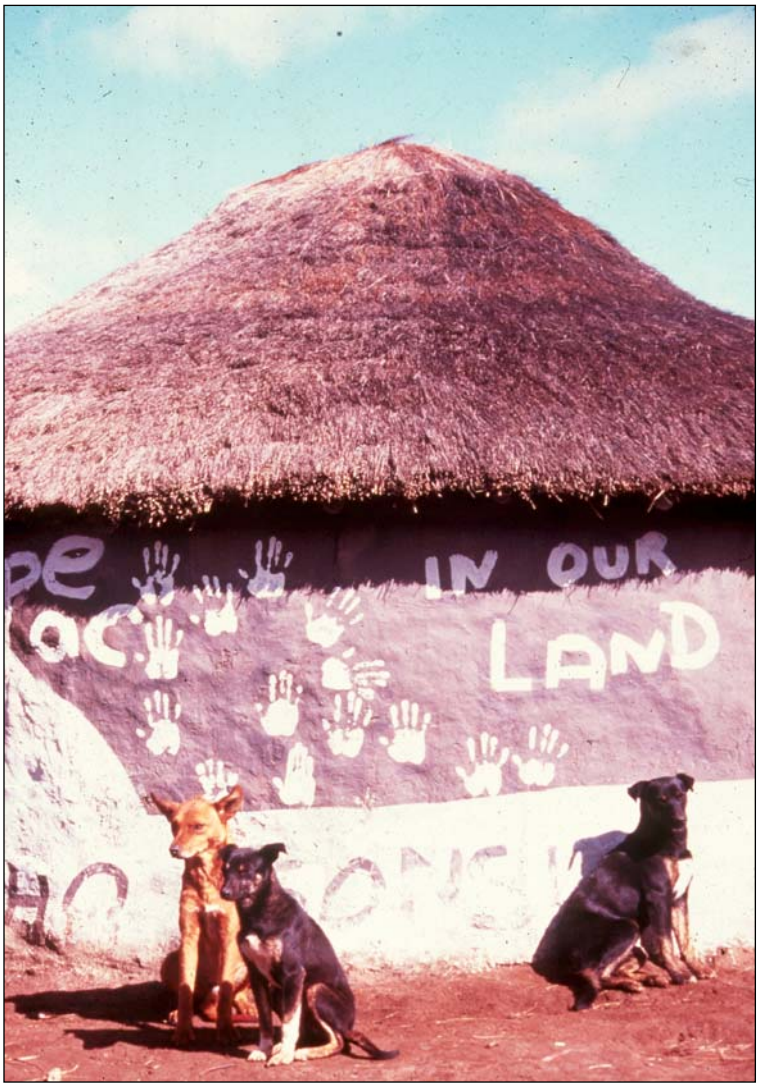
CHAPTER 4

DISCUSSION



DISCUSSION

4.1. THE USE OF PROTEIN ELECTROPHORESIS IN CANINE GENETICS

Allozyme based results have contributed tremendously to the study of canine genetics. It has provided insight into the variation and differentiation between species (Fisher *et al.*, 1976; Wayne *et al*., 1991b) and populations (Kennedy *et al.*, 1991; Wayne *et al.*, 1991a; Wayne, 1996), despite several theoretical and practical disadvantages associated with this technique (Avise, 1974; Harris and Hopkinson, 1976; Ayala, 1982). With the advance of technology in recent decades, genetic methods have become more modern and sophisticated and can provide accurate data on the genomes of organisms. These techniques include microsatellites, amino acid or nucleotide sequencing, random amplified polymorphic DNA segments (RAPDs), restriction fragment length polymorphisms (RFLPs) and amplified fragment length polymorphisms (AFLPs). Yet, allozyme studies can be distinguished by several advantages. These include objectivity, constancy of genetic characters and the identification of homology (Avise, 1974). According to Ayala (1982), protein electrophoresis allows the study of protein variation with only a moderate investment of time and money. Compared to the time constraints and costs associated with sequencing of proteins or nucleotides, protein electrophoresis has a definite advantage. Also, protein electrophoresis allows for large numbers of different individuals to be screened simultaneously in population studies using relatively simple thought precise techniques (Harris and Hopkinson, 1976). Although techniques like mini- and microsatellite loci may provide a finer grained picture of population

history and variation in some cases where neutral, rapidly evolving loci are involved, levels of variation in allozymes may be at least partly indicative of balancing selection rather than demographic history alone (Allendorf and Leary, 1986). Probably one of the greatest advantages of allozyme studies is that this technique requires no optimising, i.e. there is no need to have specific primers, as is the case for most DNA analyses.

Allozyme electrophoresis was chosen as an appropriate technique in this study based on the specific questions that needed to be answered concerning the dog populations' genetic variation and differentiation. Several other studies have shown that enough genetic variation exists between different species and populations of *Canis* for this technique to be useful in assessing the issues at hand. Protein electrophoresis has been used to distinguish several populations and species of the genus *Canis*. With the help of biochemical markers, the red wolf (*Canis rufus*) could be identified as a unique species, as well as the identification of hybrids with other members of this genus (Ferrell *et al*., 1980). Starch-gel electrophoresis has been used in canine genetics for various purposes, from assessing the genetic variability in natural populations of wolves and other canids (Fisher *et al.*, 1976; Kennedy *et al*., 1991; Wayne *et al*., 1991a, b) to calculating the allozyme divergence within the Canidae family (Wayne and O'Brien, 1987).

Genetic data, produced by protein electrophoresis, can be used by systematisists to determine if samples are from different gene pools, representing different species or, in this case, different populations or breeds (Thorpe and Solé-Cava, 1994). Genetic variation and genetic differentiation may be measured at various levels, but in this study

protein electrophoresis was used to identify the allelic variation at structural loci. Due to the large amount of structural loci present in the genome of an organism (especially higher organisms, such as mammals), it is virtually impossible to study and determine the exact amount of genetic variation in a population. Protein electrophoresis is used to estimate the total genetic variation by analysing a small proportion of genes from random sampled loci. It must be noted, however, that the "random" sampling of loci does depend on the availability of a proper staining technique for a specific encoded protein and this criterion has therefore been questioned by some authors (e.g. Leigh Brown and Langley, 1979), but is generally considered as random (Hubby and Lewontin, 1966).

The reasoning behind the analysis of protein or enzyme differences in order to understand and estimate genetic differences is derived from the fact that specific genes control the expression of specific proteins (such as enzymes or hemoglobins). Specific chemical differences in protein molecules have been shown to be directly correlated with allelic alterations in genes (Sinnott *et al*., 1958).

Furthermore, allozymes are almost invariably codominant. The calculation of gene frequencies is very simple, because the identification of hetero- and homozygotes is easy (heterozygotes have different phenotypes from the homozygotes). Molecular systematics is directly linked to the manner in which genes evolve and their distribution in the species, contrary to most morphological, numerical and chemical systematics (Van der Bank, *et. al*., 2001).

Isozyme data often constitute the largest existing genetic data sets for many organisms, both within and between species (Park and Moran, 1995). Reasons for this are likely to include the relative low cost associated with this technique (compared to other molecular genetic methods) and the quick processing time required (Van der Bank *et. al*., 2001). A laboratory can assay many hundreds of samples per day for many different loci, which translates into further reductions in labour costs.

Objectivity is also one of the major advantages of protein electrophoresis. The numeration of alleles and their frequencies are objective determinations, based solely on the mobility of bands on gels (Van der Bank *et.al.*, 2001). Going in hand in hand with this, is the matter of weighting. The initial weighting of characters is not a problem in electrophoretic data, because each locus is accorded equal value (Van der Bank *et. al.*, 2001). *A posteriori* weighting may be practised if some loci appear to be of more value than others in elucidating systematic relationships (Avise, 1974).

Techniques involving DNA-based analysis do not suggest higher accuracy or reliability. It has been shown that there are discrepancies even within these techniques. The gene trees are often incongruent with the species trees (e.g. see reviews by Doyle, 1997; Maddison, 1997). Patton and Smith (1994) concluded that patterns of species and population relationships suggested by mitochondrial cytochrome b and the nuclear genes (electropmorphs) are sometimes opposite. Another example can be seen in a study done by Karl *et. al*. (1992) where it was also found that nuclear DNA did not yield the same results as mitochondrial DNA (mtDNA). Other studies have, however, shown that

morphological, mtDNA and allozyme data can agree and support each others results (Britten *et. al.*, 1997).

4.2. GENETIC VARIATION IN THE FOUR POPULATIONS STUDIED

With regards to the overall genetic variation between the various populations studied, two main trends could be distinguished, especially concerning genetic diversity. (a) The dogs sampled from the Jericho area and those from the SPCA showed the most diversity of all the groups studied. (b) The Africanis and Saluqi populations on the other hand displayed lower levels of diversity, as can be interpreted from most of the values in this section.

The Jericho group was by far the most polymorphic, with **H** and **P** values (0.159 and 33.33 respectively) much higher than those for any of the other populations studied (Table 5). The SPCA group showed a similar pattern with regards to high levels of polymorphism and diversity, and this was emphasised by the relatively high **H** value of 0.137. Although the **P** value for the latter population (28.57) does not support this diversity completely, it doesn't deny it either. These two groups represent the mixed breeds and races of a numerically large, free-ranging population and could therefore also be potentially favourable for **H**, which seems to be the case here. This, however, does not necessarily lead to perfect or ideal Hardy-Weinberg relationships, as is shown in Table 3.

Ideal Hardy-Weinberg populations do not actually occur in nature due to various factors that can shift the equilibrium and change or disrupt the stability of a population, giving rise to change in genetic structure. These factors include random genetic drift, non-random mating, mutation and self-sorting crossing and linking (Hickman *et al.*, 1993). Furthermore, significant deviations of allele frequencies may occur owing to other factors such as inbreeding, population bottlenecks and sampling errors (Ayala, 1982).

It is not surprising that genotype classes deviated from the expected Hardy-Weinberg proportions at most loci (see Table 3). The Africanis population did not comply with Hardy-Weinberg expectations at loci **CK**, **PER** and **P-Tf** (although they met Hardy-Weinberg expectations at loci **PA-1**, **-2** and **–3**); the SPCA group deviated from Hardy-Weinberg only at two loci (**AK-2** and **Hb**), whereas the Saluqi population showed a deviation only at one locus, namely **PER**. The population sampled at Jericho fell short of Hardy-Weinberg expectations at the following loci: **AK-1** and **PER** (with conformation to Hardy-Weinberg expectations at loci **Hb**, **PA-1**, **-2**, **-3** and **P-Tf**). The deviation from Hardy-Weinberg expectation could be due to the fact that the dogs studied do not represent purely natural populations, but have instead been subjected to selection, either natural or artificial. All domestic dog breeds are descendant populations of the grey wolf (Wayne, 1993), and although they might be feral in some cases, humans to a certain degree control their reproduction and effectively influence their evolution. However, non-random mating and gene flow can be excluded since these processes generally affect all loci (Flint *et al*., 1998/1999). This fairly high level of deviation from expected Hardy-Weinberg values for the Africanis population, suggests that these dogs have been isolated

from gene flow of Western dog breeds and have possibly gone through a historical bottleneck, which caused this isolation.

Values for **H** and **P** (Table 5) can also be interpreted as a lower level of diversity for Africanis and Saluqi. Although low diversity was suggested in the latter population (the lowest **P** value of 23.81), Africanis had the smallest **H** value (0.106), confirming the low genetic diversity for these two populations in comparison with the other two, mixed groups (Jericho and SPCA).

The **P** (0.95 criterion) values obtained varied between 23.81% and 33.33% for Saluqi and Jericho respectively (Table 5). Although values ranging from 20 to 86% have been reported in most animal groups for which large numbers of protein-coding loci have been surveyed (e.g. Selander and Johnson, 1973; Lewontin, 1974), the values from this study are much higher than those reported in the literature for wolves. Kennedy *et al.* (1991) found the overall percentage of polymorphic loci to be between 13.5 and 17.9% for wolves in the north-western Canada, which was also higher than the 11.3% reported by Fisher *et al*. (1976). The differences in values could be explained by the fact to mostly monomeric and dimeric enzymes were examined in this study (compared to multimeric enzymes used in the comparable studies), which have been shown to have some influence over the amount of proportionate polymorphic loci. Both the proportions of polymorphic loci and average heterozygosity are higher for monomers than for dimers, which in turn show higher values of **P** and **H** than trimers and tetramers (Zouros, 1976; Harris *et al.*, 1977; Ward, 1977). The reason for this lower degree of polymorphism in multimeric

enzymes seems to be the higher degree of functional constrain, which stems from the fact that different polypeptides have to form a single functional protein (Koehn and Eanes, 1978; Kimura, 1983).

The values for unbiased **H** varied between 0.106 and 0.159. Jericho and SPCA had the highest **H** values and therefore have the most genetic variation. This is expected for mixed (non-descriptive) groups, as correctly assumed previously (see Chapter 2 - *Material and Methods*). Kennedy *et al.* (1991) approximated the average value for Canadian wolves at 3.6%. Average heterozygosity values ranging from 0.0 to 8.0% (with most less than 3.0%) have been reported in the literature for natural populations of other families of Carnivora (Allendorf *et al*., 1979; Hamilton and Kennedy, 1987; Manlove *et al*., 1980; Simonsen, 1982; Wathen *et al*., 1985). An unusually high level of heterozygosity (17.0%) for 10 *Martes americana* (pine marten or American sable – a weasel like animal) from a single locality has however been reported (Mitton and Raphael, 1990). The authors suggested that "sampling of related individuals may have inflated the estimate of heterozygosity" in this case.

Selander and Kaufman (1973) reported average **H** values of natural vertebrate populations (rats, seals and humans) of more or less 6%. All of the dog groups analysed in the present study showed much higher values, but this is likely due to the fact that they do not represent natural populations. One possible explanation for this (although by no means the only one) is that selection occurs in favour of heterozygosity, as opposed to homozygosity (Ford, 1964). Selection will tend to make the favourable effects of the

gene dominant and the unfavourable ones recessive; therefore, the heterozygote will have nothing but advantages and will be superior to the homozygotes (Sheppard, 1953). It is very interesting to note though, that Africanis showed the closest **H** value (10%) to the 6% mentioned above. This could be explained by the fact that these dogs are part of a rural community where domestication is limited. These dogs rely on humans to feed and care for them in a way, but can manage very well on their own to fend for themselves. They are known to help with hunting, therefore, know how to obtain food, where to find water, have their own hierarchy with individuals as leaders and live very much outdoors where they are exposed to elements from nature. In short, they live an almost natural life.

Although it is interesting to note that distinct differences were encountered at allele frequencies, it must be mentioned that distinct (fixed) markers were not found. However, pair-wise comparisons showed that significant ($P<0.05$) genotypic differences occurred between all of the groups (Table 4) at four loci. These loci (**AK-1, -2, Hb** and **P-Tf**) can serve as markers to define the groups. For example, only Africanis showed the presence of private alleles, namely **CK*B** (no heterozygotes were encountered) and **PA-2*C**; this could also be due to small sample size for Saluqi and Jericho, but not SPCA (N=92). Further studies and larger sampling of the former two groups should be attempted to verify this result. The major insight into the possible uniqueness of the Africanis breed was revealed at the **P-Tf** locus. Here it was not so much the allelic frequency differences that made this group unique, but rather the genotypic variation with regards to the allelic differentiation. The combination of alleles A and C were found in much larger numbers

than that of the other groups. Three alleles were observed at the **PA-1**, **-2** and **P-Tf** loci, whereas all of the other polymorphic protein coding loci had only two alleles.

The populations had an **A** value ranging from 1.33(±0.14) in Saluqi to 1.43(±0.15) in Jericho (Table 5). Allelic diversity (**A**) ranged from 4.4 to 4.5 in a study of Canadian wolf populations and was 4.1 in wolves from Montana (Forbes and Boyd, 1996, 1997). These values were obtained using DNA microsatellites and are therefore expected to be higher than those found using allozyme electrophoresis (Table 5). Although these latter values can not be numerically compared with those obtained from the present study, they do suggest a similar trend with that found here and is therefore useful as a comparison within species regarding certain genetic tendencies.

In a study done between five pig breeds (Van der Bank *et al*., 1997), a comparable group, being domesticated mammals as well, the **A** value was slightly higher (although in the same order) and ranged from 1.91(±0.25) to 2.91(±0.67). The Wild Pig, however, had a lower value of 1.18(±0.12), which compares well with the values found for *C. familiaris* in this study (Table 5). Allelic diversity (**A**) is more sensitive to founder events than **H**, and it is more indicative of future adaptive potential (Nei *et al.*, 1975; Leberg, 1992). This value is also biologically more significant, because it reflects loss of alleles directly, whereas **H** is more of a mathematical value. Therefore, even though the **A** values obtained in this study was found to be lower than those for wolves (Forbes and Boyd, 1996, 1997), but similar to other domestic animals (Van der Bank *et al.*, 1997), it is not an indication of a major bottleneck and indicate possible inbreeding.

4.3. GENETIC DIFFERENTIATION BETWEEN THE FOUR POPULATIONS STUDIED

Throughout the results in this section, pertaining to genetic differentiation, yet another notable trend could be identified. All these values suggest and indicate a close association between Africanis and those dogs of Middle Eastern origin, the Saluqi's. Overall, the dogs sampled from animal shelters and squatter camps (i.e. SPCA and Jericho populations) were the furthest removed, genetically, from the endemic breed, as would be expected for such a diverse, mixed sampling group.

The computer programme GENEPOP (Raymond and Rousset, 1995, 1997) was used to differentiate between all the populations on the basis of significant differences in allele classes (Table 4). The **AK-1** and **Hb** loci can be used to differentiate between all the populations. The **AK-2** locus can be used to differentiate between SPCA group and all the other populations, except Jericho, whereas **P-Tf** can be used to differentiate populations SPCA and Saluqi from the Africanis breed.

Genetic distances values are used to indicate genetic or population differentiation. There are also several statistically based measures available to reduce genetic differentiation between populations over a range of enzyme loci to a single figure level, but Nei's (1978) measure is now used predominantly (Thorpe and Solé-Cava, 1994). These values are specially adapted for small sample sizes and were calculated from the observed allele frequencies at all 21 protein-coding loci. Table 8 shows a matrix of genetic distance

coefficients. The values compared fairly well with those known in the literature. Wayne and O'Brien (1987) calculated the genetic distance (Nei, 1978) to be 0.013 between *C. familiaris* and *C. lupus*. The largest **D** value in the present study was obtained between Africanis and Jericho (0.026) and the smallest value of 0.001 was found between Africanis and Saluqi dogs. The average **D** between all the populations is 0.013.

The lower **D** value (0.001) calculated between the southern African breed and that of the Middle Eastern Saluqi dogs is notable. This reflects a similarity between these two breeds, which could indicate migration of dogs through and down the African continent from the northern parts of Africa and the Middle East, as proposed by Gallant (1999; in prep.) from archaeological evidence (Camps, 1977; Hoffman, 1984; Roubet and Carter, 1984; Boessneck, 1988). It is further proof that the endemic breed is descendant from immigrants to this continent thousands of years ago, and not through western influences. If Nei's (1975, 1987) formula for estimating the time since divergence between two populations or species is used, where time (in years) = 5×10^6 D, it shows that these two populations (Africanis and Saluqi) diverged about 5 000 years ago. Archaeological evidence suggests exactly the same estimate (Gallant, in prep).

The phylogenetic relationships between the breeds/populations studied are summarised in Figure 8 . The close relationship between the Africanis breed and the dogs from the Middle East support the notion that southern African endemic breeds came from the Middle East and are closely linked to them with regards to their ancestry (Gallant, 1999;

in prep.). The high bootstrap value (generated by the DISPAN computer programme) of 85 at the node between the Africanis and Saluqi breeds, supports this strong relationship.

Fixation index values also provide informative measures of population structure. The most important of these is the $\mathbf{F_{ST}}$ value, the proportion of total variation that is due to differences between subpopulations (if $\mathbf{F_{ST}}$ = 1, subpopulations have no alleles in common; if $\mathbf{F_{ST}}$ = 0, allele frequencies in all subpopulations are identical). The mean $\mathbf{F_{ST}}$ value of 0.090 for the polymorphic loci (Table 6) indicates little genetic differentiation between populations resulting from genetic drift. This level of differentiation is comparable to that among populations of *Canis latrans* (coyote) and subunits of *C. lupus*, with $\mathbf{F_{ST}}$ values of 0.080 (Hamilton and Kennedy, 1986) and 0.074 (Kennedy *et al*., 1991) respectively. In a study done by Kennedy *et al*. (1991), considerably less differentiation among wolves from different regions ($\mathbf{F_{ST}}$ = 0.029) was found. But, the $\mathbf{F_{ST}}$ value for dogs in this study (0.090) falls within the moderate range of differentiation of 0.05 to 0.15 (Hartl, 1980; Nei, 1987; Hartl and Clark, 1989).

The relatively small $\mathbf{F_{IS}}$ value of 0.021 approaches zero, which – in natural populations – is indicative of random mating within a subpopulation (Nei, 1986). The total inbreeding coefficient estimate ($\mathbf{F_{IT}}$) is 0.108 and quantifies inbreeding due to population subdivision and is an indication of minimal barriers to gene flow between the populations studied. The inbreeding in the individuals relative to the total population is fairly small and this could also be an indication of little genetic drift among the populations.

Although comparisons between the pairwise $\mathbf{F_{ST}}$ values for the dogs in this study and those for cattle breeds (0.035 – 0.069) (Kotze *et al.*, 2000) yielded similar results, the pairwise $\mathbf{F_{IS}}$ and $\mathbf{F_{IT}}$ values were much lower in the present study than those reported for cattle, where $\mathbf{F_{IS}}$ ranged from 0.840 to 0.857 and $\mathbf{F_{IT}}$ values were between 0.849 and 0.862 (Kotze *et al.*, 2000). Of all these comparisons, the one that seems to be the most significant to discuss, is the $\mathbf{F_{ST}}$ value. The 'overall' $\mathbf{F_{ST}}$ value is 0.09, which indicates a significant amount of divergence between the dogs in this study. These high, pair-wise $\mathbf{F_{ST}}$ values (Table 7) consistently occur between the Jericho group and all the other populations: Jericho/Africanis = 0.098, Jericho/SPCA = 0.069 and Jericho/Saluqi = 0.080. In contrast to this, the pair-wise $\mathbf{F_{ST}}$ values between Saluqi/Africanis and Saluqi/SPCA are 0.014 and 0.022 respectively, which would suggest that the Jericho population stands out and is the furthermost removed from the rest. Although the precise, numerical value of all the F-statistics is far less important than the trend it suggests, it is interesting that a low value such as 0.014 for the pair-wise $\mathbf{F_{ST}}$ between Africanis and the Saluqi breed confirm their close genetic relationship.