A phylogeographic assessment of the greater kudu (*Tragelaphus strepsiceros*) across South Africa

R. Jacobs 1 · W. G. Coetzer 1,2 · J. P. Grobler 1

Received: 26 January 2022 / Accepted: 28 June 2022 / Published online: 21 July 2022
© The Author(s), under exclusive licence to Springer Nature B.V. 2022

Abstract
The greater kudu (*Tragelaphus strepsiceros*) is widespread throughout South Africa and occurs in higher densities in the north-eastern regions, with isolated populations in the Eastern Cape Province and in the Kimberley area in the Northern Cape Province. This study aimed to assess the phylogeographic structure of regional South African greater kudu groups, based on neutral mitochondrial DNA regions as well as adaptive variation. A total of 116 kudu were sampled from across the South African distribution range, separated by geographic features and distance, from six South African provinces. Sampling was mainly 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. Short tandem repeat (STR) sequences were identified at the 3′-UTR of the bone morphogenetic protein 4 (*BMP4*) gene and used for downstream analyses. Twenty-six haplotypes were identified from 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 result was consistent with 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. Overall, these results can provide support for future management decisions to ensure the conservation of natural patterns of diversity in this majestic antelope species in South Africa.

Keywords Ungulate · Mitochondrial DNA · Divergence dating · Adaptive genetic variation · Control region · *HSPA5* · *BMP4*

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). The greater kudu used to be found over a wide area across South Africa stretching from the Limpopo Province down to Cape Town in the Western Cape Province (Furstenburg 2016). Early reports of kudu in South Africa also noted the presence of kudu in the Eastern Cape Province of South Africa, specifically in the Albany or Grahamstown district (Furstenburg 2022). 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 bottleneck, coupled with potential fragmentation, have led to speculation 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 towards 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) and wildebeest (Codron and Brink 2007) to name a few]. These studies all found some population fragmentation in the various species, with evidence of Pleistocene refugia and historic population declines and recoveries. 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, management strategies aimed at improving trophy quality has become a common practice amongst commercial wildlife ranchers in South Africa with kudu amongst the most popular trophy-hunting species (Taylor et al. 2015).

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 South African population. Furstenburg (2022) reported that kudu populations differ morphologically (body size; horn shape) between regions, with the Eastern Cape kudu smaller than elsewhere. The only greater kudu population in South Africa recognized as potentially distinct is that of the valley bushveld areas of the Sundays River and Great Fish River valleys in the Eastern Cape Province (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. This marker is widely used in animal population genetic studies, including greater kudu (Nersting and Arctander 2001). 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). The HSP70 family of genes has been strongly linked to 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. 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). 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; van Aswegen et al. 2012; McCord and Westneat 2016). It was previously reported that the BMP4 gene showed variation based on short tandem repeats (STRs) at the 3′-end of the fragment in springbok, which could potentially be linked to morphological variation in springbok (van Aswegen 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. 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 was 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 (Fig. 1). Most samples (n = 90) were collected from taxidermists, with the remainder obtained from previous postgraduate studies on kudu from the Eastern Cape Province. The samples originated from animals hunted on private farms or culled in 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 1 × 1 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

---

**Fig. 1** Sample locations of the greater kudu (*Tragelaphus strepsiceros*) specimens collected across the South African distribution range (shaded area). Sample localities are indicated by circles, squares or triangles as indicated in the map legend. The central group specimens originated from regions on the periphery of the shown distribution range, which are not clearly visible on the map. The three black circles represent the three genetic clades (S-W—South-Western; I—Intermediate; E—Eastern) identified by Nersting and Arctander (2001). The dashed circles represent the two mtDNA control region clades identified in the current study. Not all sampling localities are within the dashed lines as some specimens showed signs of admixture. 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 (DEA; https://egis.environment.gov.za). The DEA map is based on long-term presence absence data, giving a finer scale for the South African distribution than that of the IUCN data which provides information on the larger African distribution.
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 specimens were grouped in to five regional groups according to geographical location. The northern group consists of all Limpopo specimens, the southern group consists of all Eastern Cape specimens, the central group is comprised of the majority of the Free State specimens, the eastern group consists of all KwaZulu-Natal specimens, with the western group containing all the Northern Cape and North-West Province specimens including two Free State specimens originating from the western border of the province. These grouping were used for all down-stream genetic diversity and genetic structure analyses.

**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). 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; Pantelophos 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 Bayssian method in Phase v2.1 (Stephens et al. 2001; Stephens and Scheet 2005) as implemented in DnaSP. 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 regional group. The software GenAlEx 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 GenAlEx 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. GenAlEx was used to perform an analysis of molecular variance (AMOVA) and pairwise PhiPT based on the CR and HSPA5 sequence results. PhiPT estimates are analogues estimate to Fst used in GenAlEx for haploid data (Peakall and Smouse 2006).

**Phylogenetic analysis**

Thirty-two control region sequences were downloaded from GenBank, 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 692 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 substitu-

### Supplementary Table S1.

```report

| Region       | Haplotype Frequency | Haplotype Frequency |
|--------------|---------------------|---------------------|
| Northern     |                      |                     |
| Southern     |                      |                     |
| Central      |                      |                     |
```
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 outgroup 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, age = 5.72 Mya (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 generations was 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 had 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 for each of our sample locations from Diva-GIS (Hijmans et al. 2012), were used with the HSPA5 haplotype presence per sample for the MANOVA analysis. For specimens where the exact location is unknown, the closest known town was taken to extract the environmental data. A Spearman’s rank correlation (rho) analysis (Spearman 1904) was additionally performed in SPSS to assess the direction of the possible effects.

BMP4 STR genetic diversity

The GenAlEx software was used to determine measures of genetic diversity for the BMP4 tandem repeat data, including the number of alleles, observed and expected heterozygosity (H0 and He), 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 (Fis) was used to further explore deviations from expected HWE per population. A principal coordinates analysis (PCoA) was also performed to visualize the pairwise PhiPT and Nei’s individual pairwise genetic distance (D; Nei et al. 1983) results generated from the BMP4 tandem repeat fragments. A MANOVA was performed to investigate the significance of the allelic size distribution across the kudu sample range. Alleles were first grouped according to geographical origin with northern, southern, central, eastern and western groups, and then according to the three genetic clusters obtained from our ML results to assess possible similarities in genetic structure.

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 successfully amplified. The CR sequence assemblies were trimmed to a final sequence length of 692 bp. A HSPA5 sequence of 900–1100 bp was amplified from the 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 (Eastern, n = 6; Central, 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 (Southern, n = 2; Western, n = 1).
Population genetic diversity

The level of nucleotide diversity (π) at the CR for each group ranged from 0.002 in the southern group, to 0.021 in the western and central groups (Table 1). The northern group showed the highest haplotype diversity, with the southern group showing the lowest number of haplotypes and haplotype diversity. For HSPA5, the nucleotide diversity ranged from no diversity in the northern group to 0.0004 in the central group. The highest haplotype diversity levels at HSPA5 were observed for the eastern, central and southern regions (Table 1).

An AMOVA was performed after grouping all individuals according to geographical origin. For the CR, slightly more genetic variation was found among the different groups (51%) than within groups (49%). 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 97%, with 3% of the variation occurring between geographical groups. A second AMOVA was performed after excluding the limited samples from the Free State and North West Provinces, producing similar results.

The CR-based pairwise PhiPT comparisons (below diagonal in Table 2) of the groups from the eastern, southern and northern regions were significantly different from all others (p < 0.05). The only pairwise comparison showing no statistically significant differentiation was observed between the western and central groups (p = 0.321). The HSPA5 pairwise combinations (above the diagonal in Table 2) showed that the western group differed significantly from the southern group, although the corresponding PhiPT 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 southern sequences and most of the western 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 the northern parts of the Northern Cape Province. The second major clade (designated the Eastern clade) contained all the sequences from the eastern group and most of the northern 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 central group and a number of representative sequences from the western group were spread across clades with no apparent trend, possibly due to admixture.

| Regional group | Control Region | Heat Shock Protein Sub-unit 5 |
|---------------|---------------|-----------------------------|
|               | Number of sequences | Nucleotide diversity (π) | Number of haplotypes | Unique haplotypes |
| Central       | 7             | 0.021                       | 4                    | 0.0004            |
| Eastern       | 29            | 0.0003                      | 5                    | 0.0021            |
| Northern      | 18            | 0.021                       | 5                    | 0.0003            |
| Southern      | 26            | 0.0003                      | 4                    | 0.0002            |
| Western       | 28            | 0.0001                      | 2                    | 0.0001            |

The number of HSPA5 sequences with synonymous single nucleotide polymorphisms (SNPs), as well as the A/G non-synonymous SNP are also provided.
The Bayesian phylogenetic tree, estimated from BEAST, identified two main lineages, similar to that observed for our ML tree. The divergence dates calculated for the Bovidae and Cervidae split was 20.748 Mya [95% highest posterior density (HPD) = 17.718–23.758 Mya], 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 Mya), during the early Pleistocene, with the MRCA 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 (Fig. 4). Cluster 1 consisted of all samples from the southern group and the majority of sequences from the western and central groups. Cluster 2 consisted of all samples from the northern and eastern groups, and some samples from the central and western groups. One haplotype (Hap_12) consisting of two specimens appeared as an outlier, with 25 mutational steps between this haplotype and the nearest neighbouring haplotypes. The specimens from this haplotype were from the central group sampled from the eastern Free State and the other from the western group sampled in the western Free State. These neighbouring haplotypes originated from Northern Cape, North-West and Limpopo provinces.

### Table 2

|          | Central | Eastern | Northern | Southern | Western |
|----------|---------|---------|----------|----------|---------|
| Central  | *       | 0.000 (0.340) | N/A      | 0.011 (0.577) | 0.000 (0.281) |
| Eastern  | 0.647 (0.001) | *       | 0.007 (0.299) | 0.034 (0.055) | 0.000 (0.292) |
| Northern | 0.208 (0.010) | 0.183 (0.001) | *       | 0.055 (0.174) | 0.000 (0.099) |
| Southern | 0.607 (0.001) | 0.871 (0.001) | 0.619 (0.001) | *       | 0.075 (0.015) |
| Western  | 0.038 (0.321) | 0.532 (0.001) | 0.619 (0.001) | 0.278 (0.001) | *       |

p-values are listed in parentheses. Bold type indicates values that were statistically significant (p < 0.05).
(western and northern groups). A second possible outlier haplotype (Hap_9) were also observed, originating from the southern group (Eastern Cape Province), with 13 mutational steps to the closest neighbouring haplotype. Both these haplotypes still grouped with the western clade observed with our ML analysis. Haplotypes from the northern, western and central groups were observed throughout the haplotype network. Overall, the mitochondrial diversity from the CR reflects the western and eastern geographical origin of the sample groups.

The haplotype network based on HSPA5 haplotypes (Fig. 5) showed one major haplotype (Kudu_HSPA5_1) for the entire dataset. Two additional minor haplotypes were observed. Kudu_HSPA5_2 consisted of two sequences from the southern group and one sequence from the western group. Kudu_HSPA5_3 consisted of six sequences from the eastern group and one sequence from the central group, representing the non-synonymous mutation observed at this gene region. The MANOVA results identified a significant association between HSPA5 haplotype occurrence and the tested bioclimatic variables \([F(36, 154) = 1.505; \text{Wilks' lambda} = 0.547, \text{p-value} = 0.047]\). 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; Supplementary Table S3).

The Spearman’s rank correlation (rho) 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; Supplementary Table S4).
**Table 3** The divergence dates estimated from mitochondrial CR sequences for *Tragelaphus strepsiceros* haplotypes and 21 outgroup ungulate species

| 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                 |
| Antilocapinae/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                     |
| *Bubalus/Syncerus* MRCA  | h           | 6.33         | 3.179–9.847      | 0.993                 |
| Crown Tragelaphini      | i*          | 6.081        | 5.135–6.699      | 1                     |
| *Tragelaphus strepsiceros* split | j | 4.476 | 3.449–5.644     | 0.999 |
| RSA *T. strepsiceros* East/West split | k | 2.237 | 1.309–3.214     | 1                     |
| RSA *T. strepsiceros* West MRCA | l | 1.641 | 0.879–2.488     | 0.866                 |
| Crown *Connochaetes*     | m*          | 1.491        | 1.150–2.067      | 1                     |
| RSA *T. strepsiceros* East MRCA | n | 0.996 | 0.512–1.704     | 1                     |

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

![Fig. 4](image)

**Fig. 4** The median joining haplotype network representing the mitochondrial control region (CR) sequence data for 108 kudu samples. The hatch marks between the various nodes is an indication of the number of mutational steps between each haplotype observed. The black nodes represent possible haplotypes not sampled during the current study. The size of the circles indicates the number of individuals associated with that specific haplotype.

![Fig. 5](image)

**Fig. 5** The median joining haplotype network for the heat shock protein sub-unit 5 (HSPA5) sequence data obtained from 100 kudu samples. The hatch marks between the various nodes is an indication of the number of mutational steps between each haplotype observed. The size of the circles indicates the number of individuals associated with that specific haplotype.
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 western (n = 3) and eastern (n = 1) groups. The ANOVA results showed that the samples from the western and southern groups (western CR clade) had significantly

| Population | n     | Na   | Private Alleles (PA) | Ne    | Ho   | He   | FIS   | HWE deviation significance |
|------------|-------|------|----------------------|-------|------|------|-------|-----------------------------|
| Central    | 6     | 3.000| 0                    | 2.323 | 0.667| 0.569| −0.171| NS                          |
| Eastern    | 34    | 9.000| 1                    | 4.533 | 0.353| 0.779| 0.547 | p < 0.001                   |
| Northern   | 20    | 7.000| 0                    | 3.704 | 0.600| 0.730| 0.178 | p < 0.001                   |
| Southern   | 26    | 5.000| 0                    | 3.976 | 0.538| 0.749| 0.281 | p < 0.01                    |
| Western    | 25    | 12.000| 3                   | 4.883 | 0.680| 0.795| 0.145 | NS                          |

The number of samples (n), total number of alleles (Na), number of private alleles (PA), number of effective alleles (Ne), observed heterozygosity (Ho), expected heterozygosity (He), inbreeding coefficient (FIS), and deviation from Hardy–Weinberg equilibrium (HWE) values are provided.

NS not significant

**Fig. 6** The estimated marginal means of the *BMP4* allele sizes **a** per geographical region and **b** per CR clade. A clear increase in allele size was observed moving from the western regions of South Africa to the east.
(p-value < 0.05) shorter alleles than samples from the eastern group (eastern CR clade; Fig. 6).

All groups except for the central group had lower observed heterozygosity values ($H_O$) than expected ($H_E$; Table 4). The eastern kudu group showed the lowest $H_O$. Populations from both the central and western groups showed higher $H_O$ levels, although values for the central group should be interpreted with caution due to low sample size. Furthermore, with the samples taken from taxidermists, sampling error may be present leading to genetic diversity values that are not representative of values within specific natural populations.

An AMOVA was performed after grouping all individuals according to their geographical region of origin. It was observed that 85% of the genetic variation is found within the different groups, with lower levels of variation observed between the groups at 15%. Almost all pair-wise PhiPT combinations showed statistically significant differentiation ($p < 0.05$; Table 5). Only the eastern and northern pairwise comparison did not show statistically significant differentiation. The PCoA based on the PhiPT and Nei’s D, however, provided no clear patterns of genetic differentiation between groups (Supplementary Fig. 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 speculation 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 of South Africa, 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.

**Table 5** The pairwise PhiPT values based on allele differentiation at the bone morphogenetic protein 4 (BMP4), among pairs of kudu groups, with the corresponding p-values listed in parentheses.

|        | Central | Eastern | Northern | Southern | Western |
|--------|---------|---------|----------|----------|---------|
| Central | *       |         |          |          |         |
| Eastern | 0.198 (0.005) | *       |          |          |         |
| Northern| 0.266 (0.011) | 0.000 (0.391) | *       |          |         |
| Southern| 0.151 (0.030) | 0.070 (0.009) | 0.054 (0.043) | *       |         |
| Western | 0.340 (0.001) | 0.189 (0.001) | 0.253 (0.001) | 0.209 (0.001) | *       |

Statistically significant estimates are presented in bold type ($p < 0.05$)

**Genetic diversity and phylogenetic structuring: CR data**

The highest level of mitochondrial CR genetic diversity was observed for the northern kudu 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 area (Nersting and Arctander 2001; Sakwa 2005; Furstenburg 2016). The kudu from this area 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 facilitated 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 populations from the western (Northern Cape) and southern (Eastern Cape) regions 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 northern and eastern parts of South Africa (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 692 bp mitochondrial DNA fragment. The estimated split between Bovidae and Cervidae of 20.748 Mya is younger than that 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
The observed divergence time of Tragelaphini from the other bovid groups are in line with that observed by Willows-Munro et al. (2005). 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), but younger than that reported by Willows-Munro et al. (2005). Our estimates for the MRCA of the South African greater kudu lineages indicate an early Pleistocene (2.237 Mya) divergence. 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 suggest 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 colonisation 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 declining with distance from the ancestral populations. The two distinct genetic clades observed for kudu in South Africa, could also explain the reports of phenotypic differences between kudu from the Eastern Cape compared to kudu from the north-east of the country (Furstenburg 2016). The central populations showed high CR diversity estimates, and this could be explained by migration and admixture of animals from the Western and Eastern clades. This result should, however, be taken with caution as the sample size for this group is small.

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. The occurrence of the two outlier CR haplotypes could either be explained by incomplete sampling of populations during field collections, or due to individuals being translocated from other populations possibly outside of South Africa. 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 diverse patterns of observed genetic diversity for the *HSPA5* and *BMP4* data, are 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 (Coetzee 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 promoter 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, especially increased precipitation during the warmer periods of the year (MANOVA and Spearman’s rho). The association of *HSPA5* Haplotype 3 with high precipitation, especially increased precipitation during the warmer periods of the year (MANOVA and Spearman’s rho).
Regions with warm, high rainfall seasons are known to have higher pathogen and parasite abundance (Chapman et al. 2010; Van Wyk and Boomker 2011; Lovera et al. 2017). Helminth loads in several artiodactyl species, including kudu, were previously reported at lower abundance levels in the drier regions of the Limpopo Province, compared to other higher rainfall regions (Van Wyk and Boomker 2011). Such 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 differed 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 processes and translation efficiency (de Smit & van Duin, 1994; Zhong et al., 2010). The BMP4 gene’s role in kudu morphological differentiation can, therefore, not be excluded.

More in-depth sampling, as well as sequencing of the complete HSPA5 gene, including the promoter 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 BMP4 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 those from the east of the country. Data from all three genetic markers therefore are consistent with 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.

Supplementary Information  The online version contains supplemental material available at https://doi.org/10.1007/s10592-022-01464-4.

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.

Author contributions  Both WGC and JPG contributed to the project conception and design. Sample collection and laboratory work was performed by RJ. Data analysis was performed by RJ and WGC. The manuscript is based on the MSc dissertation of RJ, and the first draft of this manuscript was written by RJ. Funding was provided by National Research Foundation incentive funding received by JPG. All authors commented and edited the subsequent versions of the manuscript.

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.

Declarations

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

References

Albertson RC, Kocher TD (2006) Genetic and developmental basis of cichlid trophic diversity. Heredity (edimh) 97:211–221. https://doi.org/10.1038/sj.hdy.6800864
