A peer-reviewed version of this preprint was published in PeerJ on 4 June 2018.

View the peer-reviewed version (peerj.com/articles/4953), which is the preferred citable publication unless you specifically need to cite this preprint.

Adaptive genetic variation at three loci in South African vervet monkeys (*Chlorocebus pygerythrus*) and the role of selection within primates

Willem G Coetzer Corresponding Author: Willem G Coetzer
Email address: coetzerwg@ufs.ac.za

Vervet monkeys (*Chlorocebus pygerythrus*) are one of the most widely distributed non-human primate species found in South Africa. They occur across all the South African provinces, inhabiting a large variety of habitats. These habitats vary sufficiently that it can be assumed that various factors such as pathogen diversity could influence populations in different ways. In turn, these factors could lead to varied levels of selection at specific fitness linked loci. The Toll-like Receptor (*TLR*) gene family, which play an integral role in vertebrate innate immunity, is a group of fitness linked loci which has been the focus of much research. In this study, we assessed the level of genetic variation at partial sequences of two *TLR* loci (*TLR*4 and 7) and a reproductively linked gene, *acrosin* (*ACR*), across the different habitat types within the vervet monkey distribution range. Gene variation and selection estimates were also made among 11 - 21 primate species. Low levels of genetic variation for all three gene regions were observed within vervet monkeys, with only two polymorphic sites identified for *TLR*4, three sites for *TLR*7 and one site for *ACR*. *TLR*7 variation was positively correlated with high mean annual rainfall, which was linked to increased pathogen abundance. The observed genetic variation at *TLR*4 might have been influenced by numerous factors including pathogens and climatic conditions. The *ACR* exonic regions showed no variation in vervet monkeys, which could point to the occurrence of a selective sweep. The *TLR*4 and *TLR*7 results for the among primate analyses was mostly in line with previous studies, indicating a higher rate of evolution for *TLR*4. Within primates, *ACR* coding regions also showed signs of positive selection, which was congruent with previous reports on mammals. Important additional information to the already existing vervet monkey knowledge base was gained from this study, which can guide future research projects on this highly researched taxon as well as help conservation agencies with future management planning involving possible translocations of this
species.
Adaptive genetic variation at three loci in South African vervet monkeys (*Chlorocebus pygerythrus*) and the role of selection within Primates

Willem G. Coetzer* a, Trudy R. Turner b, Christopher A. Schmitt c and J. Paul Grobler a

a Department of Genetics, University of the Free State, Bloemfontein 9300, South Africa
b Department of Anthropology, University of Wisconsin-Milwaukee, Milwaukee, WI 53201
c Department of Anthropology, Boston University, Boston, MA, USA

*Corresponding author:

Dr Willem G. Coetzer
Department of Genetics,
University of the Free State,
PO Box 339,
Bloemfontein,
9300, South Africa

Email: coetzerwg@ufs.ac.za
Abstract

Vervet monkeys (Chlorocebus pygerythrus) are one of the most widely distributed non-human primate species found in South Africa. They occur across all the South African provinces, inhabiting a large variety of habitats. These habitats vary sufficiently that it can be assumed that various factors such as pathogen diversity could influence populations in different ways. In turn, these factors could lead to varied levels of selection at specific fitness linked loci. The Toll-like Receptor (TLR) gene family, which play an integral role in vertebrate innate immunity, is a group of fitness linked loci which has been the focus of much research. In this study, we assessed the level of genetic variation at partial sequences of two TLR loci (TLR4 and 7) and a reproductively linked gene, acrosin (ACR), across the different habitat types within the vervet monkey distribution range. Gene variation and selection estimates were also made among 11–21 primate species. Low levels of genetic variation for all three gene regions were observed within vervet monkeys, with only two polymorphic sites identified for TLR4, three sites for TLR7 and one site for ACR. TLR7 variation was positively correlated with high mean annual rainfall, which was linked to increased pathogen abundance. The observed genetic variation at TLR4 might have been influenced by numerous factors including pathogens and climatic conditions. The ACR exonic regions showed no variation in vervet monkeys, which could point to the occurrence of a selective sweep. The TLR4 and TLR7 results for the among primate analyses was mostly in line with previous studies, indicating a higher rate of evolution for TLR4. Within primates, ACR coding regions also showed signs of positive selection, which was congruent with previous reports on mammals. Important additional information to the already existing vervet monkey knowledge base was gained from this study, which can guide future research projects on
this highly researched taxon as well as help conservation agencies with future management planning involving possible translocations of this species.
Introduction

Data on the level of genetic adaptation within animal populations is an important aspect of conservation biology. Population or habitat specific adaptations should be considered when studying animal population dynamics and evolution, as these aspects of population genetic diversity will undoubtedly provide important information to guide conservation management (Funk et al. 2012). Neutral genetic variation has long been the marker of choice but can overestimate the amount of genetic diversity within a population and therefore overestimate a population’s viability. The consideration of both adaptively linked and neutral markers during population level studies is therefore important (Funk et al. 2012; Hartmann et al. 2014).

Vervet monkeys (Chlorocebus pygerythrus) are one of the most widely distributed non-human primate species found in South Africa, with only Papio ursinus found in a wider range (IUCN 2008). This species occurs across all the South African provinces, inhabiting a large variety of habitats (Skinner & Chimimba 2005; Kingdon et al. 2008), including xeric and mesic zones. These primates are listed as ‘Least concern’ in the Red list of mammals of South Africa, Swaziland and Lesotho (Turner et al. 2016a). The genus Chlorocebus is a taxon of significant scientific interest within research areas such as immunodeficiency virus pathology (Ma et al. 2013), parasite ecology (Gaetano et al. 2014; Wren et al. 2015; Chapman et al. 2016), neuroscience (Woods et al. 2011; Mendell et al. 2014), social learning (Botting et al. 2018) and sexual selection (Borgeaud et al. 2015; Rodríguez et al. 2015), and these studies require a clear understanding of genetic boundaries and connectivity between populations, using both neutral and adaptive markers. Vervet monkeys also often find their way to primate rehabilitation centres following human primate conflicts (Wimberger et al. 2010) and re-introduction of these animals to suitable recipient populations without considering the provenance of the rehabilitated animals.
should be discouraged. A recent study focusing on South African vervet monkey phylogeography identified mtDNA structuring among populations linked to current and past habitat distributions, geographic barriers, distance and female philopatry (Turner et al. 2016b). Three broad genetic clusters were identified, corresponding to i) the northern part of the distribution range, including the northern part of the Indian Ocean coastal belt, ii) the central regions of South Africa and iii) the southern part of the Indian Ocean coastal belt and adjacent inland areas. Turner et al. (2016b) urged that the addition of nuclear loci should be considered in future research to provide a better understanding of the influence of selection on the observed genetic structuring seen among South African vervet monkey populations. A genome wide study by Svardal et al. (2017) showed that viruses played an important role during Chlorocebus evolution, which is in line with the findings of Enard et al. (2016) who identified viruses as the main driver of evolutionary adaptation in mammals. Strong signals for selection were specifically identified for genes involved in cell signalling and transcriptional regulation following viral exposure in Chlorocebus monkeys (Svardal et al. 2017). Further research into the adaptive genetic diversity of vervet monkeys will not only be of value for future studies on this widely researched taxon but can also provide conservation authorities with the needed information to make informed decisions with regards to possible translocations. The different habitats within the vervet distribution range differ sufficiently that it can be assumed that different environmental factors could drive adaptation in local populations. This could then lead to selection at specific fitness linked loci. In recent years, there has been an increasing trend to use adaptive variation in population and phylogenetic analyses (Vasemägi & Primmer 2005; Gatesy & Swanson 2007; Ayoub et al. 2009; Gonzalez-Quevedo et al. 2015; Quéméré et al. 2015). Fitness-linked genes are influenced by external elements such as host-
pathogen interactions (Vasemägi & Primmer 2005; Holderegger et al. 2006), which directly affect population fitness. The study of variation at adaptive loci can therefore provide valuable information on interaction between environment, genotype and the selective influences that shape patterns of diversity.

Pathogen-mediated selection is widely accepted as an important evolutionary driver in nature (Little 2002). In this regard, the Toll-like Receptor (TLR) gene family, which play an integral role in vertebrate innate immunity, is a group of fitness linked loci which has been the focus of much phylogeographic and phylogenetic research (Wlasiuk & Nachman 2010; Tschirren et al. 2012; Hartmann et al. 2014). TLRs respond to a wide range of pathogen-associated molecular patterns (PAMPs) (exogenous and endogenous, Roach et al. 2005); for example, lipoteichoic acids (gram-positive bacteria, Takeuchi et al. 1999), lipopolysaccharides (gram-negative bacteria, Poltorak et al. 2000), CpG DNA (bacterial, Hemmi et al. 2000; viral, Lund et al. 2003) and single-stranded RNA (viral, Diebold et al. 2004; Lund et al. 2004) to name a few. There are generally two classes of TLRs; one recognising mainly extracellular threats such as lipopolysaccharides and expressed on the outer cell membrane, and the other being expressed within intracellular organelles and recognising mainly intracellular threats such as CpG DNA (Takeda et al. 2003; Roach et al. 2005; Vinkler & Albrecht 2009; Fornůsková et al. 2013). Two TLR genes were selected for this study on the basis of the PAMPs associated with each. First, TLR4 was selected due to its responds to a range of PAMPs including lipopolysaccharide from gram-negative bacteria as well as the fusion protein of respiratory syncytial virus (Kurt-Jones et al. 2000), whereas TLR7 was selected due to its involved in immune responses to viral ssRNA (Diebold et al. 2004; Lund et al. 2004) and phagosomal bacteria (Mancuso et al. 2009). These two TLR genes covers a wide range of possible pathogen mediated responses. Gram-negative
Bacteria can cause various infections, from pathogenic *Escherichia coli* causing gastrointestinal infections (Manning 2010) to *Klebsiella pneumoniae* causing pneumonia (Aujla et al. 2008; Twenhafel et al. 2008) and *Neisseria meningitidis* being a causal agent of meningococcal meningitis (Brandtzaeg 2006).

Non-human primates are also affected by numerous viral pathogens. African green monkey simian immunodeficiency virus (SIVagm) is a ssRNA lentivirus (Kaup et al. 2005) associated with primates from the genus *Chlorocebus*. These viruses do, however, not cause AIDS symptoms in their host species and is therefore studied at length to identify the reasons for this lack of symptoms (Ma et al. 2013). Other ssRNA viruses related to *Chlorocebus* species include simian hepatitis A virus (Balayan 1992), influenza (Webster et al. 1992) and a range of simian arteriviruses and simian pegiviruses (Bailey et al. 2014a; Bailey et al. 2014b; Bailey et al. 2016). Using these two genes would therefore provide valuable information with regards to vervet monkey adaptations to viral and bacterial pathogens.

The use of a reproductively linked gene can provide additional information with regards to local adaptations among populations, especially long separated populations. The acrosin (*ACR*) gene encodes for the serine proteinase acrosin, which is linked to fertility in animals and thought to be involved in secondary binding of the sperm cell to the egg zona pellucida leading to proteolysis of the extracellular egg coat (Keime et al. 1990; Gatesy & Swanson 2007). Reports indicated that this gene show significant signs of adaptive change within mammals (Swanson et al. 2001), making it a good candidate for the assessment of the influences of population isolation within a species.

Here, we assessed the level of genetic variation at partial sequences of one reproductively-linked gene (*ACR*) and two immune-linked loci (*TLR4* and *TLR7*) across eight
habitat types within the vervet monkey distribution range, to complement the work by Turner et al. (2016) using a neutral mtDNA locus. We also investigated the effect of natural selection on the evolution of these gene segments within a wider sample of primates. This study will add substantially to our current knowledge of vervet monkey evolution in South Africa, which could aid future research projects as well as conservation management plans.

**Methods**

**Ethics**

Sampling approval and research ethical clearance was obtained for a previous study using mtDNA analysis (Turner et al. 2016) from the Institutional Care and Use Committee of the University of Wisconsin-Milwaukee (Ref no. 07-08 #32) and the Inter-Faculty Animal Ethics Committee of the University of the Free State, South Africa (Ref no. UFS-AED13/2010). All samples from this study were sourced from the Turner et al. (2016) study and no new additional samples were included.

**Samples and DNA extraction**

Samples from a total of 81 vervet monkeys were included in this study (Supplementary Table S1), representing nine localities from across the distribution range (Turner et al. 2016b), located in five biomes (Nama-Karoo, Savanna, Grassland, Albany Thicket and Indian Ocean Coastal Belt biomes; Table 1; Figure 1A) and eight bioregions / vegetation types. These localities represent significantly different environmental conditions, based on parameters such as the mean annual precipitation, mean annual ambient temperature and maximum annual temperature (Table 1). Due to the philopatric nature of vervet monkey troops, there is expected to be a certain level of relatedness among members of the same troop. We therefore aimed to select specimens from multiple troops, if possible. All samples were sourced from the Department of Genetics,
University of the Free State, Biobank and were previously collected by Turner et al. (2016b). All DNA extracts were obtained from ear biopsies. DNA extraction was performed using the Roche High Pure PCR Template Preparation Kit (Roche Diagnostics, Indianapolis, IN). Sample preparation included a re-hydration step, followed by the manufacturer’s instructions. The initial digestion step was carried out overnight. All DNA samples were stored at -20 °C.

DNA amplification and sequencing

Primers for amplification of partial segments from the *Acrosin* (ACR), *Toll-like receptor* (TLR) 4 and TLR7 genes were designed using Primer3 (Rozen & Skaletsky 1999) as implemented in Geneious v9 software (Kearse et al. 2012). Primers were designed using the online Genbank genome sequences for *Chlorocebus sabaeus* (ACR, Chromosome 19: NC_023660; TLR4, Chromosome 12: NC_023653; TLR7, Chromosome X: NC_023671.1). See Table 2 for the primer sequences and annealing temperature information. The *ACR* amplicons covered a partial segment of exon 4 and the complete intron 4 (Chl_ACR2) and complete exon 5 (Chl_ACR1), totalling a ~ 1000 bp sequence. This region was selected because it was reported that exon 5 is the most rapidly evolving segment of the mammalian *ACR* gene (Gatesy & Swanson 2007). The *ACR* intronic region was a by-product from sequencing the ACR exon 4 and exon 5 regions. The *TLR4* primers covered a partial ~700 bp segment of exon 3, with the *TLR7* primers amplifying a partial ~1350 bp segment of exon 3. The first ~700bp of the *TLR4* exon 3 gene was selected as it is the largest of the three *TLR4* exons observed in primates and it contains the extracellular leucine-rich repeat (LRR) domain, which is responsible for ligand recognition (Janssens & Beyaert 2003). The *TLR7* protein is mainly encoded by the third exon of this gene (Astakhova et al. 2009). The region targeted in our study covers approximately a third of the 5’ end of exon 3, which also contains the extracellular domain. Two polymerase chain reaction (PCR) kits (KAPA
HiFi HotStart ReadyMix PCR kit, KAPA Biosystems, South Africa and Ampliqon TEMPase Hot Start 2x Master Mix, Denmark) were used, due to product availability. The manufacturer protocols were followed for both kits. The annealing temperatures ($T_{a}$) for the different gene segments are provided in Table 2.

Amplification success was assessed on a 1% Agarose gel. PCR amplicons were purified using the BioSpin PCR Purification Kit (BioFlux, Japan). The ABI PRISM® BigDye® Terminator v3.1 Cycle Sequencing Kit (Applied Biosystems Division, USA) was used for all sequencing reactions. Sequences were analysed with an ABI 3130 Genetic Analyzer, following sequence product clean-up with the ZR DNA Sequencing Clean-up™ Kit (Zymo Research, USA).

**Sequence diversity within vervet monkeys**

All raw sequences were assembled and edited using Geneious. Sequence alignments for each gene region were performed with ClustalW (Thompson et al. 1994) as implemented in Geneious. Diploid genotypes were resolved into haplotypes using the Bayesian method in Phase v2.1 (Stephens et al. 2001; Stephens & Scheet 2005) implemented in DnaSP v5.10.01 (Librado & Rozas 2009). TLR7 is located on the X chromosome, and we therefore only performed haplotype inference analysis for the sequences from females, as only one gene copy is present in males. The hemizygous male sequences were then added to the phased female sequences for downstream analyses. The number of iterations were set to 50,000, with a burn-in of 5000, and with all other settings at default values. Samples were grouped according to eight sampling regions for all analyses. Data from NC2 and NC3 were combined due to small sample sizes, the close proximity between sites and similar environmental conditions observed between these sites (Table 1). Summary statistics performed with DnaSP included the number of polymorphic sites
(S), number of non-synonymous sites, number of haplotypes (h), haplotype diversity (Hd) and nucleotide diversity (π, rounded to four decimal places due to small magnitude). The program PopArt (Leigh & Bryant 2015) was used to assess the relationship between haplotypes by generating a minimum spanning network of each gene region.

Protein homology modelling was used to provide a clearer perspective of the position of each partial sequence used in the current study. The partial sequences were translated to amino acid sequences using Geneious software. Homology modelling was then performed in SWISS-MODEL (Biasini et al. 2014), and the partial protein structures were superimposed on reference protein structures using DeepView/Swiss-PdbViewer (Guex & Peitsch 1997; http://www.expasy.org/spdbv/). No complete Chlorocebus ACR mRNA or amino acid sequences were available for comparison. We therefore used SWISS-MODEL to identify the best fit template for model building. An ACR protein structure for Ovis aries beta-acrosin (PDB ID: 1fiw; Tranter et al.) was identified as the best fit. Our vervet partial ACR sequence was then superimposed onto this structure using DeepView/Swiss-PdbViewer (Guex & Peitsch 1997; http://www.expasy.org/spdbv/). A Chlorocebus TLR4 protein structure was modelled for use as a reference structure using a Chl. sabaeus amino acid sequence downloaded from the Uniprot Knowledgebase (UniProtKB accession: A0A0D9RL22), with a published human TLR4 protein reference structure (PDB ID: 4G8G; Ohto et al. 2012). The partial sequences generated from our study was then modelled according to, and then superimposed onto, this Chl. sabaeus TLR4 protein structure. A reference protein structure for Chlorocebus TLR7 was constructed using a Chl. sabaeus amino acid sequence from the Uniprot Knowledgebase (UniProtKB accession: A0A0D9SDT9) and a Macaca mulatta protein structure as a reference (PDB ID: 5gmf; Zhang et
Our partial *Chl. pygerythrus* TLR7 sequences were then superimposed on this *Chl. sabaeus* scaffold.

A Mantel test was performed in Arlequin v3.5 (Excoffier & Lischer 2010) to estimate the correlation between geographic distance and the observed genetic diversity. The pairwise $F_{ST}$ values generated in Arlequin were used as measure of genetic distance, with the geographic distance calculated using an online tool for latitude/longitude to geographic distance conversions (Veness 2017). The correlation between the estimated haplotype diversity (Hd) against each of three environmental factors, mean annual precipitation (MAP), mean annual ambient temperature (MAT) and maximum annual temperature (MaxT), were assessed via generalized linear model (GLM) analysis, using PAST v3 software (Hammer et al. 2001). The environmental data were taken from Mucina & Rutherford (2006), which is the most widely used and complete vegetation and biome data for South Africa. A GLM analysis was used as our Hd data not conforming to normality. PAST implements the basic version of the GLM, and a least squares linear regression was utilized due to its robustness. Correlation analyses using Hd values calculated from only the non-synonymous sites were also performed. A correlation analysis was also conducted using the Hd values for the mitochondrial *D-loop* region assessed by Turner et al. (2016b) and the MAP, MAT and MaxT values for the corresponding regions assessed in the current study. This assessment was done to compare neutral genetic variation patterns to that observed from adaptive markers.

**Assessment of natural selection within South African vervet monkeys**

The TLR4 and TLR7 sequence alignments for the South African vervet monkeys were used to assess the influence of natural selection on these gene fragments within the South African
populations. The ACR locus was excluded from the selection analyses within vervets due to lack of variation in both exons.

To estimate the presence of positive and/or negative selection, two codon-based methods as implemented in the Datamonkey web server were used (Kosakovsky Pond & Frost 2005b). The methods included the MEME method for identifying sites under episodic diversifying or positive selection via a maximum likelihood approach (Murrell et al. 2012), and the Bayesian FUBAR model described by Murrell et al. (2013) which is a superior model to the related likelihood-based FEL model. Rather than reporting evidence of positive selection using *p*-values, FUBAR uses posterior probabilities due to the Bayesian algorithm used. FUBAR is also capable of analysing large alignments at a high speed, and might outperform FEL if positive selection is present but weak (Murrell et al. 2013; HyPhy 2018a). Datamonkey ignores identical sequences and therefore only calculate selection from the phased haplotype data. Sites were accepted as candidates for selection with *p*-values < 0.1 for MEME and posterior probability (Post Pr) > 0.9 for FUBAR. The optimal model of nucleotide substitution was selected for each dataset using the Akaike information criterion (AIC; Akaike 1974) in jModelTest (Darriba et al. 2012), prior to the site-by-site selection analyses in Datamonkey. Default settings were used for all other parameters.

An additional selection analysis was conducted in Codeml (Kohlhase 2006), as implemented in PamlX v1.3.1 (Yang 2007; Xu & Yang 2013), using the model pair M7 and M8 to assess site-by-site selection for the three Primate sequence alignments. The codon is considered the unit of evolution (Goldman & Yang 1994) in these models and use the nonsynonymous/synonymous rate ration (*ω = dN/dS*) for selection analysis. It was shown that the likelihood-ratio tests (LRT) for this model pair identified sites of positive selection more readily
than the alternative M1 and M2 pair. This was attributed to the unrealistic nature of the strict neutral model (M1) which does not account for sites with $0 < \omega < 1$ (Yang et al. 2000). The F3x4 model of codon frequencies were used in all analyses. A NJ tree obtained from Datamonkey for each gene was used as a working topology. These models were compared for each gene region using a LRT calculated in the program Impact_S (Maldonado et al. 2014). A Bayes empirical Bayes approach (BEB; Yang et al. 2005) was used to identify codons under selection for model M8. Sites with a posterior probability of > 90% were considered as candidates for selection, as identifying specific sites under selection is more difficult than to identify a specific proportion of sites (Wlasiuk & Nachman 2010).

Assessment of natural selection within Primates

A second set of analyses focused on the ACR, TLR4 and TLR7 sequence alignments of 23 Primate taxa and the South African vervet haplotypes from this study. Primate sequences for ACR and both TLR coding regions were downloaded from GenBank. Sequences from 21 primate species were downloaded for TLR4 and TLR7. Only 11 GenBank sequences covered our amplified ACR coding region. The exonic regions for the ACR sequences were identified via the Chl. sabaeus (NC_023660) acrosin gene region viewed through the graphical viewer tool from GenBank. See Supplementary Table S2 for details on sequence accession numbers. The ACR, TLR4 and TLR7 vervet monkey haplotype were added to the GenBank sequences. Sequence alignments were performed using the online version of MAFFT (Katoh et al. 2002; Katoh et al. 2017), which is more accurate when analysing datasets containing insertions or deletions (Golubchik et al. 2007).
The same Datamonkey and Codeml selection models used in the vervet monkey analyses were implemented to identify sites under selection in the primate dataset. Identical assessment methods were also implemented when considering candidate sites.

**Results**

*Genetic diversity within vervet monkey populations*

More than 90% of all vervet monkey samples successfully amplified at the three gene regions. Following trimming of the aligned sequences, a total of 585bp of *TLR4* (partial exon 3), 1296bp of *TLR7* (partial exon 3) and 1034bp of *ACR* (partial exon 4, intron 4 and partial exon 5) were available for downstream analyses. All haplotypes were deposited on Genbank (Accession numbers: *ACR*, MG014710 - MG014711; *TLR4*, MG014712 - MG014714; *TLR7*, MG014715 - MG014719). A moderate amount of genetic diversity was observed for the two TLR exonic gene regions, with only one polymorphic site observed within the intronic region of the vervet *ACR* gene region (SNP frequencies: T/T = 0.76; T/G = 0.2; G/G = 0.04). No polymorphisms were observed in either amplified *ACR* exonic region. For the *TLR* gene regions we observed two polymorphic sites for *TLR4*, with one non-synonymous position identified in the vervet population from the C2 region, and three polymorphic sites for *TLR7*, with one non-synonymous position observed in vervet populations from the N1, NC1 and NC2&3 regions. Contrasting patterns of Hd was observed for the two TLR genes (Table 3). The highest *TLR4* haplotype diversities were estimated for the northern (N1, Hd = 0.337) and central (C1, Hd = 0.294) regions, with the highest *TLR7* values observed for the northern and northern coastal belt regions (N1, Hd = 0509; NC1, Hd = 0.425; NC2&3, Hd = 0.660; Figure 2). Similar patterns were observed for the nucleotide diversity estimates. *ACR* haplotype diversity estimates followed a similar pattern to *TLR4* diversity, with the highest levels seen in the northern and central regions.
(N1, Hd = 0.519; C2, Hd = 0.337) and the dominant haplotype occurring at a frequency of 0.86.

One dominant haplotype was observed for both TLR4 (92.47%) and TLR7 (80.77%) across all populations (Figure 1B and 1C; Figure 3). The partial ACR, TLR4 and TLR7 sequences generated for this study were modelled onto protein models to view the position of these segments and the identified non-synonymous mutations in TLR4 and TLR7 (Figure 4). Both non-synonymous mutations resulted in amino acid changes in loops on the protein structure close to ligand recognition sites (Figure 4B and 4C).

No significant correlation was found between Hd and geographic distance between vervet monkey sampling localities, following the Mantel analyses (ACR, r = -0.006, P > 0.05; TLR4, r = 0.197, P > 0.05; TLR7, r = 0.208, P > 0.05). The GLM analysis identified a negative, although nonsignificant, correlation between TLR4 Hd and MAP (Slope = -0.0002; p > 0.05; Figure 5) and TLR4 Hd and MAT (Slope = -0.004; p > 0.5), with a positive correlation to MaxT (Slope = 0.014; p > 0.05). For TLR7 a positive correlation was observed between Hd and MAP (Slope = 0.001; p < 0.05; Figure 5). A positive and a negative, although non-significant, correlation was observed for MAT and MaxT respectively (MAT; Slope = 0.053; p > 0.05; MaxT; Slope = -0.031; p > 0.05). The same trends were observed with the GLM analyses based on the non-synonymous TLR Hd estimates. No significant correlations were observed between the neutral D-Loop region and MAP (Slope = -317.24, p > 0.05) and MAT (Slope = 1.867, p > 0.05) A significantly positive correlation was, however, identified between D-loop Hd and MaxT (Slope = -0.093, p < 0.05). See Supplementary Table S3 for full GLM statistics.

Role of selection within vervet monkeys

For the vervet TLR4 alignment, only one codon was identified by the FUBAR method as possibly being under negative (purifying) selection, which is a non-synonymous mutation at
codon 144 (human codon 359; nucleotide 432, T -> C). A strong indication of negative selection at one TLR7 codon (codon 419, human codon 500, nucleotide 432, C -> T) was identified with FUBAR (Supplementary Table S4). This polymorphism is, however, a synonymous mutation and the non-synonymous mutation at TLR7 codon 402 (human codon 483) was not identified as a candidate for selection. These codon positions are based on our amplified fragments lengths. No codons under positive selection were identified.

Genetic diversity within Primates

All three Primate sequence alignments showed high levels of genetic diversity, with 176, 245 and 361 polymorphic sites observed for ACR, TLR4 and TLR7 respectively (Table 3). The highest number of polymorphic sites were observed for TLR7, and the highest Hd levels were observed for TLR4 (Hd = 0.996; Table 3).

Role of selection within Primates

Within Primates, eight codons showed signs of positive, and possibly episodic, selection for the ACR coding region following the MEME results. FUBAR only identified one candidate site for positive selection, with M8 identifying two candidate sites. Furthermore, 21 ACR codons were identified as potentially under negative selection. Two and seven sites of potential positive selection were identified for TLR4 using FUBAR and M8 respectively. Twelve candidate sites for episodic positive selection was identified by the MEME analysis. Five sites of possible negative selection were identified for TLR4 following the FUBAR analysis (Supplementary Table S4). Between the two immune-linked genes, TLR7 displayed the highest amount of candidate sites for positive selection (n = 6) identified with FUBAR, but only three sites were identified with M8. Ten sites possibly under episodic positive selection was identified with MEME, with 52 candidate sites under possible negative selection.
Discussion

The genetic diversity estimates for South African vervet monkey immune-linked genes showed significant correlations to environmental factors associated with specific bioregions within South Africa. Host-pathogen interactions can shape the genetic variation observed in host immune-linked genes (Hamilton 1982; Little 2002; Zhan et al. 2002; Villaseñor-Cardoso & Ortega 2011). Pathogen prevalence and abundance has also been linked to climatic factors such as precipitation and temperature (Guernier et al. 2004; Harvell et al. 2009; Lafferty 2009; Dunn et al. 2010).

Vervet monkey variation and selection

The higher TLR7 diversity observed along the Indian Ocean Costal belt had a strong positive correlation to the high mean rainfall observed in this area. The only non-synonymous mutation observed in this gene region was also seen in the Indian Ocean Costal belt populations. A number of studies have linked higher rainfall to increased human and non-human primate pathogen prevalence (Kupek et al. 2000; Guernier et al. 2004; Chapman et al. 2010; Gaetano et al. 2014). A study on African green monkey simian immunodeficiency virus (SIVagm) infections in South African vervet monkeys, showed a distinct geographic structuring of SIVagm genetic strains across South Africa (Ma et al. 2013). Three main clusters were identified; one with only Free State (Grassland biome) strains, one with Eastern Cape (Albany thicket biome) and KwaZulu-Natal (Indian Ocean coastal belt) strains and a third with mostly KwaZulu-Natal strains. According to the Ma et al. (2013) study one can infer that the vervet monkeys from the Indian Ocean costal belt region are exposed to a wider variety of SIVagm strains. Villaseñor-Cardoso & Ortega (2011) suggested that a higher level of variation at immune-linked loci in a population could lead to an increased potential for a population to respond to a wider range of pathogen-associated molecules. A higher level of pathogen diversity could therefore support the
occurrence of a higher \textit{TLR7} gene diversity in this area. It was also observed that the non-
synonymous mutation identified in our \textit{TLR7} dataset occurs close to a uridine-5'-monophosphate
ligand recognition site (Figure 4C). This site is crucial for \textit{TLR7} recognition of viral ssRNAs
(Diebold et al. 2006; Zhang et al. 2016) originating from viruses such as SIV (ref), HIV-1 (Heil
et al. 2004; Simmons et al. 2013), vesicular stomatitis virus (Lund et al. 2004), hepatitis C virus
(Lee et al. 2006) and influenza virus (Lund et al. 2004; Wang et al. 2008).

The negative correlation between vervet monkey \textit{TLR4} regional genetic variation and the
mean annual precipitation, although insignificant, was interesting. Higher rainfall is generally
linked to higher pathogen diversity (Lafferty 2009), which in turn could lead to higher \textit{TLR}
diversity. Higher \textit{TLR4} Hd values were, however, observed in the central regions than seen in the
Indian Ocean coastal belt regions and the only non-synonymous mutation was also observed in
the central region. A number of possible ligands were observed close to the non-synonymous
mutation site identified for the vervet \textit{TLR4} sequences (Figure 4B). All of these ligands are
associated with gram negative bacteria. The monosaccharide N-acetyl-D-glucosamine (NAG or
GalNAc) forms part of the O-specific polysaccharide of gram-negative bacterial LPS (Kilár et al.
2013). NAG is also an essential component of fungal cell walls (Lee et al. 2008) as well as some
parasites (Araujo et al. 1993) and is excreted by mucolytic bacteria (Sicard et al. 2017). The
ulosonic acid 3-deoxy-D-manno-oct-2-ulosonic acid (KDO) is a component of the LPS inner
core of most gram negative bacteria (Kilár et al. 2013), as well as the inner core of lipooligosaccharide (LOS) of enteric gram-negative bacilli like \textit{Neisseria meningitidis} (Tzeng et
al. 2002). The saturated fatty-acids lauric acid (DAO) and Myristic acid (MYR) forms part of the
Lipid A component of LPS (Lee et al. 2001; Bäckhed et al. 2003). Lipid A is known as an
endotoxin responsible for the pathogenicity of gram negative bacteria (Raetz et al. 1991; Ribeiro et al. 1999; Tzeng et al. 2002).

Bacterial abundance and diversity will vary among the different regions, with some species more prevalent in drier areas than others. Patz et al. (1996) observed that meningococcal meningitis epidemics, in sub-Saharan African human populations, predominantly occur during the hot dry season. This disease is mainly caused by the gram-negative bacteria Neisseria meningitidis (Brandtzaeg 2006). The highest maximum annual temperatures were also notably observed for the central regions (Table 1) and could affect TLR4 diversity in vervet monkeys from this region. Numerous heat shock proteins are associated with TLR4 activity and vice versa (Ohashi et al. 2000; Asea 2003; Roelofs et al. 2006). Small Heat Shock Protein B8 (HSP22) have been identified as a ligand for TLR4 during inflammation caused by rheumatoid arthritis (Roelofs et al. 2006) and HSP70 is known to have chaperokin activity by interacting with TLR4 following cellular stress such as pathogen exposure or exercise (Asea 2003). Ohashi et al. (2000) suggested that TLR4’s murine ortholog, Tlr4, is a mediator for heat shock protein 60 (hsp60) in mice. Further analysis of different HSP genes could provide more insights to the relationship between heat shock proteins and TLR4 diversity and regional temperature differences across the vervet monkey distribution range. A combination of TLR4 mediation of heat shock proteins and exposure to gram-negative bacteria more prevalent in drier areas could in theory drive the retention of different haplotypes in these areas. Divergence of vervet monkeys from the central region from those in the coastal regions was estimated at 1.099 million years ago (MYA; Turner et al. 2016b). Immune genes generally evolve at a faster rate than other genes (Hurst & Smith 1999), and therefore the time since divergence would have been sufficient to support the occurrence of the observed haplotype variation between the two regions. The different patterns
of genetic diversity observed between TLR4 and TLR7 highlights the intricate role different environmental factors or pathogen types (bacterial vs viral) can play on immune gene evolution. More in-depth sampling and larger sequence lengths, as well as a study on bacterial diversity in vervet monkey populations, might provide a better understanding of TLR4 diversity in these primates and the role these factors might play.

The positive correlation observed between the mitochondrial D-loop and maximum annual temperature could be due to the mitochondrial involvement in cell metabolism (Wilson et al. 1985). Lovegrove (2003) showed that ambient temperature has a strong effect on basal metabolic rate. The correlation of mtDNA D-loop variation and MaxT could thus be due to the role this region plays in replication and expression of mitochondrial genes (Sbisà et al. 1997), such as cytochrome b which is involved in metabolism and forms an important component of the mitochondrial electron-transfer chain (Zhang et al. 1998). This result will need to be assessed in further detail, either through sequencing of targeted metabolic genes or through whole mitochondrial genome sequencing of specimens from the study localities. These additional sequencing data will make it possible to better understand the link between climate and mtDNA variation.

Failure to identify signs of positive selection at the two vervet TLR gene segments can be attributed to the low number of haplotypes (TLR4 haplotype n = 3; TLR7 haplotype n = 5) identified for analysis. The lack of positive selection at TLR genes within species could also be linked to the occurrence of episodic selection (Wlasiuk & Nachman 2010). Episodic positive selection is a known consequence of pathogen-mediated selection as pathogen infections might be more sporadic in nature than long established events (Wlasiuk & Nachman 2010; Grueber et al. 2014). Sites under episodic positive selection will mainly be under strong purifying selection.
during their evolution, with bursts of strong positive selection occurring in some lineages (Murrell et al. 2012). These sites can, however, be masked by the overpowering signal of purifying selection that is at play. The MEME analysis for the detection of episodic selection, however, did not identify any sites under episodic selection. Small sample size could, however, also have affected this analysis.

A number of studies have suggested that diversity at synonymous positions could in fact have adaptive significance and influence protein expression (Chamary et al. 2006; Hunt et al. 2014) by either changing splice sites (Pagani et al. 2005; Parmley et al. 2006; Batista et al. 2017) or mRNA secondary structure (Duan et al. 2003; Chamary & Hurst 2005; Chen et al. 2017). The observation of possible purifying selection at a synonymous position at TLR7 further supports this view. We were, however, unable to identify the possible effect these synonymous mutations have on TLR7 expression. This will have to be assessed in further analyses.

The ACR gene show significant levels of variation between mammalian species, with exon 5 being the most rapidly evolving segment (Gatesy & Swanson 2007). The mammalian ACR gene is, however, functionally conserved across mammalian species, highlighting its role in mammalian fertility (Raterman & Springer 2008). Lack of variation within the amplified vervet ACR fragment and the disproportional occurrence of one gene variant, could be linked to this conservation of functionality within the vervet monkey species. Natural selection would favour sperm which contains a functionally stronger gene variant, which will outperform other variants leading to a selective sweep, resulting in a dominant variant remaining in the species. Metz et al. (1998) suggested that the occurrence of a selective sweep might have led to the observed lack of genetic diversity at the reproductive gene lysin in red abalone species (Haliotis rufescens). Further analysis of gene regions located close to the vervet ACR gene will be required to verify
the occurrence of a selective sweep at this gene, as genetic hitchhiking is a known consequence of selective sweeps (Choudhuri 2014). An assessment of the level of ACR gene differentiation between and within all Chlorocebus species would also be an interesting future study. We only had one ACR sequence from Chl. sabaeus, which showed a 21 bp indel in the proline rich region of exon 5 when compared to our vervet ACR sequences. Additional analyses could shed more light on the evolution of this gene within the genus Chlorocebus.

Variation and selection within Primates

The high number of sites under negative selection at TLR7 supports the observation by Nakajima et al. (2008) that TLR7 is subject to negative/purifying selection in Old World primates. Within primates, TLR7 had, overall, the lowest amount of positively selected sites compared to TLR4 which is also congruent with previous findings Wlasiuk & Nachman (2010). The small difference in number of sites under positive selection between TLR4 and TLR7 in our study can also be attributed to the smaller sequence length of the TLR4 region in our study compared to TLR7. The work by Wlasiuk & Nachman (2010) focused on the complete coding regions of primate TLR genes and found a larger difference in the number of sites under selection for these two genes. Fornůsková et al. (2013) observed a similar trend between rodent Tlr4 (Tlr4 polymorphic sites = 545) and Tlr7 (Tlr7 polymorphic sites = 466) sequences while assessing the genetic variation and evolutionary processes involved in these two genes and 23 Murinae species. A study assessing the autosomal sequence variation of European and African Bos taurus and Asian B. indicus revealed that TLR7 showed no genetic variation in any of the breeds, whereas TLR4 showed higher than average signs of genetic variation in B. indicus (Murray et al. 2010). The higher level of positive selection observed for TLR4 vs TLR7 can be associated with
stronger selective pressures playing a role on non-viral TLR loci (TLR4) compared to viral TLRs (TLR7) in primates (Nakajima et al. 2008; Wlasiuk & Nachman 2010).

Between-species variation for ACR was prominent, with numerous indels observed within the proline rich region of exon 5. This proline rich region is cleaved during the conversion of the zymogen form proacrosin to the mature form (Adham et al. 2004). A comparison between various artiodactyl ACR exon 5 sequences by Gatesy & Swanson (2007), however, showed no signs of indels within this region. This could be explained by different evolutionary and life histories of artiodactyls and primates. Different reproductive strategies could be a driver for this difference, with artiodactyls producing small numbers of large, well-developed, offspring capable of escaping predators and primates giving birth to a few large offspring which needs to be carried for protection. This behaviour limits the number of offspring primates can produce per season (Sibly & Brown 2009). Sibly & Brown (2007) also showed that artiodactyls have higher reproductive rates than primates. This was linked to primates occupying areas with lower predation, which is linked to the number of offspring needed for population growth (Sibly & Brown 2007). Raterman & Springer (2008) observed no sites under significant positive selection at ACR for placental mammals. In contrast to the results from Raterman & Springer (2008), signs of positive selection for ACR in mammals were identified by Swanson et al. (2003) supporting the nine sites of possible positive selection identified within our primate ACR dataset.

Signs of strong negative selection was observed in all three genes. Negative or purifying selection is responsible for the removal of deleterious mutations from the genome, and thereby preventing the accumulation of these mutations (Fay et al. 2001). This can be linked to the importance of maintaining stable functional copies of immune-linked and reproductive genes in a
population. These strong signals of negative selection can also potentially mask the occurrence of sites under episodic positive selection (Murrell et al. 2012).

Conclusion

Results from the current study confirm that adaptation to some environmental factors (rainfall and pathogen prevalence) can be linked to genetic variation observed at different immune-linked genes. Specifically, vervet TLR diversity could be shaped by environmental drivers linked to pathogen abundance and prevalence. Vervet monkey TLR4 gene diversity might be shaped by gram-negative bacteria linked to drier climates, as well as its involvement in heat shock protein activity. Higher rainfall and possible increased pathogen prevalence in high rainfall areas, were linked to the observed TLR7 gene variation for South African vervet monkeys. These contrasting patterns of gene diversity coincides with the strong genetic structuring previously observed in vervet monkey mtDNA diversity. All three genes (ACR, TLR4 and TLR7) are under strong selective pressures within primates. The clear differences in evolutionary patterns observed between the two TLR genes might best be explained by the location of expression, as well as the different types of pathogens they respond to. The observations from this study add valuable information to the already existing knowledge surrounding South African vervet monkey evolution and could be valuable for future research in biogeography or host-pathogen ecology. These results will also help to assist conservation agencies to better plan re-introduction or translocation programs for rehabilitated animals, if needed. The genetic structuring linked to selective pressures associated to different environments should therefore be considered when identifying appropriate re-introduction sites or recipient populations.

Acknowledgements
The authors would like to thank all those involved in the original sample collection for the Turner et al. (2016b) study. Specifically, Dr Joseph G. Lorenz and Dr Nelson B. Freimer for their contributions towards the larger vervet monkey project. We would also like to thank Elmarie Blom and Rudi Lombaard for their assistance in the laboratory.


https://doi.org/10.1016/j.tig.2014.04.006


http://www.nature.com/ni/journal/v10/n6/suppinfo/ni.1733_S1.html


ungulate: the European roe deer (Capreolus capreolus). Molecular Ecology 24:3873-
3887.

endotoxin: an extraordinary lipid with profound effects on eukaryotic signal transduction.
The FASEB Journal 5:2652-2660. 10.1096/fasebj.5.12.1916089

Molecular Reproduction and Development 75:1196-1207. 10.1002/mrd.20868

Ribeiro AA, Zhou Z, and Raetz CRH. 1999. Multi-dimensional NMR structural analyses of
doi:10.1002/(SICI)1097-458X(199909)37:9<620::AID-MRC517>3.0.CO;2-Q

Roach JC, Glusman G, Rowen L, Kaur A, Purcell MK, Smith KD, Hood LE, and Aderem A.
Academy of Sciences of the United States of America 102:9577-9582.

Rodriguez RL, Cramer JD, Schmitt CA, Gaetano TJ, Groblier JP, Freimer NB, and Turner TR.
2015. The static allometry of sexual and nonsexual traits in vervet monkeys. Biological
Journal of the Linnean Society 114:527-537.

Roelofs MF, Boelens WC, Joosten LAB, Abdollahi-Roodsaz S, Geurts J, Wunderink LU,
Schreurs BW, van den Berg WB, and Radstake TRDJ. 2006. Identification of Small Heat
Shock Protein B8 (HSP22) as a Novel TLR4 Ligand and Potential Involvement in the
Pathogenesis of Rheumatoid Arthritis. The Journal of Immunology 176:7021-7027.
10.4049/jimmunol.176.11.7021

Rozen S, and Skaletsky H. 1999. Primer3 on the WWW for General Users and for Biologist
Programmers. In: Misener S, and Krawetz SA, eds. Bioinformatics Methods and

loop region structural analysis: identification of new conserved sequences and their

Sibly RM, and Brown JH. 2007. Effects of body size and lifestyle on evolution of mammal life
10.1073/pnas.0707725104

Sibly RM, and Brown JH. 2009. Mammal reproductive strategies driven by offspring mortality-

Bacteria with Components of the Intestinal Mucus. Frontiers in Cellular and Infection
Microbiology 7. 10.3389/fcimb.2017.00387

Simmons RP, Scully EP, Groden EE, BENEDICT KF, Chang JJ, Lane K, Liason J, Rosenberg E,
Lauffenburger DA, and Altfeld M. 2013. HIV-1 infection induces strong production of

Skinner JD, and Chimimba CT. 2005. The mammals of the southern African sub-region. 3rd
Edition: Cambridge University Press.

Inference and Missing-Data Imputation. The American Journal of Human Genetics
76:449-462. http://dx.doi.org/10.1086/428594

Reconstruction from Population Data. The American Journal of Human Genetics 68:978-
989. http://dx.doi.org/10.1086/319501


Tranter R, Read JA, Jones R, and Brady RL. Effector Sites in the Three-Dimensional Structure of Mammalian Sperm β-Acrosin. *Structure* 8:1179-1188. 10.1016/S0969-2126(00)00523-2


Figure 1

South African maps providing sampling locality information and haplotype frequencies for the two TLR loci.

(A) Biome map indicating vervet monkey sampling localities across different biomes. The vervet monkey distribution range is indicated by the grey shading. The software DIVA-GIS (http://www.diva-gis.org/) was used to create the vervet distribution range and biome map, which is based on information obtained from IUCN (2008a) and Mucina & Rutherford (2006). (B) Vervet TLR4 haplotype frequencies per locality and (C) vervet TLR7 haplotype frequencies per locality.
Figure 2

Haplotype diversity estimates for the different vervet monkey sampling localities as calculated in DnaSP.

The black bars represent values calculated for TLR4, red bars TLR7 and blue bars ACR haplotype diversity values.
Figure 3

The minimum spanning haplotype networks estimated for the three gene regions sequenced in the current study.

(A) Acrosin (ACR) haplotype network calculated from the intronic and two exonic regions amplified, (B) Toll-like receptor (TLR) 4 haplotype network and (C) TLR7 haplotype network.

The size of each circle is in relation to the number of sequences analysed.
Figure 4

Protein structures of the three genes assessed for genetic variation in South African vervet monkeys.

(A) The partial ACR segment (blue) investigated in this study covered 32% of the reference β-acrosin protein (yellow). (B) The partial vervet TLR4 segment (blue) covered 31% of the Chl. sabaeus TLR4 reference protein B-chain. The known ligand structures are shown as: NAG – purple, MYR – turquoise, DAO – orange, KDO – green. (C) The partial vervet TLR7 segment (blue) covered 52% of the Chl. sabaeus TLR7 reference protein A-chain. The identified U-ligand structure is shown in green.

*Note: Auto Gamma Correction was used for the image. This only affects the reviewing manuscript. See original source image if needed for review.
Figure 5

The generalised linear model (GLM) results for the correlation analysis between vervet *TLR4* and *TLR7* haplotype diversity and mean annual precipitation.

(A) mean annual precipitation (mm) vs *TLR4* haplotype diversity (Hd) and (B) mean annual precipitation (mm) vs *TLR7* Hd. N1 – Northern; NE1 – North eastern; NW1 – North western; C1 – Central 1; C2 – Central 2; NC1 – Northern coastal belt 1; NC2 – Northern coastal belt 2; NC3 – Northern coastal belt 3; SC1 – Southern coastal belt 1.
Table 1

Sampling sites of vervet monkeys within the South African distribution range.

General geographic locality, average annual precipitation, average ambient temperature and maximum annual temperature are also provided. Site and geographic locality information follows that from Turner et al. (2016). The environmental data were taken from Mucina and Rutherford (2006), unless otherwise stipulated. The provided environmental data are the values provided for the specific Bioregion unit or Veld type.
<table>
<thead>
<tr>
<th>Locality:</th>
<th>Site:</th>
<th>Geographic locality:</th>
<th>Sample size:</th>
<th>Biome:</th>
<th>Bioregion unit / Vegetation type:</th>
<th>Average annual precipitation (mm):</th>
<th>Average annual ambient temperature (°C):</th>
<th>Maximum annual temperature (°C):</th>
</tr>
</thead>
<tbody>
<tr>
<td>Northern</td>
<td>N1</td>
<td>Central Limpopo Province</td>
<td>11</td>
<td>Savanna</td>
<td>Polokwane Plateau Bushveld</td>
<td>500</td>
<td>16.9</td>
<td>33.2</td>
</tr>
<tr>
<td>Northeastern</td>
<td>NE1</td>
<td>Mpumalanga Lowveld</td>
<td>10</td>
<td>Savanna</td>
<td>Granite Lowveld</td>
<td>633</td>
<td>20.9</td>
<td>38.75</td>
</tr>
<tr>
<td>Northwestern</td>
<td>NW1</td>
<td>Orange River, Northern Cape</td>
<td>10</td>
<td>Nama-Karoo</td>
<td>Lower Gariep Alluvial Vegetation</td>
<td>131</td>
<td>20.5</td>
<td>39.5</td>
</tr>
<tr>
<td>Central</td>
<td>C1</td>
<td>Northwestern Free State</td>
<td>10</td>
<td>Savanna</td>
<td>Kimberley Thornveld</td>
<td>407</td>
<td>17.2</td>
<td>37.45</td>
</tr>
<tr>
<td></td>
<td>C2</td>
<td>Central Free State</td>
<td>10</td>
<td>Grassland</td>
<td>Highveld Alluvial Vegetation</td>
<td>495</td>
<td>16.6</td>
<td>32°</td>
</tr>
<tr>
<td>Southern coastal belt</td>
<td>SC1</td>
<td>Shamwari</td>
<td>10</td>
<td>Albany Thicket</td>
<td>Kowie Thicket</td>
<td>517</td>
<td>17.5</td>
<td>35</td>
</tr>
<tr>
<td>Northern coastal belt</td>
<td>NC1</td>
<td>St. Lucia</td>
<td>10</td>
<td>Indian Ocean Coastal belt</td>
<td>Northern Coastal Forest</td>
<td>1044</td>
<td>21</td>
<td>35.3</td>
</tr>
<tr>
<td></td>
<td>NC2</td>
<td>Thorny Park</td>
<td>5</td>
<td>Indian Ocean Coastal belt</td>
<td>KwaZulu-Natal Coastal Belt</td>
<td>989</td>
<td>19.6</td>
<td>32.6</td>
</tr>
<tr>
<td></td>
<td>NC3</td>
<td>Blythedale</td>
<td>5</td>
<td>Indian Ocean Coastal belt</td>
<td>KwaZulu-Natal Coastal Belt</td>
<td>989</td>
<td>19.6</td>
<td>32.6</td>
</tr>
</tbody>
</table>

* Average maximum temperature observed for Bloemfontein over the last nine years, taken from www.worldweatheronline.com.
Table 2 (on next page)

Primer sequences and annealing temperatures for the three gene regions amplified in this study.

The ACR primer pair was only amplified with the Ampliqon TEMPas e Hot Start 2x Master Mix. All primer sequences are provided in 5’ to 3’ direction.
<table>
<thead>
<tr>
<th>Primer</th>
<th>Forward sequence (5'-3')</th>
<th>Reverse sequence (5'-3')</th>
<th>KAPA HiFi Ta</th>
<th>Ampliqon TEMPase Ta</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chl_ACR1</td>
<td>CAGCAGAAACCATGTGACT</td>
<td>TTTTCTCAGCAGCTAGGGG</td>
<td>-</td>
<td>64°C</td>
</tr>
<tr>
<td>Chl_ACR2</td>
<td>TATGCTGATGGAGGCGACGTGGA</td>
<td>GCGGCTTCTTTGCTGTTTTGCAC</td>
<td>-</td>
<td>64°C</td>
</tr>
<tr>
<td>Chl_TLR4</td>
<td>ACAGAAGCTGGTGGCTGTGG</td>
<td>TTGAAGCAACTCTCTGTGTG</td>
<td>66°C</td>
<td>64°C</td>
</tr>
<tr>
<td>Chl_TLR7a</td>
<td>ACCATGTGATGGTCAGACTGC</td>
<td>GGGGCACATGCTGAAGAGAG</td>
<td>66°C</td>
<td>64°C</td>
</tr>
<tr>
<td>Chl_TLR7b</td>
<td>TGCCCCATTTCTTTGACGC</td>
<td>AAGCCGTTGTTGGAGAAGTC</td>
<td>62°C</td>
<td>64°C</td>
</tr>
</tbody>
</table>
Table 3 (on next page)

Genetic diversity estimates for TLR4 and TLR7 at the eight sampling regions for South African vervet monkeys.

Number of samples (N), sequence size in base pairs (bp), number of polymorphic sites (S), number of haplotypes (h), haplotype diversity (Hd) and nucleotide diversity (π) estimates are provided.
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>TLR4</strong></td>
<td><em>All primates:</em></td>
<td>24</td>
<td>585</td>
<td>245</td>
<td>23</td>
<td>0.996</td>
<td>0.0838</td>
</tr>
<tr>
<td></td>
<td><em>All vervet:</em></td>
<td>73</td>
<td>585</td>
<td>2</td>
<td>3</td>
<td>0.141</td>
<td>0.0002</td>
</tr>
<tr>
<td></td>
<td>N1</td>
<td>10</td>
<td>1</td>
<td>2</td>
<td></td>
<td>0.337</td>
<td>0.0006</td>
</tr>
<tr>
<td></td>
<td>NE1</td>
<td>9</td>
<td>1</td>
<td>2</td>
<td></td>
<td>0.111</td>
<td>0.0002</td>
</tr>
<tr>
<td></td>
<td>C1</td>
<td>9</td>
<td>1</td>
<td>2</td>
<td></td>
<td>0.294</td>
<td>0.0005</td>
</tr>
<tr>
<td></td>
<td>C2</td>
<td>10</td>
<td>2</td>
<td>3</td>
<td></td>
<td>0.195</td>
<td>0.0003</td>
</tr>
<tr>
<td></td>
<td>NW1</td>
<td>9</td>
<td>0</td>
<td>1</td>
<td></td>
<td>0.000</td>
<td>0.0000</td>
</tr>
<tr>
<td></td>
<td>SC1</td>
<td>10</td>
<td>0</td>
<td>1</td>
<td></td>
<td>0.000</td>
<td>0.0000</td>
</tr>
<tr>
<td></td>
<td>NC1</td>
<td>8</td>
<td>0</td>
<td>1</td>
<td></td>
<td>0.000</td>
<td>0.0000</td>
</tr>
<tr>
<td></td>
<td>NC2&amp;3</td>
<td>8</td>
<td>0</td>
<td>1</td>
<td></td>
<td>0.125</td>
<td>0.0002</td>
</tr>
<tr>
<td><strong>TLR7</strong></td>
<td><em>All primates:</em></td>
<td>26</td>
<td>1299</td>
<td>361</td>
<td>23</td>
<td>0.988</td>
<td>0.0496</td>
</tr>
<tr>
<td></td>
<td><em>All vervet:</em></td>
<td>80</td>
<td>1296</td>
<td>3</td>
<td>4</td>
<td>0.330</td>
<td>0.0004</td>
</tr>
<tr>
<td></td>
<td>N1</td>
<td>11</td>
<td>3</td>
<td>4</td>
<td></td>
<td>0.509</td>
<td>0.0007</td>
</tr>
<tr>
<td></td>
<td>NE1</td>
<td>9</td>
<td>1</td>
<td>2</td>
<td></td>
<td>0.282</td>
<td>0.0002</td>
</tr>
<tr>
<td></td>
<td>C1</td>
<td>10</td>
<td>1</td>
<td>2</td>
<td></td>
<td>0.118</td>
<td>0.0001</td>
</tr>
<tr>
<td></td>
<td>C2</td>
<td>10</td>
<td>0</td>
<td>1</td>
<td></td>
<td>0.000</td>
<td>0.0000</td>
</tr>
<tr>
<td></td>
<td>NW1</td>
<td>10</td>
<td>0</td>
<td>1</td>
<td></td>
<td>0.000</td>
<td>0.0000</td>
</tr>
<tr>
<td></td>
<td>SC1</td>
<td>10</td>
<td>0</td>
<td>1</td>
<td></td>
<td>0.000</td>
<td>0.0000</td>
</tr>
<tr>
<td></td>
<td>NC1</td>
<td>10</td>
<td>2</td>
<td>2</td>
<td></td>
<td>0.425</td>
<td>0.0007</td>
</tr>
<tr>
<td></td>
<td>NC2&amp;3</td>
<td>10</td>
<td>3</td>
<td>3</td>
<td></td>
<td>0.660</td>
<td>0.0010</td>
</tr>
<tr>
<td><strong>ACR</strong></td>
<td><em>All primates:</em></td>
<td>12</td>
<td>615</td>
<td>176</td>
<td>11</td>
<td>0.985</td>
<td>0.1075</td>
</tr>
<tr>
<td></td>
<td><em>All vervet:</em></td>
<td>75</td>
<td>1034</td>
<td>1</td>
<td>2</td>
<td>0.242</td>
<td>0.0002</td>
</tr>
<tr>
<td></td>
<td>N1*</td>
<td>11</td>
<td>1</td>
<td>2</td>
<td></td>
<td>0.519</td>
<td>0.0005</td>
</tr>
<tr>
<td></td>
<td>NE1*</td>
<td>9</td>
<td>0</td>
<td>1</td>
<td></td>
<td>0.000</td>
<td>0.0000</td>
</tr>
<tr>
<td></td>
<td>C1*</td>
<td>9</td>
<td>1</td>
<td>2</td>
<td></td>
<td>0.294</td>
<td>0.0003</td>
</tr>
<tr>
<td></td>
<td>C2*</td>
<td>10</td>
<td>1</td>
<td>2</td>
<td></td>
<td>0.337</td>
<td>0.0003</td>
</tr>
<tr>
<td></td>
<td>NW1*</td>
<td>8</td>
<td>0</td>
<td>1</td>
<td></td>
<td>0.000</td>
<td>0.0000</td>
</tr>
<tr>
<td></td>
<td>SC1*</td>
<td>9</td>
<td>1</td>
<td>2</td>
<td></td>
<td>0.209</td>
<td>0.0002</td>
</tr>
<tr>
<td></td>
<td>NC1*</td>
<td>10</td>
<td>0</td>
<td>1</td>
<td></td>
<td>0.000</td>
<td>0.0000</td>
</tr>
<tr>
<td></td>
<td>NC2&amp;3*</td>
<td>9</td>
<td>1</td>
<td>2</td>
<td></td>
<td>0.209</td>
<td>0.0002</td>
</tr>
</tbody>
</table>

* Diversity estimates were calculated from the complete amplified fragment. The single polymorphic site for the ACR region was observed in the intronic segment of the sequence.

* Diversity estimates only determined for the partial exon 4 and exon 5 regions.