18 million years ago (Mya; 16-20 Mya); Crown Cervidae age = 16.6 Mya (16.6-28.4 Mya); Crown Tragelaphini 5.72 (4.7–6.7 Mya); Crown *Connochaetes* spp., mean age = 1.15 Mya (1.15–2.15 Mya) (Thomas 1977; Gentry 1978, 2010; Solounias et al. 1995; Vrba 1997; Vrba and Schaller 2000; Haile-Selassie and WoldeGabriel 2009; Bibi 2013). A single simulation of 20 million generation were run, with a sampling frequency of 10 000 trees. The effective sample size (ESS) values were assessed in Tracer v.1.7.1 (Rambaut et al. 2018) to ensure that the simulation has reached stationarity, with values above 200 deemed acceptable (Drummond et al. 2006). TreeAnnotator v.2 was used to estimate the maximum clade credibility (MCC) tree following a 25% burn-in. FigTree was used to assess the MCC tree and obtain the divergence date estimates.

Environmental factors and HSPA5 diversity

A Multivariate Analysis of Variance (MANOVA) was performed in SPSS (IBM Corp. 2017) to assess the effect of different environmental factors on the abundance of the observed *HSPA5* haplotypes. The 19 Worldclim bioclimatic variables (Hijmans et al. 2005), extracted from our sample location information through Diva-GIS (Hijmans et al. 2012), were used with the *HSPA5* haplotype presence per sample for the MANOVA analysis. A Spearman's rank correlation (ρ) analysis (Spearman 1904) was additionally performed in SPSS to assess the direction of the possible effects.

BMP4 STR genetic diversity

GenAlEx was used to determine measures of genetic diversity for the *BMP4* tandem repeat data, including the number of alleles, observed and expected heterozygosity (H_O and H_E), effective number of alleles and the unbiased heterozygosity calculated to account for sample size. The *BMP4* tandem repeat alleles were also subjected to a Hardy-Weinberg equilibrium (HWE) analysis using the same software to determine the possible signature of genetic drift, mutation, migration, and natural and sexual selection. Finally, inbreeding coefficient (F_{IS}) was used to further explore deviations from expected HWE per population. A Principal Coordinates Analysis (PCoA) was also performed to visualize the pairwise F_{ST} and Nei's individual pairwise genetic distance (D ; Nei et al. 1983) results generated from the *BMP4* tandem repeat fragments. An Analysis of Variance (ANOVA) was performed to investigate the significance of the allelic size distribution across the kudu sample range. Alleles were first grouped into provincial groups, and then according to geographical origin with a western, middle and eastern group following our ML

Results

CR and HSPA5 sequence diversity

From the total of 116 samples collected for this study, 108 samples were successfully sequenced for the mtDNA CR and 100 samples for the *HSPA5* gene region. A ~800 bp CR sequence was amplified from each of 108 kudu samples, before assemblies were trimmed to a final sequence length of 692 bp. A *HSPA5* sequence of 900 - 1100 bp was amplified from 100 kudu samples and these were trimmed to a sequence length of 582 bp, which only included exon 8 and 30 bp of the 3'-UTR. Sequences were trimmed to this extent due to sequencing difficulties at the 5' end of the target region. All unique haplotypes for CR and *HSPA5* were uploaded to Genbank (Accession numbers: OK642751- OK642779).

The total sequencing success rate for the CR was 93.1%, and 86.2% for the *HSPA5* region. A total of 60 polymorphic sites were observed for the mitochondrial CR sequences, with only two polymorphic sites for the *HSPA5* amplicon (synonymous, $n = 1$; non-synonymous, $n = 1$). The non-synonymous polymorphic site was identified as an A/G single nucleotide polymorphism (SNP). The majority of samples (93%) had the A/A genotype, with 7% identified as the A/G heterozygote (KwaZulu-Natal Province, $n = 6$; Free State Province, $n = 1$), with no G/G genotypes observed. This heterozygous SNP is located at the 3' end of the exon, 21 bp from the stop codon. It causes an amino acid change of glutamic acid (Glu) to glycine (Gly). The synonymous polymorphic site was identified as a C/T SNP, leading to no amino acid changes, and was identified in only three samples (Eastern Cape Province, $n = 2$; Northern Cape Province, $n = 1$).

Population genetic diversity

The level of nucleotide diversity (π) at the CR for each population ranged from 0.002 in the Eastern Cape Province population, to 0.027 in the Free State Province population (Table 1). The Limpopo Province showed the highest number of haplotypes and haplotype diversity, with the Eastern Cape Province population showing the lowest level of haplotypes and haplotype diversity. For *HSPA5*, the nucleotide diversity ranged from no diversity to 0.00029. The highest haplotype diversity levels at *HSPA5* were observed for the KwaZulu-Natal, Free State and Eastern Cape Provinces (Table 1).

An AMOVA was performed after grouping all individuals according to their province of origin. For the CR, slightly more genetic variation was found within the different groups (55%) than between populations (45%). This indicates some degree of genetic structuring within the overall kudu sample population at the CR. The *HSPA5* AMOVA results showed that the majority of the genetic variation was found within the different groups at 86%, with 14% of the variation occurring between populations. A second

AMOVA was performed after excluding the limited samples from the Free State and North West Provinces, producing similar results.

All CR-based pairwise F_{ST} comparisons (below diagonal in Table 2) of the populations from Limpopo, Northern Cape, Free State and North West Provinces showed no to very low genetic differentiation. Populations from both the KwaZulu-Natal and Eastern Cape Provinces showed significant differentiation compared to all other populations for the CR ($p=xxx$), with no statistically significant differentiation when compared to each other ($p = 0.111$). The *HSPA5* pairwise combinations (above the diagonal in Table 2) showed that the North West Province differed significantly from all the other provincial groupings, although the corresponding F_{ST} values were low.

Phylogenetic analysis

A phylogenetic tree estimated from the ML analysis showed two major clades when analysing the combined CR sequence data (Fig. 2), with strong bootstrap support (99.1%). One clade (designated the Western clade) contained all the Eastern Cape Province sequences and most of the Northern Cape Province sequences. These samples grouped neatly with sequences from the Nersting and Arctander (2001) South-Western clade from Namibia and Botswana, as well as the single South African sample used in that study originating from northern parts of the Northern Cape Province. The second major clade (designated the Eastern clade) contained all the KwaZulu-Natal Province sequences and most of the Limpopo Province sequences. The sequences from this clade grouped with representative sequences from the Nersting and Arctander (2001) Intermediate clade from Botswana, Zimbabwe and Zambia. The sequences from the North West and Free State Provinces were spread across clades with no apparent trend.

The Bayesian phylogenetic tree estimated from Beast shows similar clustering of our South African kudu sequences than observed for the ML tree, with two main lineages observed. The divergence dates calculated for the Bovidae and Cervidae split was 20.748 Mya (95% highest posterior density (HPD) = 17.718-23.758), and the Bovini-Tragelaphini split was estimated at 14.752 Mya (95% HPD = 11.041-18.553 Mya; Fig. 3; Table 3). The *T. strepsiceros* split from the other *Tragelaphus* species was calculated at around 4.476 Mya (95% HPD = 3.449-5.644 Mya). The most recent common ancestor (MRCA) for the South African kudu was estimated at 2.237 Mya (95% HPD = 1.309-3.214), during the early Pleistocene, with the MRCAs for the Western and Eastern clades estimated at 1.641 (95% HPD = 0.879-2.488 Mya) and 0.996 Mya (95% HPD = 0.512-1.704 Mya), respectively.

For the haplotype network based on the CR sequences, 26 haplotypes were identified, which grouped to two major clusters, separated by 20 mutational steps (Supplementary Figure S1). Cluster 1 consisted of all samples from the populations from the Eastern and Northern Cape Provinces with some individuals from the Free State and North West Provinces. Cluster 2 consisted of all samples from the Limpopo and KwaZulu-Natal Provinces, and some samples from the Free State and North West Provinces. One Free State specimen sampled from the eastern Free State appeared as an outlier, with 25 mutational steps between this haplotype and the nearest neighbouring haplotypes sampled. These neighbouring haplotypes originated from Northern Cape, Limpopo and North-West Province. No clustering was shown between samples from the Free State, North West and Limpopo Provinces as they were spread throughout the haplotype network. Mitochondrial diversity from the CR thus reflects the western (Eastern and Northern Cape Provinces) and eastern (Limpopo and KwaZulu-Natal Provinces) geographical origin of the sample populations, with the exception of samples from the Free State and North West Provinces found in both clusters.

The haplotype network based on *HSPA5* haplotypes (Supplementary Figure S2) showed one major haplotype (Haplotype 1) for the entire dataset. Two additional minor haplotypes were observed. Haplotype 2 consisted of two sequences from the Eastern Cape Province and one sequence from the Northern Cape Province. Haplotype 3 consisted of six sequences from KwaZulu-Natal Province and one sequence from the Free State Province, representing the non-synonymous mutation observed at this gene region. The MANOVA results identified a significant association between *HSPA5* haplotype occurrence per locality and the tested bioclimatic variables ($F(36, 154) = 1.505$; Wilks' lambda = 0.547, p -value = 0.047). Significant links were observed between *HSPA5* haplotype occurrence and the total annual precipitation, precipitation of the wettest month, precipitation in the wettest quarter and precipitation in the warmest quarter of the sampling localities (p -value < 0.05). The post hoc multiple comparisons showed that *HSPA5* Haplotype 3 occurrence was significantly different from the other two haplotypes when considering the total annual precipitation, precipitation of the wettest month, precipitation in the wettest quarter and precipitation in the warmest

quarter bioclimatic variables (p -value < 0.05). The Spearman's rank correlation (ρ) analysis supported the MANOVA results, and indicated a significant positive correlation between *HSPA5* Haplotype 3 occurrence and the total annual precipitation, precipitation of the wettest month, precipitation in the wettest quarter and precipitation in the warmest quarter bioclimatic variables (p -value < 0.05).

BMP4 STR data

A total of 111 samples were successfully genotyped for the *BMP4* STR. These tandem repeats displayed a variable number of dinucleotide repeats of cytosine (C) and adenine (A) bases (CA_n). Following fragment analysis, 14 distinct alleles were identified (Table 4). Private alleles were observed for the Northern Cape ($n = 2$), KwaZulu-Natal ($n = 1$) and North West Provinces ($n = 1$). The ANOVA results showed that the samples from the Northern Cape and Eastern Cape (western group) had significantly (p -value < 0.05) shorter alleles than samples from Limpopo and KwaZulu-Natal (eastern group; Fig. 4).

Populations from the Northern Cape, Eastern Cape, Limpopo, and KwaZulu-Natal Provinces had lower observed heterozygosity values (H_o) than what was expected (H_e). The KwaZulu-Natal Province population showed the lowest H_o . Populations from both the Free State and North West Provinces showed higher H_o levels, although this should be interpreted with caution as both these populations were represented by low sample sizes. Furthermore, with the samples taken from taxidermists, sampling error may be present leading to F_{IS} values that are not representative of values within specific natural populations.

An AMOVA was performed after grouping all individuals according to their province of origin. It was observed that more genetic variation is found within the different individuals at 61%, with lower levels of variation between the individuals at 10%, and among the populations at 28%. Almost all pair-wise F_{ST} combinations showed statistically significant differentiation ($p < 0.05$; Table 5). Only pairwise combinations of the populations from Limpopo and KwaZulu-Natal Province, and Northern Cape and North West Provinces, did not show statistically significant differentiation. The PCoA based on the F_{ST} and Nei's D , however, provided no clear patterns of genetic differentiation between groups (Supplementary Figure S3).

Discussion

The South African greater kudu population is believed to have experienced relatively recent bottlenecks (Nersting and Arctander 2001; Sakwa 2005; Furstenburg 2016), coupled with potential fragmentation. This has led to speculations about the genetic structure of the South African kudu populations (Sakwa 2005; Stuart and Stuart 2015; Furstenburg 2016). Morphologically distinct populations have previously been reported in the south-east, distributed along the thickets of the valley bushveld areas of the Sundays and Great Fish River valleys of the Eastern Cape Province (Sakwa 2005; Furstenburg 2016). In the current study, the population genetic structures observed varied depending on which marker was considered. An east-west genetic structure was observed when assessing the mitochondrial DNA data, with less clear structuring seen with the two adaptive loci.

Genetic diversity and phylogenetic structuring: CR data

The highest level of mitochondrial CR genetic diversity was observed for the Limpopo Province group, with genetic diversity gradually decreasing moving south. Previous reports on the southern African greater kudu noted that this species suffered significant population declines in the late 1800s, but higher population densities remained in Limpopo and Mpumalanga Provinces due to early conservation efforts in the now Kruger National Park (KNP) area (Nersting and Arctander 2001; Sakwa 2005; Furstenburg 2016). This population could therefore have served as a refuge population, retaining a higher level of genetic diversity. The kudu's ability to traverse high fences (Dyirakumunda et al. 2017; Pirie et al. 2017) could have facilitate gene flow between such protected areas and the surrounding habitats.

Our ML analysis with the Nersting and Arctander (2001) sequences showed that South African kudu most likely originated from two distinct colonisation events (Fig. 1), and not strictly from the south-west as proposed by Sakwa (2005). Our Western clade grouped with the South-Western Nersting and Arctander (2001) clade from Namibia. It can therefore be assumed that the Northern Cape and Eastern Cape populations migrated south from this Namibian clade. This further supports the notion from Nersting and Arctander (2001) that the south-western region of the kudu distribution range could have acted as a Pleistocene refugium for this

arid adapted species. Similarly, our Eastern clade was most similar to individuals from the Intermediate Nersting and Arctander (2001) clade, most likely colonising the Limpopo, Mpumalanga and KwaZulu-Natal provinces from Botswana and Zimbabwe.

Our molecular clock analysis gave divergence dates similar to other studies, in spite of only using a short 692bp mitochondrial DNA fragment. The estimated split between Bovidae and Cervidae of 20.748 Mya is younger than what was reported by Guha et al. (2007), based on cytochrome b (*cytb*) and 16S rDNA (16S) gene regions and one fossil calibration point, but older than the estimates from Bibi (2013), based on full mitochondrial genomes sourced from Hassanin et al. (2012) and 16 fossil calibration points. The split of *T. strepsiceros* from the other *Tragelaphus* sp. estimated in our study (4.476 Mya) is similar to the estimates from Hassanin et al. (2018). Our estimates for the MRCA of the South African greater kudu lineages indicate an early Pleistocene (2.237 Mya) divergence. It was previously hypothesised that the region in modern day Namibia could have been a Pleistocene refugium for greater kudu, from where the species migrated into the current distribution range (Nersting and Arctander 2001; Sakwa 2005). Our results further supports the recommendation by Sakwa (2005) for the management of the eastern and western South African greater kudu lineages as separate management units. The molecular dating results estimated, could in actual fact hint at the possibility of these two lineages being unique taxonomic units, equivalent to species or subspecies, but for now they should at least be considered as distinct management units. Other well-known antelope species are also estimated to have diverged during the early Pleistocene. Hassanin et al. (2018) dated the MRCA for the two eland species (*T. oryx* and *T. derbianus*) at 1.94 Mya, and the split between the two bushbuck species (*T. sylvaticus* and *T. scriptus*) at 2.12 Mya. These bushbuck divergence dates were similar to estimates from Moodley and Bruford (2007) confirming the estimates from Hassanin et al. (2018). Similarly, the split between blue (*Connochaetes taurinus*) and black wildebeest (*C. gnou*) is also estimated to have occurred during the early to mid-Pleistocene at 1.15 Mya (1.15-2.15 Mya; (Gentry 1978, 2010; Vrba 1997). Further analysis of the South African greater kudu populations is, however, needed to confirm our speculation of two unique taxonomic groups of greater kudu inhabiting South Africa. The addition of alternative genetic markers such as microsatellites or single nucleotide polymorphism (SNP) arrays can provide the necessary molecular data to confirm this theory.

The observation of higher genetic diversity in the northern regions of each clade's distribution, could be linked to migratory patterns following a north to south pattern as seen in other African mammals (Lawes 1990; Sithaldeen et al. 2009; Turner et al. 2016), with genetic diversity lowering the further the population in question is located from the ancestral populations. The two distinct genetic clades observed for kudu in South Africa, could also explain the anecdotal reports of phenotypic differences between kudu from the Eastern Cape compared to kudu from the north-east of the country. The central populations (FS and NW) on average showed higher CR diversity estimates, and this could be explained by migration and admixture of animals from the Western and Eastern clades.

The popularity of the greater kudu in the commercial wildlife trade in South Africa can also lead to the observation of some admixture in our dataset. High numbers of private wildlife ranches are found in South Africa and wildlife translocations are common amongst ranches involved in breeding and trophy hunting (Taylor et al. 2015; Pitman et al. 2017). We observed possible signs of such translocations, with specimens from the Northern Cape Province grouping with the Eastern Clade, mostly grouping with Limpopo specimens. Four specimens from the Limpopo Province were also observed in the Western Clade, grouping with specimens from the North-West, Northern Cape and Eastern Cape provinces. One CR haplotype from the eastern clade, and sampled from the eastern Free State Province, was identified as an outlier. The large number of mutational steps between this haplotype and its nearest neighbour could either be as a result of incomplete sampling during field collections, or due to individuals being translocated from other populations. Translocations and reintroductions from other regions or provinces may have served as a highly effective mechanism of artificial gene flow, leading to greater genetic diversity observed. Such a phenomenon in South African wildlife was previously reported in impala by Grobler and Van der Bank (1994), who observed that a small impala population of 150 animals, founded from diverse origins, showed genetic diversity equal to a control group of ~11,000 animals.

Genetic diversity and phylogenetic structuring: Adaptive loci

The genetic diversity patterns observed for the *HSPA5* and *BMP4* data, is as expected from the different types of genetic regions investigated. The exon regions of adaptively linked genes involved in regulatory roles such as heat shock or innate immunity, can show low levels of genetic diversity due to selective sweeps driving the retention of the most advantageous gene variant (Coetzer

et al. 2018). The increased level of genetic diversity observed for tandem repeat regions, in turn is known to be driven by increased mutation rates at these regions (Fan and Chu 2007). Tandem repeats located in gene promotor regions or 5' and 3' flanking regions can influence gene expression and function (Myers 2007; Gemayel et al. 2012; Mayr 2016, 2019).

The genetic diversity and haplotype structure observed for the *HSPA5* dataset supports the western and eastern clades identified by our CR analysis. The majority of samples were linked to one ancestral clade, with two smaller haplotypes uniquely linked to the Western and Eastern CR clades. The only non-synonymous mutation was observed in the eastern haplotype. This haplotype was significantly linked to high precipitation, especially increased precipitation during the warmer periods of the year (MANOVA and Spearman's rho). The association of *HSPA5* Haplotype 3 with high precipitation in the warmest quarter of the year could be linked to the increased pathogen activity in this region during this time of the year. The association of heat shock genes and innate immunity genes like Toll-like receptors (TLRs) are well documented (Ohashi et al. 2000; Koliński et al. 2016; Coetzer et al. 2018). Regions with warm, high rainfall seasons are known to have higher pathogen and parasites abundances (Chapman et al. 2010; Lovera et al. 2017). These assumptions should, however, be more thoroughly tested by assessments of pathogen and parasite prevalence on kudu from both arid and mesic regions during low and high rainfall seasons. This information, in combination with genetic screening of TLR and additional heat shock protein loci could provide valuable information with regards to kudu evolution in southern Africa and how future climate changes could influence this and other species.

No significant geographically linked genetic structure was observed for the *BMP4* genotypes, as the majority of groups different significantly from each other. In contrast, when assessing the allele length it was seen that the Western groups had significantly shorter fragments than that observed for the Eastern groups. A study on southern African springbok (*Antidorcas marsupialis*) reported distinct *BMP4* tandem repeat lengths distinguishing the Karoo and Kalahari springbok populations (van Aswegen et al. 2012). The exact effect of these 3'-UTR tandem repeat regions on *BMP4* protein expression is unclear, but variation in the 3'-UTRs of some genes are known to affect mRNA processes such as translation, mRNA stability, and mRNA localization (Mayr 2016). It has also been shown that 3'-UTRs can establish protein-protein interactions (PPIs), thereby regulating protein features that are not mediated by the amino acid sequence (Mayr 2019). Zhong et al. (2010) found that variation in the *BMP4* gene caused by a CA-dinucleotide tandem repeat situated at the 3'-UTR may be associated with differences in body morphology and body size in domestic cattle (*Bos taurus*), as it may likely also affect mRNA structure and translation efficiency (de Smit & van Duin, 1994; Zhong et al., 2010). The *BMP4* gene's role in morphological differentiation can, however, not be excluded.

More in-depth sampling, as well as sequencing of the complete *HSPA5* gene, including the promotor regions, could further our understanding of this gene's possible involvement in phenotypic differentiation in kudu. Additional candidate genes in the BMP gene family could be considered to investigate the possible difference in body size of South Africa's eastern and western kudu populations. This further research should, however, be accompanied by the documentation of morphological features of kudu from the different geographical areas representing the two mitochondrial clades.

Conclusion

Our study shed new light on the possible phylogeographic history of southern African greater kudu. The phylogenetic assessment using CR data, supports the presence of two distinct genetic clades in South Africa. The patterns of genetic diversity and inclusion of CR sequence data from Nersting and Arctander (2001) brought to light the possibility of two independent colonisation events taking place after the Pleistocene. We highlighted the possibility of two unique taxonomic groups, either species or subspecies, of kudu in South Africa. The *HSPA5* sequence data further supports an east/west split in South African kudu, with a unique haplotype resulting from a non-synonymous mutation found in the eastern kudu populations. Our *BPM4* data, to a lesser extent, also indicated some degree of differentiation. Kudu from the western regions of South Africa showed statistically significantly shorter *BMP4* tandem repeat alleles than population to the east of the country. Data from all three genetic markers therefore support the possible migration of kudu south into South Africa from two directions. Further research into the evolutionary and adaptive significance of these two genetic clades are needed to properly support and guide future conservation management efforts of this species.

Declarations

Funding

We would like to acknowledge Inqaba Biotec for the 2018 SeedIT Research Award granted to Ruan Jacobs.

The study was additionally funded by National Research Foundation incentive funding awarded to J. Paul Grobler.

Competing Interests

The authors declare that there are no competing interests related to this manuscript.

Author Contributions

Both Willem G. Coetzer and J. Paul Grobler contributed to the project conception and design. Sample collection and laboratory work was performed by Ruan Jacobs. Data analysis was performed by Ruan Jacobs and Willem G. Coetzer. The manuscript is based on the MSc dissertation of Ruan Jacobs, and the first draft of this manuscript was written by Ruan Jacobs. Funding was provided by National Research Foundation incentive funding received by J. Paul Grobler. All authors commented and edited the subsequent versions of the manuscript.

Acknowledgements

We would like to thank all the owners and managers of the sampling localities. The assistance of the taxidermies is much appreciated, and we would like to specifically thank African Artistry and Marakalalo Taxidermy in Bloemfontein, Zululand Taxidermy in Pongola and Spirit of Africa Taxidermy in Howick. We also would like to thank Hesmari Bindeman for sharing DNA samples from the Eastern Cape for this study.

References

1. Albertson RC, Kocher TD (2006) Genetic and developmental basis of cichlid trophic diversity. *Heredity* (Edinb) 97:211–221. doi: 10.1038/sj.hdy.6800864
2. Alpers DL, Van Vuuren BJ, Arctander P, Robinson TJ (2004) Population genetics of the roan antelope (*Hippotragus equinus*) with suggestions for conservation. *Mol Ecol* 13:1771–1784. doi: <https://doi.org/10.1111/j.1365-294X.2004.02204.x>
3. Bibi F (2013) A multi-calibrated mitochondrial phylogeny of extant Bovidae (Artiodactyla, Ruminantia) and the importance of the fossil record to systematics. *BMC Evol Biol* 13:1–15
4. Bouckaert R, Vaughan TG, Barido-Sottani J, et al (2019) BEAST 2.5: An advanced software platform for Bayesian evolutionary analysis. *PLOS Comput Biol* 15:e1006650
5. Chapman CA, Speirs ML, Hodder SAM, Rothman JM (2010) Colobus monkey parasite infections in wet and dry habitats: implications for climate change. *Afr J Ecol* 48:555–558. doi: <https://doi.org/10.1111/j.1365-2028.2009.01123.x>
6. Chen D, Zhao M, Mundy GR (2004) Bone Morphogenetic Proteins. *Growth Factors* 22:233–241. doi: 10.1080/08977190412331279890
7. Clerico EM, Tilitsky JM, Meng W, Gierasch LM (2015) How Hsp70 Molecular Machines Interact with Their Substrates to Mediate Diverse Physiological Functions. *J Mol Biol* 427:1575–1588. doi: <https://doi.org/10.1016/j.jmb.2015.02.004>
8. Codron D, Brink JS (2007) Trophic ecology of two savanna grazers, blue wildebeest *Connochaetes taurinus* and black wildebeest *Connochaetes gnou*. *Eur J Wildl Res* 53:90–99. doi: 10.1007/s10344-006-0070-2
9. Coetzer WG, Turner TR, Schmitt CA, Grobler JP (2018) Adaptive genetic variation at three loci in South African vervet monkeys (*Chlorocebus pygerythrus*) and the role of selection within primates. *PeerJ* 6:e4953. doi: 10.7717/peerj.4953
10. Darriba D, Taboada GL, Doallo R, Posada D (2012) jModelTest 2: more models, new heuristics and parallel computing. *Nat Methods* 9:772. doi: 10.1038/nmeth.2109
11. Dugaard M, Rohde M, Jäättelä M (2007) The heat shock protein 70 family: Highly homologous proteins with overlapping and distinct functions. *FEBS Lett* 581:3702–3710. doi: <https://doi.org/10.1016/j.febslet.2007.05.039>
12. Drummond AJ, Ho SYW, Phillips MJ, Rambaut A (2006) Relaxed phylogenetics and dating with confidence. *PLoS Biol* 4:e88

13. Dyrakumunda B, Saidi B, Mbanga J (2017) Identification of foot and mouth disease virus isolates using Vp1 gene sequencing. *Zimbabwe J Sci Technol* 12:15–23
14. Fan H, Chu J-Y (2007) A Brief Review of Short Tandem Repeat Mutation. *Genomics Proteomics Bioinformatics* 5:7–14. doi: [https://doi.org/10.1016/S1672-0229\(07\)60009-6](https://doi.org/10.1016/S1672-0229(07)60009-6)
15. Frost W (2014) *The antelope of Africa*. Jacana Media, Sunnyside, Auckland Park, South Africa
16. Furstenburg D (2016) Kudu (*Tragelaphus strepsiceros*). In: Oberem P, Oberem P (eds) *The New Game Rancher*. Briza Publications, Pretoria, pp 180–188
17. Galtier N, Nabholz B, Glémin S, Hurst GDD (2009) Mitochondrial DNA as a marker of molecular diversity: a reappraisal. *Mol Ecol* 18:4541–4550. doi: <https://doi.org/10.1111/j.1365-294X.2009.04380.x>
18. Gemayel R, Cho J, Boeynaems S, Verstrepen KJ (2012) Beyond Junk-Variable Tandem Repeats as Facilitators of Rapid Evolution of Regulatory and Coding Sequences. *Genes* 3
19. Gentry A (2010) Bovidae. In: Werdelin L, Sanders W (eds) *Cenozoic mammals of Africa*. University of California Press, Berkeley, pp 747–803
20. Gentry AW (1978) The fossil bovidae of the Baringo Area, Kenya. *Geol Soc London, Spec Publ* 6:293 LP – 308. doi: [10.1144/GSL.SP.1978.006.01.21](https://doi.org/10.1144/GSL.SP.1978.006.01.21)
21. Gething M-J (1999) Role and regulation of the ER chaperone BiP. In: *Seminars in cell & developmental biology*. Elsevier, pp 465–472
22. Gissi C, Iannelli F, Pesole G (2008) Evolution of the mitochondrial genome of Metazoa as exemplified by comparison of congeneric species. *Heredity (Edinb)* 101:301–320. doi: [10.1038/hdy.2008.62](https://doi.org/10.1038/hdy.2008.62)
23. Grobler JP, Van der Bank FH (1994) Isozyme variation in South African impala (*Aepyceros melampus*) populations under different management regimes. *SA J Wild Res*, 24B(4): 89-94.
24. Guha S, Goyal SP, Kashyap VK (2007) Molecular phylogeny of musk deer: A genomic view with mitochondrial 16S rRNA and cytochrome b gene. *Mol Phylogenet Evol* 42:585–597. doi: <https://doi.org/10.1016/j.ympev.2006.06.020>
25. Guindon S, Dufayard J-F, Lefort V, et al (2010) New algorithms and methods to estimate maximum-likelihood phylogenies: Assessing the performance of PhyML 3.0. *Syst Biol* 59:307–321. doi: [10.1093/sysbio/syq010](https://doi.org/10.1093/sysbio/syq010)
26. Haile-Selassie Y, WoldeGabriel G (2009) *Ardipithecus kadabba: late miocene evidence from the Middle Awash, Ethiopia*. Univ of California Press
27. Hassanin A, Delsuc F, Ropiquet A, et al (2012) Pattern and timing of diversification of Cetartiodactyla (Mammalia, Laurasiatheria), as revealed by a comprehensive analysis of mitochondrial genomes. *Comptes Rendus - Biol* 335:32–50. doi: [10.1016/j.crv.2011.11.002](https://doi.org/10.1016/j.crv.2011.11.002)
28. Hassanin A, Houck ML, Tshikung D, et al (2018) Multi-locus phylogeny of the tribe Tragelaphini (Mammalia, Bovidae) and species delimitation in bushbuck: Evidence for chromosomal speciation mediated by interspecific hybridization. *Mol Phylogenet Evol* 129:96–105. doi: <https://doi.org/10.1016/j.ympev.2018.08.006>
29. Henle KJ, Dethlefsen LA (1978) Heat fractionation and thermotolerance: A review. *Cancer Res* 38:1843–1851
30. Hijmans RJ, Cameron SE, Parra JL, et al (2005) Very high resolution interpolated climate surfaces for global land areas. *Int J Climatol A J R Meteorol Soc* 25:1965–1978
31. Hijmans RJ, Guarino L, Mathur P (2012) DIVA-GIS 7.5. In: DIVA-GIS. http://www.diva-gis.org/docs/DIVA-GIS_manual_7.pdf. Accessed 29 May 2019
32. IBM Corp. (2017) *IBM SPSS Statistics for Windows, Version 25.0*
33. IUCN SSC Antelope Specialist Group (2020) *Tragelaphus strepsiceros* (amended version of 2016 assessment). In: *The IUCN Red List of Threatened Species 2020*. p e.T22054A166487759
34. Kearse M, Moir R, Wilson A, et al (2012) Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* 28:1647–1649. doi: [10.1093/bioinformatics/bts199](https://doi.org/10.1093/bioinformatics/bts199)
35. Koliński T, Marek-Trzonkowska N, Trzonkowski P, Siebert J (2016) Heat shock proteins (HSPs) in the homeostasis of regulatory T cells (Tregs). *Cent J Immunol* 41:317

36. Kotzé A, Smith RM, Moodley Y, et al (2019) Lessons for conservation management: Monitoring temporal changes in genetic diversity of Cape mountain zebra (*Equus zebra zebra*). *PLoS One* 14:e0220331
37. Lanave C, Preparata G, Saccone C, Serio G (1984) A new method for calculating evolutionary substitution rates. *J Mol Evol* 20:86–93
38. Lawes MJ (1990) The distribution of the samango monkey (*Cercopithecus mitis erythrarchus*) and forest history in South Africa. *J Biogeogr* 17:669–680
39. Lefort V, Longueville J-E, Gascuel O (2017) SMS: Smart Model Selection in PhyML. *Mol Biol Evol* 34:2422–2424. doi: 10.1093/molbev/msx149
40. Leigh J, Bryant D (2015) POPART: full-feature software for haplotype network construction. *Methods Ecol Evol* 6:1110–1116. doi: <https://doi.org/10.1111/2041-210X.12410>
41. Lorenzen ED, Arctander P, Siegismund HR (2006) Regional Genetic Structuring and Evolutionary History of the Impala *Aepyceros melampus*. *J Hered* 97:119–132. doi: 10.1093/jhered/esj012
42. Lorenzen ED, Arctander P, Siegismund HR (2008) High variation and very low differentiation in wide ranging plains zebra (*Equus quagga*): insights from mtDNA and microsatellites. *Mol Ecol* 17:2812–2824
43. Lorenzen ED, Masembe C, Arctander P, Siegismund HR (2010) A long-standing Pleistocene refugium in southern Africa and a mosaic of refugia in East Africa: insights from mtDNA and the common eland antelope. *J Biogeogr* 37:571–581. doi: <https://doi.org/10.1111/j.1365-2699.2009.02207.x>
44. Lovera R, Fernández MS, Jacob J, et al (2017) Intrinsic and extrinsic factors related to pathogen infection in wild small mammals in intensive milk cattle and swine production systems. *PLoS Negl Trop Dis* 11:e0005722
45. Mayr C (2016) Evolution and Biological Roles of Alternative 3'UTRs. *Trends Cell Biol* 26:227–237. doi: <https://doi.org/10.1016/j.tcb.2015.10.012>
46. Mayr C (2019) What are 3' UTRs doing? *Cold Spring Harb Perspect Biol* 11:a034728
47. McCord CL, Westneat MW (2016) Phylogenetic relationships and the evolution of BMP4 in triggerfishes and filefishes (Balistoidea). *Mol Phylogenet Evol* 94:397–409. doi: <https://doi.org/10.1016/j.ympev.2015.09.014>
48. Moodley Y, Bruford MW (2007) Molecular Biogeography: Towards an Integrated Framework for Conserving Pan-African Biodiversity. *PLoS One* 2:e454
49. Myers PZ (2007) Tandem repeats and morphological variation. *Nat Educ* 1:1
50. Nei M, Tajima F, Tateno Y (1983) Accuracy of estimated phylogenetic trees from molecular data. *J Mol Evol* 19:153–170
51. Nersting LG, Arctander P (2001) Phylogeography and conservation of impala and greater kudu. *Mol Ecol* 10:711–719. doi: 10.1046/j.1365-294X.2001.01205.x
52. Ohashi K, Ohashi K, Burkart V, et al (2000) Cutting edge: heat shock protein 60 is a putative endogenous ligand of the toll-like receptor-4 complex. *J Immunol* 164:558–561. doi: 10.4049/jimmunol.164.2.558
53. Osmers B, Petersen B-S, Hartl GB, et al (2012) Genetic analysis of southern African gemsbok (*Oryx gazella*) reveals high variability, distinct lineages and strong divergence from the East African *Oryx beisa*. *Mamm Biol* 77:60–66. doi: <https://doi.org/10.1016/j.mambio.2011.08.003>
54. Peakall R, Smouse PE (2006) GENALEX 6: Genetic analysis in Excel. Population genetic software for teaching and research. *Mol Ecol Notes* 6:288–295. doi: 10.1111/j.1471-8286.2005.01155.x
55. Pirie TJ, Thomas RL, Fellowes MDE (2017) Game fence presence and permeability influences the local movement and distribution of South African mammals. *African Zool* 52:217–227
56. Pitman RT, Fattebert J, Williams ST, et al (2017) The Conservation Costs of Game Ranching. *Conserv Lett* 10:403–413. doi: <https://doi.org/10.1111/conl.12276>
57. Rambaut A (2018) FigTree, version 1.4.4. <http://tree.bio.ed.ac.uk/software/figtree/>
58. Rambaut A, Drummond AJ, Suchard M (2018) Tracer v1.7
59. Raza SHA, Hassanin AA, Dhshan AIM, et al (2021) In silico genomic and proteomic analyses of three heat shock proteins (HSP70, HSP90- α , and HSP90- β) in even-toed ungulates. *Electron J Biotechnol* 53:61–70. doi:

60. Rozas J, Ferrer-Mata A, Sánchez-DelBarrio JC, et al (2017) DnaSP 6: DNA sequence polymorphism analysis of large data sets. *Mol Biol Evol* 34:3299–3302
61. Sakwa JS (2005) Molecular phylogeography and evolutionary history of the greater kudu (*Tragelaphus strepsiceros*). University of Pretoria
62. Simpson CD (1972) An evaluation of seasonal movement in greater kudu populations-*Tragelaphus strepsiceros* Pallas-in three localities in southern Africa. *African Zool* 7:197–205
63. Sithaldeen R, Bishop JM, Ackermann RR (2009) Mitochondrial DNA analysis reveals Plio-Pleistocene diversification within the chacma baboon. *Mol Phylogenet Evol* 53:1042–1048
64. Solounias N, Barry JC, Bernor RL, et al (1995) The oldest bovid from the Siwaliks, Pakistan. *J Vertebr Paleontol* 15:806–814
65. Spearman CE (1904) The proof and measurement of association between two things. *Am J Psychol* 15:72–101
66. Stephens M, Scheet P (2005) Accounting for decay of linkage disequilibrium in haplotype inference and missing-data imputation. *Am J Hum Genet* 76:449–462
67. Stephens M, Smith NJ, Donnelly P (2001) A new statistical method for haplotype reconstruction from population data. *Am J Hum Genet* 68:978–989
68. Strub A, Röttgers K, Voos W (2002) The Hsp70 peptide-binding domain determines the interaction of the ATPase domain with Tim44 in mitochondria. *EMBO J* 21:2626–2635. doi: <https://doi.org/10.1093/emboj/21.11.2626>
69. Stuart C, Stuart M (2015) *Stuarts' Field Guide to Mammals of Southern Africa: Including Angola, Zambia & Malawi*, 5th edn. Penguin Random House South Africa, Cape Town
70. Taylor WA, Lindsey PA, Davies-Mostert H (2015) An assessment of the economic, social and conservation value of the wildlife ranching industry and its potential to support the green economy in South Africa. The Endangered Wildlife Trust, Johannesburg
71. Thomas H (1977) Les bovidés du Miocène supérieur des couches de Mpesida et de la formation de Lukeino (district de Baringo, Kenya). In: Proceedings of the 8th Panafrican congress of Prehistory, Nairobi. pp 82–91
72. Turner TR, Coetzer WG, Schmitt CA, et al (2016) Localized population divergence of vervet monkeys (*Chlorocebus* spp.) in South Africa: Evidence from mtDNA. *Am J Phys Anthropol* 159:17–30. doi: 10.1002/ajpa.22825
73. van Aswegen E, Labuschagne C, Grobler JP (2012) Phenotypic differences, spatial distribution and diversity at the *Cytb* and *BMP4* genes in springbok (*Antidorcas marsupialis*). *Mamm Biol* 77:391–396. doi: 10.1016/j.mambio.2011.11.006
74. van der Merwe P, Saayman M (2003) Determining the economic value of game farm tourism. *Koedoe* 46:103–112
75. Vrba ES (1997) New fossils of Alcelaphini and caprinae (Bovidae: Mammalia) from Awash, Ethiopia, and phylogenetic analysis of Alcelaphini. *Palaeontol Africana* 34:127–198
76. Vrba ES, Schaller GB (2000) Phylogeny of Bovidae based on behavior, glands, skulls, and postcrania. In: Vrba ES, Schaller GB (eds) *Antelopes, Deer, and Relatives. Fossil Record, Behavioral Ecology, Systematics, and Conservation*. Yale University Press, New Haven, CT, USA, pp 203–222
77. Zhong X, Zan LS, Wang HB, Liu YF (2010) Polymorphic CA microsatellites in the third exon of the bovine BMP 4 gene. *Genet Mol Res* 9:868–874

Tables

Table 1. Nucleotide diversity (π), number of haplotypes, number of unique haplotypes (control region (CR) only) and haplotype diversity levels of the mitochondrial control region (CR) and Heat Shock Protein Sub-unit 5 (*HSPA5*) loci in six populations of kudu.

Table 2. Pairwise F_{ST} values of the mitochondrial control region (CR) (below diagonal) and Heat Shock Protein Sub-unit 5 region (*HSPA5*) (above diagonal) among pairs of kudu populations. The p -values are listed in parentheses. Bold type indicate values that

Population	Control Region				Heat Shock Protein Sub-unit 5			
	Nucleotide diversity (π)	Number of haplotypes	Unique haplotypes	Haplotype diversity	Nucleotide diversity (π)	Number of haplotypes	Haplotype diversity	Non-synonymous A/G SNP
KwaZulu-Natal Province	0.00313	7	5	0.522	0.00029	2	0.168	6
Eastern Cape Province	0.00200	3	2	0.151	0.00028	2	0.162	0
Limpopo Province	0.01909	11	4	0.935	0	1	0	0
Northern Cape Province	0.01505	6	2	0.490	0.00019	2	0.108	0
Free State Province	0.02722	5	3	0.861	0.00029	2	0.167	1
North West Province	0.02312	5	2	0.786	0	1	0	0

were statistically significant ($p < 0.05$). (KZN = KwaZulu-Natal, EC = Eastern Cape, LIM = Limpopo, NC = Northern Cape, FS = Free State, NW = North West)

	KZN	EC	LIM	NC	FS	NW
KZN	*	0.000 (0.330)	0.000 (0.323)	0.000 (0.313)	0.028 (0.114)	0.083 (0.012)
EC	0.023 (0.111)	*	0.000 (0.341)	0.000 (0.316)	0.027 (0.154)	0.080 (0.013)
LIM	0.681 (0.001)	0.767 (0.001)	*	0.000 (0.334)	0.025 (0.171)	0.073 (0.026)
NC	0.702 (0.001)	0.786 (0.001)	0.000 (0.306)	*	0.026 (0.144)	0.078 (0.020)
FS	0.744 (0.001)	0.854 (0.001)	0.000 (0.347)	0.000 (0.380)	*	0.032 (0.131)
NW	0.711 (0.001)	0.834 (0.001)	0.000 (0.326)	0.004 (0.315)	0.001 (0.292)	*

Table 3. The divergence dates estimated from mitochondrial CR sequences for *Tragelaphus strepsiceros* haplotypes and 21 outgroup ungulate species. The mean estimated divergence time is provided (million years ago; Mya), with the 95% highest posterior density (95% HPD) and the Bayesian posterior probabilities given for each numbered node. The node numbers correspond to Fig. 3. MRCA - most recent common ancestor

Node ID:	Node number:	Mya:	95% HPD:	Posterior probability:
Bovidae-Cervidae MRCA	a	20.748	17.718-23.758	1
Crown Bovidae	b*	18.054	16.226-19.973	1
Crown Cervidae	c*	17.509	16.602-19.266	1
Bovini-Tragelaphini MRCA	d	14.752	11.041-18.553	0.912
Antilopinae/Aepycerotinae/Alcelaphinae MRCA	e	14.61	11.107-18.086	0.999
Bovini MRCA	f	10.138	6.330-14.306	1
Alcelaphini MRCA	g	6.506	3.481-9.387	1
<i>Bubalus/Syncerus</i> MRCA	h	6.33	3.179-9.847	0.993
Crown Tragelaphini	i*	6.081	5.135-6.699	1
<i>Tragelaphus strepsiceros</i> split	j	4.476	3.449-5.644	0.999
RSA <i>T. strepsiceros</i> East/West split	k	2.237	1.309-3.214	1
RSA <i>T. strepsiceros</i> West MRCA	l	1.641	0.879-2.488	0.866
Crown <i>Connochaetes</i>	m*	1.491	1.150-2.067	1
RSA <i>T. strepsiceros</i> East MRCA	n	0.996	0.512-1.704	1

Table 4. The summary statistics for the Bone Morphogenetic Protein 4 (*BMP4*) gene region for each population. The number of samples (n), total number of alleles (Na), number of private alleles (PA), number of effective alleles (Ne), observed heterozygosity (H_O), expected heterozygosity (H_E), inbreeding coefficient (FIS), and deviation from Hardy-Weinberg Equilibrium (HWE) values are provided. (KZN = KwaZulu-Natal; Lim = Limpopo; EC = Eastern Cape; NC = Northern Cape; FS = Free State; NW = North West)

Population	n	Na	Private Alleles (PA)	Ne	H_O	H_E	FIS	HWE deviation significance (%)
Limpopo	20	7	0	3.704	0.6	0.73	0.178	<0.001
NC	17	8	2	5.026	0.647	0.801	0.192	<0.010
EC	26	5	0	3.976	0.538	0.749	0.281	<0.010
KZN	34	9	1	4.533	0.353	0.779	0.547	<0.001
FS	6	3	0	2.323	0.667	0.569	-0.171	NS
NW	8	7	1	3.282	0.75	0.695	-0.079	NS

NS = not significant

Table 5. The pairwise F_{ST} values based on allele differentiation at the Bone Morphogenetic Protein 4 (*BMP4*), among pairs of kudu populations, with the corresponding p-values listed in parentheses. Statistically significant estimates are presented in bold type. (KZN = KwaZulu-Natal; Lim = Limpopo; EC = Eastern Cape; NC = Northern Cape; FS = Free State; NW = North West)

	KZN	EC	LIM	NC	FS	NW
KZN	*					
EC	0.051 (0.001)	*				
LIM	0.000 (0.386)	0.034 (0.042)	*			
NC	0.132 (0.001)	0.135 (0.001)	0.158 (0.001)	*		
FS	0.155 (0.001)	0.099 (0.018)	0.172 (0.003)	0.206 (0.001)	*	
NW	0.159 (0.001)	0.165 (0.001)	0.205 (0.001)	0.026 (0.142)	0.310 (0.001)	*

Figures

Figure 1

Sample locations of the greater kudu (*Tragelaphus strepsiceros*) specimens collected across the South African distribution range. Sample localities are indicated by circles, squares or triangles as indicated in the map legend. The three black circles represents the three genetic clades identified by Nersting and Arctander (2001). The dashed circles represents the two mtDNA Control Region clades identified in the current study. The dashed arrows, indicates the possible direction of colonisation of the observed kudu clades from the south-western Nersting and Arctander (2001) clade. The map is derived from information sourced from IUCN SSC Antelope Specialist Group (2020) and the South African Department of Environmental Affairs (<https://egis.environment.gov.za>)

Figure 2

Maximum likelihood (ML) analysis of kudu mitochondrial control region (CR) sequences estimated through PhyML. Annotations to the right of each branch name indicates the number of individuals that each carried the specific haplotype. Bootstrap support is provided for the major clades observed. (KZN = KwaZulu-Natal; LIM = Limpopo; EC = Eastern Cape; NC = Northern Cape; FS = Free State; NW = North West)

Figure 3

Maximum clade credibility tree estimated using mitochondrial control region (CR) sequence data. The node numbers (a-n) is associated with the divergence date estimates from Table 3. The asterisk (*) indicates the nodes used as calibration points.

Figure 4

The estimated marginal means of the BMP4 allele sizes a) per province and b) per region. A clear increase in allele size was observed moving from the western regions of South Africa to the east.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [SupplementaryFigures.docx](#)
- [Supplementarytables.docx](#)