CHAPTER 3

MICROSATELLITE DIVERSITY

Chapter 3: Microsatellite Genetic Diversity and Apportionment

Introduction

Classes of simple repetitive DNA, called microsatellites, are the most powerful molecular markers used to answer both evolutionary genetic and behavioral queries in population studies today (Goldstein and Pollock 1998). Microsatellites are highly variable sequences composed of tandem repeat motifs made up of di, tri, tetra, or penta nucleotides (Moore 1991). These sequences are found scattered throughout eukaryotic genomes as pure repeat units, compound (two or more repeat motifs), or interrupted repeat motifs. The inherent properties of microsatellite loci provide for a rich set of data in application to population studies. Firstly, they have an extremely high mutation rate in the order of (10^{-2} –5 x 10^{-6}) that results in a high number of alleles for a locus found in one population (Dallas et al. 1992, Weber and Wong 1993, Goldstein et al. 1995a). Microsatellites are co-dominant nuclear markers that occur in all chromosomal regions. Multi-locus genotypes are easily scored using PCR and genotyping analyses. Finally, primers developed from one species are usually able to amplify homologous loci in related taxa (Bruford and Wayne 1993). The conservation of the sequences flanking repeat loci allows for amplification of homologous regions among highly divergent species of ruminants (Pepin et al. 1995), felids (Menotti-Raymond and O'Brien 1995), canids (Gottelli et al. 1994) and marine turtles (FitzSimmons et al. 1995).

Genetic studies in a diverse array of species have revealed patterns of evolution of microsatellite loci. Allelic variability at a locus involves changes in the number of repeat motifs (Hamada 1982, Litt and Luty 1989, Tautz 1989, Weber and May 1989). Therefore, it was first assumed that microsatellites mutate by a stepwise mutation model (SMM), whereby an allele is formed by loss or gain of a repeat unit (Ohta and Kimura 1973).

However, additional mutations have been found within microsatellite repeat motifs, creating a departure from this model (Garza et al. 1995, Estoup al 1995). In these cases, alleles mutate by insertion-deletions, or base substitutions under the infinite alleles model (IAM) making the repeat imperfect (Kimura and Crow 1964). The two-phase model (TPM) predicts that mutations tend to be larger jumps between repeat units (Di Rienzo et al. 1994). Research has also revealed a mutational bias in loci of different repeat length. For instance, small repeat loci tend to mutate slower and towards a higher repeat number, while large repeat loci are prone to mutate faster towards a smaller repeat. Mutation may be in a single step (following the strict mutation model) or multiple steps. In many cases alleles are shortened by two repeat units and lengthened by only one. Allele size is constrained by an upper and lower boundary (Garza et al. 1995, Nauta and Weissing 1996). These boundaries restrict variation at a locus. This constraint is either set by the mutational process or by natural selection (Bowcock et al. 1994). No model yet described can explain mutations at all loci.

Microsatellite markers were first successfully used to establish relatedness, measure genetic diversity and assess population substructure (see Bruford and Wayne 1993, Goldstein and Pollock 1997 for review). Microsatellites are now being exploited further for assessing: social structure (Morin et al. 1994), male mating success (Coltman et al. 1999), neonatal fitness (Coltman et al. 1998, Coulson et al. 1998), hybridization (Reich et al. 1999), bottleneck events (Luikart et al. 1998a,b), demographic history (Goldstein et al. 1999) and evolutionary relationships (Grant et al. 1999). Furthermore, the combination of microsatellite analysis and non-invasive sampling has opened up a broader spectrum of inquiry into the behaviors of free-ranging mammals. DNA analysis of dung samples has been used for estimating population size in coyotes (Kohn et al. 1999), tracking movements of brown bear (Taberlet et al. 1997), revealing infanticide

behavior in langur monkeys (Borries et al. 1999), as well as determining dispersal patterns and paternity success in bonobos (Gerloff et al. 1999).

Microsatellite data has been used to investigate hybridization events between taxa of several mammalian species. These markers have been used to detect hybridization between species such as Ethiopian wolf and domestic dog (Gottelli et al. 1994), grey wolf and coyote (Roy et al. 1996), taurine and Zebu cattle (MacHugh 1997), and sika and red deer (Goodman et al. 1999). Microsatellites have also been used to identify introgression below the species level in populations of the great cormorant (Goostrey et al. 1998) and subspecies of tiger (Wentzel et al. 1999). Assignment of individuals into a population is based upon three different methods of classification: the multi-locus genotype, genetic distance estimate and maximum-likelihood score.

A proportion of alleles can be shared between taxa while other alleles can be unique (private) to each taxon. Private alleles or allele frequency differences can be used to detect hybridization in admixture studies (Nason and Ellstrand 1993). In this case, the multi-locus genotype of the individual is used to assign it to as specific category (ie. hybrid or pure taxa). The disadvantage to this method is that the error rate of classifying individuals after the F_1 generation will increase even if all loci are diagnostic (Epifanio and Phillip1997).

Phylogenetic methods have been used to group related individuals into categories according to distance estimates calculated from microsatellite data. This approach assumes that individuals from the same population or origin will have similar genotypes and cluster together. Individuals who share common alleles or display small genetic distances are joined together by a tree-joining algorithm (Bowcock et al. 1994, MacHugh et al. 1996, Blott et al. 1999).

Likelihood-based methods use genotypes to assign an individual to a population with the highest likelihood over all other populations (Shriver et al. 1997). Either a frequency method or Baysian approach is employed in the assignment test. The frequency-based method uses the computed allele frequencies in each population and likelihood estimates of each individual multilocus genotype. In effect, the likelihood of a multilocus genotype existing in a population is the product of all likelihoods for each locus (Cornuet et al. 1999). The Baysian approach is similar to the frequency method however; it incorporates a probability distribution of allele frequencies when the distribution is unknown (in the case of rare, or missing alleles).

Although the distance and likelihood methods both utilize observed allele frequencies of each population, the actual assignments are tested by different probability estimates. Simulation models have predicted that in order for an individual to be correctly classified, at least 10 polymorphic loci with heterozygosity values of at least ($H = 0.6$) and Fst estimates of 0.1 are needed (Cornuet et al. 1999). An accurate estimation of admixture is highly dependent upon the allelic distribution of the loci as well as power of the statistical assignment test. Miller (2000) has recommended that a high number of loci analyzed under a maximum-likelihood method will greatly increase the probability of correct assignment. The major limitation to the multi-locus genotype assignment test is that it only reflects the observed alleles found in the study, and not all possible alleles in the populations. Simulation tests are able to overcome this disadvantage by generating allele frequencies of missing alleles within populations.

Microsatellite analysis allows for a sensitive prediction of demographic expansions or contractions within populations (Goldstein et al. 1999). In this effect, the variances of allele repeat score across microsatellite loci will be reduced in a population under expansion or become inflated under equilibrium. Bottleneck events can be

determined by examination of frequency distributions. The loss of low frequency alleles is indicative of populations that have experienced a large demographic collapse (Nei 1975).

This microsatellite study was carried out in order to explore nuclear diversity within *D. pygargus* and to provide a finer resolution to the examination of genome-wide variation. The allelic distributions of these makers were applied to detect genetic sub-structuring as well as individual classification. Further exploration of microsatellite data was used to infer demographic changes within each subspecies.

MATERIALS AND METHODS

Bovine and Ovine Markers

A panel of bovine primers was tested for amplification in *D. pygargus* (Table 6). A total of 8 microsatellite loci were chosen for the study based upon reliable amplification and high heterozygosity values measured in cattle (*Bos taurus*) and sheep (*Ovis ovis*). All microsatellite loci were assumed to mutate neutrally with the exception of MDRB3, which is found within intron 2 of the MHC-DRB locus. The mutation of this microsatellite locus is influenced by the evolution of the coding DRB gene in ruminant species such as cattle (Ammer et al. 1992, Ellegren et al. 1993, van Haeringen et al. 1999), bighorn sheep (Patterson 1998) and domestic sheep (Schwaiger et al. 1993).

PCR Conditions

The forward primer for each locus was labeled at the 5' end with a fluorescein tag of the HEX (yellow), TET (green) or 6-FAM (blue) molecule (Perkin-Elmer). The PCR

Table 6. Microsatellite primers tested.

Marker	Source Species	Amplification
TGLA142	Cattle	*
TGLA53	Cattle	*
DINC	Cattle	*
CSSM18	Cattle	A, 2 alleles
INRA26	Cattle	A, 2 alleles
BM3215	Cattle	A, 1 alleles
MDRB3	Cattle	*
RBP3	Cattle	N
MAF46	Sheep	*
MAF50	Sheep	*
OarCP26	Sheep	*
OarCP64	Sheep	*
OarFCB304	Sheep	N
AGLA269	Cattle	N
AGLA218	Cattle	N
TGLA48	Cattle	A, 2 alleles
TGLA263	Cattle	N
TGLA57	Cattle	A, 2 alleles
TGLA73	Cattle	A, 2 alleles
TGLA227	Cattle	N
TGLA126	Cattle	A, 1 alleles
TGLA122	Cattle	N
MGTG4	Cattle	A, wrong size
MGTG7	Cattle	N

* = chosen for study, A = amplification of only 1 or 2 alleles

N = no amplification

reaction conditions are as follows: PCR amplification of individual microsatellite loci was performed in 15 µl reactions on a PEC 2400 thermocycler. Each reaction consisted of 14 pmol of each primer, 10X PCR buffer (10 mM Tris-hydrochloric acid [pH 8.3], 50 mM potassium chloride), 250 µM of each deoxyribonucleoside 5'triphosphates (dATP, dCTP, dGTP, dTTP) (Promega), 0.4 Unit Amplitaq DNA polymerase (Promega), and 100 ng of DNA template. The forward primer of each primer pair was labeled with a fluorescent dye phosphoramite. The PCR conditions are as follows: initial denaturation step at 94° C for 3 min; ten cycles of 94° C for 15 sec, 48 - 52° C for 15 sec, 72° C for 30 sec; followed by 20 cycles of 89° C for 15 sec, 48 - 52° C for 15 sec, 72° C for 30 sec, and terminating in a 10 min extension step at 72° C.

Genescan Electrophoresis and Analysis

Multiplexing of PCR products is made possible by staggering microsatellite loci having different allele size ranges and also by labeling products with the selection of tag colors. For each gel, microsatellite products generated from 4-5 different loci were electrophoresed for each individual. The gel lanes can accommodate either 36 or 64 samples; therefore several gels were run for each panel of microsatellites. In order to establish consistency across multiple plates, PCR products from 2-4 animals from each gel were run as controls across secondary and tertiary gels.

PCR products for each animal were pooled and diluted in a volume of approximately 50 µl ddH_2O. Due to the varying strength in fluorescent signal of each dye, different amounts of product were added to the ddH_20 dilution. Only 1-2 µl of product labeled with 6-FAM was added to the dilution, while 3-4 µl of the TET labeled product and 7-8 µl of the HEX products. After the PCR products were pooled, 1 µl of the

dilution was mixed with 1.5 µl of loading buffer cocktail that consists of: 2.0 µl de-ionized formamide, 1.0 µl ABI PRISM™ Genescan-350 TAMRA internal lane standard and 0.5 µl of ABI Genescan loading dye. The TAMRA lane standard is made up of a ladder of DNA fragments (350, 300, 250, 200, 160, 150, 139, 100, 75, 50), which is used in estimating PCR fragment sizes. The mixture is then denatured at 94° C for 3 min then snap-cooled on ice. The denatured products (2.0 µl) were loaded on a 6% denaturing polyacrylamide-sequencing gel housed in an automated ABI 377 Sequencer. Gel electrophoresis was run through a 1X TBE buffer solution at 2000 volts, 400 mA, and 25 W for 2.5 hours. The raw data for each sample was stored as a collection file that was created by the Gene Scan Collection software (version 1.2.2-1). The collection file was subsequently analyzed using the Genescan software package (ABI). Allele sizes were estimated using the Genotyper software package (version 1.1). The fragment lengths were called according to the Local Southern method (Elder and Southern 1987) to generate a best- fit curve from the internal lane size standards.

Descriptive Statistical Analyses

Genetic diversity found in the species and each subspecies was estimated as the mean number of alleles per locus (A) and heterozygosity levels both observed (H_O) and expected (H_E). The observed level of heterozygosity was compared to the expected level then analyzed by the Markov chain method (Guo and Thompson 1992) for departure from Hardy-Weinberg equilibrium using the Arlequin (v. 2.000) software package (Schneider et al. 2000).

The partitioning of genetic diversity into the two subspecies was measured by two F statistic measures. The *Fst* (Weir and Cockerham 1984) statistic can detect substructure in a population using loci that mutate under the infinite allele model. Since most

microsatellite loci are assumed to mutate under the step-wise model, an analogous statistic Rst, has been recommended to take the place of *Fst* in the measurement of substructuring (Slatkin 1995). Estimations of Rst values were carried out using the RSTCALC program (Goodman 1997).

Clustering Analysis

Genetic distances were calculated from allele frequency data using the MICROSAT (version 1.5) created by Minch et al. (1995). Two distance estimators, (Ps) proportion of alleles shared and (Dkf) kinship co-efficient (Bowcock et al. 1994) were chosen to estimate all pair-wise individual distances. These distance measures were chosen on the basis of their accuracy for estimating distances at population and individual levels (Takezaki and Nei 1996, Goldstein and Pollock 1997, Goldstein et al. 1999).

1. The proportion of shared alleles as distance is represented by:

$$(D_{ps}') - 1\text{-ps}$$

where ps is defined as:

$$ps = (\sum \mathbf{MIN}\ \{P[A(i)], P[B(i)]\})/n$$

for i = 1, 2, 3, ...n alleles where n is the total number of alleles for all loci.

2. Kinship co-efficient distance:

$$D_{kf} = 1\text{-kf}$$

$$kf = (\sum P[A(i)]P[B(i)]$$

where P[A(i)] is the relative frequency of allele i in taxon A for i= 1,2, 3, ...n alleles. The Kf value relates the probability that two genes taken at random from the same locus in two individuals are identical by descent.

The computed distance matrices were used in a cluster analysis employing two tree-building methods. The first method of phylogenetic reconstruction, neighbor-joining (NJ) is a method of joining the two most similar taxa (Saitou and Nei 1987). This method is best used when evolutionary rates vary among taxa. The second method unweighted pair group method with arithmetic mean (UPGMA) is normally applied when rates of evolution are assumed to be constant (Sneath and Sokal 1973, Nei 1987).

Bootstrap re-sampling was used in order to gain a confidence estimate of the tree topologies generated by all distance measures and tree building methods. The bootstrap test involves re-sampling of the data set by drawing points of data with replacement. A confidence level is given as a percentage out of 500 iterations.

Class Assignment of Individuals

Multi-locus genotype data for each individual was used to create a self-classification data set within the GeneClass program (Cornuet et al. 1998). Self-classification was carried by using two methods: likelihood estimates (frequency, Baysian) and distance measures (DA) average distance (Nei et al. 1983), (DAS) allele shared (Chakaborty and Jin 1983), (Dm) minimum (Nei 1972), (Ds) standard (Nei 1978), (DM) delta mu^2 (Goldstein et al. 1995), chord (Cavalli-Sforza and Edwards 1967). Variations of each assignment method were computed directly from the multilocus genotypes or simulated. Simulations were performed for each method incorporating 1000 simulated individuals per population with frequencies of (0.01) for each locus. After each

individual was assigned to a class, the self-classification data set was used as reference set for assigning the test individuals to an appropriate class.

Loss of genetic diversity

The BOTTLENECK program (Luikart and Cornuet 1998) was used to detect recent reductions in the effective population sizes (*Ne*) of both bontebok and blesbok. This method uses allele frequency estimates in order to detect loss of alleles through genetic drift or founder effect. The principle of BOTTLENECK operates under the assumption that if a population has had a constant *Ne* (under mutation-drift equilibrium) then there will be an equal probability of finding an excess or deficit in heterozygosity at a locus. In the case of reduction of *Ne*, the observed heterozygosity (H_{OBS}) will exceed that of the expected heterozygosity (H_{EX}). The heterozygosity values are used in three statistical tests (sign test, standardized differences test and a Wilcoxon sign-rank test) in order to detect heterozygosity excess (Cornuet and Luikart 1996, Luikart et al. 1997a). A reduction in the effective population size will also cause a faster loss of alleles at polymorphic loci. Therefore, the program calculates the distribution of allelic frequencies in order to detect a mode shift (indicating a bottleneck) from the expected L-shaped distribution (stable population) of the allelic range (Luikart et al. 1998b).

Results

A panel of cattle and sheep derived primers were tested for amplification in blesbok and bontebok antelope (see Table 6). A subset of these markers failed to amplify in the antelope, which may be due to mutations in the flanking regions of the microsatellite loci. The remaining set of primers was rejected from the study due to lack

individual was assigned to a class, the self-classification data set was used as reference set for assigning the test individuals to an appropriate class.

Loss of genetic diversity

The BOTTLENECK program (Luikart and Cornuet 1998) was used to detect recent reductions in the effective population sizes (*Ne*) of both bontebok and blesbok. This method uses allele frequency estimates in order to detect loss of alleles through genetic drift or founder effect. The principle of BOTTLENECK operates under the assumption that if a population has had a constant *Ne* (under mutation-drift equilibrium) then there will be an equal probability of finding an excess or deficit in heterozygosity at a locus. In the case of reduction of *Ne*, the observed heterozygosity (H_{OBS}) will exceed that of the expected heterozygosity (H_{EX}). The heterozygosity values are used in three statistical tests (sign test, standardized differences test and a Wilcoxon sign-rank test) in order to detect heterozygosity excess (Cornuet and Luikart 1996, Luikart et al. 1997a). A reduction in the effective population size will also cause a faster loss of alleles at polymorphic loci. Therefore, the program calculates the distribution of allelic frequencies in order to detect a mode shift (indicating a bottleneck) from the expected L-shaped distribution (stable population) of the allelic range (Luikart et al. 1998b).

Results

A panel of cattle and sheep derived primers were tested for amplification in blesbok and bontebok antelope (see Table 6). A subset of these markers failed to amplify in the antelope, which may be due to mutations in the flanking regions of the microsatellite loci. The remaining set of primers was rejected from the study due to lack

of polymorphism in *D. pygargus*. Therefore, out of the 25 bovine and ovine markers tested, eight loci were chosen for this population study. The loci ranged in number of alleles from 3 – 8 alleles per locus. Each locus was typed for 34 bontebok, 42 blesbok, and 9 unknown (possible hybrid) animals. The multi-locus composite genotypes are presented in Table 7. Allele frequency distributions were computed for each locus found within each subspecies (Figure 13). The eight loci together produced 41 allelic states, of which 66% were found to be specific to one subspecies (44% blesbok, 22% bontebok) and the other 34% shared between the two subspecies (Table 8).

Table 7.a. Microsatellite alleles found within blesbok

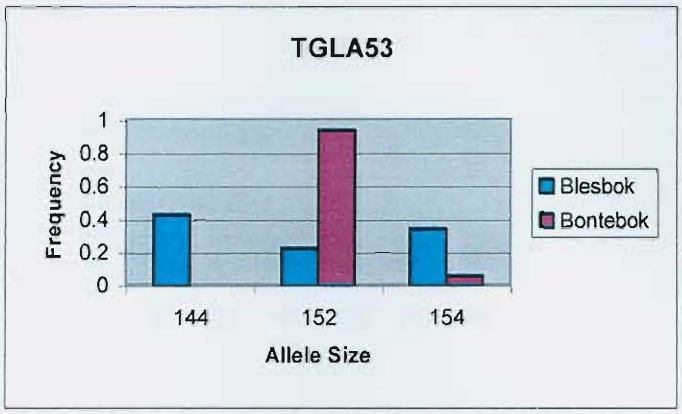
	TGLA53	TGLA142	DINC	MDRB3	OarHH64	OarCP26	MAF46	MAF50
BL1	152152	157161	106108	174174	108112	122122	103103	150156
BL3	144154	157161	106106	000000	112112	122124	095095	150150
BL4	144152	157159	106108	174174	108110	122124	097103	150150
JP1	144154	157159	106106	174176	108108	122124	097103	156156
JP2	144154	151153	106106	174174	108112	124124	097103	156156
JP3	144144	159161	106106	174176	108108	122124	097103	156160
JP4	154154	157161	106108	176176	108108	122122	097097	152160
JP5	144154	161161	106108	174174	108112	122124	097097	156160
JP6	152154	157159	108108	176176	108112	122124	097103	156160
JP7	144144	161161	106106	000000	108112	122122	097103	156156
JP8	144144	157161	106108	174174	108108	122124	097103	156160
JP9	144154	157161	106106	176176	108112	122124	097103	152156
JP10	154154	159161	106108	175176	116116	130130	107107	156156
JP11	144154	157159	106108	174174	108108	122124	103103	156156
JP13	144154	157159	106108	174174	108108	122124	103103	156156
JP14	144144	159161	106106	174174	108112	122124	097103	156156
JP15	144154	157161	106108	174174	108108	122122	097103	156160
JP16	154154	157159	106108	176176	108108	122122	097103	156156
JP17	154154	157161	106108	174176	108108	122122	097103	152156
JP18	154154	157159	106108	174176	108112	122122	097103	156156
JP19	154154	159161	108108	174174	108112	122124	097097	156160
JP20	144154	157159	106106	176176	108108	122122	097097	152160
JP21	144144	159159	106106	174174	108112	122122	097097	152156
JT2	152154	159161	106106	000000	112112	122124	103103	150158
JT3	152154	157161	106106	000000	112112	122124	095103	152154
JT4	144154	157161	106106	174174	112112	122122	095095	156160
JT5	144152	159161	108108	174174	112112	122122	095097	150150
JT6	153154	157159	106106	182182	108112	124124	095097	150158
DB2	144144	157157	106108	175175	108112	122124	095097	158160
DB3	144144	157159	106106	175175	112112	122124	095097	158160
DB4	144144	157157	106108	179179	108110	122124	095097	158160
DB5	144144	157159	106106	175175	112112	122122	097103	158160
SR1	152152	157157	106106	175175	112112	122122	095103	156160
SR2	144144	157157	108108	000000	108112	122122	095097	158158
SW1	144152	161161	106108	170175	112112	122122	095097	152158
SW2	152154	161161	106108	000000	108112	122122	093093	156156
GG26	144154	161161	106108	175175	112112	122124	097097	152160
GG30	144152	159159	106108	175175	108112	122122	095097	152156
GG31	144154	153153	108108	186186	112112	122124	095103	158160
CR1	152152	159159	106106	174174	112112	126128	095097	152158
CR2	152152	159159	106106	174174	112112	126128	097103	150158
CR3	152152	159159	106106	174174	112112	124126	095103	150158

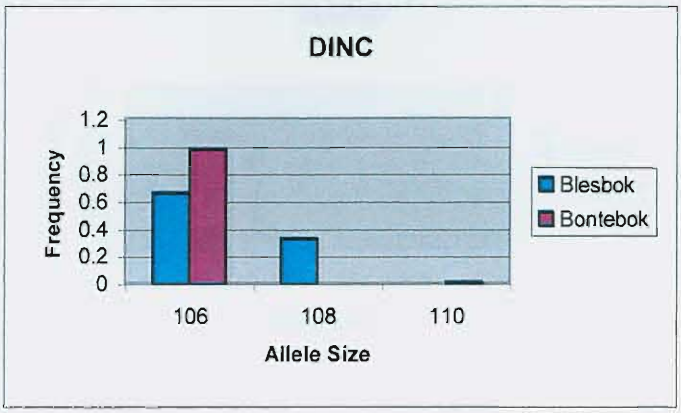
Table. 7.b. Microsatellite alleles found within bontebok

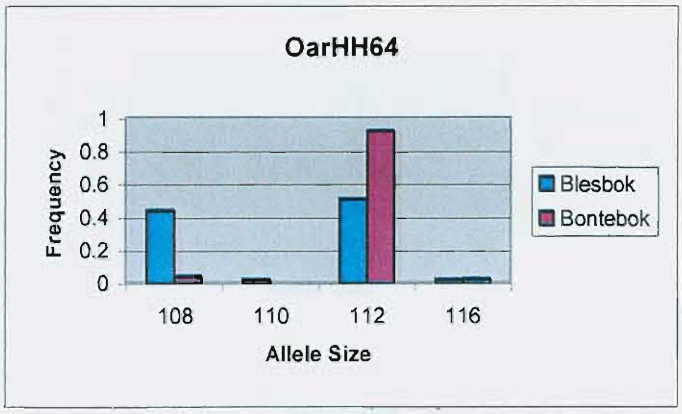
	TGLA53	TGLA142	DINC	MDRB3	OarHH64	OarCP26	MAF46	MAF50
BB1	152152	159159	106106	000000	108112	128128	095095	158158
BB2	152152	159159	106106	186186	112112	126128	095095	158158
BB3	152152	159159	106106	186186	108112	128128	095095	158158
BB4	152152	159159	106106	186186	112112	126128	095095	158158
BB5	152152	159159	106106	175186	112112	128128	095095	158158
BB6	152152	159159	106106	175175	112112	126128	095095	158158
BB7	152154	159159	106106	175186	112112	126128	095095	158158
BB8	152154	159159	106106	175175	112112	128128	095095	158158
BB9	152152	159159	106106	186186	116116	128128	095095	158158
BB10	152152	159159	106106	000000	112112	126128	095103	158158
EL3	152152	159159	106106	186186	112112	128128	095095	158158
EL6	152152	159159	106106	169186	112112	126128	095095	158158
EL8	152152	159159	106106	175175	108112	126128	095095	158158
EL9	152152	161161	106106	000000	112112	126128	095095	158158
EL10	152152	161161	106106	186186	112112	126128	095095	158158
HB2	152152	159159	106110	000000	112112	126128	095095	158158
JO1	152152	159159	106106	000000	112112	126128	095095	158158
JO2	152154	159159	106106	186186	112112	128128	095095	158158
JO3	152152	159159	106106	169169	112112	126128	095095	158158
JO4	152152	159159	106106	186186	112112	128128	095095	158158
JO5	152152	159159	106106	175169	112112	128128	095095	158158
JO6	152152	159159	106106	175169	112112	128128	095095	158158
JO7	152152	159159	106106	186186	112112	126128	095095	158158
JO8	152152	159159	106106	169169	112112	126128	095095	158158
WC1	152152	159159	106106	169186	112112	126128	095095	158158
WC2	152152	159159	106106	186186	112112	128128	101101	158158
WC3	152152	159159	106106	169186	112112	126128	095095	158158
WC4	152152	159159	106106	186186	112112	128128	095095	158158
WC11	152152	159158	106106	169186	112112	126128	095095	158158
WC12	152152	159158	106106	186186	112112	126128	095095	158158
BN1	152152	159167	106106	169169	112112	128128	095095	158158
BN3	152152	159159	106106	000000	112112	126128	095095	158158
BN4	152152	159167	106106	000000	112112	126128	095095	158158
BN5	152154	159159	106106	000000	112112	128128	095095	158158

Table 7.c. Microsatellite alleles found within the test group.

	TGLA53	TGLA142	DINC	MDRB3	OarHH64	OarCP26	MAF46	MAF50
Bb1	152152	159159	106106	175175	112112	128128	000000	158158
Bb3	152152	159161	106106	186186	112112	122128	000000	158158
WB2	152152	159159	106106	186186	108112	128128	095095	158158
WB3	152152	159159	106106	172172	108112	126126	095095	158158
WB4	152154	161161	106106	175186	108112	126126	095095	158158
WB5	154154	159159	106106	186186	108112	126128	095103	158158
WB6	152152	159159	106106	186186	108108	126128	095103	158158
WB7	152154	161161	106106	186186	116116	128130	101101	158158
WB8	152152	159159	106106	186186	114114	126128	095095	158158







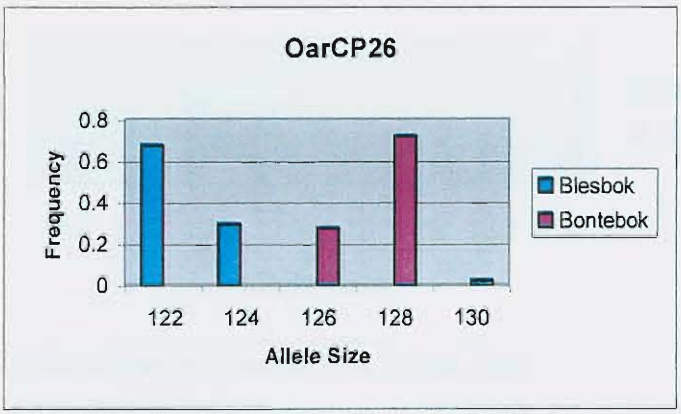
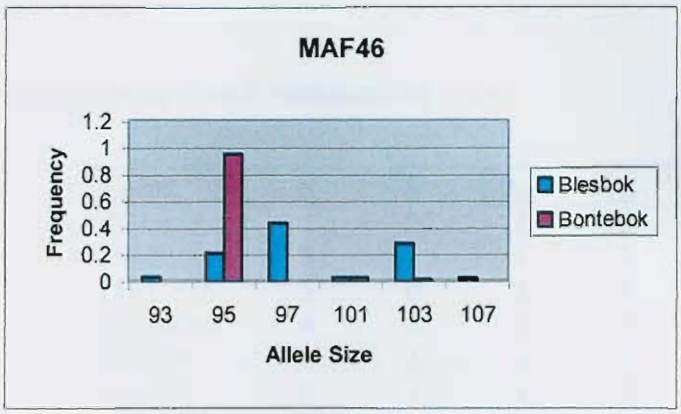
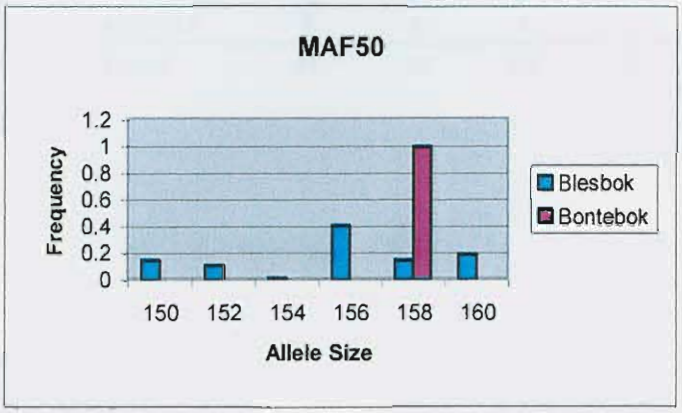
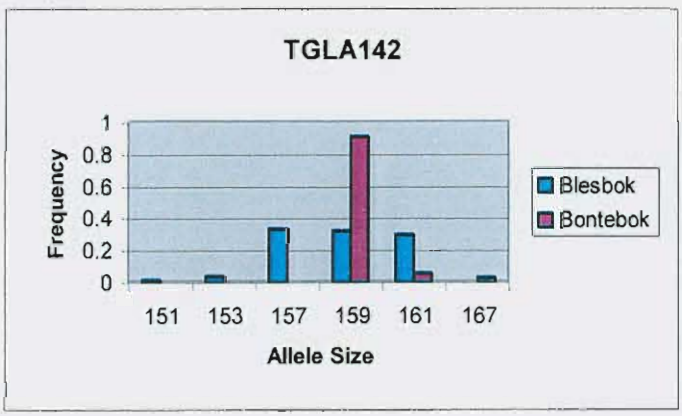
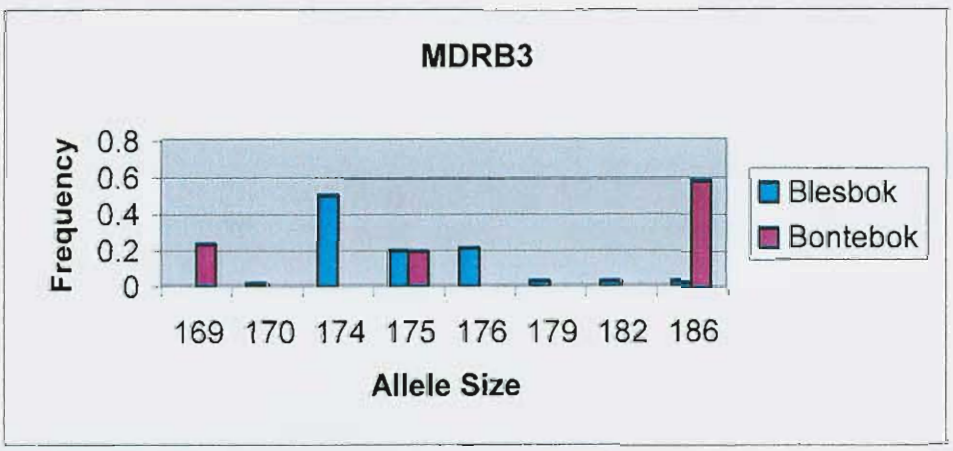


Figure 13.a. Allele frequency distribution









*MDRB3 is an imperfect di-nucleotide repeat locus

Figure 13.b. Allele frequency distribution

Table 8. Allelic states of each microsatellite locus.

Locus	N	S	P_{BL}	P_{BO}
TGLA53	3	2	1	0
DINC	3	1	1	1
OarCP26	5	0	2	3
OarHH64	4	3	1	0
MAF46	6	3	3	0
MAF50	6	1	4	1
TGLA142	6	2	2	2
MDRB3	8	2	4	2
Total	**41**	**14**	**18**	**9**

A comparison of genetic diversity between the two subspecies revealed a lack of polymorphism in the bontebok (Table 9a). The average observed levels of heterozygosity (H_O) were compared to the expected levels (H_E) separately for each subspecies. The observed values for the blesbok did not deviate significantly from Hardy-Weinberg equilibrium while the values for bontebok did exhibit heterozygosity deficiency across 6 out of 8 loci. Significant substructure was detected using the *Fst* and Rst calculations (Table 9b).

Genetic distances were estimated between all pairwise comparisons of individual multilocus genotypes by two measures: proportion of alleles shared (Dps) and kinship co-efficient (Dkf). Phylogenetic trees were built using both the UPGMA and NJ methods. Similar trees were reconstructed using both distance measures and phylogenetic algorithms (Figures 14-17). In all cases, the bontebok and blesbok clustered into separate groups that are supported by high bootstrap values. The bontebok cluster is characterized by shorter branch lengths and fewer bifurcations compared to that of the blesbok cluster. Microsatellite data from animals of unknown origin (Bb1, Bb3, WB2, WB3, WB4, WB5, WB6, WB7, WB8) were added to the composite multilocus genotype data from the bontebok and blesbok data set for phylogenetic analysis. All samples clustered within the bontebok group, including Bb1 and Bb3 that are derived from a hybrid herd (Figure 18).

Subspecies affiliation was executed using the variations of the GeneClass program (see Methods section). Table 10 lists each statistical method used to assign each individual to a class (subspecies). A reference data set was first computed by all statistical methods directly then under simulation. These results were then used for comparison against the test (unknown origin) sample set. All test samples were assigned to the bontebok class under all statistical methods.

Table 9. a. Heterozygosity values for 8 polymorphic loci.

Blesbok	# of Alleles	Het (e)	Het (o)	Bontebok	# of Alleles	Het (e)	Het (o)
TGLA53	3	0.654	0.500	TGLA53	2	0.112	0.118
DINC	2	0.450	0.428	DINC	2	0.029	0.029
OarCP26	3	0.456	0.500	OarCP26	2	0.409	0.559
OarHH64	4	0.549	0.405	OarHH64	3	0.141	0.088
MAF46	5	0.685	0.667	MAF46	3	0.086	0.029
MAF50	6	0.756	0.667	MAF50	1	0.000	0.000
TGLA142	4	0.704	0.643	TGLA142	4	0.221	0.118
MDRB3	7	0.676	0.167	MDRB3	3	0.588	0.308

Table 9.b. Microsatellite diversity indices.

Locus	Rst	Fst
TGLA53	0.527	0.243
DINC	0.976	0.775
OarCP26	0.066	0.891
OarHH64	0.223	0.294
MAF46	0.245	0.295
MAF50	1.013	0.265
TGLA142	6.024	0.584
MDRB3	0.253	0.771

Figure14. UPGMA-Ps Phylogenetic tree

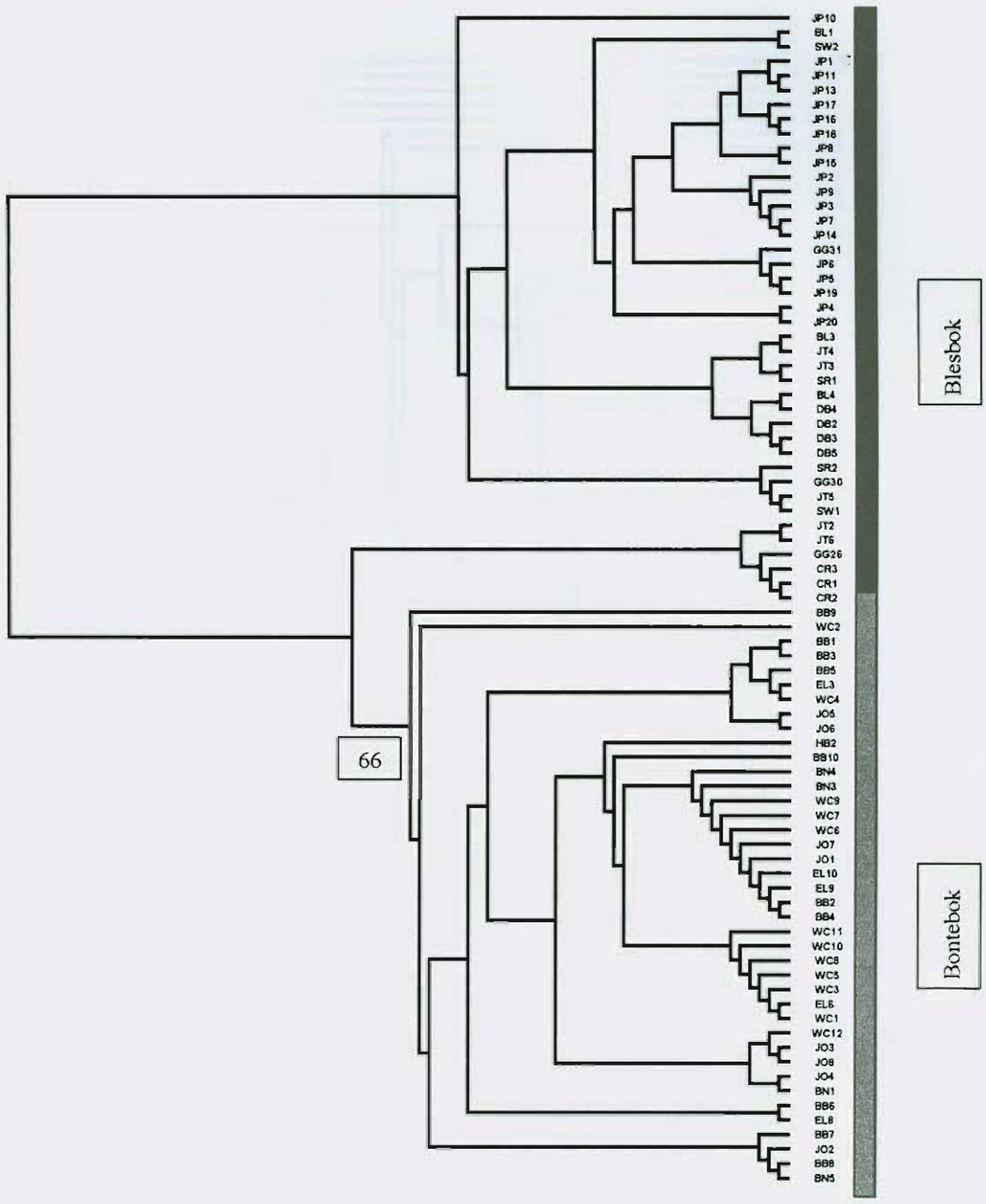


Figure 15. NJ-Ps Phylogenetic Tree

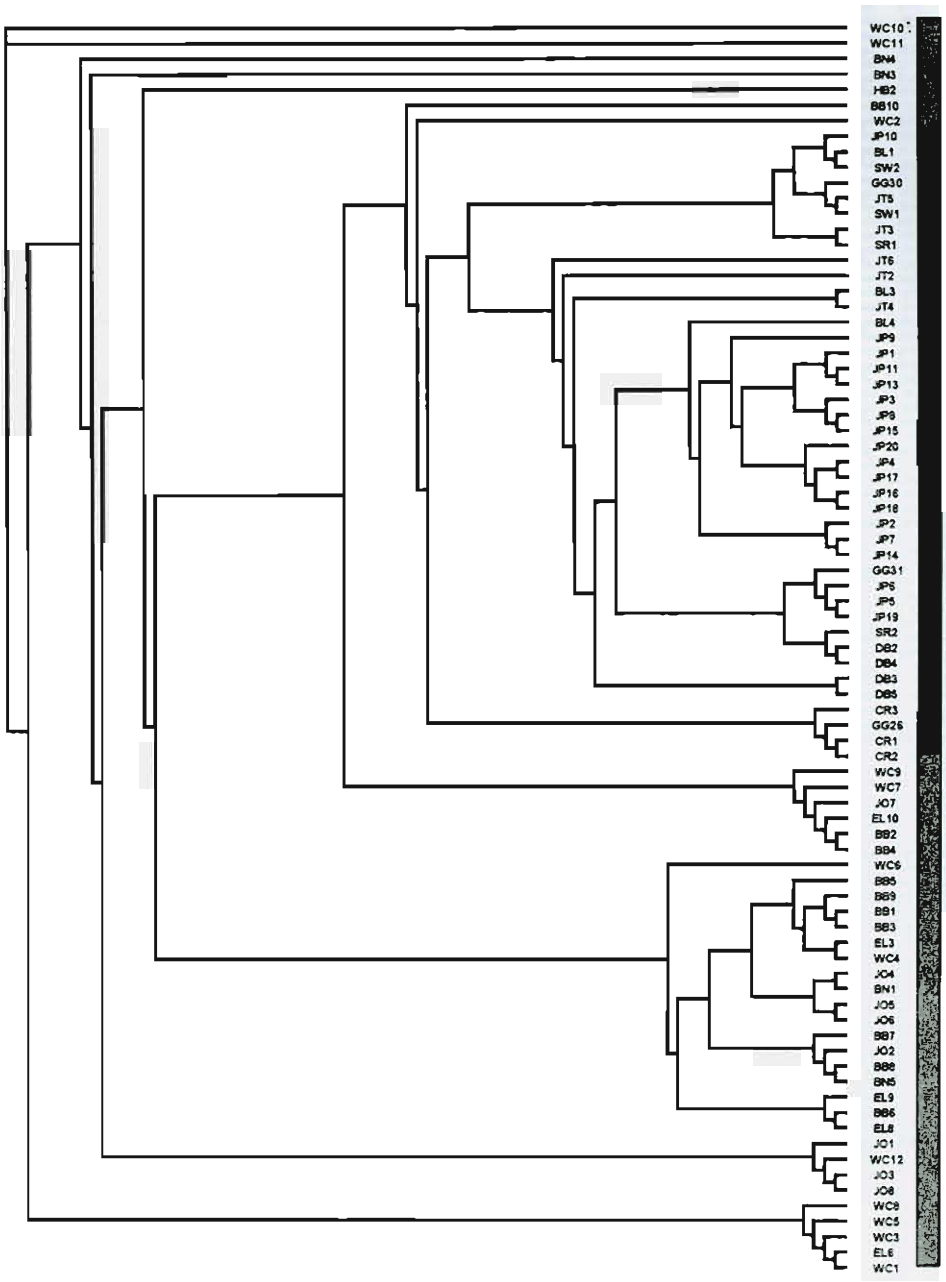


Figure 16. UPGMA-Kf Phylogenetic Tree

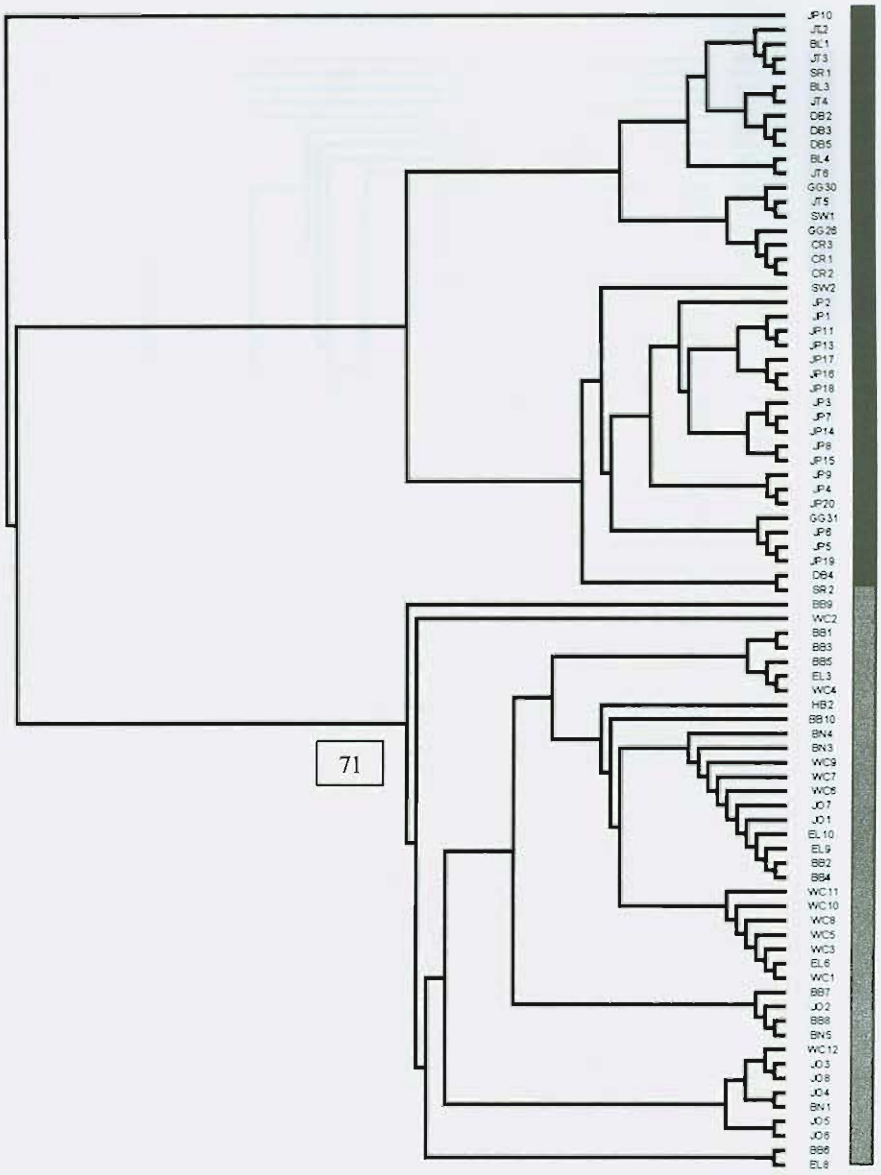


Figure 17. NJ-Kf Phylogenetic Tree

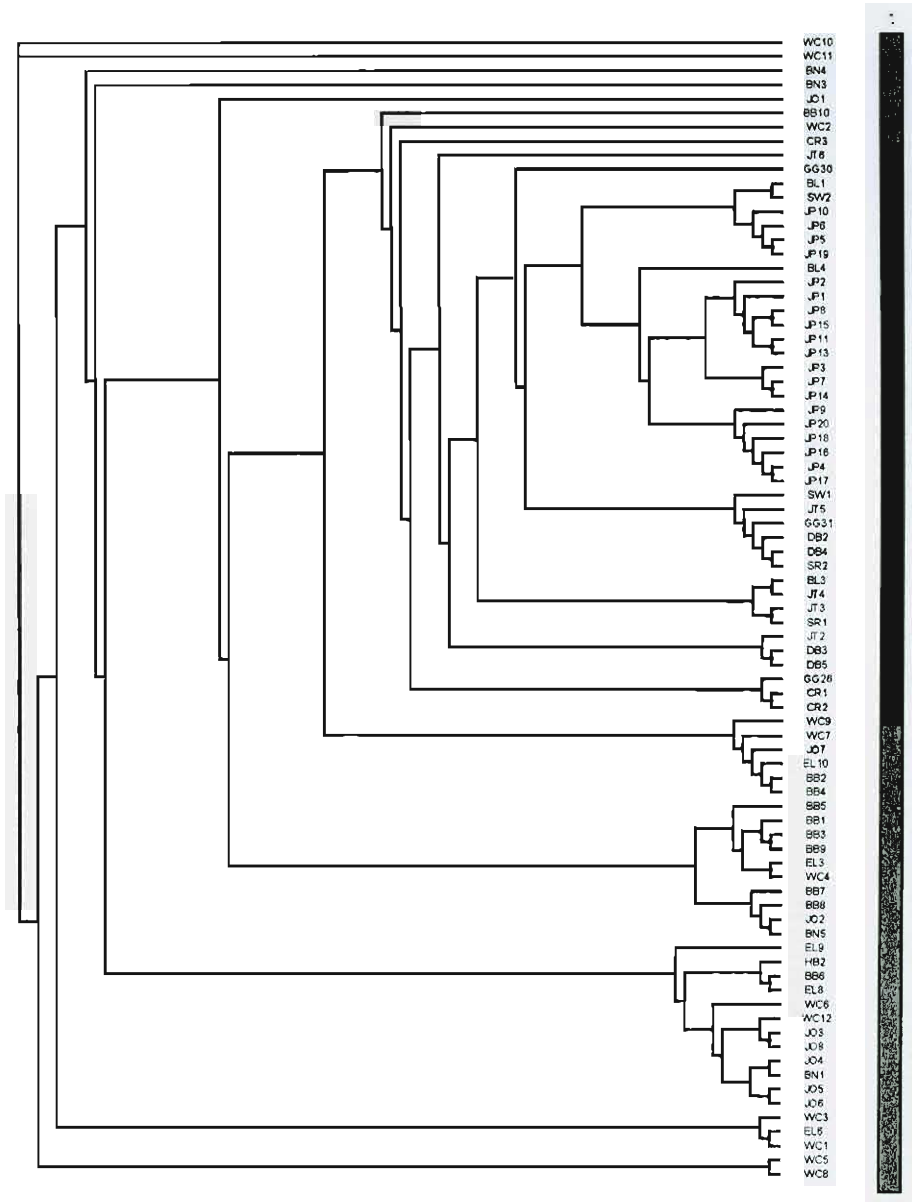


Figure 18. Test Animal Tree (NJ-Ps)

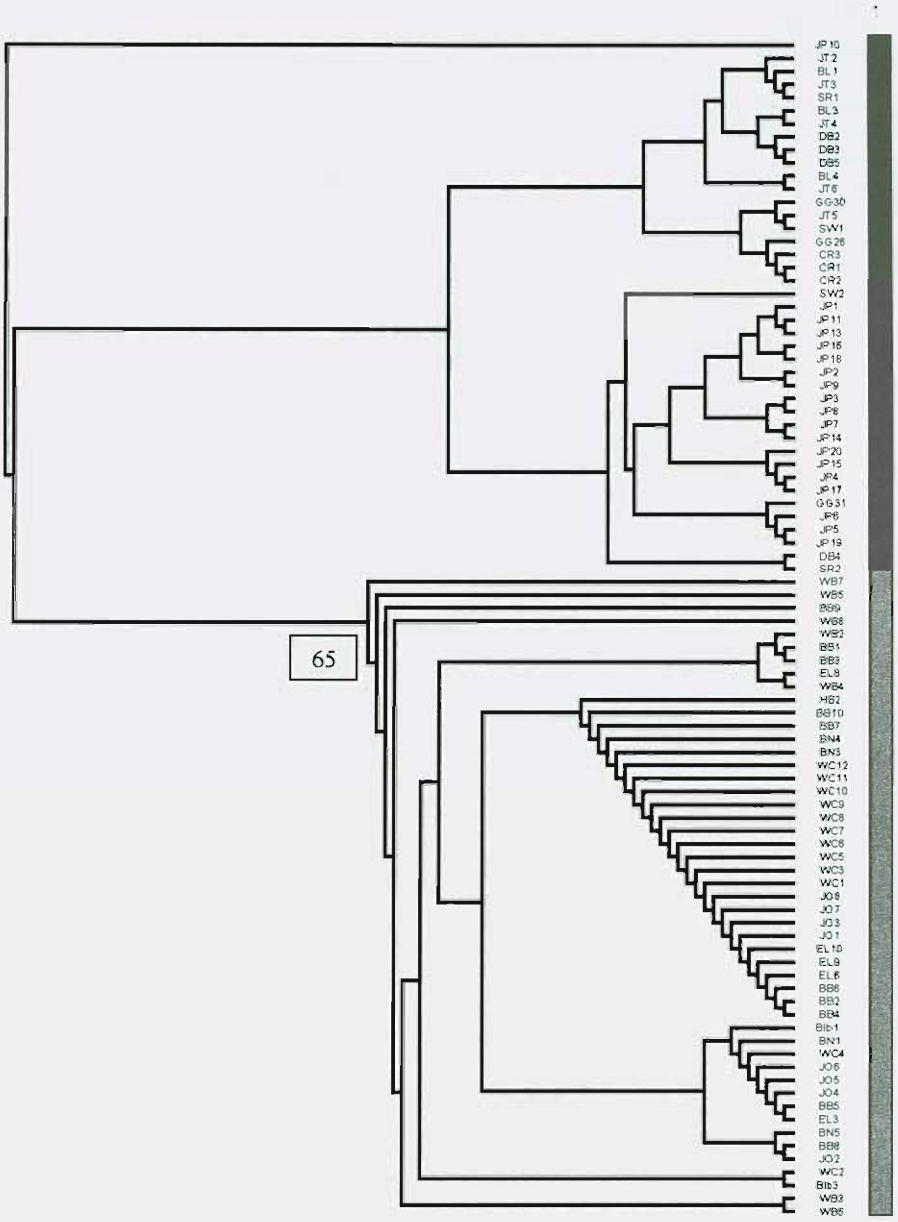


Table 10. GeneClass results for reference class and test group.

Self Assignment METHOD	CLASSIFIED CORRECTLY	SAMPLES MIS-CLASSIFIED	SAMPLES NOT CLASSIFIED		*TEST GROUP ASSIGNMENT*	
CHORD DISTANCE	96%	CR1, CR2, CR3	0		All Bontebok	
NEI'S STANDARD	97%	CR2, CR3	0		All Bontebok	
NEI'S MINIMUM	100%	0	0		All Bontebok	
NEI'S AVERAGE	100%	0	0		All Bontebok	
ALLELES SHARED	93%	JT2, SR1, CR1, CR2, CR3	0		All Bontebok	
DELTA MU	88%	JT6, GG31, EL6, JO3, JO8, WC1, WC3, WC11, BN1	0		WB3 Blesbok Others bontebok	
FREQUENCY	100%	0	0		All Bontebok	
BAYSIAN	100%	0	0		All Bontebok	
					TEST GROUP	
SIMULATION OF METHODS	**CLASSIFIED CORRECTLY**	**SAMPLES MIS-CLASSIFIED**	**SAMPLES NOT CLASSIFIED**		**NOT CLASSIFIED**	**Bontebok class**
CHORD DISTANCE	100%	0	JP10, GG31, BB9, WC2		WB3-7	Bb1, Bb3, WB2, WB8
NEI'S STANDARD	100%	0	JP10, JT6, DB4 GG31		WB3-7	Bb1, Bb3, WB2, WB8
NEI'S MINIMUM	100%	0	JP10, GG31, BB9, WC2		WB3-7	Bb1, Bb3, WB2, WB8
NEI'S AVERAGE	100%	0	JP10, JT6, DB4, GG31, BB9, EL9, EL10 WC2		WB3-8	Bb1, Bb3, WB2
ALLELES SHARED	100%	0	JP10, GG31		WB4, 5, 7	Bb1, Bb3, WB2, WB3, WB6, WB8
DELTA MU	88% samples assigned to both classes	GG31	JP10		WB3	Both classes
FREQUENCY	100%	0	JP10, JT6, DB4, GG31, BB9, EL9, EL10 WC2		WB3-8	Bb1, Bb3, WB2
BAYSIAN	100%	0	JP10, JT6, DB4, GG31, BB9, EL9, EL10 WC2		WB 3-8	Bb1, Bb3, WB2

*Note – Two assignment methods (self assignment and simulation) were performed to categorize samples into subspecies groups. Both types of assignment used either a distance-based or a likelihood-based method (frequency or Baysian). Seven distance estimators were used to assign an individual to the population that is genetically closest to it. The likelihood methods assign the individual to the group that has the highest likelihood compared to all other populations. The first 4 columns indicate the assignment of all samples from the bontebok and blesbok populations. The last 2 columns represent the classification of the test animals using all methods.

Microsatellite allele frequencies were imported into the Bottleneck program to quantitatively estimate the amount of genetic diversity lost through the demographic reductions in both subspecies. The blesbok data set did not show a highly significant excess of heterozygosity under the SMM model when tested by both the sign and Wilcoxon methods (Table 11.a). Within the bontebok data set, two loci (OarCP26, MAF50) demonstrated heterozygosity excess under the SMM model ($P = 0.028$, $P = 0.01$, respectively. The results of the sign and Wilcoxon tests do not show significant excess of heterozygosity under any model (Table 11.b.). The mode-shift test produced an L-shaped allele distribution that was expected under equilibrium for blesbok; however a slight mode shift was observed in the distribution of alleles found in bontebok (Figure 19).

		Observed		Under the I.A.M.				Under the S.M.M.			
Locus	**n**	**ko**	**He**	**Heq**	**S.D.**	**DH/sd**	**Prob**	**Heq**	**S.D.**	**DH/sd**	**Prob**
TGLA 53	84	4	0.659	0.450	0.169	1.238	0.0890	0.589	0.111	0.628	0.2990
TGLA 142	84	5	0.704	0.523	0.156	1.163	0.0920	0.675	0.082	0.356	0.4390
DINC	84	2	0.450	0.196	0.169	1.502	0.1370	0.227	0.168	1.330	0.1700
MDRB3	72	7	0.676	0.652	0.122	0.193	0.4870	0.774	0.054	-1.831	0.0550
OarHH64	84	4	0.549	0.433	0.175	0.665	0.3110	0.588	0.112	-0.348	0.2900
OarCP26	84	3	0.456	0.335	0.181	0.664	0.3480	0.466	0.136	-0.074	0.3760
MAF46	84	5	0.685	0.515	0.159	.064	0.1370	0.579	0.083	0.072	0.4380
MAF50	84	6	0.756	0.587	0.145	1.166	0.0800	0.729	0.068	0.407	0.4240

SIGN TEST

Assumptions: all loci fit I.A.M., mutation-drift equilibrium.
Expected number of loci with heterozygosity excess: **4.41**
0 loci with heterozygosity deficiency and 8 loci with heterozygosity excess.
Probability: 0.00811

Assumptions: all loci fit S.M.M., mutation-drift equilibrium.
Expected number of loci with heterozygosity excess: **4.66**
3 loci with heterozygosity deficiency and 5 loci with heterozygosity excess.
Probability: 0.55265

WILCOXON TEST

Assumptions: all loci fit I.A.M., mutation-drift equilibrium.
Probability (one tail for H deficiency): 1.00000
Probability (one tail for H excess): 0.00195
Probability (two tails for H excess and deficiency): 0.00391

Assumptions: all loci fit S.M.M., mutation-drift equilibrium.
Probability (one tail for H deficiency): 0.76953
Probability (one tail for H excess): 0.27344
Probability (two tails for H excess or deficiency): 0.54688

Table 11.a. Bottleneck test simulations for blesbok. The number of alleles for each locus is represented by (n). Expected heterozygosity values were generated from the observed number of alleles (k) for IAM and SMM. Standard deviations (SD) were computed. The standard difference is given as ((Hobs/Hexp)/SD). All tests indicated that the *Ne* of blesbok have not experienced a demographic reduction.

		Observed		Under the I.A.M.				Under the S.M.M.			
Locus	**n**	**ko**	**He**	**Heq**	**S.D.**	**DH/sd**	**Prob**	**Heq**	**S.D.**	**DH/sd**	**Prob**
TGLA 53	68	2	0.112	0.205	0.166	-0.559	0.4500	0.240	0.170	-0.751	0.3700
TGLA 142	68	5	0.221	0.541	0.151	-2.109	0.0410	0.680	0.081	-5.645	0.0000
DINC	68	2	0.029	0.201	0.168	-1.022	0.2230	0.245	0.167	-1.296	0.1360
MDRB3	52	3	0.588	0.374	0.173	1.236	0.1020	0.481	0.132	0.815	0.2180
OarHH64	68	3	0.141	0.354	0.176	-1.213	0.1890	0.470	0.137	-2.409	**0.0280**
OarCP26	68	2	0.409	0.203	0.169	1.217	0.2010	0.253	0.170	0.915	0.2820
MAF46	68	3	0.086	0.347	0.174	-1.492	0.1040	0.471	0.134	-2.873	**0.0100**
MAF50	68	1	0.000	MONOMORPHIC LOCUS.							

SIGN TEST

Assumptions: all loci fit I.A.M., mutation-drift equilibrium.
Expected number of loci with heterozygosity excess: **3.42**
5 loci with heterozygosity deficiency and **2** loci with heterozygosity excess.
Probability: 0.24249

Assumptions: all loci fit S.M.M., mutation-drift equilibrium.
Expected number of loci with heterozygosity excess: **3.82**
5 loci with heterozygosity deficiency and **2** loci with heterozygosity excess.
Probability: 0.15553

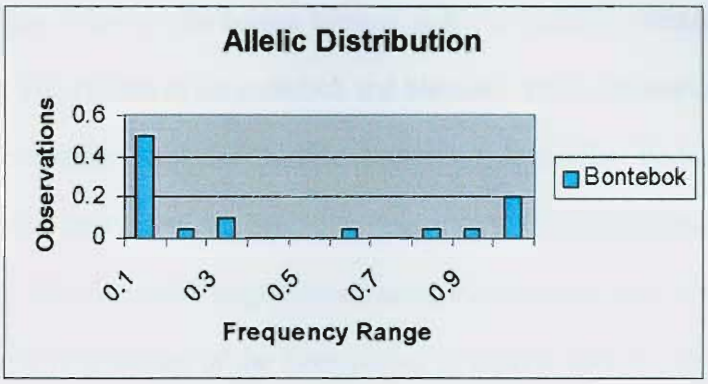
WILCOXON TEST

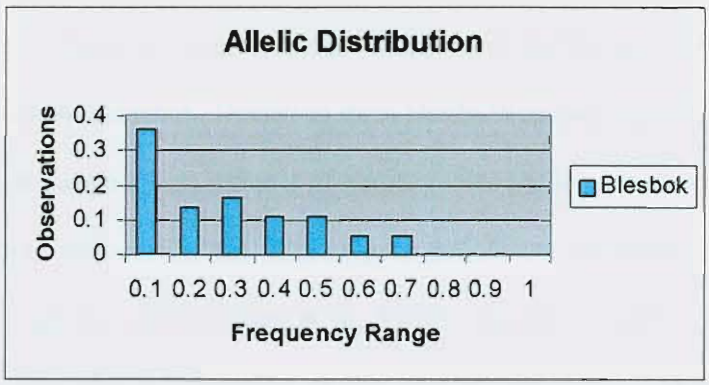
Assumptions: all loci fit I.A.M., mutation-drift equilibrium.
Probability (one tail for H deficiency): 0.23438
Probability (one tail for H excess): 0.81250
Probability (two tails for H excess and deficiency): 0.46875

Assumptions: all loci fit S.M.M., mutation-drift equilibrium.
Probability (one tail for H deficiency): **0.03906**
Probability (one tail for H excess): 0.97266
Probability (two tails for H excess or deficiency): 0.07813

Table 11.b. Bottleneck test simulations for bontebok. All tests indicated that the *Ne* of bontebok have not experienced a demographic reduction.

Figure 19. Distribution of microsatellite alleles for all loci





Discussion

A panel of bovine and ovine primers was tested for cross-species amplification in *D. pygargus*. Many of the bovine primers failed to amplify, produce strong products or exhibit polymorphism in the bontebok and blesbok. The ovine markers proved to amplify more successfully than the bovine primers. A similar finding was shown in a microsatellite study of the closely related hartebeest (*Alcelaphus buselaphus*) (Flagstad et al. 1999). The successful amplification using the ovine primers is most likely explained by the close relationship of the alcelaphines to caprine species rather than to the bovids (Matthee and Robinson 1999a).

The microsatellite diversity ranged in the amount of variation found within subspecies. The lack of genetic diversity appears to be the result of the bottleneck events experienced by bontebok. Deviation from Hardy-Weinberg equilibrium was not found in bontebok although a high number of homozygotes per locus was observed. A number of events could account for the heterozygote deficiency, including selection, non-random mating or genetic drift (Callen et al. 1993). Given the demographic history of the bontebok and lack of variation at the control region, it is most plausible that the bottlenecks have caused a great loss of allelic diversity within microsatellite loci.

In order to assess the loss of nuclear genetic variation in the bontebok, the outbred blesbok subspecies should provide a reliable indicator of diversity. The estimated diversity may be slightly reduced, partially biased, since the microsatellite primers were designed from homologous loci in cattle and sheep (Goldstein and Pollock 1997). However, this ascertainment bias should not influence the relationship between the amount of variation that is maintained within each subspecies. Blesbok demonstrated a moderate level of allelic diversity and heterozygosity. Table 12 illustrates the range of heterozygosity values for ungulate species having different demographic histories.

Table 12. Microsatellite heterozygosity values and demographic histories of various ruminant species.

Taxa	# Loci Typed	Average H	Demographic history	Reference
Moose (Canada)	5	0.219	Founder event with 6 individuals	Broders et al. 1999
African buffalo	6	0.648	Slight bottleneck event	Simonsen et al. 1998
Hartebeest	16	0.700	Outbred	Flagstad et al. 1999
Arabian oryx	6	0.544	Captive herd, founder effect	Marshall et al. 1999
Muskox	12	0.059	Possible historic or prehistoric bottleneck	Holm et al. 1999
Blesbok	8	0.623	Outbred	This study
Bontebok	8	0.201	Bottleneck events	This study

Clustering methods and assignment tests were preformed for the assignment of individuals into subspecies classes. The phylogenetic trees depicted clustering of individuals according to their subspecies designation. Unknown samples were placed within the bontebok subspecies class. It appears that the genotypes of the test samples consisted of many of the shared alleles rather than the subspecies specific (private) alleles. Samples (Bb1, Bb3) were derived from a "bontebok" population that is suspected to contain hybrid animals based on the presence of blesbok control region haplotypes. The clustering test places them within the bontebok class since no blesbok private alleles were evident.

Multilocus genotypes were used for assignment using distance estimates and likelihood scores. The reference data set showed high success in correct classification using all variations of each method (96%-100%) with delta mu^2 being the exception displaying the lowest percentage (88%). Delta mu^2 is successfully used to estimate distances between taxa at the species level (Goldstein et al. 1995a,b). The samples that were mis-classified (CR1, CR2, CR3) are derived from a blesbok population located within the Eastern Cape. This population could possess old bontebok alleles from the time before the species split into two subspecies. A more likely scenario is that bontebok were translocated into that herd of blesbok. The data set and each method was simulated and shown to classify individuals correctly (100%), however, the methods failed to place several samples into a category (see Table 10). The genotypes of many of the un-classified animals consist of rare, low frequency alleles. The rare alleles may create larger distances and smaller likelihood values for these animals and make placement difficult compared to all other samples. The test group data set was compared to the reference set and all samples were classified to the bontebok subspecies by direct

assignment. The simulated test showed similar results but failed to classify animals "test" animals (WB 3 - WB7) from one population.

The classification tests and phylogenetic clustering methods demonstrate that individuals of known origin can be correctly placed into their subspecies category. However, the degree to which the assignment tests were able to successfully assign an unknown was not confirmed due to the limitations of the study. More animals documented as hybrids are required to serve as positive controls. In this study, animals were only suspected as being hybrid by morphological characteristics. Secondly, a higher number of polymorphic loci (10 – 20 loci) would provide greater allocation success of known animals (Blott et al. 1999).

It has been predicted that bottlenecked populations will demonstrate excess in heterozygosity that will be higher than expected at equilibrium (Luikart et al. 1998a,b). The blesbok data did not display significant heterozygosity excess when tested under the assumptions of the SSM model. Furthermore, the distribution of allele frequencies did not show the signature shift that is indicative of a bottlenecked population. These results suggest that blesbok have had a stable demographic history.

Two loci demonstrated significant heterozygosity excess within the bontebok data set. The results of the sign and Wilcox tests do not indicate a bottleneck event in bontebok, although historical records have recorded two severe reductions in (*Ne*). Three possible explanations may elucidate the results of the bottleneck tests. Firstly, the historical records may have been incorrect and the actual founding population of bontebok may have been larger than the 20-30 individuals estimated. Secondly, the power of bottleneck detection relies more on the assumption of heterozygosity excess than on allelic diversity estimates. It has been suggested that reduction in (*Ne*) has a greater impact on the allelic diversity of a population than on heterozygosity (Nei et al.

1976). Therefore, the distribution of alleles appears to be a more powerful in detecting recent and severe bottleneck events. Lastly, the bottleneck tests may be dependent on the size of the bottleneck and the number of generations that have passed since the time of the event. Luikart and colleagues (1998a) have simulated bottleneck events based on varying number of founders and have reported that small bottleneck sizes (<20) are more likely to be detected.

A genetic bottleneck event will quickly result in the loss of low frequent alleles, which will in turn, cause the increase of the intermediate and high frequency alleles (Nei et al. 1975, Luikart et al. 1998a). Figure 19 clearly illustrates the loss of rare and intermediate alleles in the distribution of bontebok microsatellite alleles compared to that of blesbok. The graphical representation also reveals a distortion of allele frequencies, however, the results of the Bottleneck program indicate that the distribution of bontebok alleles is L-shaped. Clearly, there can be other alternatives to distribution shape other than the L-shape and mode-shift distributions. The generated data set from Luikart et al. 1998a also revealed that a mode shift is only detected after 40 generations with 20 founders. I predict that there is a mode shift in the distribution and will be apparent once the number of generations exceeds 40.

Microsatellite diversity was examined in order to assess the nuclear gene variation within each subspecies. The results of this study confirm the mtDNA findings that revealed extremely low levels of genetic variation in bontebok and moderate levels within blesbok and a high degree of substructure between the subspecies. Both studies revealed patterns of genetic evolution at non-coding loci. In the next chapter, I present the results from a coding nuclear gene that is presumed to evolve under selective forces in vertebrate species.