ANALYSIS OF THE POPULATION GENETICS OF IMPALA, *AEPYCEROS MELAMPUS*, IN SOUTHERN AFRICA USING PROTEIN ELECTROPHORESIS

Gavin John Fleming

A Dissertation Submitted to the Faculty of Science of the University of the Witwatersrand, Johannesburg, in fulfilment of the requirements for the Degree of Master of Science.

Johannesburg, 1996.

DECLARATION

I declare that this dissertation is my own unaided work. It is being submitted for the degree of Master of Science in the University of the Witwatersrand, Johannesburg. It has not been submitted before for any degree or examination in any other university, nor has it been prepared under the aegis or with the assistance of any other body or organisation outside the University of the Witwatersrand, Johannesburg.

Gavin John Fleming

26th day of January, 1995

ABSTRACT

Impala are an ancient and successful species whose biogeography differs from other bovids. A detailed electrophoretic investigation of genetic variability within and between subpopulations found six polymorphic loci, *CK-C**, *GPI**, *MPI**, *PEP-B**, *PGM-2** and *PROT-2** in a sample of 464 impala collected from 10 localities in southern Africa. Average gene diversity was 0,047. Between-population gene diversity was normal for bovid species. Allele frequency differences and genetic distances revealed low levels of subdivision into three broad regions. Wright's F_{ST} (0,035) revealed a significant yet low level of population subdivision. The distributions of single-locus heterozygosities and allele frequencies were significantly different to those predicted during mutation-drift equilibrium, indicating that non-equilibrium conditions may prevail and that the population may be recovering from a recent bottleneck.

To my wife, Bridget

ACKNOWLEDGMENTS

For impala samples, I would like to thank staff at the following nature reserves: Savannah Wildlife, Central Estates, Zimbabwe; Nylsvlei Game Reserve; Klaserie Private Nature Reserve; Kruger National Park; Natal Parks Board / kwaZulu Bureau of Natural Resources (Mkuzi, Albert Falls and Tembe); Londolozi Game Reserve; Mala Mala Game Reserve; Marloth Park Local Government Affairs Council; Pongolapoort Game Reserve.

Stewart Grant provided me with a firm footing in the field of population genetics, and widened my perspective of other biological questions. I am grateful for comments on earlier manuscripts by Stewart Grant, Jason van Boukering, Freddy Sitas, Nancy van Schaik and Angus and Bridget Fleming. David Russell helped produced some cluster diagrams. Vincent Naidoo helped with printing. Willem Ferguson examined final drafts, suggested corrections and kindly made his time and computer available. Bridget, my adoring wife, helped produce presentable figures and kept me going through a significant number of testing times.

Funding was provided by the FRD, the University of the Witwatersrand and the Richard Ward Endowment Fund.

CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1

INTRODUCTION

Population genetic mechanisms are fundamental to evolution. Therefore, genetic arguments are important arguments to use in any discussion of evolutionary mechanisms. Population genetic theory, which aims to describe and account for genetic variability within and among populations, has been extensively developed this century. However, much of population genetic theory is based on simplified and untested assumptions. For an improved understanding of evolutionary mechanisms, it is important to test hypotheses designed to clarify details in population genetic models (Weir 1990; Grant 1993). One of the most effective ways of doing this is to pose hypotheses based on aspects of theory and to test them with gene frequency data from natural populations. Population genetic analysis is also an essential aid to systematics (Hillis and Moritz 1990) and to the conservation of genetic resources (Soulé and Wilcox 1980; Frankel & Soulé 1981; Vrijenhoek et al. 1985).

Climate change and other environmental factors are the underlying causes of evolution. In the case of the Bovidae (antelope) in Africa, changes in environment, such as a proposed spreading of grassland near 2,5 million years ago, have been responsible for major spurts of evolution (Vrba 1985a; 1985b). These spurts have caused most bovid tribes to have more than one member. However, the impala is the only member of the tribe Aepycerotinae, extant or extinct (Vrba 1985b). One of the aims of this study was to investigate reasons for the lack of divergence within the impala lineage. Chapter 4 addresses this question, while investigating within- and between-population genetic variability in detail.

Protein electrophoresis is a desirable method with which to measure genetic variability, since we assume that genotype can be directly inferred from electrophoretic phenotype. Allelic frequencies thus obtained can

provide detailed information on the distribution of variability within species. In this study, allelic frequency data are used to make statistical inferences regarding genetic variability in impala.

Before an effective electrophoretic study can take place, the variability in expression of protein loci must be considered. The analysis of the tissue distribution of gene expression and the number of loci coding for each protein are covered in Chapter 3. By comparing gene expression in different species, inferences regarding bovid evolution can be made.

The social structure of a species has a strong bearing on two of the most important population genetic parameters, gene flow and random genetic drift (Slatkin 1987). One important aspect of social structure is mating behaviour. For example, male impala disperse more than females and are thus expected to contribute more to gene flow than are females. The hypothesis that male impala contribute more to gene flow than do females is also investigated in Chapter 4. The results of this investigation into the effects of sex-biased dispersal show how genetic processes cannot always be predicted from life history observations.

Since population genetics is a quantitative science, large sample sizes are required to make powerful statistical tests. For this study one of the largest bovid tissue samples yet assembled was used to test hypotheses regarding biogeography, population genetics and sex-biased dispersal in impala.

Chapters 3 and 4 are natural divisions of the main body of work. They were written with the aim of submitting them for publication in peer-reviewed journals, so they are self-contained. Literature reviews and the background and aims of each part of the study are therefore in the introductions to each chapter.

Materials and methods were combined to avoid repetition. Chapter 2 describes the collection of samples, enzyme electrophoresis and statistical methods used for the descriptions and analyses of variability in Chapters 3 and 4. General conclusions are presented in Chapter 5.

provide detailed information on the distribution of variability within species. In this study, allelic frequency data are used to make statistical inferences regarding genetic variability in impala.

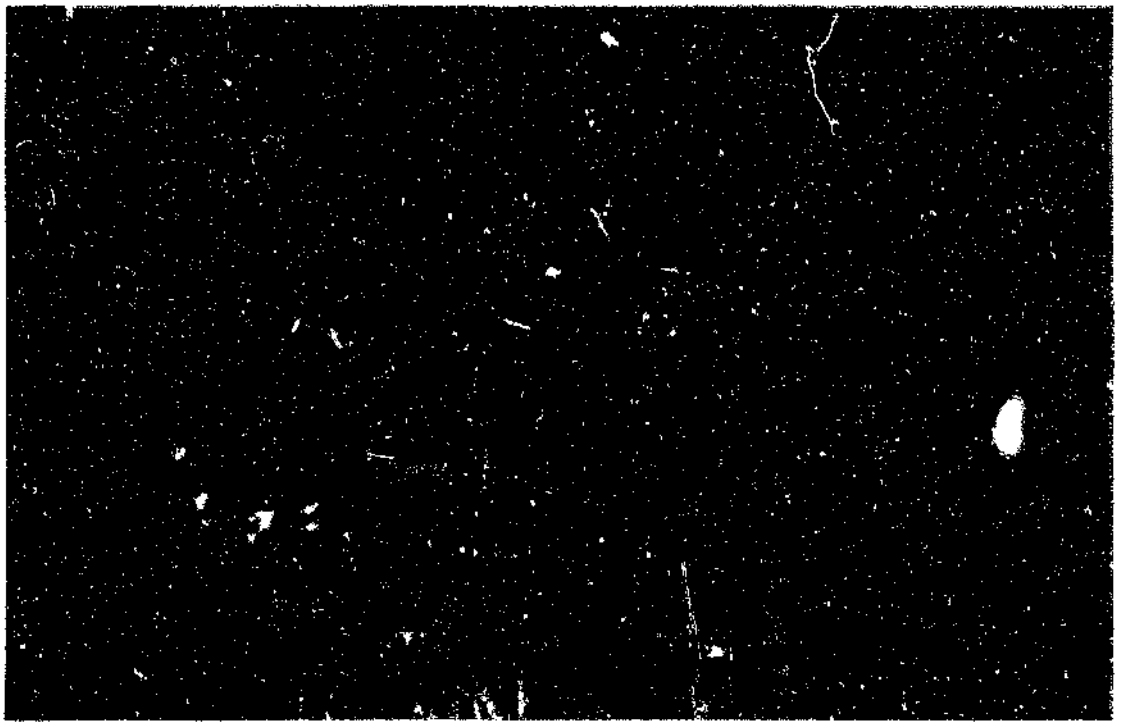
Before an effective electrophoretic study can take place, the variability in expression of protein loci must be considered. The analysis of the tissue distribution of gene expression and the number of loci coding for each protein are covered in Chapter 3. By comparing gene expression in different species, inferences regarding bovid evolution can be made.

The social structure of a species has a strong bearing on two of the most important population genetic parameters, gene flow and random genetic drift (Slatkin 1987). One important aspect of social structure is mating behaviour. For example, male impala disperse more than females and are thus expected to contribute more to gene flow than are females. The hypothesis that male impala contribute more to gene flow than do females is also investigated in Chapter 4. The results of this investigation into the effects of sex-biased dispersal show how genetic processes cannot always be predicted from life history observations.

Since population genetics is a quantitative science, large sample sizes are required to make powerful statistical tests. For this study one of the largest bovid tissue samples yet assembled was used to test hypotheses regarding biogeography, population genetics and sex-biased dispersal in impala.

Chapters 3 and 4 are natural divisions of the main body of work. They were written with the aim of submitting them for publication in peer-reviewed journals, so they are self-contained. Literature reviews and the background and aims of each part of the study are therefore in the introductions to each chapter.

Materials and methods were combined to avoid repetition. Chapter 2 describes the collection of samples, enzyme electrophoresis and statistical methods used for the descriptions and analyses of variability in Chapters 3 and 4. General conclusions are presented in Chapter 5.



CHAPTER 2

MATERIALS AND METHODS

2.1 Sample Collection

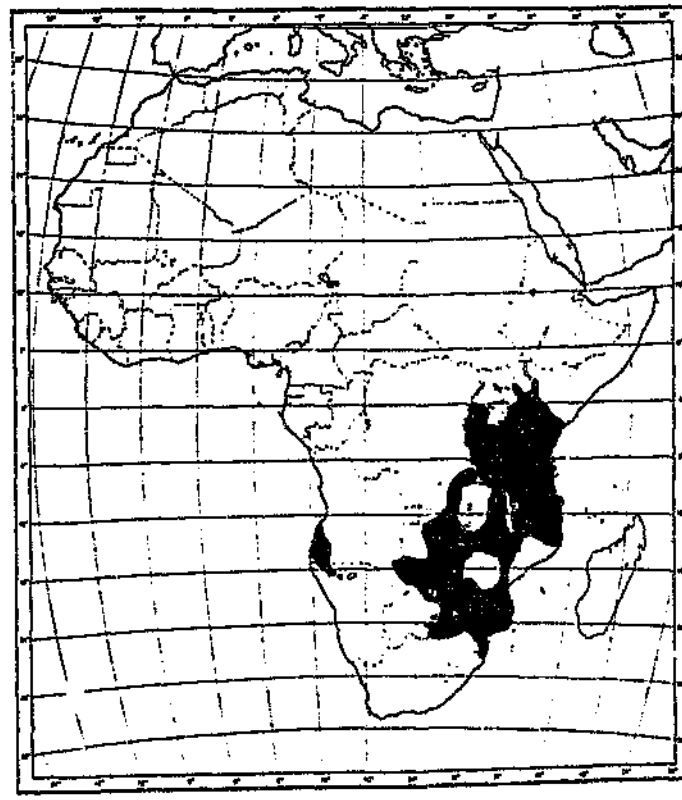
Tissue samples were collected from ten cull localities in South Africa and Zimbabwe soon after the death of the impala (Table 2.1; Figure 2.1). Ten tissues from ten impala from four localities were collected for the investigation of tissue expression (Table 3.1). Heart, liver, kidney and skeletal muscle showed the best enzyme expression, so samples of these four tissues were collected from 464 individuals.

Impala in Tembe and Albert Falls came originally from Mkuzi and impala in Pongola are a mix of local and Mkuzi impala. None of the other populations has received translocations of impala, so any exchange has been due to historical dispersal between reserves.

No animals were killed specifically for this study; samples were collected from routine culling operations. The project was approved by the Ethics committee of the University of the Witwatersrand.

Table 2.1 Sample sizes, localities and collection dates. Numbers in parentheses are the locality numbers indicated in the presentation of some results.

Locality	Coordinates longitude	Coordinates latitude	Sample Size total	Sex males	Sex females	Sex unknown	Collection dates	Latest Census
Zimbabwe								
Central Estates (1)	30°29'E	19°10'S	100	68	31	1	Oct '93	7000
South Africa								
Nyalvlei (2)	28°42'E	24°39'S	23	22	1	0	May '93	250
Klaserie (3)	31°10'E	24°17'S	21	2	9	10	Aug '93	12000
Herd 1			11	2	9	0		
Herd 2			10	0	0	10		
Sabi Sands(4)	31°30'E	24°47'S	83	42	41	0	Jun '93	17000
Londolozi	31°30'E	24°50'S	7	6	1	0		
Marthly	31°30'E	24°47'S	44	19	25	0		
Eyrefield	31°34'E	24°46'S	31	16	15	0		
Kruger National Park (5)	31°54'E	25°08'S	39	27	12	0	May '93	-
Skukuza	31°35'E	24°59'S	24	24	0	0		
Sabie	32°14'E	25°18'S	15	3	12	0		
Marloth Park (6)	31°47'E	25°21'S	70	34	35	1	Jun '93	500
Pongola-poort (7)	31°56'E	27°21'S	83	14	69	0	Sep '93	1100
Herd 1			18	2	16	0		
Herd 2			46	11	35	0		
Herd 3			19	1	18	0		
Mkuzi (8)	32°14'E	27°38'S	23	23	0	0	Aug '93 Sep '93	5500
Tembe (9)	32°21'E	27°00'S	9	9	0	0	Jun '93	-
Albert Falls (10)	30°26'E	27°26'S	11	8	3	0	Aug '93 Jun '93	-
North West								
Rooibokkraal	26°54'E	24°12'S	1	0	1	0	Apr '93	-
Vaalwater	28°07'E	24°12'S	1	0	0	1	Mar '93	-
Totals:			464	249	202	13		



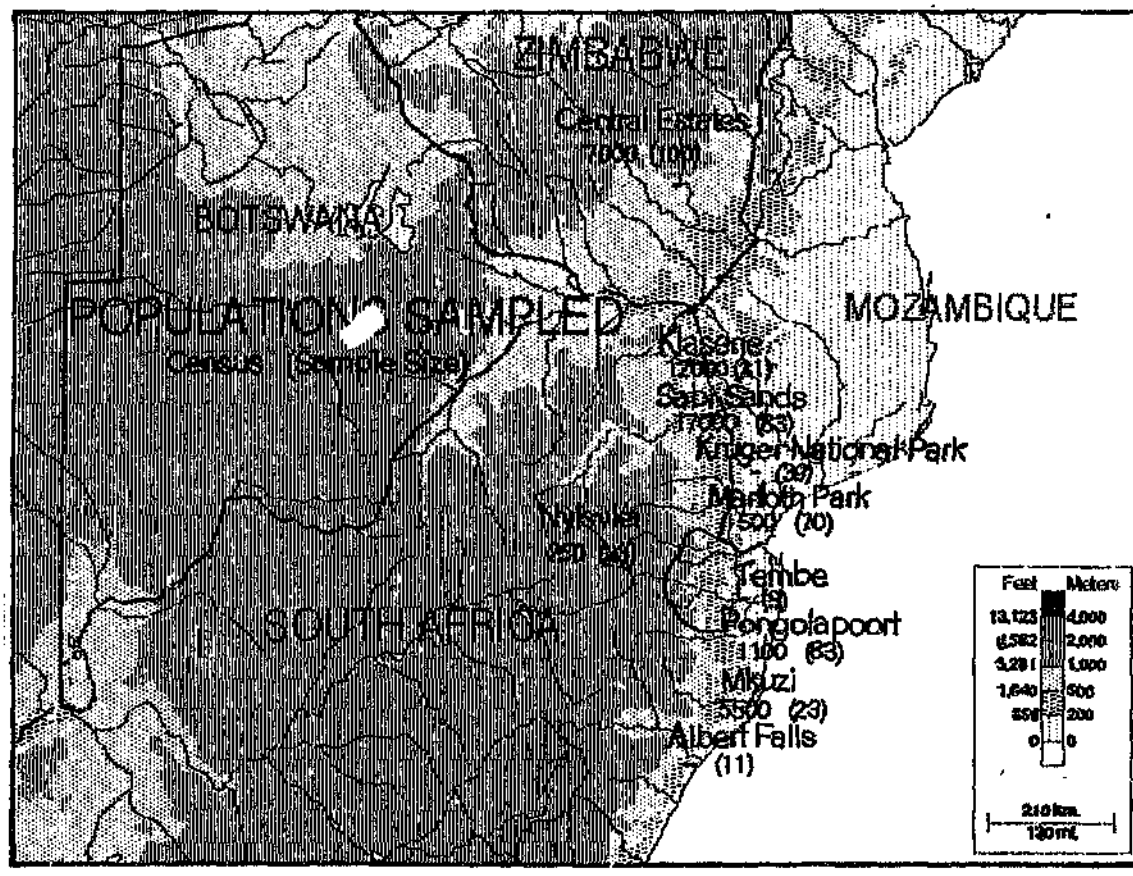


Figure 2.1 Map showing the distribution of impala (Smithers 1983; above)

2.2 Electrophoresis

After collection, the tissues were placed immediately on ice for transport to the laboratory where they were stored at -70°C. Subsequent freezing and thawing were kept to a minimum to preserve enzyme activity. I extracted soluble proteins in 0.05M Tris-HCl pH 8.0 with 0.2 mM pyridoxal-5-phosphate and performed starch gel electrophoresis (12% w/v; StarchArt Corporation, Smithville, Texas) on all extracts. Red food colouring was used as a marker dye on all gels. Buffer-tissue combinations which provided the best banding resolution are listed in Tables 3.1 and 3.2. I stained gels for enzyme activity according to Harris & Hopkinson (1976) and Grant (1988).

A sample of 200 individuals, that consisted of up to 25 individuals from each locality, was surveyed with 40 enzyme substrates for Mendelian variability at 54 presumptive loci. This was sufficient to detect rare alleles in the sample. The remainder of the total sample (an average of 170 more individuals per locus) was then surveyed for variability at polymorphic loci only, where the frequency of the most common allele, p, was $p \leq 0.95$ (Table 3.2).

2.3 Locus and Allele Nomenclature

Locus and allele nomenclature followed Shaklee et al. (1989) and Shows et al. (1987). Where subcellular origin or orthology of an enzyme with other mammals was not certain, I numbered the loci, beginning at the cathodal end of the gel. If subcellular origin was certain, I prefixed locus abbreviations with an *m* (mitochondrial) or *s* (supernatant). If orthology was well established for multilocus enzymes, a letter indicating the orthology was used. Where a suffix is used, but only one system is described or scored, it implies that there were other bands visible which I could not interpret. Enzyme or allele product abbreviations are in plain type: AAT, 100. Locus abbreviations are italicised and followed by an asterisk (e.g. *AAT**). Alleles are denoted by an asterisk followed by the italicized relative mobility of the corresponding band on the gel

(e.g. *"100*). *AAT"100* denotes the 100 allele of the AAT gene. Phenotypes are written without italics (e.g. LDH-A 107/100) and genotypes in italics (e.g. *LDH-A"107/"100*). I used the relative mobility system for designating alleles, and used the most common allele at each locus as the electrophoretic standard for that locus. Conditions for defining allelic mobilities were as in Table 3.2, unless otherwise mentioned.

2.4 Description and Interpretation of Banding Patterns

Mendelian variability was inferred from banding patterns according to the principles of Harris and Hopkinson (1976), Harris (1980), Allendorf et al. (1977), Grant (1988), Nei et al. (1978) and Zouros (1976), because no family studies were possibl For example, there should be parallel expression of true genetic variability amongst tissues expressing the same locus, and expression should be similar in closely related taxa (Fisher and Whitt 1978; Fisher et al. 1980; Tabachnick and Howard 1982). Genotypic frequencies in samples with correctly interpreted patterns should fit Hardy-Weinberg proportions if they were taken from a single randomly mating population. Banding patterns were scored only if individual bands were distinct. The presence of an alternate allele was confirmed if: apparent heteromeric forms were observed; a definite alternate homomer was observed; and if banding patterns conformed to the patterns expected for the known subunit structure of that enzyme (Darnall and Klotz 1972). Most enzyme systems were run on different buffers to resolve isoloci or cryptic allozymes.

2.5 Microsatellite Analysis

Microsatellite analysis (Weber and May 1989; Litt and Luty 1989) is an effective way of measuring variability in nuclear DNA in large numbers of individuals. Primer sequences were extracted from GENBANK and suitable primers were designed from these using the program PRIMER. Seven primer pairs were synthesized by the department of Biochemistry at the University of Cape Town. DNA was extracted from impala and cow (positive control) tissue (Ali and Jameel 1993). A large amount of time was

spent performing PCR on these samples with the microsatellite primers, and attempting to resolve PCR products on Metaphor agarose (FMC products) or polyacrylamide sequencing gels, using ethidium bromide or silver staining to visualise bands. However, only small quantities of PCR product were obtained from some of the primer pairs, and resolution was not sufficient to resolve bands that differed in length by as little as two base pairs. Therefore, although variability was evident, this technique was regretfully abandoned.

2.6 Statistical Analyses

2.6.1 Tests for fit to Hardy-Weinberg equilibrium

I performed a contingency-table analysis of genotypic counts and used the log-likelihood ratio test (G-test, Sokal and Rohlf 1995) to test for departures from Hardy-Weinberg proportions. I used Levene's (1949) unbiased correction for small samples and performed a sequential Bonferroni test ($\alpha' = \alpha / k$, where k is the rank of the probability) to obtain an experimentwise error rate of $\alpha = 0,05$ (Cooper 1968; Rice 1989; Sokal and Rohlf 1995). I performed the tests with and without pooling of low frequency alleles.

2.6.2 Tests for allele-frequency homogeneity

To test for genetic subdivision among subpopulations I used log-likelihood ratio G-tests on a nested sample design that was based on geographical distances between sample localities and physical barriers such as fences and rivers (Figure 2.2). Referring to Figure 2.2, 'countries' are separated by more than 570km. 'Regions' are separated by between 190km and 390km. 'Major localities' are separated by between 30km and 70km as well as by physical barriers. Pongola is grouped with Natal because it is closer to Natal populations than to Eastern Transvaal populations and its large local population has received some impala from Mkuzi. 'Minor localities' in Natal are geographically distant from each other, but are effectively close because of founding events two decades

ago. So I considered their level of subdivision to be intermediate between that of 'herds' and that of 'major localities'. Impala at the next level were sampled less than 10km apart, except at Kruger National Park (38km), and were not separated by any barriers. 'Herds' were breeding groups of females with one or two males that were corralled together. 'Within park samples' were not necessarily herds and were taken on separate occasions from distinct sites. Table 2.2 is the model pictured in Figure 2.2. It excludes a sixth level, 'herds and within park samples', because this level was only used in one test.

Table 2.2 Model of comparisons and their degrees of freedom for nested contingency table analysis of allele-frequency homogeneity. This model is shown graphically in Fig 2.2.

Level/Comparison	d.f.
1 Total	9
5 Between Zimbabwe and South Africa	1
Within RSA	(8)
4 Among three regions (Nylsvlei, Lowveld, Natal)	2
3 Among parks in Lowveld region	3
3 Between Natal and Pongola	1
2 Among parks in Natal	2

I made a two-way contingency table analysis of allelic frequencies at each level, and used sequential Bonferroni α values (α') across loci. Large sample sizes are needed to detect small differences, especially between intermediate allele frequencies (Grant 1988). For example, for a 60% chance ($1-\beta=0.6$) of detecting a difference of 0,05 between frequencies of 0,10 and 0,05, sample size must be 156, and for a 90% chance ($1-\beta=0.9$), it must be 310 (Sokal and Rohlf 1995). My maximum sample size was 100, so the power of my tests to detect small allele frequency differences was not large.

Figure 2.2 Nested sample design for analysis of allele frequency homogeneity and gene diversity analysis. Localities are arranged approximately NW (Zimbabwe) to SE (Natal). Nested design is described in the text. Mpumalanga is a province in the Lowveld region of southern Africa. Each asterisk represents a locus at which there were significant

2.6.3 Geographic allele frequency distributions

Common allele frequencies at each locality were plotted for easy graphical comparison of geographic variability (Hubbs and Hubbs 1953). Localities were arranged in a predetermined geographical order approximating that in Figure 2.2, and not according to the pattern of allele frequencies. Error bars representing approximate 95% confidence intervals were calculated as the allele frequency plus or minus two standard errors, so 95% C.I. $\approx p \pm 2[\sqrt{(pq/2n)}]$ where p is the allele frequency, $q = (1 - p)$ and n is the sample size.

2.6.4 Cluster analysis of genetic distances

UPGMA (Sneath and Sokal 1973) and neighbour-joining (Saitou and Nei 1987) clustering and multidimensional scaling analysis were performed on genetic distances with programs by W.S. Grant and NTSYS-pc (Exeter Software). I estimated genetic distances with Nei's (1978) unbiased genetic distance with Hillis' (1984) correction for small sample sizes (D_N*), Rogers genetic distance (Rogers 1972) with Wright's (1978, p91) modification for small genetic distances (D_m), and Cavalli-Sforza & Edward's chord distance (D_c) (Cavalli-Sforza and Edwards 1967; Wright 1978 p92). Cluster analysis based on Cavalli-Sforza and Edwards (1967) chord distance gives the greatest separation between samples, since this distance is more appropriate than Nei and Rogers' distances and overcomes some of their limitations. Neighbour-joining makes fewer assumptions about the data, merely that they are additive, and not ultrametric, which the UPGMA assumes (Swofford and Olsen 1990).

2.6.5 Gene diversity analysis

Observed heterozygosity, H, was calculated from genotype frequencies. Expected gene diversities H_e and H_e(unbiased) (Crow and Kimura 1970; Nei 1978) and their variances (Nei and Roychoudhury 1974) were calculated for each locality and for the whole population. Estimates of H in small samples were not compromised by the sample size, because many loci were

used (Nei 1978). I also estimated the variance of gene diversity by using M from $H = M / (1 + M)$, where $M = N_e v$, in which N_e is the effective population size and v is the mutation rate (Kimura and Crow 1964; Fuerst et al. 1977). Variance of single locus heterozygosity from this estimate of M was calculated according to Watterson (1974), Stewart (1976), Li and Nei (1975) and Fuerst et al. (1977). Nested analysis of gene diversity based on 18 variable loci (Nei 1973, 1987; Chakraborty 1980; Chakraborty et al. 1982) was performed according to the hierarchy in Figure 2.2, excluding the 'herds and within-park samples' level. This analysis is a measure of differentiation between populations and describes the distribution of heterozygosities within and between subpopulations at all levels of the hierarchy.

2.6.6 Estimation of gene flow

I used three methods to estimate gene flow. Firstly, I used equations 1 to 4 of Weir and Cockerham (1984) to estimate the parameters F, f and θ, which are unbiased estimates of the parameters F_{IT}, F_{IS} and F_{ST} (Wright 1951). F_{IT} is the correlation between two randomly chosen homologous alleles in an individual relative to alleles at that locus in the whole population. It is known as the overall inbreeding coefficient as it takes into account both division into subpopulations and nonrandom mating within subpopulations. F_{IS} is the correlation between two randomly chosen homologous alleles in an individual relative to alleles at that locus in the subpopulation. It measures the deviation from random mating within subpopulations. F_{ST} is the correlation between two randomly chosen homologous alleles in a subpopulation relative to alleles at that locus in the entire population. In other words, F_{ST} measures the extent of between-population allele-frequency differentiation (due to random drift) as a function of the relatedness of genes within populations. θ explicitly takes into account different sample sizes and the number of populations sampled, and therefore depends only on population size and history, unlike Wright's F statistics (Weir and Cockerham 1984). Every allele at every locus provides an estimate of the same quantity θ. The variances of these estimates for each locus are added, and the ratio of

the sums of the observed and expected variances provides a single estimate of θ with minimal bias and variance (Weir 1990, p389).

Numerical resampling methods were used to estimate variances and confidence intervals of the *F* statistics (Weir 1990 pp383-384). Jackknifing involves dropping one observation at a time till each has been dropped once, and calculating new estimates of the *F* statistics in each case. The variances of the *F* statistics in the population are then estimated as the variances of these new sets of sample statistics. Less biased sample statistics are also calculated from the jackknifed estimates. In this study, jackknifing was performed over populations and then over loci. Bootstrapping involves making estimates by constructing many new samples of the same size by random sampling with replacement. Thus a distribution of estimates is created from which a 95% confidence interval is constructed. This way, *F* statistics could be compared to zero and to each other. In this study, bootstrapping was performed over loci. Both jackknifing and bootstrapping were performed with the Fortran program DIPLOID.FOR (Weir 1990 p405).

To estimate gene flow from F_{ST}, the infinite-island model was assumed (Weir 1990 pp391-392). *m* is the probability each generation that any gene sampled from a population migrated from any one of an infinite number of other populations. If these 'islands' have a finite size, *N*, the gain of variation by migration and the loss of variation due to drift within islands establish an equilibrium. The equilibrium value of θ, if *m* is small (Wright 1978; Wright 1951; Slatkin and Barton 1989), is

$$\theta = 1/(1+4Nm).$$

$Nm(F_{ST})$, which is the number of migrants entering each population per generation if mutation and selection are assumed to be unimportant, can be estimated from this. However, estimates of *Nm* obtained this way are only rough indications of the level of gene flow, as the infinite-island model is not realistic. Impala are distributed widely, but in scattered concentrations, often separated by barriers such as distance, fences and

rivers, none of which is impossible for an impala to traverse. The isolation by distance model (Wright 1943) was considered. However, there are sufficient barriers to prevent impala living in a continuum and I did not sample a high enough density of populations to use the isolation by distance model, so the infinite island model was considered appropriate for this study.

Secondly, a measure of the spread of the gene frequency distribution is $p(1)$, which is the average frequency of alleles that are found in only one sample (Slatkin and Barton 1989). $Nm(p(1))$ can be estimated from: $\log_{10}[p(1)] = a\log_{10}(Nm) + b$ (Slatkin and Barton 1989). So, $Nm \approx 10\exp[(\log_{10}[p(1)] - b)/a]$, where a and b depend on the sample sizes from each population, and were determined empirically by Slatkin and Barton (1989) for a set of three sample sizes. If Nm_{est} is calculated with the sample size from this set (N_{ref}) that is closest to the actual sample size (N_{sam}), then a more accurate estimate of Nm which takes into account the actual sample size, is $Nm_{adj} = (N_{ref}/N_{sam}).Nm_{est}$. I used the average of sample sizes as N_{sam}. Then $Nm(p(1)) = Nm_{adj}$. Slatkin and Barton (1989) found that the number of loci sampled has a greater effect on the accuracy of estimates of gene flow than the does the number of samples, and that the relationship between $p(1)$ and Nm depends only weakly on mutation and selection at a loc[illegible].

Although they should produce similar results most of the time, F_{ST} can be more reliable than $p(1)$. While F_{ST} uses all of the gene-frequency data, $p(1)$ is sensitive to errors from small sample sizes and misreading gels, which lead to misclassification of, or missing, low-frequency private alleles. Furthermore, F_{ST} can be reasonably estimated from any polymorphic locus, whereas to be estimated accurately, $p(1)$ requires a large number of loci to be sure of finding any private alleles that may exist. In this study, large sample sizes and many loci were surveyed for private alleles and band interpretations were carefully confirmed, so the use of $p(1)$ was deemed appropriate.

Unbiased estimates of $Nm(F_{ST})$, namely $Nm(F_{ST})^*$, were made by jackknifing over loci (Weir 1990 p383-384; Grant and Little 1992). Unbiased estimates of $Nm(p(1))$, namely $Nm(p(1))^*$, were made by jackknifing over private alleles.

The relative contributions to gene flow by either sex were estimated with $Nm(F_{ST})$, $Nm(F_{ST})^*$, $Nm(p(1))$ and $Nm(p(1))^*$ after splitting allele frequencies by sex at each locality. In this way, all-male or all-female populations were considered.

Thirdly, the distributions of conditional average allele frequencies provided a qualitative gauge of relative gene flow in either sex (Slatkin 1981). If the occupancy number i of an allele is the number of demes in which it is present, then the conditional average frequency $p(i)$ of each occupancy number is the average frequency of all alleles with that occupancy number. The Kolmogorov-Smirnov goodness-of-fit test was used to test the null hypothesis that $p(i)_{males}$ and $p(i)_{females}$ were distributed identically (Sokal and Rohlf 1995 pp434-439).

2.6.7 Tests for fit to neutral expectations

The Kolmogorov-Smirnov goodness-of fit test was used to test the null hypothesis that the observed distribution of single-locus gene diversities was identical to that expected under the drift-mutation hypothesis for $H = 0,05$ (Fuerst et al. 1977; Sokal and Rohlf 1995 pp434-439, 708-715; Sokal and Rohlf 1969 pp571-575). I also compared relationships between average gene diversity and the interlocus variance of gene diversity, and between gene diversity and the proportion of polymorphic loci, to those expected under the stepwise mutation model, the infinite alleles model and the infinite alleles model with varying mutation rate (Fuerst et al. 1977).

The observed allele frequency distribution was tested for goodness-of-fit to the distribution expected in an equilibrium population with the same level of gene diversity, assuming random mating and infinite neutral alleles. The null hypothesis was that the sample values came from the same distribution as the theoretical values. The test of this hypothesis

was as for the distribution of single-locus gene diversities. In species with low gene diversities, an excess of rare alleles may reflect a recent increase in population size after a major bottleneck (Nei et al. 1975; Maruyama and Fuerst 1984; Watterson 1984).

2.6.8 Correlation of population size with heterozygosity

The correlation of population size with heterozygosity was estimated to test the hypothesis that heterozygosity is directly proportional to population size in impala populations. Any deviation from this expectation could mean that impala populations are influenced by factors other than the inbreeding and random genetic drift that are expected in restricted populations.

CHAPTER 3

GENE EXPRESSION IN IMPALA

3.1 Introduction

As one of the most diverse mammal groups, bovids have the potential to foster understanding of evolutionary and biogeographic mechanisms. Molecular and biochemical studies on bovids have been directed towards systematics (Corbet and Robinson 1991; Georgiadis et al. 1990; Gatesy et al. 1992; Lowenstein 1986). As a result, little is known about the genetics of micro-evolutionary processes within the bovids. Knowledge of the amounts of within- and between-population genetic variability can shed light on past population events and enhance management decisions. In this chapter, I report on the distribution of tissue expression of enzymes in impala. Evolution of impala differs from that of other antelopes, which makes impala particularly interesting. The impala lineage, of which impala are the only member, has remained virtually unchanged since the Pliocene while other antelope have diverged (Vrba 1984).

A study of gene expression is important, as it aids in the choice of proteins to be used in the population genetic study and identifies the tissues that express them best. A gene expression study provides a basis for the genetic interpretation of banding variability and can also detect evolutionary changes in tissue expression, such as gene duplication, that might have occurred during speciation. Tissue expression patterns, gene duplications and the ability to form heteropolymers are classed as isozyme characters. Allozyme characters, on the other hand, represent gene variants segregating at a locus (Murphy et al. 1990).

Gene expression in a tissue depends on the requirement for a particular enzyme in the metabolic pathways in that tissue. Genes encoding metabolic enzymes are therefore expressed at different levels in different tis-

sues. Paralogous loci, which arise from gene duplication events, may be expressed in different tissues in different species (e.g. Shaklee et al. 1973 in fishes). In more ancient species, expression is more generalised, while in more recent species, expression is more tissue restrictive. This phenomenon is consistent with either changes in gene regulation or with a response to natural selection (Kettler and Whitt 1986).

To detect as many loci as possible, one must use the tissues that show the best expression and activity of the products of those loci. Strong, consistent expression of an enzyme is required for confident detection of variability. I sampled tissues representing each embryonic tissue type: ectoderm, mesoderm and endoderm. Differential expression is likely to occur among these tissues (Bezy and Sites 1987). At the same time levels of polymorphism at the loci encoding these proteins were estimated, and polymorphic loci were identified for statistical analysis. Although other molecular methods are available, protein electrophoresis is a cost-effective, well established and refined technique of investigating genetic variability (Murphy et al. 1990). To test the hypothesis that impala show no differences from other species in variability of tissue expression, heteropolymer assembly and gene number, I compared these isozyme characters with those in other bovids and with members of the Cervidae (deer), a family closely related to the Bovidae (antelope).

Methods and Materials - see Chapter 2

3.2 Results and discussion

The tissues whose extracts stained best for each enzyme were used for measuring variability. Table 3.2 shows the proteins used in this study. The products of 39 presumptive loci showed sufficient resolution and activity for scoring. If p is the frequency of the most common allele at a locus, the products of six loci were polymorphic ($p \leq 0.95$), 12 loci had rare variants ($0.95 < p < 1$) and 21 loci were monomorphic ($p = 1.00$) (Table 3.2). No isoloci or cryptic allozymes were detected. Staining was unsuccessful for 15 enzymes, namely: AH-1; AH-2; ADA-2; AO;

CAT; FBP; G6PDH; GDA; MEP-3; LAP; PEP-C; PGDH-1; PGDH-2; TAT; XO. Table 3.1 shows the relative expression of each protein in each tissue.

Descriptions of variability follow in alphabetical order of their abbreviations. A heading introduces each enzyme and a subheading provides other information in this format: Enzyme abbreviation; IUCBN (International Union of Biochemistry's Nomenclature Committee) number ('E.C.' number); quarternary structure; and alternative name, according to Harris and Hopkinson (1976) and Shaklee et al. (1989) unless otherwise mentioned.

For each enzyme, a discussion of genetic variability in impala that was observed in this study is followed by discussion of genetic variability in other mammals. Variability in foetal tissue is described in some cases. Comparison with variability in other species is necessary for the correct interpretation of variability in impala (Shaklee et al. 1989; Shows et al. 1987) and for drawing conclusions about impala evolution.

Table 3.1 Tissue distribution of gene expression in impala. +,++,+++,++++ = quality of banding (++++ = best); * = good for scoring; - = no staining; -+ = no staining in some individuals; (+) = stained, but not resolvable; n = not tested with this tissue. Buffer and enzyme abbreviations are described in Table 3.2.

		Tissue											
ENZYME	BUFFER	Brain	Eye	Muscle	Tongue	Kidney	Heart	Liver	Gut	Spleen	Lung	Pancreas	Mammary
sAAT	TCBL	+++	+(+)	++++*	+++	+++	++++*	++++*	++	++	++	+	+
mAAT	TCBL	+	-	++	-	++	++	++	+	+	+	(+)	(+)
AH	TBE	n	n	(++)	n	++++	-+++	n	n	n	n	n	n
ADA	TBE	+++	+	+++*	++	+	++	++	+++	++++	++	++	++
AK-1	TC	-	-	+++	+	+++*	++	++	-	-	-	-	-
ALAT	TCBL	n	n	++++	n	n	n	n	n	n	n	n	n
ADH	TCBL	-	-	-	-	-	-	+++*	-	-	-	-	-
AO	TCBL	n	n	n	n	n	n	-	n	n	n	n	n
CAT	TCBL	n	n	n	n	n	n	-	n	n	n	n	n
CK-A	TCBL	+++*	(+)	++++*	+++*	+	++++*	+	+	+	++	-	+
CK-C	TCBL	++++*	++	+	-	++	+++*	+	+++	+++	+	-	-
CBR	TCBL	-	-	+++*	-	+++	+	+(+)	-	+	+	(+)	(+)
EST-5	TCBL	(+)	(+)	(+)	(+)	+	(+)	++++*	+	+	++	-	-
EST-D	TCBL	-	-	-	-	-	-	++++*	-	-	-	-	-
FBALD-1	TCBL	(+)	-	++++	-	+(++)	+(+)	-	-	(-+)	(-+)	-	-
FBALD-2	TCBL	(+)	-	-	-	(+)	(+)	-	-	(-+)	(-+)	-	+
FBP	TBE	n	n	n	n	(+++)	n	n	n	n	n	n	n
FH	TBE	++	(+)	+++	+++	+++	+++	+++	++	++	++	++	++
PROT-2	TCBL	++	+	+++*	(+)	+++	+	++	+	+++	++	+	+
PROT-3	TCBL	++	+	+++*	+	+++*	+	++	+	++++	++++	++	++
GLBDH	TBE	+(+)	(+)	+	+(+)	+++*	+	+++	+	-	+	+	+
G6PDH-1	TC	-	-	+	-	-	++	+	+	+	+	+	+
G6PDH-2	TC	-	-	++	++	++	++	-	-	-	-	-	-
GPI	TCBL	++++*	++(+)*	(++++)	+++*	++++*	++++	+++*	++++*	+++*	++++*	++++	+
GAPDH	TC+NAD	-++	-	-++++*	-	-+++	-	-++	-	-++	-	-	-
G3PDH-1	TC	-	-	++	+	+++	(+)	+++	-	-	-	-	-
G3PDH-1	TC+NAD	-	-	+++	-	++++*	-	+++	-	-	-	-	-
G3PDH-2	TC	-	-	++++	++	++++	+++	+++	-	-	-	-	-
G3PDH-2	TC+NAD	-	-	+++	-	++++*	-	+(+)	-	-	-	-	-
GDA	TBE	-	-	-	-	-	-	-	-	-	-	-	-
HK-2	TCBL	+	(+)	+	(+)	++	-+	+++	+	++	+(+)	(+)	(+)
HK-3	TCBL	++	(+)	++	+	+++	(+)	-	(+)	(+)	(+)	(+)	(+)
IDDH	TCBL	-	-	-	-	+++	-	++++*	-	-	-	-	-
IDHP-1	TC	-	-	+	++	++++*	++	-	-	-	-	-	-
IDHP-2	TC	-	-	+	++	++++*	+++	+++	+++	+++	+++	-	-
LDH-A	TCBL	+(+)	(+)	+++*	(+)	-	-	-	-	-	-	-	-
LDH-B	TCBL	++++	+++*	+++*	+++*	(++++)	(++++)	+++	+++	+++	+++	+	++++
MDH-1	TC	+++	+	+++	++	++++*	++	++	++	++	++	++	++
MDH-1	TC+NAD	++	+	++++*	++(+)	+++	++	++	+	(+)	(+)	(+)	(+)
MDH-2	TC	+	-	+	++	+++*	+	+	+	+	+	+	+
MDH-2	TC+NAD	+++	++	++++*	+++	+++	+++*	+++*	+	++	+++	++	++

...continued...

Table 3.1 continued

		Tissue											
ENZYME	BUFFER	Brain	Eye	Muscle	Tongue	Kidney	Heart	Liver	Gut	Spleen	Lung	Pancreas	Mammary
MEP-1	TC	-	(++)	+++*	-	+++	++(+)	+(+)	-+	++	-	-	-
MEP-2	TC	++	-	+	-	-	+++	+-	+	+	++	-	-
MEP-3	TC	+(++)	-	-+++	-	++	+++	+++	-++	+++	+++	n	n
MPI	TBE	(++)	(+)	++++*	+++	+++*	+++	+++	+++	+++	+++	n	n
NP	TBE	(+)	(++)	-	-	++++*	+	+	++	+++	+++	++	++
PEP													
Lap	TBE	-	-	-	-	-	-	-	-	-	-	-	-
Lgg-1	TBE	+++	++	+++	++	++++*	++	+++*	+++	+++	+++	+++	++
Lt-1	TBE	++	+	+	+	(++++)	++	+++*	++	+++	+++	+++	n
Phg	TBE	+	+	++	+	+++*	+	(+++)	++	++	++	+	+
PGM-1	TCBL	++	-	++++	++	+	++	++(+)*	-	+	+	++	-
PGM-2	TCBL	++	(+)	+++	+++	++++*	+	+++(+)*	++(++)*	++++*	++++*	++++*	+++
PGDH-1	TC	+	+	++	+++	++++*	+++	++	++	++	++	+++	+++
PGDH-2	TC	-	-	++	+++	++++*	++	+	+	+	+	-	-
PK-2	TC	-	-	-(+++)	-	-	-	-	-	-	-	-	-
PK-3	TC	+++	-	+++	-+	+++	+++	+++	-	-++	-	++	++
PK-4	TC	+++	-	+++	-++	+++	-	-	-	-	++	++	++
SOD	TBE	n	n	++	n	++	n	++	n	n	n	n	n
TAT	TBE	-	-	(++)	-	-	-	-	-	-	-	-	-
XD	TBE	-	-	-	-	-	-	+(++)	-	-	-	-	-

Table 3.2 Genes and proteins studied in impala, and allele frequencies and sample sizes at non-monomorphic loci. Numbers in italics are sample sizes (n) for each locus at each locality. Number of genes sampled is $2n$. Buffers and tissues were those used for scoring and confirmation. TCBL = tris-citrate-borate-lithium (Ridgway et al. 1970; gel pH 8.7; electrode pH 8.0); TBE = tris-borate-EDTA (Markert and Faulhaber 1965; pH 8.6); TC = tris-citrate (Whitt 1970; pH 6.9); TCP = tris-citrate-phosphate (Grant 1988; pH 6.7); TG = tris-glycine (Grant 1988; pH 8.5); TM = tris-maleic acid (Georgiadis et al. 1990; pH 7.4). L = liver, H = heart, K = kidney, M = skeletal muscle. The following proteins failed to appear as bands on a gel or were not resolved: AH-1; AH-2; ADA-2; AO; CAT; PEP; G6PDH; GDA; MEP-3; LAP; PEP-C; PGDH-1; PGDH-2; TAT; XO. continued...

PROTEIN NAME	IUBCN NUMBER	BUFFER	TISSUE	LOCUS	ALLELE	Allele frequencies and sample sizes: Zimbabw	Nylsvlei	Klaserie	Sabi Sands	Kruger Park	Marloth Park	Pongola	Tembe	Albert Falls	Mkuze	TOTAL
Polymorphic proteins						*79*	*23*	*19*	*74*	*34*	*60*	*56*	*8*	*10*	*22*	*383*
Creatine kinase	2.7.3.2	TCBL,TG	H	*CK-C**	*100	0.873	0.848	0.974	0.973	0.985	0.967	0.982	1.000	1.000	0.955	0.947
					*105	0.127	0.152	0.026	0.027	0.015	0.033	0.018	0.000	0.000	0.045	0.053
						100	*23*	*19*	*83*	*39*	*68*	*81*	*9*	*11*	*23*	*459*
Glucose-6-phosphate isomerase	5.3.1.9	TCBL,TG	L,H	*GPI**	*-100	1.000	1.000	0.868	0.988	0.987	0.971	0.840	0.722	0.682	0.587	0.925
					*100	0.000	0.000	0.132	0.012	0.013	0.029	0.160	0.278	0.318	0.413	0.075
						100	*23*	*10*	*83*	*39*	*70*	*80*	*9*	*8*	*23*	*447*
Mannose-6-phosphate isomerase	5.3.1.8	TBE	M,K	*MPI**	*100	0.710	0.761	0.600	0.687	0.705	0.714	0.650	0.611	0.500	0.609	0.686
					*113	0.285	0.239	0.350	0.307	0.282	0.279	0.325	0.389	0.500	0.391	0.304
					*76	0.000	0.000	0.050	0.005	0.013	0.007	0.025	0.000	0.000	0.000	0.009
					*103	0.005	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.001
						100	*23*	*21*	*83*	*39*	*68*	*83*	*9*	*11*	*23*	*462*
Tripeptide aminopeptidase	3.4.11.4	TBE	K,H	*PEP-B**	*100	0.975	1.000	0.857	0.958	0.962	0.941	0.976	0.833	0.727	0.891	0.949
					*82	0.025	0.000	0.143	0.042	0.038	0.059	0.024	0.167	0.273	0.109	0.051
						100	*23*	*19*	*83*	*39*	*70*	*83*	*9*	*11*	*23*	*482*
Phosphoglucomutase	5.4.2.2	TBE,TCBL	L,K	*PGM-2**	*100	0.775	0.609	0.553	0.693	0.615	0.764	0.542	0.500	0.409	0.435	0.654
					*89	0.185	0.391	0.395	0.277	0.308	0.214	0.392	0.333	0.591	0.478	0.301
					*95	0.040	0.000	0.053	0.030	0.077	0.021	0.066	0.167	0.000	0.087	0.045
						98	*21*	*21*	*83*	*39*	*65*	*82*	*7*	*11*	*22*	*451*
General (unidentified) protein	-	TBE,TG,TCBL	K,M	*PROT-2**	*100	0.770	0.976	0.667	0.699	0.692	0.685	0.640	0.571	0.773	0.682	0.713
					*71	0.230	0.024	0.333	0.247	0.295	0.277	0.280	0.429	0.227	0.316	0.256
					*87	0.000	0.000	0.000	0.042	0.013	0.038	0.073	0.000	0.000	0.000	0.028
					*120	0.000	0.000	0.000	0.012	0.000	0.000	0.006	0.000	0.000	0.000	0.003
Proteins with rare variants						*59*	*23*	*19*	*26*	*24*	*25*	*21*	*9*	*11*	*20*	*239*
Aspartate aminotransferase	2.6.1.1	TG,TCBL	M,H	*sAAT**	*100	0.992	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.998
					*114	0.008	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.002
						100	*23*	*19*	*83*	*39*	*70*	*83*	*9*	*11*	*23*	*462*
Esterases	3.1.1.-	TCBL	L	*EST-D**	*100	1.000	1.000	1.000	1.000	1.000	1.000	0.994	1.000	1.000	1.000	0.999
					*114	0.000	0.000	0.000	0.000	0.000	0.000	0.006	0.000	0.000	0.000	0.001
						100	*23*	*21*	*83*	*39*	*70*	*83*	*9*	*11*	*23*	*444*
	3.1.1.-	TCBL	L	*EST-5**	*100	1.000	0.978	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.999
					*95	0.000	0.022	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.001
						17	*25*	*10*	*32*	*24*	*6*	*15*	*3*	*3*	*7*	*142*
Glyceraldehyde-3-phosphate dehydrogenase	1.2.1.12	TC+NAD	M	*GAPDH**	*-100	1.000	0.978	1.000	0.984	1.000	1.000	0.933	1.000	1.000	1.000	0.986
					*-80	0.000	0.022	0.000	0.016	0.000	0.000	0.067	0.000	0.000	0.000	0.014
						23	*21*	*10*	*25*	*14*	*25*	*20*	*9*	*8*	*20*	*176*
Alanine aminotransferase	2.6.1.2	TCBL	M	*ALAT**	*100	1.000	1.000	1.000	1.000	1.000	0.980	0.975	1.000	1.000	1.000	0.994
					*120	0.000	0.000	0.000	0.000	0.000	0.020	0.025	0.000	0.000	0.000	0.006
						23	*21*	*21*	*25*	*24*	*25*	*21*	*9*	*11*	*20*	*200*
Hexokinase	2.7.1.1	TBE	K	*HK-3**	*100	1.000	1.000	0.976	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.998
					*80	0.000	0.000	0.024	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.002
						4	*23*	*18*	*9*	*3*	*70*	*8*	*1*	*3*	*4*	*74*
Isocitrate dehydrogenase (NADP+)	1.1.1.42	TCP,TC	K	*IDHP-1**	*-100	1.000	1.000	0.972	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.993
					*-60	0.000	0.000	0.028	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.007
						100	*23*	*10*	*83*	*39*	*70*	*82*	*9*	*8*	*23*	*449*

Table 3.2 Genes and proteins studied in impala, and allele frequencies and sample sizes at non-monomorphic loci. Numbers in italics are sample sizes (n) for each locus at each locality. Number of genes sampled is $2n$. Buffers and tissues were those used for scoring and confirmation. TCBL = tris-citrate-borate-lithium (Ridgway et al. 1970; gel pH 8.7; electrode pH 8.0); TBE = tris-borate-EDTA (Markert and Faulhaber 1965; pH 8.6); TC = tris-citrate (Whitt 1970; pH 6.9); TCP = tris-citrate-phosphate (Grant 1988; pH 6.7); TG = tris-glycine (Grant 1988; pH 8.5); TM = tris-maleic acid (Georgiadis et al. 1990; pH 7.4). L = liver, H = heart, K = kidney, M = skeletal muscle. The following proteins failed to appear as bands on a gel or were not resolved: AH-1; AH-2; ADA-2; AO; CAT; FBP; G6PDH; GDA; MEP-3; LAP; PEP-C; PGDH-1; PGDH-2; TAT; XO. continued...

PROTEIN NAME	IUBCN NUMBER	BUFFER	TISSUE	LOCUS	ALLELE	Allele frequencies and sample sizes: Zimbabw	Nyleviei	Kleserie	Sabi Sands	Kruger Park	Mariath Park	Pongola	Temba	Albert Falls	Mkuze	TOTAL
Polymorphic proteins																
						79	*23*	*19*	*74*	*34*	*60*	*56*	*9*	*10*	*22*	*386*
Creatine kinase	2.7.3.2	TCBL,TG	H	*CK-C**	*100	0.873	0.848	0.974	0.973	0.985	0.967	0.982	1.000	1.000	0.955	0.947
					*105	0.127	0.152	0.026	0.027	0.015	0.033	0.018	0.000	0.000	0.045	0.053
						100	*23*	*19*	*83*	*39*	*68*	*81*	*9*	*11*	*23*	*458*
Glucose-6-phosphate isomerase	5.3.1.9	TCBL,TG	L,H	*GPI**	*-100	1.000	1.000	0.868	0.988	0.987	0.971	0.840	0.722	0.682	0.587	0.925
					*100	0.000	0.000	0.132	0.012	0.013	0.029	0.160	0.278	0.318	0.413	0.075
						100	*23*	*10*	*83*	*39*	*70*	*80*	*9*	*8*	*23*	*447*
Mannose-6-phosphate isomerase	5.3.1.8	TBE	M,K	*MPI**	*100	0.710	0.761	0.600	0.687	0.705	0.714	0.650	0.611	0.500	0.609	0.688
					*113	0.265	0.239	0.350	0.307	0.282	0.279	0.325	0.389	0.500	0.391	0.304
					*76	0.000	0.000	0.050	0.006	0.013	0.007	0.025	0.000	0.000	0.000	0.009
					*103	0.005	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.001
						100	*23*	*21*	*83*	*39*	*68*	*83*	*9*	*11*	*23*	*462*
Tripeptide aminopeptidase	3.4.11.4	TBE	K,H	*PEP-B**	*100	0.975	1.000	0.857	0.958	0.962	0.941	0.976	0.833	0.727	0.891	0.949
					*82	0.025	0.000	0.143	0.042	0.038	0.059	0.024	0.167	0.273	0.109	0.051
						100	*23*	*19*	*83*	*39*	*70*	*83*	*9*	*11*	*23*	*462*
Phosphoglucomutase	5.4.2.2	TBE,TCBL	L,K	*PGM-2**	*100	0.775	0.609	0.553	0.693	0.615	0.764	0.542	0.500	0.409	0.435	0.654
					*89	0.185	0.391	0.395	0.277	0.308	0.214	0.392	0.333	0.591	0.478	0.301
					*95	0.040	0.000	0.053	0.030	0.077	0.021	0.066	0.167	0.000	0.087	0.045
						98	*21*	*21*	*83*	*39*	*65*	*82*	*7*	*11*	*22*	*451*
General (unidentified) protein	-	TBE,TG,TCBL	K,M	*PROT-2**	*100	0.770	0.976	0.667	0.699	0.692	0.685	0.640	0.571	0.773	0.682	0.713
					*71	0.230	0.024	0.333	0.247	0.295	0.277	0.280	0.429	0.227	0.318	0.256
					*87	0.000	0.000	0.000	0.042	0.013	0.038	0.073	0.000	0.000	0.000	0.028
					*120	0.000	0.000	0.000	0.012	0.000	0.000	0.006	0.000	0.000	0.000	0.003
Proteins with rare variants																
						59	*23*	*19*	*26*	*24*	*25*	*21*	*9*	*11*	*20*	*239*
Aspartate aminotransferase	2.6.1.1	TG,TCBL	M,H	*sAAT**	*100	0.992	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.998
					*114	0.008	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.002
						100	*23*	*19*	*83*	*39*	*70*	*83*	*9*	*11*	*23*	*462*
Esterases	3.1.1.-	TCBL	L	*EST-D**	*100	1.000	1.000	1.000	1.000	1.000	1.000	0.994	1.000	1.000	1.000	0.999
					*114	0.000	0.000	0.000	0.000	0.000	0.000	0.006	0.000	0.000	0.000	0.001
						100	*23*	*21*	*83*	*39*	*70*	*83*	*9*	*11*	*23*	*444*
	3.1.1.-	TCBL	L	*EST-5**	*100	1.000	0.978	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.999
					*95	0.000	0.022	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.001
						17	*23*	*10*	*32*	*24*	*6*	*15*	*3*	*3*	*7*	*142*
Glyceraldehyde-3-phosphate dehydrogenase	1.2.1.12	TC+NAD	M	*GAPDH**	*-100	1.000	0.978	1.000	0.984	1.000	1.000	0.933	1.000	1.000	1.000	0.986
					*-80	0.000	0.022	0.000	0.016	0.000	0.000	0.067	0.000	0.000	0.000	0.014
						23	*21*	*10*	*25*	*14*	*25*	*20*	*9*	*8*	*20*	*175*
Alanine aminotransferase	2.6.1.2	TCBL	M	*ALAT**	*100	1.000	1.000	1.000	1.000	1.000	0.980	0.975	1.000	1.000	1.000	0.994
					*120	0.000	0.000	0.000	0.000	0.000	0.020	0.025	0.000	0.000	0.000	0.006
						23	*21*	*21*	*25*	*24*	*25*	*21*	*9*	*11*	*20*	*200*
Hexokinase	2.7.1.1	TBE	K	*HK-3**	*100	1.000	1.000	0.976	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.998
					*80	0.000	0.000	0.024	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.002
						4	*23*	*18*	*9*	*3*	*70*	*8*	*1*	*3*	*4*	*74*
Isocitrate dehydrogenase (NADP+)	1.1.1.42	TCP,TC	K	*IDHP-1**	*-100	1.000	1.000	0.972	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.993
					*-60	0.000	0.000	0.028	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.007
						100	*23*	*10*	*83*	*39*	*70*	*82*	*9*	*8*	*23*	*449*

Table 3.2 continued

PROTEIN NAME	IUBCN NUMBER	BUFFER	TISSUE	LOCUS	ALLELE	Allele frequencies and sample sizes: Zimbabw	Nylsvlei	Klaserie	Sabi Sands	Kruger Park	Mariotti Park	Pongola	Tembe	Albert Falls	Mkuze	TOTAL
Lactate dehydrogenase	1.1.1.27	TBE	M	*LDH-A**	*-100	1.000	1.000	0.900	0.982	0.987	1.000	1.000	1.000	1.000	1.000	0.993
					*-150	0.000	0.000	0.100	0.018	0.013	0.000	0.000	0.000	0.000	0.000	0.007
						100	*23*	*10*	*83*	*39*	*70*	*92*	*9*	*8*	*23*	*449*
	1.1.1.27	TBE	M	*LDH-B**	*100	1.000	1.000	1.000	1.000	1.000	0.993	1.000	1.000	1.000	1.000	0.999
					*80	0.000	0.000	0.000	0.000	0.000	0.007	0.000	0.000	0.000	0.000	0.001
						100	*23*	*21*	*83*	*39*	*68*	*83*	*9*	*11*	*23*	*462*
Dipeptidase	3.4.-.-	TBE	K,L	*PEP-A**	*100	1.000	0.978	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.999
					*77	0.000	0.022	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.001
						100	*23*	*10*	*83*	*39*	*70*	*93*	*9*	*11*	*23*	*459*
Phosphoglucomutase	5.4.2.2	TBE,TCBL	M	*PGM-1**	*100	1.000	0.978	1.000	0.988	1.000	1.000	0.994	1.000	1.000	1.000	0.996
					*65	0.000	0.022	0.000	0.006	0.000	0.000	0.006	0.000	0.000	0.000	0.003
					*120	0.000	0.000	0.000	0.006	0.000	0.000	0.006	0.000	0.000	0.000	0.001
						100	*23*	*21*	*83*	*39*	*67*	*83*	*9*	*11*	*23*	*481*
Nucleoside phosphorylase	2.4.2.1	TBE	K	*NP**	*100	0.980	0.957	0.929	0.934	0.974	0.933	1.000	1.000	1.000	1.000	0.966
					*136	0.020	0.043	0.071	0.066	0.026	0.067	0.000	0.000	0.000	0.000	0.034
Monomorphic proteins																
Aspartate aminotransferase	2.6.1.1	TG,TCBL	M,H	*mAAT**	*-100											
Adenosine deaminase	3.5.4.4	TCBL,TBE	M	*ADA-1**	*100											
Adenylate kinase	2.7.4.3	TC	K	*AK-1**	*100											
Alcohol dehydrogenase	1.1.1.1	TCBL	L	*ADH**	*100											
Creatine kinase	2.7.3.2	TCBL,TG	H	*CK-A**	*100											
Cytochrome b5 reductase	1.6.2.2	TCBL,TBE,TM	M	*CBR**	*100											
Fructose bisphosphate aldolase	4.1.2.1	TCBL,TG,TM	M,H	*FBALD-1**	*100											
Fumarate hydratase	4.2.1.2	TBE	K,L	*FH**	*100											
General (unidentified) protein	-	TBE,TG	K,M	*PROT-3**	*100											
Glutamate dehydrogenase	1.4.1.3	TBE	K,L	*GLUDH**	*100											
Glycerol-3-phosphate dehydrogenase	1.1.1.8	TC	K,M	*G3PDH-1**	*100											
	1.1.1.8	TC	K,M	*G3PDH-2**	*100											
L-iditol dehydrogenase	1.1.1.14	TCBL	L	*IDDH**	*100											
Isocitrate dehydrogenase(NADP	1.1.1.42	TCP,TC	K	*IDHP-1**	*100											
Malate dehydrogenase	1.1.1.37	TC+NAD	K,M	*MDH-1**	*-100											
	1.1.1.37	TC+NAD	K,M	*MDH-2**	*100											
Malic enzyme(NADP+)	1.1.1.40	TC,all	M	*MEP-1**	*-100											
Proline dipeptidase	3.4.13.9	TBE	K	*PEP-D**	*100											
Pyruvate kinase	2.7.1.40	TC	K,M	*PK-3**	*100											
Superoxide dismutase	1.15.1.1	TBE	K,L,M	*SOD-1**	*-100											
	1.15.1.1	TBE	K,L,M	*SOD-2**	*100											

3.2.1 Aspartate aminotransferase

AAT; 2.6.1.1; dimer; glutamic oxaloacetic transaminase

Banding patterns were typical of a nuclear and mitochondrial two-locus system with no interlocus heterodimer (Harris and Hopkinson 1976; Pemberton and Smith 1985; Roed 1985; Shaklee *et al.* 1989). A clear three-banded phenotype (*s*AAT 100/114) was observed in an individual from Zimbabwe. AAT activity declined in intensity after long storage. Two loci were reported in kidney and liver of British fallow deer *Dama dama* L. (Pemberton and Smith 1985), muscle of Norwegian semi-domestic reindeer *Rangifer tarandus* L. (Roed 1985), kidney of roe deer *Capreolus capreolus* L. (Hartl and Reimoser 1988), liver and muscle of white-tailed deer *Odocoileus virginianus* (Sheffield et al. 1985), springbok *Antidorcas marsupialis* and blesbok *Damaliscus dorcas phillipsi* (Bigalke et al. 1993). Two AAT loci were expressed in liver in 27 bovid species (including impala) and giraffe (Georgiadis et al. 1990).

3.2.2 Adenosine deaminase

ADA; 3.5.4.4; monomer

Two anodal bands were observed, representing presumptive loci *ADA-1** and *ADA-2**. However, no independent variability appeared in either zone to support the interpretation of two loci, so this interpretation remains tentative. Only ADA-1 was scorable; it was monomorphic. ADA was reported in roe deer liver (Hartl and Reimoser 1988). Two ADA loci were reported in muscle and kidney of four subspecies of red deer *Cervus elaphus* L. (Gyllensten et al. 1983). One ADA locus was reported in springbok (Bigalke et al. 1993) and blesbok kidney. Harris and Hopkinson (1976) found one locus in humans.

3.2.3 Alcohol dehydrogenase

ADH; 1.1.1.1; dimer

ADH activity was easily distinguishable from LDH (Lactate dehydrogenase) activity when the latter appeared (LDH can cross-react with ethanol as a

substrate). Only two cathodal bands were observed. One would expect an interallelic dimer if two alleles were present, but these do not always form due to incompatibility resulting from divergence between alleles (Burnett and Felder 1978). So the most likely explanation for the second band which appeared in some individuals was that it was a secondary band due to allozyme degradation. Burnett and Felder (1978) performed genetic crosses with *Peromyscus* and concluded that *ADH*" was a single locus, coding for a dimeric enzyme. I interpreted th econd band to be a secondary product, and the first band was monomorphic. Grobler and van der Bank (1994b) concluded that there were two ADH loci in impala, using liver and TBE and TCBL buffers (TBE for ADH-1 and TCBL for ADH-2). They found both to be monomorphic. Sheffield et al. (1985) reported three ADH loci in white-tailed deer. In muscle these loci were all polymorphic, while in liver, ADH-2 was polymorphic. However, genotype frequencies for ADH-2 deviated significantly from Hardy-Weinberg equilibrium, so scoring may have been inaccurate. In humans three loci code for the α, ß and τ subunits of ADH (Harris and Hopkinson 1976).

3.2.4 Adenylate kinase

AK; 2.7.4.3; monomer

Only one strong anodal band was detected in impala, with no variants. Thus *AK-1*" was scored as a monomorphic locus. Grobler and van der Bank (1994b) reported two loci using liver in TC in impala. Both were monomorphic. There are two AK loci in humans (Harris and Hopkinson 1976). One locus was reported in blue (*Connochaetes taurinus*) and black (*C. gnou*) wildebeest (Corbet et al. 1994). Products of two loci were reported by Sheffield et al. (1985) in white-tailed deer. Products of "ADK-2" were anodal (buffer conditions not described). Gyllensten et al. (1983) reported two loci in red deer. Pemberton and Smith (1985) reported two loci in fallow deer. Hartl and Reimoser (1988) reported two loci in liver of roe deer. Georgiadis et al. (1990) reported one locus in bovids (with 1 allele unique to impala). AK-1 and AK-2 were also reported by Bigalke et al. (1993) in springbok and blesbok.

3.2.5 Alanine aminotransferase

ALAT; 2.6.1.2; dimer; glutamic-pyruvate transaminase

ALAT stained rapidly in impala, and had to be scored before the bands merged. Narrow-banded phenotypes were interpreted to be homozygotes and broad-banded, more anodal phenotypes were interpreted to be heterozygotes. ALAT banding was reported in red deer muscle (Gyllensten *et al.* 1983) and in heart and liver in wildebeest (Corbet et al. 1994). Georgiadis *et al.* (1990) reported banding variability across 27 bovids and Giraffidae; 1 allele was unique to impala.

3.2.6 Creatine kinase

CK; 2.7.3.2; dimer

CK-1 was interpreted as the primary gene product of *CK-A**, orthologous to the human 'M' allozyme (Harris and Hopkinson 1976) and CK-2 as the primary gene product of *CK-C**, orthologous to the human 'B' allozyme (Harris and Hopkinson 1976). The interpretation of the two CK bands as CK-A and CK-C was also supported by Shaklee et al. (1989) and Grant (1988). CK-1 occasionally showed two bands, either due to one being a secondary product or to SOD (superoxide dismutase) decolorising the middle of the band. The intense CK-C heterodimer band in heterozygotes migrated slightly faster than the common allozyme, with fainter regions ahead of and behind it representing the homodimers. The homodimer of alternate homozygotes migrated faster then the common homodimer. Variability was first detected with TCBL, but resolution improved with TG buffer, which was used for confirmation of scoring. However, as samples aged, some individuals did not stain intensely enough to score CK-C. Also, activity was highly variable in foetal samples. Although variability did exist, it was impossible to score. Foetal allozymes at both loci showed similar electrophoresis patterns to adult enzymes, and similar levels of activity. In fresh samples, a faint band between CK-A and CK-C sometimes appeared and was assumed to be the interlocus heterodimer. One CK monomorphic locus was reported in impala in muscle (Grobler and van der Bank 1994b). One locus was also reported in red

deer muscle (Gyllensten et al. 1983) and in roe deer liver (Hartl and Reimoser 1988). CK-A and CK-C were detected in blue and black wildebeest by Corbet et al. (1994). CK-1 and -2 were reported in liver of white-tailed deer (Sheffield et al. 1985), springbok and blesbok (Bigalke et al. 1993). The transition to two loci occurred early in Chordate line and most species have CK-A and CK-C only (Buth et al. 1985; Fisher and Whitt 1978). Expressions of the CK loci vary among tissues between reptiles and amphibians, and the allozymes expressed by the two loci vary in their ability to form intra- and interlocus heterodimers. These are interpreted to be derived states (Buth et al. 1985; Ferris and Whitt 1978). CK-A,B,C and D have established orthology in fishes (Shaklee et al. 1989).

3.2.7 Cytochrome b_5 reductase

CBR; 1.6.2.2; monomer; NADH Diaphorase

Anodally-migrating bands appeared in spleen and liver but these were presumably secondary products, not CBR-3, because human CBR-3 is slower migrating at a high pH and occurs mainly in reproductive tissue (Harris and Hopkinson 1976). CBR-3 is one of the two NADH (nicotinamide adenine dinucleotide, reduced) dependent allozymes and is expressed by another CBR locus in humans, predominantly in reproductive tissue (Harris and Hopkinson 1976). For impala, muscle was used for scoring and usually produced one band, but sometimes with a slow-migrating, diffuse band. Occasionally an even slower-migrating band appeared. This presumably represented CBR-3 and not a variant allele. No alternative homozygotes were found, and the CBR-1 band did not diminish in intensity as would be expected in a heterozygote. Banding variability could also have been caused by prolonged storage (Cepica and Stratil 1978). Emerson and Tate (1993) reported two CBR loci in 10 deer taxa. Two CBR loci were reported in liver of 28 bovid and giraffid species examined by Georgiadis et al. (1990). Two loci were reported in roe deer by Hartl and Reimoser (1988).

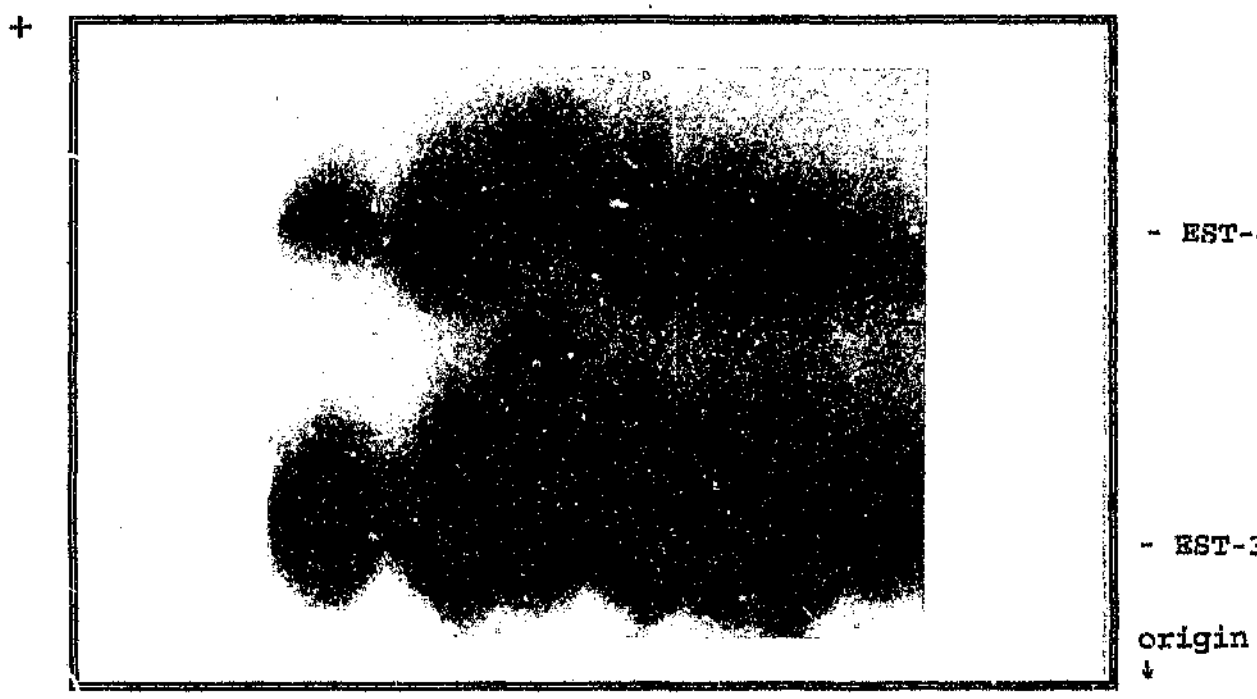


Figure 3.1 Gel showing typical EST-3 and -4 banding as well as faster-migrating bands of both EST-3 and -4 from impala #452 from Pongola, which had been dead several hours before tissue collection. Buffer = TCL; substrate = α-Naphthyl acetate; liver.

3.2.8 Esterases

EST; 3.1.1.-; 'ESA_1' monomeric, 'ESD' dimeric in humans (Harris and Hopkinson 1976)

Figure 3.1 shows typical banding patterns for EST-3 and -4, including the variant bands due to enzyme degradation in a long-dead individual from Pongola. Esterases run in TCBL buffer were detected with α-Naphthyl acetate and 4-methylumbelliferyl acetate. Two bright bands and three less intense bands appeared with 4-methylumbelliferyl acetate as a substrate. None of the three less intense bands attenuated the main band, as would be expected for a three-banded heterozygote, so they were presumed not to be variant alleles. Their presence or absence and intensity were highly variable among individuals. These bands may represent the products of human *ESA_4** and *ESD**, or may have been due to secondary products (Peters and Nash 1978; Peters 1982; Harris and Hopkinson 1976). I interpreted the main band to be EST-D. The bands observed in impala migrated anodally, and therefore one most likely was EST-D, as ESA_4 migrates cathodally in humans (at pH 7.2) (Harris and Hopkinson 1976). α-Naphthyl acetate could have detected in liver the products of the impala orthologues of human *ESA_4**, which is widespread in tissues, and possibly of *ESA_1**, *ESA_5**, *ESA_7** *and ESC** which predominate in other tissues or developmental stages (Harris and Hopkinson 1976; Peters and Nash 1978; Peters 1982). Liver shows some activity of nearly all locus products on both substrates in humans (Harris and Hopkinson 1976), so it was used for scoring in impala. α-Naphthyl acetate routinely produced five main anodal bands, EST-1 to -5. All showed some activity in most tissues, but good activity in liver. EST-4 also showed strong activity in kidney. EST-D and EST-3 migrated at the same rate and showed the same banding pattern, including variants, so I interpreted EST-3 to be EST-D. Impala alleles in this study were *EST-D*100* and **114*, and *EST-5*100* and **95*. Grobler and van der Bank (1994b) reported alleles with similar electrophoretic mobilities for these loci, which they named *EST-1** and *EST-2**. One variant phenotype appeared at EST-3 in an individual from Pongola (#452) and the same variant appeared at EST-4, indicating that EST-4 was a secondary product of EST-3. This individual

was dead several hours before tissue collection, so I assumed that this a resulted from enzyme degradation (Figure 3.1). A slower variant appeared in EST-5 in an individual from Nylsvlei. I interpreted this as a heterozygote. I concluded that only two bands, EST-3 and EST-5 represented presumptive loci and that EST-3 was the same as EST-D.

Roed (1985) designated eight EST loci in reindeer, 5 of which had insufficient activity to score and one of which was polymorphic. Four presumptive loci were reported in liver or plasma of white-tailed deer (Sheffield et al. 1985). EST-1 and EST-2 (numbered from cathode to anode) were reported in fallow deer (Pemberton and Smith 1985). Only EST-D was reported in roe deer by Hartl and Reimoser (1988). Two esterase loci were reported by Georgiadis et al. (1990) in 28 bovid and giraffid species. '*ES-1***A*' was unique to impala and '*ES-2***C*' was unique to impala and one Reduncini species. One locus, *EST-D**, was reported in blue and black wildebeest (Corbet et al. 1994). *EST-D** was reported in springbok and blesbok (Bigalke et al. 1993). Their EST-1 and -2 appear to represent my EST-3 and -5.

3.2.9 Fructose-bisphosphate aldolase

FBALD; 4.1.2.13; tetramer; aldolase

Three main bands of FBALD appeared, two anodal and one cathodal. Since bands of GAPDH may also appear with the stain used, and since the pattern of cathodal banding was similar to that of GAPDH, I assumed the cathodal zone of banding was GAPDH. Thus I designated the two anodal bands FBALD-1 and -2. However, TG and TM caused FBALD-1 to migrate cathodally. FBALD-1 appeared in extracts from muscle, kidney and heart. FBALD-2 appeared in kidney and heart, but was usually fainter and band intensities more variable than FBALD-1. FBALD-2 bands were unscorable, and although FBALD-1 bands were not tight (possibly due to interlocus tetramers), there was no lightening, moving and spreading of the bands that would be typical of interallelic heterotetramers, so I scored them monomorphic. Bigalke et al. (1993) found one FBALD locus in springbok and blesbok. Georgiadis et al. (1990) found one locus in 28 bovid and

giraffid species. Two putative loci were reported in liver and muscle of white-tailed deer (both monomorphic) (Sheffield et al. 1985) and muscle of reindeer (insufficient activity) (Hartl and Reimoser 1985). Three loci were observed in humans (Harris and Hopkinson 1976). Their products all migrate anodally.

3.2.10 Fumarate hydratase

FH; 4.2.1.2; tetramer; fumarase

No variability was evident, so the impala were considered to be monomorphic. FH was reported in springbok and blesbok (Bigalke et al. 1993). One FH locus was reported in roe deer (Hartl and Reimoser 1988). Two presumptive loci were found in white-tailed deer (Sheffield et al. 1985). There are two sets of allozymes in humans, one mitochondrial and one cytosolic, both probably encoded by the same autosomal locus (Harris and Hopkinson 1976).

3.2.11 L-Glutamate dehydrogenase

GLUDH; 1.4.1.3

NAD (nicotinamide adenine dinucleotide) in the stain caused LDH bands to appear, so NADP (NAD phosphate) was used as a cofactor, after which only one clear GLUDH band appeared, sometimes with a slow-migrating, diffuse band. GLUDH was monomorphic in impala. GLUDH was reported in springbok and blesbok (Bigalke et al. 1993). Products of two presumptive loci ('GDH-1 and -2') were reported in 28 bovid and giraffid species (Georgiadis et al. 1990). One locus was reported in roe deer (Hartl and Reimoser 1988) and red deer (Gyllensten et al. 1983). 'GDH-1' and 'GDH-2' were reported in white-tailed deer (Sheffield et al. 1985). In these studies the scoring of GLUDH as monomorphic was probably conservative, as variants known to exist in other species have only been detected by DEAE column chromatography (Pryor 1974).

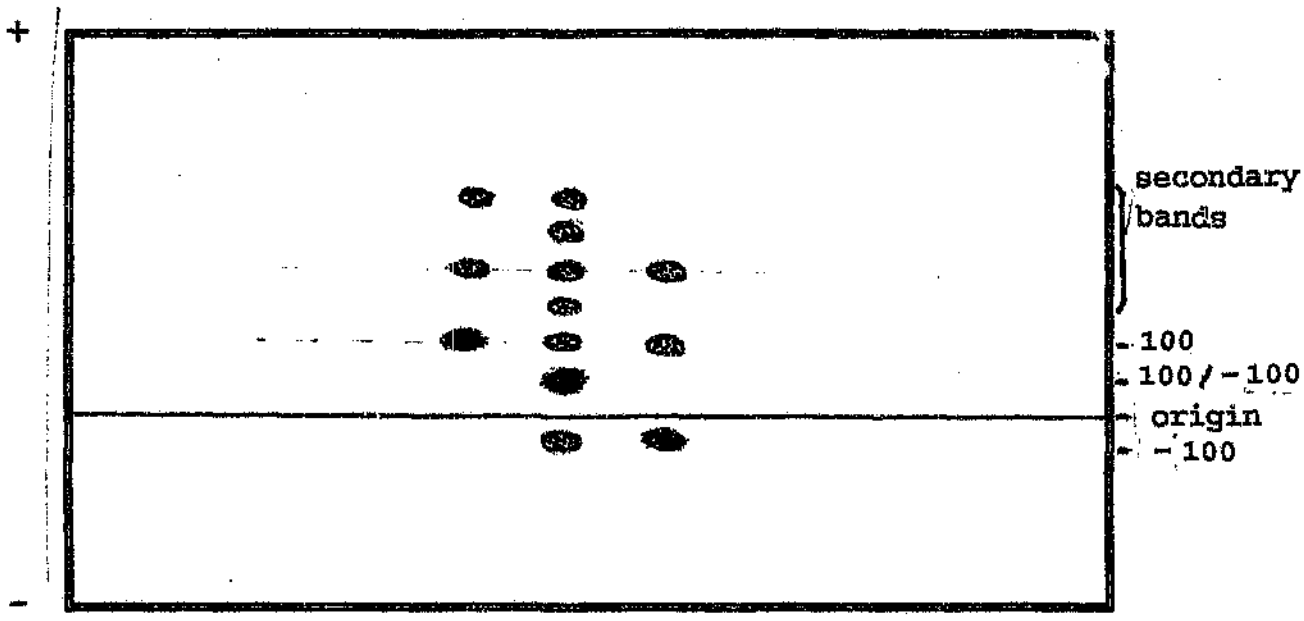


Figure 3.2 GPI banding patterns. Showing GPI 100/100, -100/100 and -100/-100 from left to right. Most anodal bands did not always appear. With other buffers, patterns were constant but shifted anodally or cathodally. Buffer = TCL; liver.

3.2.12 Glucose-phosphate isomerase

GPI; 5.3.1.9; dimer; phosphoglucose isomerase

Figure 3.2 shows the banding patterns I observed in impala. Homozygotes show a main band with two equidistant secondary bands running more anodally. The interpretation of only one main locus product was confirmed when I detected both heterozygotes and alternate homozygotes, each with patterns consistent with a one locus explanation. The common three banded phenotype I observed in impala was the same in all tissues examined. In TCBL and TG, the common main band was cathodal, and its two secondary bands anodal. The variant homodimer and its secondary allozymes were anodal. The main heterodimer band appeared just cathodally of the origin and also had two secondary bands, accounting for the seven bands observed in the presumptive heterozygotes (the second and third band of the common allelic product migrated at the same rate as the first and second bands of the alternate allele respectively).

My interpretation of a one locus dimeric system with one variant allele is the simplest explanation for these patterns but contrasts with interpretations of this enzyme in some other studies. Banding patterns in this study were similar to those for red deer (Cervidae) in Gyllensten et al. (1983; Fig. 2). Since Bovidae and Cervidae are closely related families I suggest that the two or three locus system they and others propose is incorrect. While Gyllensten et al. (1983) interpreted the second band to represent the product of a second locus, and the third band to be a secondary product, I interpreted both the second and third bands to be secondary products. A two locus system was proposed for moose *Alces alces* (Ryman et al. 1980), impala (Grobler and van der Bank 1994b; monomorphic), sable *Hippotragus niger* (Grobler and van der Bank 1994a; Grobler and van der Bank 1993), red deer *C. elaphus* (Gyllensten et al. 1983), springbok and blesbok (Bigalke et al. 1993), white-tailed deer (Sheffield et al. 1985) and roe deer (Hartl and Reimoser 1988). Three loci were proposed in fallow deer (Pemberton and Smith 1985) and reindeer (Roed 1985). Roed's (1985) estimate of the number of loci is "conservative be-

cause [it] reflects the minimum number of loci involved (cf. Allendorf et al. 1977)". I disagree with his interpretation on the basis of my results and those of others cited here for bovids and cervids.

A one locus GPI system was proposed in blue and black wildebeest (Corbet et al. 1994); in 10 deer taxa (Emerson and Tate 1993); and in 28 bovid and giraffid species (Georgiadis et al. 1990). Humans also express one GPI locus with a banding phenotype consisting of one primary band with two anodal secondary bands (Harris and Hopkinson 1976). Mice *Mus musculus* express one locus, confirmed by comparing F_1 and backcross progeny of inbred strains to determine genetic control and autosomal variation. There are multiple alleles at this locus; they agree with GPI being a dimer (DeLorenzo and Ruddle 1969).

3.2.13 Glyceraldehyde-phosphate dehydrogenase

GAPDH; 1.2.1.12; tetramer

GAPDH showed parallel expression in all tissues. The bands appeared cathodally near the origin. However, when resolution was good, there was a rare variant that was consistently slower than the common band and which appeared in impala from more than one locality (including Nylsvlei, Mala Mala and Pongola). The region of the variant band in the same position as the common band was lightened, and the whole band was elongated, with a dark centre, consistent with the interpretation of a tetrameric heterozygote. Impala were thus scored monomorphic, with one rare variant allele. Grobler and van der Bank (1994b) reported one locus in impala, with no variants. Sheffield et al. (1985) reported two presumptive loci in white-tailed deer. Gyllensten et al. (1983) reported one locus in red deer. Corbet et al. (1994) reported two presumptive loci in blue and black wildebeest.

3.2.14 Glycerol-3-phosphate dehydrogenase

G3PDH; 1.1.1.8; dimer; α-glycerophosphate dehydrogenase

The TC+NAD and TC buffers were adequate in resolving G3PDH, but TC gave the best results. The TM and TG buffers yielded intense bands but gave

inadequate resolution. Activity was present only in muscle, kidney, heart and liver, being highest and most consistent among individuals in kidney under all buffer conditions. All four tissues showed a clear 3-banded anodal pattern. Muscle and kidney occasionally showed faint, irregular, more anodal bands. Kidney was used for scoring. The presence of three equidistant bands supports the hypothesis that they represent the homodimers and heterodimer of the products of two loci. No variants were detected to confirm this interpretation.

One monomorphic locus was previously reported for impala by Grobler and van der Bank (1994b). Two autosomal loci are expressed in humans (Harris and Hopkinson 1976). The products of two loci were reported by Georgiadis et al. (1990) in 28 bovid and giraffid species ('αGPD-1 and -2'; equivalent to G3PDH-1 and -2). One band was resolved by Pemberton and Smith (1985) in fallow deer and by Gyllensten et al. (1983) in red deer. The products of two presumptive loci were reported in white-tailed deer (Sheffield et al. 1985). One zone of banding was reported in springbok and blesbok (Bigalke et al. 1993).

3.2.15 Hexokinase

HK; 2.7.1.1; monomer

Three anodal zones of banding appeared in fresh tissue extracts. Variant bands were tissue specific, indicating they were not Mendelian variants coded for by alleles at one locus. So the three bands appear to represent three loci. HK-3 was present in brain, muscle and kidney; HK-2 appeared in heart, tongue, liver, kidney, lung and spleen. HK-1 appeared in kidney and liver, but was too faint to be sure about activity in other tissues. HK-3 stained most consistently, and therefore I scored only that one locus with confidence. A rare HK-3 variant band sometimes appeared between the normal positions of HK-2 and -3, i.e. they appeared to move to the centre, and this was interpreted to represent a variant at one locus.

The products of one monomorphic hexokinase locus were previously resolved in impala (Grobler and van der Bank 1994b). Three presumptive

loci were reported in springbok and blesbok (Bigalke et al. 1993). The products of one locus were resolved in white-tailed deer (Sheffield et al. 1985). Two loci were reported in red deer (Gyllensten et al. 1985) and roe deer (Hartl and Reimoser 1988). Even though more than one HK locus was apparently expressed in 28 bovid and giraffid species, only one was scored (Georgiadis et al. 1990). Humans show banding patterns possibly representing four loci with different tissue specificities (Harris and Hopkinson 1976). Some human HK allozymes appear to be monomeric (Harris and Hopkinson 1976).

3.2.16 L-Iditol dehydrogenase

IDDH; 1.1.1.14; tetramer (Op't Hof et al. 1969); sorbitol dehydrogenase

Bands were cathodal with diffuse, slow-migrating bands in TCBL buffer. There was some non-Mendelian variability in mobility amongst individuals, but the main band did not lighten as would be expected in a five-banded heterozygote banding pattern. Impala were monomorphic. The products of four alleles encoded by a single locus were reported in 28 bovid and giraffid species, one allele being unique to impala (Georgiadis et al. 1990). One locus was reported in roe deer (Hartl and Reimoser 1988), fallow deer (Pemberton and Smith 1985), red deer (Gyllensten et al. 1983) and impala (Grobler and van der Bank 1994b). A single zone of barding was reported in white-tailed deer (Sheffield et al. 1985). One locus was reported in a sample of springbok and blesbok (Bigalke et al. 1993)

3.2.17 Isocitrate dehydrogenase NADP+

IDHP; 1.1.1.42; dimer

With NADP rather than NAD as a cofactor, the stain detects the cytosolic form and one mitochondrial form in humans (Harris and Hopkinson 1976). We used NAD as a cofactor, but obtained no banding since NADP is a cofactor for nuclear-encoded enzymes only. The two forms I detected with NADP are both nuclear encoded, even though one may be associated with mitochondria (Henderson 1965; Harris and Hopkinson 1976). So I called them IDHP-1 and -2 instead of mIDHP and sIDHP, respectively. No interlocus

heterodimer band appeared on the gels, indicating that the two bands in impala represent proteins from two loci (Henderson 1968). Activity of IDHP-1 appears only in muscle, kidney and heart, while IDHP-2 activity appears in muscle, kidney, heart, liver, gut, spleen and lung. IDHP-2 had more intense banding than did IDHP-1 in all tissues, and resisted storage degradation for longer periods. IDHP-2 also did not show as much variability in activity levels as did IDHP-1. TBE yielded good activity, but provided poor resolution. I used TCP later, and it provided excellent resolution, but by that time activity had decreased, presumably because of storage (Harris and Hopkinson 1976). IDHP-2 was monomorphic in all buffers tested, but IDHP-1 showed variability in TC and TCP.

3.2.18 Lactate dehydrogenase

LDH; 1.1.1.27; tetramer

Patterns of gene expression in impala are similar to those in humans for a two-locus system, LDH-A (cathodal) and LDH-B (anodal), including expected interlocus heterotetramers. LDH-A is active only in brain, eye and muscle, but LDH-B showed activity in all tissues tested, with best resolution in muscle. More precisely, LDH-B_4, and -B_3A_1 allozymes stained in all tissues, LDH-B_2A_2 in brain, eye and muscle, and LDH-A_4 mainly in muscle. LDH-B_1A_3 was usually missing, but occasionally appeared faintly in muscle. LDH-A_4 and -B_4 were sometimes thick or even closely double-banded, but this variation was deemed non-genetic, as it did not produce the expected interlocus heterodimeric bands.

Two monomorphic loci (*LDH-1** and *-2**) were reported in impala (Grobler and van der Bank 1994b), springbok and blesbok (the same allele in these latter two species) (Bigalke et al. 1993), white-tailed deer (Sheffield et al. 1985), red deer (Gyllensten et al. 1983), fallow deer (Pemberton and Smith 1985) and blue and black wildebeest (Corbet et al. 1994). One monomorphic and one polymorphic locus (*'LDH-2'*; 2 alleles) were reported in reindeer (Roed 1985a) and roe deer (*'LDH-2'*; 2 alleles) (Hartl and Reimoser 1988). Two loci were reported in 28 bovid and giraffid species (Georgiadis et al. 1990). 'LDH-1' had 5 alleles and 'LDH-2' 7 alleles. At

both loci, one allele was unique to impala. One locus was reported in 10 cervid taxa (Emerson and Tate 1993). One species was polymorphic (2 alleles).

3.2.19 Malate dehydrogenase

MDH; 1.1.1.37; dimer

In humans, two autosomal loci code for mMDH and sMDH, but this was not reported in any species related to impala, so I designated the two bands I resolved MDH-1 (cathodal) and MDH-2 (anodal), respectively. MDH-1 and -2 were active in all tissues tested. The addition of NAD to the TC buffer improved the activity of MDH-1 in some tissues and MDH-2 most tissues. MDH-1 activity was greatest in muscle and kidney and MDH-2 in muscle, kidney, heart and liver. Kidney and muscle were used for scoring. MDH-1 appeared more prone to storage degradation, as its activity decreased with time in samples from some localities. Both bands were monomorphic.

Two loci were reported in red deer, *MDH-1** and *MDH-2** (Gyllensten et al. 1983). Two loci were reported in blue and black wildebeest (Cortet *et al.* 1994), roe deer (Hartl and Reimoser 1988), fallow deer (Pemberton and Smith 1985), white-tailed deer (Sheffield et al. 19[illegible] springbok and blesbok (Bigalke et al. 1993) and impala (Grobler and van der Bank 1994b). Three loci were reported in reindeer (Roed 1985). One locus was reported in ten deer taxa (Emerson and Tate 1993). Two loci were reported in 28 bovid and giraffid species (Georgiadis et al. 1990).

3.2.20 Malic enzyme

MEP; 1.1.1.40; tetramer

Three bands were resolved in impala. One was cathodal (MEP-1) and the other two anodal (MEP-2 and -3). Since there are two loci in humans, cervids and other bovids, I considered MEP-1 and -3 to represent the products of two loci. MEP-1 was probably equivalent to human mME and MEP-3 to sME. MEP-3 ran just ahead of MEP-2. MEP-2 lost activity

quickly with storage and could have been an interlocus heterotetramer or the product of another locus. It was only present in muscle, heart, liver, brain, lung, gut and spleen. Muscle was used for scoring. MEP-1 showed no variability and MEP-3 appeared highly variable, yet several buffers did not resolve bands sufficiently to score them.

3.2.21 Mannose-phosphate isomerase

MPI; 5.3.1.8; monomer; phosphomannose isomerase

Figure 3.3 shows typical MPI banding patterns and the interpretation of most phenotypes. MPI showed parallel tissue expression of four clearly defined allozymes at one locus. One allele, **103*, occurred in only one individual from Zimbabwe. Grobler and van der Bank (1994b) reported only two alleles, **100* and **130*, for impala. However, we detected four alleles, **100*, **113*, **75*, and **103*, in samples from some of the same localities that they sampled. Genotypic frequencies in some samples of the present study were out of Hardy-Weinberg proportions but this may have been due to small frequencies of some genotypes; when alleles were pooled, frequencies did not depart from Hardy-Weinberg proportions. I confirmed scoring for all individuals under various conditions.

One locus has also been found in humans (Harris and Hopkinson 1976), red deer (Gyllensten et al. 1983), roe deer (Hartl and Reimoser 1988), blue and black wildebeest (Corbet et al. 1994), ten deer taxa (Emerson and Tate 1993), 28 bovid and giraffid species (Georgiadis et al. 1990) and impala (Grobler and van der Bank 1994b). The products of two presumptive loci were reported in white-tailed deer (Sheffield et al. 1985), and springbok and blesbok (Bigalke et al. 1993).

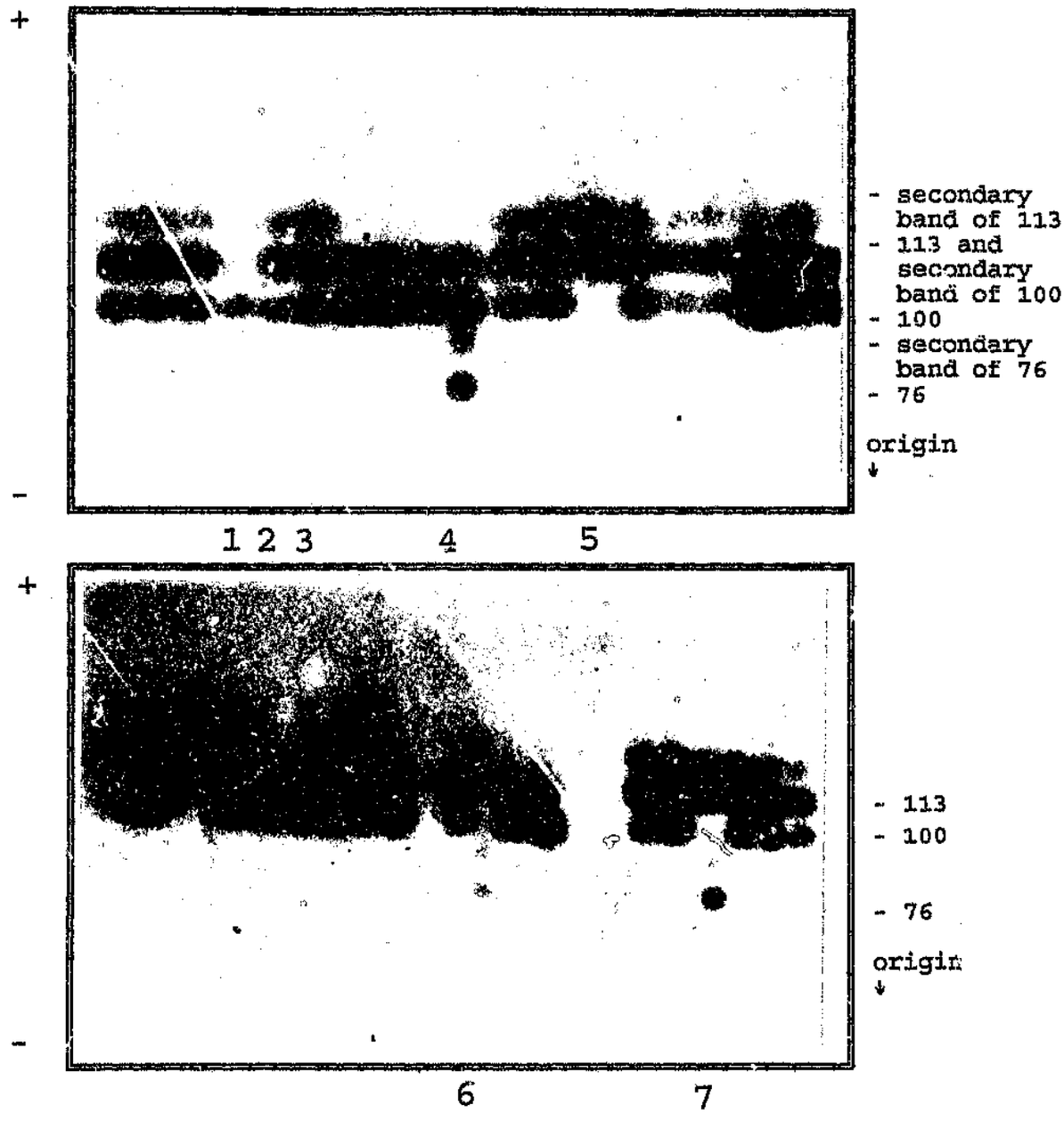


Figure 3.3 Typical banding patterns of MPI. Phenotypes in lanes labelled 1–7 are given as examples of interpretation: 1 and 3 = 100/100; 2 = 100/113; 4 = 100/76; 5 = 113/113; 6 and 7 = 76/113.

3.2.22 Nucleoside phosphorylase

NP; 2.4.2.1; trimer (Ward et al. 1979); purine nucleoside phosphorylase

Mendelian variability within the main zone of banding was observed. Phenotypes with light homotrimer bands and two heavier heterotrimer bands were interpreted to represent heterozygotes. No alternate homotrimer patterns representing homozygotes were observed to confirm this but the variant pattern was consistent among individuals and repeatable. One locus has been found in humans (Harris and Hopkinson 1976), impala (Grobler and van der Bank 1994b) and white-tailed deer (Sheffield et al. 1985).

3.2.23 Cytosolic aminopeptidases

PEP; 3.4.-.-

Many peptidase loci are orthologous among mammals and fish, and substrate specificity is a reliable indicator of identity (Frick 1983; Frick 1984; Laurie-Ahlberg 1982). Pep A,B,C,D,S are also products of independent loci and have distinct adult tissue and substrate specificities (Wyban 1982).

I compared banding patterns produced with four substrates, L-G-G (leucyl-glycyl-glycine), L-T (leucyl-tyrosine), P-P (prolyl-phenylalanine) and G-L (glycyl-leucine) by applying each substrate separately to one of four of the same individual sequences on one gel. I also compared the effects of different buffers on peptidase banding patterns. With L-T, two anodal zones of banding appeared and were interpreted as PEP-A and PEP-C (PEP-C most anodal) (Harris and Hopkinson 1976). PEP-A showed one rare Mendelian variant. PEP-C was highly polymorphic, yet I could not resolve its variability enough to score this locus with confidence. L-G-G produced one band which I interpreted as PEP-B. This showed Mendelian variability typical of a monomer. P-P produced one band, PEP-D. G-L was a substrate for the same allozymes as L-T, but showed lower activity. While relative band mobility varied greatly with different buffers, banding patterns and interpretation remained consistent across buff-

ers and tissues. Peptidases showed parallel tissue expression of all variability in impala, and results were repeatable. Figure 3.4 shows banding patterns and interpretation of PEP-B in three different buffers, demonstrating the effects of the buffer on resolution.

One monomorphic PEP locus was reported in red deer (Gyllensten et al. 1983). Hartl and Reimoser (1988) reported two loci in roe deer using L-A. PEP-A and -C are specific for L-A, so these authors' PEP-1 may be PEP-A and their PEP-2 may be PEP-C, if relative motility is similar to that in impala. Georgiadis et al. (1992) reported four PEP loci in 28 bovid and giraffid species, '*PEP-A**', '*PEP-B**', '*PEP-C**' and '*PEP-D**'. No substrates were listed. Bigalke et al. (1993) reported two monomorphic loci in springbok and blesbok. Corbet et al. (1994) used G-L and L-G-L as substrates. G-L stained PEP-A, and L-G-L stained PEP-B and PEP-X. All three enzymes were monomorphic. Grobler and van der Bank (1994b) reported one allozyme that stained with L-G-G (PEP-1 [B?]), two that stained with L-T (PEP-2 and -3 [A and C?]), and one that stained with P-P (PEP-4[D?]). Their PEP-3[C?] had two variants, 95 and 100.

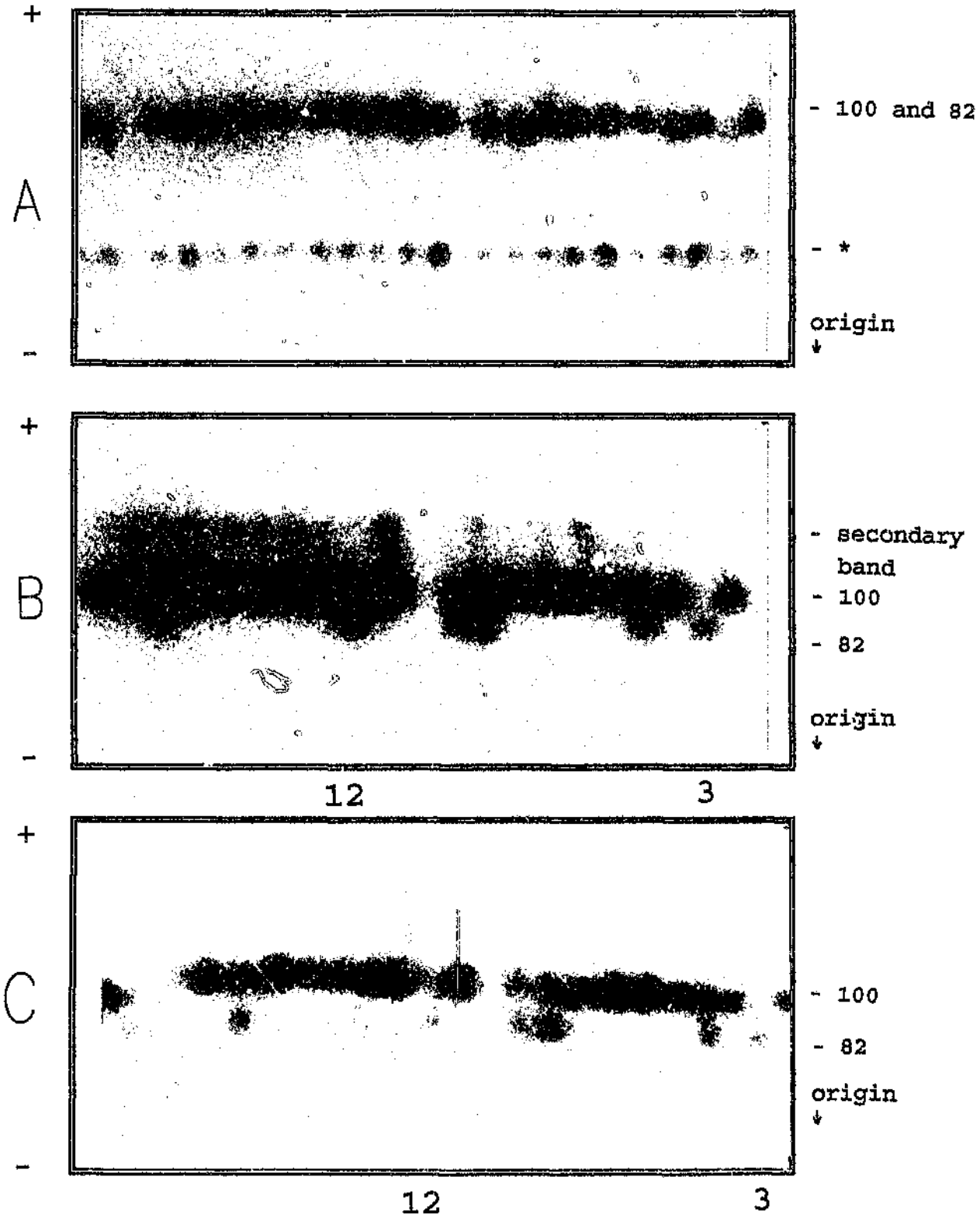


Figure 3.4 PEP-B. Gels showing the same series of individuals stained with L-G-G. Buffer in *a* is TCL, *b* is TM and *c* is TC. Note how buffer conditions affect resolution. Variant phenotypes are well resolved in TM and TC, and secondary banding is minimal in TC. Phenotypes in lanes labelled 1-2 are given as examples of interpretation: 1 = 100/100; 2 = 100/82; 3 = 82/82. '*' is probably a secondary band.

3.2.24 General or unidentified protein
PROT

Figure 3.5 shows typical banding of PROT, including variants of PROT-2. Three main zones consistently appeared, with similar migration to haemoglobin (which was too faint to score), but with variants that were not typical for multimeric haemoglobin variants. So I assumed they were not haemoglobin, but other soluble proteins. PROT-1 was unscorable, yet appeared to show no variability. PROT-3 was clear and more intense than PROT-1, and was monomorphic. PROT-2 stained clearly and was polymorphic (4 alleles). Interpretation of variants of PROT-2 was supported by three observations. Firstly, the common band in presumptive heterozygotes was attenuated. Secondly, in the presumptive heterozygotes two bands were present, which is typical of monomeric heterozygotes. Alternate homozygotes appeared as a single dark band at the same position as the alternate band in the presumptive heterozygotes for all variants except PROT 120, which was too rare to be present as a homozygote. Grobler and van der Bank (1994b) reported three presumptive loci in impala blood, with three alleles of PROT-2 (GP-2): **100*, **71* and **86*, which appear to be the same as my alleles: **100*, **71* and **87*. Roed (1985) reported five presumptive loci in muscle and liver of reindeer. Georgiadis et al. (1990) found one locus in 28 bovid and giraffid species. Blue and black wildebeest showed three loci ('PT-1, -2 and -3') (Corbet et al. 1994).

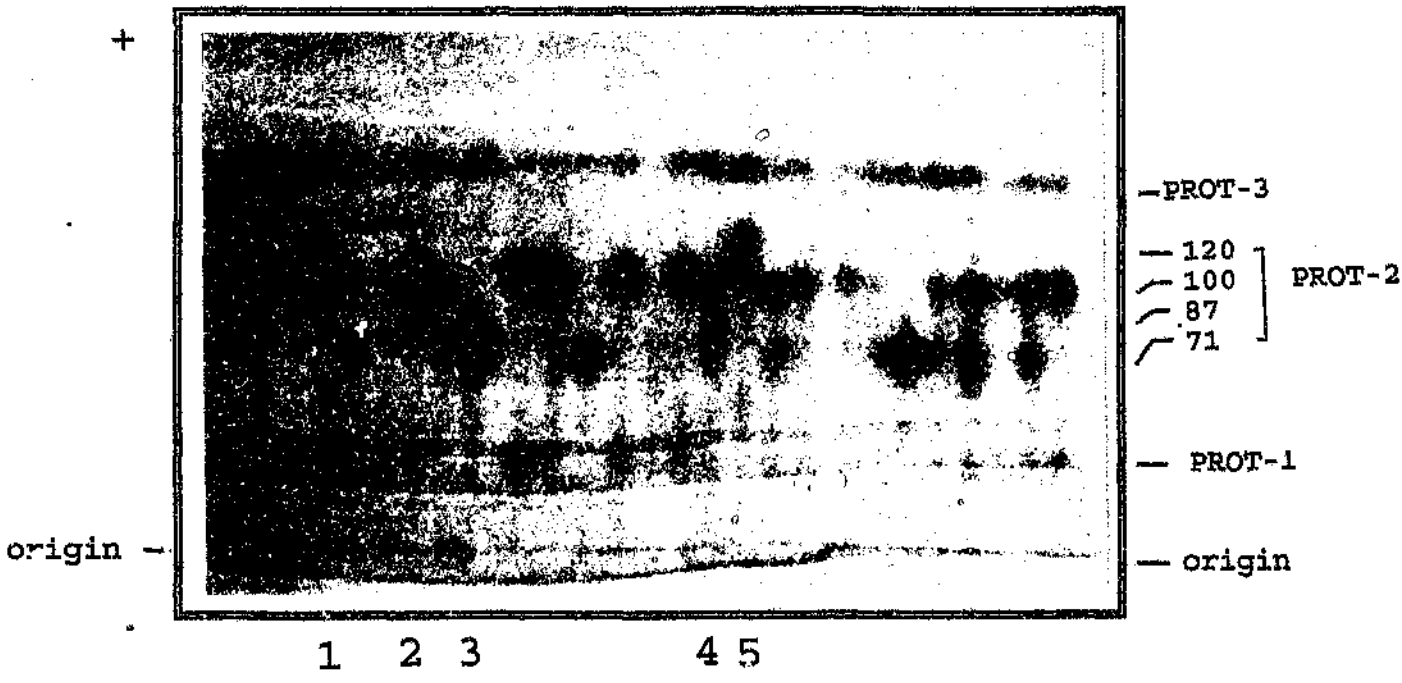


Figure 3.5 Gel showing typical PROT banding and variability. Horizontal lines are artefacts of gel slicing. Phenotypes in lanes labelled 1-5 are given as examples of interpretation: 1 = 100/100; 2 = 100/87; 3 = 71/71; 4 = 100/71; 5 = 100/120. Buffer = TBE; kidney.

3.2.25 Phosphoglucomutase

PGM; 2.7.5.1; monomer

I resolved the products of two presumptive loci in impala. Both zones of banding were anodal. PGM-1 and -2 activity showed markedly different tissue distributions. PGM-1 showed activity in all tissues tested except eye, gut and mammary gland. It was most active in muscle and liver. PGM-2 showed activity in [illegible] tissues tested and was generally more active than PGM-1. PGM-2 activity was strongest in kidney and liver, which were used for routine scoring. Variability was parallel across tissues. Three PGM-2 phenotypes were detected, present as double-banded heterozygotes and single-banded alternate homozygotes. PGM-1 showed fewer variant bands than did PGM-2. Bands were resolved best on TBE. Bands representing variants were distinguished from secondary bands by the attenuation of the main band, by the appearance of a new band in a different position to the secondary bands, and by the appearance of their own secondary bands.

Humans express three autosomal loci in all tissues (Harris and Hopkinson 1976). Two loci were reported in red deer (Gyllensten et al. 1983), blue and black wildebeest (Corbet et al. 1994), 28 bovid and giraffid species (Georgiadis et al. 1990), roe deer (Hartl and Reimoser 1988), white-tailed deer (Sheffield et al. 1985) and fallow deer (Pemberton and Smith 1985). One locus was reported in ten deer taxa (Emerson and Tate 1993). Three loci were reported in reindeer (Roed 1985), springbok and blesbok (Bigalke et al. 1993). Grobler and van der Bank (1994b) reported two loci in impala. PGM-1 was scored in muscle and was monomorphic, and PGM-2 was scored in kidney and was polymorphic with two alleles, *PGM-2*100* and *PGM-2*84*. These alleles appear to correspond to *PGM-2*100* and *PGM-2*89* in the present study.

3.2.26 Pyruvate kinase

PK; 2.7.1.40; complex tetramer (also Rigaut and Chalumeau 1984)

Four complex zones of banding were observed on gels for impala. PK-1 (cathodal) was observed in some individuals in kidney extracts in TC

buffer. PK-2 was just cathodal and was evident only in muscle. PK-3 (anodal) appeared in brain, muscle, kidney, heart, liver, spleen, pancreas and mammary gland. PK-4 (most anodal) appeared in brain, muscle, kidney, lung, pancreas and mammary gland. Muscle and kidney were used for routine scoring in TC and TC+NAD gels, as these tissues showed activity for most bands. The intensities of some bands between individuals was not consistent. Only PK-3 showed consistent activity no banding variability consistent with tetrameric subunit structure was apparent.

There are probably three loci in humans, producing enzymes with different tissue distributions (Harris and Hopkinson 1976). One locus was reported in springbok and blesbok (Bigalke et al. 1992), white-tailed deer (Sheffield et al. 1985) and red deer (Gyllensten et al. 1983).

3.2.27 Superoxide dismutase

SOD; 1.15.1.1; human sSOD dimeric, mSOD tetrameric

Two zones of light-coloured bands appeared on TBE gels soaked in only PMS (phenazine methosulphate) and MTT (3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide) and were interpreted to reflect the activity of SOD. The same bands also appeared on gels stained for ADA and GDA and these gels were used for routine scoring of SOD. SOD appeared in liver, kidney and muscle with TBE, but I did not test other tissues. Faint but definite cathodal and anodal bands were resolved, which appeared to be monomorphic. The cathodal band may be orthologous to human mSOD (SOD_m) and the anodal band to human sSOD (SOD_s) (Harris and Hopkinson 1976).

Two monomorphic loci were previously reported in impala (Grobler and van der Bank 1994b), in springbok and blesbok (Bigalke et al. 1993), fallow deer (Pemberton and Smith 1985), roe deer (Hartl and Reimoser 1988) and blue and black wildebeest (Corbet et al. 1994). Four loci were reported by Roed (1985) in reindeer. Three loci were found in white-tailed deer (Sheffield et al. 1985). At least one locus was reported in 28 bovid and giraffid species (Georgiadis et al. 1990) and ten deer taxa (3 alleles) (Emerson and

Tate 1993). Two loci were reported in red deer (Gyllensten et al. 1983). I designated the two bands I observed as reflecting the products of *SOD-1** and *-2**, because no other bovid or cervid researcher reviewed has assigned a cellular or evolutionary origin for either.

3.3 Conclusions

Thirty-nine presumed loci were resolved in this study. I observed Mendelian variability at 18 of these loci, and 21 loci were monomorphic. This number of variable loci is sufficient for the statistical analyses on inter- and intra-population variability that are presented in subsequent chapters. Little work has been done on genetic variability in Bovidae, especially antelope, so I compared expression in impala with expression and interpretation in the few bovids available, and in a close relative taxon, the Cervidae. My interpretation of variability at most loci concurred with that of other workers in other bovids and cervids. However, there were some interpretations, particularly of the polymorphic loci *GPI** and *MPI**, with which I disagreed and for these I offered alternatives. To ensure the correct interpretation of banding patterns,I have attempted to substantiate my interpretations thoroughly with such criteria as consistency, agreement of genotype frequencies with Hardy-Weinberg equilibrium and known subunit structure.

Tissue distribution of gene expression in impala did not differ greatly from that in other bovids. Perhaps this was because differences in the tissue distribution of gene expression between cervids and bovids were not as great as might be expected in different families. It follows that differences between taxa in the same family would be even smaller. Furthermore, bovids in Africa underwent speciation within tribes during such short geological timespans (Vrba 1985b) that neither fossil (Vrba 1979), ribosomal RNA (Gatesy et al. 1992; Allard et al. 1992) immunological (Lowenstein 1986) nor allozyme (Georgiadis et al. 1990) differences have accumulated sufficiently between species to allow estimation of clear phylogenies. Thus bovid taxa are so closely related that few isozyme differences would be expected amongst them.

Gene number, ability to form heteropolymers and the tissue specificity of gene expression are only useful for comparisons between species if they vary at the taxonomic level under investigation, which in this case is families and above (Murphy et al. 1990). Nevertheless, the similarity of gene expression patterns among bovid species does not refute the hypothesis that bovids diverged recently and are thus closely related.

It is possible to deduce evolutionary patterns by tracing presumptive gene duplication events and trends from generalised to restrictive tissue expression. Where two loci (such as the *LDH** loci) are present in many species, the probability that gene duplication events occurred independently in each species is very small. Only if a locus is evident in one taxon and not in another, or if interlocus heteromers cease forming in one taxon while they still form in the other, can one draw conclusions about the time and manner in which that locus has evolved. Detailed impala isozyme data were presented in this study. To make comparisons with other bovids, appropriate data must be obtained from more species.

It was difficult to comment on evolutionary patterns based on the tissue specificity of expression in impala, as no surveys of tissue expression in other bovids were encountered for comparison. Tissues used in other studies were often the only ones available or were chosen because they had shown good results in other studies. However, it would be worthwhile to do such surveys in other bovids, as it would provide additional data with which to support bovid phylogenies.

In derived groups, one would expect to find more loci coding for the same enzymes (paralogous loci) than in more basal groups. Therefore impala, if primitive or basal as suggested by Georgiadis et al. (1990), should have fewer loci per enzyme than more derived groups, such as wildebeest and sable. The hypothesis that there are no differences in the number of loci per enzyme in different bovid species was not contradicted by isozyme data in impala. GPI, an example of a protein that did represent one presumptive locus in impala in this study and two in

sable, also represented two presumptive loci in impala in another study (Grobler and van der Bank 1994b). So this difference was considered to be one of interpretation and not evidence of paralogous loci in a derived species. These differences in interpretation in the same species emphasize the need to support interpretations with supplementary evidence such as other studies or fit to Hardy-Weinberg proportions.

CHAPTER 4

POPULATION GENETICS OF IMPALA

4.1 Introduction

Antelope lineages are characterized by a diverse array of extinct and living species. Most of these species were created by repeated vicariant events in the arid areas of eastern and southern Africa in the late Pliocene and early Pleistocene (Gentry 1978; Vrba 1984). However, only one fossil species from the tribe or subfamily Aepycerotinae (Artiodactyla: Bovidae) is known, and it is morphologically indistinguishable from present-day impala, *Aepyceros melampus*. Impala appear in the fossil record towards the end of the Miocene, about five million years ago (Vrba 1985b). While many fossil lineages appear and others disappear during the late Pliocene and early Pleistocene, around 2,5 million years ago, impala appear to have remained unchanged. In addition, at least one study has found impala to be the bovid group most divergent from other groups (Georgiadis et al. 1990). It is unclear why geographic subdivision failed to initiate speciation in the lineage leading to present-day impala.

Impala inhabit the abundant, widespread ecotone that forms the border between grassland and woodland near surface water (Smithers 1983; Vrba 1984). They can adapt their feeding habits during climatic cycles by grazing or browsing as food availability varies (Vrba 1984). Grassland and woodland have expanded or contracted during periods in climatic cycles (Tyson 1986) that coincide with bovid lineage turnover (Vrba 1985a). While environmental change may have driven speciation and extinction among ecologically specialised species that were pure grazers or pure browsers, impala apparently thrived, as feeding preference is thought to be a major factor in bovid evolution (Vrba 1985a; 1985b). Impalas' habitat, although it has shrunk, expanded and moved as the climate has

changed, has been more persistent than the homogeneous habitat of specialist bovids (Vrba 1984 p74). One possible reason for the success of impala and the lack of divergence in its lineage is that local adaptability and persistence have led to only a small amount of genetic subdivision among impala populations and no significant loss of variability due to population bottlenecks. A related hypothesis is that impala shows more population subdivision than other species such as springbok, *Antidorcas marsupialis*, which migrate and mix (Smithers 1983; Skinner 1993).

Since bovids are one of the most divergent mammal taxa and occupy such diverse habitats, population genetics of a bovid such as impala could contribute independent evidence of the nature and timing of environmental changes (Vrba 1985b; Gatesy et al. 1992). Genetic analysis of populations may also reveal effects of past population events. For example, population size restrictions ('bottlenecks') may reduce variability (Nei et al. 1975) or cause long-term deviations from mutation-drift equilibrium (Watterson 1984).

Hardy-Weinberg equilibrium of genotype frequencies is the biological model on which most population genetic hypotheses are ultimately based (Weir 1990). To be in Hardy-Weinberg equilibrium, a population is assumed to be infinitely large, undergo random mating, receive no migrants and experience no mutation or selection. Many statistics have been developed around this model, to test whether a lack of equilibrium, for example, is due to a violation of one or more of these assumptions, and to help point to what could be causing that violation. In this way the process of microevolution, the infinitesimal changes that take place at the level of the population, can be dissected. An understanding of microevolutionary processes and their effects is fundamental to the formulation of adequate hypotheses concerning macroevolution, or the maintenance and change of species.

Impala are distributed from northern and eastern South Africa to Kenya and Uganda and from Mozambique to Botswana. Their extensive natural

range and abundance make impala an ideal subject for a population genetic study, since large samples can be obtained from a wide range of populations with varied sizes and histories. Sampling is facilitated by the hunting and culling of large numbers of impala on a regular basis. Though even sample sex ratios would be ideal for a study such as this, they depended entirely on culling policies, so were skewed in some cases.

Osterhoff et al. (1972) found little electrophoretic variability in three proteins in impala in the Kruger National Park. Grobler and van der Bank (1994b) found fairly low levels of heterozygosity (H = 0,037) in South African impala on the basis of somewhat small sample sizes and found 88% of genetic diversity within populations. By analysing mtDNA restriction site variation, Georgiadis (pers. comm.) found substantial variation and subdivision in populations in northern Tanzania and Botswana, consistent with isolation by distance. A small, isolated population of black-faced impala (*Aepyceros melampus petersi*) exists in northern Namibia (Smithers 1983). Despite geographic isolation and small morphological differences between black-faced and common impala (Smithers 1983), no clear genetic subdivision has been found between these two groups of impala (N. Georgiadis, pers. comm.). In this chapter I test the null hypothesis that southern African impala comprise one randomly mating population.

Knowledge of the population genetics of impala has implications for the management of impala and other bovid populations (Gyllensten et al. 1983). Impala is a key herbivore in managed nature reserves and has the potential to become an important alternative to cattle in marginal areas. Impala are highly managed in southern Africa (D. Rowe-Rowe, R. Smith, pers. comm.). Although females are preferentially captured and translocated, hunting and culling usually favour males (Brooks 1975). So together, these processes should have a minimal effect on variability unless they reduce populations drastically. A small proportion of males mate (Smithers 1983), resulting in small effective population sizes, and this may cause inbreeding to become a problem even in fairly large populations. In this study I estimated the effects of these processes on

managed populations. I also tested the hypothesis that smaller sub-populations and those which are heavily culled, particularly those that are fenced in nature reserves, are not inbred.

An interesting question concerns the population genetics of sex-biased dispersal: Which sex contributes more to gene flow?. Since male mammals generally, and impala males in particular, disperse more than females (Greenwood 1980; Schenkel 1966; Jarman 1974; Murray 1981, 1982a, 1982b; Estes 1993; various reserves, pers. comm.), one expect males to contribute more to gene flow and would thus expect fewer allele-frequencies differences among males than among females. To properly test the hypothesis that males contribute more to gene flow than females, through greater effective dispersal (Greenwood 1980), one would analyse allele-frequency distributions within and between male and female impala, utilizing large sample sizes from many herds or groups in one area where impala can move freely. If, as is the case with impala, very few males mate while many females do, and if one sex disperses more than the other, one would expect allele frequency differences to be maintained each generation, as non-random mating would be occurring. Even with random mating, two generations are required to reach Hardy-Weinberg equilibrium with initial allele-frequency differences between the sexes.

In this study, at any one locality, samples sufficiently large to detect the small allele-frequency differences required were unavailable. Therefore, I performed analyses on the entire population represented in this study. The Eastern Transvaal samples and the Natal samples each represent groups of impala with only minor barriers to dispersal. The Eastern Transvaal impala only have distance, rivers or scalable fences between them. The Natal impala in this study, although separated by fences and much 'hostile' land, have been connected in the recent past by translocation. So gene flow revealed in this part of the study would be due in part to effective dispersal within these areas and in part to dubitable dispersal among far-flung areas like Zimbabwe and Nylsvlei. I used Wright's *F* statistics, Slatkin and Barton's (1989) private allele method and Slatkin's (1981) conditional average allele frequency method to com-

pare allele frequencies in each sex and thus compare gene flow arising from effective dispersal. These first two methods are consistent over a wide range of assumptions about selection, mutation and population structure (Slatkin 1987).

4.2 Results

4.2.1 Hardy-Weinberg equilibrium

In the pooled sample only genotypic frequencies of CK-C differed significantly from Hardy-Weinberg proportions (p = 0,002). In the individual samples, only genotypic frequencies of GPI at Mkuzi differed significantly from Hardy-Weinberg proportions (p = 0,0025). Genotypic frequencies of MPI at Sabi Sands also deviated significantly from expected proportions (p = 0,0027) when low frequency alleles were pooled.

Table 4.1 Results of log-likelihood ratio G-tests from the nested contingency table analysis of allele-frequency homogeneity.

Level/d.f.		CK-C	PROT-2	GPI	MPI	PEP-B	PGM-2
1	9	29,12**	31,26**	124,70**	8,12	30,46**	50,88**
5	1	15,67**	4,37	28,63**	0,80	4,46	17,35**
4	2	11,29*	24,55**	71,74**	4,66	2,19	21,99**
3	3	0,64	0,19	10,83	1,15	5,46	8,91
3	1	1,50	0,54	12,23**	0,89	15,57**	2,28
2	2	0,03	1,63	1,27	0,63	2,78	0,35

* $p < 0,05$ (with sequential Bonferroni adjustment)
** $p < 0,01$ (with sequential Bonferroni adjustment)

4.2.2 Allele-frequency homogeneity

The first contingency-table analysis performed included the 'herds and within park' level (Figure 2.2). Allele frequencies between herds at

Klaserie were significantly different at one locus, *MPI** (p = 0,005; d.f. = 1; not shown in Table 4.1). However, sample sizes from Klaserie were small (n = 10 and 11). To remove the effects on the overall analysis of small sample sizes in herds and within-park samples, the analysis was repeated without the 'herds and within-park samples' level (see model in Table 2.2).

In spite of the low power of tests for homogeneity, several significant allele frequency differences were detected among localities. In Figure 2.2, asterisks indicate loci that showed significant departures from homogeneity after sequential Bonferroni adjustment. There were significant differences among individual localities at 5 loci, *CK-C** (p = 0,0006), *PROT-2** (p = 0,0003), *GPI** (p = 1,46 x 10^{-22}), *PEP-B** (p = 0,0004) and *PGM-2** (p = 7,36 x 10^{-8}); between Pongola and Natal at two loci, *GPI** (p = 0,0005) and *PEP-B** (p = 0,00008); among 'regions' at four loci, *CK-C** (p = 0,0035), *PROT-2** (p = 0,000005), *GPI** (p = 2,65 x 10^{-16}) and *PGM-2** (p = 0,00007); and between 'countries' at three loci, *CK-C** (p = 0,00008), *GPI** (p = 8,76 x 10^{-8}) and *PGM-2** (p = 0,00003).

4.2.3 Geographic allele frequency distributions

Relationships between allele frequency and geographic distance are shown graphically in Figure 4.1. The distribution of CK-C alleles indicates a difference between northern and southern populations. For example, 95% confidence intervals do not overlap between Zimbabwe (p = 0,873) and most other populations besides Nylsvlei. *CK-C*100* is fixed in Natal, emphasizing differences between Natal impala and other impala. The common allele at PROT-2 at Nylsvlei (p = 0,976) is closer to fixation than in all other populations. Its 95% confidence interval does not overlap with that in almost all other subpopulations. This suggests that the population at Nylsvlei has lost variability, possibly due to inbreeding. No geographic trend is apparent from PROT-2. Differences between Tembe and Albert Falls and Mkuzi reflect sampling of genes for a small founder populations from Mkuzi. The common allele at GPI is fixed in Northern populations, almost fixed in Eastern populations (e.g. Sabi Sands p =

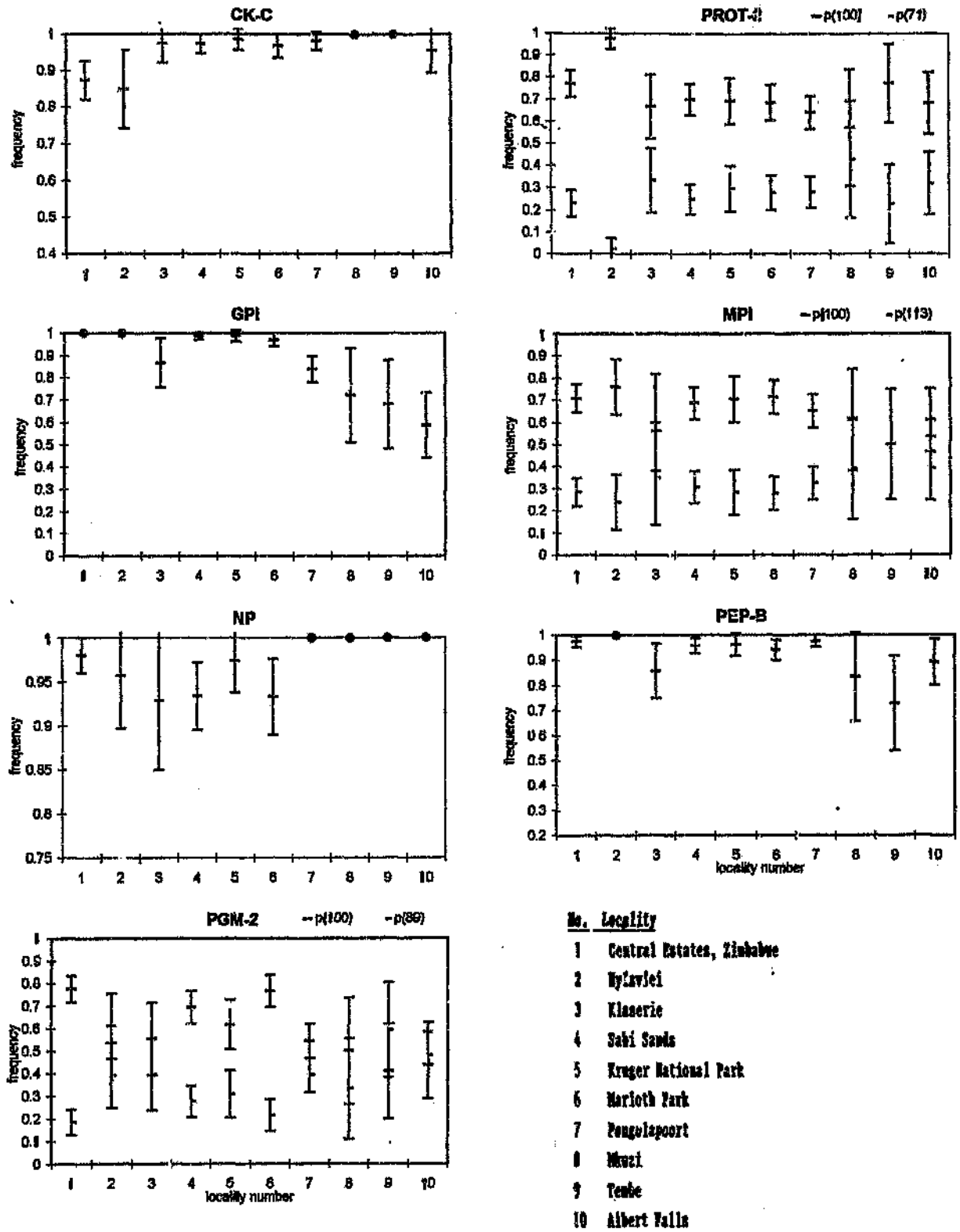
0,988), and its frequency differs significantly from the frequency of the same allele in Natal (e.g. Mkuzi p = 0,587). Klaserie's low frequency and high variance reflects its small sample size. The intermediate frequency of *GPI*100* at Pongola (p = 0,840) is expected because of its geographical position between impala populations in Natal and the Transvaal and the mix of impala from both areas. PEP-B and NP show a similar though less marked pattern. *NP*100* is fixed in Natal and Pongola, further emphasizing differences between Natal impala and other impala. PGM-2 shows a complex pattern of variation, but three rough groups are apparent: Zimbabwe, Transvaal, and Natal and Pongola. At Mkuzi and Albert Falls, p(**100*) < p(**89*). A slight trend of Northern versus Natal populations is observed with MPI, and the frequencies of both alleles present at Albert Falls are the same.

4.2.4 Cluster analysis of genetic distances

Figures 4.2, 4.3 and 4.4 show clustering results and multidimensional scaling plots based on each distance measure. Distances were calculated for a matrix of ten localities using all non-monomorphic loci. UPGMA clusters based on Nei's and Rogers' distances both show the same structure (Figure 4.2 a and b), while clusters based on the chord distance are slightly different, one difference being that the largest genetic distance is between Nylsvlei and all other populations(Figure 4.2 c). Clusters based on Nei's and Rogers' distances show the largest distance as that between Natal and other populations, and show Nylsvlei clustered with Transvaal.

Neighbour-joining of each genetic distance matrix produced similar clustering, except for the inclusion of Klaserie and Pongola in the Transvaal group with Nei's distance (Figure 4.3 a). There was also ambiguity between neighbour-joining and UPGMA about the placing of Klaserie and Pongola, and neighbour-joining was not as sensitive to the small population size in Nylsvlei as was UPGMA.

Two-dimensional multidimensional-scaling analysis produced similar plots with all distance matrices (Figure 4.4). Plotted points appeared to roughly represent the sample populations' geographic arrangement.



No.	Locality
1	Central Estates, Zimbabwe
2	Nylsvlei
3	Klaserie
4	Sabi Sands
5	Kruger National Park
6	Marloth Park
7	Pongolapoort
8	Mkuzi
9	Tembe
10	Albert Falls

Figure 4.1 Diagrams showing geographical distributions of frequencies of the common allele at each polymorphic locus. Frequencies of the two common alleles at a locus are plotted if there are more than two alleles at that locus. Vertical bars represent four standard errors and approximate a 95% confidence interval.

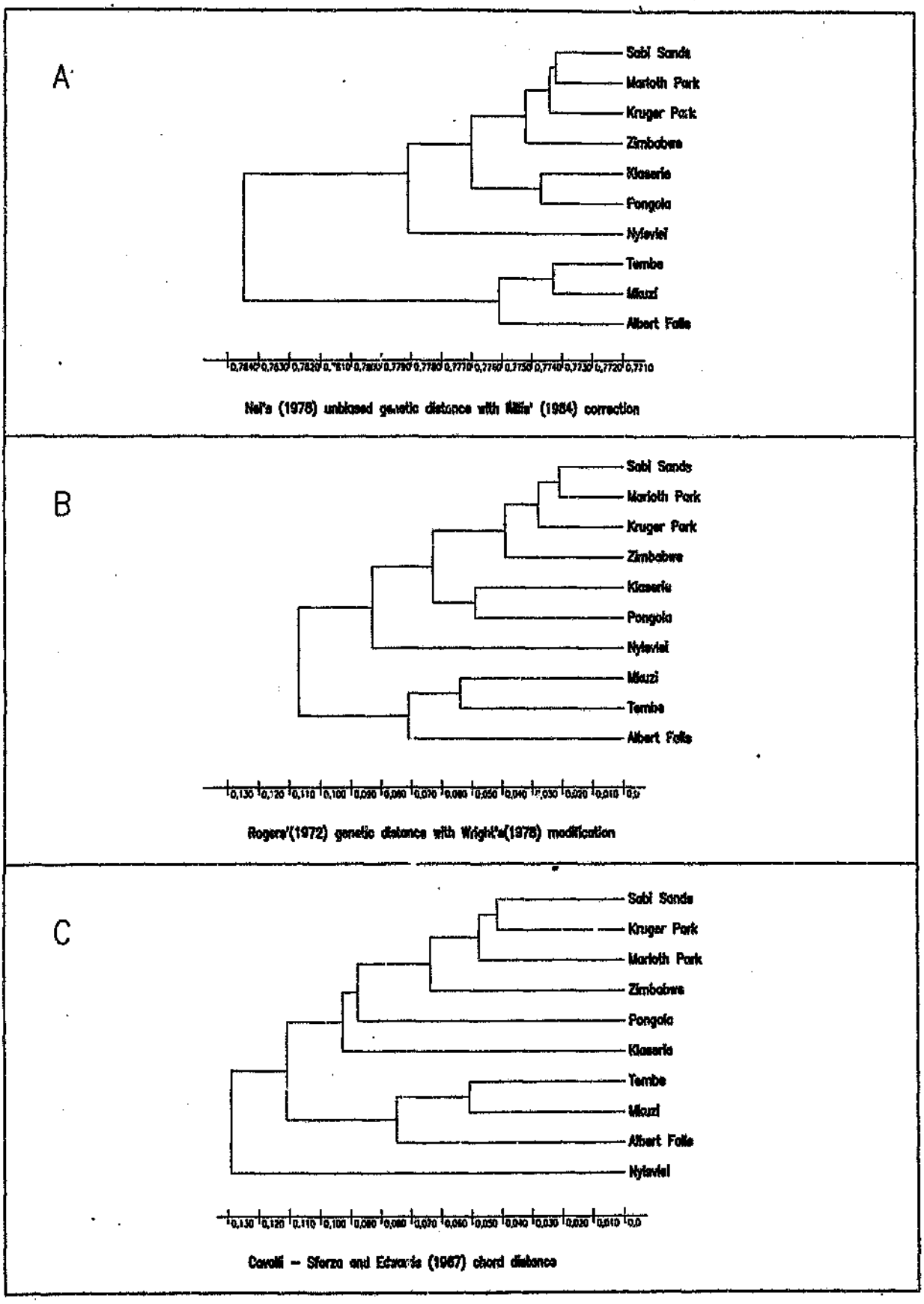


Figure 4.2 UPGMA clustering diagrams based on three genetic distance measures.

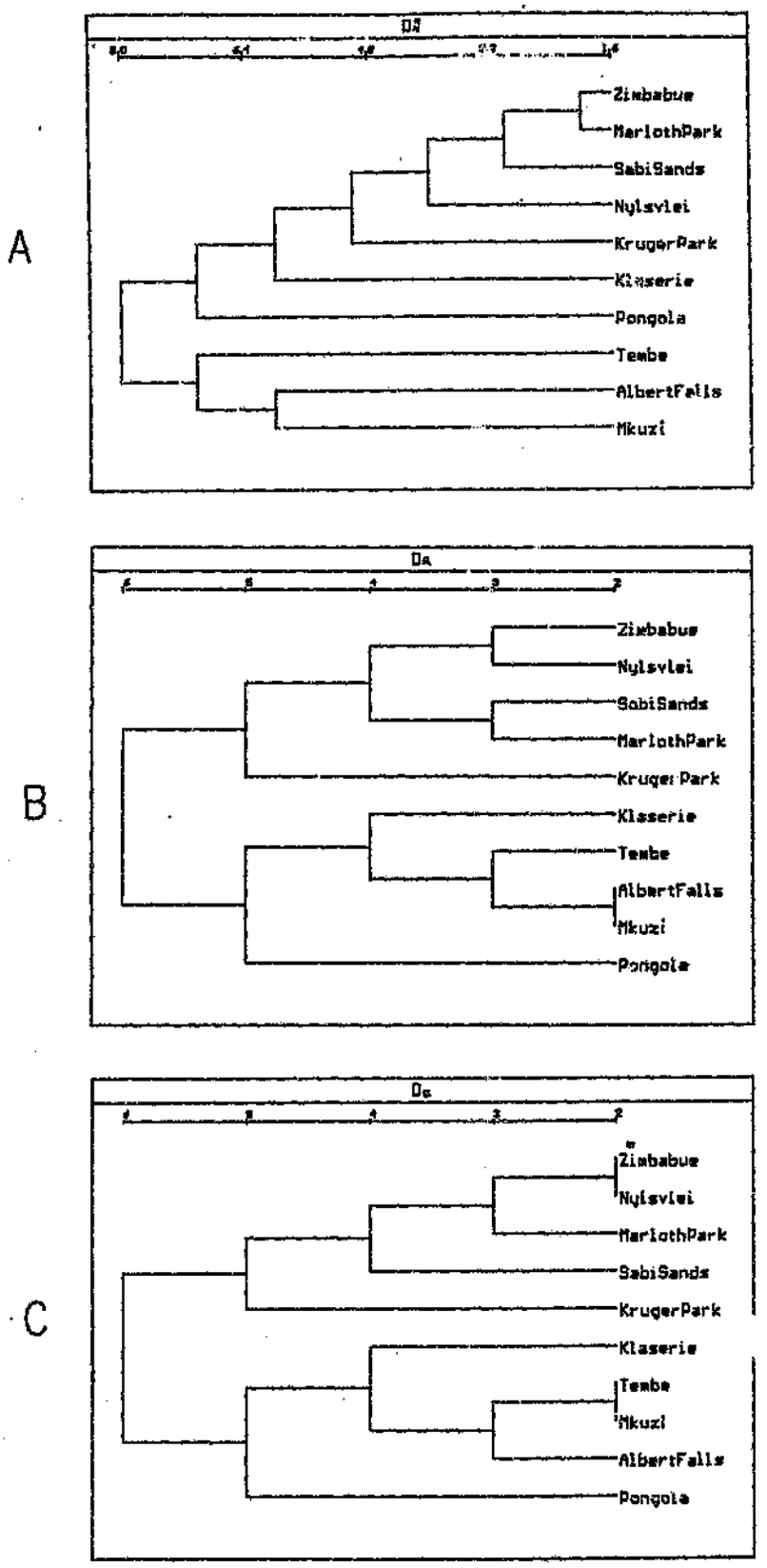


Figure 4.3 Neighbour-joining diagrams based on three genetic distance measures: Nei's (1978) unbiased genetic distance with Hillis' (1984) correction (D_N^*), Rogers (1972) genetic distance with Wright's (1978) modification (D_R), and Cavalli-Sforza & Edward's (1967) chord distance (D_C). Scales refer to topology only, not actual genetic distance.

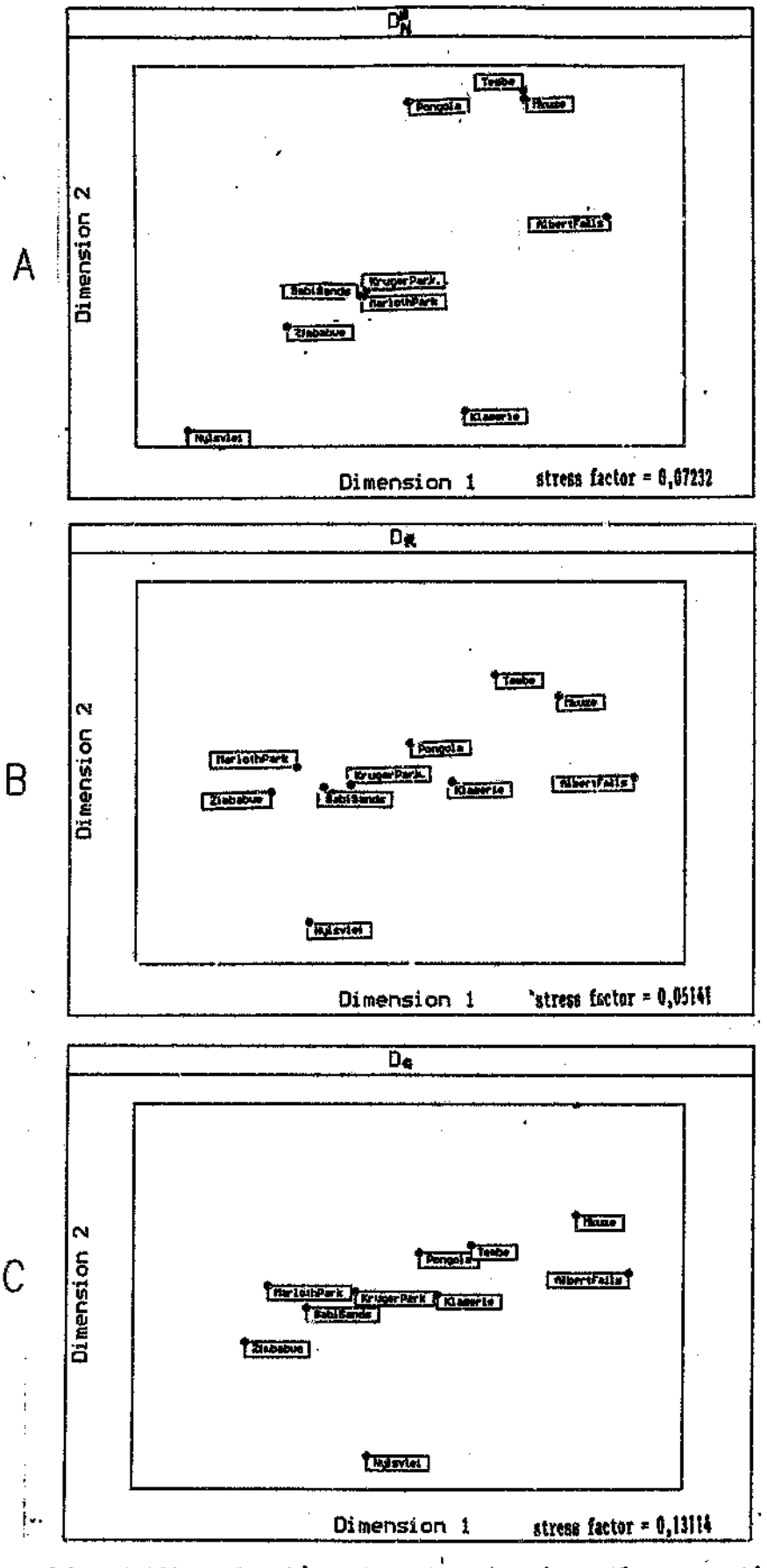


Figure 4.4 Multidimensional scaling plots based on three genetic distances: (A) Nei's (1978) unbiased genetic distance with Hillis' (1984) correction (D_N^*), (B) Rogers (1972) genetic distance with Wright's (1978) modification (D_R), and (C) Cavalli-Sforza & Edward's (1967) chord distance (D_C).

4.2.5 Gene diversity analysis

Table 4.2 Various measures of heterozygosity in impala, and their variances.

Measure	Value	Reference
H_o	0,046	this study
$V(H_o)$	0,0000748	this study
H_s	0,047045	Nei (1978)
$V(h_s)$	0,013512	Nei and Roychoudhury (1974)
H_s (unbiased)	0,047102	Nei (1978)
$V(h_s$(unbiased))	0,013541	Nei and Roychoudhury (1974)
M(biased)	0,049368	Kimura and Crow (1964);
$V(h$(biased))	0,014348	Watterson (1974); Stewart
M(unbiased)	0,049431	(1976); Li and Nei (1975)
$V(h$(unbiased))	0,014364	and Fuerst et al. (1977)

Population gene diversities and standard errors for impala. From gene diversity analysis (Chakraborty 1980).

	H_s	s.e.
Zimbabwe	0,0474	0,0170
Nylsvlei	0,0459	0,0163
Klaserie	0,0624	0,0228
Sabi Sands	0,0435	0,0191
Kruger Park	0,0406	0,0199
Marloth Park	0,0678	0,0304
Pongola	0,0532	0,0227
Tembe	0,0579	0,0253
Albert Falls	0,0555	0,0237
Mkuzi	0,0577	0,0246
Average (H_s):	0,0511	0,0198

Table 4.2 shows various measures of average heterozygosity or gene diversity and their variances, as well as population gene diversities and their standard errors. Estimates of H_s(biased) and H_s(unbiased) were supported by jackknifing over single-locus heterozygosities. Expected heterozygosity or gene diversity (H_s = 0,047) was only slightly larger than observed heterozygosity (H_o = 0,046), indicating that it was safe to assume Hardy-Weinberg proportions. Unbiased gene diversity (H_s(unbiased)

= 0,0471) was almost the same as biased gene diversity (H_s(biased) = 0,0470), indicating that the estimate of heterozygosity was not adversely affected by small sample sizes. Population heterozygosities are highest in Klaserie and Marloth Park and lowest in Nylsvlei and Zimbabwe (Table 4.2). Klaserie has the largest population among those studied, and Nylsvlei the smallest.

Table 4.3 Gene diversity analysis (Chakraborty 1980), including absolute and relative gene diversities for each locus and averages over loci. Levels referred to are those in Figure 2.2.

Absolute gene diversities

Locus	Within levels H_4	Between 4 within 3 D_{43}	Between 3 within 2 D_{32}	Between 2 within 1 D_{21}	Between 1 D_{1T}	Total H_T
AAT-2	0,00159	0,00000	0,00000	0,00000	0,00001	0,00160
CK-C	0,07974	0,00027	0,00004	0,00311	0,00152	0,08468
EST-D	0,00119	0,00000	0,00001	0,00000	0,00000	0,00120
EST-5	0,00430	0,00000	0,00000	0,00009	0,00000	0,00439
GAPDH	0,01995	0,00000	0,00071	0,00009	0,00002	0,02078
PROT-2	0,39615	0,00409	0,00122	0,01505	0,00051	0,41703
GPI	0,19237	0,00192	0,00665	0,02926	0,00408	0,23428
ALAT	0,00880	0,00000	0,00015	0,00001	0,00001	0,00896
HK	0,00469	0,00000	0,00009	0,00002	0,00000	0,00479
IDHP-1	0,10544	0,00000	0,07372	0,01469	0,00118	0,19502
LDH-A	0,02410	0,00000	0,00124	0,00048	0,00004	0,02586
LDH-B	0,00139	0,00000	0,00001	0,00001	0,00000	0,00140
MPI	0,44856	0,00161	0,00254	0,00591	0,00063	0,45924
NP	0,05524	0,00000	0,00027	0,00136	0,00002	0,05688
PEP-B	0,14747	0,00277	0,00524	0,00415	0,00088	0,16051
PEP-A	0,00430	0,00000	0,00000	0,00009	0,00000	0,00439
PGM-1	0,00787	0,00000	0,00002	0,00008	0,00000	0,00797
PGM-2	0,49259	0,00518	0,00554	0,01213	0,00711	0,52254
	Mean and s.e. including monomorphic loci					
Mean	0,05117	0,00041	0,00250	0,00222	0,00041	0,05671
s.e.	0,01977	0,00019	0,00189	0,00094	00021	0,02115
	Mean and s.e. excluding monomorphic loci					
Mean	0,11088	0,00088	0,00542	0,00480	0,00089	0,12286
s.e.	0,03877	0,00038	0,00405	0,00189	0,00044	0,04109

continued...

Table 4.3 continued

Relative gene diversities (G_{st})

Locus	Within levels H_4/H_T	Between 4 within 3 H_{43}	Between 3 within 2 H_{32}	Between 2 within 1 H_{21}	Between 1 H_{1T}	Total
AAT-2	0,99284	0,00000	0,00000	0,00000	0,00716	1,00000
CK-C	0,94171	0,00319	0,00042	0,03673	0,01795	1,00000
EST-D	0,99458	0,00000	0,00442	0,00090	0,00010	1,00000
EST-5	0,98013	0,00000	0,00000	0,01965	0,00023	1,00000
GAPDH	0,96029	0,00000	0,03425	0,00428	0,00118	1,00000
PROT-2	0,94996	0,00982	0,00293	0,03609	0,00121	1,00000
GPI	0,82112	0,00821	0,02838	0,12488	0,01742	1,00000
ALAT	0,98163	0,00000	0,01717	0,00069	0,00051	1,00000
HK	0,97833	0,00000	0,01805	0,00335	0,00028	1,00000
IDHP-1	0,54067	0,00000	0,37799	0,07532	0,00603	1,00000
LDH-A	0,93211	0,00000	0,04798	0,01844	0,00147	1,00000
LDH-B	0,99370	0,00000	0,00516	0,00102	0,00013	1,00000
MPI	0,97674	0,00351	0,00552	0,01286	0,00137	1,00000
NP	0,97108	0,00000	0,00470	0,02389	0,00034	1,00000
PEP-B	0,91876	0,01724	0,03267	0,02584	0,00550	1,00000
PEP-A	0,98013	0,00000	0,00000	0,01965	0,00023	1,00000
PGM-1	0,98775	0,00000	0,00271	0,00916	0,00039	1,00000
PGM-2	0,94268	0,00991	0,01060	0,02320	0,01361	1,00000
Mean and s.e. excluding monomorphic loci						
Mean	0,90247	0,00717	0,04406	0,03910	0,00724	1,00000
s.e.	0,03996	0,00155	0,03352	0,01210	0,00261	

Gene diversity analysis was performed according to the *a priori* model of population subdivision described in Table 2.2. Components of relative gene diversity are shown in Table 4.3. The within-population coefficient of gene diversity was 90,2%. The between-population coefficient of gene diversity was 9,8%. This was broken down into 0,7% among Natal populations (a low value was expected, since Tembe and Albert Falls impala were derived from Mkuzi impala), 4,4% among Eastern Transvaal populations and between Natal and Pongola populations, 3,9% among Nylsvlei, Eastern Transvaal and Natal, and 0,7% among Zimbabwe and all other populations.

4.2.6 Estimation of gene flow

I used allelic frequencies of 10 alleles at seven polymorphic loci (including NP) to estimate *F* statistics. Average sample size from 10 localities was 46. Jackknifed estimates were as follows (with 95% confidence intervals derived from bootstrapping): F_{ST} 0,0353 (0,0109 to 0,0899), F_{IT} 0,0355 (-0,0225 to 0,1154), and F_{IS} 0,0001 (-0,0557 to 0,0552) (Table 4.4). Zero is not included in the 95% confidence interval of F_{ST}, so it follows that F_{ST} is significantly different from zero. From the value of F_{ST} obtained before jackknifing (0,0377), Nm = 6,75. The estimate of V(Nm) obtained by jackknifing Nm over loci was 8,70 and the jackknife-adjusted estimate of Nm, Nm^*, was 5,88 (Table 4.4). F_{IT} and F_{IS} were not significantly different from zero nor were any *F* statistics significantly different from each other. So there is no evidence of a significant level of inbreeding in impala populations represented in this study.

The population subdivision revealed by F_{ST} was also evident in most of the combinations of single sex samples analysed (Table 4.4). However, no values of F_{ST} were significantly different from each other, so there is no evidence to refute the hypothesis that males, who disperse more than females, also contribute more to gene flow. Nm values provide a rough estimate of migration rate but they cannot be compared statistically as their confidence intervals are not known. Their confidence intervals are likely to be large and to overlap, since V(Nm) is very large in some cases.

Table 4.4 Wright's F-statistics calculated according to Weir and Cockerham (1984) and estimates of migration (Nm (F_{ST})) for impala subpopulations. f^*, F^* and θ^* are jackknife-adjusted estimates of f, F and θ (Weir 1990). Numbers in parentheses are 95% confidence limits found by bootstrapping (Weir 1990). V(Nm) is the jackknife estimate of the variance of Nm. Nm' is the jackknife-adjusted estimate of Nm. ** next to a value indicates that it is significantly different from zero ($p < 0,05$).

Sample description	n_{avg}	f	f^*	F	F^*	θ	θ^*	Nm	V(Nm)	Nm'
Ten localities, sexes combined	46	0,0385	0,0081 (-0,0557 0,0552)	0,0651	0,0355 (-0,0225 0,1154)	0,0357	0,0351** (0,0105 0,0899)	6,75	8,70	5,88
1) Males, ten localities	25	0,0141	-0,0206 (-0,1210 0,0875)	0,0575	0,0229 (-0,0601 0,1515)	0,0440	0,0426** (0,0148 0,1243)	5,43	4,74	5,03
Females, eight localities	25	0,0533	0,0178 (-0,0293 0,0694)	0,0920	0,0493 (-0,0136 0,1092)	0,0409	0,0322 (-0,0065 0,0770)	5,86	50,12	2,70
2) Males, eight localities, n≥9	39	0,0134	-0,0208 (-0,1298 0,0842)	0,0541	0,0229 (-0,0518 0,1401)	0,0412	0,0420** (0,0144 0,1274)	5,82	6,67	5,23
Females, six localities, n≥9	33	0,0459	0,0140 (-0,0312 0,0695)	0,0814	0,0433 (-0,0130 0,0935)	0,0371	0,0300** (0,0013 0,0664)	6,49	71,83	1,12
3) Five localities, n≥12 for both sexes: Males	37	-0,028	-0,0308 (-0,0987 0,0926)	-0,013	-0,0198 (-0,0944 0,1301)	0,0152	0,0102 (-0,0032 0,0408)	16,2	248,0	6,06
Females	38	0,0439	0,0169 (-0,0372 0,0916)	0,0744	0,0422 (-0,0138 0,1093)	0,0319	0,0262** (0,0007 0,0595)	7,59	61,84	3,05

The estimate of Nm from Slatkin and Bartons' (1989) private allele method was 12,3, from the nine private alleles observed (Table 4,5). Slatkin and Barton (1989) suggested that errors, such as misclassifying private alleles, may decrease the accuracy of the private allele method for estimating Nm. However, I surveyed up to 25 individuals per locality for variation at non-polymorphic loci and surveyed all individuals for variation at polymorphic loci. Therefore, I should have detected most rare alleles at the 18 non-monomorphic loci I used for this analysis.

To compare gene flow by sex with this method I used the same locality combinations as in Table 4.4. Removing localities with small sample sizes from the tests produced a different value of Nm_{mmt} in only one case, when five localities with $n < 12$ were removed from the male sample. Otherwise only one value is shown. Variances of $Nm(p(1))$ (Table 4.5) were low compared to variances of $Nm(F_{mt})$ (Table 4.4) and jackknife-adjusted values of $Nm(p(1))$ were close to the unbiased estimates. Although the statistical significance of these results is questionable, they support a higher contribution to gene flow by females than by males.

Table 4.5 Estimates of $Nm(p(1))$ from Slatkin and Barton's (1989) private allele method. $Nm_{mt} = Nm(p(1))$. n_{mm} is the average sample size on which Nm is based. $V(Nm_{mt})$ is the jackknife estimate of the variance of Nm_{mt}. Nm^* is the jackknife-adjusted value of Nm_{mt}.

Sample	no. of private alleles	n_{mm}	Nm_{mt}	$V(Nm_{mt})$	Nm^*
1) combined sexes	9	47	12,34	8,73	11,12
2) males - all localities	7	38	9,29		
- 5 localities, n≥12 for both sexes	5	48	18,33	13,30	9,29
3) females	5	57	28,03	8,96	19,22

To estimate gene flow from the distribution of conditional average allele frequencies (Slatkin 1981), I calculated $p(i)$ with the same combinations of localities and sexes as those in Table 4.4, with all alleles at 18 non-monomorphic loci. However, subpopulations with extremely small single-sex samples were excluded from these analyses, as were alleles in a sex class for which there were missing data at any locality. The distribution of $p(i)$ (conditional average allele frequency) for the pooled

sexes, as well as for individual sexes, has a reverse-L shape (Figure 4.5). This shape is expected for a relatively subdivided population because low frequency alleles are present mainly in the populations in which they originated, while only the most common allele at each locus was present in all subpopulations (Slatkin 1981).

Slatkin (1981) found that differences between high and low levels of migration could be detected if the minimal sample size was ten or more. $p(i)$'s for the single-sex cases were based on a minimal sample size of 12. For the pooled-sex case for the same five localities used in the single-sex cases, $p(i)$ was based on a minimal sample size of 39. I plotted distributions of occupancy numbers $i = 1$ to $i = 5$ (Figure 4.5), with localities with large samples ($n > 12$) of both sexes. Distribution patterns were the same as when all ten localities ($i = 1$-10) were used, but with the five localities common to each sex the distributions could be plotted on the same axes and so could be directly compared. The distribution of p(i) in males was not significantly different from the distribution of p(i) in females (Kolmogorov-Smirnov $D = 0{,}40$, m=5, n=5; $p>0{,}05$).

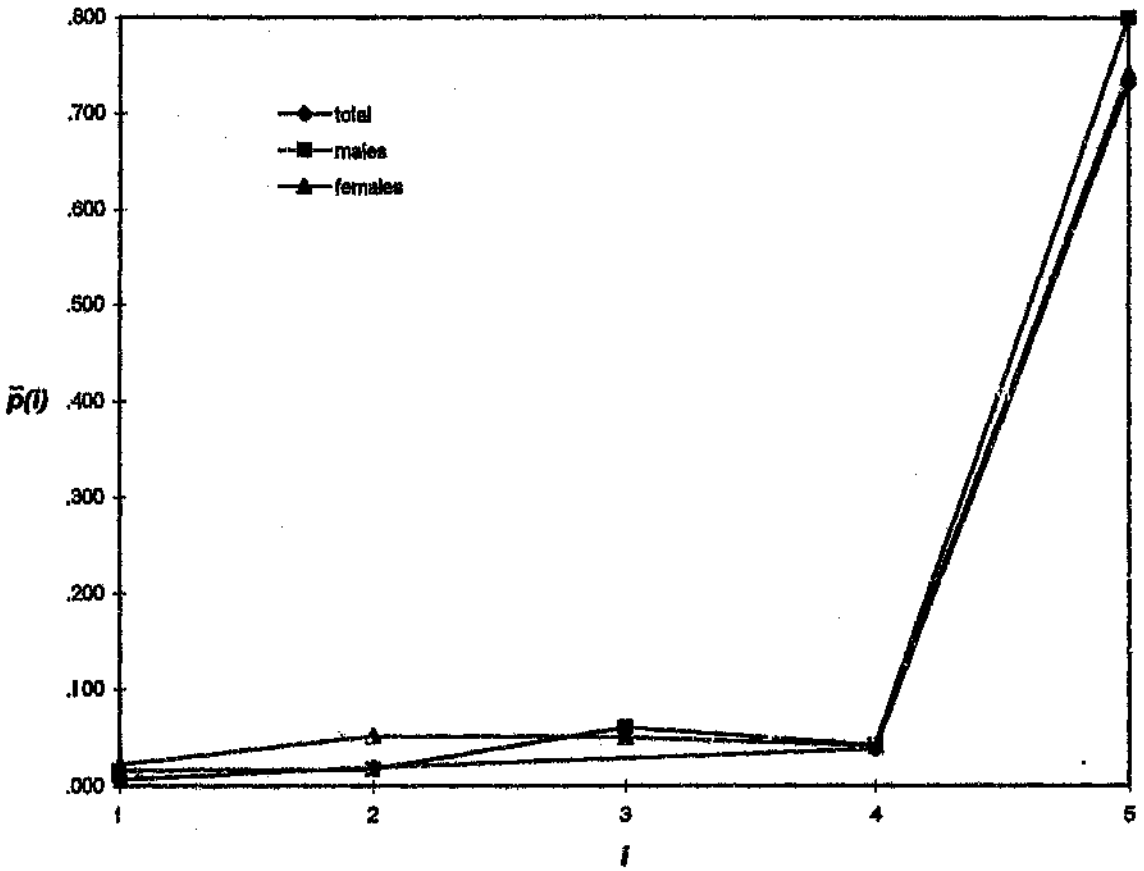


Figure 4.5 Distributions of conditional average frequency. i is occupancy number. For the pooled-sex distribution, occupancy number 4 had no alleles, so it was not included and points adjacent to it were interpolated (Slatkin 1981).

4.2.7 Fit to neutral expectations

Figure 4.6 shows the observed and theoretical distributions of single locus heterozygosities. A Kolmogorov-Smirnov goodness-of-fit test showed that the observed distribution is significantly different from the theoretical distribution (D = 0,733, m=15, n=15; p<0,001). There is an excess of intermediate and low single locus heterozygosities and a deficit of other classes.

When plotted against H (H_e = 0,047), $V(h)$ ($V(h)$ = 0,014) is only just below the 95% significance intervals for three models of mutation (Figure 4.7; from Fuerst 1977, Figure 1 p458). The proportion of polymorphic loci in impala (P = 0,154) is close to the theoretical prediction (0,0138) for H in this study (Eq. 3 and Fig 8, p473 Fuerst et al. 1977).

The observed distribution of allele frequencies in impala was compared to the distribution expected in a population in drift-mutation equilibrium (Figure 4.8). The distributions were significantly different from each other (Kolmogorov-Smirnov D = 0,750, m=20, n=20; p<0,001) and there was an excess of alleles in the lowest frequency class (8,3 expected versus 12 observed). The distribution was J-shaped.

4.2.8 Correlation of population size with heterozygosity

Impala populations showed a slight inverse correlation between population size and heterozygosity, based on population gene diversities (Table 4.2) and census sizes (Table 2.1). However, this correlation was not significant (r = -0,09562; d.f. = 5; p>0.05).

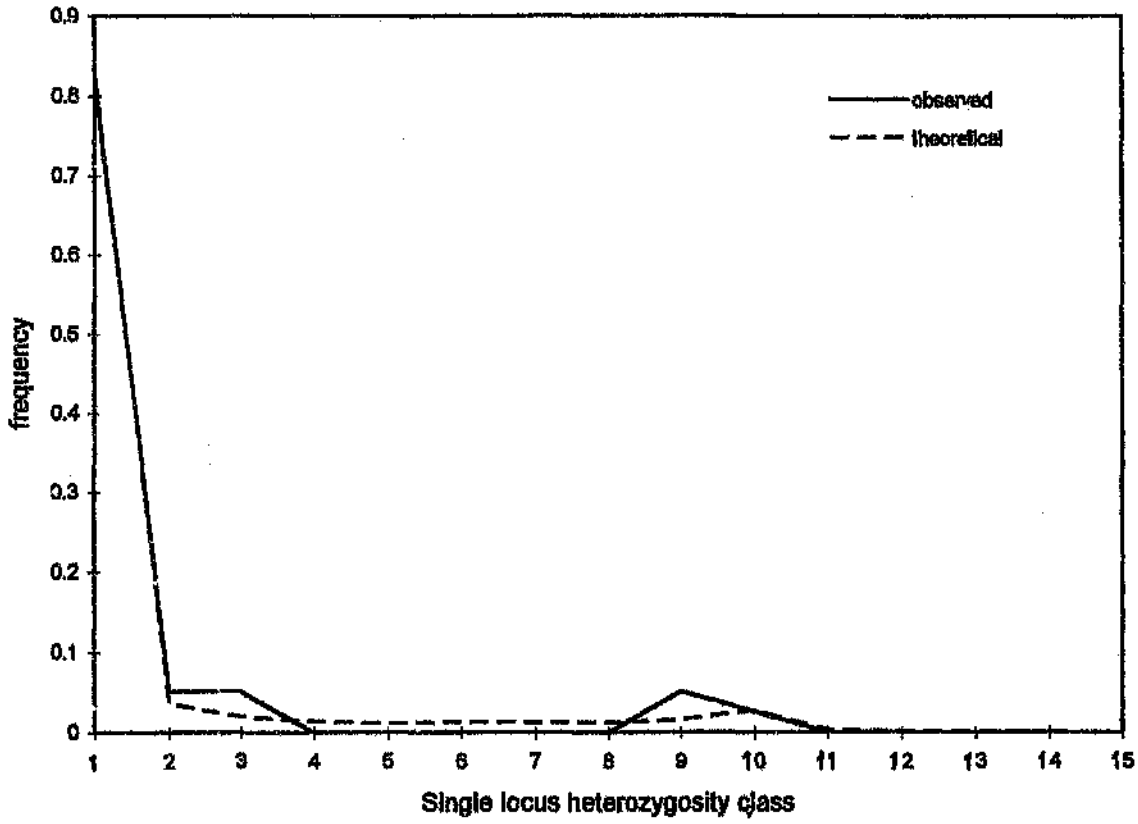


Figure 4.6 Observed and theoretical distributions of single locus heterozygosity. The theoretical distribution is based on $H = 0,05$ (Fuerst et al. 1977). Heterozygosity classes cover the interval (0;0,75] in 0,05 increments. The observed distribution is significantly different from the theoretical distribution (Kolmogorov-Smirnov $D = 0,733$, m=15, n=15; p=0,001). There is an excess of intermediate and low single locus heterozygosities and a deficit of other classes.

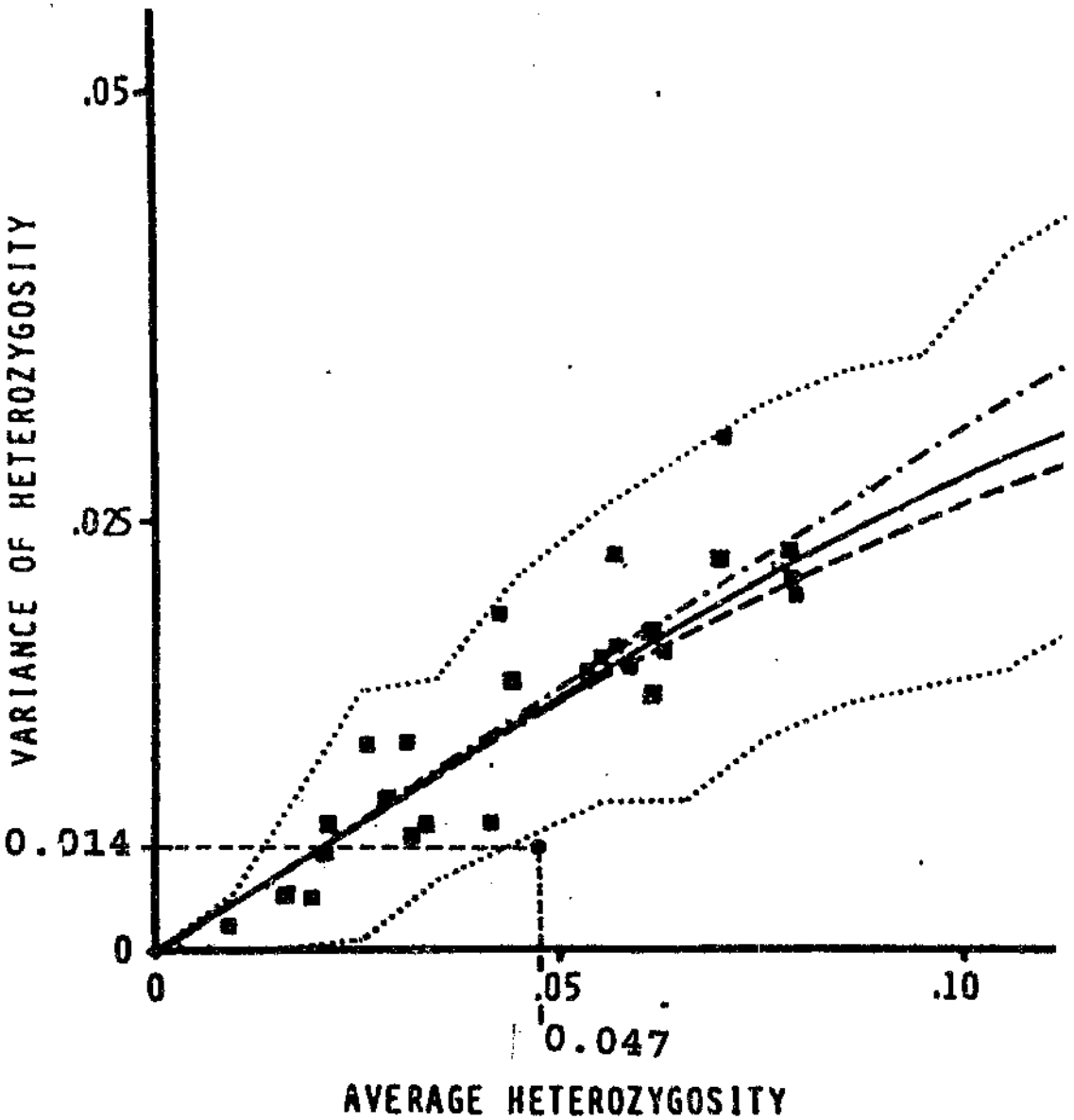


Figure 4.7 Relationship between the observed interlocus variance of heterozygosity and the average heterozygosity of impala, compared to the expected relationships, from Fuerst et al. (1977, Fig 1 p458). ——— : theoretical relationship for the infinite alleles model. — — — : theoretical relationship for the stepwise mutation model. —•—•— : theoretical relationship for the infinite allele model with varying mutation rate. ••••• : 95% significance intervals of the variance obtained empirically by Fuerst et al. (1977). Squares = other mammals.

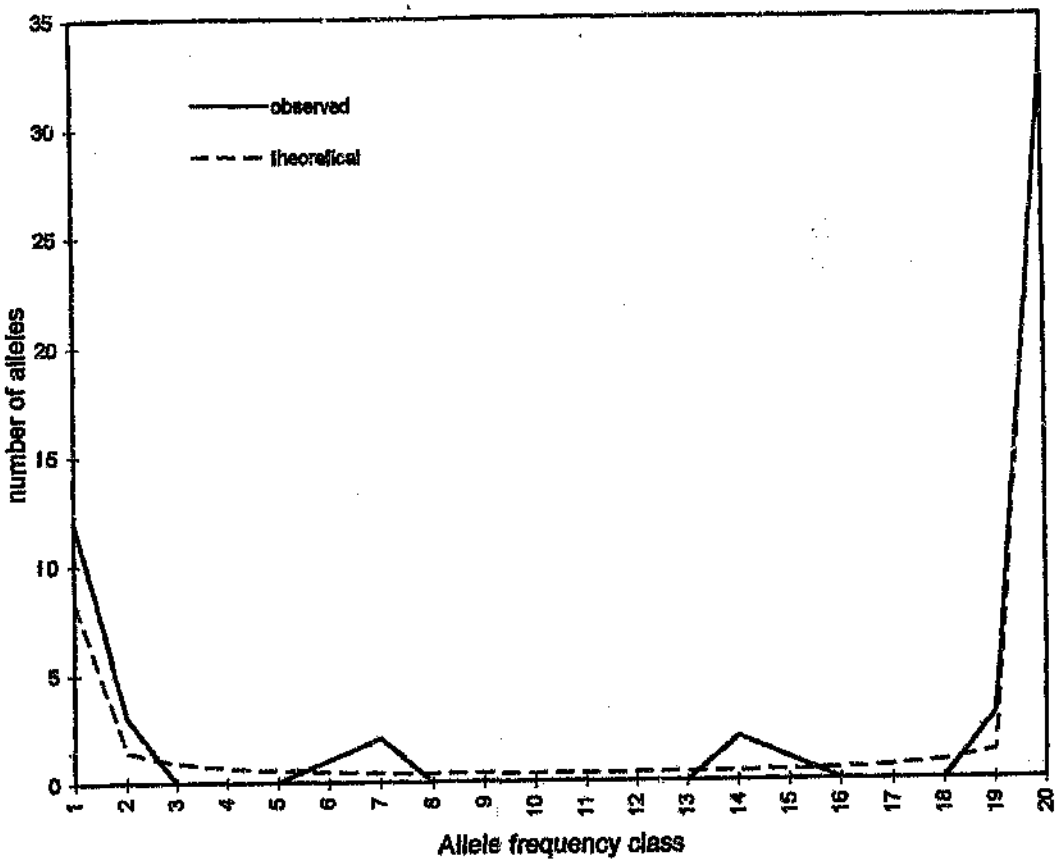


Figure 4.8 Observed and theoretical distributions of allele frequencies (Watterson 1984). The theoretical distribution was based on 39 loci, H = 0,047 and 800 genes sampled per locus. Allele frequency classes cover the interval (0;1] in 0,05 increments. The sum over classes of observed alleles is 63, and over classes of expected alleles is 52,8. The oberved distribution is significantly different from the theoretical distribution (Kolmogorov-Smirnov D = 0,750, m=20, n=20; p<0,001) and there is an excess of alleles in the lowest frequency class (8,3 expected versus 12 observed).

4.3 Discussion

4.3.1 Population structure

The lack of fit to Hardy-Weinberg proportions (4.2.1) points to subpopulation differentiation. Although this is only supported by one locus (CK-C), it is a conservative test with low power, so little chance exists of type I error. Also, it is unlikely that these significant differences are due to errors in my interpretation of electrophoretic variability, as genotypic frequencies showed almost no significant departures from Hardy-Weinberg proportions in samples from subpopulations. At CK-C the deficiency of heterozygotes is consistent with Wahlund's effect which arises when individuals from genetically different subpopulations are included in a sample. Significant departures from Hardy-Weinberg genotypic proportions together with deficits of heterozygotes at one locus each at Mkuzi and Sabi Sands suggest that these subpopulations may not be randomly mating. Since the Sabi Sands population is large, the deficit of heterozygotes suggests that some structure due to isolation by distance or herd subdivision exists. Because Hardy-Weinberg analysis relies on many assumptions, I discuss support for these and other conclusions with the results of other analyses.

G-tests for allele frequency homogeneity reveal significant allele frequency differences among samples at the ten localities (4.2.2). These differences are partitioned among homogeneous subpopulations of impala in three areas broadly classified as northern (Zimbabwe and Nylsvlei), Lowveld and Natal. The cluster analysis of genetic distances (4.2.4) indicates a similar structure, and also indicates that Pongola is genetically intermediate between the Eastern Transvaal and Natal, which probably reflects the translocations of impala from Mkuzi that have occurred in the last 20 years.

To compare genetic distances with those from other studies, I calculated the same distance measures that were used in those studies (4.2.4). The average Nei's (1972) biased genetic distance among pairs of impala popula-

tions is 0,038. This is compared to 0,906 among 11 bovid and cervid species (Baccus et al. 1983). Average unbiased genetic distance (Nei 1978) between impala populations studied by Grobler and van der Bank (1994b) was 0,003. Most of their samples were small and came from small, highly managed populations that were formed by or supplemented with translocated impala. In this study, average unbiased genetic distance (Nei 1978) is 0,025, which is over eight times larger.

The results of the two methods discussed above indicate that subpopulation differentiation is due in part to isolation by distance (Wright 1943). Multidimensional scaling plots correspond roughly to relative geographic positions of each subpopulation (4.2.4) and the distributions of allele frequencies show definite geographic trends (4.2.3).

Average heterozygosity in impala is similar to or greater than levels in other bovids and cervids (Table 4.6). This suggests that impala populations have remained large for a long time without losing genetic diversity through such events as bottlenecks in population size. For example, Grobler and van der Bank (1994b) found a maximum H of 0,043 in their smaller samples of impala populations. H in springbok, *Antidorcas marsupialis*, is 0,042 (Bigalke et al. 1993) and in blue wildebeest is 0,081 (Corbet et al. 1994). H in various species of deer ranges up to 0,041, but is usually lower. H in impala in this study (H = 0,046) is also high compared to H in mammals generally (H = 0,036; Table 4.6).

However, variability at some localities warrants further discussion. Heterozygosity in the Tembe and Albert Falls samples was as high as that in the sample from their source population in Mkuzi, despite Tembe and Albert Falls having very small populations. Sampling error is an unlikely reason for this, as the standard errors of heterozygosity estimates from all three localities were similar. A more likely explanation is that their small founding populations sampled most of the genetic diversity present in the Mkuzi population and have not yet lost variability through subsequent drift.

Table 4.6 Comparative data for heterozygosity, proportion of polymorphic loci, number of alleles per locus and fixation indices (F_{ST} was calculated with different methods by different authors)

Species	Mean indices H	H_e	P	A	F_{ST}	Reference
impala	0,046	0,047	0,154	1,615	0,036	this study
impala	0,030					Grobler & v d Bank 1995
springbok	0,042	0,046	0,123			Bigalke et al. 1993
	G_{ST} within = 88,5%					
wildebeest						
black	0,018					Corbet et al. 1995
blue	0,081					Corbet et al. 1995
	0,012					Georgiadis et al. 1990
topi	0,010					Georgiadis et al. 1990
Damaliscus lunatus						
Coke's hartebeest	0,019					Georgiadis et al. 1990
Alcelaphus buselaphus						
Thomson's gazelle	0,055					Georgiadis et al. 1990
Gazella thomsonii						
roe deer	0,041	0,047	0,156	1,175	0,115	Lorenzini et al. 1993
red deer						
	0,021					Gyllensten et al. 1983
	G_{ST} within = 73%					
	0,004-0,023					Ströhlein et al. 1993
moose	0,020				0,096	Ryman et al. 1980
	G_{ST} within = 90%					
reindeer	0,022		0,059		0,029	Roed 1985
mammals						
in general	0,036					Nevo et al. 1978
	0,039					Wooten and Smith 1985
large grazing						
	0,033					Baccus et al. 1983
22 vertebrates	0,058					Nonnell & Selander 1974

Heterozygosity of impala at Nylsvlei was low, as expected in a small population. Heavy management or inbreeding in a small effective population can cause this condition. Heterozygosity in Zimbabwe was low despite a large sample from the large population at Central Estates. This population has possibly been through a bottleneck or has been heavily managed. The remaining samples with high heterozygosity come from reserves where populations are mainly indigenous and large. These populations undergo

large fluctuations from culling and natural population dynamics (pers. comms., reserve staff and records), but population sizes have been large enough to retain high levels of heterozygosity.

The large within-population coefficient of gene diversity (91,2%) indicates a low level of population structure. Implicit in this conclusion is that no populations sampled are small or isolated enough to have diverged from the others through inbreeding or random genetic drift.

Overall population subdivision as indicated by F_{ST}, while significant, was low compared to, for example, roe deer (*Capreolus capreolus*) from populations of different origin (Lorenzini et al. 1993). This corresponds to high levels of gene flow. $Nm > 2$ is an indication of a lack of strong differentiation (Wright 1978). With $Nm > 1$, gene flow is strong enough to prevent differentiation due to random genetic drift (Slatkin and Barton 1989). F_{IS} is not significantly different from zero, indicating low levels of inbreeding in different localities. On the other hand, distribution of conditional average frequencies (Slatkin 1981) suggests that the population is fairly subdivided. Population recoveries, which entail dispersal by both sexes (Murray 1982), could have caused considerable gene flow and thus helped bring about the low level of population differentiation observed in this study. Population sizes can fluctuate drastically from year to year due to natural and anthropomorphic causes. Consecutive annual censuses at Klaserie have varied by a factor of four to five (Klaserie census; E. Leibnitz). In Mkuzi, heavy hunting took place last century, and this century rinderpest and nagana extermination campaigns have decimated the population. Densities dropped from 22 to four impala/km^2 in Mkuzi in the 1960's, and have since shown an erratic recovery (Goodman 1990).

When certain localities were removed from the analysis (this was done with males only), the estimate of F_{ST} dropped (e.g. F_{ST} dropped from 0,0412 to 0,0279 when Mkuzi was removed), implying that gene flow among remaining populations is greater than that among all localities. This indicates that these localities, including Nylsvlei, Tembe and especially

Mkuzi, could be fairly isolated from the rest of the species (Slatkin 1985), although this effect in the case of Tembe could be due to the removal of a small sample from the analysis (at Tembe, $n = 9$).

The observed variance of heterozygosity in impala falls just outside the 95% confidence limits of the theoretical variance under stepwise, infinite alleles and infinite-alleles-with-variable-mutation-rate models (Fuerst et al. 1977). The observed variance is not likely to be significantly different from that expected under assumptions of neutrality, because the 95% confidence interval of the observed variance, though it is unknown, is likely to overlap with the 95% confidence interval of the theoretical variance.

The distribution of single-locus heterozygosity in impala was significantly different to that expected in the drift-mutation model. The distribution of allelic frequencies in impala was also not consistent with the hypothesis that impala populations are in mutation-drift equilibrium. There was an excess of rare alleles signifying a recent bottleneck (Watterson 1984). Also, the allele frequency spectrum was J-shaped, which is typical soon after a bottleneck or with a low mutation rate (Watterson 1984). As the mutation rate or time since the bottleneck increase, the distribution becomes U- then L- shaped.

The mutation rate in impala is unlikely to be unusually low, or to have changed. A more likely explanation for the lack of fit of the two distributions discussed here is that impala numbers have been severely reduced or populations severely fragmented in the recent past. Extermination campaigns such as those that took place in Natal this century (Goodman 1990), heavy culling regimes and large population fluctuations due to climatic variations are among the possible causes of these reductions. These reductions could have been severe enough to reduce the total number of alleles, which has since exceeded the number expected for the gene diversity evident in impala, and which is not yet in equilibrium with the loss of alleles by drift (Watterson 1984). Population reductions may not have been severe or sustained enough to reduce heterozygosity.

4.3.2 Sex-biased dispersal

While Slatkin and Barton's (1989) private alleles method indicated that females perform more effective dispersal than males, this could not be tested statistically. Two other methods used to test the hypothesis that males are the more effective dispersers failed to refute that hypothesis. Firstly, Wright's F_{ST} for males was not significantly greater than F_{ST} for females. Secondly, the distribution of conditional average allele frequency, which is fairly sensitive for comparing gene flow (Slatkin 1981), failed to detect a clear difference in the contribution to gene flow by either sex. So, the hypothesis that males disperse more effectively than females was neither refuted nor supported by these methods. Some results hinted that females may contribute more to gene flow. This needs to be tested with more powerful methods.

A more specific null hypothesis, that females do not contribute to gene flow, could be tested with mtDNA restriction site or sequence data. MtDNA is maternally inherited and has a high mutation rate, so it should reveal the amount of fine population structure caused by the effective gene flow due to females. If females contribute little or nothing to gene flow, mtDNA variability should show strong subdivisions among areas and possibly among clans.

4.3.3 Biogeography

The limited genetic subdivision evident in impala and their fairly high levels of genetic variability could be evidence that impala have not been affected over their whole range nor for sustained periods by long term climate change (as opposed to short-term, localised climatic variations). Although the present-day distribution of impala is irregular and clumped (Estes 1993), current low levels of population genetic structure are due to high levels of gene flow. Impala are natal dispersers but do not make long distance migrations (Murray 1982), which may partly explain current levels of subdivision in terms of isolation by distance.

I propose that while less adaptable bovids underwent speciation during climate driven turnover events (Vrba 1985a), impala may have been subdivided, but did not undergo sufficient genetic drift or differential selection to speciate. Hence, they persisted unchanged except by small amounts of genetic drift, until their habitat expanded and they could mix again. This process repeated periodically over the last five million years could have given rise to the present genetic structure. Impala may thus have remained evolutionarily unchanged for so long because of their high levels of gene flow, adaptability and abundant, wide distribution.

A contemporary analogy and test of this hypothesis is that African bovid food specialists should show greater population subdivision than do impala. One food specialist for which population genetic data are available is the springbok, albeit from only two populations, one of which was very small (Bigalke et al. 1993). ΣD_{ST}, absolute between-population diversity, is independent of the gene diversity within populations, so it can be used to compare genetic differentiation in different species (Nei 1973). Gene differentiation relative to the total population is $G_{ST} = D_{ST} / H_T$. This measure depends on the population used, so cannot be used to compare species. I would expect ΣD_{ST} to be greater among impala than among springbok. However, ΣD_{ST} = 0,0055 in impala (this study; Table 4.3; $\Sigma D_{ST} = D_{43} + D_{32} + D_{21} + D_{1T}$) and ΣD_{ST} = 0,006 in springbok, which does not support this hypothesis. Between population G_{ST} in sable antelope was 7,73%, but D_{ST} was not furnished by Grobler and van der Bank (1994a). Greater samples from more species will allow this hypothesis to be tested more rigorously. A related hypothesis to be tested with more populations is that although springbok are specialist grazers, they may be homogeneous because of the treks that used to take place, in which large fractions of the entire population aggregated (Skinner 1993).

The extensive degree of genetic subdivision that Georgiadis (pers. comm.) found in the analysis of impala mtDNA in East Africa could be due to

mtDNA revealing finer population structure than allozymes do, or it could be due to differences in the mating system of tropical impala. I propose that mating systems in tropical impala cause less gene flow than temperate breeding systems, as no major cycle of breeding and dispersal occurs there as it does in southern Africa (Estes 1993). Analysis of allozyme variability in tropical impala will allow direct comparisons to be made to test this hypothesis.

4.3.4 Management

High levels of variability, adaptability and the absence of strong population subdivision help make impala an ideal antelope for managing as part of a reserve ecosystem or as a food source. Genetic considerations are important for game management. For example, chromosomal mutations could be deleterious in heterozygotes of outbred populations. Some variation in size and coloration of impala exists in South Africa. For example, impala from North West province are relatively large (G. Siddle, pers. comm.) and impala in Natal are relatively dark (G. Cooper, pers. comm.). This morphological variation indicates that there could be genetic differences that would make it inadvisable to translocate impala because of potential outbreeding depression. Analysis of larger samples from more subpopulations is needed to detect these differences, if they exist. Conversely, inbreeding and loss of variability can cause a loss of fitness in small populations. However, results of this study indicate that present levels of dispersal and translocation are probably sufficient to prevent significant levels of inbreeding and differentiation among populations.

Small populations are more susceptible to inbreeding and random genetic drift than are large populations. So small populations should have lower average heterozygosity than large populations. Therefore, a strong positive correlation is expected between population size and heterozygosity. The slightly negative, non-significant correlation found in this study seems to be due to higher than expected levels of inbreeding in Zimbabwe, which has a large population, and to higher than expected levels

of heterozygosity in Marloth Park and Pongola, which have small populations. Marloth Park impala can cross the Crocodile river into the large population in the Kruger National Park. The population in Pongola has received translocations from Mkuzi and is thus outbred. Data discussed above also reveal a loss of variability at Nylsvlei and Central Estates, Zimbabwe.

An hypothesis formulated in this study was that seasonal dispersion in temperate zones results in greater gene flow than in tropical zones, where breeding takes place for most of the year in territories that are less variable than those in temperate zones. This hypothesis could be tested with direct and indirect estimates of dispersal in areas where these quite different mating patterns occur. MtDNA restriction site variation among impala in East Africa already indicates a substantial amount of population subdivision (N. Georgiadis, pers. comm.). However, similar techniques should be used to compare variability in the different regions, as allozymes evolve much more slowly than does mtDNA.

CHAPTER 5

GENERAL CONCLUSIONS

- The products of 39 protein-encoding loci were resolved by starch gel electrophoresis. Six loci, *CK-C**, *GPI**, *MPI**, *PEP-B**, *PGM-2** and *PROT-2** were polymorphic and there were rare alleles at 12 loci.

- Interpretation of variability at two loci, *GPI** and *MPI** differed from previous studies.

- No clear pattern of apparent gene duplication events, tissue specific gene expression or ability to form heteropolymers was evident. More bovid species should therefore be surveyed for electrophoretic variability in more tissues.

- A slight Wahlund effect among genotypes, significant allele frequency differences among some localities, geographic trends in allele frequencies, a small proportion of between population gene diversity, small genetic distances and a significant yet small F_{ST} estimate revealed weak yet definite population structure.

- Impala in southern Africa can be grouped into three rough subpopulations: kwaZulu/Natal, Mpumalanga and Northern Province/Zimbabwe.

- High levels of gene flow account for the lack of strong population structure. Natural dispersal and translocation are sufficient to maintain genetic homogeneity in the southern African impala population.

- There is not a deficit of genetic variability in impala in southern Africa.

- Population heterozygosity was not correlated with population size, indicating that the populations studied have a variety of natural histories.

- Genetic variability and inbreeding coefficients indicated that most populations studied have not been adversely affected by reductions in numbers. Nylsvlei and Zimbabwe showed lower levels of variability than the other populations. Nylsvlei has a small population, so this may be cause for concern.

- Variability among populations revealed no differences large enough to warrant concern about outbreeding depression or mixing unique types.

- Distributions of single locus heterozygosities and allele frequencies indicated a departure from mutation-drift equilibrium.

- The number of rare alleles exceeded expectations. This is expected in a population that is recovering from a recent bottleneck.

REFERENCES

Ali, N. and S. Jameel. 1993. Rapid and efficient processing of tissue specimens for PCR using glass fibre filter paper. *Biotechniques* 15(1): 90-92.

Allard M.W., M.M. Miyamoto, L. Jarecki, F. Kraus and M.R. Tennant. 1992. DNA systematics and evolution of the Artiodactyl family Bovidae. *Proc. Natl. Acad. Sci. USA* 89: 3972-3976.

Baccus, R., N. Ryman, M.H. Smith, C. Reuterwall and D. Cameron. 1983. Genetic variability and differentiation of large grazing mammals. *J. Mammal.* 64: 109-120.

Bezy, R.L. and Sites, J.W., Jr. 1987. Preliminary study of allozyme evolution in the lizard family Xantusidae. *Herpetologica* 43: 280-292.

Bigalke R.C., G.B. Hartl, M.P.S. Berry and H.J. van Hensbergen. 1993. Population genetics of the springbok *Antidorcas marsupialis*, a preliminary study. *Acta Theriologica* 38, suppl.2: 103-111.

Brooks, P.M. 1975. The sexual structure of an impala population and its relationship to an intensive game removal programme. *The Lammergeyer* 22: 1-8.

Burnett, K.G. and M.R. Felder. 1978. Genetic regulation of liver alcohol dehydrogenase in *Peromyscus*. *Biochem. Genet.* 16: 443-454.

Buth, D.G., R.W. Murphy, M.M. Miyamoto and C.S. Lieb. 1985. Creatine kinases of amphibians and reptiles: evolutionary and systematic aspects of gene expression. *Copeia* 2: 279-284.

Cavalli-Sforza, L.L. and A.W.F. Edwards. 1967. Phylogenetic analysis: Models and estimation procedures. *Evolution* 21: 550-570.

Cepica, S. and A. Stratil. 1978. Further studies on sheep polymorphic erythrocyte diaphorase. *Anim. Blood Grps biochem. Genet.* 9:239-243

Chakraborty, R. 1980. Gene diversity analysis in nested subdivided populations. *Genetics* 96: 721-726.

Chakraborty, R., M. Haag, N. Ryman and G. Stèhl. 1982. Hierarchical gene diversity analysis and its application to brown trout population data. *Hereditas* 97: 17-21.

Cooper, D.W. 1968. The significance level in multiple tests made simultaneously. *Heredity* 23: 614-617.

Corbet S.W., W.S. Grant and T.J. Robinson. 1994. Genetic divergence in South African wildebeest: Analysis of allozyme variability. *J. Hered.* 85: 479-483.

Corbet, S.W. and T.J. Robinson. 1991. Genetic divergence in South African wildebeest: Comparative cytogenetics and analysis of mitochondrial DNA. *J. Hered.* 82: 447-452.

Crow, J.F. and M. Kimura. 1970. *An introduction to population genetics theory.* Harper and Row, London.

Darnall, D.W. and I.M. Klotz. 1972. Protein subunits: A table (Revised Edition). *Arch. Biochem. Biophys.* 149: 1-14.

DeLorenzo, R.J. and F.H. Ruddle. 1969. Genetic control of two electrophoretic variants of glucosephosphate isomerase in the mouse (*Mus musculus*) *Biochem. Genet.* 3: 151-162.

Emerson, B.C and M.L. Tate. 1993. Genetic analysis of evolutionary relationships among deer (subfamily Cervinae). *J. Hered.* 84: 266-273.

Estes, R.D. 1993. Impala, Chap 12 in *The safari companion: A guide to watching African mammals.* Russell Friedman, Halfway House, RSA, pp 127-132

Ferris, S.D. and G.S. Whitt. 1978. Genetic and molecular analysis of non-random dimer assembly of the creatine kinase isozymes of fishes. *Biochem. Genet.* 16: 811-829.

Fisher, S.E., J.B. Shaklee, S.D. Ferris and G.S. Whitt. 1980. Evolution of five multilocus isozyme systems in the chordates. *Genetica* 52/53: 73-85.

Fisher, S.E. and G.S. Whitt. 1978. Evolution of isozyme loci and their differential tissue expression--creatine kinase as a model system. *J. Mol. Evol.* 12: 25-55.

Frankel, O.H. and M.E. Soulé. 1981. *Conservation and Evolution.* Cambridge University Press.

Frick, L. 1983. An electrophoretic investigation of the cytosolic di- and tripeptidases of fish: Molecular weights, substrate specificities and tissue and phylogenetic distributions. *Biochem. Genet.* 21: 309-322.

Frick, L. 1984. Isolation and characterisation of five cytosolic di- and tripeptidases from the skipjack tuna (*Katsuwonis pelamis*). *Comp. Biochem. Physiol.* 77B: 533-540.

Fuerst, P.A., R. Chakraborty and M. Nei. 1977. Statistical studies of protein polymorphism in natural populations I: Distribution of single locus heterozygosity. *Genetics* 86: 455-483.

Gatesy, J., D. Yelon, R. DeSalle and E.S. Vrba. 1992. Phylogeny of the Bovidae (Artiodactyla, Mammalia), based on mitochondrial ribosomal DNA sequences. *Mol. Biol. Evol.* 9: 433-446.

Gentry, A.W. 1978. Bovidae. Chap 27, p540-572 *in* Maglio, V.J. and H.G.S. Cooke (eds) *Evolution of African Mammals.* Harvard University Press, Cambridge, Massachussetts.

Georgiadis, N.J., P.W. Kat, H. Oketch and J. Patton. 1990. Allozyme divergence within the Bovidae. *Evolution* 44: 2135-2149.

Goodman, P.S. 1990. *Soil, vegetation and large herbivore relations in Mkuzi game reserve, Natal.* PhD thesis, University of the Witwatersrand, Johannesburg.

Grant W.S. 1988 (1994 revision). *Protein Electrophoresis in Population Genetics.* University of the Witwatersrand, Johannesburg.

Grant, W.S. 1993. Population genetics in southern Africa: What are the questions? *S. Afr. J. Sci.* 89: 113-118.

Grant, W.S. and R.M. Little. 1992. How sedentary are Greywing Francolins (*Francolinus africanus*)? *Evolution* 46: 1477-1491.

Greenwood, P.J. 1980. Mating systems, philopatry and dispersal in birds and mammals. *Anim. Behav.* 28: 1140-1162.

Grobler, J.P. and F.H. van der Bank. 1993. Genetic diversity and differentiation of the three extant southern African species of the subfamily Hippotraginae (Family: Bovidae). *Biochem. Syst. Ecol.* 21: 591-596.

Grobler, J.P. and F.H. van der Bank. 1994a. Genetic heterogeneity in sable antelope (*Hippotragus niger* Harris 1838) from four southern African regions. *Biochem. Syst. Ecol.* 22: 781-789.

Grobler, J.P. and F.H. van der Bank. 1994b. Allozyme variation in South African impala populations under different management regimes. *S. Afr. J. Wildl. Res.* 24: 89-94.

Gyllensten, U., N. Ryman, C. Reuterwall and P. Dratch. 1983. Genetic differentiation in four European subspecies of red deer (*Cervus elaphus* L.). *Heredity* 51: 561-580.

Harris, H. 1980. *The Principles of Human Biochemical Genetics.* Elsevier, Holland.

Harris, H. and D.A. Hopkinson. 1976. *Handbook of Enzyme Electrophoresis in Human Genetics.* North-Holland Publishing Company, Amsterdam.

Hartl, G.B. and F. Reimoser. 1988. Biochemical variation in roe deer (*Capreolus capreolus* L.): are r-strategists among deer genetically less variable than K-strategists? *Heredity* 60: 221-227.

Henderson, N.S. 1965. Isozymes of isocitrate dehydrogenase: Subunit structure and intracellular location. *J. Exp. Zool.* 158: 263-274.

Hillis, D.M. 1984. Misuse and modification of Nei's genetic distance. *Syst. Zool.* 33: 238-240.

Hillis, D.M and C. Moritz (eds). 1990. *Molecular Systematics.* Sinauer Associates, Sunderland, Massachussetts. pp 373-502.

Hubbs, C.L. and C. Hubbs. 1953. An improved graphical analysis and comparison of series of samples. *Systematic Zoology* 8: 50-55.

Jarman, P.J. 1974. The social organisation of antelope in relation to their ecology. *Behaviour* 48: 215-267.

Kettler, M.K., and G.S. Whitt. 1986. An apparent progressive and recurrent evolutionary restriction in tissue expression of a gene, the lactate dehydrogenase-C gene, within a family of bony fish (Salmoniformes: Umbridae). *J. Mol. Evol.* 23: 95-107.

Kimura, M. and J.F. Crow. 1964. The number of alleles that can be maintained in a finite population. *Genetics* 49: 725-738.

Laurie-Ahlberg, C.C. 1982. Genetic, ontogenic and tissue-specific variation of dipeptidases in *Drosophila melanogaster*. *Biochem. Genet.* 20: 407-424.

Levene, H. 1949. On a matching problem arising in genetics. *Ann. Math. Statist.* 20: 91-94.

Li, W.-H. and M. Nei. 1975. Drift variances of heterozygosity and genetic distance in transient states. *Genet. Res.* 25: 229-248.

Litt, M. and J.A. Luty 1989. A hypervariable microsatellite revealed by in vitro amplification of a dinuleotide repeat within the cardiac muscle actin gene. *Am. J. Hum. Genet.* 44: 397-401.

Lorenzini, R., M. Patalano, M. Apollonio and V. Mazzarone. 1993. Genetic variability of roe deer *Capreolus capreolus* in Italy: electrophoretic survey on populations of different origin. *Acta Theriol.* 38 Suppl. 2: 141-151.

Lowenstein J.M. 1986. Bovid relations based on serum immunology. *S. Afr. J. Sci.* 82: 77-78.

Markert, C.L and I. Faulhaber. 1965. Lactate dehydrogenase isozyme patterns of fish. *J. Exp. Zool.* 159: 319-332

Maruyama, T. and P.A. Fuerst. 1984. Population bottlenecks and non-equilibrium models in populations genetics. I. Allele numbers when populations evolve from zero variability. *Genetics* 108: 745-763.

Murphy, R.W., J.W. Sites, D.G. Buth and C.H. Haufler. 1990. Proteins I: Isozyme electrophoresis. Chap 10 *in* D.M. Hillis and C. Moritz (eds), *Molecular Systematics*. Sinauer Associates, Sunderland, MA., pp 45-103.

Murray, M.G. 1981. Structure of association in impala, *Aepyceros melampus*. *Behav. Ecol. Sociobiol.* 9: 23-33.

Murray, M.G. 1982. Home range, dispersal and the clan system of impala. *Afr. J. Ecol.* 20: 253-269.

Nei, M. 1972. Genetic Distance Between Populations. *Amer. Nat.* 106:283-292.

Nei, M. 1973. Analysis of gene diversity in subdivided populations. *Proc. Nat. Acad. Sci. USA* 70: 3321-3323.

Nei, M. 1978. Estimation of average heterozygosity and genetic distance from a small number of individuals. *Genetics* 89: 583-590.

Nei, M. 1987. *Molecular Population Genetics and Evolution.* North Holland, Amsterdam, ppi-288.

Nei, M., P.A. Fuerst and R. Chakraborty. 1978. Subunit molecular weight and genetic variability of proteins in natural populations. *Proc. Natl. Acad. Sci. USA.* 75: 3359-3362.

Nei, M., T. Maruyama and R. Chakraborty. 1975. The bottleneck effect and genetic variability in populations. *Evolution* 29: 1-10.

Nei, M. and A.K. Roychoudhury. 1974. Sampling variances of heterozygosity and genetic distance. *Genetics* 76: 379-390.

Nevo, E. 1978. Genetic variation in natural populations: patterns and theory. *Ther. Popl. Biol.* 13: 121-177.

Op't Hof, J., U. Wolf and W. Krone. 1969. Studies on isozymes of sorbitol dehydrogenase in some vertebrate species. *Humangenetik* 8: 178-182.

Osterhoff, D.R., I.S. Ward-Cox and E. Young 1972. Biochemical polymorphism in the South African impala (*Aepyceros melampus*). *Koedoe* 15: 55-59.

Patton, J.L. and J.H. Feder. 1981. Microspatial genetic heterogeneity in pocket gophers: Non-random breeding and drift. *Evolution* 35: 912-920.

Pemberton, J.M. and R.H. Smith. 1985. Lack of biochemical polymorphism in British fallow deer. *Heredity* 55: 199–207.

Peters, J. 1982. Nonspecific esterases of *Mus musculus*. *Biochemical Genetics* 20: 585–606.

Peters, J. and H.R. Nash. 1978. Esterases of *Mus musculus*: substrate and inhibition characteristics, new isozymes and homologies with man. *Biochemical Genetics* 16: 553–569.

Pryor, A.J. 1974. Allelic glutamic dehydrogenase isozymes in maize—a single hybrid isozyme in heterozygotes? *Heredity* 32: 397–419.

Rice, W.R. 1989. Analyzing tables of statistical tests. *Evolution* 43: 223–225.

Ridgway, G.J, S.W. Sharburne and R.D. Lewis. 1970. Polymorphism in the esterase of Atlantic herring. *Trans. Amer. Fish. Soc.* 99: 147–151

Rigaut, J-P. and M-T. Chalumeau. 1984. Pyruvate kinase isozyme patterns of fish. *Comp. Biochem. Physiol.* 77B: 451–458.

Roed, K.H. 1985. Genetic variability in Norwegian semi-domestic reindeer (*Rangifer tarandus* L.). *Hereditas* 102: 177–184.

Rogers, J.S. 1972. Measures of genetic similarity and genetic distance. Studies in genetics. VII. *Univ. Texas Publ.* 7213: 145–153.

Ryman, N., C. Reuterwall, K. Nygrén and T. Nygrén. 1980. Genetic variation and differentiation in Scandinavian moose (*Alces alces*): are large mammals monomorphic? *Evolution* 34: 1037–1049.

Saitou, N. and M. Nei. 1987. The neighbour-joining method: A new method for reconstructing phylogenetic trees. *Mol. Biol. Evol.* 4: 406–425.

Schenkel, R. 1966. On sociology and behaviour in impala (*Aepyceros melampus* Lichtenstein). *E. Afr. Wildl. J.* 4: 99–114.

Shaklee, J.B., F.W. Allendorf, D.C. Morizot and G.S. Whitt. 1989. Genetic nomenclature for protein-coding loci in fish: proposed guidelines. *Trans. Amer. Fish Soc.* 118: 218–227.

Shaklee, J.B., K.L. Kepes and J.S. Whitt. 1973. Specialised lactate dehydrogenase isozymes: The molecular and genetic basis for the unique eye and liver LDHs of teleost fishes. *J. Exp. Zool.* 185: 217–240.

Sheffield, S.R., R.P. Morgan II, G.A. Feldhamer and D.M. Harman. 1985. Genetic variation in white-tailed deer (*Odocoileus virginianus*) populations in western Maryland. *J. Mamm.* 66: 243–255.

Shows, T.B., and 24 co-authors. 1987. Guidelines for human gene nomenclature. An international system for human gene nomenclature (ISGN, 1987). *Cytogenet. Cell Genet.* 46: 11–28.

Skinner, J.D. 1993. Springbok (*Antidorcas marsupialis*) treks. *Trans. Roy. Soc. S. Afr.* 48: 291–305.

Slatkin, M. 1981. Estimating levels of gene flow in natural populations. *Genetics* 99:323–335.

Slatkin, M. 1985. Gene flow in natural populations. *Ann. Rev. Ecol. Syst.* 16: 393–430.

Slatkin, M. 1987. Gene flow and the geographic structure of natural populations. *Science* 236: 787–792.

Slatkin, M. and N.H. Barton. 1989. A comparison of three indirect methods for estimating average levels of gene flow. *Evolution* 43: 1349–1368.

Smithers, R.H.N. 1983. *Mammals of the Southern African Subregion.* University of Pretoria, Pretoria, pp 627-631 and 647-651.

Sneath, P.H.A. and R.R. Sokal. 1973. *Numerical Taxonomy.* W.H. Freeman, San Francisco.

Sokal, R.R. and F.J. Rohlf. 1969. *Biometry.* W.H. Freeman and Co., San Francisco.

Sokal, R.R. and F.J. Rohlf. 1995. *Biometry.* W.H. Freeman and Co., San Francisco.

Soulé, M.E. and B.A. Wilcox (eds.). 1980. *Conservation Biology - an Evolutionary Ecological Perspective.* Sinauer Associates, Stamford.

Stewart, F.M. 1976. Variability in the amount of heterozygosity maintained by neutral mutations. *Theoret. Pop. Biol.* 9: 188-201.

Ströhlein, H., S. Herzog, W. Hecht and A. Herzog. 1993. Biochemical genetic description of German and Swiss populations of red deer *Cervus elaphus.* *Acta Theriol.* 38, Suppl. 2: 153-161

Swofford, D.L. and G.J. Olsen. 1990. Phylogeny reconstruction, Chap 11 *in* D.M. Hillis and C. Moritz (eds), *Molecular Systematics.* Sinauer Associates, Sunderland, MA., pp 411-501.

Tabachnick, W.J. and D.J. Howard. 1982. Genetic control of hexokinase variation in insects. *Biochem. Genet.* 20: 47-57.

Tyson, P.D. 1986. *Climatic change and variability in southern Africa.* Oxford University Press, Cape Town.

Vrba, E.S. 1979. Phylogenetic analysis and classification of fossil and recent Alcelaphini Mammalia: Bovidae. *Biol. J. Linn. Soc.* 11: 207-228.

Vrba, E.S. 1984. Evolutionary pattern and process in the sister group Alcelaphini-Aepycerotini (Mammalia: Bovidae), *in* N. Eldredge and S.M. Stanley (eds), *Living fossils*. Springer Verlag, New York, pp 62-79.

Vrba, E.S. 1985a. Environment and evolution: Alternative causes of the temporal causes of evolutionary events. *S. Afr. J. Sci.* 81: 229-236.

Vrba, E.S. 1985b. African Bovidae: Evolutionary events since the miocene. *S. Afr. J. Sci.* 81: 263-266.

Vrijenhoek, R.C, M.E. Douglass and G.K. Meffe. 1985. Conservation genetics of endangered populations in Arizona. *Science* 229: 400-402.

Ward, R.D., B.J. McAndrew and G.P. Wallace. 1979. Purine nucleoside phophorylase variation in the brook lamprey, *Lampetra planeri* (Bloch) (Pteromyzone, Agnatha): Evidence for a trimeric enzyme structure. *Biochem. Genet.* 17: 251-256.

Watterson, G.A. 1974. Models for the logarithmic species abundance distributions. *Theoret. Pop. Biol.* 6: 217-250.

Watterson, G.A. 1984. Allele frequencies after a bottleneck. *Theor. Pop. Biol.* 26: 387-407.

Weber, J.L. and P.E. May 1989. Abundant class of human DNA polymorphisms which can be typed using the polymerase chain reaction. *Am. J. Hum. Genet.* 44: 388-396.

Weir, B.S 1990. Intraspecific differentiation, Chap 10 *in* D.M. Hillis and C. Moritz (eds), *Molecular Systematics*. Sinauer Associates, Sunderland, MA., pp 373-410.

Weir, B.S. and C.C. Cockerham. 1984. Estimating F-statistics for the analysis of population structure. *Evolution* 38: 1358-1370.

Westneat, D.F., P.C. Frederick and R.H. Wiley. 1987. The use of genetic markers to estimate the frequency of successful alternative reproductive tactics. *Behav. Ecol. Sociobiol.* 21: 35–45.

Whitt, G.S. 1970. Developmental genetics of the lactate dehydrogenase enzymes of fish. *J. Exp. Zool.* 175: 1–35.

Wooten, M.C. and M.H. Smith. 1985. Large mammals are genetically less variable? *Evolution* 39: 212–215.

Wright, S. 1943. Isolation by distance. *Genetics* 28: 114–138.

Wright, S. 1951. The genetical structure of populations. *Ann. Eugen.* 15: 323–354.

Wright, S. 1978. Variability within and among natural populations. Vol. 4 in *Evolution and the genetics of populations.* University of Chicago Press, Chicago, pp 90–92.

Wyban, J.A. 1982. Soluble peptidase isozymes of the Japanese Medaka (*Oryzias latipes*): Tissue distributions and substrate specificities. *Biochem. Genet.* 20: 849–858.

Zouros, E. 1976. Hybrid molecules and the superiority of the heterozygote. *Nature* 262: 227–229.