Field Notes: A Journal of Collegiate Anthropology

Integrating Geographic Information into the Analysis of the Genetic Distribution of South African Vervet Monkeys

Kerry McAuliffe Dore

Trudy R. Turner

Joseph G. Lorenz

J Paul Grobler

Follow this and additional works at: https://dc.uwm.edu/fieldnotes

Integrating Geographic Information into the Analysis of the Genetic Distribution of South African Vervet Monkeys

Kerry McAuliffe Dore[1], Trudy R. Turner[1], Joseph G. Lorenz[2], J. Paul Grobler[3] [1]University of Wisconsin-Milwaukee, [2]University of Central Washington, [3]University of the Free State, South Africa

Abstract

This project uses the program Geneland to reanalyze McAuliffe's (2008) thesis data on genetic variability in three South African vervet monkey populations (Polokwane, Oribi and Blyde). Using information on the geographic location and genetic variability of these populations, the spatially explicit Geneland program shows that the three populations are weakly differentiated. These findings oppose the results of previous genetic studies of South African vervet monkeys as well as the results obtained by McAuliffe with the spatially implicit Structure program, which found that the 34 individuals all come from one population. Based on this historic data and the fact that other studies have found the same number of subpopulations with both Structure and Geneland, I conclude that Polokwane, Oribi and Blyde are slightly differentiated, though not distinct enough to be considered separate populations (Latch et al. 2008). These results need to be supported by an analysis of the entire sample of South African vervet monkey genetic data from up to 200 animals prior to suggesting policy changes regarding genetic structuring in South African vervet monkeys.

Introduction

This project reassesses McAuliffe's (2008) Master's thesis on genetic variability in three South African vervet monkey populations by incorporating geographic information into the analysis. I begin by discussing the long-term South African vervet monkey genetics project and why it began as well as how McAuliffe's data fits in to this project. I then review McAuliffe's (2008) results and conclusions before introducing

Field Notes: A Journal of Collegiate Anthropology, Vol. 1 [2009], Iss. 1, Art. 11

Integrating Geographic Information into the Analysis of the   113
Genetic Distribution of South African Vervet Monkeys
K. Dore, T. Turner, J.G. Lorenz, J.P. Grobler

why it is important to integrate geographic data into this analysis. The shortcomings of Structure, the Bayesian assignment test utilized by McAuliffe, are discussed, as well as specific details on how Geneland, the new statistic used for the this analysis, uses geographic and genetic information to reach conclusions. This is followed by a review of the Geneland results and what new information has been gained from this analysis. Finally, I discuss the assumptions of Geneland and how the results from this project compare to those obtained by Latch et al. (2008), who utilized both Structure and Geneland statistical programs in a similar study.

Background
In 2001, Grobler et al. (2006) began a large project with the goal of better understanding the genetic structuring of South African vervet monkeys (*Cercopithecus aethiops*). To date, the team has collected samples from almost 200 vervet monkeys from more than 12 sites that cover the geographic range of vervet monkeys in South Africa. This research project began for two main reasons. First, the taxonomy of vervets in Africa is uncertain. There are generally believed to be six sub-species of *Cercopithecus aethiops* throughout the continent; however, this classification has undergone many changes over the years (Lorenz et al. in press). In fact, Meester et al. (1986) cite six additional subspecies from the southern African subregion alone (see Skinner and Chimimba 2005). Understanding the genetic distribution of the animals can help quantify the true number of subspecies. Second, governmental regulations in South Africa currently restrict the mixing of vervet monkeys from different geographic areas due to potential negative genetic effects. This is a problem considering mixing occurs at many rehabilitation centers which aim to re-release orphaned vervet monkeys back into the wild. There are many orphaned vervets in South Africa; either their parents were killed because many South Africans see vervet monkeys as pests or they were

taken in as pets and became unmanageable. Therefore, there are many rehabilitated vervet monkeys that are unable to be re-released back into the wild. A better understanding of the animals' genetic distribution can help effectively manage this problem.

The entire data set of 200 animals has not yet been analyzed; however, smaller projects investigating a few of these populations have been completed. Preliminary research (Grobler and Matlala 2002; Grobler et al. 2006) provides little evidence to support the hypothesis of genetic structuring within South African vervet monkey populations, however the authors suggest additional analyses using more markers and additional populations. Therefore, McAuliffe (2008) investigated genetic variability in three populations that had not yet been analyzed: Polokwane, Blyde and Oribi.

Results from McAuliffe (2008) support the results obtained by Grobler and Matlala (2002) and Grobler et al. (2006), finding little evidence of genetic structuring. Microsatellites, highly variable neutral genetic markers, have been examined from two perspectives, population genetics and landscape genetics, to reach this conclusion (McAuliffe 2008). Population genetics techniques require that populations be designated beforehand and looks for differences between them, while landscape genetics uses the individual as the unit of measure and tests whether the genetic makeup of individuals clusters them into distinct populations. The population genetics statistics results are as follows. An Analysis of Molecular Variance (AMOVA), conducted to compare the percentage of variation occurring within and between populations, shows that over 95% of the genetic variation in these three populations occurs within, rather than between populations (Figure 1). R_{ST} values, which quantify the genetic differences between populations, were determined (Table 1) with the program R_{ST} Calc (Goodman 1997). The only significant R_{ST} value occurs between the Oribi and Polokwane populations. In order to clarify this significant R_{ST} value, a landscape genetics statistic, the Bayesian assignment test, was

Integrating Geographic Information into the Analysis of the
Genetic Distribution of South African Vervet Monkeys
K. Dore, T. Turner, J.G. Lorenz, J.P. Grobler

115

employed. Using no information about the geographic location or population membership of the individuals, the Bayesian assignment test, run through the statistical program Structure (Pritchard et al. 2000), determined that one population was the best fit for the data (Figure 2). Figure 3, the Structure output for two populations, explains this result because it shows the individuals are extremely genetically similar. This result led to the conclusion that the confounding R_{ST} result was due to sample size; only 34 animals were investigated. The author concludes that while it is likely that these animals come from one population, running the tests with all sampled animals is important before making any conclusions (McAuliffe 2008).

AMOVA design and results :

Source of variation	d.f.	Sum of squares	Variance components	Percentage of variation
Among populations	3	15.608	0.14413 Va	4.93
Within populations	64	177.789	2.77795 Vb	95.07
Total	67	193.397	2.92208	
Fixation Index	FST :	0.04932		

Figure 1: Analysis of molecular variance results.

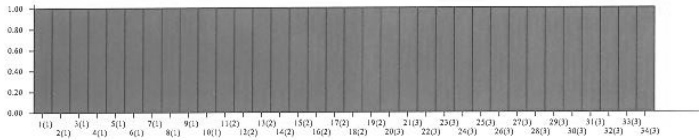| **Pairwise comparison**	**R_{ST} value**	**P-value**	**Nm**
Blyde-Oribi	0.241	0.001	0.784
Blyde-Polokwane	0.000	0.670	Infinite
Oribi-Polokwane	0.123	0.040	1.78

Table 1: Results from R_{ST} Calc.

Figure 2: Histogram from Structure illustrating K = 1 population.
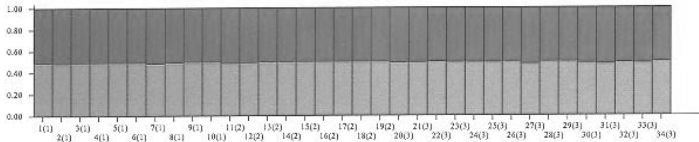

Figure 3: Histogram from Structure illustrating K = 2 populations.

Integrating Geographic Information

The statistical tests run by McAuliffe (2008) do not take into consideration the geographic locations of the vervet monkeys. This is important, for Blyde and Polokwane are close geographically while Oribi is more distant from these two populations (See Figure 4). We therefore expect Blyde and Polokwane to be more genetically similar to each other than to Oribi (and in fact the R_{ST} Calc test found significant genetic differences between Polokwane and Oribi). This is important in determining the number of populations present, for it has been shown that spatially implicit models like Structure are sometimes unable to correctly identify the number of populations in a dataset or to unambiguously assign individuals to populations as allele frequency distributions between two populations become more similar (Evanno et al. 2005, Latch et al. 2006). This is likely due to a combination of similar allele frequency distributions among subpopulations (which can be caused by recent isolation or contemporary gene flow) as well as a corresponding lack of power to differentiate among subpopulations (Latch et al. 2006).

Integrating Geographic Information into the Analysis of the   117
Genetic Distribution of South African Vervet Monkeys
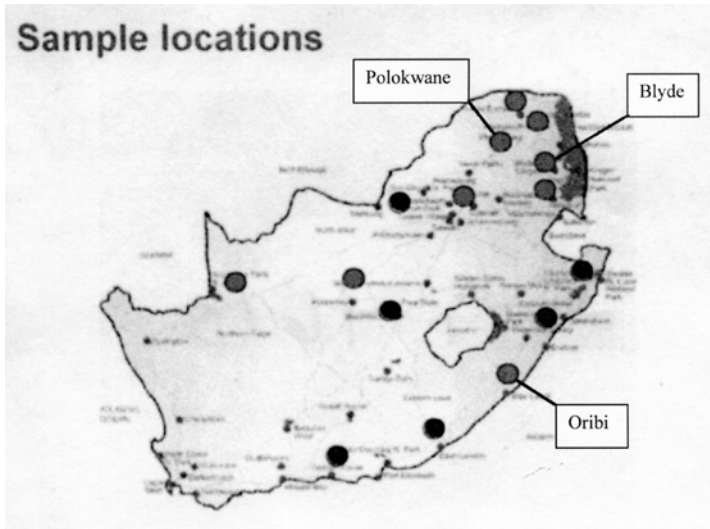K. Dore, T. Turner, J.G. Lorenz, J.P. Grobler

Figure 4: Trapping locations in South Africa.

The program Geneland (Guillot et al. 2005), a spatially explicit model, was therefore utilized to incorporate geographic data into the analysis of genetic structuring in these three populations. Geneland has been developed to optimize the delineation of subpopulations by incorporating spatial coordinates for each sample into the model. This is in contrast with the Structure approach, where all clustering solutions are equally probable. This may make sense in some situations where geographic barriers do not mimic genetic partitions, however it is often true that genetically differentiated populations exist in geographically distinct areas. Geneland was designed to incorporate spatial coordinates at an earlier stage of model development so that geographic boundaries among populations can be better defined (Latch et al. 2008). This program uses the Poisson-Voronoi tessellation model to define these boundaries. The Poisson-Voronoi model assumes that the spatial domain of each population can be approximated by the union of a few polygonal domains (see examples in Figure 5). These kinds of spatial patterns can be expected

when genetic differentiation occurs by limited gene flow induced by the presence of physical barriers such as roads, rivers, mountain ranges and human activity. Formally, the Poisson-Voronoi tessellation model assumes that there are an unknown number of polygons that approximate the true pattern of population spread across space. These polygons are centered around spatial points and each polygon is assigned to one of the populations (Figure 6) (Guillot et al. 2005).
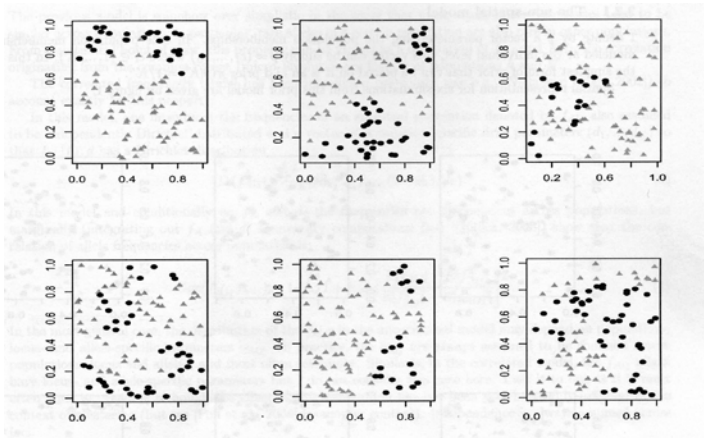

Figure 5: Six examples of 100 individuals belonging to two populations where the spatial domain of each population can be approximated by the union of a few polygonal domains (Guillot et al. 2005).
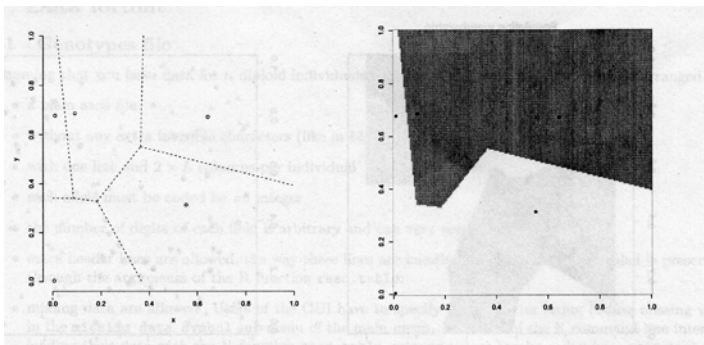

Figure 6: Example of Poisson-Voronoi tessellation. Left panel: location of cell "center" and voronoi cells induced. Right panel: an example of colored tessellation (Guillot et al. 2005).

Since all of the 34 animals in this study were trapped at three single sites, the spatial coordinates were treated as uncertain in Geneland. When the dataset has samples sharing the same coordinates, allowing some uncertainty in the coordinates allows the program to assign samples with the same coordinates to different populations (Guillot et al. 2005). Since the animals were baited to these trapping locations with food, the trapping location is not necessarily a true location for each animal, so assigning an uncertainty on coordinates is an important consideration.

Results

When using the program Geneland, one must first consider the possible number of subpopulations that could be represented in the sample. The program then determines the most likely number of subpopulations based on the genetic and geographic data. On the first run, the potential number of populations in the sample was varied from 1 to 6. The uncertainty of coordinates was first set to 0.025 because at these latitude and longitude levels (in decimal degrees: Blyde = -24.58333 Lat, 30.81667 Long; Oribi = -30.7 Lat, 30.28333 Long; Polokwane = -23.9 Lat, 29.45 Long) one degree of latitude or longitude is equal to approximately 100 km (Hutchison, 2008; zodiacal.com). Since vervets move as much as 2.5 km in a day (Struhsaker 1967), the uncertainty of coordinates was estimated at 2.5/100, or 0.025, accounting for their longest dispersal distances.

The results with an uncertainty of coordinates of 0.25 (Figure 7) indicate that the number of subpopulations is three. Once the number of populations was determined, the test was re-run for three populations to obtain a map of population membership (Figure 8) as well as three separate maps illustrating the likelihood of population membership to each of the three clusters (Figures 9-11). (The dots in figures 8-12

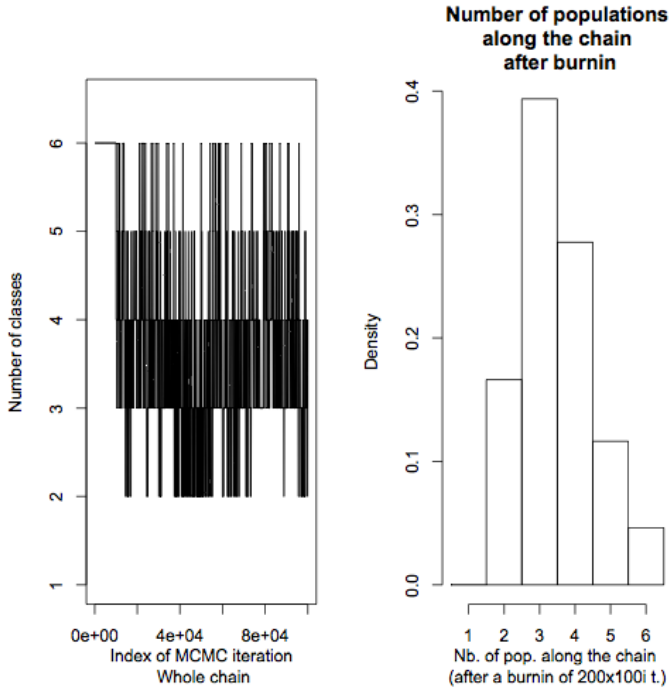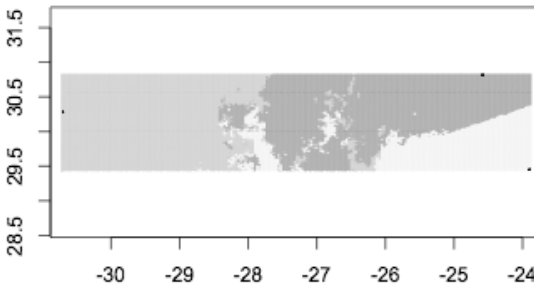represent the actual geographical coordinates of the three populations.)

Number of populations along the chain after burnin

Figure 7: Number of populations as determined by Geneland.

Posterior mode of population membership

Figure 8: Map of population membership.

Integrating Geographic Information into the Analysis of the
Genetic Distribution of South African Vervet Monkeys
K. Dore, T. Turner, J.G. Lorenz, J.P. Grobler

121

Map of posterior probability to belong to cluster 1

Figure 9

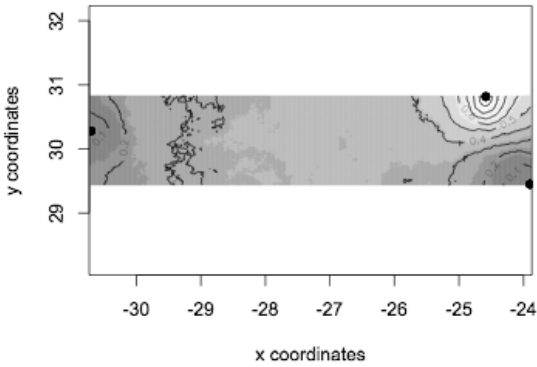**Map of posterior probability to belong to cluster 2**

Figure 10
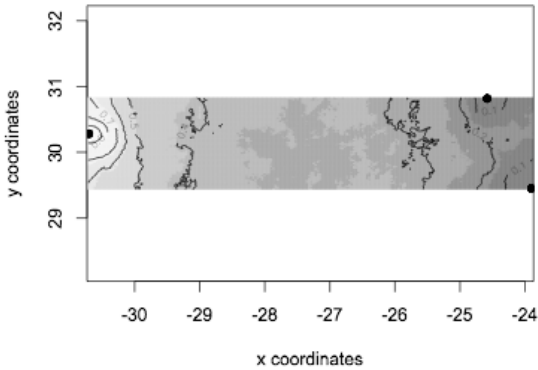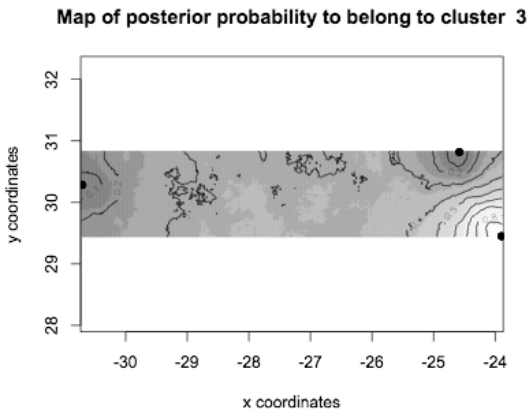
Map of posterior probability to belong to cluster 3

Figure 11

Figures 9-11: The probability that a sample belongs to a particular cluster when the degree of uncertainty is 0.025 (probability ranges from low (dark color) to high (light color)).

Another test was conducted with an uncertainty of coordinates of 1 (or approximately 100 km) to account for the fact that the actual dispersal distances for the trapped populations are unknown and to increase the probability that individuals from different geographical areas could be assigned to the same population. This is to ensure that the spatial data alone is not driving the results. The results of this high spatial uncertainty test show the number of populations is still three, though there are some interesting differences. Once the number of populations was again determined to be three, the test was re-run for three populations. The resulting map of population membership (Figure 12) shows less mixing between populations (this is inferred – there is no data for this area). Also, in the three maps illustrating the likelihood of population membership (Figures 13-15), one can see that Blyde and Polokwane now have a much higher likelihood of belonging to the same population.

Integrating Geographic Information into the Analysis of the   123
Genetic Distribution of South African Vervet Monkeys
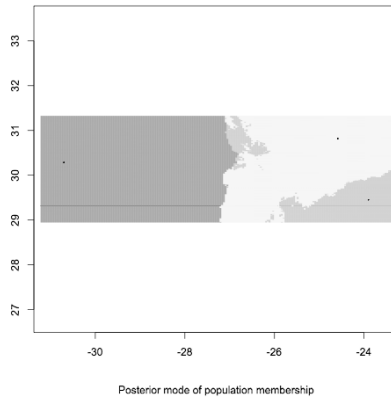K. Dore, T. Turner, J.G. Lorenz, J.P. Grobler

Posterior mode of population membership

Figure 12: Map of population membership with a degree of spatial uncertainty.

Figure 13

Map of posterior probability to belong to cluster 2
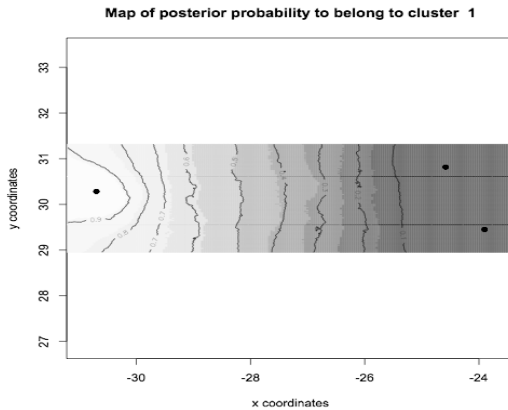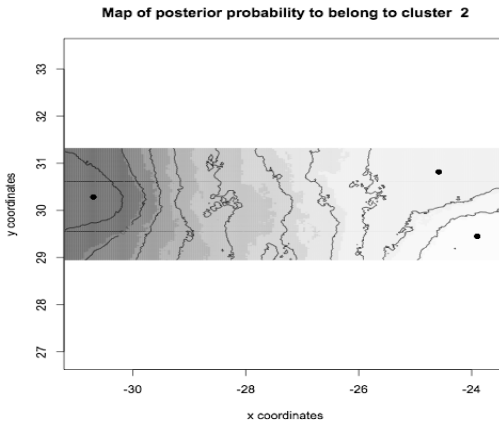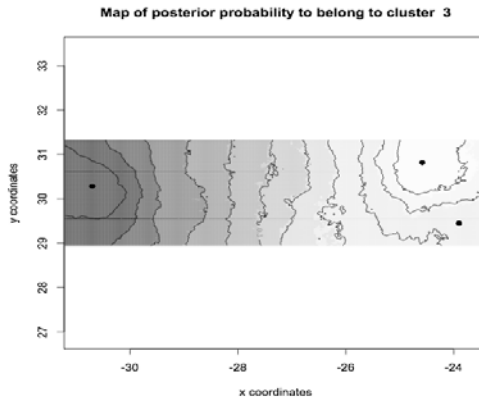

Figure 14

Map of posterior probability to belong to cluster 3

Figure 15

Figures 13–15: The probability that a sample belongs to a particular cluster when the degree of uncertainty is 1 (probability ranges from low (dark color) to high (light color)).

Discussion and Conclusion

The incorporation of geographic information into this analysis with the program Geneland has shown that, in contrast to the Structure results, Blyde, Oribi, and Polokwane are weakly differentiated. It is important to consider some of the assumptions of Geneland before concluding that these are three distinct populations. First, Geneland assumes uniformity of sampling over the entire study area. It is clear that for this investigation there are only three trapping sites across a rather

large geographic area; this forces the program to make conclusions based on limited coverage. Also, the way the program is designed, the geographic information is analyzed before the genetic data is considered. This imposes a "penalty" for assigning animals from geographically different areas to the same population. Furthermore, it is even more difficult to assign animals from different trapping sites to the same population when you place a low uncertainty on the coordinates. In the first run, an uncertainty of 0.025 was used to mimic the longest dispersal distances of vervet monkeys. This resulted in three very distinct populations. When the uncertainty was increased to an entire coordinate, the individuals from the Blyde and Polokwane populations became more likely to be members of the same population. This shows that as specificity on geographic location decreases, these populations become more likely to be considered similar.

This is further illustrated by the fact that when geographic data are not considered, such as when analyzed in Structure, the animals from these three areas are considered members of the same population. These results are in contrast with the results obtained by Latch et al. (2008). In this investigation of the North American river otter in Louisiana, both the spatially implicit Structure and the spatially explicit Geneland found the individuals to be clustered into three populations. Since Structure and Geneland do not coincide on this dataset, it seems likely that these individuals are differentiated, though not enough to be considered distinct populations. This makes sense; animals located hundreds of kilometers apart, as Oribi is from Polokwane and Blyde, should be genetically different. This is especially true when one considers that in the country of South Africa dispersal is not unlimited.

These data are preliminary. There are additional data available for up to 12 trapping sites spanning the geographic range of South African vervet monkeys. It will be very

interesting to see the results obtained when all of the information is analyzed. Considering the current policies regarding the translocation of vervet monkeys around South Africa, continuing the restrictions on the mixing of vervets from different geographical areas could result in further differentiation between these populations and the eventual development of subspecies. Studies such as this, that quantify genetic differences between populations, can be very effective in making the best conservation decision(s) for a primate species.

References Cited

Evanno, G., S. Regnaut and J. Goudet
2005   Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Molecular Ecology* 14: 2611-2620.

Grobler, P.J. and M.J. Matlala
2002   Regional genetic variability among South African vervet monkey (*Chlorocebus aethiops*) populations. *Acta Theriologica* 47(2):113-124.

Grobler, P.J., M. Jacquier, H. de Nys, M. Blair, P. Whitten and T. Turner
2006   Primate sanctuaries, taxonomy and survival: a case study from South Africa. *Ecological and Environmental Anthropology* 2: 12-16.

Goodman, S.J.
1997   RST CALC: A collection of computer programs for calculating unbiased estimates of genetic differentiation and determining their significance for microsatellite data. *Molecular Ecology* 6: 881-885.

Guillot, G., A. Estoup, F. Mortier, and J-F. Cosson
2005   A spatial statistical model for landsdcape genetics. *Genetics* 170: 1261-1280.

Latch, E.K., D.G. Scognamillo, J.A. Fike, M.J. Chamberlain and O.E. Rhodes, Jr.

Integrating Geographic Information into the Analysis of the   127
Genetic Distribution of South African Vervet Monkeys
K. Dore, T. Turner, J.G. Lorenz, J.P. Grobler

2008   Deciphering ecological barriers to North American
        river otter (*Lontra canadensis*) gene flow in the
        Louisianna
        landscape. *Journal of Heredity* 99: 265-274.
Lorenz, J.G., J.P. Grobler, K. McAuliffe Dore, and T.R. Turner
(In press) Preserving Primate Biomaterials for Future
        Research: Field Research and the Integrated Primate
        Biomaterials and Information Resource. In: *Molecular
        Primatology*, ed. A. DiFiore.

Meester, J., I.L. Rautenback, N.J. Dippenaar and C.M. Baker
1986   Classification of southern African mammals. *Transvaal
        Museum Monograph* 5:1-359.
McAuliffe, K.
2008   *Genetic variability in three South African vervet
        monkey (*Cercopithecus aethiops*) populations*. Master's
        Thesis: University of Wisconsin-Milwaukee.

Pritchard, J.K., M. Stephens and P. Donnelly
2000   Inference of population structure using multilocus
        genotype data. *Genetics* 155: 945-959.

Skinner, J.D. and C.T. Chimimba
2005   *The Mammals of the Southern African Subregion*. New
        York: Cambridge University Press.
Struhsaker, T.T.
1967   Social structure among vervet monkeys *(Cercopithecus
        aethiops). Behaviour* 29: 83-121.