Chapter 1

General introduction

1. Introduction

1.1 The Genus *Syncerus*

The African buffalo (*Syncerus caffer*, Sparrman 1779) consists of three recognized subspecies, two savannah subspecies, *S. c. caffer*, *S.c. bracyceros* and a forest form *S.c. nanus*, that are morphologically quite distinct, with horn configuration, coat colour and size being the primary indicators of a particular phenotype (Grubb, 1972, Mloszewski, 1983). The Cape buffalo *S. c. caffer* or southern savanna buffalo is found on the savannahs of eastern and southern Africa and is the most widely distributed subspecies. It has a shoulder height of more than 1.6 metres, may weigh in excess of 835 kilograms (Mloszewski, 1983) and its coat colour is predominantly black. Male buffalo continue growing in size until they are about 9-10 years old, and their horns continue developing until at least 7-8 years of age, (Grimsdell, 1973; Sinclair, 1977; Jolles, 2004). In terms of longevity, buffalo seldom live beyond 20 years, although in extreme cases animals have reached the age of 25 years (Grimsdell, 1973). The smallest member of the three subspecies, found in the rain forests of western and central Africa, is *S.c. nanus*, also referred to as the forest, dwarf or Congo buffalo. It stands just under a meter at the shoulder and weighs less than 300 kilograms, while its coat colour ranges from red to black (Mloszewski, 1983; Roure and Ledger, 1968, as cited by Spinage and Brown, 1988). *S.c. brachyceros* also referred to as the Sudan or West African savannah buffalo is intermediate and occurs in the Sahel-Sudan savannahs (Buchholtz, 1990; Kingdon, 1997), and the size of an adult male of this subspecies is comparable to that of a female Cape buffalo (Mloszewski, 1983). Some authors recognize a relict 'mountain buffalo' subspecies *S.c. mathewsi* (Kingdon, 1997), whilst others recognize a third savannah subspecies, *S.c. aequinoctialis* from central Africa (East, 1999), but these remain contentious subspecific designations. The three formally recognized African buffalo subspecies exhibit karyotype variability (2n = 52-54); with *S. c. caffer* having 52

chromosomes and *S. c. nanus* 54 chromosomes. Whilst the differences in chromosome number support their sub-specific status (Hsu and Benirschkem, 1970; Gallagher and Womack, 1992), hybrids between Cape and forest buffalo, are known to occur in captivity (Cribiu and Popescu, 1980).

1.2 Distribution and habitat

African buffalo occur in habitats ranging from dense lowland tropical forests, savannah grasslands, and tropical savannah woodland to dry bush (Simonsen *et al*., 1998). Historically, African buffalo (*S. caffer*) inhabited nearly the whole of sub-Saharan Africa where they constituted up to 35 % of the large herbivore biomass (Prins, 1996). Their geographical distribution has been affected by historical climatic fluctuations on the continent, much like other mammals such as hartebeest (*Alcelaphus buselaphus),* topi (*Damaliscus lunatus)* and wildebeest (*Connochaetes taurinus*), to name a but a few (Arctander *et al.,* 1999; Flagstad *et al.,* 2001; Van Hooft *et al.,* 2002). Of the more than 3 million buffalo that once roamed the African continent, only about 400000 animals were left by the 19th century (Lessard *et al.,* 1990). In 1998, it was estimated that there were approximately 111900 (±10000) Cape buffalo in southern Africa (excluding Mozambique, Winterbach, 1998), and today the largest concentrations of the Cape buffalo are found in Tanzania, Zimbabwe, Zambia, South Africa and Uganda, within designated conservation areas. The largest free-ranging populations occur in northern Namibia (Caprivi area), the Okovango area in Botswana and the Matetsi–Hwange area in Zimbabwe (Winterbach, 1998). In South Africa, the two largest populations are confined to the Kruger National Park (KNP) and Hluhluwe-ImFolozi Park (HiP).

1.2.1 KNP

KNP is one of the largest conservation parks inside the borders of South Africa. It comprises close to 2 million hectares and is approximately 350 km from north to south and on average about 60 km wide (Fig. 1.2). The park, situated on the north-eastern border of the country and formerly known as the Sabie Game Reserve, was proclaimed in 1898

(Mabunda *et al*., 2003). Two climatic transitional zones span across the park, *viz*. a tropical to subtropical northern region, and a more temperate southern region. The average long-term annual rainfall for the entire park is 530 mm, of which most is received from October to March (Mabunda *et al*., 2003). The vegetation of the park is classified as sub-arid to arid wooded savannah, with the exception of the wettest part of the park which is the southwest (Pretoriuskop area, Mabunda *et al*., 2003). Today the park is home to, amongst others, 147 species of mammals, 505 species of birds and close to 2000 plant species (Mabunda *et al*., 2003)

At the time of proclamation the park was in essence unfenced, which allowed the free movement of animals across its geographic borders. Fencing was only initiated and completed much later, and comprised different phases. The southern boundary along the Crocodile river was fenced by 1959, the western boundary by 1961, the eastern boundary by 1976 and the northern boundary by 1980. These fences subsequently served as barriers that restricted historical and seasonal migration of large mammals (Mabunda *et al*., 2003). The fences, particularly those in the far northern parts of the park were however often obliterated by floods, resulting in the unrestricted movement of animals and their diseases, into and out of the park. The unrestricted movement, of buffalo in particular, is of particular concern from an animal health perspective as it has been shown that buffalo-precipitated outbreaks of foot-and-mouth disease in cattle have occurred at times when fences were compromised (Vosloo *et al*., 1992; Vosloo *et al*., 2002). Parts of the western fence that separate KNP and private reserves have been dropped in order to permit the free movement of wild animals within this region (De Vos *et al*., 2001). Separation of wildlife and livestock is however maintained by outer buffalo-proof fences that are regularly inspected for breaks. The eastern border of KNP has subsequently been opened up in order to integrate the park into the Great Limpopo Transfrontier Park, which links KNP with Gonarezhou National Park in Zimbabwe and the Limpopo National Park in Mozambique.

The buffalo population of the KNP presently consists of approximately 28500 animals (Ian Whyte, 2005, personal communication), comprising 100 herds, with an average herd size of 244 individuals. The latter may however vary depending on the availability of water and food. Dramatic increases in buffalo numbers, due to increased reproductive success under favourable conditions were recorded in the late 1960's to early 1980's (de Vos *et al*., 1983).

Other marked increases in buffalo numbers include the remarkable recovery of buffalo following the rinderpest pandemic at the turn of the 19th century. From an estimated 20 individuals in 1902 (Stevenson-Hamilton, 1957), a staggering 15758 were recorded in 1967 (de Vos *et al*., 1983) when the first complete aerial census of buffalo in KNP was carried out (Owen-Smith and Ogutu, 2003). This prompted park managers to initiate population control measures in order to stabilize population numbers and prevent damage to the ecosystem and resulted in the culling of 25857 buffalo between 1967 and 1981. Remarkably however, their numbers increased to an astounding 34912 by 1981, despite this drastic intervention. It was estimated that the buffalo population would have reached more than 89 000 if these measures had not been implemented (de Vos *et al*., 1983). Subsequent to these interventions, the population size fluctuated, with a dramatic population crash occurring from 1992 to 1995 due to a severe drought which reduced buffalo numbers from 29359 to 14123 (De Vos *et al.*, 2001).

1.2.2 HiP

The Hluhluwe and Imfolozi game reserves were proclaimed in 1895, and together with the corridor area that connects them, today constitute HiP. Records show that the fate of the Imfolozi park often hung in the balance with the park being proclaimed and then de-proclaimed several times between 1930 and 1950 (Brooks and Macdonald, 1983). HiP is located in the province of KwaZulu-Natal and falls within the Zululand thornveld subcategory of the coastal tropical forest. The altitude of the park ranges between 60 and 750 metres above sea level and it has a unimodal rainfall that peaks in midsummer (December to February) (Brooks and Macdonald, 1983). HiP is significantly smaller in size than KNP, and it comprises approximately 96000 hectares. The entire park was completely fenced in the late 1960's to early 1970's. Prior to fencing of the Hluhluwe area, which commenced during 1941, animals could move freely in and out of this area. Little movement of animals is however believed to have taken place due to human settlement and activities, and the anti-nagana campaigns adjacent to the area (Brooks and Macdonald, 1983). The park also experienced extensive poaching prior to 1929, prompting park officials to adopt a firm anti-poaching attitude (Brooks and Macdonald, 1983). Today the park is home to a vertebrate fauna of more than 500 species (KZN Wildlife, 2002) and has

a carrying capacity in the order of 5000-6000 for buffalo (depending on rainfall patterns). Census records reveal that in 2002 HiP had 3440 buffalo. Prior to this, buffalo numbers were significantly depressed due to extensive anti-Nagana removals that took place during 1930 and around 1942 to 1950. The Hluhluwe area however largely escaped these campaigns (Brooks and Macdonald, 1983). Like KNP, HiP did not escape population control measures, and 3949 animals were removed or culled between 1982 and 1994 in order to restrict population growth, which may have led to the perturbation of the ecosystem. The buffalo population responded rapidly to the population control measures, especially during the early 1980's, but less so to the control measures implemented between 1967 and 1982, during which time an increase from 1377 to 1903 was recorded (Brooks and Macdonald, 1983).

1.3 Ecology and behaviour

Cape buffalo are gregarious animals that occur in mixed herds of up to several thousand (Hwange, Zimbabwe: 2500; Savuti, Botswana: 3000). The herds have well defined home ranges which may vary between 40 and 1455 km² and which have been shown to be herd-size dependant (Sinclair, 1977; Prins, 1996). The home ranges seldom overlap with that of other herds (Funston *et al.,* 1994). Jolles (2004) showed that home ranges of the HiP buffalo were stable throughout both the wet season and dry seasons; although in the dry season the home range may be expanded to some extent. Still, following expansion, the herd would stay centered in the same area as in the previous wet season. Neighbouring herds also remained neighbours, and fragmented herds reconvened to form the same breeding herd each year, despite dry season changes (Jolles, 2004). Herd sizes for buffalo in HiP range between 30 and 250 animals, with most herds numbering around 100. The average herd size for KNP is 244. Remarkably, a herd may survive as a single recognizable entity for up to 35 years (Prins, 1996).

Herding behaviour, movement and size are largely dictated by the availability of food and water (Hunter, 1996), and fluctuations thereof result in herds responding with fission-fusion behaviour, splitting up and reconvening again in search of food and water. Availability of food furthermore may result in buffalo adjusting from general non-selective bulk grazers

(Sinclair, 1977; Prins, 1996) to selective feeders, depending on the quantity and quality of grass (Sinclair, 1977; Macandza *et al*., 2004). This adjusting behaviour is particularly prevalent during the critical periods within the dry season.

The behaviour of the two sexes is quite differentiated. Like many mammal species, female buffalo are philopatric, while the males are the dispersing sex. Contradicting results have however been demonstrated for different study areas. In Manyara for instance, females (making up the vast majority of mixed herds - in excess of 70 %) have never been observed outside their native herd, showing extreme fidelity to their home range over many generations (Prins, 1996). In a recent study in northern Botswana however, using telemetry, females have been shown to frequently switch between herds, travelling distances of up to 133 kilometres. These females also never returned to their native herd again (Halley *et al*., 2002). Females also display recognisable bonding affections towards one another, and form distinctive familial cohesions that last into adulthood (Prins, 1996, Jolles, 2004). In HiP, herd membership of females and immature animals has been shown to be stable from year to year (Jolles, 2004). Bonding relationships between females and subadult males are however terminated when the males reach the age of about three. Females also exhibit a structured hierarchy in the herd, and their position is correlated with body condition. Animals in the front of the herd display the best body condition, while body condition deteriorates towards the back of the herd (Prins, 1996).

Male behaviour is apparently attributed to a combination of seasonal mating opportunities, foraging preferences (and seasonal availability thereof) and predation avoidance (Sinclair 1977; Prins, 1996; Halley and Mari, 2004). Bulls tend to stay in mixed herds until their body condition starts to deteriorate, after which they leave the herd to join bachelor groups and to regain condition, following which they rejoin the mixed herd again. Bulls are also often encountered wandering between herds, presumably looking for oestrous cows (Jolles, 2004). Batchelor herds may consist of up to 12 adult and subadult individuals. In HiP, adult males have been shown to spend as much as 77 % of their time in male groups, and only between 23 and 34 % in mixed herds throughout the year (Jolles, 2004). A higher fraction of males have been shown to leave the breeding herds in the dry season (Jolles, 2004). The herding behaviour of the males has also been implicated as playing a major role in inter-herd disease transmission. (Jolles, 2004).

Adult bulls are sexually mature at approximately five and a half years of age, although some reach sexual maturity earlier (three and a half years of age). The older dominant bulls however, seldom, if ever, allow them to mate until they are 7–8 years old (Prins, 1996). A female in oestrus may be served by several adult bulls (Pienaar, 1969a). Perhaps not unexpectedly, the dominance among the adult bulls in large herds is less well defined than in the smaller herds, while female dominancy may not exist at all in herds (Sinclair, 1977). In any case, neither individual females nor males have been shown to be leaders of a herd (Prins, 1996).

In KNP, the majority of calves are born between January and April, and the calving season peaks in January/February (Fairall, 1968; Pienaar, 1969a). This coincides with the optimum period of the growth of the grass and its nutritional value with regard to protein content. Females may have their first calf between 4 and 5 years of age (Mloszewski, 1983), and the mean gestation period is 343 days (Knechtel, 1993).

1.4 Buffalo and disease

Buffalo are host to a variety of subacute diseases, of which foot-and-mouth disease (Anderson *et al*., 1979; Keet *et al*., 1996), corridor disease and Bovine Tuberculosis (Bengis *et al*., 1996; Cooper, 1998), abbreviated FMD, CD and BTB, respectively, are arguably the most important. The animals are also prone to other infections such as parafilariosis, of which the prevalence of the infection is approximately 34 % within KNP (Keet *et al*., 1997). The importance of disease within South Africa's buffalo populations only becomes evident when you consider that only 7.7 % of all buffalo in SA are estimated to be disease-free (Winterbach, 1998).

Disease is probably one of the most important factors affecting buffalo populations throughout the African continent during the past century. It is also largely responsible for the fact that buffalo are mostly kept in enclosed areas such as conservancies, and that strict regulations are imposed on their movement. This subsequently affects their population dynamics, which is central to this study. A classical example of the impact that disease may

exert on a population is the rinderpest epidemic of 1889-1897. It was responsible for a catastrophic decline of up to 95 % of buffalo in South- and southern Africa (Plowright, 1982; Shigesada and Kawasaki, 1997; O'Ryan *et al*., 1998; Barrett and Rossiter, 2000).

Together with over hunting and livestock ranching, disease has also played a major role in the displacement of buffalo from many of their former ranges where they used to occur naturally (Hofmeyr, 2005). In order to address this phenomenon, the re-introduction of buffalo into newly established conservancies and their former "home ranges" is gaining momentum. Paradoxically, it is disease that is now hampering re-introduction efforts, since strict control measures are imposed on the movement to or from areas where certain diseases such as FMD, CD and BTB prevail. This has sparked a demand for disease-free buffalo and their subsequent establishment in disease-free areas. The demand has already culminated in the initiation of both privately owned and state-sanctioned projects aimed at the breeding of disease-free buffalo in South Africa (Hofmeyr, 2005).

1.4.1 Foot-and mouth disease (FMD)

Foot-and-mouth disease (FMD) is a contagious disease that affects cloven hoofed animal species, including more than 22 members of the Bovidae alone (Hedger, 1981; Thomson, 1994). The disease is caused by a Picornavirus and is characterised by a high morbidity and a low mortality (Thomson, 1994). Symptoms of the disease, which develop two to eight days after infection, include dullness, loss of appetite and fever. In dairy cows milk production is reduced as a result of the infection. These symptoms are soon followed by lameness, unwillingness to stand, salivation, smacking of the lips and development of lesions and vesicles on the tongue (Thomson, 1994). The latter are easily ruptured resulting in a raw and bleeding surface. Lesions also develop on the feet, specifically at the interdigital space and the bulbs of the heel (Thomson, 1994). Myocarditis may also affect young animals and may result in death (Thomson *et al.*, 2003). Wildlife is only mildly affected by FMD, and diseased animals usually recover within a week or two (Thomson *et al.*, 2003). The disease may however be more severe in some wildlife species such as impala (*Aepyceros melampus*) and even fatal in the mountain gazelle (*Gazella gazella*) (Macaulay, 1963; Shimshony *et al*., 1986). The African buffalo, although rarely developing clinical

signs of FMD under natural circumstances, is recognised as a major reservoir and host species of the SAT-type FMD viruses endemic to sub-Saharan Africa.

Although the control of FMD is considered of lesser importance due to the prevalence of more acute and lethal diseases, FMD may severely affect the livestock industry. It poses a particular threat to intensive livestock production systems, although under extensive conditions it is considered relatively benign (Thomson *et al.*, 2003). The mere presence of the disease however has a major impact on the livestock industry, and strict control measures are imposed by countries to prevent the importation of animals or animal products from areas where FMD occurs (Thomson *et al.*, 2003). The seriousness and magnitude of FMD is reflected by the fact that some outbreaks have resulted in the spread of the disease across continental borders (Thomson *et al.*, 2003). Although FMD has been eradicated in many parts of the world, it is still prevalent in South America, most African countries, the Middle East, and many parts of south, central and south-east Asia, and several major outbreaks have occurred in these countries in the recent past (Thomson *et al.*, 2003). In South Africa, the KNP is the only locality where FMD has persistently been reported in wildlife during the past 60 years and where numerous outbreaks have been reported (Thomson *et al.*, 2003).

The importance of infected wildlife species lies in the fact that they play an important role in the transmission of the disease to livestock. FMD may be spread by mechanical means through contaminated animal products or by people that are contaminated with the virus. Physical contact between infected and susceptible animals seems to be the major mechanism of transmission, while aerosols also play a role (Thomson, 1994). Transmission through the latter system over long distances is however rare (Gloster *et al*., 1982). Both acutely infected buffalo and carriers of FMD are known to transmit the disease to other species that they come into close contact with (Hedger and Condy, 1985; Gainaru *et al.*, 1986; Hunter *et al.*, 1996; Thomson, 1996; Vosloo *et al.*, 1996), and buffalo are the only wildlife species able to transmit FMD during the carrier state of the disease (Dawe *et al*., 1994a,b; Vosloo *et al*., 1996).

African buffalo are host to the SAT (South African Territories) type viruses of which three types are known, *viz*. SAT1, SAT2 and SAT3 (Hedger *et al*., 1972). These viruses have

been shown to persist in an individual buffalo for more than 5 years, and in an individual herd for up to 25 years (Condy *et al*., 1985). Based on the widespread distribution and genetic diversity of these viral types and their apparent harmonious association with the buffalo, it has been suggested that the animals have been associated with the viruses in Africa since prehistoric times (Hedger, 1976; Thomson, 1994; Bastos, 2001; Bastos *et al.*, 2001). The three SAT type viruses are only found in sub-Saharan Africa, while types O, A and C (also referred to as the European types) are prevalent in South America, Europe, Africa, the Middle and Far East, and the Balkans (Thomson, 1994). The virus types also exhibit intratypic diversity, and the different variants within a type are termed topotypes. The different topotypes have been shown to occur and evolve in particular geographic localities (Hunter *et al.*, 1996; Bastos *et al*., 2001), and these characteristics have been used to trace the source of FMD outbreaks and the origin of illegally moved animals that are infected with FMD (Vosloo *et al*., 2002a,b).

One of the major approaches for the control of FMD in most southern African countries is to separate livestock from infected buffalo through fencing and to control the movement of animals between infected and uninfected areas. Vaccination, whilst important, has been shown to be relatively inefficient due to a) the fact that immunity following primary vaccination is short-lived and b) the occurrence of antigenic diversity. With regard to the latter, it has been shown that animals that have recovered from infection with a particular virus type are still susceptible to re-infection by other types. This subsequently necessitates different vaccine strains for immunization against the different topoptypes (Hunter *et al.*, 1996; Thomson *et al.*, 2003). Another important factor complicating vaccination approaches is that information pertaining to the prevalence and distribution of the different topoptypes in different regions is still lacking for some areas/regions on the continent. It is thus vital to execute detailed molecular epidemiological studies in this regard in order to embark on efficient immunization programmes (Bastos *et al.*, 2001; Samuel and Knowles, 2001; Sangare *et al.*, 2001)

1.4.2 Corridor disease (CD)

Theileriosis is one of the major vector-borne-complex of diseases that affect both livestock and wild animals in sub-Saharan Africa. In fact, most African wild Bovidae are believed to be carriers of *Theileria* parasites, although the extent to which they affect the respective species and their host vectors is not known (Grootenhuis, 1989). Corridor disease (CD) forms part of this complex and its name is derived from the corridor area that connected the Hluhluwe and Imfolozi Game reserves in the Kwazulu Natal Province of South Africa where the disease was first diagnosed in the early 1950's. CD is caused by the protozoa *Theileria parva lawrencei* and is transmitted by the brown ear tick *Rhipicephalus appendiculatus (*Young and Purnell, 1973; Young *et al*., 1977; Nene *et al*., 1999). These ticks feed on both cattle and wild animals and are central in transmitting the disease within and among different animal species. The disease poses a serious threat to the livestock industry due to its high rates of morbidity and mortalities among livestock. Theileria-associated mortalities in 1989 amounted to the death of more than 1.1 million head of cattle in the African region (Mukhebi *et al*., 1992). In countries where CD is still prevalent such as South Africa, strict quarantine measures are imposed on populations of infected buffalo in order to contain the spread of the disease (Uilenberg, 1999). Another well-known disease that falls within the Theileriosis complex is East Coast Fever (ECF).

Clinical signs of CD, which has an incubation period of only 5 to 9 days, include a rise in body temperature, swelling of the lymph nodes, fever, listlessness, swelling of the eyelids, nasal discharge, diarrhoea, difficult breathing and dyspnea (Young *et al*., 1973, Irvin and Mwanachi, 1983; Dolan *et al*., 1984). The African buffalo (*S. caffer*) is a reservoir host of *Theileria parva lawrencei* (Potgieter *et al*., 1988), although *T. taurotragi, T. velifera* and *T. mutans* have also been isolated from buffalo, all of which are transmissible to cattle. Of these, *T. p. lawrencei* is the only species that is highly pathogenic for cattle (Grootenhuis, 1989). Other wild bovid species such as the blue wildebeest (*Connochaetes taurinus*), eland (*Tragelaphus oryx*), impala (*Aepyceros melampus*) and waterbuck (*Kobus defassa*) have been shown to be naturally infected with *Theileria* species. (Burridge, 1975), but are unlikely to play a significant role in the epidemiology of the disease as transmission of *Theileria* from impala and wildebeest to cattle could not be demonstrated under experimental conditions (Purnell *et al*., 1973; Grootenhuis *et al*., 1975).

Although buffalo and cattle are phylogenetically closely related, the parasite is non-pathogenic in buffalo but lethal to cattle (Brocklesby and Vidler, 1966). In addition to buffalo being long-term carriers of the parasite (Burridge, 1975), they also have very high infection rates, with surveys conducted in Kenya, Tanzania and Uganda revealing that almost every buffalo sampled was a carrier of *T. p. lawrencei* (Young *et al*., 1978a). It is believed that African buffalo of the sub-Saharan region lived in harmony with *T. parva* and its vector *Rhipicephalus appendiculatus* prior to the introduction of cattle into the region (Grootenhuis, 1989) when large scale deaths among cattle from theileriosis, alerted researchers to the presence of this disease agent. It is also believed that cattle gradually developed resistance and tolerance to the disease after prolonged exposure (Burridge, 1975) and it has been shown that cattle that survive *Theileria parva* infections acquire lifelong immunity against the disease (Burridge *et al.*, 1972).

CD may be passed from buffalo to cattle by infected ticks in the event that cattle and buffalo come into close contact, but not vice-versa as erythrocytic piroplasms in cattle are usually absent or too scanty to infect new ticks, rendering the disease is self-limiting in cattle (Uilenberg, 1999). Since ticks play such a central and pivotal role in the transmission of CD and other tick-borne diseases (TBD's) that form part of the Theileriosis complex, control strategies in many African countries are based mainly on the use of acaracides (Mavale, 1996). The development of resistance to acaracides in certain regions of the African continent has however necessitated the development of new generation chemotherapeutics such as organochlorines, organophosphates, carbamates, amidines and synthetic pyrethroid acaricides. In South Africa there is no official national or provincial programme for the control of ticks or TBD's, with the exception of the Kwazulu-Natal province (Potgieter, 1996), and it is thus the sole responsibility of the farmer to effect tick control.

A further measure aimed at reducing the risk of transmission from wildlife to domestic livestock is simply to exclude livestock from areas where infected wildlife may share the same grazing environment (Norval *et al*., 1992a). The third approach to CD control is through vaccination. The latter is accomplished by infecting animals with the disease, which is followed by treatment of the infection with therapeutics. Prior knowledge on the particular strains of the *Theileria* parasite that occurs in the infected region is however a prerequisite, as antigenic diversity of *T. p. lawrencei* from buffalo has been demonstrated in

several studies (Young *et al*., 1977, 1978b; Conrad *et al*., 1987; Grootenhuis *et al*., 1987a; Conrad *et al.,* 1988) and a single animal may harbour as many as four or five antigenic types (Conrad *et al*., 1987). Another factor that may impede the efficiency of vaccination is one of logistics as a cold-chain needs to be maintained which in many instances is not possible in rural Africa (Young *et al.*, 1978).

1.4.3 Bovine Tuberculosis (BTB)

BTB, caused by *Mycobacterium bovis*, has received considerable attention during the last decade due to its prevalence in buffalo in both KNP and HiP (Bengis *et al*., 1996; Cooper, 1998) and the potential risk that it poses to the cattle industry. *Mycobacterium bovis* is furthermore genetically very closely related to *M. tuberculosis*, and humans can also contract the disease from animals (Huchzermeyer *et al*., 1994; Daborn, 1995). A variety of other wild animal species have also been shown to act as hosts for the disease (Keet *et al*., 1996). The first diagnosed case of BTB in free-ranging African buffalo was in 1963 in Uganda (Guilbride *et al*., 1963), and the disease is believed to have entered KNP during the same time period (De Vos *et al.*, 2001). Positive identification of the disease agent was however only confirmed in 1990, in a 2-year-old bull that was physically in a very grave state (Bengis *et al*., 1995). This positive diagnosis coincided with outbreaks in the cattle population bordering the south of the park, and the bacterial genotype causing that outbreak was shown to be the same as that identified in the buffalo (Vosloo *et al*., 2001). The latter lead to the postulation that livestock were the source of infection that subsequently resulted in the widespread prevalence throughout the study area (Bengis *et al*., 1996). The first longitudinal study reporting BTB prevalence in the KNP (Rodwell *et al*., 2000) revealed that the south of the park had an average herd prevalence of 38 %, while in the central and northern regions the prevalence levels were 16 % and 1.5 %, respectively. Some herds from the southern area of the park contained as many as 90 % infected individuals. It was subsequently estimated that the disease spreads at a rate of approximately 6 km per year in a northerly direction.

The impact that BTB has on buffalo has been well-studied in KNP (Caron *et al*., 2003) and HiP (Jolles, 2004). In KNP, BTB may increase the mortality rate of the older age classes in the high-prevalence southern region of KNP. A positive correlation was demonstrated

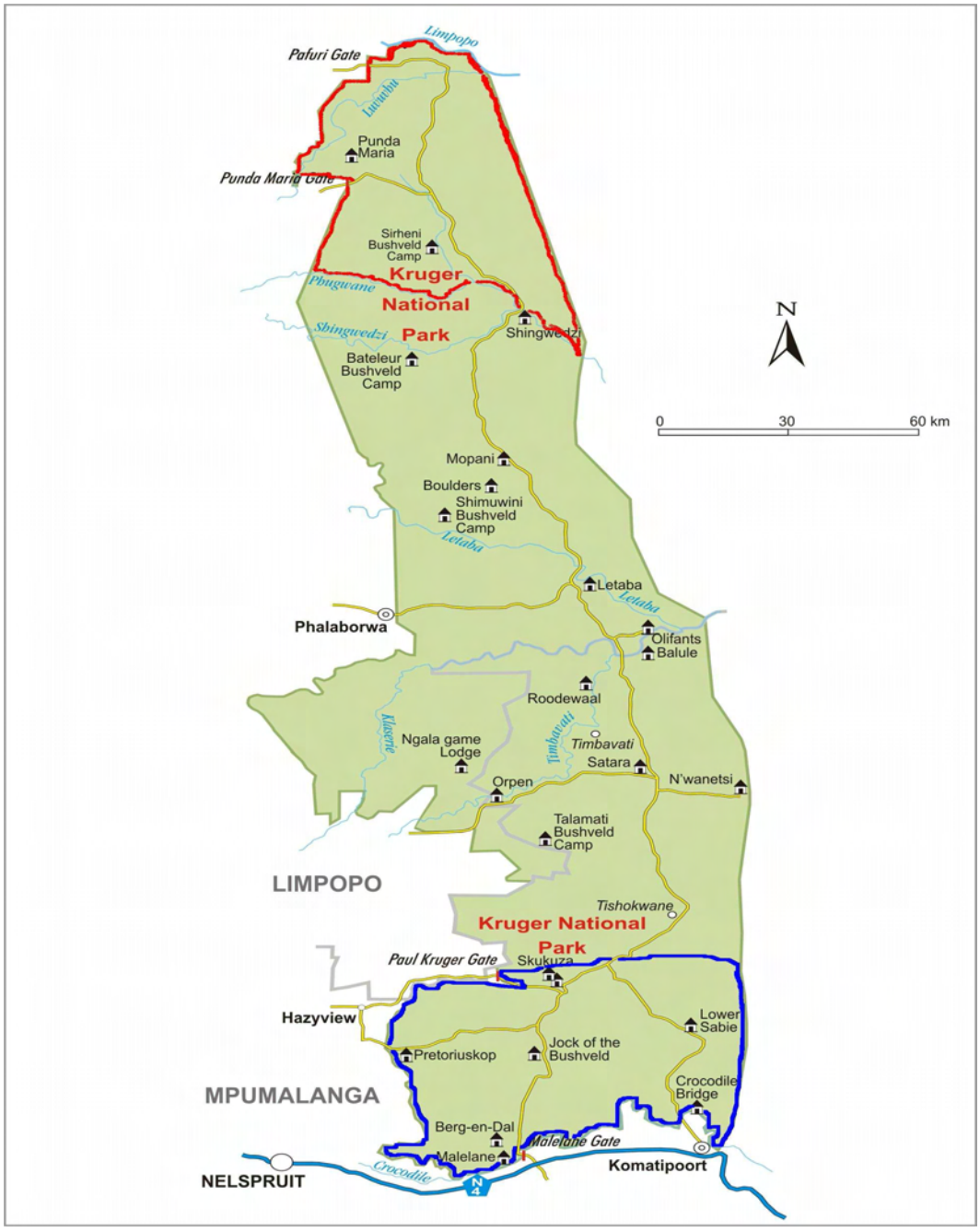
between prevalence and body condition, with high prevalence herds having the worst overall body. The condition of animals from these herds furthermore deteriorated faster during the dry season compared to herds of lower prevalence. With regard to clinical manifestations, acute symptoms are only associated with advanced cases showing characteristic histopathological signs, and a positive correlation exists between increasing prevalence values and the number of advanced BTB cases with potential mortality (De Vos *et al*., 2001). Lions have been shown to selectively kill buffalo weakened by BTB and may contract the disease in the process (Caron *et al*., 2003).

In HiP, the first case of BTB was recorded in 1970 in a black rhino, although it was never confirmed by bacterial culture techniques. A positive diagnosis was however made in a buffalo in 1986 (Cooper, 1998). BTB is also thought to have entered HiP in the early 1960's, prior to the construction of a fence around the park (Dale 1998). The mode of transmission is believed to have been from cross contamination from cattle. The prevalence in HiP, as in KNP, varies among regions. Jolles (2004) reported a prevalence of 2.2 % for Nqumeni, 15.1 % for Manzibomvu and 53.3 % for the Masinda area. However, in contrast to KNP, there is no south-north gradient in HiP, as Nqumeni, which is flanked by high prevalence Masinda to the south and Manzibomvu to the north (Fig. 1.1), has the lowest infection rate, rather than an intermediary one. The disease has been shown to affect both fecundity and adult survival in HiP. By using statistical modelling, Jolles (2004) showed that BTB may significantly reduce population growth rate and may also compromise resilience of the buffalo to disturbance. Modelling results furthermore indicated that the annual mortality rate due to the disease may be as high as 11 %. The disease does not have a strong effect on population age structure in HiP, although it is most prevalent among adult male buffalo.

Figure 1.1 Geographical map of HiP, showing the five different management areas of the park

Figure 1.2 Geographical map of KNP. The northern and southern areas are indicated by the red and blue border lines respectively

.



1.4.4 Rinderpest

Certainly one of the most devastating diseases ever to have affected Cape buffalo was rinderpest or black African cattle plague (Rossiter, 1994). It is an acute viral disease of Artiodactyls and is characterized by necrosis, pyrexia, fever and erosions in the gastro-intestinal (GI) tract that result in severe diarrhoea and dehydration (Rossiter, 1994). The disease is caused by an RNA morbillivirus, a member of the subgroup paramyxoviridae. The latter includes the virus that causes measles in humans (Warren, 1960). A wide range of vertebrate species are susceptible to the disease, and it is particularly fatal in bovines (Stevenson-Hamilton, 1957). The virus is excreted in the air through the respiratory tract, through nasal and oral secretions and also in the faeces. Airborne transmission occurs when infected droplets are inhaled and the virus penetrates the mucosa of the upper respiratory tract of an animal. Transmission also occurs during close contact between infected and susceptible animals.

Rinderpest is believed to have originated in Asia, from where it spread to Europe during wars (Rossiter, 1994). It eventually entered north Africa via Eritrea between 1887 and 1889 through infected Indian and Arabian cattle that the Italian army brought with them (Rossiter, 1994). From Eritrea, the disease subsequently spread to East and southern Africa, embarking on an unprecedented path of destruction. It eventually resulted in the eradication of more than 5 million cattle south of the Zambezi River (Rossiter, 1994), and in South Africa it was estimated to have killed more than 95 % of the buffalo population (Stevenson-Hamilton, 1957). A mere 20 animals were reported to have survived in KNP in 1902 (Stevenson-Hamilton, 1957). The pathogenicity was so severe that some authors estimated that for every 10000 killed by the disease, only one buffalo survived (Roosevelt and Heller, 1914, as cited by Estes, 1991). The gregarious nature of buffalo and their intimate social and behavioural characteristics probably enhanced the transmission and spread of the disease. Other wildlife species were differentially affected. Eland were also almost eradicated entirely, while impala (*Aepyceros malampus*), blue wildebeest (*Connochaetes gnou*), sable (*Hippotragus niger niger*) and roan (*Hippotragus equinus*), tsessebe (*Damaliscus lunatus*) and waterbuck (*Kobus ellipsiprymnus*) were abundant during the same period (Stevenson-Hamilton, 1957). The total eradication of the tsetse fly in certain regions of Africa has been postulated to be an indirect effect of rinderpest (Stevenson-Hamilton, 1957).

Remarkably, the buffalo populations recovered within the two decades following the epidemic and animals were abundant in areas where they were hardly seen just after the epidemic subsided (Estes, 1991, as cited by Simonsen *et al*., 1998). Sinclair (1977) attributed the dramatic recovery to their reproductive success. Animals that survive the disease usually acquire permanent immunity (Scott, 1981, as cited by Prins, 1996), and an effective vaccine has since been developed against the virus (Barret and Rossiter, 2000).

1.5 Molecular markers as tools for population genetic studies

It is not within the scope of this thesis to review the large body of literature dealing with the application of molecular tools in population genetic studies. Their characteristics and modes of evolution however warrant a brief review since it has bearing on the calculated parameters and the interpretation thereof. The population genetic study carried out in this thesis is based on two molecular markers, firstly on nuclear microsatellite DNA and secondly on mitochondrial DNA sequences. These markers are widely used and are considered to be suitable targets for unraveling factors that affect the genetic status of a species.

1.5.1 Microsatellites (Msats)

1.5.1.1 Msat characteristics

Microsatellites are short tandemly repeated nuclear DNA sequences, usually 1–5 bp in length, such as (CA)n or (ATT)n, that are most abundant in the non-coding part of the genome (Beckmann and Weber 1992). They are relatively evenly spaced with trimeric and tetrameric repeats occurring at a rate of one in every 10 to 20 kilobases on the X-chromosome (Edwards *et al*., 1992). These markers are also selectively neutral in most instances, making them compatible with the assumptions of population genetic theory. The sequence of the repeat units arise and mutate predominantly through slippage synthesis during DNA replication (Schlotterer, 2000), although unequal crossing over during meiosis

has also been shown as a mode of mutation (Levinson and Gutman, 1987a). Msats are inherited in a co-dominant Mendelian fashion, and heterozygosities for loci commonly exceed 70 % (Webster *et al*., 2002). Another very beneficial characteristic of Msats is that they are relatively easily amplified through the polymerase chain reaction (PCR), which means that not only can very little starting material be used, but non-invasive samples (such as hair and excrement) and samples that are partially degraded can be analysed (Paabo, 1989), and that the analysis is also amenable to automation.

Probably the most characteristic feature of Msats is their high mutation rates, resulting in high levels of polymorphism for these markers. Mutation rates of up to 10^{-2} per generation have been reported (Jarne and Lagoda, 1996, as cited by Estoup and Angers, 1998), which is up to 10000 times higher than that of nuclear genes (Ritz *et al*., 2000; Schlotterer, 2000) and 1000 times higher than that of mitochondrial genes. The mutation rate has also been shown to be a function of repeat length and base composition (Rubinstein *et al*., 1995). Dinucleotide repeats for instance mutate faster than tri nucleotide repeats, while AT-rich sequences have superior mutation rates compared to repeats characterized by a high GC content (Schlotterer and Tautz, 1992). These super mutation rates result in a large number of alleles being present in most populations, implying that significant variation can be uncovered through Msat analysis, while the genetic relatedness between populations can be assessed, even if they have diverged as recently as 50 to 100 generations ago (Diez-Tascon *et al*., 2000). The sequences that flank the repeat units are often highly conserved, permitting cross-species application of the markers (Moore *et al*., 1991; Schlotterer *et al*., 1991; Mommens *et al*., 1998; Diez-Tascon *et al*., 2000; Cronin *et al*., 2003).

1.5.1.2 Msat evolution and homoplasy

The use of Msats in population genetic studies and phylogenetics is not without its problems and drawbacks, one of the most important being that their mutational process is still not well understood (Callen *et al*., 1993; Ellegren *et al*., 1995; Rubinstein *et al*., 1995; Ramel, 1997; Ellegren, 2000). One of the most frequent assumptions made during calculations is that the loci under study follow a stepwise mutation model or SMM. This model states that mutations involve the gain or loss of a single repeat, and is supported by the observation

that many Msats do in fact mutate in a stepwise fashion (reviewed in Estoup and Angers, 1998; Ellegren, 2000a; Schlotterer, 2000). The SMM model can however lead to homoplasy, i.e. the occurrence of different copies of a locus that are identical in state, although not identical by descent (Kimura and Ohta, 1978). Calculations based on this model will subsequently result in an underestimation of the total amount of variation and genetic distance.

The infinite allele model (IAM) on the other hand states that mutations result in an allelic state not previously encountered in a population, and may involve any number of tandem repeats (Kimura and Crow, 1964). Under the Two phase model or TPM, the state of the mutating allele changes by an absolute number of *x* repeat units (Di Rienzo *et al*., 1994), allowing for mutation that involves the gain or loss of more than a single repeat unit. The generalized stepwise mutation model (GSM) is a simplified version of the TPM mode, whilst the *K*-allele model (KAM, Crow and Kimura, 1970) states that there are exactly *K* possible allelic states and that any allele has a constant probability of mutating towards any of the other allelic states. With the exception of the IAM model, all models are prone to size homoplasy since they allow for mutation towards allelic states that may already be present in the population (Estoup *et al.,* 2002). Msats that mutate according to the IAM model have been suggested to be best suited for assessing population subdivision and genetic relationships (Estoup *et al*., 1995a; 1995b) since they will contain the lowest levels of homoplasy.

1.5.1.3 Homoplasy, F_{ST} and gene flow (Nm)

Since homoplasy may bias genetic distance calculations, it will subsequently affect estimations of gene flow (expressed as number of migrants), since this is directly related to the former. One of the most widely used estimators of genetic distance between populations is F_{ST}, which is easy to estimate from allele frequency data. F_{ST} is frequently used to estimate the level of differentiation between populations as a function of gene flow. F_{ST} is also directly related to genetic drift in the absence of gene flow between populations. It can be defined as the fractional reduction in heterozygosity as a consequence of population subdivision and is described by the following equation:

$(1-F_{IT}) = (1-F_{IS}) \times (1-F_{ST})$

where F_{IS} is the fractional reduction in heterozygosity due to non-random mating within subpopulations while F_{IT} is the fractional reduction in heterozygosity due to both non-random mating within subpopulations and fragmentation between subpopulations, and is a measure of the reduction in heterozygosity of an individual relative to the total population (Hartl and Clark, 1989). F_{IS} is also referred to as the inbreeding coefficient.

F_{ST} reaches migration–drift equilibrium quickly and is not overly dependent on the theoretical demographic model (stepping-stone versus n-island model). Under an infinite-island model, Wright showed that F_{ST} could be used to estimate the number of migrants (Nm) exchanged by the populations. A problem is that small values of F_{ST} cannot be precisely estimated because they correspond to differences in population allele frequencies that are small relative to differences that arise by chance in samples of populations (Weir and Cockerham 1984; Waples, 1998). Small F_{ST} values will also result in a large variance in the estimator of Nm, since the relationship between F_{ST} and Nm is non-linear (Waples, 1998). It is also known that the number of migrants, based on F_{ST} , is artificially high, especially if highly polymorphic loci have been used. F_{ST} values also provide little information on ancient demography.

1.5.1.4 Alternatives to F_{ST}

Alternatives to F_{ST} have been proposed, such as R_{ST}, which compares variance in allelic sizes rather than frequency (Slatkin, 1995). R_{ST} has been projected to be a more appropriate measure of distance for populations that diverged in the very distant past and over a time frame long enough to permit the introduction of new mutations. F_{ST}, on the other hand, as pointed out earlier, may be more reliable and appropriate than R_{ST} for estimating distance between more recently diverged populations where migration and genetic drift are the major factors driving differentiation (Slatkin, 1995). If the populations have a common ancestry and have diverged due to drift alone, F_{ST} and R_{ST} values should however be similar (Slatkin, 1995). This was confirmed by Harley *et al*. (2005) who showed that these two distance

estimates are similar for two populations that were recently separated after a bottleneck, and where drift was primarily responsible for population differentiation. If enough time (generations) has passed for new mutations to take place, F_{ST} will be lower than R_{ST} (Harley *et al*., 2005). It should however be noted that R_{ST} also tends to have a higher variance than other measures of population differentiation and it is sensitive to violations of the stepwise mutation model (Balloux *et al*., 2000; Richard and Thorpe 2001).

Distance measures, such as θ^2 and D_A, which account for the characteristics and evolution of Msats, have also been developed (Nei, 1987). Since the mathematical expression for D_A is simpler than for θ^2, the former is preferred simply for practical considerations (Nei and Takezaki, 1994). Simulated data showed that highly polymorphic loci provided better estimates of these genetic distances than less polymorphic loci (Kalinowski, 2002), while the classical distances of Cavalli-Sforza and Edwards (1967) and of Nei *et al.* (1983) perform better for phylogenetic reconstruction than distances that are based on allele size differences (Takezaki and Nei, 1996). This is essentially due to the lower coefficient of variation and acceptable linearity with time of these classical distances when short periods of divergence are considered.

1.5.1.5 Direct measurement of gene flow and dispersal

While F_{ST} based estimates of Nm are indirect methods for assessing gene flow over various generations, a direct approach is possible with assignment methods that estimate gene flow in the current generation. These methods have also proved valuable for distinguishing between male and female biased dispersal (Goudet, 2001). Assignment methods can be divided into likelihood- and distance-based approaches, and both try to account for allele frequency related difficulties. Likelihood-based methods have been shown to perform better than distance-based approaches. They are based on the calculation of the probability that a multilocus genotype is observed in a population for which allele frequencies are known. The accuracy of assignment methods however depends on the degree to which populations are differentiated, becoming more accurate and powerful when large genetic distances separate populations. Populations separated by small genetic distances generally require more loci to attain high levels of accuracy (Cornuet *et al.,* 1999).

A major advantage of assignment tests is that they are robust to key assumptions about the model of mutation of the markers. Breed identification can be attained with high levels of accuracy (Arranz *et al.,* 2001; Koskinen, 2003), while it has proved useful for traceability studies (Manel *et al.,* 2002), hybrid detection (Comstock *et al*., 2002) and sex-biased dispersal (Goudet *et al.,* 2002). The latter however is significantly affected by a biased intensity in dispersal between the two sexes. If only one sex for instance disperses (100:0), then biased dispersal can be detected most of the time, provided sampling was exhaustive. It however becomes difficult to detect a sex-associated dispersal in the event that both sexes disperse (Goudet *et al*., 2002).

1.5.1.6 Estimating effective population size

Effective population size is one of the most important parameters in population genetics, since it summarizes the effects of demographic parameters in determining a given genetic property of the population (Caballero, 1994). The concept of effective population size (N_e) was first introduced by Wright (1931), referring to the size of an idealised model population that has the same genetic properties as observed for the real population. The value of determining the effective population size is that it provides a number for a population in mutation-drift equilibrium, and it is one of the best predictors of a population's ability to maintain genetic diversity (Harley *et al*., 2005). Several factors are however related to the calculation of N_e. It is a function of the molecular variability within a population (Chikhi and Bruford, 2005), the sex ratio and mating system, selection, pattern of inheritance, changes of the population size over generations and population subdivision (Caballero, 1994). Mutation on the other hand is independent of population size (Bodmer and Cavalli-Sforza, 1976). In certain cases (such as inbreeding), the relationships between N_e and population parameters are unknown which may complicate its calculation (Wang, 2005).

1.5.2 Mitochondrial DNA

1.5.2.1 MtDNA Characteristics

The animal mitochondrial DNA (mtDNA) genome is a small (15-20 kb), circular molecule that contains 37 genes (coding for 22 tRNA's, two rRNA's and 13 mRNAs) that are conserved due to functional and structural constraints (Simon *et al*., 1994). It lacks introns and small intergenic spacers are found where the reading frames sometimes overlap. The control region of the molecule is the primary non-coding area, and it is involved in the regulation of heavy (H) and light (L) strand transcription and of H-strand replication. In mammals the control region, located between the tRNA-Pro and the tRNA-Phe genes (Jae-Heup *et al*., 2001), has one of the most rapid mutation rates found in mammals, varying from 3 % (horses) to 3.9 % (European bison) substitutions/million years (Slade *et al.,* 1998; Burzynska *et al.,* 1999). For higher primates however, the mutation rate has been estimated to be as high as 7 % per million years (Avise *et al*., 1987).

The average rate of evolution of the mtDNA genome is up to 10 times faster than that of coding genes in the nuclear genome (Brown *et al*., 1982; Miyata *et al*., 1982), but 1000 times slower than that of microsatellites. The more rapidly evolving areas, characterized by multiple and accumulated substitutions, have been shown to be very useful for population genetic studies. In contrast, the more slowly evolving regions tend to be useful for phylogenetic studies and comparison of more distantly related taxa (Stevens *et al*., 1989; Baker *et al*., 1993; Simon *et al*., 1994; Árnason and Gullberg 1996).

Several additional characteristics of mtDNA make it an attractive marker for population genetic studies. The molecule does not recombine and the whole haploid molecule is inherited as one linkage group (Hayashi *et al.,* 1985). Due to its small effective population size, it is very sensitive to drift and thus to the effect of population subdivision. This trait also renders mtDNA a sensitive detector of other demographic events such as bottlenecks (Hoelzel *et al.*, 1993; Goldsworthy *et al.,* 2000). One approach to detecting population decline as a result of a bottleneck is to calculate the *D* statistic of Tajima, and it is one of the most widely used and accepted statistics to demonstrate population growth or decline (Przeworski *et al.,* 2000; Ptak and Przeworski, 2002) using DNA sequence data. Population

contractions tend to generate positive *D* values, whereas expansions create a bias towards negative values (Tajima, 1989b). Overall, values of *D* are however influenced by the 'excess' or 'deficit' in rare alleles as compared with the neutral expectation (Ewens, 1972). This means that any factor that may change the importance of rare alleles will influence *D* values, and may be misinterpreted as signals for selection or ancient population size changes.

Recent population expansion, following a bottleneck may be detected using Fu's neutrality statistics (Fs very negative) and Fu and Li's statistical test. In phylogenetic trees, the gene genealogies of populations that have undergone recent expansions typically exhibit "star-like" topologies (Avise *et al*., 1984; Slatkin and Hudson 1991). It is also not uncommon for phylogenies inferred from mtDNA sequence data of species characterized by sex-biased dispersal to differ from that inferred using nuclear markers. This is due to the mode of inheritance and small effective population size of the mtDNA molecule (Bowen *et al*., 1992; Degnan, 1993; Palumbi and Baker, 1994). This also has an effect on F_{ST} values between populations. Values based on mtDNA sequence data will generally be substantially greater than values based on the nuclear markers, especially if the species exhibit a male-biased dispersal pattern.

1.5.2.2 MtDNA sequence evolution

Mutation rates within the mtDNA molecule may vary considerably across a gene sequence. In humans for instance, it is well known that mutation rates in the mitochondrial D-loop region are highly heterogeneous (Vigilant, 1986; Vigilant *et al*., 1991) with so-called 'hotspots' (hypermutable nucleotides) being identified within the rapidly evolving areas (Hasegawa *et al*., 1993; Tamura and Nei 1993; Wakeley, 1993). In order to account for this mutation rate heterogeneity during estimation of relatedness among taxa, the gamma distribution of rates across the sites is estimated (Uzzell and Corbin, 1971). The shape of the gamma distribution is controlled by the α parameter, and large values of α ($\alpha >> 1$) give a distribution curve that is bell-shaped, which is indicative of insignificant or low levels of rate heterogeneity. This is however very rarely encountered. Small values ($\alpha << 1$) are more typical and will result in an L-shaped distribution, suggesting higher levels of rate

heterogeneity. The importance of accounting for rate heterogeneity when inferring phylogenetic relatedness and genetic distance between populations has been stressed (Whelan *et al*., 2001). Simon *et al*. (1994) noted that the properties of the data have a much greater effect on the inferred phylogeny than the method used to analyse the data. The most important consequences of ignoring rate heterogeneity, is that it may lead to the underestimation of sequence distances as well as erroneous phylogeny reconstruction (Hasegawa *et al*., 1993; Tamura and Nei 1993; Yang 1996*a*). Mutational hotspots may also lead to homoplasy and needs to be taken into account when inferring phylogenies as extremely homoplasious data will produce poorly supported phylogenetic trees, irrespective of the tree inference method is used. Distance correction methods, or weighting are often implemented in order to account for intermediate levels of homoplasy (Simon *et al*., 1994). For populations that have diverged recently however, homoplasy has been shown to be less of a problem and most distance correction methods will yield approximately the same estimate if the taxa are closely related since homoplasy will be small as a result of the accumulation of few mutations since their divergence (Simon *et al*., 1994).

1.6 The application of Msats and mtDNA markers in population genetic studies

Microsatellites have become the markers of choice in parentage testing (Queller *et al*., 1993), forensics casework studies, linkage studies and gene mapping studies and population genetic studies (Barendse *et al*., 1994; Gotelli *et al*., 1994; Morin *et al*., 1994; Barker *et al*., 1997; Van Hooft *et al*., 2000) in species ranging from reptiles to mammals (Saitbekova *et al*., 1999; Diez-Tiascon *et al*., 2000; Hanslik *et al*., 2000; Eizerik *et al.,* 2001; Ivankovic *et al*., 2002; Mburu *et al*., 2002; Nyakaana *et al.,* 2002; Cronin *et al*., 2003; Malone *et al*., 2003; Muwanika *et al.,* 2003; Li *et al*., 2004; Harley *et al.,* 2005, Okello *et al.,* 2005). Mitochondrial DNA sequence analysis is widely applied for especially phylogenetic inference, determining population structure, long-term female effective population size and in the assessment of sex-biased dispersal. Its sensitivity in detecting demographic events has been highlighted in several studies.

One of the first studies on Cape buffalo that investigated genetic variation was conducted by Grobler and van der Bank (1996). Using allozymes, they found little variation and low heterozygosities in small disease-free buffalo populations in South Africa. O'Ryan *et al*., (1998) subsequently studied the genetic diversity among the buffalo populations from St. Lucia, Addo, KNP and HiP. They demonstrated a significant correlation between genetic variation and population size, with the St. Lucia and Addo populations exhibiting significantly lower variation than either KNP or HiP. The St. Lucia and Addo populations consisted of 175 and 85 animals respectively at the time, and the reduced genetic variation and significant population differentiation between the respective populations was attributed to genetic drift as a result of small effective population sizes.

Wenink *et al*., (1998), using restriction analysis of the DRB-3 gene of the major histocompatibility complex (MHC), showed that buffalo from parks situated in eastern and southern Africa exhibited high levels of genetic diversity, despite suffering the severe rinderpest bottleneck a century earlier. The high diversity levels were suggested to be the result of a rapid recolonization of the populations following the rinderpest outbreak, while the heterozygote deficiency that was found in allelic variants of the DRB3 gene was attributed to mating behaviour of bulls. The latter resulted in a higher than average relatedness between herd members due to re-entry of bulls into their native herds, at higher than average frequencies. Simonsen *et al*., (1998) found high levels of diversity among buffalo populations from southern and eastern Africa, but little differentiation between populations. The latter was supported by the work of Van Hooft (2000), who showed that low levels of population differentiation exist between populations from eastern and southern Africa, and attributed this to high rates of historical migration between these populations. As significant differentiation could only be demonstrated at the continental level, Simonsen *et al.* (1998) proposed that the pattern of distribution of genetic variation between populations at the regional level could possibly be attributed to a fragmentation of a previously panmictic population. Van Hooft *et al*. (2000), by comparing the genetic diversity of nine buffalo populations from throughout Africa, found a north-south gradient of declining gene diversity across populations in Zimbabwe (Save Valley) and northern and southern KNP. The intermediate level of genetic diversity for northern KNP was attributed to recolonization following high mortality rates caused by the rinderpest in this area, which most probably took place from Zimbabwe and Mozambique (Van Hooft *et al*., 2000). They

also demonstrated the limited impact of the rinderpest, since populations from eastern Africa had high levels of genetic diversity. The inability to detect genetic remnants of the rinderpest bottleneck was attributed to the immigration of rare alleles into the KNP population following the bottleneck. Harley *et al*., (2005), by studying the black rhinoceros (*Diceros bicornis*), found results similar to that of Van Hooft *et al.,* (2000). Although the rhino species suffered a severe population reduction of 96 % at the turn of the last century mainly due to poaching, it seems to have retained most of its diversity. Muwanika *et al*. (2003) studied the effects of severe population declines on genetic variation in wildlife species in Uganda following the complete breakdown of law and order between 1972 and 1985, and increased illegal hunting that brought many wildlife species in the country to the brink of extinction (Eltringham and Malpas, 1980). The common warthog and elephant populations were particularly severely affected, to the extent that the elephant population was reduced by approximately 99 %. These authors could clearly illustrate the impact of the population crash that the two species experienced, and showed that the common warthog exhibited extremely low genetic variation, having an mtDNA nucleotide diversity of less than 0.1 %. The elephant did not fair much better having a nucleotide diversity of only 0.4 %. Cape buffalo and hippopotamus populations were however less severely affected, possibly due to differential poaching pressures. The buffalo retained nucleotide diversities ranging from 3.7 to 5.4 %, with a haplotype diversity of 0.99, indicative of high levels of variation for the species. The differences in retained diversities among the species have also been suggested to be due to recolonization from neighbouring populations, resulting in the restoration of lost diversity. In a study by Okello *et al*. (2005) on hippo populations from various localities in Africa, it was shown that this species retained an overall nucleotide diversity of 1.9 %, which was low compared to other large mammals within the same region. They also found significant negative Fs-values (Fu's neutrality statistic), characteristic mismatch distributions and lack of geographical partitioning among the mtDNA haplotypes, clearly indicating that the species experienced a recent population expansion.

A study by Van Hooft *et al.* (2003) on the structure among herds in KNP revealed no significant differentiation among herds or subpopulations within the park. They did however demonstrate small but significant differentiation when subpopulations from northern and southern KNP were compared. In addition, significant differentiation between the adjoining

populations of Amboselli NP and Tsavo NP in East-Africa was attributed to limited migration between the two populations due to the dryness of the habitat. One of the striking results to emerge was that long-term effective population sizes of many of Africa's buffalo populations were below the census size, implying that without migration, historical levels of diversity will not be maintained in the long term.

Significant reduction in genetic variation as a result of a population crash has also been demonstrated for other mammal species such as the northern hairy-nosed wombat (Taylor *et al*., 1994), the black-footed rock-wallaby (Eldridge *et al.*, 1999), the northern elephant seal (Weber *et al.,* 2000) and elephants in South Africa (Whitehouse and Harley, 2001). It must be noted that it is not uncommon that remnants of a bottleneck are difficult to detect, since the extent to which genetic variability is lost in a population that experienced a bottleneck is complex and may be affected by a number of factors. These include the duration of the bottleneck, population numbers and the presence of fine-scale genetic structure of the post-bottleneck populations (Frankel and Soulé, 1981).

Although several studies have highlighted the negative effects of bottlenecks on genetic variation (O'Brien and Evermann, 1988; Björklund, 2003), a reduced level of genetic variation may not necessarily be associated with a compromised fitness or potential extinction. The cheetah (*Acinonyx jubatus*) has been shown to exhibit extraordinary low levels of genetic variation, which is attributed to inbreeding that followed a severe bottleneck event (O'Brien *et al*., 1983, 1987; Freeman *et al*., 2001), and reproductive and congenital abnormalities together with high cub mortalities have subsequently been put forward as effects of inbreeding depression (O'Brien, 1994). High infant mortality rates have however also been shown to occur among non-inbred cheetah, both in captivity as well as in the wild, and the notion that cheetah in the wild suffer mortalities as a result of genetic impoverishment was also laid to rest when it was shown to be primarily due to predation (Crooks *et al*., (1998). According to these authors, although the importance of processes that are genetically-linked should be recognised, demographic, ecological, and behavioral factors play an equally important role in the fitness and adaptibility of populations.

1.7 Conservation genetics

Conservation, in very broad terms, is about strategies aimed at protecting biodiversity. The importance of conservation is highlighted when a species is under threat of extinction, whether wild or domesticated. It is estimated for instance that diversity of domesticated livestock is shrinking at such an alarming pace that one to two breeds are lost per week (Schearf, 2003). This undoubtedly calls for urgent interventions to prevent further erosion of and loss of diversity (Hanotte and Jianlin, 2005). Conservation however is also about management of diversity that is currently not under threat. In these cases, initiatives are aimed at maximizing opportunities to ensure that these species thrive within their ecosystem over time and within space. The two major applications of genetics in wildlife conservation is firstly to uncover genetic variation and diversity and secondly to relate it to fitness of populations, races or species. Wildlife conservation programs are thus primarily designed to maintain biodiversity at the ecosystem level and genetic variation at the species level (Frankham, 1995). Harley *et al*. (2005) pointed out that measures of genetic diversity would assist in demarcating a genetically viable population, while measures of population differentiation will contribute to the identification of different ecological groupings.

The importance of the availability of baseline genetic information in order to monitor genetic diversity in time and space has also been highlighted, since it is vital if predictions are to be made about how a species might respond to future changes (Larson *et al*., 2002; Harley *et al*., 2005). Inbreeding for instance may affect fitness determinants such as fecundity, survival, growth and susceptibility to environmental stress (Charlesworth and Charlesworth, 1987; Crnokrak and Roff 1999; DeRose and Roff, 1999; Hedrick and Kalinowski, 2000). Empirical evidence has shown that inbred populations face a higher risk of extinction (Saccheri *et al*., 1998; Bijlsma *et al*., 2000; Nieminen *et al*., 2001). Björklund (2003) showed that a mere 5 % reduction in male lion dispersal from their natal site may lead to a substantial increase in the level of inbreeding and subsequent loss of genetic variation.

Translocation of animals is often one of the primary interventions considered for restoration and maintenance of lost genetic variation. This intervention should however go hand in hand with the monitoring of genetic variation in both the source and introduced populations. Larsen *et al*. (2002) showed that translocations of relatively small numbers of sea otters

could effectively avoid further loss of diversity within a bottlenecked population. O'Ryan *et al*. (1998) calculated that genetically impoverished buffalo populations (such as the St. Lucia and Addo populations in South Africa), can be re-juvenated by translocation of one individual per generation (7.5 years) from KNP to these populations. They also estimated that it would restore the heterozygosity of the other parks to 90 % of that found in KNP.

Translocations should however be executed with great care since it is not without associated risks such as disease transmission and outbreeding depression. The latter may result in a compromised adaptability which may have detrimental effects on the viability and sustainability of populations (Storfer, 1999). Crandall *et al.* (2000) emphasized the importance of exchanging individuals from similar environments (that is, the greatest ecological exchangeability) in order to prevent disruption of local adaptations and outbreeding depression. It has been suggested that exchange of individuals between populations should only be considered if there is evidence of at least minimal levels of historical gene flow between them ($F_{ST} < 0.2$; Forbes and Hogg, 1999). Mills and Allendorf (1996) demonstrated, through modelling, that F_{ST} values for populations that are exchanging one migrant per generation would reach an equilibrium value of 0.2.

A particularly problematic situation is where a population has experienced a loss in genetic variation as a result of a population crash (bottleneck). Populations reduced to small sizes in any event have been shown to be prone to loss of variation and to be susceptible to drift (Bonnell and Selander, 1974; Nei *et al*., 1975; Maruyama and Fuerst, 1985; Frankham, 1995a; Goldsworthy *et al.*, 2000; Weber *et al.*, 2000). For instance, Larsen *et al*. (2002) showed, through the review of several studies, that Msat diversity was consistently lower for populations of a variety of species that experienced bottlenecks compared to those that did not, whilst an earlier study (Nei *et al*., 1975) revealed that population bottlenecks usually lead to the loss of rare alleles, and that these populations may experience the fixation of deleterious genes, which can compromise fitness and adaptability.

Although Cape buffalo in South Africa are currently not threatened by extinction, conservation and management strategies for these remarkable animals are vital for ensuring their future sustainability, particularly in view of the growing pressures exerted by man on natural resources. Conservation strategies are however complicated by the fact that these

animals are host to a variety of diseases, which restrict options with regard to intervention. These animals can nevertheless serve as models for understanding the processes affecting their response to environmental change. Although genetics is but one facet of a comprehensive conservation strategy, it can aid in providing information relevant for understanding these underlying processes that govern their existence and which are vital for their sustained conservation.

1.8 Aims of this study

The overall objective of this study was to gain an understanding of how historical events, together with geographical, ecological and behavioural factors differentially impact upon the population dynamics within the Kruger National Park and Hluhluwe Imfolozi Parks. The aim was to study the buffalo at the population level (among the two parks), the sub-population level (among regions within the parks) and on the more fine scale level, *viz*. at the herd level.

1.8.1 Specific objectives were to:

1. Determine the current degree and pattern of genetic variation within and between KNP and HiP, the level to which the two populations are differentiated, and to relate this to historic, geographic, demographic and behavioural factors that influence variation.
2. Assess the pattern, extent and gender-related effect of dispersal taking place at the population as well as the herd level (fine-scale) and to relate this to population structure and potential impact on disease transmission and persistence.
3. Determine the effective population size within KNP and HiP.
4. Infer historical relationships between KNP and HiP buffalo through phylogenetic inference.

Chapter 2

Assessing the impact of the rinderpest bottleneck in KNP and HiP from mitochondrial DNA sequences

Abstract

Disease, whether acute or chronic, has played a major role in the population history of the African buffalo. While subacute diseases such as BTB are not characterised by high rates of mortality, acute diseases such as rinderpest typically result in high mortalities. The rinderpest pandemic of 1890's, besides causing the death of more than 5 million cattle in South Africa, also caused the eradication of as much as 95 % of the Cape buffalo population in KNP. This dramatic population decline may have resulted in a genetic bottleneck, which in turn may have reduced the level of genetic variation among the buffalo of the park. A reduction in variation may subsequently result in genetic drift, which in turn can negatively affect the ability of the population to sustain variation and adapt to environmental change.

The research presented in this chapter is aimed at quantifying the degree to which the bottleneck has affected genetic variation in KNP and HiP. A 452 basepair region of the D-loop region (HV1) of 161 and 97 animals from KNP and HiP, comprising 30 and 14 herds respectively, was sequenced for this purpose. While 34 haplotypes were identified in KNP (with the mean number of pairwise differences between them being 22.09 ±9.78), only 4 haplotypes could be identified for HiP (mean number of parwise differences between haplotypes = 11.14 ± 5.10). HiP consistently exhibited a reduced level of variation when compared to KNP, reflected by both nucleotide diversities (0.049 vs. 0.025) as well as haplotype diversities (0.92 ±0.009 vs. 0.48 ± 0.05). It is thus apparent that while HiP exhibits signals of a genetic bottleneck, the impact thereof in KNP appears to have been overestimated.

The effect of the bottleneck is also reflected by statistics that describe a change in population size. While KNP seems to be in equilibrium, HiP exhibits strong signals of a

population contraction (Fu's Fs = 24.03, p = 0.01). Other test statistics that indicate remnants of a bottleneck and subsequent population contraction for HiP include the D* and F* statistics of Fu and Li and Tajima's D statistic. The potential effect of small post-rinderpest populations, large-scale removal of animals from both parks and immigration into the parks are discussed. The importance of having access to baseline information regarding the genetic status of the buffalo population is stressed, particularly for making sound discisions regarding conservation management of the two parks.

1. Introduction

Both ancient and recent demographic events have played a role in the population history of the Cape buffalo. While climatic fluctuations have influenced the geographical distribution of Cape buffalo on a more ancient time scale (Van Hooft *et al*., 2002), more recently their abundance has been impacted upon quite dramatically by disease. The rinderpest pandemic of the 1890s caused a drastic population crash in Cape buffalo populations throughout sub-Saharan Africa, and resulted in a reduction of up to 95 % of buffalo in the Kruger National Park, South Africa (Stevenson-Hamilton, 1957). A population decline of this magnitude was from the outset regarded as a potential bottleneck which may have affected the population dynamics of the Cape buffalo. As bottlenecks are known to have a detrimental effect on the genetic diversity of populations, a significant reduction in genetic variation of buffalo due to rinderpest is anticipated. Certain potential risks may be associated with a reduced level of genetic variation, such as a compromised ability of a population to sustain and to maintain genetic variation. The latter is of particular importance for a species' response to environmental and demographic change. Several studies have highlighted the effects of population bottlenecks on a variety of species, and many demonstrated a concomitant reduction in genetic variation. Examples include the northern hairy-nosed wombat (Taylor *et al*., 1994), the black-footed rock-wallaby (Eldridge *et al.*, 1999), the northern elephant seal (Weber *et al.,* 2000), the common warthog and African elephant (Muwanika *et al*., 2003), the golden monkey (Su and Shi, 1995; Li *et al*., 2003) and the European lynx (Hellborg *et al.*, 2002).

The importance of understanding the effects of demographic bottlenecks for future management has been stressed by several authors (Hellborg *et al.*, 2002; Harley *et al.*, 2005). A small population size is by default an effect of a bottleneck, and this may result in genetic drift and subsequent loss of variation. Detecting and understanding bottlenecks and their effects is however by no means a simple task, and tools sensitive enough to detect genetic signals of past events are a necessity. The mitochondrial genome is one of the most widely used markers in this regard. Certain of its characteristics such as its mode of inheritance, small effective population size (only one quarter of autosomal loci), make it a sensitive marker for detecting reduction in genetic variation (Hoelzel *et al.*, 1993; Goldsworthy *et al.,* 2000). The molecule is furthermore sensitive to genetic drift and the effect of population subdivision.

The aim of this chapter is to investigate the extent to which the rinderpest epidemic affected the genetic status and level of variation among the buffalo populations in KNP and HiP. Conservation implications and future perspectives are discussed with regard to the long-term maintenance and sustainability of genetic variation within the two parks.

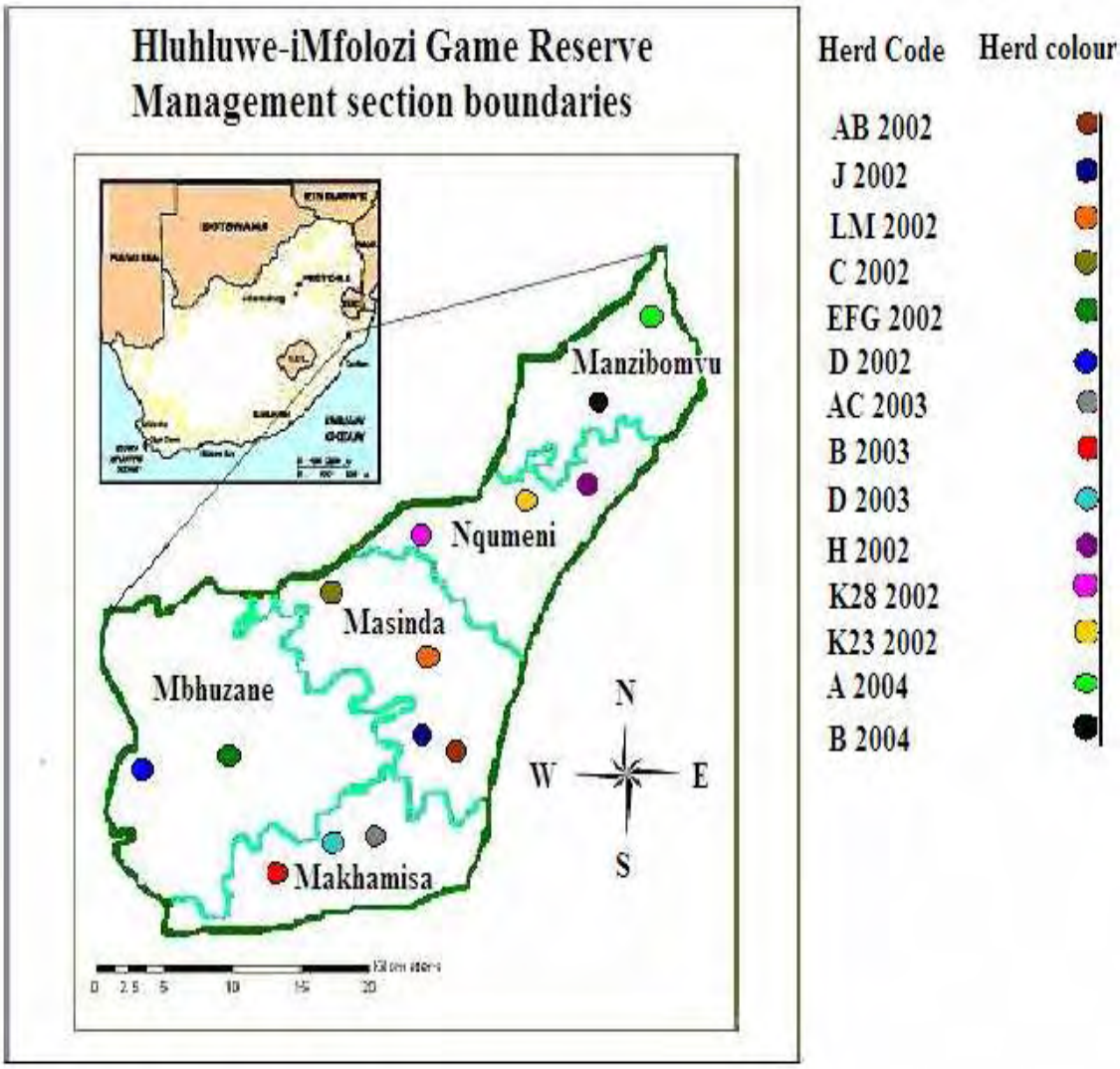
2. Materials and Methods

2.1 Sampling

2.1.1 HiP

Blood from ninety-seven animals, originating from 14 herds, was used for mtDNA sequence analysis. These herds represent the five major management regions of the park, viz. Manzibomvu, Nqumeni, Masinda, Makhamisa and Mbhuzane areas (Fig. 2.1). Young animals born in the same year were randomly sampled from each herd in order to minimize the chance of sampling full-sibs and maximize the chance of sampling as much diversity as possible.

Fig. 2.1. Geographical map of HiP showing the 5 management regions from which samples were obtained.



2.1.2 KNP

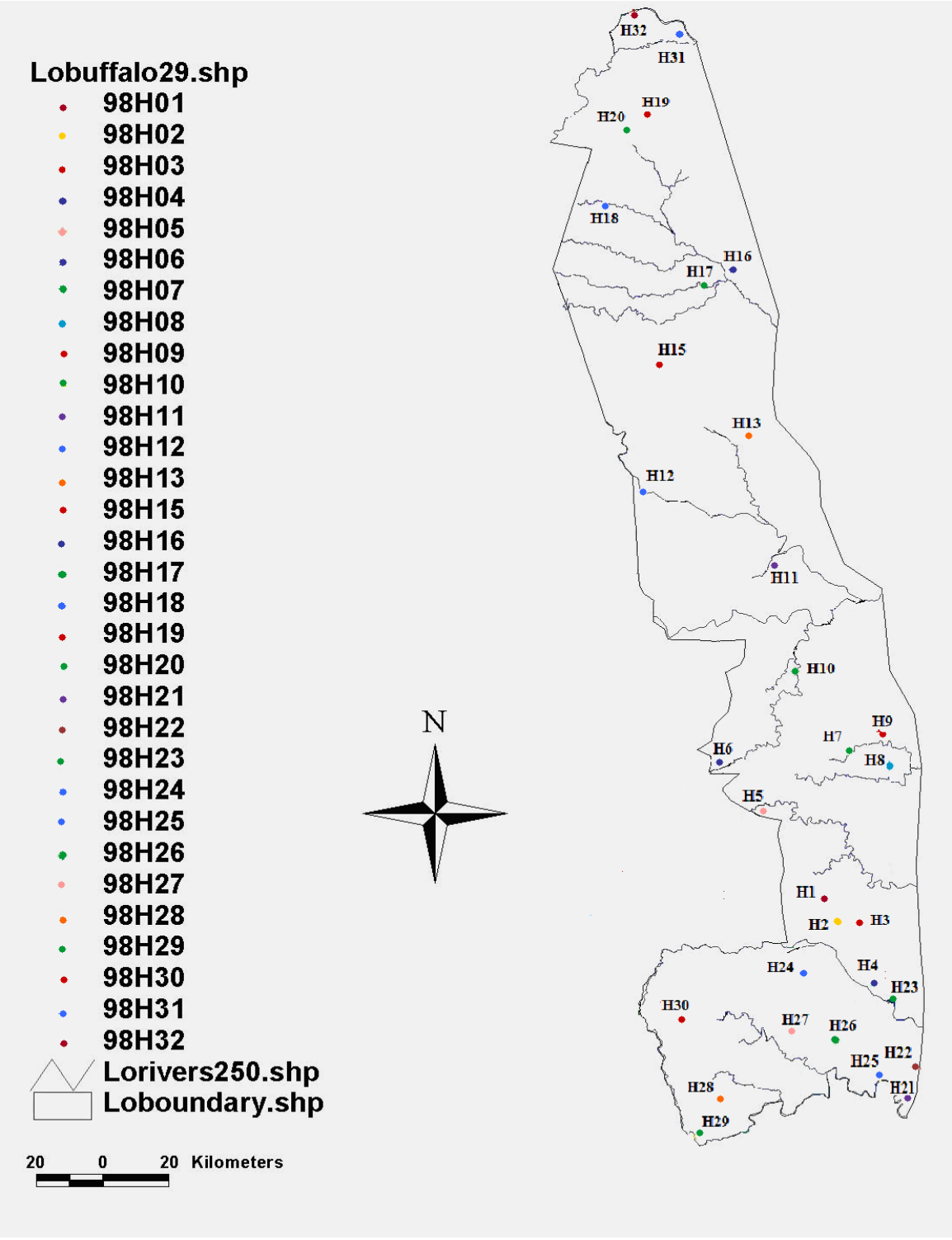
Blood from a total of 161 animals, originating from 30 herds occurring throughout the park (Fig. 2.2) was used for mtDNA sequence analysis. At least 5 animals, again primarily young animals were chosen at random from each of the 30 herds.

2.2 Genomic amplification and characterisation

DNA was extracted from whole blood (300 µl) using the Roche kit according supplier specifications and used as template for the amplification of a 452 bp region corresponding to HVRI of the D-loop region of the mitochondrial genome (Bastos, submitted). Approximately 200ng of extracted DNA, 1X PCR buffer, 0.2 mM of dNTP, 0.4 µM of each primer (DLH –BOV, 5'CCT GAA GAA AGA ACC AGA TG 3'; Thrl-BOV, 5' TAA TAT ACT GGT CTT GTA AAC C 3') and 1 unit of Taq polymerase (Biotools) in a final reaction volume of 50 µl was used to amplify the target region under the following cycling conditions: an initial denaturation step at 95^0C for 40 seconds, followed by 39 cycles consisting of denaturation at 95^0C for 20 seconds, annealing at 50^0C for 30 seconds and extension at 70^0C for 45 seconds, and a final extension step at 70^0C for 1 minute.

The amplicons were subsequently purified using the Roche High Pure PCR purification kit, according to the instructions of the manufacturer. Forty nanograms of purified amplicon were used for cycle sequencing, with the Bigdye Cycle Sequencing kit (Perkin Elmer) according to the instructions of the manufacturer. The cycle sequencing profile consisted of 25 cycles starting with a denaturation step at 96^0C for 10 seconds, followed by an annealing step at 50^0C that lasted for 5 seconds and a final extension step at 60^0C for 4 minutes. Precipitated amplicons were electrophoresed on an ABI 3100 capillary-based DNA sequencer (Applied Biosystems) and reaction profiles were visualised and edited manually using the program Chromas (Conor McCarthy, School of Biomolecular and Biomedical Science, faculty of science, Griffith University, Brisbane, Queensland, Australia).

Fig. 2.2. Geographical map of KNP indicating the location of the 30 herds from which samples were taken. Each coloured dot represents one of the 30 herds sampled. These samples were taken as part of the 1998 BTB longitudinal survey the aim of which was to assess infection rates in different regions of the park.



2.3 Statistical analysis of sequences

Edited sequences were aligned using DNAman (version 4.13, Lynnon Biosoft) and transformed into GDE or Clustal format. In order to account for mutation rate heterogeneity during subsequent analysis, the α parameter of the gamma distribution of rates of mutation across the sites was estimated using Modeltest 3.06 (Posada and Crandall, 1998). General estimates of sequence diversities and molecular evolutionary analyses were conducted using MEGA version 3.0 (Kumar *et al.*, 2004), DnaSP4.00 (Rozas *et al.*, 2003) and Arlequin (Schneider *et al.*, 1997). TCS 1.13 (Clement *et al.*, 2000) was used to construct a minimum spanning network (MSN), depicting the phylogenetic relationships among the respective haplotypes. The program estimates gene genealogies from DNA sequences based on the cladogram estimation method, also known as statistical parsimony (Templeton *et al*., 1992; Clement *et al.*, 2000). MSN's are favoured above neighbourjoining trees for intraspecific studies since they allow for multifurcations whilst not assuming that ancestral lineages are extinct (Crandall and Templeton, 1993). The latter implies that haplotypes are not forced to occupy the tip positions.

In order to detect signals of past demographic events, Arlequin was used to determine the distribution of the observed and expected pairwise nucleotide site differences (mismatch distribution), based on a sudden expansion model, while DnaSP 4.00 determined the mismatch distributions based on a stable population (at equilibrium and for no recombination), i.e. population with constant population size (Watterson 1975; Slatkin and Hudson 1991, Rogers and Harpending 1992). The site frequency spectra (SFS) which is a measure of the difference in the frequency of different mutation classes, was determined using DnaSP 4.00. A population that has recently expanded will be characterized by an excess of singleton mutations relative to the expected frequencies under neutrality and stationarity (Fu, 1997; Okello *et al.*, 2005).

Other genetic parameters describing population change, such as Tajima's D, Fu and Li's D* and F* statistics (determinates of the SFS), Strobeck's S statistic and Fu's Fs statistics were determined using DnaSP and Arlequin. Fu and Li's D* and F* statistics test the hypothesis that all mutations are selectively neutral (Kimura, 1983) and that estimates of η/a1, (n - 1)ηs/n, and of k, are unbiased estimates of theta θ, where, η, is the total number of

mutations; a1 = S (1/i) from i = 1 to n-1; n is the number of nucleotide sequences; ηs, is the total number of singletons (mutations appearing only once among the sequences); k is the average number of nucleotide differences between pairs of sequences (Tajima, 1983). The D* test statistic is based on the differences between ηs and η , while the F* statistic is based on the differences between ηs and k (Fu and Li, 1993). Strobeck's S statistic (Strobeck 1987; Fu, 1997) is based on the haplotype (gene) frequency distribution conditional on the value of θ (Ewens, 1972). The S statistic gives the probability of obtaining a sample with equal or less number of haplotypes than that observed, and is an indication of a haplotype deficit or excess within the sample population.

Tajima's D is a function of theta (θ), an important parameter that determines the degree of polymorphism at a locus. Several inferences regarding the evolution of a population are also a function of the value of theta (Fu, 1997), which is defined as 2Nμ for mitochondrial DNA, where N is the effective female population size and μ is the neutral mutation rate (per gene or per base pair) per generation. Theta can be calculated on a per site basis from Eta (η) or from S, i.e. the Watterson estimator (Watterson 1975). Eta (η) is the total number of mutations while S is simply the number of segregating (polymorphic) sites. Theta can also be calculated on a per sequence or per gene basis, from either S, or from π, the nucleotide diversity (Tajima, 1996).

3. Results and Discussion

3.1 Control region diversity

The diversity indices of the HV1 Area of the D-loop region were based on an homologous 445 basepair region (Appendix I). Of the 161 animals sequenced for KNP, 34 haplotypes were identified, and the mean number of pairwise differences between these haplotypes was 22.09 (±9.78) (Table 2.1). For HiP, only 4 haplotypes were identified among the 97 animals sampled, and the mean number of pairwise differences between them was 11.14 (±5.10). The overall mean sequence divergence among the KNP haplotypes was also markedly higher than for HiP. This reduced diversity for HiP compared to KNP is also evident from the respective nucleotide diversities (0.025 for HiP compared to 0.049 for KNP) as well as

haplotype diversities (Table 1). Although none of the haplotypes identified were shared between the two parks, the level divergence of between sequences from the two populations was small, with the mean distance between the total number of haplotypes from the respective populations being 0.051 (SE 0.007; Tamura-3 corrected, gamma shape parameter= 1.11). The average number of nucleotide differences between KNP and HiP populations was 18.9, which is smaller than that within KNP. A minority of haplotypes differed by as little as a single nucleotide substitution or insertion (Appendix 1).

Table 2.1. Summary of diversity indices based on a homologous 445 basepair region of HVR1 within the D-loop.

Diversity parameter	**KNP**	**HiP**
N	161	97
Pi (nucleotide diversity)	0.049 ± 0.024	0.025 ± 0.012
Hd	0.927 ± 0.009	0.489 ± 0.050
K	34	4
Nts	87	26
Ntv	4	0
Mean number of pairwise differences	22.09 ± 9.78	11.14 ± 5.10
Number of segregating sites	73	24
Overall mean seq divergence (Tamura-3 corrected, gamma 1.11)	0.052 (SE 0.007)	0.033 (SE 0.007)

K: The number of detected haplotypes in each population; N: sample size: N; π: nucleotide diversity; Hd: haplotype/gene diversity; Nts: number of observed transitions; Ntv: number of observed transversions

Both parks exhibited similar L-shaped haplotype frequency distributions, with most haplotypes being represented more than once (Figures 2.3a and 2.3b) and one haplotype being very dominant and occurring at a high frequency for both parks. The frequency of haplotypes per herd sampled also varied considerably, which contributed to the level of

divergence found among herds. This will be discussed further in the next chapter that deals with aspects of structure within and among the parks.

Fig. 2.3a. Haplotype frequency distribution within the KNP population. Each of the coloured bars represents a unique haplotype and follows the haplotype numbering used in Appendix I and Fig. 4.

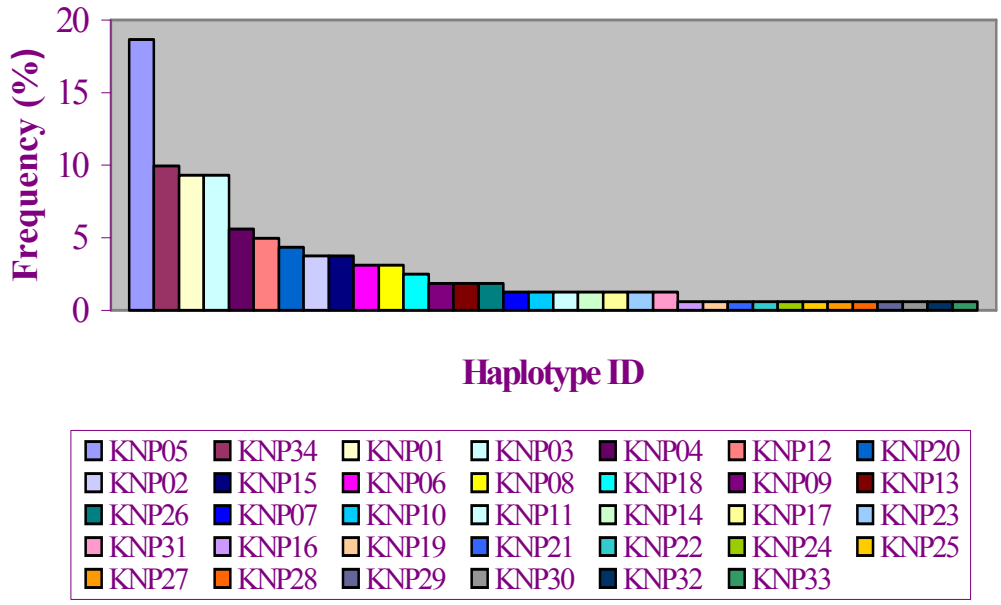
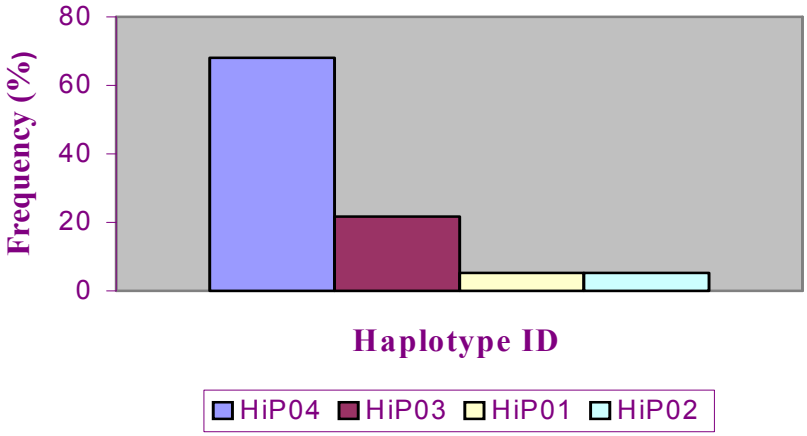


Fig. 2.3b. Haplotype frequency distribution within the HiP population. The four haplotypes are each indicated by a different colour.



While the nucleotide and haplotype diversities for KNP are moderate and high respectively, HiP is characterized by low nucleotide and haplotype diversities. This, together with the large proportion of haplotypes that are identical or have high similarity scores (small number of nucleotide differences between the haplotypes), is consistent with a reduction in genetic variation.

A correlation between a sharp reduction in genetic variation and a bottleneck has been highlighted by several studies**.** Hellborg *et al*. (2002) found a single haplotype in the Scandinavian lynx population following intense hunting in the early 20th century. Muwanika *et al.* (2003) demonstrated an acute reduction in the mtDNA diversity for warthog and elephant populations following a drastic reduction in population size in Uganda from 1972 to 1985. In the latter study, the elephant population, believed to have suffered a population decline of close to 99 %, had a nucleotide diversity of 0.4 %, while only 4 haplotypes could be identified among the 66 animals sampled in Queen Elizabeth National Park (QENP). For the warthog, only a single haplotype could be identified within the QENP individuals sampled in Uganda. Among the 102 animals sequenced from QENP, Kidepo Valley (KV) and Murchison Falls (MF) national parks in Uganda only 12 haplotypes were found, differing in total by 31 variable sites. Four and three haplotypes respectively were identified in the warthog populations from KV and MF. The buffalo on the other hand, despite experiencing a 70 % reduction in population size during the same period, retained nucleotide diversities ranging from 3.7 to 5.4 %, with a haplotype diversity of 0.99 (Muwanika *et al*., 2003). Van Hooft *et al*. (2002) found somewhat lower values for Cape buffalo populations from east, central and southern Africa. Haplotype diversities for these populations ranged from 0.89 (east Africa) to 0.94 (central Africa) and nucleotide diversities were 3.8 %, 4.9 % and 6.5 %, respectively for the three populations. Muwanika *et al*. (2003) attributed the high level of variation to recolonization from neighbouring populations that resulted in the restoration of lost diversity. These diversity indices from Uganda are of the same order as that found within the KNP population. It is possible that the current diversity in KNP is attributable to an overestimation of the severity of the rinderpest epidemic as well as recolinization of the KNP in the period before the park was completely fenced. It is for example known that KNP experienced an influx of buffalo from Mozambique as recently as the 1960's (De Vos *et al.*, 1983), possibly contributing to the restoration of diversity. KNP exhibits moderate nucleotide diversity and high haplotype diversity, which indicates that the

park has retained high levels of variation, despite the rinderpest bottleneck. Simonsen *et al.* (1998) also speculated that the rindepest bottleneck has been overestimated or that it was brief and affected only a few generations, so that a reduction in genetic variation was limited, which is in accordance with our results and that of others (O'Ryan *et al.*, 1998; Simonsen *et al.*, 1998; Van Hooft *et al*., 2000). Harley *et al.* (2005) found similar results for the black rhinoceros (*Diceros bicornis*). Although the species suffered a severe population reduction of 96 % at the turn of the last century, mainly due to poaching, it seems to have retained most of its diversity.

3.2 Genetic drift and effects of small post-bottleneck populations

A greater population decline in HiP, concomitant with the absence of immigration into the park, could explain the lower variability in HiP. The assumption that little or no immigration took place into HiP, is based on the fact that substantial human settlement prevailed around the park, while strict disease control measures were imposed since the early 1930's. HiP was also known to have had a small population size following the bottleneck. During the 1929 census, a total of 337 animals were recorded in the entire park. According to population genetics theory, small populations are prone to drift (Bonnell and Selander, 1974; Nei *et al.*, 1975; Maruyama and Fuerst, 1985; Frankham, 1995a; Goldsworthy *et al.*, 2000; Weber *et al.*, 2000), and the reduced genetic variation for HiP is consistent with a small post-rinderpest population. O'Ryan *et al.* (1998) demonstrated a significant correlation between genetic variation and population size of buffalo from the St. Lucia and Addo populations. These two populations, consisting of 175 and 85 animals respectively at the time, exhibited significantly lower variation than either KNP or HiP. Genetic drift as a result of small effective population size is postulated to have been responsible for the reduced genetic variation and significant population differentiation. Another factor that could have contributed towards a decline in diversity in HiP is the removal of animals as a result of population control measures. Close to four thousand animals were removed from the park between 1982 and 1994, in order to restrict population growth. Likewise, more than 25000 buffalo were removed from KNP between 1967 and 1981, while the population experienced a population crash from 29359 to 14123 during a

severe draught in the mid 90's (De Vos *et al.*, 2001). These removals may have negatively affected the haplotype diversity.

It is noteworthy however that the nucleotide diversity of HiP, when compared to several other studies in which population crashes occurred, is still relatively high. It thus seems that the bottleneck effect had a more notable effect on haplotype diversity, and that nucleotide diversity in the population following the botleneck may have been comparable to the current diversity. This study shows that, although the current census size for buffalo in HiP is close to 3000, it nevertheless exhibits a reduced level of genetic variation. A similar state of affairs has been shown for the snub-nosed monkey. Li *et al*. (2003) found that polymorphism at allozyme loci of this primate, numbering approximately 15000, was almost non-existent, and they attributed this to a recent bottleneck event.

3.3 Detecting bottlenecks and population change

According to population genetic theory, the effects of bottleneck and subsequent population expansion may leave certain signatures in the population under study. In genetic terms however it is necessary to distinguish between ancient expansion and expansion following a recent bottleneck. It is known for instance that a population that has been expanding in size following a recent bottleneck, will have gene genealogies that exhibit a "star phylogeny" (Avise *et al*., 1984; Slatkin and Hudson, 1991). This is due to recently evolved haplotypes that are separated by small genetic distances. Typically, these haplotypes are found on the short branches of the phylogenetic tree, since coalescences of ancestors occur within a relatively short period (Marjoram and Donnelly, 1994).

The relationships among the haplotypes within the two populations and their respective frequencies are shown in the minimum spanning network (Fig. 2.4). A considerable number of haplotypes were seperated by relatively large genetic distances, as reflected by the multiple mutational events between them. In fact, many haplotypes were sepereated by mutational steps that exceeded the confidence limits for parsimony. This is in accordance with the relatively large mean number of pairwise differences between the haplotypes (Table 2.1). The absence of a star-like structure in the haplotype network (Fig. 2.4) concurs

with results from other statistics presented in this chapter that could not detect a population expansion event. Recent expansion events are characterised by a large number of haplotypes that are separated by a limited number of mutational changes. Although the distribution of the different haplotypes is not related to their geographical location, the small level of sequence divergence among haplotypes origination from different parks suggests genetic contact between the two populations in the recent past. From the network it seems that haplotype HiP04 is the most pertinent ancestral haplotype (according to coalescent theory, ancestral haplotypes will be those that occur at the highest frequencies in a population level study; Crandall and Templeton, 1993). The high frequency of haplotype HiP04 may however be due to drift (which resulted in the fixation of haplotype HiP04) that followed the bottleneck. The frequency of this particular haplotype may also be a reflection of the variation present in the post-bottleneck population. Although non-detected/inferred haplotypes, represented by interlinking nodes, may be an artefact of the sampling process, the possibility that these haplotypes are extinct can not be excluded.

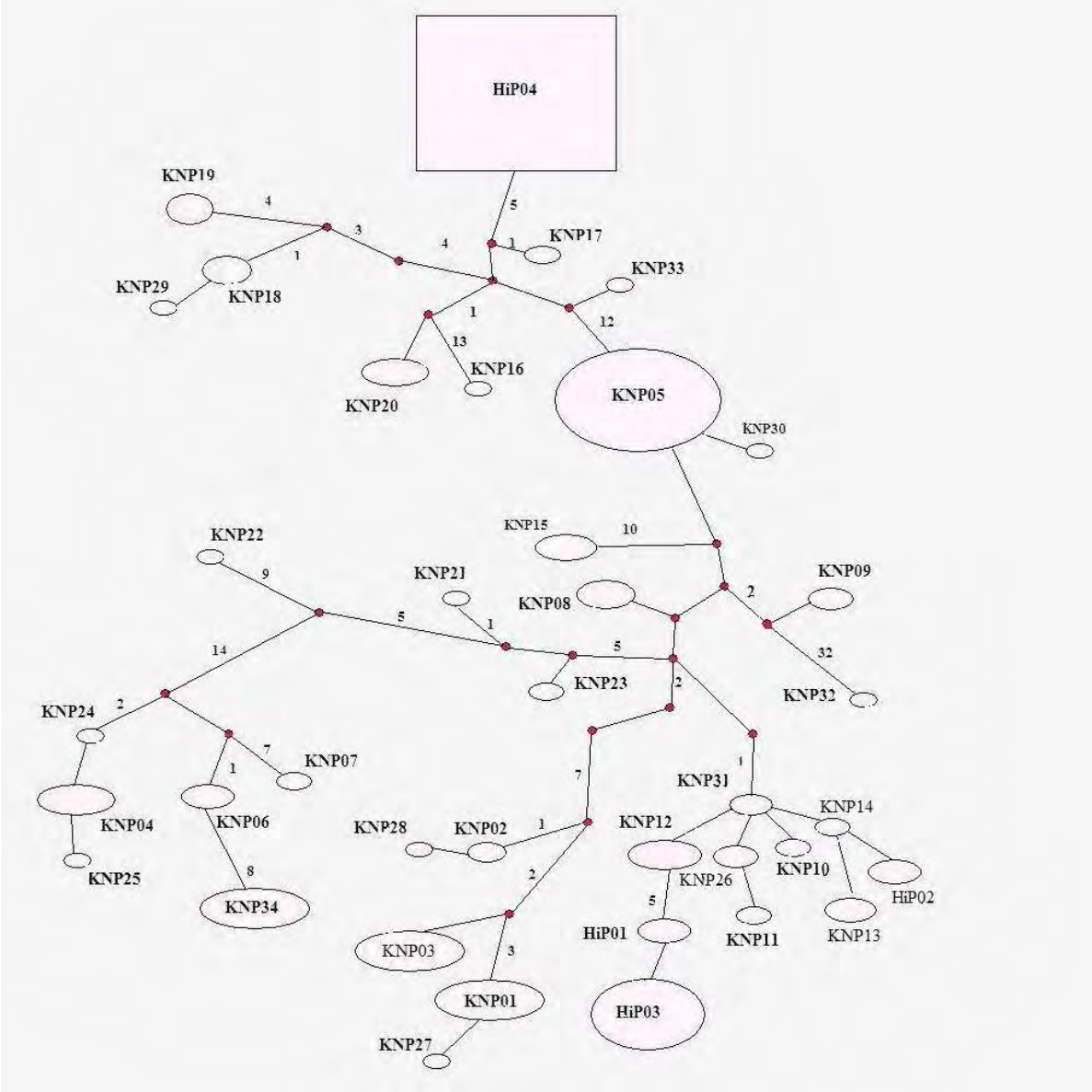


Fig. 2.4. Minimum spanning network depicting the phylogenetic relationships among the observed 38 mitochondrial control region haplotypes. Anonymous nodes represent inferred and nondetected haplotypes. The numerical numbers on the branches between the nodes represent the number of mutational steps separating nodes and haplotypes, and the size of the square and ovals corresponds to the respective haplotype frequencies. Gaps in the data, generated as a result of insertions or deletions, were treated as a fith character state.

Detecting signals of a population bottleneck and subsequent expansion with molecular data is not necessarily a trivial exercise, since the signal is a function of both the severity and the duration of the bottleneck, as well as the size of the post-bottleneck population and the time in generations that have passed since the bottleneck. This is clearly illustrated by the study of Muwanika *et al.* (2003) who showed that although the warthog and elephant were genetically severely impoverished, all genetic based tests for detecting a bottleneck were negative. Another possible explanation for the detection failure may be that the effects of a bottleneck are a transient feature, which may not be apparent beyond a few generations (Luikart and Cornuet, 1998). Populations may thus reach equilibrium and stationarity again some time after a bottleneck, eradicating signals of expansion.

This study used several genetic indices such as Tajima's D, Fu and Li's D*, and Fu and Li's F* statistics in an attempt to detect signals of recent and ancient bottlenecks and subsequent expansion (Table 2.2).

Table 2.2 Statistical parameters associated with deviations from a neutral model of evolution

Test statistic	KNP	HiP
Fu and Li's D*	1.55 ($p < 0.05$)	1.86 ($P < 0.02$)
Fu and Li's F*	1.58 ($0.10 > P > 0.05$)	2.74 ($P < 0.02$)
Fu's Fs	4.89 (95 % CI: -14.45 - 10.79)	24.03 (95 % CI: -8.06-8.95)
	P [Fs <= 4.89] = 0.88	P [Fs <= 24.03] = 0.99
Strobeck's S	P (NHap <= 30) = 0.013	P (NHap <= 4): = 0.000
Tajima's D	1.14 ($P > 0.10$)	2.973 ($P < 0.01$)

Both KNP and HiP exhibit positive Fs values, indicating no recent expansion. The Fs value for KNP is however non-significant (Fs = 4.89, p = 0.12). In contrast, HiP exhibits a highly significant and large positive Fs (Fs = 24.03, p = 0.01), which is consistent with population contraction. The significance of the Fs statistic was tested by generating random samples (using DnaSP) using a coalescent approach, based on a neutral infinite sites model (Hudson,

1990) and assuming a large constant population size. The p-value of Fs is given as the proportion of random Fs statistics less or equal to the observed value.

The positive values for Fu's Fs are indicative of a haplotype deficit, given the observed level of nucleotide diversity (Fu, 1997). The haplotype deficit is supported by Strobeck's S statistic, which tests for the presence of a haplotype excess, as experienced in an expanding population or when immigration of haplotypes has taken place into the population. From Strobeck S statistic it is evident that neither KNP nor HiP exhibits a haplotype excess. In fact, both parks exhibit a significant haplotype deficit. Fu and Li's D* statistic for both parks are also significantly positive (Table 2), and positive values indicate an excess of intermediate-frequency alleles in a population that may be the result of a population bottleneck (Akey *et al.*, 2004). Fu and Li's F* statistic and Tajima's D is only significantly positive for HiP which is indicative of a population contraction (Tajima, 1989b). In summary, and from table 2, all statistics testing for a bottleneck in HiP are significant, while only the D* statistic of Fu and Li is significantly positive for KNP. Fu and Li's F statistic is only near significant ($p < 0.1$) for KNP. These statistics suggest that, although not strong for KNP, both parks exhibit signals of a bottleneck.

3.4 Site frequency spectra

The site frequency spectrum describes the distribution of the allelic frequency at a site. Under the mutation/drift model of molecular evolution (observed diversity represents a balance between the introduction of new polymorphism by mutation and the extinction of existing polymorphism by genetic drift) as well as appropriate demographic assumptions, it is possible to predict the expected site frequency spectrum (SFS) of a region (Watterson 1975; Ewens, 1979). A number of statistical tests have been devised that compare an observed SFS (mutation frequency spectrum) against neutral theory predictions. Tajima's D (Tajima, 1989a) is frequently used to compare the nucleotide diversity estimated from the number of polymorphic sites observed against nucleotide diversity estimated from the allele frequency of the polymorphic sites. Recently expanded populations show characteristic features in the frequency of mutation classes (Donnelly *et al.*, 2001). Fu and Li's D (Fu and Li, 1993) estimate is based upon the number of singleton derived alleles observed. An

excess of singleton mutations relative to the expected frequencies under neutrality and stationarity, (thus large differences between expected and observed frequency of mutation classes), is consistent with recent population growth (Fu, 1997; Okello, 2005). The SFS of KNP and HiP is shown in figures 2.5a and 2.5b.

Fig. 2.5a. Site frequency spectrum (SFS) for HiP, based on a model of equilibrium (constant population size). The observed and expected frequencies of mutation classes differ significantly ($p < 0.001$, Wilcoxon matched pair test).

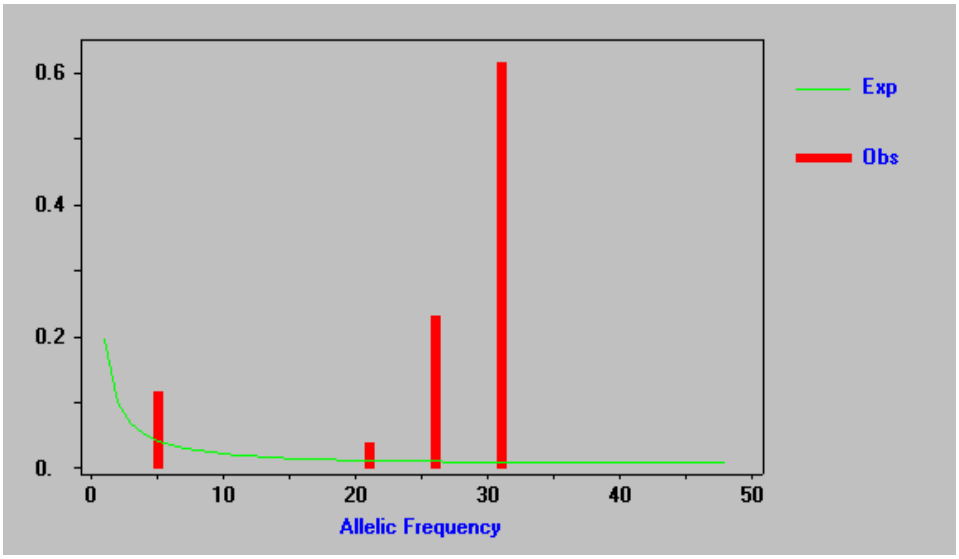
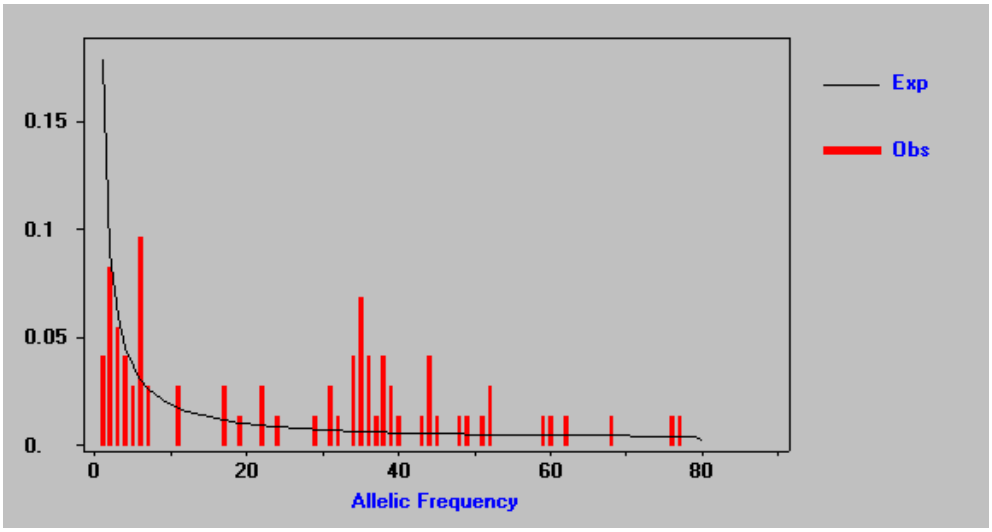


Fig. 2.5b. Site frequency spectrum (SFS) for KNP, based on a model of equilibrium (constant population size). Observed and expected frequency of mutation classes do not differ significantly ($p = 0.84$; Wilcoxon matched pair test)



The observed and expected frequencies of singleton mutations under a model of constant population size for HiP differ significantly ($p < 0.001$; Wilcoxon matched pair test). Large differences between observed and expected frequencies under a model of population stationarity generally indicate an excess of singleton mutations. The latter may be interpreted as a signal for population expansion in the recent past (Fu, 1997; Okello *et al.*, 2005). This result for HiP is however contradictory as positive values for Tajima's ' D and Fu and Li's D and F statistics were obtained for HiP, which is consistent with population contraction. It has been shown however that Fu's Fs is a more powerful statistic for detecting deviation from neutrality. The large and significant positive value for Fu's Fs (24.03, $p < 0.001$) is consistent with population contraction. The SFS for observed and expected values for KNP, under a model of stationarity, do not differ significantly ($p = 0.84$; Wilcoxon matched pair test). This indicates, together with insignificant positive values for Fu's Fs, Tajima's D and Fu and Li's F*statistics (Table 2), that the SFS of KNP is consistent with a stationary population at equilibrium. Fu and Li's D statistic however suggest expansion (1.55, $p < 0.05$).

The problem of disentangling signatures of natural selection from demographic factors such as bottlenecks and subsequent population expansion events is by no means simple. For example, both positive selection and increases in population size lead to an excess of low-frequency alleles in a population relative to what is expected under a standard neutral model (Akey,*et al.*, 2004). Positive values for Tajima's D, Fu and Li's D*, and Fu and Li's F* could thus also be due to positive selection (Akey *et al.*, 2004). A negative Tajima's *D* on the other hand may be the result of a recent selective sweep, weak negative selection or population admixture. In order to distinguish selective factors from demographic forces, it is essential to analyze multiple loci dispersed throughout the genome (Akey *et al.*, 2004). This is necessary since demographic forces affect patterns of variation at all loci in a genome in a similar manner, while natural selection acts upon specific loci (Cavalli-Sforza, 1966, Przeworski *et al.*, 2000; Nielsen, 2001).

3.5 The distribution of the pairwise nucleotide site differences (mismatch distribution)

The mismatch distribution is the distribution of the observed number of differences between pairs of haplotypes and is frequently used to detect demographic events in the more distant past. Multimodal mismatch distributions are typical of populations that are at demographic equilibrium, while populations that have experienced expansions are characterized by unimodel (bell or wave-like) distributions (Slatkin and Hudson, 1991; Rogers and Harpending, 1992). The critical parameters on which the distributions are based are tau, $theta_0$ and $theta_1$, which are estimated using a generalized least squares method under a model of sudden expansion (Schnieder and Excoffier, 1999). A parametric bootstrap approach is subsequently used to calculate approximate confidence intervals of the three parameters of expansion, and the validity of the estimated expansion model is then tested by calculation of the sum of the squared deviations (SSD) between observed and expected mismatch. The p-value of the latter test is:

the number of SSD (sim) >= SSD (obs)/B,

where B equals the number of samples simulated around the estimated parameters. The raggedness index (r) of Harpending (1994) is also estimated and describes the smoothness of the observed distribution. Expanding populations typically have smooth distributions and thus low r index values, while a stationary population at equilibrium on the other hand is characterized by a large r index and "ragged" distribution. The significance of the raggedness index is calculated in the same way as the SSD.

The observed and expected mismatch distributions for a stable and an expanding population for KNP and HiP are shown in figures 2.6a-2.6d. Neither the expansion nor the constant population size models fit the observed distributions for HiP (Figures 2.6a and 2.6b). The Wilcoxon matched pairs test also revealed that the observed and expected distributions under a model of stationarity differ significantly. This is in accordance with the positive estimates of Tajima's D, Fu's Fs and Fu and Li's D* and F* statistics, which are significantly positive and indicative of population contraction. The highly ragged (r = 0.39) distribution supports this finding, while actual and simulated values for SSD differ significantly.

The sudden expansion model also does not fit the observed distribution obtained for KNP (Figure 2.6d), as the simulated and observed values for both the SSD and raggedness index differ significantly. A prominent ragged distribution is also evident from a visual inspection of the curve (Fig. 2.6d). Under a model of constant population size and stationarity however, the observed and expected distributions are not significantly different ($p = 0.75$, Wilcoxon matched pairs test), which is consistent with a population in equilibrium. Simulated values of the raggedness index (r) based on the coalescent process for a neutral infinite-sites model and assuming a large constant population size (Hudson, 1990), indicate that the raggedness index is actually smooth and it is not different from that obtained by simulation (P [r <= 0.0140 = 0.492). It must be noted however, that the raggedness statistic has a low statistical power for detecting population expansion, and therefore statistics such as Fu's Fs are more powerful and appropriate in this regard (Fu and Li, 1993; Fu, 1997). It has also been argued that mismatch distributions may not necessarily be able to detect demographic events like expansions, since they do not use all the information accumulated in the data (Felsenstein, 1972; Okello *et al.*, 2005). Once again Fu's Fs is considered to be the more appropriate statistic for substantiating evidence for an expansion.

Fig. 2.6a. Mismatch distribution for HiP under a model of constant population size. Harpendings raggedness statistic (r) = 0.3942. The observed and expected distributions differ significantly (p = 0.0018; Wilcoxon matched pair test).

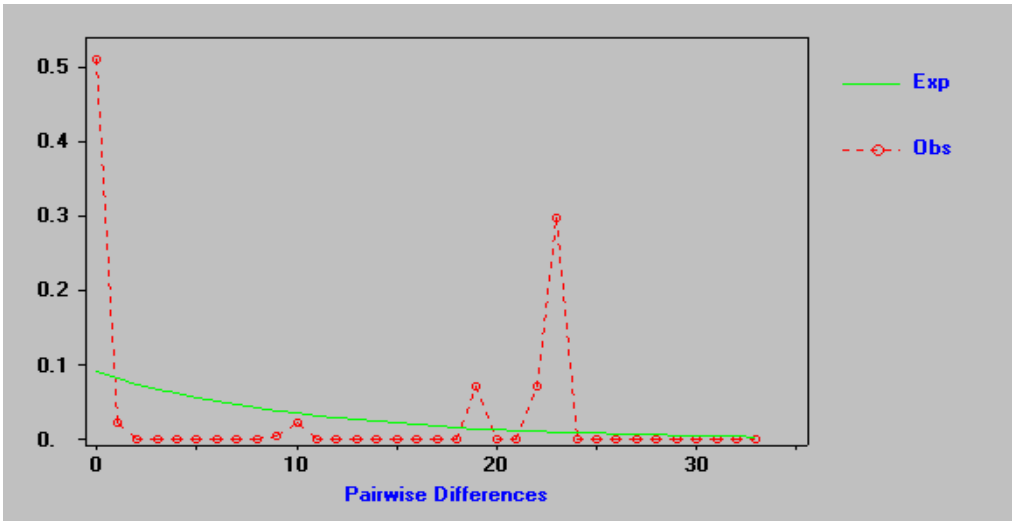


Fig. 2.6b. Mismatch distribution for HiP based on a model of sudden expansion. SSD = 0.165, P(Sim. Ssd >= Obs. Ssd) = 0.034, Harpending's raggedness index (r) = 0.394, P(Sim. Rag. >= Obs. Rag.) = 0.329. Observed and expected values are also significantly different (p = 0.002, Wilcoxon matched pair test)

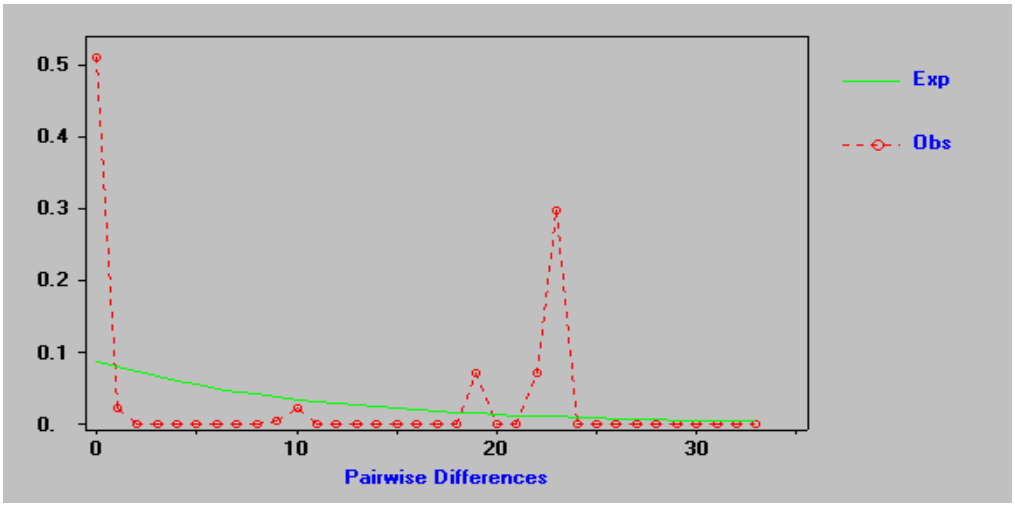


Fig. 2.6c. KNP mismatch distribution under a model of constant population size. Raggedness index (r) = 0.0145, P[r <= 0.0140] = 0.49. Observed and expected values do not differ significantly (p = 0.75, Wilcoxon matched pairs test).

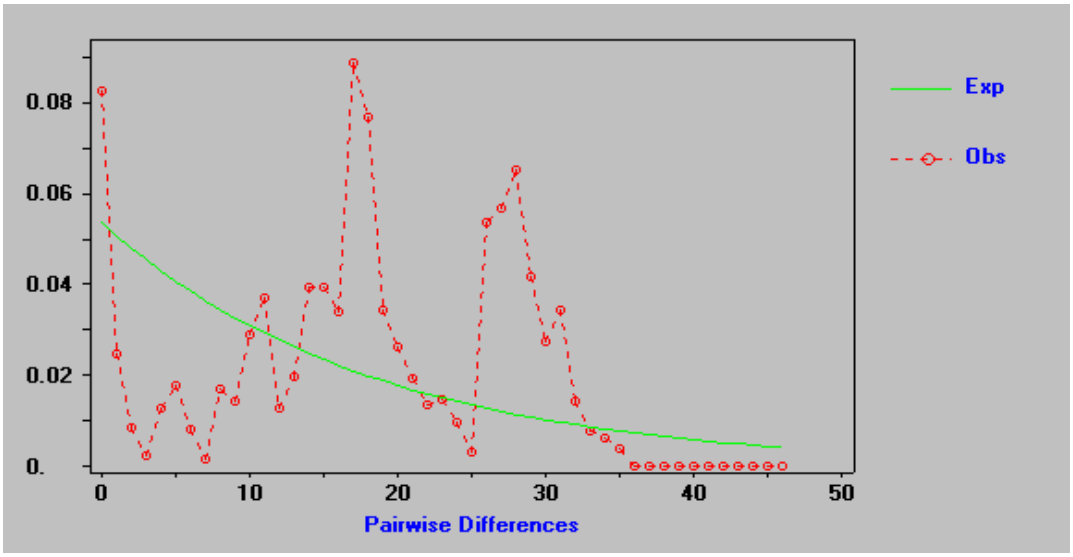
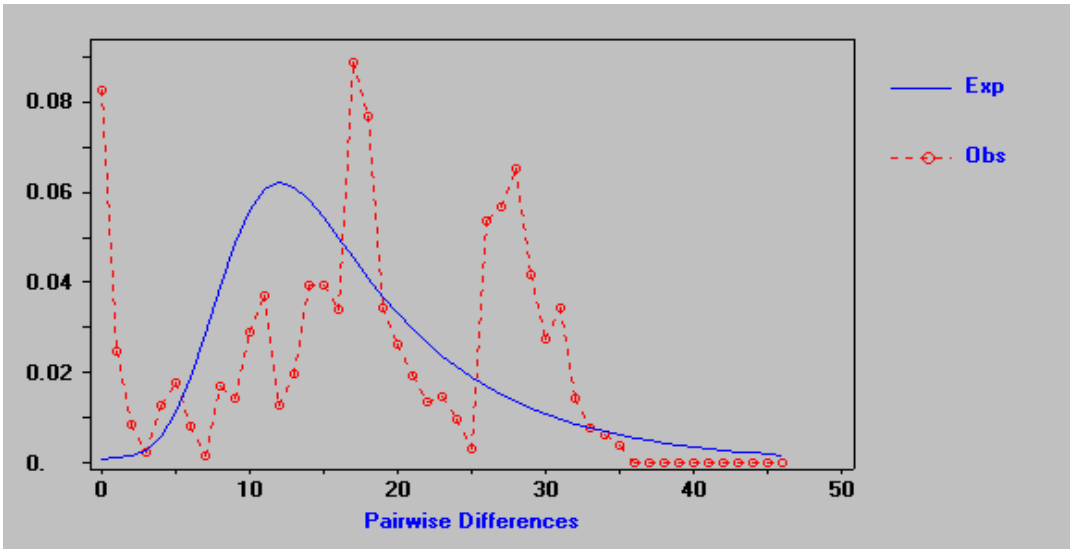


Fig. 2.6d. KNP mismatch distribution under a model of sudden expansion. SSD = 0.019, P(Sim. Ssd >= Obs. Ssd) = 0.031, Harpending's raggedness index = 0.016, P(Sim. Rag. >= Obs. Rag) < 0.001.



4. Conclusions

The advent of molecular techniques has produced the tools necessary for gaining insight into the processes that affect the maintenance of genetic variation and subsequently the dynamics of a species. This information is vital if we want to apply sound management decisions regarding the protection of biodiversity. The recent population crash due to the rinderpest epidemic has raised concerns about its potential impact on genetic variation. Current census sizes of both KNP and HiP may be misleading with regard to the ability of these populations to sustain genetic variation. This is quite evident from the fact that only 4 haplotypes were identified for HiP with concomitant low to moderate nucleotide diversity. While the current state of genetic affairs can be related to the bottleneck caused by the rinderpest, the effect of factors other than the bottleneck makes sound explanations complex. For instance, large numbers of animals were removed from both parks, which certainly could have affected the level of genetic variation to a certain degree. Furthermore, the effect that the bottleneck had on signals of a previous expansion, such as during the Pleistocene, is unknown. It is however tempting to speculate that signatures of expansion might have become lost as a result of the population crash, since the number of rare alleles that are genetically very similar and that are signatures of expansion, may have been reduced.

The moderate to high level of variation in KNP suggests that the bottleneck was overestimated. Other factors however come into play that affected variation after the bottleneck. Although it has been reported that very few animals survived the rinderpest, it is possible that the population that survived in KNP was in fact larger and that it retained a significant amount of variation from the pre-bottleneck period. It is also known that recolonization of northern KNP took place from Mozambique as well as Zimbabwe, which resulted in the immigration of rare alleles and subsequent restoration of variation (Van Hooft *et al.*, 2000). This recolonization event is also supported by statistical tests that indicate signals of admixture LD in northern KNP. The statistical tests carried out on the data show that KNP as a whole is however currently in equilibrium and no remnants of a bottleneck or an expansion, ancient or recent, can be detected. All the statistical tests carried out on HiP however point towards a population contraction, and in view of the small number of animals found in the park following the bottleneck and the large-scale removal of

animals from the park over a prolonged period of time, may have exacerbated the genetic drift. In the absence of immigration into HiP, drift will be perpetuated and it is possible that the rare haplotypes may become extinct, resulting in the fixing of the most predominant haplotype. It should be remembered however that mitochondrial sequence data gives a female-biased view of the levels of diversity, and that nuclear loci, as shown by several studies, may reveal significantly higher levels of genetic variation.

The minimum spanning network, depicting the genetic relatedness among the haplotypes, does not support a population expansion event. Clear differentiation between recent and ancient time scales and their effect on genetic parameters must however be made. By no means is it possible to detect expansion events (resulting in an increase in genetic variation) that took place during the last century, the time period that lapsed since the rinderpest bottleneck. In genetic terms it is not sufficient time to permit accumulation of mutations that may fix signatures of expansion. Effects of a bottleneck and subsequent expansion may furthermore be transient, not lasting beyond a few generations, and populations may thus reach equilibrium and stationarity again some time after a bottleneck, eradicating signals of expansion.

Van Hooft and co-workers (2000) found evidence of expansion only (Fu's Fs = -19.78, $p<0.05$) on pooling samples of Cape buffalo from different populations originating from eastern and southern Africa. They however found only slightly negative and nonsignificant values for Tajima D test of neutrality (-0.944, $p > 0.1$) when the entire region of the mtDNA that they sequenced was used in the calculation, highlighting the fact that mutation rate heterogeneity, and thus the choice of the area to be analysed, may significantly affect accurate reconstruction of the demographic history of the population. Tajima (1989a,b) has shown, through simulation results, that the higher uneven mutation rate heterogeneity shifts the D statistic towards more positive values. Tajima (1995) also showed that some forms of non-random sampling can result in positive D values, and depending on the severity of the population size change and the number of generations since that demographic event, *D* values may be transiently positive or negative (Tajima 1989b; Fay and Wu, 1999). As in excess of 250 animals were sequenced in this study and as the size D-loop region characterized exceeds that used in previous studies, the effects of

non-random sampling and sequence region on the D statistic should be less pronounced in this study.

Furnishing baseline information regarding the genetic status of the buffalo, as outlined in this chapter, is vital for understanding the differential effects that demographic processes exert on the population, and for making sound decisions regarding the genetic management of buffalo in KNP and HiP. It is evident from the research presented in this chapter that HiP exhibits strong and significant signatures of the rinderpest bottleneck, while these signatures were not as evident and in most cases insignificant in KNP.

Appendix I. DNAman (version 4.13, Lynnon Biosoft) sequence alignment of 34 haplotypes from KNP and 4 from HiP**.**

```
KNP20    CCCCCACGTT.CATGGGCCCGGAGCGAGAAGAGAAATCCCTGTCTAGCGGGTTGCTGGTTTCACGCGGCATGCTAGACAAG    80
KNP25    CCCCCACATTGTATGGGCCCGGAGCGAGAAGAGAAATCCCTGTCTAGCGGGTTGCTGGTTTCACGCGGCATGCTGAACAAG    81
KNP26    CCCCCACATTACATGGGCCCGGAGCGAGAAGAGAAATCCCTGTCTAGCGGGTTGCTGGTTTCACGCGGCATGCTAGATAAG    81
KNP27    CCCCCACATTGCATGGGCCCGGAGCGAGAAGAGAAATCCCTGTCTAGCGGGTTGCTGGTTTCACGCGGCATGCTAGGCAAG    81
KNP28    CCCCCACATTGCATGGGCCCGGAGCGAGAAGAGAAATCCCTGTCTAGCGGGTTGCTGGTTTCACGCGGCATGCTAGACAAG    81
KNP29    CCCCCACATTACATGGGCCCGGAGCGAGAAGAGAAATCCCTGTCTAGCGGGTTGCTGGTTTCACGCGGCATGCTAGGCAAG    81
KNP30    CCCCCACGTTACATGGGCCCGGAGCGAGAAGAGAAATCCCTGTCTAGCGGGTTGCTGGTTTCACGCGGCATGCTAGACAAG    81
KNP31    CCCCCACATTACATGGGCCCGGAGCGAGAAGAGAAATCCCTGTCTAGCGGGTTGCTGGTTTCACGCGGCATGCTAGACAAG    81
KNP32    CCCCCACATTGTATGGGCCCGGAGCGAGAAGAGAAATCCCTGTCTAGCGGGTTGCTGGTTTCACGCGGCATGCTGATCAAG    81
KNP33    CCCCCACATTACATGGGCCCGGAGCGAGAAGAGAAATCCCTGTCTAGCGGGTTGCTGGTTTCACGCGGCATGCTAGACAAG    81
KNP34    CCCCCACATTGTATGGGCCCGGAGCGAGAAGAGAAATCCCTGTCTAGCGGGTTGCTGGTTTCACGCGGCATGCTAAACAAG    81
KNP01    CCCCCACATTGCATGGGCCCGGAGCGAGAAGAGAAATCCCTGTCTAGCGGGTTGCTGGTTTCACGCGGCATGCTAGGCAAG    81
KNP08    CCCCCACATTACATGGGCCCGGAGCGAGAAGAGAAATCCCTGTCTAGCGGGTTGCTGGTTTCACGCGGCATGCTAGACAAG    81
KNP22    CCCCCACATTACATGGGCCCGGAGCGAGAAGAGAAATCCCTGTCTAGCGGGTTGCTGGTTTCACGCGGCATGCTAGGTAAG    81
KNP23    CCCCCACATTACATGGGCCCGGAGCGAGAAGAGAAATCCCTGTCTAGCGGGTTGTTGGTTTCACGCGGCATGCTAGACGAG    81
KNP24    CCCCCACATTGTATGGGCCCGGAGCGAGAAGAGAAATCCCTGTCTAGCGGGTTGCTGGTTTCACGCGGCATGCTAAACAAG    81
KNP02    CCCCCACATTGCATGGGCCCGGAGCGAGAAGAGAAATCCCTGTCTAGCGGGTTGCTGGTTTCACGCGGCATGCTAGACAAG    81
KNP03    CCCCCACATTGCATGGGCCCGGAGCGAGAAGAGAAATCCCTGTCTAGCGGGTTGCTGGTTTCACGCGGCATGCTAGACAAG    81
KNP04    CCCCCACATTGTATGGGCCCGGAGCGAGAAGAGAAATCCCTGTCTAGCGGGTTGCTGGTTTCACGCGGCATGCTAAACAAG    81
KNP05    CCCCCACATTACATGGGCCCGGAGCGAGAAGAGAAATCCCTGTCTAGCGGGTTGCTGGTTTCACGCGGCATGCTAGACAAG    81
KNP06    CCCCCACATTGTATGGGCCCGGAGCGAGAAGAGAAATCCCTGTCTAGCGGGTTGCTGGTTTCACGCGGCATGCTAAACAAG    81
KNP07    CCCCCACGTTGTATGGGCCCGGAGCGAGAAGAGAAATCCCTGTCTAGCGGGTTGTTGGTTTCACGCGGCATGCTAAACAAG    81
KNP21    CCCCCACATTACATGGGCCCGGAGCGAGAAGAGAAATCCCTGTCTAGCGGGTTGTTGGTTTCACGCGGCATGCTAGACAAG    81
KNP19    CCCCCACATTACATGGGCCCGGAGCGAGAAGAGAAATCCCTGTCTAGCGGGTTGCTGGTTTCACGCGGCATGCTGGACAAG    81
KNP10    CCCCCACATTACATGGGCCCGGAGCGAGAAGAGAAATCCCTGTCTAGCGGGTTGCTGGTTTCACGCGGCATGCTAGACAAG    81
KNP11    CCCCCACATTACATGGGCCCGGAGCGAGAAGAGAAATCCCTGTCTAGCGGGTTGCTGGTTTCACGCGGCATGCTAGATAAG    81
KNP12    CCCCCACGTTACATGGGCCCGGAGCGAGAAGAGAAATCCCTGTCTAGCGGGTTGCTGGTTTCACGCGGCATGCTAGACAAG    81
KNP13    CCCCCACATTACATGGGCCCGGAGCGAGAAGAGAAATCCCTGTCTAGCGGGTTGCTGGTTTCACGCGGCATGCTAGACAAG    81
KNP14    CCCCCACATTACATGGGCCCGGAGCGAGAAGAGAAATCCCTGTCTAGCGGGTTGCTGGTTTCACGCGGCATGCTAGACAAG    81
KNP15    CCCCCACATTATATGGGCCCGGGGCGAGAAGAGAAATCCCTGTCTAGCGGGTTGCTGGTTTCACGCGGCATGTTAGACAAG    81
KNP16    CCCCCACGTTGCATGGGCCCGGAGCGAGAAGAGAAATCCCTGTCTAGCGGGTTGCTGGTTTCACGCGGCATGCTAGACAAG    81
KNP17    CCCCCACATTACATGGGCCCGGAGCGAGAAGAGAAATCCCTGTCTAGCGGGTTGCTGGTTTCACGCGGCATGCTAGACAAG    81
KNP18    CCCCCACATTACATGGGCCCGGAGCGAGAAGAGAAATCCCTGTCTAGCGGGTTGCTGGTTTCACGCGGCATGCTAGGCAAG    81
KNP09    CCCCCACATTGCATGGGCCCGGAGCGAGAAGAGAAATCCCTGTCTAGCGGGTTGCTGGTTTCACGCGGCATGCTAGACAAG    81
HiP01    CCCCCACGTTGCATGGGCCCGGAGCGAGAAGAGAAATCCCTGTCTAGCGGGTTGCTGGTTTCACGCGGCATGCTAGACAAG    81
HiP03    CCCCCACGTTGCATGGGCCCGGAGCGAGAAGAGAAATCCCTGTCTAGCGGGTTGCTGGTTTCACGCGGCATGCTAGACAAG    81
HiP04    CCCCCACATTACATGGGCCCGGAGCGAGAAGAGAAATCCCTGTCTAGCGGGTTGCTGGTTTCACGCGGCATGCTAGACAAG    81
HiP02    CCCCCACATTACATGGGCCCGGAGCGAGAAGAGAAATCCCTGTCTAGCGGGTTGCTGGTTTCACGCGGCATGCTAGACAAG    81
```

```
KNP20      CTCGTGGTCTAGTGGTGGGGGAGGTCCATGCCATTAAGAGAATGCGCTTGATATAAGTTTTCATGTACAATTAATGGTAG     160
KNP25      CTCGTGATCTAGTGGTGAGGGAAGTCCATGCCATTGAAGAAATGTACTTGACGTAAGTTTTCATGTACAATTAATAATAA     161
KNP26      CTCGTGGTCTAGCGGTGGGGGAGGTCCATGCCATTGAGGGAATGTACTTGACATAGGTTTTCATGTACAGTTAATGGTGG     161
KNP27      CTCGTGGTCTAGTGGTGAGGGGAGTCCATGCCATTAAGGGGATGCGCTTGACATAAATTTTCATGTACAATTAATGGTGG     161
KNP28      CTCGTGGTCTAGTGGTGAGGGGAGTCCATGCCATTAAGGGAATGCGCTTGACATAAGTTTTCATGTACAATTAATGGTGG     161
KNP29      CTCGTGGTCTAGTGGTGAGGGAGGTTCATGTCATTAAGGGGGTACGCTTGATATAAGTTTTCATGTACAATTAATGGTAG     161
KNP30      CTCGTGGTCTAGCGGTGGGGGAAGTCCATGTCGTTGAGAGAATGTACTTGACATAAGTTTTCATGTACAATTAATGGTGG     161
KNP31      CTCGTGGTCTAGCGGTGGGGGAGGTCCATGCCATTGAGGGAATGTACTTGACATAGGTTTTCATGTACAGTTAATGGTGG     161
KNP32      CTCGTGGTCTAGCGGTGGGGGAAGTCCATGCCGT.GGAGGAATGTACTTGACATAGGTTTCCATGTACAATTAATAATAA     160
KNP33      CTCGTGATCTAGTGGTGGGGGAGGTCCATGCCATTGAGAGAATGCGCTTGATATAAGTTTTCATGTACAATTAATGGTAG     161
KNP34      CTCGTGATCTAGTGGTGAGGGAAGTCCATGCCATTGAAGAAATGTACTTGACGTAAGTTTTCATGTACAATTAATAATAA     161
KNP01      CTCGTGGTCTAGTGGTGAGGGGAGTCCATGCCATTAAGGGGATGCGCTTGACATAAATTTTCATGTACAATTAATGGTGG     161
KNP08      CTCGTGGTCTAGCGGTGGGGGAAGTCCATGTCGTTGAGGGAATGTACTTGACGTAAGTTTTCATGTACAATTAATGGTGG     161
KNP22      CTCGTGGTCTAGTGGTGGGGGAAGTCCATGCCATTGAGGGAATGCGCTTGACGTAAATTTTCATGTACAATTGATAGTGG     161
KNP23      CTCGTGGTCTAGTGGTGGGGGAAGTCCATGCCATTGAGGGAATGCACTTGACGTAAGTTTTCATGTACAATTAATGGTGG     161
KNP24      CTCGTGATCTAGTGGTGAGGGAAGTCCATGCCATTGAAGAAATGTACTTGACGTAAGTTTTCATGTACAATTAATAATAA     161
KNP02      CTCGTGGTCTAGTGGTGAGGGGAGTCCATGCCATTAAGGGAATGCGCTTGACATAAGTTTTCATGTACAATTAATGGTGG     161
KNP03      CTCGTGGTCTAGTGGTGAGGGGAGTCCATGCCATTAAGGGGATGCGCTTGACATAAGTTTTCATGTACAATTAATGGTGA     161
KNP04      CTCGTGATCTAGTGGTGAGGGAAGTCCATGCCATTGAAGAAATGTACTTGACGTAAGTTTTCATGTACAATTAATAATAA     161
KNP05      CTCGTGGTCTAGCGGTGGGGGAAGTCCATGTCGTTGAGAGAATGTACTTGACATAAGTTTTCATGTACAATTAATGGTGG     161
KNP06      CTCGTGATCTAGTGGTGAGGGAAGTCCATGCCATTGAAGAAATGTACTTGACGTAAGTTTTCATGTACAATTAATAATAA     161
KNP07      CTCGTGATCTAGTGGTGAGGGAAGTTCATGCCATTGAAGAAATGTACTTGACGTAAGTTTTCATGTACAATTAATAATAA     161
KNP21      CTCGTGGTCTAGTGGTGGGGGAAGTCCATGCCATTGAGGGAATGCACTTGACGTAAGTTTTCATGTACAATTAATGGTGG     161
KNP19      CTCGTGGTCTAGCGGTGAGGGAGGTTCATGCCATTAAGGGGGTACGCTTGATATAAGTTTTCATGTACAGTTAATGGTAG     161
KNP10      CTCGTGGTCTAGCGGTGGGGGAGGTCCATGCCATTGAGGGAATGTACTTGACATAGGCTTTCATGTACAGTTAATGGTGG     161
KNP11      CTCGTGGTCTAGCGGTGGGGGAGGTCCATGCCATTGAAGGAATGTACTTGACATAGGTTTTCATGTACAGTTAATGGTGG     161
KNP12      CTCGTGGTCTAGCGGTGGGGGAGGTCCATGCCATTGAGGGAATGTACTTGACATAGGTTTTCATGTACAGTTAATGGTGG     161
KNP13      CTCGTGGTCTAGCGGTGGGGGAGGTCCATGCCATTGAGGGAATGTACTTGACATAGGTTTTCATGTACAGTTAATGGTGG     161
KNP14      CTCGTGGTCTAGCGGTGGGGGAGGTCCATGCCATTGAGGGAATGTACTTGACATAGGTTTTCATGTACAGTTAATGGTGG     161
KNP15      CTCGTGGTCTAGCGGTGAGGGAAGTCCATGCCATTAAGAGAATGTGCTTGACATAAGTTTTCATGTACAGTTAATGGTGG     161
KNP16      CTCGTGGTCTAGTGGTGGGGGAGGTTCATGCCATTAAGGGGATGCGCTTGATATAAGTTTTCATGTACGATTAATGATGG     161
KNP17      CTCGTGGTCTAGTGGTGGGGGAGGTTCATGCCATTAGGGAAATGCGCTTGATATAAATTTTCATGTACAGTTAATGGTGG     161
KNP18      CTCGTGGTCTAGTGGTGAGGGAGGTTCATGTCATTAAGGGGGTACGCTTGATATAAGTTTTCATGTACAATTAATGGTAG     161
KNP09      CTCGTGGTCTAGCGGTGGGGGAAGTCCATGTCGTTGAGGGAATGTACTTGACATAAGCTTTCATGTACGATTAATGGTGG     161
HiP01      CTCGTGGTCTAGCGGTGGGGGAGGTCCATGTCATTGAGGGAATGTACTTGACATAGGTTTTCATGTACAGTTAATGGTGG     161
HiP03      CTCGTGGTCTAGCGGTGGGGGAGGTCCATGTCATTGAGGGAATGTACTTGACATAGGTTTTCATGTACAGTTAATGGTGG     161
HiP04      CTCGTGGTCTAGTGGTGAGGGAGGTTCATGCCATTAAGGAAATGCGCTTGATATAAATTTTCATGTGCAGTTAATGGTGG     161
HiP02      CTCGTGGTCTAGCGGTGGGGGAGGTCCATGCCATTGAGGGAATGTACTTGACATAGGTTTTCATGTACAGTTAATGGTGG     161
```

```
KNP20       TATGTACTATA.........CACTGTTAGTAAGAGCTTGAAGTATCGTACTTGCTTATATGCATGGGGTATATAATGTAA     231
KNP25       TATGTACTATAT.........ATTGTTAGTAAAAATTTGAGTTATTGTACTTGCTTATATGCATGGGGTATGTAATGTAA     232
KNP26       TATGTACTGTA.........CACTGTTAGTAAGAACTTGAAGTGTTGTACTTGCTTATATGCATGGGGTATATAATGTAA     232
KNP27       TATGTACTGTA.........CACTGTTAGTAAGAACTTGAAGTATTGTACTTGCTTATATGCATGGGGCATATAATGTAA     232
KNP28       TATGTACTGTA.........CACTGTTAGTAAAAACTTGAAGTGTTGTACTTGCTTATATGCATGGGGTATATAATATAA     232
KNP29       TATGTACTGTA.........CACTGTTAGTAAGAGCTTGAAGTATTGTACTTGCTTATATGCATGGGGTATATAATGTAA     232
KNP30       TATGTACTGTA.........CACTGTTAGTAAGAGCTTGAAATATTGTACTTGCTTATATGCATGGGGTATATAATGTAA     232
KNP31       TATGTACTGTA.........CACTGTTAGTAAGAACTTGAAGTGTTGTACTTGCTTATATGCATGGGGTATATAATGTAA     232
KNP32       CATGTACTGTG.........TACTGTTGGTGAGGATTTAGGATATTATACTTGCTTATATGCATGGGGCATATAATGTAA     231
KNP33       TATGTACTATA.........CACTGTTAGTAAGAGCTTGAAGTATTGTACTTGCTTATATGCATGGGGTATATAATGTAA     232
KNP34       TATGTACTATATGTACTATATATTGTTAGTAAAAATTTGAGTTATTGTACTTGCTTATATGCATGGGGTATGTAATGTAA     241
KNP01       TATGTACTGTA.........CACTGTTAGTAAGAACTTGAAGTATTGTACTTGCTTATATGCATGGGGCATATAATGTAA     232
KNP08       TATGTACTGTA.........CACTGTTAGTAAGAACTTGAAGTATTGTACTTGCTTATATGCATGGGGTATATAATGTAA     232
KNP22       TATGTACTGTAT........CATTGTTAGTAAGAGTTTGAAGTATTGTACTTGCTTATATGCATGGGGTATGTAATGTAA     233
KNP23       TATGTACTATA.........CACTATTAGTAAGAACTTGAAATATTGTACTTGCTTATATGCATGGGGTATATAATGTAA     232
KNP24       TATGTACTATAT.........ATTGTTAGTAAGAATTTGAGTTATTGTACTTGCTTATATGCATGGGGTATGTAATGTAA     232
KNP02       TATGTACTGTA.........CACTGTTAGTAAAAACTTGAAGTGTTGTACTTGCTTATATGCATGGGGTATATAATATAA     232
KNP03       TATGTACTGTA.........CACTGTTAGTAAGAACTTGAAGTATTGTACTTGCTTATATGCATGGGGTATATAATGTAA     232
KNP04       TATGTACTATAT.........ATTGTTAGTAAAAATTTGAGTTATTGTACTTGCTTATATGCATGGGGTATGTAATGTAA     232
KNP05       TATGTACTGTA.........CACTGTTAGTAAGAGCTTGAAATATTGTACTTGCTTATATGCATGGGGTATATAATGTAA     232
KNP06       TATGTACTATAT.........ATTGTTAGTAAAAATTTGAGTTATTGTACTTGCTTATATGCATGGGGTATGTAATGTAA     232
KNP07       TATGTACTATAT.........ATTGTTAGTAAAGGTTTGGGTTATTGTACTTGCTTATATGCATGGGGCATGTAATGTAA     232
KNP21       TATGTACTATA.........CATTATTAGTAAGAACTTGAAATATTGTACTTGCTTATATGCATGGGGCATATAATGTAA     232
KNP19       TATGTACTATA.........CACTGTTAGTAAGAGCTTGGAGTATTGTACTTGCTTATATGCATGGGGTATATAATGTAA     232
KNP10       TATGTACTGTA.........CACTGTTAGTAAGAACTTGAAGTGTTGTACTTGCTTATATGCATGGGGTATATAATGTAA     232
KNP11       TATGTACTGTA.........CACTGTTAGTAAGAACTTGAAGTGTTGTACTTGCTTATATGCATGGGGTATATAATGTAA     232
KNP12       TATGTACTGTA.........CACTGTTAGTAAGAACTTGAAGTGTTGTACTTGCTTATATGCATGGGGTATATAATGTAA     232
KNP13       TATGTACTGTA.........CACTGTTAGTAAGAACTTGAAGTGTTGTACTTGCTTATATGCATGGGGTATATAATGTAA     232
KNP14       TATGTACTGTA.........CACTGTTAGTAAGAACTTGAAGTGTTGTACTTGCTTATATGCATGGGGTATATAATGTAA     232
KNP15       TATGTACTGTA.........CACTGTTAGTAAGAACTTGAAATATTGTACTTGCTTATATGCATGGGGTATATAATGTAA     232
KNP16       TATGTACTGTA.........CACTGTTAGTAAGAGCTTAAGGTATTGTACTTGCTTATATGCATGGGGTATATAATGTAA     232
KNP17       TATGTACTATA.........CACTGTTAGTAAGAGCTTGAAGTATTGTACTTGCTTATATGCATGGGGCATGTAATGTAA     232
KNP18       TATGTACTATA.........CACTGTTAGTAAGAGCTTGAAGTATTGTACTTGCTTATATGCATGGGGTATATAATGTAA     232
KNP09       TATGCACTGTG.........CACTGTTAGTAAGAACTTGAAATATTGTACTTGCTTATATGCATGGGGCATATAATGTAA     232
HiP01       TATGTACTGTA.........CGCTGTTAGTAAGAACTTGAAGTGTTGTACTTGCTTATATGCATGGGGCATATAATGTAA     232
HiP03       TATGTACTGTA.........CGCTATTAGTAAGAACTTGAAGTGTTGTACTTGCTTATATGCATGGGGCATATAATGTAA     232
HiP04       TATGTACTGTAT........CACTGTTAGTAAGAACTTGAAGTATTGTACTTGCTTATATGCATGGGGTATGTAATGTAA     233
HiP02       TATGTACTGTA.........CACTGTTAGTAAGAGCTTGAAGTGTTGTACTTGCTTATATGCATGGGGTATATAATGTAA     232
```

```
KNP20    TGTACTATATACATATTATGTCTTTATTACATTAATGTTATGTACCTACCCCGCTTGGGGTAGTAGGTCTATGTGTTGTG   311
KNP25    TGTACTACATACATATTATGTCTTTATTACATTAATGTTATGTACCCACCCCGCTTGGGGTAGTAGGTCT.TGTGTCGTG   311
KNP26    TGTACTATATACATATCATGTCCTTATTACATTAATGTTATGTACCTACCCCGCTTGGGGTAGTAGGTCT.TGTGTCGTG   311
KNP27    TGTACTATATACATATCATGTCTTTGTTACATTAATGTTATGTACCTACCCCGCTTGGGGTAGTAGGCCT.TGTGTCGTG   311
KNP28    TGTACTATATACATATCATGTCTTTATTACATTAATGTTATGTACCTACCCCGCTTGGGGTAGTAGGCCT.TGTGTCGTG   311
KNP29    TGTACTATATACATATTATGTCTTTATTACATTAATGTTATGTACCTACCCCGCTTGGGGTAGTAGGTCTATGTGTTGTG   312
KNP30    TGTACTATATACATATCATGTCTTTATTACATTAATGTTATGTACCTACCCCGCTTGGGGTAGTAGGTCT.TGTGTCGTG   311
KNP31    TGTACTATATACATATCATGTCCTTATTACATTAATGTTATGTACCTACCCCGCTTGGGGTAGTAGGTCT.TGTGTCGTG   311
KNP32    TGTACTATATACATATCATGTCTTTATTACATTAATGTTATGTACCTATCCCGCTTGGGGTAATAGGTTT..GTGTGGTG   309
KNP33    TGTACTATATACATATTATGTCTTTATTACATTAATGTTATGTACCTACCCCGCTTGGGGTAGTAGGTCTATGTGTTGTG   312
KNP34    TGTACTATATACATATTATGTCTTTATTACATTAATGTTATGTACCCACCCCGCTTGGGGTAGTAGGTCT.TGTGTCGTG   320
KNP01    TGTACTATATACATATCATGTCTTTATTACATTAATGTTATGTACCTACCCCGCTTGGGGTAGTAGGCCT.TGTGTCGTG   311
KNP08    TGTACTATATACATATCATGTCTTTATTACATTAATGTTATGTACCTACCCCGCTTGGGGTAGTAGGTCT.TGTGTCGTG   311
KNP22    TGTACTATATACATATTATGTCTTTATTACATTAATATTATGTACATACCCCGCTTGGGGTAGTAGGTCT.TGTGTCGTG   312
KNP23    TGTACTATATACATATCATGTCTTTATTACATTAATGTTATGTACCTACCCCGCTTGGGGTAGTAGGTCT.TGTGTCGTG   311
KNP24    TGTACTACATACATATTATGTCTTTATTACATTAATGTTATGTACCCACCCCGCTTGGGGTAGTAGGTCT.TGTGTCGTG   311
KNP02    TGTACTATATACATATCATGTCCTTATTACATTAATGTTATGTACCTACCCCGCTTGGGGTAGTAGGCCT.TGTGTCGTG   311
KNP03    TGTACTATATACATATCATGTCCTTATTACATTAATGTTATGTACCTACCCCGCTTGGGGTAGTAGGCCT.TGTGTCGTG   311
KNP04    TGTACTACATACATATTATGTCTTTATTACATTAATGTTATGTACCCACCCCGCTTGGGGTAGTAGGTCT.TGTGTCGTG   311
KNP05    TGTACTATATACATATCATGTCTTTATTACATTAATGTTATGTACCTACCCCGCTTGGGGTAGTAGGTCT.TGTGTCGTG   311
KNP06    TGTACTATATACATATTATGTCTTTATTACATTAATGTTATGTACCCACCCCGCTTGGGGTAGTAGGTCT.TGTGTCGTG   311
KNP07    TGTACTATATACATATTATGTCTTTATTACATTAATGTTATGTACCTACCC.GCTTGGGGTAGTAGGTCT.TGTGTCGTG   310
KNP21    TGTACTATATACATATTATGTCTTTATTACATTAATGTTATGTACCTACCCCGCTTGGGGTAGTAGGTCT.TGTGTCGTG   311
KNP19    TGTACTATATACATATTATGTCCTTATTACATTAATGTTATGTACCTACCCCGCTTGGGGTAGTAGGTCTATGTGTTGTG   312
KNP10    TGTACTATATACATATCATGTCCTTATTACATTAATGTTATGTACCTACCCCGCTTGGGGTAGTAGGTCT.TGTGTCGTG   311
KNP11    TGTACTATATACATATCATGTCCTTATTACATTAATGTTATGTACCTACCCCGCTTGGGGTAGTAGGTCT.TGTGTCGTG   311
KNP12    TGTACTATATACATATCATGTCCTTATTACATTAATGTTATGTACCTACCCCGCTTGGGGTAGTAGGTCT.TGTGTCGTG   311
KNP13    TGTACTATATACATATCATGTCCTTGTTACATTAATATTATGTACCTACCCCGCTTGGGGTAGTAGGTCT.TGTGTCGTG   311
KNP14    TGTACTATATACATATCATGTCCTTATTACATTAATATTATGTACCTACCCCGCTTGGGGTAGTAGGTCT.TGTGTCGTG   311
KNP15    TGTACTATATACATATCATGTCTTTATTACATTAATGTTATGTACCTACCCCGCTTGGGGTAGTAGGCCT.TGTGTCGTG   311
KNP16    TGTACTATATACATATCATGTCTTTATTACATTAATGTTATGTACTTATCCCGCTTGGGGTAGTAGGTCTATGTGTTGTG   312
KNP17    TGTACTATATACATATTATGTCTTTATTACATTAATGTTATGTACCTACCCCGCTTGGGGTAGTAGGTCTATGTGTTGTG   312
KNP18    TGTACTATATACATATTATGTCTTTATTACATTAATGTTATGTACCTACCCCGCTTGGGGTAGTAGGTCTATGTGTTGTG   312
KNP09    TGTACTATATACATATCATGTCTTTATTACATTAATGTTATGTACCTATCCCGCTTGGGGTAGTAGGTCT.TGTGTCGTG   311
HiP01    TGTACTATATACATATTATGTCCTCATTACATTAATGTTATGTACCTACCCCGCTTGGGGTAGTAGGTCT.TGTGTCGTG   311
HiP03    TGTACTATATACATATTATGTCCTCATTACATTAATGTTATGTACCTACCCCGCTTGGGGTAGTAGGTCT.TGTGTCGTG   311
HiP04    TGTACTATATACATATTATGTCTTTATTACATTAATGTTATGTACCTACCCCGCTTGGGGTAGTAGGTCTATGTGTTGTG   313
HiP02    TGTACTATATACATATCATGTCCTTATTACATTAATATTATGTACCTACCCCGCTTGGGGTAGTAGGTCT.TGTGTCGTG   311
```

```
KNP20    TGCTGTGGGTTCGTAGTGGTGTGTTGAGGTTGATAGTTATTTTTAATGTTGGGTTAGAGTGCAAAGTCTGTGTTGAATTG    391
KNP25    TGCTGTGGGTTCATAGTGGTGTGTTGAGGTTGATAGTTATTTTTAATGTTGGGTTAGAGTGCAAAGTCTGTGTTGAATTG    391
KNP26    TGCTGTGGGTTCGTAGTGGTGTGTTGAGGTTGATAGTTATTTTTAATGTTGGGTTAGAGTGCAAAGTCTGTGTTGAATTG    391
KNP27    TGCTGTGGGTTTGTAGTGGTGTGTTGAGGTTGGTAGTTATTTTTAATGTTGGGTTAGAGTGCAAAGTCTGTGTTGAATTG    391
KNP28    TGCTGTGGGTTTGTAGTGGTGTGTTGAGGTTGATAGTTATTTTTAATGTTGGGTTAGAGTGCAAAGTCTGTGTTGAATTG    391
KNP29    TGCTGTGGGTTCGTAGTGGTGTGTTGAGGTTGATAGTTATTTTTAATGTTGGGTTAGAGTGCAAAGTCTGTGTTGAATTG    392
KNP30    TGCTGTGGGTTCGTAGTGGTGTGTTGAGGTTGATAGTTATTTTTAATGTTGGGTTAGAGTGCAAAGTCTGTGTTGAATTG    391
KNP31    TGCTGTGGGTTCGTAGTGGTGTGTTGAGGTTGATAGTTATTTTTAATGTTGGGTTAGAGTGCAAAGTCTGTGTTGAATTG    391
KNP32    CGCTGTGGGTTTGTAATGGTGTGTTGAGGTTGATAATTATTTTTAATGTTGGGCTAGAGTGCAAAGTCTGTGTTGAATTG    389
KNP33    TGCTGTGGGTTCGTAGTGGTGTGTTGAGGTTGATAGTTATTTTTAATGTTGGGTTAGAGTGCAAAGTCTGTGTTGAATTG    392
KNP34    TGCTGTGGGTTCATAGTGGTGTGTTGAGGTTGATAGTTATTTTTAATGTTGGGTTAGAGTGCAAAGTCTGTGTTGAATTG    400
KNP01    TGCTGTGGGTTTGTAGTGGTGTGTTGAGGTTGGTAGTTATTTTTAATGTTGGGTTAGAGTGCAAAGTCTGTGTTGAATTG    391
KNP08    TGCTGTGGGTTCGTAGTGGTGTGTTGAGGTTGATAGTTATTTTTAATGTTGGGTTAGAGTGCAAAGTCTGTGTTGAATTG    391
KNP22    TGCTGTGGGTTCGTAGTGGTGTGTTGAGGTTGATAGTTATTTTTAATGTTGGGTTAGAGTGCAAAGTCTGTGTTGAATTG    392
KNP23    TGCTGTGGGTTCGTAGTGGTGTGTTGAGGTTGATAGTTATTTTTAATGTTGGGTTAGAGTGCAAAGTCTGTGTTGAATTG    391
KNP24    TGCTGTGGGTTCATAGTGGTGTGTTGAGGTTGATAGTTATTTTTAATGTTGGGTTAGAGTGCAAAGTCTGTGTTGAATTG    391
KNP02    TGCTGTGGGTTTGTAGTGGTGTGTTGAGGTTGATAGTTATTTTTAATGTTGGGTTAGAGTGCAAAGTCTGTGTTGAATTG    391
KNP03    TGCTGTGGGTTTGTAGTGGTGTGTTGAGGTTGGTAGTTATTTTTAATGTTGGGTTAGAGTGCAAAGTCTGTGTTGAATTG    391
KNP04    TGCTGTGGGTTCATAGTGGTGTGTTGAGGTTGATAGTTATTTTTAATGTTGGGTTAGAGTGCAAAGTCTGTGTTGAATTG    391
KNP05    TGCTGTGGGTTCGTAGTGGTGTGTTGAGGTTGATAGTTATTTTTAATGTTGGGTTAGAGTGCAAAGTCTGTGTTGAATTG    391
KNP06    TGCTGTGGGTTCATAGTGGTGTGTTGAGGTTGATAGTTATTTTTAATGTTGGGTTAGAGTGCAAAGTCTGTGTTGAATTG    391
KNP07    TGCTGTGGGCTCATAGTGGTGTGTTGAGGTTGATAGTTATTTTTAATGTTGGGTTAGAGTGCAAAGTCTGTGTTGAATTG    390
KNP21    TGCTGTGGGTTCGTAGTGGTGTGTTGAGGTTGATAGTTATTTTTAATGTTGGGTTAGAGTGCAAAGTCTGTGTTGAATTG    391
KNP19    TGCTGTGGGTTCGTAGTGGTGTGTTGAGGTTGATAGTTATTTTTAATGTTGGGTTAGAGTGCAAAGTCTGTGTTGAATTG    392
KNP10    TGCTGTGGGTTCGTAGTGGTGTGTTGAGGTTGATAGTTATTTTTAATGTTGGGTTAGAGTGCAAAGTCTGTGTTGAATTG    391
KNP11    TGCTGTGGGTTCGTAGTGGTGTGTTGAGGTTGATAGTTATTTTTAATGTTGGGTTAGAGTGCAAAGTCTGTGTTGAATTG    391
KNP12    TGCTGTGGGTTCGTAGTGGTGTGTTGAGGTTGATAGTTATTTTTAATGTTGGGTTAGAGTGCAAAGTCTGTGTTGAATTG    391
KNP13    TGCTGTGGGTTCGTAGTGGTGTGTTGAGGTTGATAGTTATTTTTAATGTTGGGTTAGAGTGCAAAGTCTGTGTTGAATTG    391
KNP14    TGCTGTGGGTTCGTAGTGGTGTGTTGAGGTTGATAGTTATTTTTAATGTTGGGTTAGAGTGCAAAGTCTGTGTTGAATTG    391
KNP15    TGCTGTGGGTTCGTAGTGGTGTGTTGAGGTCGATAGTTATTTTTAATGTTGGGTTAGAGTGCAAAGTCTGTGTTGAATTG    391
KNP16    TGCTGTGGGTTCGTAGTGGTGTGTTGAGGTTGATAGTTATTTTTAATATTGGGTTAGAGTGCAAAGTCTGTGTTGAATTG    392
KNP17    TGCTGTGGGTTCGTAGTGGTGTGTTGAGGTTGATAGTTATTTTTAATGTTGGGTTAGAGTGCAAAGTCTGTGTTGAATTG    392
KNP18    TGCTGTGGGTTCGTAGTGGTGTGTTGAGGTTGATAGTTATTTTTAATGTTGGGTTAGAGTGCAAAGTCTGTGTTGAATTG    392
KNP09    TGCTGTGGGTTCGTAGTGGTGTGTTGAGGTTGATAGTTATTTTTAATGTTGGGTTAGAGTGCAAAGTCTGTGTTGAATTG    391
HiP01    TGCTGTGGGTTCGTAGTGGTGTGTTGAGGTTGATAGTTATTTTTAATGTTGGGTTAGAGTGCAAAGTCTGTGTTGAATTG    391
HiP03    TGCTGTGGGTTCGTAGTGGTGTGTTGAGGTTGATAGTTATTTTTAATGTTGGGTTAGAGTGCAAAGTCTGTGTTGAATTG    391
HiP04    TGCTGTGGGTTCGTAGTGGTGTGTTAAGGTTGATAGTTATTTTTAATGTTGGGTTAGAGTGCAAAGTCTGTGTTGAATTG    393
HiP02    TGCTGTGGGTTCGTAGTGGTGTGTTGAGGTTGATAGTTATTTTTAATGTTGGGTTAGAGTGCAAAGTCTGTGTTGAATTG    391
```

```
KNP20       TTGCAAGTTTTAGTGAATTTAATACTGGGGAGGCTCTTTATATT    435
KNP25       TTGCAAGTTTTAGTAAATTTAATACTGGGGAGGCTCTTTATATT    435
KNP26       TTGCAAGTTTTAGTGAATTTAATACTGGGGAGGCTCTTTATATT    435
KNP27       TTGCAAGTTTTAGTGAATTTAATACTGGGGAGGCTCTTTATATT    435
KNP28       TTGCAAGTTTTAGTGAATTTAATACTGGGGAGGCTCTTTATATT    435
KNP29       TTGCAAGTTTTAGTGAATTTAATACTGGGGAGGCTCTTTATATT    436
KNP30       TTGCAAGTTTTAGTGAATTTAATACTGGGGAGGCTCTTTATATT    435
KNP31       TTGCAAGTTTTAGTGAATTTAATACTGGGGAGGCTCTTTATATT    435
KNP32       TTGCAAGTTTTAGTGAATTTAATACTGGGGAGGCTCTTTATATT    433
KNP33       TTGCAAGTTTTAGTGAATTTAATACTGGGGAGGCTCTTTATATT    436
KNP34       TTGCAAGTTTTAGTAAATTTAATACTGGGGAGGCTCTTTATATT    444
KNP01       TTGCAAGTTTTAGTGAATTTAATACTGGGGAGGCTCTTTATATT    435
KNP08       TTGCAAGTTTTAGTGAATTTAATACTGGGGAGGCTCTTTATATT    435
KNP22       TTGCAAGTTTTAGTGAATTTAATACTGGGGAGGCTCTTTATATT    436
KNP23       TTGCAAGTTTTAGTGAATTTAATACTGGGGAGGCTCTTTATATT    435
KNP24       TTGCAAGTTTTAGTAAATTTAATACTGGGGAGGCTCTTTATATT    435
KNP02       TTGCAAGTTTTAGTGAATTTAATACTGGGGAGGCTCTTTATATT    435
KNP03       TTGCAAGTTTTAGTGAATTTAATACTGGGGAGGCTCTTTATATT    435
KNP04       TTGCAAGTTTTAGTAAATTTAATACTGGGGAGGCTCTTTATATT    435
KNP05       TTGCAAGTTTTAGTGAATTTAATACTGGGGAGGCTCTTTATATT    435
KNP06       TTGCAAGTTTTAGTAAATTTAATACTGGGGAGGCTCTTTATATT    435
KNP07       TTGCAAGTTTTAGTAAATTTAATACTGGGGAGGCTCTTTATATT    434
KNP21       TTGCAAGTTTTAGTGAATTTAATACTGGGGAGGCTCTTTATATT    435
KNP19       TTGCAAGTTTTAGTGAATTTAATACTGGGGAGGCTCTTTATATT    436
KNP10       TTGCAAGTTTTAGTGAATTTAATACTGGGGAGGCTCTTTATATT    435
KNP11       TTGCAAGTTTTAGTGAATTTAATACTGGGGAGGCTCTTTATATT    435
KNP12       TTGCAAGTTTTAGTGAATTTAATACTGGGGAGGCTCTTTATATT    435
KNP13       TTGCAAGTTTTAGTGAATTTAATACTGGGGAGGCTCTTTATATT    435
KNP14       TTGCAAGTTTTAGTGAATTTAATACTGGGGAGGCTCTTTATATT    435
KNP15       TTGCAAGTTTTAGTGAATTTAATACTGGGGAGGCTCTTTATATT    435
KNP16       TTGCAAGTTTTAGTGAATTTAATACTGGGGAGGCTCTTTATATT    436
KNP17       TTGCAAGTTTTAGTGAATTTAATACTGGGGAGGCTCTTTATATT    436
KNP18       TTGCAAGTTTTAGTGAATTTAATACTGGGGAGGCTCTTTATATT    436
KNP09       TTGCAAGTTTTAGTGAATTTAATACTGGGGAGGCTCTTTATATT    435
HiP01       TTGCAAGTTTTAGTGAATTTAATACTGGGGAGGCTCTTTATATT    435
HiP03       TTGCAAGTTTTAGTGAATTTAATACTGGGGAGGCTCTTTATATT    435
HiP04       TTGCAAGTTTTAGTGAATTTAATACTGGGGAGGCTCTTTATATT    437
HiP02       TTGCAAGTTTTAGTGAATTTAATACTGGGGAGGCTCTTTATATT    435
```

Chapter 3

Development and evaluation of a multiplex STR system for genotyping Cape buffalo

Abstract

Estimating genetic parameters, of which the level of genetic variation within and among buffalo populations is one of the most important, is central to both population and conservation genetics. In fact, management strategies aimed at the sustainable utilization and conservation of Cape buffalo is to a large extent dependant on the availability of information pertaining to indices of genetic variation. Developments in the field of molecular genetics during the past few decades have resulted in the availability of microsatellite markers that are valuable tools for furnishing such information. These markers can also serve a vital role in the field of animal forensics, paternity verification and traceability studies. From a forensic point of view, individual animal identification is crucial for combating poaching on the one hand while tracing the illegal translocation of animals or their products on the other. Illegal translocation of animals is of particular importance when it threatens the internationally recognised disease-free zones in South Africa.

Population genetic studies using microsatellite (Msat) markers however generally require genotyping many samples with numerous markers, which inevitably is costly, time consuming and laborious. There is therefore a need for a high-throughput, high-resolution typing system for this species that is both cost and time effective. This chapter reports on the development and evaluation of a panel of 17 cattle Msat markers, amplified in a multiplex system and co-electrophoresed as a single injection on a capillary-based DNA sequencer, for genotyping Cape buffalo.

The robustness of the typing system was confirmed by the fact that artefacts of multiplexing such as allelic dropout and the amplification of multi-allelic peaks were minimal.

Considerable genetic variation was furthermore uncovered for sample populations of Cape buffalo and Bonsmara cattle. The number of alleles per locus for the buffalo varied between 2 and 15, with the mean number of alleles per locus (allelic diversity) being 8.24 (SD = 4.12). For the Bonsmara, the number of alleles varied from 3 to 10, with a mean allelic diversity of 6.47 (SD:2.1). Nei's unbiased gene diversity was 0.63 (SD: 0.05) and 0.67 (SD: 0.03) for buffalo and Bonsmara cattle, respectively. Four loci and one locus, respectively for the buffalo and Bonsmara exhibited heterozygote deficiencies. Minor non-random association of alleles between different loci was observed between 4 pairs of loci for the buffalo. The total combined exclusionary power with either one or both parents being genotyped was 0.9994 and 0.9999 respectively for the buffalo and 0.9977 and 0.9999 respectively for the Bonsmara. The combined cumulative probability of identity or PI (probability that two individuals in the population share identical genotypes) was 6.473×10^{-17} for the buffalo and 1.03×10^{-16} for the Bonsmara respectively, confirming that this panel is well-suited for individual animal identification in forensic traceability studies.

Our results show that although multiplexing is technically challenging, when optimised it offers a cost effective, less time consuming and less laborious option to conventional single locus typing. This, together with the ease with which this technique was successfully applied to the Bonsmara, indicates that this approach may have a broader applicability to other members of the Bovini tribe.

Part of the results included in this chapter has been submitted for publication to: *African Journal of Biotechnology*

1. Introduction

One of the major applications of molecular genetics in general is to uncover variation within and among populations, and the distribution of allele frequencies within and among populations form the basis for estimating several genetic parameters. Numerous types of markers are available for calculation of these parameters, but with the advent of the Polymerase Chain Reaction (PCR) and the availability of nuclear markers such as microsatellites (Msats), large-scale genetic screening of individuals has become a reality. Information pertaining to the genetic status of populations can subsequently be furnished timeously and on a large scale. Although Msat profiling is not without its drawbacks, it has proven very useful and suitable for a range of intended studies, especially for analysis of entire populations (Bruford and Wayne, 1993). Msats have subsequently also become the marker of choice for individual genetic identification and parentage analysis, estimation of population diversity and differentiation of populations, forensics casework studies, calculation of genetic distances and subsequent genetic relationships among population and genetic mapping (Weissenbach *et al.*, 1992; Barendse *et al.*, 1994; Gotelli *et al.*, 1994; Morin *et al.*, 1994; Barker *et al.*, 1997; Beaumont and Bruford, 1999; Van Hooft *et al.*, 2000). Msats are furthermore suitable for the estimation of effective population size (Allen *et al*., 1995), and for assessing the degree of population structure and migration between populations (Gotelli *et al*., 1994).

Msats have filled the gaps in many instances where other types of markers were limited due to their inherent characteristics. The hypervariable mutation rates of Msats enable them to reveal levels of variation superseding that found through analysis of allozymes and mitochondrial genome analysis (Paetkau and Strobeck, 1994; Estoup *et al.,* 1995a, 1995b, 1996). Their ability to detect reduced gene flow and subtle population structure also supersedes that of mtDNA and allozyme markers (Hughes & Queller 1993; Jarne *et al.,* 1994 - reviewed in Estoup & Angers 1998). As a result, Msats profiling has been applied to numerous species, ranging from reptiles to mammals (Saitbekova *et al.*, 1999; Diez-Tiascon *et al*., 2000; Hanslik *et al*., 2000; Ivankovic *et al*., 2002; Mburu *et al*., 2002; Cronin *et al*., 2003; Malone *et al*., 2003; Li *et al.*, 2004). A very useful feature of Msats is that sequences flanking the repeat unit are often conserved across related species, permitting cross-species amplification of the marker loci (Moore *et al.*, 1991; Schlotterer *et al.*, 1991; Mommens *et*

al., 1998, Diez-Tascon *et al.*, 2000; Cronin *et al.*, 2003). Cattle and buffalo, both belonging to the tribe Bovini, of the subfamily Bovinae, are related by an ancestor from which they diverged more than 4 million years ago (Buntjer *et al*., 1997; Gatesy *et al*., 1997), and a number of loci isolated from cattle have been shown to be conserved in buffalo (O'Ryan *et al.*, 1998; Simonsen *et al*., 1998; Van Hooft *et al.*, 1999). The number of loci used to profile buffalo in previous studies was however generally low and as each locus was amplified and eletrophoresed individually (O'Ryan *et al.*, 1998; Simonsen *et al*., 1998; Van Hooft *et al.*, 1999), the typing was also costly and time-consuming.

Large-scale profiling of bovine microsatellites in buffalo should permit assessment of various genetic parameters that are associated with dynamics of the buffalo populations of KNP and HiP. The ability to furnish these population genetic parameters will be of great value to the emerging buffalo breeding industry that aims to breed disease-free animals and to re-introduce animals to areas formerly within their distributional range (Hofmeyr, 2005).

Executing a population genetic study on a large scale however requires typing of a large number of samples using numerous Msat markers. The latter can be very costly, time-consuming and laborious. Msat analysis however lends itself favourably towards automation, and it is possible to design multiplex reactions that include several loci in a single PCR reaction and to co-electrophorese the resultant amplicons. This means that financial costs associated with processing large sample sizes need not necessarily be an impediment. Furthermore, considerable gains are made with regard to the time it takes to process the samples as well as the cost of running multiplex reactions as opposed to analysing samples individually.

The aim of the research presented in this chapter was to select a panel of bovine Msats that would be suitable for a population genetics study of Cape buffalo, with the added advantage of carrying out paternity verification and individual identification. A secondary aim was to ensure that these Msats could be amplified in multiplex reactions and subsequently co-eletrophoresed as a single injection in a capillary gel system. Once established, the suitability of the panel for profiling a cattle breed developed in South Africa was attempted

in order to assess the broader applicability of this approach to both domestic and wild members of the Bovini.

2. Materials and methods

2.1 Samples

DNA extracted from blood samples that were collected from 60 individuals (culled during the 1998 buffalo BTB survey in KNP) were kindly provided by Paul van Helden (University of Stellenbosch). An equal number of male and female individuals presumed to be unrelated (2 individuals from each herd), on the basis of the sampling design, and originating from 30 geographically separated herds distributed throughout the park, were selected at random and included in the sample population. In addition, DNA was extracted from hair samples of 34 unrelated Bonsmara, a cattle breed developed in South Africa that comprises 5/8 indigenous Afrikaner and 3/8 exotic Shorthorn/Hereford (Bergh and Gerhard, 2000). Non-relatedness among the Bonsmara cattle was verified with the cattle ISAG panel and the program Prest (McPeek and Sun, 2000).

2.2 Loci selected

Primers were selected from the literature based on their level of polymorphism, ability to be co-amplified, ease of scoring and allelic size ranges. The characteristics and primer sequences for the respective loci are shown in Table 3.2. Four additional loci, *viz*. ABS10, BM2113, INRA23 and TGLA53 were initially also selected to form part of the panel, but these were omitted mainly due to either the amplification of multi-allelic peaks or the fact that their allelic size ranges overlapped with other members of the panel. The latter made it technically impossible to run these loci as part of a single injection on the DNA sequencer, which formed part of the aims of the research presented in this chapter. Three multiplex core reactions were designed from the selected panel of 17 Msats (Table 3.1).

Table 3.1. Loci constituents of core multiplex reactions.

Multiplex M1	Multiplex M2	Multiplex M3
TGLA227	TGLA57	BM3517
BM1824	DIK020	BM719
ETH225	INRA006	ILSTS026
ETH10	TGLA263	BM3205
SPS115	BM4028	CSSM19
	INRA128	TGLA159

2.3 PCR

Prior to amplification of the entire sample set with the three core PCR plexes, single locus amplifications were carried out on 5 individual animals in order to establish optimum cycling conditions and concentrations of PCR constituents. This was followed by amplification of the same individuals with the respective three core plexes and reaction conditions were optimised in order to establish a balance between amounts of amplification product generated for the different loci (scored in terms of peak heights). Optimum conditions were regarded as those where sufficient amplification products were generated while allelic dropout was not observed and whilst minimising artefacts such as non-specific amplification.

The 17 autosomal microsatellite markers were size-selected and fluorescently labelled to permit co-amplification in the three multiplex reactions. The characteristics and primer sequences for the respective loci are shown in Table 3.2. Each multiplex reaction contained 50 –100 ng of genomic DNA, 2 units of *Taq* DNA polymerase (Super Therm Gold, Southern Cross Biotechnology), 10nM to 80nM of each primer, 1.5 mM $MgCL_2$, 300 µM of dNTP and 1X Super Therm Gold buffer (Southern Cross Biotechnology) in a final reaction volume of 10 µl. An enzyme activation step at 94°C for 10 min preceded 35 cycles of 94°C for 45 seconds, 58°C for 45 seconds and 72°C for 1 minute, and a final extension step at 72°C for one hour. Core multiplexes were diluted 5 fold with water prior to pooling 1 µl of

each multiplex with 7 µl of loading mix (0.0125 ul Liz size standard/µl formamide). Reaction mixtures were loaded as a single injection onto an ABI 3100 DNA sequencer following denaturation at 94^{o}C for 4 minutes. Amplification products were analysed using Genescan Analysis software version 3.7 and Genotyper 3.7 (Applied Biosystems).

TABLE 3.2. Summary of the 17 cattle Msat markers used to type African buffalo *S. caffer* detailing the bovine chromosome marker location and primer sequence, label and concentration used for each multiplex

Msat ID	Label	Plex	Chrom no.	Forward primer sequence (5' to 3')	Reverse primer sequence (5' to 3')	Primer Conc
[1]BM1824	Fam	M1	1	gAg CAA ggT gTT TTT CCA ATC	CAT TCT CCA ACT gCT TCC TTg	20
[1]ETH225	Vic	M1	29	gAT CAC CTT gCC ACT ATT TCC T	ACA TgA CAg CCA gCT gCT ACT	10
[5]ETH 10	Ned	M1	5	gTT CAg gAC Tgg CCC TgC TAA CA	CCT CCA gCC CAC TTT CTC TTC TC	17
[7]SPS115	Pet	M1	15	AAA gTg ACA CAA CAg CTT CTC Cag	AAC gAg TgT CCT AgT TTg gCT gTg	50
[9]TGLA227	Fam	M1	18	CgA ATT CCA AAT CTg TTA ATT TgC T	ACA gAC AgA AAC TCA ATg AAA gCA	67
[1]BM 4028	Ned	M2	29	ACg gAA gCA gCA TCT CTT AC	ATg gAA ACA Tgg TCT CCT gC	20
[1]INRA 006*	Fam	M2	3	Agg AAT ATC TgT ATC AAC CTC AgT C	CTg AgC Tgg ggT ggg AgC TAT AAA TA	30
[4]DIK 020*	Vic	M2	10	AAC CAg TAA TCg TgA gAg gA	AAg AAA gTC CCT ACC ATg Ag	50
[7]TGLA 263*	Pet	M2	3	CAA gTg CTg gAT ACT ATC TgA gCA	TTA AAg CAT CCT CAC CTA TAT ATg C	80
[8]INRA 128*	Ned	M2	1	TAA gCA CCg CAC AgC AgA TgC	AgA CTA gTC Agg CTT CCT AC	35
[9]TGLA 057*	Vic	M2	1	gCT TTT TAA TCC TCA gCT TgC Tg	gCT TCC AAA ACT TTA CAA TAT gTA T	35
[1]BM 3205*	Pet	M3	1	TCT TgC TTC CTT CCA AAT CTC	TgC CCT TAT TTT AAC AgT CTg C	25
[1]BM 3517*	Ned	M3	20	gTg TgT Tgg CAT CTg gAC Tg	TgT CAA ATT CTA TgC Agg ATg g	30
[1]BM 719*	Ned	M3	16	TTC TgC AAA Tgg gCT AgA gg	CAC ACC CTA gTT TgT AAG Cag C	30
[9]CSSM 19*	Fam	M3	1	TTg TCA gCA ACT TCT TgT ATC TTT	TgT TTT AAg CCA CCC AAT TAT TTg	30
[6]ILSTS026*	Pet	M3	2	CTg AAT Tgg CTC CAA Agg CC	AAA CAg AAg TCC Agg gCT gC	55
[9]TGLA 159*	Vic	M3	4	gCA TCC Agg gAA CAA ATT ACA AAC	TTT ATT TCg AAT CTC TTg AgT ACA g	35

References for the Msat markers were as follows: [1] Bishop *et al.*, (1994); [2] Schnabel *et al.*, (2000) (http://sol.marc.usda.gov); [3,9]Barendse *et al.*, (1994); [4] Hirano *et al.*, (1996); [5] Luikart *et al.*, (1999); [6] Kemp *et al.*, (1995); [7] Mommens *et al.*, (1998); [8] Vaiman *et al.*, (1994b). * Denotes markers common to this study and that of Van Hooft *et al.*, (2000).

2.4 Statistical analyses

Allele frequency distributions and general diversity indices were calculated with MS Toolkit (Park, 2001). Observed and expected heterozygosities (unbiased and calculated from allele frequencies assuming Hardy-Weinberg equilibrium) and null allele frequencies were calculated with the software package CERVUS 2.0 (Marshal *et al*., 1998). This package is freely available at http:/helios.bto.ed.ac.uk/evolgen. Provided that at least 10 individuals are typed, Cervus estimates the frequency of any null allele segregating at each locus, using an iterative algorithm based on the difference between observed and expected frequency of homozygotes. In the absence of a null allele, the estimated frequency will be close to zero, and may be slightly negative (negative values imply an excess of observed heterozygote genotypes). A locus with a large positive estimate of null allele frequency (large relative to other loci in the analysis) indicates an excess of homozygotes. It does not however necessarily imply that a null allele is present in such a case, and this will be discussed in more detail later on in this chapter.

Deviations from Hardy-Weinberg (HW) proportions as a result of heterozygote deficiencies were determined with Genepop 3.3. (Raymond and Rousset, 1995). The program *c*alculates an exact test for markers with four and less alleles, while for more alleles it runs the Markov Chain Method to obtain an unbiased estimate of the exact probability (Guo and Tompson, 1992). The non-random associations between genotypes across all loci and subsequent Bonferroni correction was calculated with FSTAT (Goudet, 2001), using Fisher's exact test. Weir and Cockerham's estimator of F_{IS} (or smallf; Weir and Cockerham, 1984) was estimated for each allele, locus and overall with FSTAT.

Exclusion probabilities with regard to parentage verification were also calculated with CERVUS, and such probabilities may be calculated taking into account that the genotypes of either one or both parents are known. The program calculates the average probability of excluding a single randomly-chosen unrelated individual from parentage for each of these two cases. The average exclusion probability is calculated by summing the individual exclusion probabilities across all combinations of genotypes, weighted by genotype frequencies and assuming Hardy-Weinberg equilibrium. The polymorphic information

content or PIC, which is a measure of informativeness related to expected heterozygosity, was also calculated with CERVUS.

The probability of identity or PI (probability that two individuals in the population share identical genotypes) was computed with the program GIMLET (Valière, 2002). Unbiased PI values were computed for each locus (PI unbias/loc) and across the entire panel {Prod (PI unbias)} by multiplying sequentially the PI values for the individual loci using the equations of Waits *et al.,* (2001).

3. Results and Discussion

3.1 Measures of genetic diversity

Measures of genetic diversity for both buffalo and Bonsmara cattle are summarised in Table 3.3, while allele frequency distributions for the buffalo and cattle are shown in appendix A and B, respectively. The large sample size of 60 unrelated individuals, selected at random, should increase the reliability of gene frequency estimates and also closely reflect that of the entire population (Lewis *et al*., 2004). The number of alleles (k) per locus for the buffalo varied between 2 and 15, with the mean number of alleles per locus (allelic diversity) being 8.24 (SD = 4.12), reflecting the high level of overall polymorphism of the loci. For the Bonsmara cattle, the number of alleles varied from 3 to 10, while the mean allelic diversity was 6.47 (SD:2.1). Nei's unbiased gene diversity was 0.63 (SD: 0.05) and 0.67 (SD: 0.03) for buffalo and Bonsmara cattle, respectively. Gene diversities for individual loci for the buffalo ranged between 0.14 and 0.89 and between 0.24 and 0.81 for the Bonsmara. The overall inbreeding coefficient (F_{IS}) was 0.028 and 0.037 for the buffalo and Bonsmara respectively. The PIC or polymorphic information content, a measure of the informativeness of a locus and which is a function of the expected heterozygosity, ranged from 0.13 to 0.87 (mean = 0.61) for the buffalo and from 0.22 to 0.78 for the Bonsmara cattle (mean = 0.63).

Table 3.3. Diversity indices for a sample population of 60 buffalo. K: allelic richness; H_o: Observed heterozygosity; H_e: Expected heterozygosity (Nei, 1987); PIC: polymorphic information content; Excl(1): exclusion probability with one parent genotyped; Excl(2): exclusion probability with both parents genotyped; bp: base pairs, F_{IS}: inbreeding coefficient (indices for Bonsmara cattle given in brackets).

Locus	K	Allelic range	H_o	H_e	PIC	Excl(1)	Excl(2)	Null allele Freq.	F_{IS}
BM1824	14 (6)	169-199 (177-191)	0.783 (0.719)	0.878 (0.777)	0.860 (0.730)	0.600 (0.373)	0.751 (0.552)	0.053 (0.034)	0.109 (0.076)
CSSM19	11 (7)	128-154(134-154)	0.783 (0.441)	0.788 (0.572)	0.754 (0.534)	0.415 (0.183)	0.593 (0.358)	-0.0006 (0.1145)	0.006 (0.231)
INRA006	7 (6)	107-119 (101-115)	0.617 (0.647)	0.621 (0.689)	0.573 (0.635)	0.214 (0.268)	0.385 (0.443)	0.015 (0.031)	0.007 (0.062)
TGLA227	4 (10)	70-76 (76-96)	0.350 (0.765)	0.442 (0.819)	0.401 (0.788)	0.100 (0.468)	0.238 (0.644)	0.112 (0.028)	0.210 (0.067)
DIK20	15 (6)	164-208 (172-184)	0.898 (0.788)	0.894 (0.737)	0.876 (0.681)	0.630 (0.318)	0.774 (0.493)	-0.008 (-0.040)	-0.005 (-0.071)
ETH225	2 (7)	133-137 (135-155)	0.467 (0.697)	0.410 (0.818)	0.324 (0.779)	0.082 (0.445)	0.162 (0.622)	-0.069 (0.073)	-0.141 (0.15)
TGLA159	8 (5)	223-237 (209-243)	0.700 (0.467)	0.799 (0.645)	0.763 (0.585)	0.421 (0.226)	0.599 (0.392)	0.059 (0.145)	0.124 (0.28)
TGLA57	7 (5)	75-101 (83-97)	0.783 (0.706)	0.779 (0.641)	0.737 (0.570)	0.381 (0.218)	0.559 (0.372)	-0.007 (-0.059)	-0.005 (-0.102)
BM3517	7 (10)	84-96 (98-118)	0.533 (0.727)	0.544 (0.725)	0.506 (0.688)	0.163 (0.333)	0.330 (0.520)	0.012 (-0.004)	0.019 (-0.003)
BM4028	3 (10)	126-134 (102-124)	0.117 (0.824)	0.143 (0.818)	0.137 (0.781)	0.010 (0.452)	0.072 (0.627)	0.0891 (-0.0085)	0.184 (-0.007)
BM719	12 (8)	136-160 (140-158)	0.867 (0.706)	0.835 (0.740)	0.807 (0.700)	0.493 (0.343)	0.664 (0.528)	-0.020 (0.013)	-0.038 (0.047)
ETH10	2 (7)	204-206 (206-218)	0.183 (0.931)	0.269 (0.779)	0.231 (0.731)	0.036 (0.378)	0.116 (0.557)	0.184 (-0.102)	0.320 (-0.199)
INRA128	7 (4)	166-182 (174-180)	0.433 (0.765)	0.538 (0.755)	0.513 (0.697)	0.168 (0.322)	0.347 (0.497)	0.145 (-0.013)	0.196 (-0.012)
BM3205	12 (3)	198-220 (204-208)	0.833 (0.231)	0.859 (0.247)	0.834 (0.228)	0.540 (0.029)	0.703 (0.124)	0.011 (0.068)	0.030 (0.065)
IILSTS26	11 (6)	143-167 (151-165)	0.867 (0.667)	0.868 (0.690)	0.846 (0.636)	0.561 (0.271)	0.721 (0.446)	-0.002 (-0.004)	0.002 (0.035)
SPS115	12 (4)	223-249 (245-257)	0.900 (0.323)	0.832 (0.362)	0.809 (0.331)	0.503 (0.065)	0.674 (0.191)	-0.046 (0.057)	-0.082 (0.11)
TGLA263	6 (6)	114-130 (108-124)	0.600 (0.676)	0.522 (0.683)	0.474 (0.622)	0.142 (0.260)	0.295 (0.429)	-0.102 (-0.020)	-0.151 (0.009)

3.2 Conformance to Hardy Weinberg equilibrium

No significant deviation from Hardy Weinberg equilibrium (HWE) for the buffalo was observed when an analysis was carried out across all loci (F_{IS} = 0.028, P = 0.074). A per locus analysis however revealed that four loci (BM1824, TGLA227, TGLA159, ETH10) may deviate from HWE due to a heterozygote deficit ($p < 0.05$, Table 3.3), although these were not significant after Bonferroni correction. A heterozygote deficit for locus BM4028 was close to significant (p = 0.0512). With regard to the Bonsmara cattle, no significant deviation from HWE across all loci was observed (F_{IS} = 0.037, p = 0.08). Locus CSSM19 did however reveal a significant heterozygote deficit (p = 0.0066; SE = 0.0018), which may be attributed to the presence of a segregating null allele. After Bonferroni correction however deviation from HW equilibrium for this locus was non-significant.

Heterozygote deficiencies at Msat loci may be due to inbreeding or the presence of segregating null alleles. The latter, usually caused by a mutation in the primer-binding site resulting in the non-annealing of primer during PCR, is a common cause of heterozygote deficiencies resulting in Hardy Weinberg disequilibrium (Barker *et al*., 1997). The occurrence of null alleles at a low frequency is also not an uncommon characteristic of Msats (Paetkau and Strobeck, 1995). Of the four loci that exhibit significant heterozygote deficiencies for the buffalo, ETH10 and TGLA227 have high null allele frequencies when compared to the rest of the loci (Table 3.3).

Greater than zero estimates of null allele frequencies however does not necessarily imply the presence of null alleles (Marshall *et al*., 1998), since other factors may also contribute to a heterozygote deficit. Selective factors acting on loci due to genetic subdivision of the population, non-random association between alleles (Marshall *et al*., 1998) or selection against heterozygous individuals in the population may also affect allele frequencies on which genetic parameters are based (Cronin *et al*., 2003). The fact that buffalo live in herds and that male dominance is extremely strong, suggest non-random mating. The latter would contribute to a heterozygote deficiency. Segregating null alleles at a low frequency can however not be excluded, which is supported by the relatively large F_{IS} values that accompany these loci and the fact that no non-random association of alleles (linkage disequilibrium) among the loci could only be demonstrated after Bonferroni correction (p =

0.0003 per locus pair). Eight (5.8 %) out of 136 pairwise comparisons among the respective loci (all of which are located on different bovine chromosomes) had p-values less than 0.05, which is what is expected by chance, excluding the presence of subgroups within the sample population (Ohta, 1982).

Table 3.4. Single locus Hardy Weinberg tests for heterozygote deficiency for buffalo. For less than five alleles, the complete enumeration method (Louis and Dempster, 1987) is used to calculate an exact P-value, while standard error is not computed. For more than five alleles the Markov chain method is used. (Guo and Thompson, 1992)

Locus	P-value	S.E.	Null freq	Fis (W&C)
BM 1824	0.0037	0.0028	0.0533	0.109
CSSM 19	0.0746	0.0122	-0.0006	0.006
Inra 006	0.7403	0.0195	0.015	0.007
TGLA 227	0.0035	/	0.1123	0.21
DIK 20	0.2858	0.0263	-0.0084	-0.005
ETH 225	0.9263	/	-0.0694	-0.141
TGLA 159	0.0009	0.0005	0.0598	0.124
TGLA 57	0.5465	0.0189	-0.0073	-0.005
BM 3517	0.466	0.0233	0.0126	0.019
BM 4028	0.0512	/	0.0891	0.184
BM 719	0.9037	0.0169	-0.0202	-0.038
ETH 10	0.0287	/	0.1849	0.32
INRA 128	0.1615	0.013	0.1458	0.196
BM 3205	0.3345	0.0272	0.0114	0.03
IILSTS 026	0.5693	0.0254	-0.0028	0.002
SPS 115	0.9367	0.0134	-0.0466	-0.082
TGLA 263	0.7991	0.0157	-0.1027	-0.151

3.3 Exclusionary power and probability of identity

Although loci with expected heterozygosities of 0.5 or less are not very informative with regard to paternity verification, the cumulative contribution that they make towards the overall exclusionary power of the entire panel merit their inclusion. Monomorphic loci, if encountered, should however be omitted from paternity analysis since they do not add any information that would increase the power of resolution. More powerful measures of the ability of the loci to resolve parentages are the exclusion probabilities. The average exclusion probability per locus, when one (first parent) or both parents (second parent) were genotyped, ranged from 0.01 to 0.63 and from 0.072 to 0.774, respectively for buffalo (Table 3.3), and from 0.029 to 0.468 (first parent) and 0.124 to 0.644 (second parent) for the Bonsmara cattle. The total combined exclusionary power was 0.99941 (first parent) and 0.99999 (second parent) for buffalo and 0.9977 (first parent) and 0.9999 (second parent) for the Bonsmara cattle, with the 17 Msat panel, whilst for cattle profiled with both the 17 microsatellite panel and the ISAG panel, a combined exclusionary probability of 0.9999 was obtained when just one parent was genotyped. These values indicate an extremely high power of resolution, meaning that the chance of allocating a candidate parent wrongly to an offspring is less than 0.01%.

The combined cumulative probability of identity value (PI, i.e. the resolving power of molecular markers with regard to the ability to distinguish between individuals) across all loci (*P*) was 6.5×10^{-17} and 1.03×10^{-16} for the buffalo and cattle, respectively. This greatly exceeds the total number of buffalo and cattle present in the world today and is also orders of magnitude higher than the value obtained with the ISAG panel (1.17×10^{-13}).

The combined cumulative probability of identity or PI value is one of the most common statistics used to quantify the resolving power of molecular markers with regard to the ability to distinguish between individuals. Single locus and combined cumulative PI values for the buffalo are shown in Table 3.5. The PI values for individual loci ranged between 6.86×10^{-1} for locus BM4028 (the least informative locus) and 1.83×10^{-2} for locus DIK20 (the most informative locus). The combined cumulative PI value across all loci (Prod. unbiased) was 6.473×10^{-17}. For traceability and individual identification purposes this level of resolution would conform to the most stringent requirements, rendering the panel suitable for forensic case studies involving either poaching or illegal movement of animals or their

products. The discriminatory power of the loci was also demonstrated when applied to the Bonsmara cattle, with the cumulative PI being 1.03×10^{-16} which exceeds the power of resolution of currently employed loci prescribed by ISAG (International Society for Animal Genetics).

It should be noted that when HW assumptions are violated, due to for instance, the sampling of close relatives, exclusion probabilities may be overestimated as a result of biased allele frequencies (Luikart *et al*., 1999). Such errors should however be minimal, provided the loci exhibit approximate agreement with HW expectations (Schnabel *et al*., 2000).

Table 3.5. Single locus (unbiased/locus) and combined cumulative (Prod.unbiased) probabilities of identity values for buffalo, as computed by Gimlet. Computations are based on equations as described by Waits *et al*., (2001). PI over entire panel of loci was calculated by multiplying sequentially the PI values over individual loci.

Locus	PI per locus (unbiased per locus)	Cumulative PI (Prod. unbiased)
BM1824	2.25E-02	2.25E-02
CSSM19	7.03E-02	1.58E-03
INRA006	2.09E-01	3.31E-04
TGLA227	3.37E-01	1.11E-04
DIK20	1.83E-02	2.04E-06
ETH225	4.31E-01	8.81E-07
TGLA159	6.20E-02	5.46E-08
TGLA57	8.08E-02	4.41E-09
BM3517	2.89E-01	1.27E-09
BM4028	6.86E-01	8.73E-10
BM719	4.76E-02	4.16E-11
ETH10	6.03E-01	2.51E-11
INRA128	2.29E-01	5.75E-12
BM3205	3.56E-02	2.05E-13
IILSTS026	3.15E-02	6.46E-15
SPS115	4.21E-02	2.72E-16
TGLA263	2.38E-01	6.47E-17

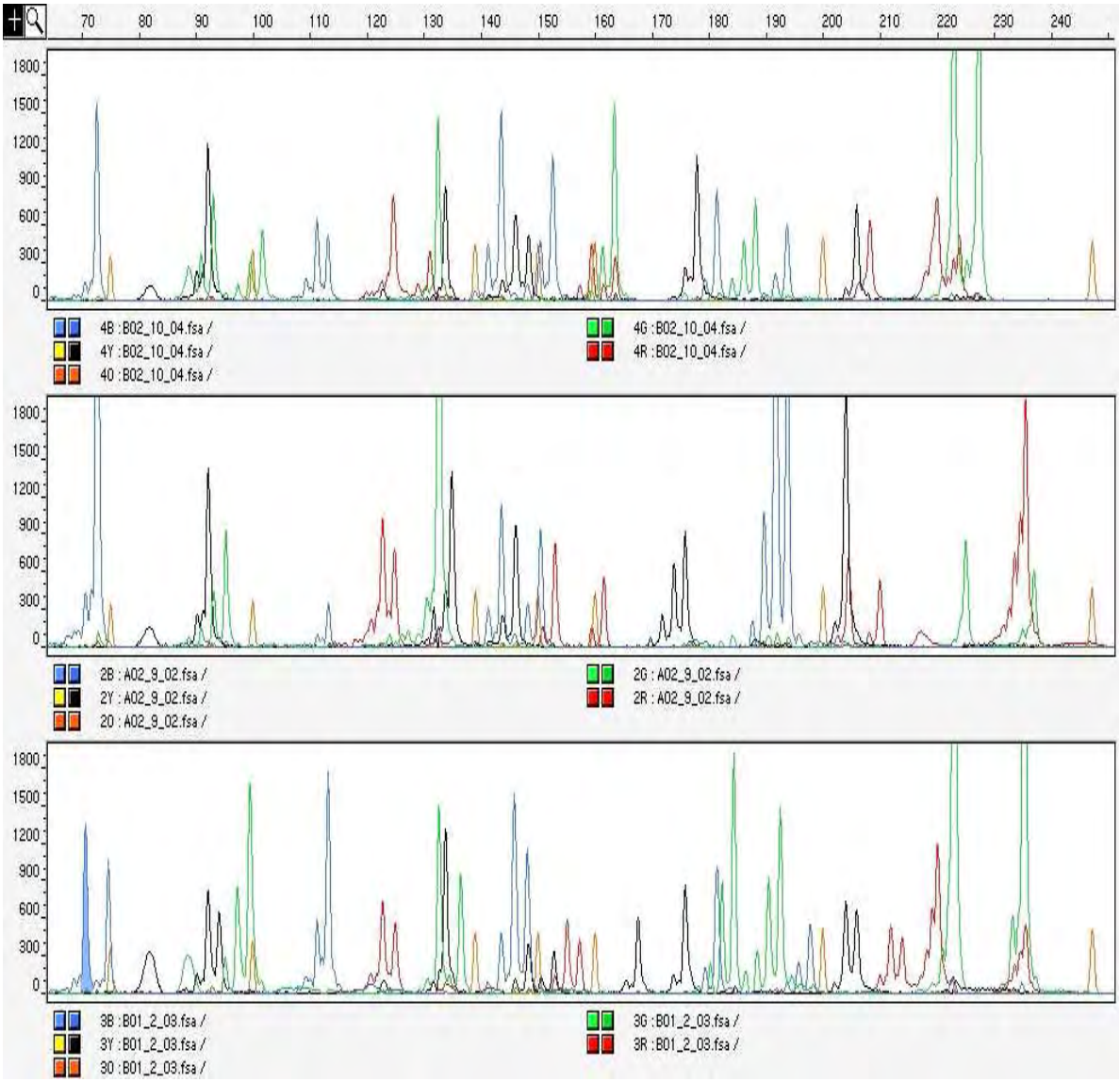
3.4 Multiplexing and co-electrophoresis

Although the primary requirement for the panel of Msats was to uncover variation for subsequent population genetic studies, in view of the cost, time and labour involved in genotyping large numbers of individuals, the secondary aim was to co-amplify and co-electrophorese the panel of Msats. A typical image of the profile obtained following

capillary electrophoresis is shown in Figure 3.1. Multiplexing is technically quite demanding and, when compared to single locus amplification, often may generate undesirable artefacts such as non-specific amplicons, primer-dimers and allelic dropout (Luikart *et al.*, 1999). Depending on the severity of the latter, it may result in a heterozygote deficit that will be interpreted as segregating null alleles, whilst introducing bias in the allele frequency distributions that are the basis on which many calculations are based. Allelic dropout or false alleles (non-specific amplicons) were not encountered with the high quality DNA obtained from blood samples. In order to test for dropped alleles as an artefact of multiplexing, we compared the genotypes of 5 individuals typed with both the multiplex system and in a single locus typing format. The profiles were identical, indicating that the occurrence of dropped alleles or false alleles is likely to be minimal. Using non-invasive samples such as hair or faeces, which in general contain very little DNA as well as inhibitors of PCR, will however require establishing the degree to which dropped alleles would manifest during multiplexing.

The ability of the capillary based DNA sequencer (3100) to detect four different fluorescent labels opened up the possibility to analyse all the amplicons simultaneously. Co-eletrophoresis however requires that allele size ranges of loci labelled with the same colour do not overlap, since this will make accurate assignment of alleles problematic. Co-electrophoresis of the 17 Msat markers revealed that allele size ranges for loci BM4028 and BM719 came within 2 basepairs of each other, while a small percentage (0.8%) of the alleles of locus BM3205 fell within the allelic range of locus SPS 115. This did not prove problematic as the peaks from the different loci had characteristic and very different stutterband profiles, which made it easy to distinguish between alleles from different loci. For the Bonsmara cattle, however, loci BM3517 and BM4028 did show overlapping allelic size ranges, pointing to a need to either change the fluorescent label of one of the loci (e.g. BM4028 could be labelled with the green fluorescent label VIC), or to add a few based to one of the primer pairs in question in order to separate the allelic ranges of the two loci with a larger margin.

Fig. 3.1. Complete Genotyper 3.7 allelic profiles for three individual buffalo. The profiles were generated by co-electrophoresis of the amplicons of the entire panel of 17 microsatellite loci that were amplified in three separate core multiplex PCR's.



4. Conclusions

The aim of this study was to develop a cost effective and less time-consuming profiling system using a panel of 17 Msats suitable for population genetics studies of buffalo in South Africa, following which the panel was evaluated in terms of its suitability for paternity verification and individual animal identification. The multiplex system approach proved to be robust and the occurrence of artefacts were negligible. Using this multiplex approach, together with the co-electrophoresis of the entire panel of 17 Msats, it is possible to genotype 96 animals within 2 days (excluding DNA extractions, which would take up a full day's work). This results in an enormous cost saving when compared to single-locus genotyping. It should be noted that the system was optimized for DNA extracted from blood, which was of both high quality and quantity. In the event of using DNA extracted from faeces or hair samples, artefacts such as non-specific amplification products, null alleles and the inhibition of PCR due to the presence of inhibitors, may be more pronounced, and should first be evaluated.

The moderate to high levels of genetic variation uncovered reflects the suitability of the selected loci for population genetic studies. It should be noted however that selection for polymorphism alone ultimately affects the amount of variation in a sample population. Excluding less polymorphic loci from analyses for instance would result in an increase in measures of diversity such as the expected heterozygosity, allelic richness and polymorphic information content. Simulated data have shown that highly polymorphic loci in many cases provided better estimates of genetic distances than less polymorphic loci (Kalinowski, 2002). Highly polymorphic loci furthermore have superior power of resolution, enabling the assessment of genetic relationships between populations that diverged as recently as 50 to 100 generations ago (Diez-Tascon *et al*., 2000). As discussed in chapter one, highly polymorphic loci are however prone to homoplasy, which when present affects genetic parameters such as genetic distances and phylogenetic relatedness.

Prior to implementing a selection of loci for analyses, it is vital to test the degree to which they conform to assumptions of population genetics theory, since violations of assumptions may affect subsequent calculations. The Hardy Weinberg equilibrium concept is central to population genetics theory, and is a barometer of the extent to which assumptions are

violated and possible causes thereof. For instance, erroneous allele frequencies due to non-conformance to Hardy Weinberg equilibrium (HWE) will result in the overestimation of exclusion probabilities (Luikart *et al*., 1999). This is of particular importance for the emerging buffalo breeding industry where accurate paternity verification and pedigrees are used as management tools in breeding programmes. An exclusion probability in excess of 0.999 (one candidate parent genotyped) means that paternity can be verified with an extremely high level of confidence which is a very useful tool for establishing paternity in cases where one of the candidate parents is unknown, as is often the case for wild animals under extensive conditions that are similar to a multi-sire mating setup.

No consistent deviation from HW equilibrium across all 17 loci was found for both the buffalo and the Bonsmara, suggesting that parameters calculated from allele frequencies are unbiased. The presence of segregating null alleles at low frequencies for four loci could not however be excluded. The fact that several loci are located on the same chromosomes (Table 3.2) may result in the potential linkage between alleles of these loci. The panel however did not exhibit any linkage disequilibrium (LD), excluding selection for certain multilocus genotypes. As far as forensic case studies are concerned, calculation of match probabilities is based on loci that also conform to HWE. The strength of DNA evidence may be overstated when match probabilities are based on biased allele frequencies due to non-conformance. The PI value obtained with the panel of loci is such that an individual can be identified beyond reasonable doubt, which is invaluable for traceability purposes.

It is concluded that the panel of 17 Msats, amplified in a multiplex system and co-electrophoresed as a single injection on an ABI 3100 DNA sequencer, meet with the intended application requirements.

Appendix A. Allele frequency distributions of the panel of 17 bovine Msats obtained from a sample size of 60 African buffalo sampled at random from throughout the Kruger National Park.

Locus	**Allele/Frequency**														
BM 1824	*Allele*	169	175	177	179	181	183	185	187	189	191	193	195	197	199
	Freq	3.3	3.3	1.7	2.5	27.5	6.7	8.3	5.8	6.7	4.2	8.3	6.7	13.3	1.7
Inra 006	*Allele*	107	109	111	113	115	117	119							
	Freq.	0.8	15.0	5.8	56.7	19.2	0.8	1.7							
CSSM 19	*Allele*	128	136	138	140	142	144	146	148	150	152	154			
	Freq.	1.7	2.5	2.5	2.5	1.7	35.8	14.2	24.2	6.7	6.7	1.7			
TGLA227	*Allele*	70	72	74	76										
	Freq.	4.2	72.5	17.5	5.8										
DIK 20	*Allele*	164	170	174	176	178	180	182	184	186	188	190	192	196	198	208
	Freq.	13.6	6.8	0.8	9.3	2.5	1.7	2.5	8.5	9.3	12.7	1.7	4.2	2.5	21.2	2.5
ETH 225	*Allele*	133	137												
	Freq.	71.7	28.3												
TGLA159	*Allele*	223	225	227	229	231	233	235	237						
	Freq.	24.2	7.5	31.7	3.3	16.7	2.5	11.7	2.5						
TGLA 57	*Allele*	75	89	93	95	97	99	101							
	Freq.	0.8	5.8	24.2	25.0	30.0	10.8	3.3							

BM 3517	*Allele*	84	86	88	90	92	94	96								
	Freq.	1.7	11.7	3.3	2.5	65.0	15.0	0.8								
BM 719	*Allele*	136	138	140	142	144	146	148	152	154	156	158	160			
	Freq.	8.3	0.8	3.3	4.2	0.8	11.7	27.5	20.0	17.5	0.8	4.2	0.8			
ETH 10	*Allele*	204	206													
	Freq.	84.2	15.8													
TGLA263	*Allele*	114	120	122	124	126	130									
	Freq.	1.7	0.8	65.8	19.2	10.8	1.7									
BM 4028	*Allele*	126	132	134												
	Freq.	3.3	4.2	92.5												
Inra 128	*Allele*	166	168	170	174	176	178	182								
	Freq.	5.8	7.5	7.5	7.5	66.7	4.2	0.8								
BM 3205	*Allele*	198	200	202	204	206	208	210	212	214	218	220	222			
	Freq.	1.7	2.5	19.2	16.7	16.7	6.7	2.5	7.5	20.8	2.5	2.5	0.8			
IILSTS26	*Allele*	143	147	149	151	153	155	157	159	161	163	167				
	Freq.	6.7	1.7	17.5	15.8	19.2	2.5	8.3	9.2	1.7	15.8	1.7				
SPS 115	*Allele*	219	223	227	229	231	233	235	237	239	241	245	249			
	Freq.	3.3	5.0	1.7	10.8	7.5	5.8	11.7	13.3	34.2	1.7	0.8	4.2			

Appendix B. Allele frequency distributions of the panel of 17 bovine Msats obtained from a sample size of 34 unrelated Bonsmara cattle.

Locus	**Allele/Frequency**															
BM 1824	*Allele*	177	179	181	185	187	191									
	Freq	9.3	34.3	10.9	3.1	15.6	26.5									
Inra 006	*Allele*	101	103	105	107	109	115									
	Freq.	20.6	7.4	1.5	48.5	19.1	2.9									
CSSM 19	*Allele*	134	138	140	148	150	152	154								
	Freq.	1.5	14.7	1.5	4.4	63.2	10.3	4.4								
TGLA227	*Allele*	76	78	80	82	86	88	90	92	94	96					
	Freq.	36.8	2.9	8.8	2.9	7.4	14.7	7.4	7.4	1.5	10.3					
DIK 20	*Allele*	172	174	176	180	182	184									
	Freq.	31.8	12.1	37.9	6.1	10.6	1.5									
ETH 225	*Allele*	135	141	143	145	147	149	155								
	Freq.	25.8	6.1	10.6	28.8	9.1	6.1	13.6								
TGLA159	*Allele*	209	211	227	231	243										
	Freq.	25.0	53.3	11.7	3.3	6.7										
TGLA 57	*Allele*	83	87	93	95	97										
	Freq.	8.8	32.4	50.0	7.4	1.5										

BM 3517	*Allele*	98	100	102	104	106	108	110	112	116	118					
	Freq.	48.5	10.6	1.5	3.0	1.5	7.6	16.7	3.0	1.5	6.1					
BM 719	*Allele*	140	142	144	146	148	150	156	158							
	Freq.	1.5	4.4	11.8	5.9	14.7	14.7	1.5	45.6							
ETH 10	*Allele*	206	208	210	212	214	216	218								
	Freq.	5.2	27.6	1.7	5.2	34.5	12.1	13.8								
TGLA263	*Allele*	108	114	118	120	122	124									
	Freq.	47.1	2.9	1.5	29.4	8.8	10.3									
BM 4028	*Allele*	102	106	108	110	114	116	118	120	122	124					
	Freq.	4.4	1.5	30.9	7.4	23.5	2.9	1.5	8.8	2.9	16.2					
Inra 128	*Allele*	174	176	178	180											
	Freq.	19.1	25.0	26.5	29.4											
BM 3205	*Allele*	204	206	208												
	Freq.	86.5	7.7	5.8												
IILSTS26	*Allele*	151	157	159	161	163	165									
	Freq.	3.0	3.0	48.5	16.7	22.7	6.1									
SPS 115	*Allele*	245	251	253	257											
	Freq.	79.0	11.3	1.6	8.1											

Chapter 4

Inter- and Intra-population structure, genetic variation and dispersal of HiP and KNP buffalo

Abstract

Population genetic studies revolve mainly around genetic variation and the factors that impact upon its maintenance. These include dispersal, habitat fragmentation, population genetic structure, demographic isolation and the degree to which genetic bottlenecks affect its distribution. The Kruger National Park (KNP) and Hluhluwe-Imfolozi Park (HiP) in South Africa makes for a special case study in that both parks not only contain the two largest buffalo populations in the country, but they also have been fragmented for more than a century, precluding any gene flow. The two parks furthermore have quite diverse histories, and ecological studies also suggest that their buffalo populations exhibit behavioural differences. This chapter reports on research, executed using both microsatellite markers (Msats) and mtDNA sequences of the D-loop region, aimed at qualifying and quantifying genetic indices that relate genetic variation to factors that influence its maintenance.

Using a panel of 17 Msats, a total of 485 and 401 individuals were profiled from KNP and HiP respectively. KNP exhibited significantly higher levels of genetic variation than HiP, as reflected in Nei's unbiased gene diversity (0.54 *vs.* 0.64). The park also has twice the total number of alleles than that found in HiP (78 *vs.* 158), and only 1.8 % of the total number of alleles found in both parks was exclusive to HiP. KNP on the other hand exhibited 89 (53.2%) private alleles. The effective population size (N_e) for KNP was also notably larger than that of HiP, based on both the SMM and IAM models of mutation (4292 *vs.* 2354 and 1469 *vs*. 2238, respectively). The differential level of variation between the parks was also reflected in their respective haplotype diversities (0.92, ± 0.009 and 0.48, ± 0.050).

The results suggest that while KNP has retained much of its variation since the rinderpest bottleneck, HiP has experienced a significant reduction in genetic variation. This reduction may be attributed to the bottleneck, although other factors such as drift, small population size, removal of animals from the park(s) and pre-bottleneck levels of variation may also have played a role. HiP also exhibited a steady decline in genetic variation between 1986 to 2004, which suggests episodes of low N_e. A positive inbreeding coefficient (F_{IS}) was found for KNP (0.034), which may be due to segregating null alleles, population substructure and non-random mating of males.

With regard to population structure, KNP and HiP are highly differentiated. Based on Msat data, the two populations are separated by a large genetic distance (F_{ST} value = 0.159; $p < 0.001$), which was also in accordance with mtDNA sequence data (F_{ST} = 0.275; $p < 0.001$). The high level of differentiation may be attributed primarily to genetic drift as a result of the bottleneck. Differentiation among subpopulations and herds within the two parks was however small, based on Msat data: F_{ST} = 0.012 (95%CI: 0.008-0.016) and 0.014 (95% CI: 0.009-0.019) respectively for KNP and HiP (among herds). MtDNA data revealed similar results, with F_{ST} values among herds in KNP and HiP being 0.067 ($p = 0.006$) and 0.005 ($p = 0.42$), respectively. Interestingly however was the fact that pair-wise analyses at the herd level revealed high and significant ($p \leq 0.05$) levels of differentiation among certain herds, suggesting that female gene flow between these herds is limited. On the subpopulation level, genetic distances between the Pafuri and Malelane/Crocodile bridge subpopulations in KNP, separated by more than 300 kilometres, and based on mtDNA haplotype frequencies, was low (F_{ST} = 0.026, $p = 0.017$). For the Manzimbomvu and Mbhuzane areas in HiP, no differentiation could be demonstrated based on mtDNA data (F_{ST} = 0.016; $p = 0.26$). Sequence data clearly indicated the absence of geographical partitioning of haplotypes to their respective populations of origin, and the haplotypes from the two parks are also separated by a small genetic distance (0.051: SE: 0.007). This may be attributed to genetic contact that took place between the two populations in the more distant past. Currently the buffalo population of KNP seems to be in equilibrium, while that of HiP exhibits signs of contraction. Finally, the relevance of the results presented in this chapter is discussed in the context of genetic and conservation management of the buffalo populations of KNP and HiP.

1. Introduction

The Kruger National Park (KNP) and Hluhluwe-Imfolozi Park (HiP) contain the two largest buffalo populations in S.A., giving them benchmark status when compared to other smaller populations. In order to devise strategies for the management and protection of the buffalo of these two parks, it is vital to gain knowledge regarding the genetic status of these populations. Understanding the role of buffalo behaviour within the parks on the one hand and gaining knowledge about the effects of demographic and environmental processes on the other will enable us not only to explain the current state of affairs regarding their genetic status, but also to predict their potential responses to future changes. One of the key components of any genetic assessment is the level of variation within and between populations. It is also an important measure of the fitness of a population, and is also used as a basis from which interventions such as translocations are planned in order to restore variation in genetically impoverished populations. The effective population size (N_e) is another important indicator of the ability of a population to sustain variation, while geneflow is probably one of the most important processes that influence the dynamics within a population over time (Turchin, 1998, as cited by Berry *et al.*, 2004), affecting structure and differentiation. It is also an important process from an epidemiological point of view, as it influences the rate of spread of disease.

Detailed large scale and especially fine scale (among herd and sub-populations) studies within and among the HiP and KNP populations have not been carried out before. These two parks are geographically quite diverse, differing both in size as well as the number of buffalo they harbour. The buffalo of these parks have been fragmented for more than a century, excluding gene flow between them. They furthermore exhibit quite diverse histories with regard to the anthropological interventions exerted on them and ecological studies suggest marked behavioural differences in their respective buffalo populations (Jolles 2004). In HiP for instance it has been shown that some herds form tight coherent entities that do not mix with neighbouring herds. It should be noted from the onset that the herd-concept is much more flexible than what was previously believed (Cross, 2005; personal communication), and it does not necessarily occupy a definable clear-cut geographical boundary or home range. It is of interest to draw comparisons between the two populations and to assess the degree to which factors may have impacted differentially on the two populations. New

insights into the behaviour of buffalo can also be gained by drawing parallels between ecological observations and genetic data.

A range of molecular markers (microsatellites or Msats and mitochondrial D-loop sequences) and statistical software have been developed and are available for qualifying and quantifying genetic indices which will contribute to our understanding of the effects of many of the above-mentioned processes. These have been used here to assess differences in genetic variation, structure and dispersal at two different levels, namely those at (i) the inter populational level and (ii) at the intra-populational level.

2. Materials and Methods

2.1 Msat analyses

2.1.1 Inter-population level analyses

A total of 485 and 401 individuals were profiled from KNP and HiP respectively, according to the methodology outlined in chapter three. General indices of genetic variation (H_o, unbiased heterozygosity or H_e, F_{IS} and total number of alleles were calculated using FSTAT (Goudet *et al.*, 1995). Allelic richness (*Rs*), which is an estimation of the alleles per locus independent of sample size, was also calculated using FSTAT. Arlequin 2.000 (Schneider *et al.,* 2000) and Agarst (Harley *et al.*, 2002) were used to estimate the genetic distance (in terms of F_{ST} and R_{ST}) between KNP and HiP. The allele frequency distribution within both parks and across all Msat loci was calculated with MS Toolkit. Agarst was used to calculate long term effective population size, assuming both the SMM and IAM models of Msat mutation, based on expected heterozygosities at each locus and assuming that the populations are in mutation-drift equilibrium. An upper bound mutation rate (μ) of 2.05 x 10^{-4} was used to calculate the long-term effective population sizes according to the following equations (Harley *et al.*, 2005; Rooney *et al.*, 1999; Lehmann *et al.*, 1998):

1. $N_e = \{1/(1-H_e)^2-1\}/8\mu$ (SMM)
2. $N_e = H_e/\{4\mu\ (1-H_e)\}$, (IAM)

where μ is the mutation rate, and H_e the unbiased expected heterozygosity.

The program Geneclass (Cornuet *et al*., 1999) was used as an indirect approach to establish the level of differentiation between KNP and HiP. Individuals were assigned to the population in which the likelihood of their genotype was highest using the Bayesian method (Rannala and Mountain, 1997). In order to avoid biases when estimating population allelic frequencies, the individual being assigned was omitted from calculations. Agarst was used in a similar way, but in this case the distribution of likelihood values (Log-values) of assignments of individuals to any sampled population was calculated. Populations that exhibit high levels of genetic variation will have lower likelihood values of assignment than populations that exhibit reduced levels of genetic variation (Harley *et al*., 2005). Exact tests of conformance to Hardy Weinberg equilibrium (HWE) were executed with Genepop 3.3 (Raymond and Rousset, 1995), while FSTAT was used to test for linkage disequilibrium among all pairs of loci in order to establish the independence among alleles. Sequential Bonferroni corrections were subsequently applied in order to compensate for the increased chance of a Type I error when conducting multiple significance tests (Rice, 1989).

In order to test for the remnants of the rinderpest bottleneck, the programs Bottleneck (Cornuet and Luikart, 1996) and Agarst were used. Population theory predicts that in a population that is at mutation-drift equilibrium (i.e., a population of which the effective size has remained constant in the recent past), there is approximately an equal probability for a locus to exhibit either a gene diversity excess or deficit (Cornuet and Luikart 1996). When populations experience a decline in their effective population size, they exhibit a correlative reduction in their number of alleles (k) and gene diversity (H_e, or Hardy-Weinberg heterozygosity). Allele numbers are however reduced faster than gene diversity, and in a recently bottlenecked population, the observed gene diversity is thus higher than the expected equilibrium gene diversity (*Heq*, Cornuet and Luikart, 1996). The program Bottleneck calculates the probability of the expected gene diversity (*He*) being higher than the expected equilibrium gene diversity (*Heq*). Calculations are based on the observed number of alleles for each locus in each population and assuming mutation-drift equilibrium. Although three tests can be used to identify loci that exhibit a gene diversity excess, the Wilcoxon sign-rank test is the most powerful (Luikart and Cornuet, 1998). For each population and each locus, the distribution, obtained through simulating the coalescent

process of *n* genes (under SMM, IAM and TPM models of mutation), of the gene diversity expected from the observed number of alleles (k), given the sample size (n), was calculated. Finally the allele frequency distribution was established in order to determine whether it was approximately L-shaped (as expected under mutation-drift equilibrium) or not. A recent bottleneck provokes a mode shift in the distribution. Agarst follows a slightly different approach. It calculates the M-statistic of Garza and Williamson (2001), which tests for gaps in the allele size distributions by measuring the mean ratio of the number of alleles to the total range in allele size (Harley, 2002). A value larger than 0.8 is typical of an outbred population, while populations that have recently experienced severe bottlenecks, have M values below 0.7. The benefit of this approach is that the remnants of a bottleneck may be detectable beyond many generations.

2.1.2 Intra-population level analyses

Estimates of genetic variation and distances among herds and subpopulations, in terms of F-statistics, were carried out with MS Toolkit, FSTAT and Arlequin. Gene diversities were calculated per herd and among both sexes and age classes within herds. Data from herds that shared a common geographical location were pooled to constitute a sub-population. This was done in order to assess the level of genetic variation among subpopulations and the degree to which they are differentiated. For KNP 83 and 182 samples were pooled from herds located in the Pafuri/Klopperfontein (northern sub-population) and Crocodile Bridge/Malelane (southern sub-population) areas, respectively. For mtDNA analyses, 36 and 48 animals were pooled to represent the northern and southern sub-populations respectively. Sub-populations for HiP, intended for Msat analyses, consisted of 79 animals from neighbouring herds in the Mbhuzane (south) and Manzimbomvu areas (north) respectively. For mtDNA analyses 18 and 21 animals respectively were pooled from the Manzimbomvu and Mbhuzane areas.

Several tests for sex-biased dispersal were executed primarily using FSTAT. Population theory predicts that the dispersing sex should display less structure than the philopatric sex, and hence the dispersing sex should exhibit a larger heterozygote deficit (Goudet *et al.*, 2002) and concomitant elevated inbreeding coefficient or $F_{IS.}$ (the probability that two genes

drawn at random from a population share a common ancestor). Sex-biased dispersal can also be detected by comparing the assignment index (Ai), F_{ST} and mean level of relatedness between members of the two sexes. The assignment index is an indication of the frequency with which an individual's genotype occurs in a given population, and a lower index indicates that the individual is less likely to have come from that population and could be thus be an immigrant (Dallimer *et al.*, 2002). A positive Ai value for a particular sex indicates that the sex is resident to its population, while a negative value is associated with potential dispersers. F_{ST} values should also be higher for the philopatric sex, while gene diversity (expected heterozygosity or H_e) should be lower for the sex dispersing (Goudet *et al.*, 2002). Immigrants in a population should display lower levels of relatedness than resident members. The relatedness statistic is related to F_{ST} through the following equation (Hamilton 1971; Queller and Goodnight 1989):

$$r = 2F_{ST}/(1 + F_{IT})$$

SPAGeDi (Hardy and Vekemans, 2002), a software program primarily designed to characterise the spatial genetic structure of mapped individuals and/or mapped populations using genotype data, was used to calculate r, the relationship coefficient of Lynch and Ritland (1999). The relationship coefficient was calculated for all pairwise comparisons between individuals within populations (entire KNP or HiP) as well as among herds within the two parks. Males and females were also analysed separately in order to determine dispersal patterns between the two sexes.

In order to test for a correlation between genetic distance and geographical distance (isolation by distance), a Mantel g-test was performed on associated genetic distance (F_{ST}) and geographic distance (kilometres) matrices among all herds from KNP and for 9 herds from HiP for which exact geographical coordinates were available, using POPTOOLS 2.6.2 (Greg Hood, CSIRO, Canberra, Australia). The test was aimed at detecting the degree to which dispersal takes place among all herds, over both small (less than approximately 20 km) and large (more than 300 km) distances.

With regard to age-biased dispersal, older male animals are presumed to make up a large percentage of the dispersers/immigrants within a population, and these animals should have a lower likelihood of being assigned to the herd or population from which they were sampled compared to resident non-dispersing individuals. In order to test for a correlation between age and dispersal, the likelihood of assignment of individuals of different age classes to the herds in which they were sampled was calculated using the program Spassign (Palsson, 2004). The program calculates the assignment probabilities of an individual's genotype to its population of origin and compares it to the probability of being assigned to another population. Calculations are based on either the method developed by Paetkau *et al.,* (1995) or the Bayesian method from Rannala and Mountain (1997). Both methods assume HWE and linkage equilibrium across the loci used. The program calculates the proportion of individuals that are more likely to belong to a population other than the one from which they were sampled, and it gives the weighted mean, *P* ass, of the two. If an allele does not exist in one population it is assumed that its frequency is $1/(n+1)$, where n is the sample size (Palsson, 2004).

2.2 MtDNA sequence analyses

The mitochondrial haplotypes identified and described for KNP and HiP in chapter two were used to calculate indices of genetic variation and to infer phylogenetic relationships. In total, the dataset consisted of 161 individuals from KNP and 97 from HiP. These individuals represented 30 and 14 herds respectively from KNP and HiP, and in both cases the herds were widely distributed throughout the parks. General indices of diversity such as haplotype diversities (the probability that two haplotypes chosen at random from a population are different), nucleotide diversities and the mean number of pair-wise differences (Nei and Li, 1979) between populations were calculated with Arlequin 2.000 and DNASP (Rozas *et al.*, 2003). The amount of population structure among and within each population was assessed by differential hierarchical analyses of AMOVA and by calculating transformed indexes of genetic distances, F_{ST} and Slatkin's distance. These distance estimates can be used to indicate short–term genetic distances between populations (Reynolds *et al*., 1983; Slatkin, 1995). Indices of genetic structure were based both on the allelic content and frequency of the respective haplotypes (Excoffier *et al*., 1992). Mantel tests were performed on pairwise

genetic distance and geographical distance matrices using Poptools in order to assess female-biased isolation by distance.

Phylogenetic and molecular evolutionary analyses were conducted using MEGA version 3.0 (Kumar *et al.*, 2004), ModelTest (Posada and Crandall, 1998), PAUP* 4.08 (Swofford, 1998) and Arlequin 2.000 (Schneider *et al*., 2000). The sequence data was subjected to hierarchical likelihood ratio testing using Modeltest in order to determine the model of sequence evolution that fits the data the best. Parameters from this test were subsequently used during phylogenetic tree inference using either Neighbour-Joining (NJ) or Maximum Parsimony (MP) algorithms. With MP, gaps (indels) were treated as a fifth character state and homoplasious characters were down-weighted on the basis of the rescaled consistency indices (RC), following an initial unweighted MP analyses. All MP analyses made use of the tree-bisection-reconnection (TBR) branch swapping algorithm, and nodal support was assessed by 10000 bootstrap replications.

A midpoint-rooted phylogenetic tree was inferred using the neighbour-joining algorithm of Satou and Nei (1987), with alignment gaps being completely omitted from analyses. The evolutionary distance between pairs of sequences was based on the Tamura-3 parameter model which accounts for multiple hits, differences in transitional and transversional rates and a G + C-content bias (Tamura, 1992). As midpoint rooting assumes that there is no rate heterogeneity across taxa, this was assessed in PAUP, prior to midpoint rooting. The reliability of tree-topology was assessed following 10000 bootstrap replications. According to Simon *et al*., (1994), most distance correction methods will yield approximately the same estimate if the taxa are closely related, since homoplasy will in all probability be small and the substitution rate among them will not vary. In such a case, it is proposed that the method with the least complexity and variance should be used in order to increase the likelihood of inferring the correct tree.

3. Results and Discussion

3.1 General indices of diversity: Inter-populational analyses

The general indices of diversity amongst HiP and KNP buffalo, based on Msat data, are shown in Table 4.1. Values are based on Msat data from 401 and 485 individuals respectively. The allele frequency distributions within KNP and HiP are shown in Figure 4.1 (at the end of the chapter). From Table 4.1 it is evident that KNP exhibits notably higher levels of genetic variation than HiP. One of the primary indicators of variation, expected heterozygosity (Nei's unbiased gene diversity) is significantly different between the two populations (0.54 *vs.* 0.64, $p < 0.01$, Wilcoxon matched pair test). This differential level of Msat diversity is also reflected in the respective mtDNA gene diversity indices, with KNP having a haplotype diversity of 0.92 (± 0.009) and HiP a significantly lower haplotype diversity (0.48, ± 0.050). It was also reflected in the nucleotide diversities, which were 0.049 and 0.03 for KNP and HiP, respectively (see Chapter 2). The latter is comparable to that of buffalo populations from eastern and southern Africa (Simonsen *et al.*, 1998), and in the same order as that found in other species such as Warthog ($\pi = 0.04$, Muwanika *et al.*, 2003).

Contrary to expectation however, the inbreeding coefficient (F_{IS}) for KNP is 0.034, which is significantly higher than the -0.001 for HiP ($p = 0.004$; Wilcoxon matched-pair test). KNP also has twice the total number of alleles than that present in the HiP population (78 *vs.* 158), although the number of animals sampled in KNP was larger than that sampled in HiP. KNP also has 34 mtDNA haplotypes compared to only 4 identified in HiP, while the allelic richness (Rs) for KNP, an indicator of the level of genetic variation within a population, was approximately two-fold higher than that of HiP ($p < 0.001$, Wilcoxon Matched Pair Test).

Table 4.1. Inter population level diversity indices based on Msat data. Calculations were based on pooled data from 401 and 485 individuals from HiP and KNP respectively.

Pop.	Census size	N	$Loci_n$	N_a	H_o	H_e	F_{IS}	Rs	M	N_e (SMM)	N_e (IAM)
HiP	3000	401	17	78	0.54 (0.006)	0.54 (0.048)	-0.001	4.57	0.74 (0.039)	2354	1469
KNP	28500	485	17	158	0.62 (0.005)	0.64 (0.054)	0.034	9.47	0.86 (0.015)	4292	2238

Ne was calculated for both SMM and IAM assuming mutation-drift equilibrium and a mutation rate of 2.05 x 10^{-4} (Rooney *et al.*, 1999; Lehmann *et al.*, 1998, Harley *et al.*, 2005); Pop: population; N: number of animals; $Loci_n$: number of loci analysed; N_a: total number of alleles; H_o: observed heterozygosity and standard deviation; H_e: expected heterozygosity (Nei's unbiased gene diversity) and standard deviation; F_{IS}: inbreeding coefficient; Rs: allelic richness calculated independent of sample size with FSTAT M: Garza and Williamson's M-statistic and variance; N_e : effective population size (calculated based on stepwise mutation model or SMM and infinite alleles model or IAM).

The moderate to high levels of H_e of both populations indicates large and long-term stable effective population size (Spong *et al.*, 2000). H_e is also comparable to that of buffalo populations from elsewhere in Africa (Simonsen *et al.*, 1998; Van Hooft *et al.*, 2003), while it is high compared to the St Lucia and Addo populations (O'Ryan *et al.*, 1998). It is however important to note that H_e will vary depending on the level of polymorphism of the Msat loci used in the analyses.

Unbiased estimates of Hardy-Weinberg exact P-values by the Markov chain method (dememorization: 1000; batches: 20; iterations per batch: 1000) revealed that KNP deviated significantly ($p < 0.01$) from HWE at 7 loci due to a heterozygote deficit. Five loci (BM1824, TGLA227, TGLA159, BM4028, Inra128) still exhibited a heterozygote deficit after Bonferroni correction (Table 4.2). HiP on the other hand showed no overall significant deviation from HWE, although one locus (TGLA227) exhibited a heterozygote deficit after Bonferroni correction (Table 4.2). Loci TGLA 227 and INRA 128 exhibit relatively large F_{IS} values in both populations, suggesting the presence of segregating null alleles. The excess of homozygotes exhibited at these particular loci may indicate one of the following: (i) that they are under selection, (ii) that segregating null alleles are present, (iii) inbreeding has occurred among members in the population or (iv) that population substructure is marked. One possible explanation may be that the mating behaviour of the bulls, in particular, may result in a type of inbreeding effect, that leads to higher than expected F_{IS} values as a result of a heterozygote deficit (Chesser, 1991). Small but significant substructure within the KNP population may further contribute to non-conformance to HWE (see section 3.5) dealing with population differentiation.

Table 4.2. Unbiased estimates of Hardy-Weinberg exact P-values by the Markov chain method (dememorization: 1000; batches: 20; iterations per batch: 1000) across all loci for KNP and HiP.

	KNP			HiP		
Locus	**P-value**	**S.E.**	F_{IS}	**P-value**	**S.E.**	F_{IS}
BM 1824	0.00*	0.00	0.08	0.64	0.03	-0.01
CSSM 19	0.13	0.02	-0.01	0.83	0.02	-0.03
INRA 006	0.65	0.03	0.03	0.76	0.02	-0.04
TGLA 227	0.00*	0.00	0.17	0.00*	.0.00	0.10
DIK 20	0.06	0.01	0.01	0.28	0.02	0.01
ETH 225	0.25	0.03	0.03	0.51	0.00	0.01
TGLA 159	0.00*	0.00	0.08	0.31	0.02	0.02
TGLA 57	0.08	0.01	0.02	0.49	0.02	0.02
BM 3517	0.28	0.02	0.00	1.00	0.00	-0.11
BM 4028	0.00*	0.00	0.14	0.66	0.01	0.00
BM 719	0.03*	0.01	0.00	0.43	0.03	0.05
ETH 10	0.02*	0.00	0.11	0.88	0.01	-0.06
INRA 128	0.03*	0.00	0.11	0.21	0.00	0.11
BM 3205	0.44	0.04	0.02	0.11	0.01	0.02
ILLSTS 026	0.55	0.03	0.00	0.57	0.02	0.00
SPS 115	0.29	0.03	-0.02	0.69	0.02	-0.01
TGLA 263	0.18	0.01	-0.02	0.36	0.02	-0.01

* Indicate values at the $p \leq 0.05$ significance level

The allele frequency distributions across all loci for both KNP and HiP are shown in Fig. 4.1, which can be found at the end of this chapter. It was striking that only 3 (1.8 %) alleles out of a possible total of 167 alleles occurring in the two parks were private/exclusive to HiP, while 89 (53.2 %) private alleles were recovered from KNP. The fact that KNP has a large percentage of private alleles not shared with HiP suggests that KNP has retained much of its variation since the rinderpest bottleneck. HiP on the other hand has experienced a significant reduction in allelic diversity, which may be due not only to the bottleneck, but also to other factors such as drift and small population size. The reduced level of genetic variation within HiP may also be due to the rate of recovery of the population. It has been shown that populations will experience a greater reduction in genetic variation if they are unable to recover rapidly following a bottleneck (Nei *et al*., 1975). Such populations may also take longer to restore heterozygosity in the absence of a rapid recovery rate, and the introduction of new alleles into these populations will ultimately require thousands of generations as a result of the mutational process of Msats (Norton and Ashley, 2004).

The differential allele frequency distributions are also reflected in the mtDNA data. None of the KNP mtDNA haplotypes occurred in HiP, and the recovery of just 4 haplotypes from HiP suggests a dramatic reduction in diversity following the rinderpest pandemic. The small population size in HiP following the rinderpest is reflected by the low haplotype diversity observed today. Similar results were found for west- and east-African kob populations (Birungi and Arctander, 2000), where the occurrence of private alleles in particular populations was attributed to drift in past populations rather than restriction of gene flow. O'Ryan (1998) also attributed the lower levels of heterozygosity and reduced allelic diversity in the Addo and St. Lucia populations to genetic drift. The drift experienced by HiP may furthermore have been exacerbated by the lack of immigration into the park over the last six decades.

The reduced level of diversity of HiP is also reflected in a lower N_e (Table 4.1) The effective size of a population is not only an important measure of genetic variation, it is also particularly important for predicting its ability to maintain genetic diversity, while it measures the effect of genetic drift on a population (Lehmann *et al.*, 1998). The method used to estimate N_e is based on the mutational model of Msats and thus estimates the long-term effective size in the past on a time-scale of the order of N_e generations. It furthermore

provides a minimum threshold for a population in mutation-drift equilibrium for a species/population, and once this number has been attained, further loss of genetic diversity should cease (Harley *et al.*, 2005). Random genetic processes also occur at a rate inversely related to population size, which makes its estimation vital for establishing the extent to which processes impact on a population. Assuming IAM and SMM, the minimum and maximum long term N_e for HiP is 48 % and 76 % respectively of its census size, while for KNP it constitutes 7.5 % and 14 % respectively of its census size. It should be noted that N_e was calculated based on an intermediate mutation rate ($2.05x10^{-4}$) for mammalian taxa (Waldick *et al.*, 2002; Harley *et al.*, 2005), and an overestimation of the mean mutation rate across the entire panel of loci will result in an underestimation of long term N_e. In order to obtain a realistic view of the size range of N_e and to relate it to census size, N_e should also be calculated based on a minimum and maximum mutation rate. It should be noted that populations that have experienced recent bottlenecks may not be in mutation-drift equilibrium, which may also result in an underestimation of the long-term effective population size. While KNP seems to be in equilibrium, HiP on the other hand may violate the mutation-drift equilibrium assumption. The latter, together with factors such as the sex ratio, mating system, selection, pattern of inheritance, changes in the population size over generations (e.g. during bottlenecks) and population subdivision may affect the prediction of effective population size (Caballero 1994) and makes it notoriously difficult to estimate (Wang, 2005)..

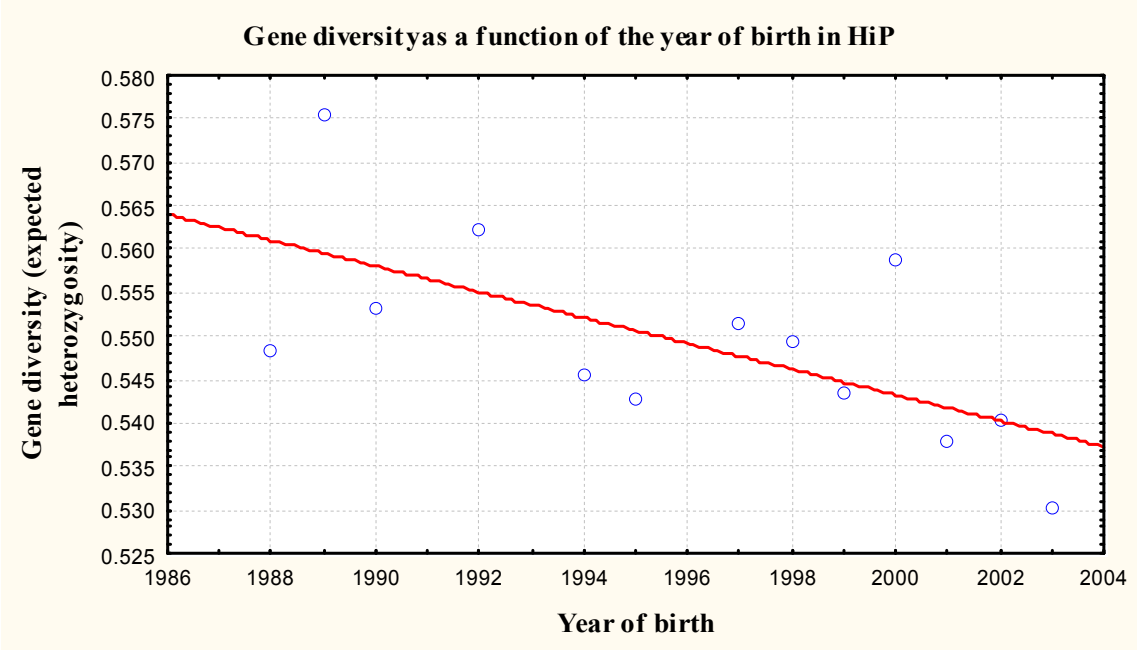
It has also been suggested that N_e, when calculated based on SMM, is a more realistic reflection of the true effective population size (Lehmann *et al.*, 1998). A low long term N_e may reflect a period of low N_e over a few generations or a steady situation (Lehmann *et al.*, 1998). Considering the known low census sizes of HiP in especially the 1930's, the former may be a plausible explanation for HiP's lower N_e. Although controversial, it has been postulated that an effective population size of between 500 and 1000 is required in order to maintain long-term adaptability whilst preserving populations from short-term genetic risks (Lynch and Lande, 1998).

A steady decline in the level of variation during the period between 1986 to 2004 was found in HiP, suggesting episodes of low N_e (Fig. 4.2a) characterized by a loss of alleles and declining heterozygosities (Lehmann *et al.*, 1998). Sampling artefacts may contribute to this

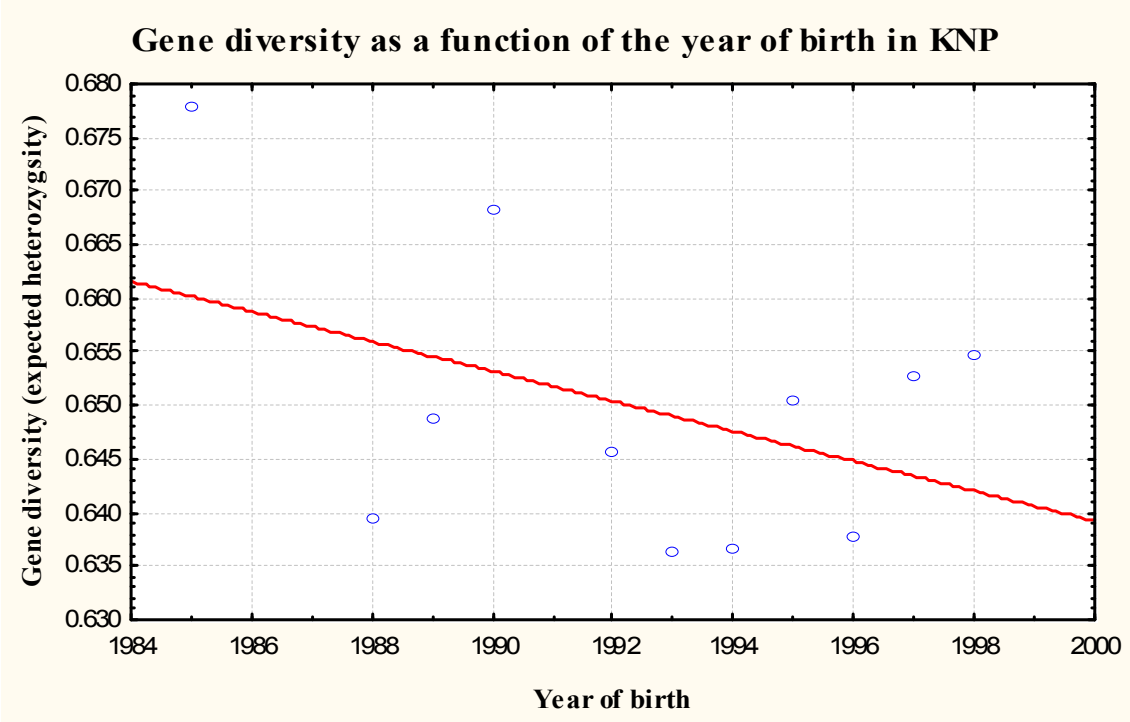
decline, since young animals that have not dispersed yet may result in elevated F_{ST} levels among herds and subsequent reduced levels of gene diversity. This is however unlikely since the F_{ST} values (based on Msat data) between herds were low. Periodical low N_e in HiP is also supported by fact that HiP experienced a loss of alleles (53 % of alleles found in KNP are private). No decline in diversity over time (1986-2004) could however be demonstrated for KNP (Figure 4.2b). It should however be noted that the time span covered is relatively short (18 and 16 years for KNP and HiP, respectively) and spans less than two generations, which may be too short to detect a signal of a significant decline in genetic variation in KNP.

Fig. 4.2a,b. Gene diversity as a function of year of birth within HiP and KNP. The steady decline in genetic diversity between 1988 and 2004 for HiP (a) is evident and also statistically significant (r = -0.64, p = 0.015, Spearman rank correlation). No significant decline in genetic diversity over time (year of birth) could be demonstrated for KNP (b) (r = -0.42, p = 0.18, Spearman rank correlation).

(a)



(b)



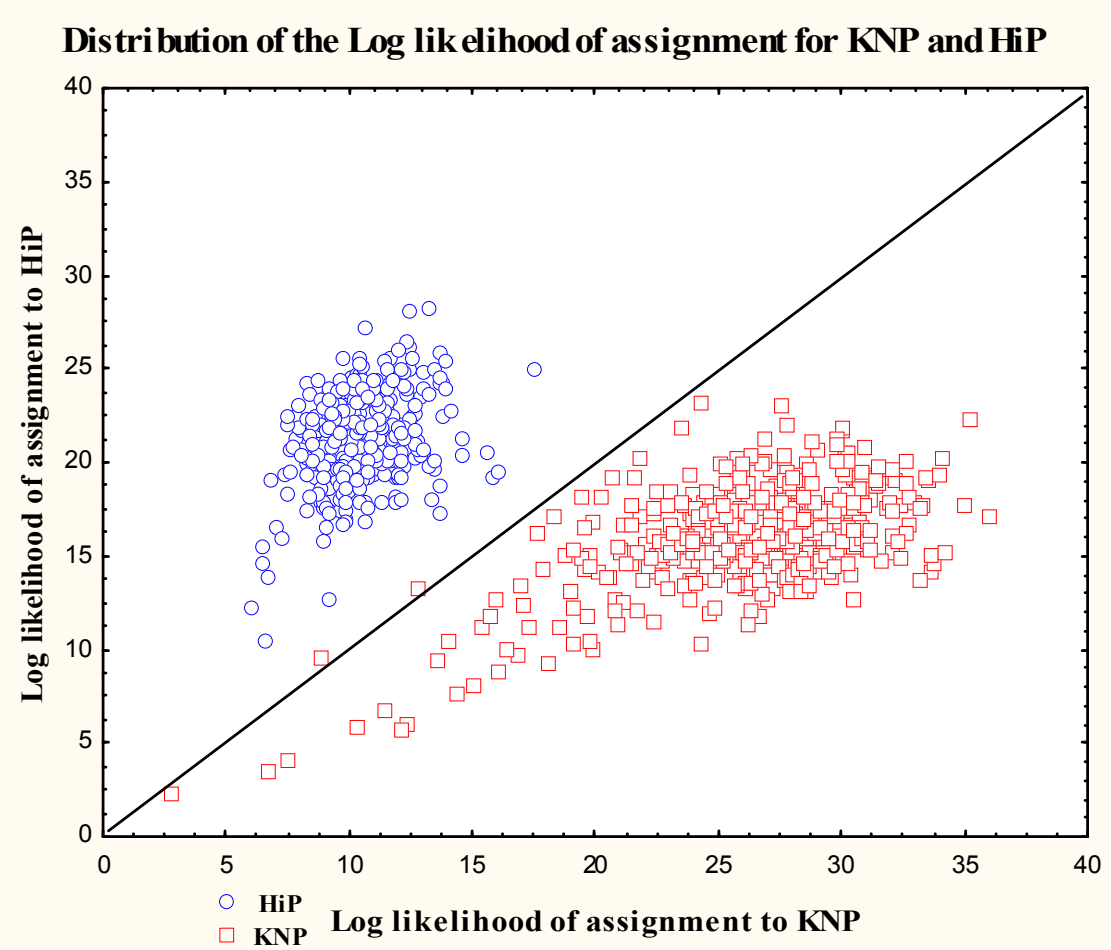
KNP exhibits a small fractional reduction in heterozygosity (F_{IS} = 0.034), which may be due to several factors, including segregating null alleles and the presence of population substructure in the park. The latter would result in the Wahlund effect (Hartl and Clark, 1989). Twenty one herds had positive F_{IS} values ranging from 0.008 to 0.095, while only 9 had slightly negative values. In view of the extensive sampling carried out over a wide geographic range within the park, and from 30 herds, population substructure may be excluded as a possible confounding factor that contributes to an elevated inbreeding coefficient. Chesser (1991) pointed out that the social and reproductive behaviour of buffalo may affect the level of inbreeding or F_{IS}. The phylopatry amongst females juxtaposed to a random dispersal and mating of bulls between the herds would result in an expected inbreeding coefficient of close to zero (Chesser, 1991). Reproducing and dispersing adult bulls may however enter their native herds at high frequencies and in a non-random fashion, which may contribute to a higher F_{IS} (Chesser, 1991). It should be noted that the reproductive behaviour may be complex and not fully understood. For spotted hyenas for instance it has been shown that reproductive skew with regard to resident and immigrant males was influenced by the dispersal status of the male and his length of residence in the population, while female choice played an important role in patterns of paternity (Engh *et al.,* 2002). Immigrant males among the hyena have subsequently been shown to father as much as 97 % of all offspring, and this form of mating behaviour has been suggested to be a form of kin selection and potential way of avoiding the deleterious consequences of inbreeding (Engh *et al.*, 2002). For buffalo in KNP (and HiP) it has been shown that male dominance may be very strong among herds and that the male effective population size can be very low as a result of a very small percentage of reproducing males (Van Hooft, 2005, unpublished results), which may contribute to a higher level of F_{IS}.

3.2 Indices of inter-population differentiation

A high level of inter-populational differentiation was demonstrated between KNP and HiP, based on both mtDNA sequence data (F_{ST} = 0.275; p < 0.001, R_{ST} = 0.176, SEM = 0.03) as well as Msat data (F_{ST} = 0.159, p < 0.001). The 95 % confidence intervals (CI) of the F_{ST} value based on Msat data ranged between 0.120 and 0.196. This high level of differentiation is also reflected by the distribution of the likelihood values (Log-values) of assignment of

individuals to their populations of origin (Figure 4.3). The graphs clearly illustrate the high level of affinities that individuals have for the populations they were sampled from.

Figure 4.3. Log likelihood distributions of assignments for HiP and KNP.



The lower mean likelihood value for KNP (7.39×10^{-17}) compared to HiP (7.65×10^{-11}) is a result of the greater genetic diversity for KNP. Assignment tests carried out with Geneclass confirm these results, as 99.4 % of all individuals were correctly assigned to their respective populations of origin. This high level of accuracy is dependent on the level of differentiation between populations. High levels of differentiation, such as between KNP and HiP (F_{ST} =

0.159, Msat data) enable assignment to be highly accurate and reliable, which has been confirmed by empirical data as well as long-term mark-and-recapture data. Berry *et al*. (2004) showed that for the grand skink (*Oligosoma grande)*, that assignment tests based on Msat data could correctly assign up to 100 % of individuals to their natal populations. They also found similar estimates of the proportions of dispersing skinks based on long-term mark-and-recapture data and assignment tests based on Msat data.

The high level of differentiation between KNP and HiP may be attributed primarily to genetic drift as a result of the bottleneck, and the magnitude of the effects of drift is inversely proportional to population size (Bodmer and Cavalli-Sforza, 1976). The relatively small N_e for HiP, and the fact that the two populations have been isolated from each other for many decades has probably been the primary contributor towards drift and subsequent divergence. O'Ryan *et al.,* (1999) also attributed the level of observed divergence between buffalo populations, including the St. Lucia population, to drift as did authors of a study on Bavarian red deer (*Cervus elaphus;* Kuehn *et al*., 2003). They ascribed the level of differentiation between populations to drift as result of a dramatic population reduction.

Our F_{ST} estimate, based on Msat frequency data, of 0.159 is notably higher than the 0.103 found by O'Ryan *et al*. (1998). This may be attributed to the use of different loci which most likely resulted in different levels of polymorphism, and to differences in the number of individuals sampled and analysed in this study, which was much higher than that in the earlier study. The substantially larger F_{ST} based on mtDNA sequence data in this study is not unexpected or uncommon. The small effective population size of the mitochondrion renders it more sensitive to drift than autosomal nuclear markers such as Msats (Wilson *et al.*, 1985). Since genetic distances based on mtDNA sequence data are female-biased, differentiation will also be accentuated in the event that the females are philopatric. Male biased dispersal and gene flow on the other hand usually results in a reduction of genetic differentiation based on nuclear loci such as Msats.

R_{ST}, which compares variance in allelic sizes rather than frequency (Slatkin, 1995), is not significantly higher than F_{ST} (Wilcoxon matched pair test: $p = 0.46$). It should also be noted that R_{ST} is more appropriate for estimating the genetic distance between populations sharing a common ancestor in the more distant past and for which the accumulation of new

mutations has contributed towards the observed level of differentiation. F_{ST} may thus be more appropriate and reliable as an indicator of differentiation between KNP and HiP, where genetic drift, as opposed to introduction of new mutations, was primarily responsible for the observed level of differentiation.

3.3 Intra-population diversity

Gene diversities within the two parks varied among the herds as well as among sub-populations (data pooled from a number of herds sharing the same deme/geographical area). For KNP, the northern sub-population (n = 83) had a H_e of 0.66 (SD = 0.037), differing significantly (p = 0.002, Wilcoxon matched-pair test) from that of the southern sub-population (H_e = 0.62; SD = 0.035; n = 182), supporting the hypothesis that the northern population exhibit signals of admixture. Haplotype diversity between the northern and southern populations also differed, although the difference was small (0.91 ± 0.02 and 0.89 ± 0.02 for individuals representative of the northern (n = 36) and southern (n = 48) sub-populations, respectively). These two sub-populations consisted of pooled data from individuals from the Pafuri and Crocodile bridge/Malelane areas respectively.

The higher diversity in the north of KNP supports immigration from the east and north of the park during the pre-fencing period. Van Hoof *et al*. (2003) also suggested high migration rates between buffalo populations in southern Africa, as supported by the fact that weak or no-isolation by distance could be demonstrated between these populations. They also showed the absence of significant differentiation between northern KNP and neighbouring Save Valley in Zimbabwe, suggesting gene flow between the two populations. Immigration of animals into an area may result in admixture, depending on the magnitude thereof, which may subsequently result in the population experiencing a transient status of linkage disequilibrium (LD). Statistical tests did point towards signals of LD for northern KNP, as 10.3% of pair-wise comparisons showed a significant ($p < 0.05$) association between alleles. This association was however non-significant after Bonferroni correction.

Subpopulations from HiP, consisting of pooled data from herds from Manzimbomvu in the north and Mbhuzane in the south, also showed differences with regard to haplotype

diversities. The mtDNA haplotype diversity of Mbhuzane (n = 21) was 0.52 (± 0.10) and that of Manzimbomvu (n = 18) was 0.30 (± 0.13). Genetic variation of these two areas also differed significantly based on Msat data. H_e of the northern population (n = 79) was 0.55, while that of the southern population (n = 91) was 0.51 (p = 0.02; Wilcoxon matched pair test).

At the herd level, the number of haplotypes varied between 1 and 4 for HiP and 2 and 7 for KNP (Table 4.3a,b). The average number of haplotypes per herd in HiP was 2.3 while for KNP it was 4.3. The gene diversity for KNP varied between 0.6 and 1 (all samples had different haplotypes), while it varied between 0 (meaning a single haplotype among entire sample/herd) and 0.7 for HiP. The fact that only a single haplotype could be found in particular herds in HiP suggest that females are all related by a single matrilineal line, indicating limited dispersal among the females. The low level of haplotype diversities within herds may however be an artefact of small sample size and low overall haplotype diversity. Relatedness among members in these herds must be mediated through the male line, as suggested by Van Hooft *et al.,* (2003). H_e per herd was consistently lower for HiP, ranging between 0.51 and 0.56, while for KNP it ranged between 0.57 and 0.67.

Table 4.3a. Diversity indices for HiP based on mtDNA haplotypic data. Hd = haplotype/gene diversity; N = sample size, H_n = number of haplotypes identified per sample size.

Herd ID	N	H_n	Hd	Nucleotide diversity (π)
A4	9	3	0.55 ± 0.16	0.025 ± 0.014
B4	9	1	0	0
AB2	6	3	0.73 ± 0.15	0.031 ± 0.019
J2	6	2	0.53 ± 0.17	0.027 ± 0.016
LM2	12	4	0.69 ± 0.09	0.028 ± 0.015
H2	5	1	0	0
B3	6	2	0.33 ± 0.21	0.016 ± 0.010
AC3	8	3	0.60 ± 0.16	0.027 ± 0.015
C2	5	2	0.40 ± 0.23	0.020 ± 0.013
D2	9	2	0.50 ± 0.12	0.025 ± 0.014
D3	4	2	0.50 ± 0.26	0.025 ± 0.017
EFG2	13	3	0.58 ± 0.12	0.025 ± 0.013
K2	5	2	0.60 ± 0.17	0.025 ± 0.016

Table 4.3b. Diversity indices for KNP based on mtDNA haplotypic data. Hd = haplotype/gene diversity; N = sample size, H_n = number of haplotypes identified per sample size.

Herd ID	N	H_n	Hd	Nucleotide diversity (π)
1	9	7	0.91 ± 0.09	0.044 ± 0.024
2	7	6	0.95 ± 0.09	0.044 ± 0.025
3	6	4	0.80 ± 0.17	0.047 ± 0.028
4	5	4	0.90 ± 0.16	0.022 ± 0.014
5	6	3	0.60 ± 0.21	0.029 ± 0.018
6	4	4	1.00 ± 0.17	0.086 ± 0.057
7	6	5	0.93 ± 0.12	0.030 ± 0.018
8	5	4	0.90 ± 0.16	0.040 ± 0.025
9	5	4	0.90 ± 0.16	0.043 ± 0.027
10	8	7	0.96 ± 0.07	0.039 ± 0.022
11	4	3	0.83 ± 0.22	0.040 ± 0.027
14	7	4	0.71 ± 0.18	0.047 ± 0.027
16	7	4	0.80 ± 0.12	0.049 ± 0.028
17	3	3	1.00 ± 0.27	0.063 ± 0.048
18	6	5	0.93 ± 0.12	0.029 ± 0.017
19	10	6	0.88 ± 0.07	0.061 ± 0.033
20	2	2	1.00 ± 0.50	0.041 ± 0.042
21	6	4	0.86 ± 0.12	0.034 ± 0.020
22	8	7	0.96 ± 0.07	0.053 ± 0.029
23	5	5	1.00 ± 0.12	0.055 ± 0.034
25	6	6	1.00 ± 0.09	0.050 ± 0.030
26	4	3	0.83 ± 0.22	0.035 ± 0.024
27	3	3	1.00 ± 0.27	0.038 ± 0.029
28	6	5	0.93 ± 0.12	0.050 ± 0.030
29	5	2	0.60 ± 0.17	0.051 ± 0.032
30	11	6	0.87 ± 0.07	0.040 ± 0.021
31	4	3	0.83 ± 0.22	0.055 ± 0.036
32	5	4	0.90 ± 0.16	0.060 ± 0.037

3.4 Intra-population level structure

3.4.1 Herd differentiation

Small but significant differentiation (F_{ST} = 0.012; 95%CI: 0.008-0.016), based on Msat data, was found among all the KNP herds (average herd size = 245). F_{ST} among the herds based on mtDNA was significantly higher (F_{ST} = 0.067; p = 0.006). The fact that the F_{ST} value across all herds and based on mtDNA data is more than four times larger than that based on nuclear Msat data, is indicative of strong male biased gene flow. Several herds exhibited high and significant differentiation ($p < 0.05$) based on mtDNA haplotype data (Table 4.4a). When the two sexes were analysed separately in terms of Msat data, differentiation across all herds was 0.008 (95%CI: 0.004-0.013) and 0.013 (95%CI: 0.009-0.018) for the males and females, respectively.

Differentiation among-herds in HiP (average herd size = 170) was small based on Msat data (F_{ST} = 0.014, 95% CI: 0.009-0.019), while no differentiation across all herds could be demonstrated based on mtDNA sequence data (F_{ST} = 0.005; p = 0.42). The absence of differentiation among herds may be an artefact of the low level of mtDNA diversity among the herds, and larger sample sizes may be needed to detect herd differentiation Pair-wise herd analyses however revealed that some herds are highly and significantly differentiated (Table 4.4b). The latter suggests little or no female dispersal is taking place between these herds, which is in accordance with ecological observations stating that these herds exhibited very coherent entities, hardly migrating out of their home range and showing little or no dispersal towards other herds.

The significant differentiation among herds based on both mtDNA and Msat data is in contrast to the study by Van Hooft *et al.* (2003). They could only demonstrate differentiation between pooled mtDNA data from northern and southern KNP, and attributed their observation to small sample size in combination with high haplotype diversities. The much higher F_{ST} between herds based on mtDNA compared to Msat data may not only be attributed to the smaller effective population size for mtDNA, since the effective population sizes for females are much larger than that of the males (Van Hooft *et*

al., 2003). The high F_{ST} based distances between herds may also be attributed to male biased dispersal and female phylopatry, both of which may attenuate mtDNA based F_{ST} estimates.

Table 4.4a. F_{ST} among herds in K NP based on mtDNA data. Values in red are significant ($p \leq 0.05$), and were calculated with Arlequin.

	herd1	herd2	herd3	herd4	herd5	herd6	herd7	herd8	herd9	herd10	herd11	herd14	herd16	herd17	herd18	herd19	herd20	herd21	herd22	herd23	herd25	herd26	herd27	herd28	herd29	herd30	herd31	herd32
herd1																												
herd2	-0.11																											
herd3	-0.04	-0.07																										
herd4	0.10	0.12	-0.01																									
herd5	-0.02	-0.04	0.14	0.41																								
herd6	-0.02	0.00	-0.03	0.16	0.09																							
herd7	-0.02	-0.02	0.09	0.24	-0.04	0.10																						
herd8	-0.13	-0.12	0.00	0.24	-0.10	-0.02	-0.05																					
herd9	-0.04	0.02	0.08	0.23	0.08	-0.06	0.05	-0.03																				
herd10	0.09	0.09	0.12	0.28	0.13	0.11	0.07	0.08	0.04																			
herd11	0.18	0.23	0.21	0.49	0.39	0.04	0.46	0.22	0.25	0.42																		
herd14	0.27	0.29	0.32	0.51	0.37	0.17	0.45	0.26	0.30	0.41	0.00																	
herd16	-0.10	-0.08	-0.04	0.09	0.07	-0.05	0.07	-0.07	0.00	0.15	0.10	0.18																
herd17	0.11	0.14	0.18	0.46	0.23	-0.06	0.32	0.08	0.15	0.32	-0.19	-0.18	0.02															
herd18	0.00	0.02	0.07	0.17	0.12	0.08	-0.04	0.04	-0.07	0.03	0.47	0.47	0.07	0.38														
herd19	0.04	0.05	0.07	0.20	0.14	0.00	0.17	0.04	0.07	0.16	0.02	0.04	-0.03	-0.07	0.19													
herd20	0.01	0.02	0.07	0.36	0.13	-0.10	0.06	0.04	-0.07	-0.16	0.41	0.37	0.07	0.22	0.00	-0.01												
herd21	-0.02	-0.05	-0.08	-0.01	0.15	0.10	0.03	0.03	0.13	0.07	0.41	0.44	0.02	0.35	0.02	0.15	0.08											
herd22	-0.11	-0.10	-0.05	0.08	-0.02	-0.05	-0.02	-0.12	-0.05	0.08	0.12	0.19	-0.11	0.02	0.00	-0.03	-0.04	0.00										
herd23	-0.12	-0.10	-0.07	0.06	0.03	-0.07	0.02	-0.11	-0.04	0.05	0.13	0.16	-0.16	0.02	0.02	-0.08	-0.11	-0.05	-0.14									
herd25	-0.12	-0.11	-0.10	0.04	0.02	-0.07	-0.01	-0.10	-0.07	0.01	0.18	0.27	-0.10	0.12	-0.05	-0.01	-0.19	-0.08	-0.12	-0.16								
herd26	0.09	0.07	0.12	0.36	0.14	0.08	0.11	0.11	0.10	-0.03	0.44	0.42	0.16	0.32	0.10	0.13	-0.19	0.11	0.07	0.04	0.00							
herd27	-0.14	-0.16	-0.12	0.06	-0.05	-0.07	-0.19	-0.13	-0.11	-0.07	0.38	0.39	-0.05	0.23	-0.25	0.08	-0.13	-0.18	-0.14	-0.13	-0.19	-0.04						
herd28	-0.10	-0.09	-0.05	0.12	-0.02	-0.06	-0.03	-0.11	-0.11	-0.07	0.20	0.26	-0.05	0.12	-0.07	0.02	-0.21	-0.05	-0.10	-0.11	-0.16	-0.05	-0.21					
herd29	0.02	0.01	0.13	0.38	-0.01	0.02	0.14	-0.04	0.11	0.22	0.13	0.05	-0.02	-0.12	0.24	-0.02	0.13	0.21	-0.05	-0.05	0.05	0.20	0.08	0.04				
herd30	0.01	0.00	0.11	0.27	-0.03	0.11	-0.01	-0.03	0.05	0.02	0.36	0.35	0.07	0.25	0.05	0.09	-0.19	0.07	-0.01	-0.02	-0.04	0.00	-0.09	-0.06	0.07			
herd31	-0.07	-0.04	-0.09	0.05	0.17	-0.09	0.14	-0.03	-0.01	0.15	0.09	0.15	-0.18	0.00	0.09	-0.06	0.05	0.04	-0.11	-0.19	-0.11	0.18	-0.04	-0.04	0.02	0.11		
herd32	-0.02	0.00	-0.01	0.17	0.15	-0.05	0.17	-0.03	0.04	0.20	-0.02	-0.01	-0.10	-0.15	0.17	-0.08	0.09	0.11	-0.08	-0.14	-0.03	0.19	0.03	0.00	-0.08	0.13	-0.17	

Table 4.4b. F_{ST} among herds in HiP, based on mtDNA data. Values in red are significant ($p \leq 0.05$), and were calculated with Arlequin.

	A4	AB2	AC3	B3	B4	C2	D2	D3	EFG2	H2	J2	K2	LM2
A4													
AB2	-0.12												
AC3	-0.10	-0.11											
B3	-0.10	-0.01	-0.09										
B4	0.17	0.36	0.21	0.07									
C2	-0.14	-0.07	-0.13	-0.22	0.13								
D2	-0.10	-0.10	-0.11	-0.08	0.25	-0.13							
D3	-0.19	-0.15	-0.19	-0.24	0.22	-0.28	-0.20						
EFG2	-0.03	-0.02	-0.09	-0.02	0.18	-0.06	-0.01	-0.10					
H2	0.07	0.23	0.11	-0.03	0.00	0.00	0.15	0.06	0.10				
J2	-0.13	-0.14	-0.14	-0.11	0.29	-0.17	-0.16	-0.24	-0.05	0.16			
K2	-0.04	-0.06	0.03	0.08	0.39	0.04	0.09	-0.01	0.06	0.25	0.06		
LM2	-0.01	-0.12	-0.05	0.08	0.36	0.03	-0.03	-0.05	0.03	0.27	-0.06	0.08	

3.4.2 Sub-population differentiation

A less pronounced degree of differentiation was found between subpopulations separated by large geographical distances in KNP. The F_{ST} between the northern (Pafuri area, n = 36) and southern (Malelane/Crocodile bridge area, n = 48) populations, separated by a geographical distance of more than 300 kilometres, based on mtDNA haplotype frequencies, was 0.026 (p = 0.017). This indicates that these two subpopulations exchange 37.5 migrants per generation (absolute number of migrants exchanged between the two subpopulations, and M = Nm for haploid populations, M assumes equilibrium values for F_{ST}). Based on Msat data, the F_{ST} between the two populations (Pafuri area, n = 83; Malelane/Crocodile bridge area, n = 182) was 0.012 (p = 0.01). Van Hooft *et al*. (2001) found similar results for Tsavo National Park in Kenya, where significant differentiation could be demonstrated between Tsavo West and Tsavo East. Both restricted female migration, as well as herd differentiation were put forward as factors that contributed to the observed level of differentiation. Van Hooft *et al*. (2003) however pointed out that small levels of population differentiation may also be attributed to demographic events in the evolutionary history of the population. Previously, these authors could not demonstrate significant differentiation between northern and southern KNP. This could be due to their reduced sample size and different panel of loci used. For the Manzimbomvu (n = 18) and Mbhuzane (n = 21) areas in HiP, the level of differentiation based on mtDNA haplotype frequencies was small and non-significant (F_{ST} = 0.016; p = 0.26). It is interesting however that differentiation among subpopulations in HiP is comparable to that in KNP, considering the much smaller size of the former park.

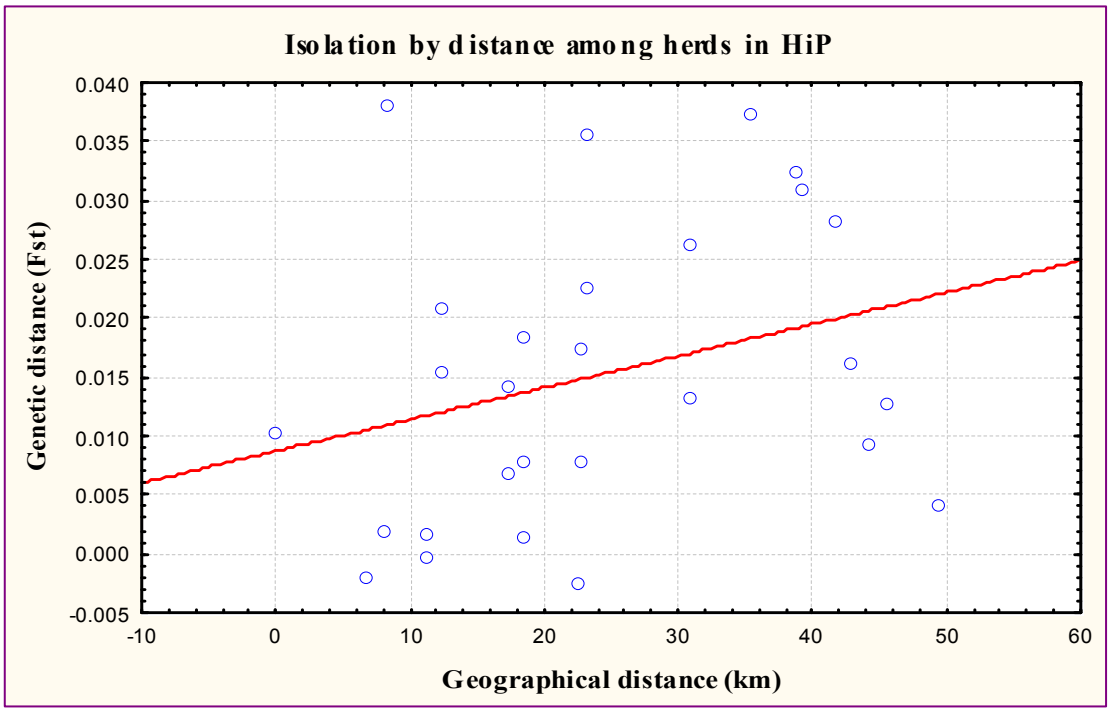
3.4.3 Isolation by distance

A Mantel test revealed a significant correlation (correlation coefficient: 0.36, $p < 0.01$) between geographical distance and genetic distance within KNP (Figure 4.4b), indicating that the degree of dispersal declines with increasing geographical distance. This relationship was evident for both males and females within KNP. A different picture was however obtained for HiP, as no significant signal of isolation by distance could be demonstrated. The positive outcome for the test for KNP suggests that limited dispersal over geographical distance may contribute to the small but significant level of differentiation among the herds.

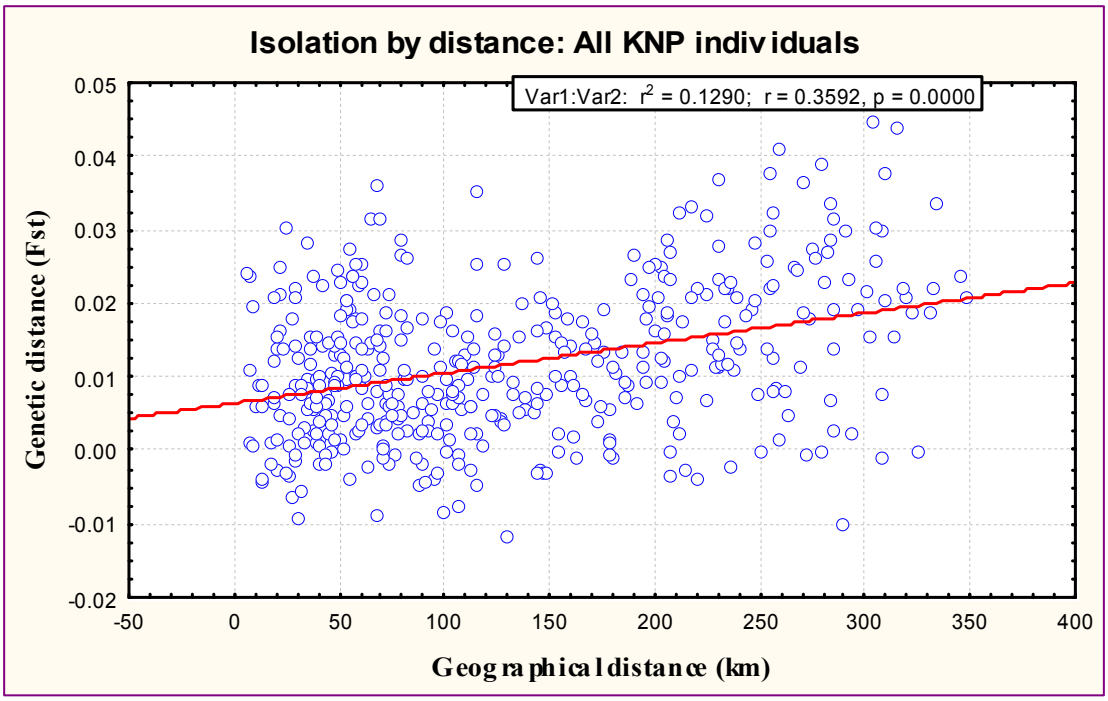
The opposite is true for HiP, where no structure/differentiation across all herds coincided with the absence of isolation by distance. In both parks however it is evident that certain herds (Table 4.4 a,b) are highly differentiated based on mtDNA sequence data, suggesting limited female gene flow between these herds, even for herds separated by relatively small geographical distance. It should be noted however that the absence of isolation by distance in HiP may be affected by the sample size and the small size of the park, and that by increasing the sample size, this bias may be addressed.

Fig. 4.4 a,b. Isolation by distance among herds within HiP and KNP, based on Msat data.

(a)



(b)



3.5 Sex and age-biased dispersal

It is a common tendency for males and females to exhibit biased dispersal (Greenwood, 1980). Advantages to dispersal *vs*. philopatry are thought to be related to risk of inbreeding and kin competition, which often differs between sexes due to mating systems and general life history (Perrin and Mazalov, 2000). Sex-biased dispersal is also a major factor influencing population dynamics and structure. Philopatry on the one hand contributes towards population divergence, while dispersal on the other counteracts divergence. A number of tests were executed in order to detect signals of sex-biased dispersal, and to assess whether it differs between KNP and HiP. Several tests using FSTAT revealed significant levels of biased dispersal in KNP. One sided tests based on 4900 randomisations revealed that males exhibited lower mean overall assignment values than females (-0.418 *vs*. 0.278, p = 0.02). Females also exhibited more structure, assessed in terms of $F_{ST,}$ than males (0.014 *vs*. 0.008, p = 0.05), while the mean degree of relatedness for the males was significantly lower than for the females (0.015 *vs*. 0.027, p = 0.05).

The above tests for sex-biased dispersal in KNP all indicate significant male-biased dispersal and strong female structure at the herd level. The fact that F_{ST} based on mtDNA data is large compared to F_{ST} based on Msat data in KNP, supports these finding of male-biased dispersal. Strong male-biased dispersal and a non-random re-entry of males into their native herds may result in positive F_{IS} values (Chesser, 1991) while the absence of negative F_{IS} values also indicates that males mate with females from their native herds (Van Hooft *et al*., 2000). Twenty one out of the 30 herds sampled in KNP exhibit positive F_{IS} values (Table 4.5), which may suggest that males are mating with females from their native herds. In HiP however, all but one herd (herd AC 2003, Table 4.6) exhibited negative F_{IS} values or values close to zero. The relationship coefficient *r* for males within KNP was significantly higher than that among the males within herds (0.001 *vs*. -0.0064, Wilcoxon matched pairs test). This suggests that males within herds are more closely related to each other than to males from different herds. These results also suggest that dispersal in KNP is not completely random, which may contribute to the positive F_{IS} values within herds. The relationship coefficient for females within KNP was 0.018, which was significantly larger than among the herds (-0.005, $p < 0.001$, Wilcoxon matched pairs test), indicating strong matrilineal herd structure.

Sex-biased dispersal tests for HiP (using FSTAT) also indicate that the males are the dispersing sex. F_{ST} for females were close to significantly larger than for males among herds (0.017 *vs*. 0.007; p= 0.056). The relatedness test furthermore revealed that females within herds are more related than males (0.035 *vs*. 0.015; p = 0.054). The relationship coefficient for males within HiP was 0.0089, which was not significantly higher than that among herds in the park (-0.0077. p = 0.33). This indicates high levels of dispersal among the males and in a random fashion. The relationship coefficient for among females within HiP was 0.026, which was significantly higher than the -0.0075 for females within herds ($p < 0.001$, Wilcoxon matched pairs test). The latter indicates strong matrilineal herd structure among the females.

In an attempt to qualify the age at which dispersal commences, the Log-likelihood of the assignment scores for individuals from both sexes and per age group were calculated. No significant difference in the Log likelihood of assignment between the different age classes could however be demonstrated due to high levels of variation in the assignment scores within each age and sex class.

Table 4.5. Gene diversities and inbreeding coefficients per herd in KNP based on Msat data.

Herd ID	Sample size	H_e	H_e SD	F_{IS}
1	13	0.61	0.07	0.171
2	19	0.65	0.06	0.046
3	15	0.67	0.06	-0.015
4	20	0.64	0.06	0.015
5	19	0.62	0.06	0.035
6	18	0.65	0.05	0.016
7	17	0.67	0.06	0.075
8	18	0.66	0.05	0.041
9	18	0.67	0.04	0.014
10	19	0.63	0.06	0.047
11	18	0.62	0.06	0.032
12	9	0.65	0.06	0.023
14	16	0.67	0.05	0.078
15	14	0.66	0.06	-0.031
16	16	0.67	0.05	0.008
17	16	0.63	0.06	0.016
18	15	0.67	0.05	0.029
19	7	0.65	0.07	-0.026
20	19	0.66	0.05	0.009
21	15	0.62	0.06	-0.017
22	16	0.63	0.06	0.026
23	16	0.66	0.06	-0.015
24	15	0.65	0.05	0.014
25	15	0.62	0.06	-0.001
26	16	0.62	0.06	-0.02
27	15	0.62	0.07	0.088
28	16	0.58	0.05	-0.013
29	17	0.58	0.06	0.023
30	22	0.63	0.05	0.015
31	10	0.64	0.05	-0.021

H_e: Nei's unbiased gene diversity; F_{IS} : inbreeding coefficient (Weir and Cockerham, 1984); H_e SD: inter locus standard deviation.

Table 4.6. Gene diversities and inbreeding coefficients per herd in HiP based on Msat data.

Herd ID	Sample size	H_e	H_e SD	F_{IS}
AB 2002	25	0.55	0.05	-0.015
J 2002	25	0.55	0.05	-0.007
LM 2002	27	0.56	0.05	-0.023
C 2002	25	0.57	0.05	-0.06
EFG 2002	47	0.52	0.05	0.009
D 2002	44	0.51	0.05	0.001
AC 2003	25	0.52	0.06	0.052
B 2003	26	0.52	0.05	-0.065
D 2003	25	0.55	0.05	-0.012
H 2002	44	0.55	0.04	-0.008
K 28 2002	5	0.62	0.05	-0.177
K23 2002	4	0.55	0.07	-0.118
A 2004	37	0.56	0.05	-0.026
B 2004	42	0.53	0.04	-0.006

H_e: Nei's unbiased gene diversity; F_{IS} : inbreeding coefficient (Weir and Cockerham, 1984); H_e SD: inter locus standard deviation.

3.6 Detection of past demographic events: Bottlenecks

Genetic bottlenecks may be induced after severe population crashes, resulting in a potentially dramatic reduction in genetic variation (Bonnell and Selander, 1974; Goldsworthy *et al.*, 2000; Weber *et al.*, 2000), while both a bottleneck and reduced population size may result in increased genetic drift and subsequent reduction in genetic variation. The degree to which variation is lost is however complex and may be a function of several factors (Nei, 2005). The inability to detect known bottleneck effects is not uncommon (Kuehn *et al.*, 2003; Harley *et al., 2005*, Muwanika *et al*., 2003; Nyakaana *et al*., 2001; Simonsen *et al*, 1998; Waldick *et al.,* 2002), and a primary reason for this is due to the fact that its impact may be obliterated after a few generations. Methods have subsequently been developed in order to circumvent this. Populations that experience a significant reduction in their effective size may develop a gene diversity excess, in which case the observed gene diversity is higher than the expected equilibrium gene diversity (*Heq*) due to the removal of rare alleles (Waldick *et al.*, 2002). This excess may be a transient feature that may not last beyond only a few generations, after which equilibrium is obtained again (Cornuet and Luikart, 1996).

Populations that have not recovered from the effects of a bottleneck may be in a state of disequilibrium (Slatkin, 1994). A LD (linkage disequilibrium) test revealed that 12 and 15 pair-wise comparisons respectively out of a total of 136 comparisons were found to be associated for KNP and HiP, while only 7 out of 136 pair wise comparisons are expected to be associated by chance. After Bonferroni correction, however, only 3 and 4 loci respectively were significantly associated for KNP and HiP ($p < 0.001$). It can thus be concluded that both populations are in linkage equilibrium and have recovered from the effects of the bottleneck.

The heterozygosity excess test carried out with the software package BOTTLENECK supported the LD test. Although 3 and 5 loci respectively in HiP showed a higher heterozygosity ($p < 0.05$) than expected at equilibrium under SMM and IAM, the Wilcoxon sign-rank test revealed no significant overall excess or deficiency. The mode-shift test, the descriptor of the allele frequency distribution, also indicated a normal L-shaped distribution under both SMM and IAM. The latter also indicates a state of equilibrium. Under the TPM

(with a 70 % proportion of SMM) only one locus (TGLA 227) exhibited a higher than expected heterozygosity ($p < 0.01$) while no excess across all loci could be demonstrated (probability for excess = 0.00467). The TPM also revealed a normal L-shaped distribution. It has been shown that most simulated datasets fit the TPM model better than either the SMM or IAM models (Di Rienzo *et al.,* 1994) and results based on TPM should be a more realistic reflection of the extent to which a heterozygosity excess prevails.

For the KNP population, the heterozygosity at 4 and 8 loci respectively under IAM and SMM deviated from expected values under equilibrium. The Wilcoxon sign-rank test, assuming IAM and mutation-drift equilibrium, revealed an overall heterozygote deficiency (probability of deficiency = 0.99), while no excess could be demonstrated (probability for excess = 0.0001). Under SMM however the probability for a heterozygote deficiency was 0.00004, while for an excess it was 0.99, suggesting a deficit in rare alleles and possible remnant of a bottleneck in the recent past. However, a normal L-shaped allele frequency distribution was found, contradicting this. When the KNP population was analysed assuming TPM, only one locus (ILSTS 026) exhibited a significantly ($p = 0.01$) higher than expected heterozygosity. Across all loci however, neither a significant deficit nor excess of heterozygosity could be demonstrated (probability for deficiency = 0.77; probability for excess = 0.24), and the mode-shift test revealed a normal L-shaped distribution of allele frequencies. Similar results were obtained for the common warthog that experienced a recent, severe bottleneck. Overall, the loci in the warthog (*Phacochoerus africanus)* exhibited high levels of allelic richness (> 4 alleles per locus) and expected heterozygosity, and the bottleneck test revealed that the expected gene diversity (*He*) was greater than the expected equilibrium gene diversity (*Heq*) at four out of seven loci in two populations from Queen Elizabeth National Park and Murchison Falls National Park in Uganda (Muwanika *et al*., 2003). The Wilcoxon sign-rank test however revealed no mode shift in the distribution of allele frequencies that would indicate a bottleneck event in these populations.

Several explanations can be put forward for the absence of a detection signal for remnants of the rinderpest bottleneck in KNP. Close to 14 generations have passed since the bottleneck, which may have enabled the populations to reach equilibrium again. The severity and duration of the bottleneck was short, and may have been overestimated (Simonsen *et al*., 1998, O'Ryan *et al.*, 1998). Harley *et al.* (2005) found that although the

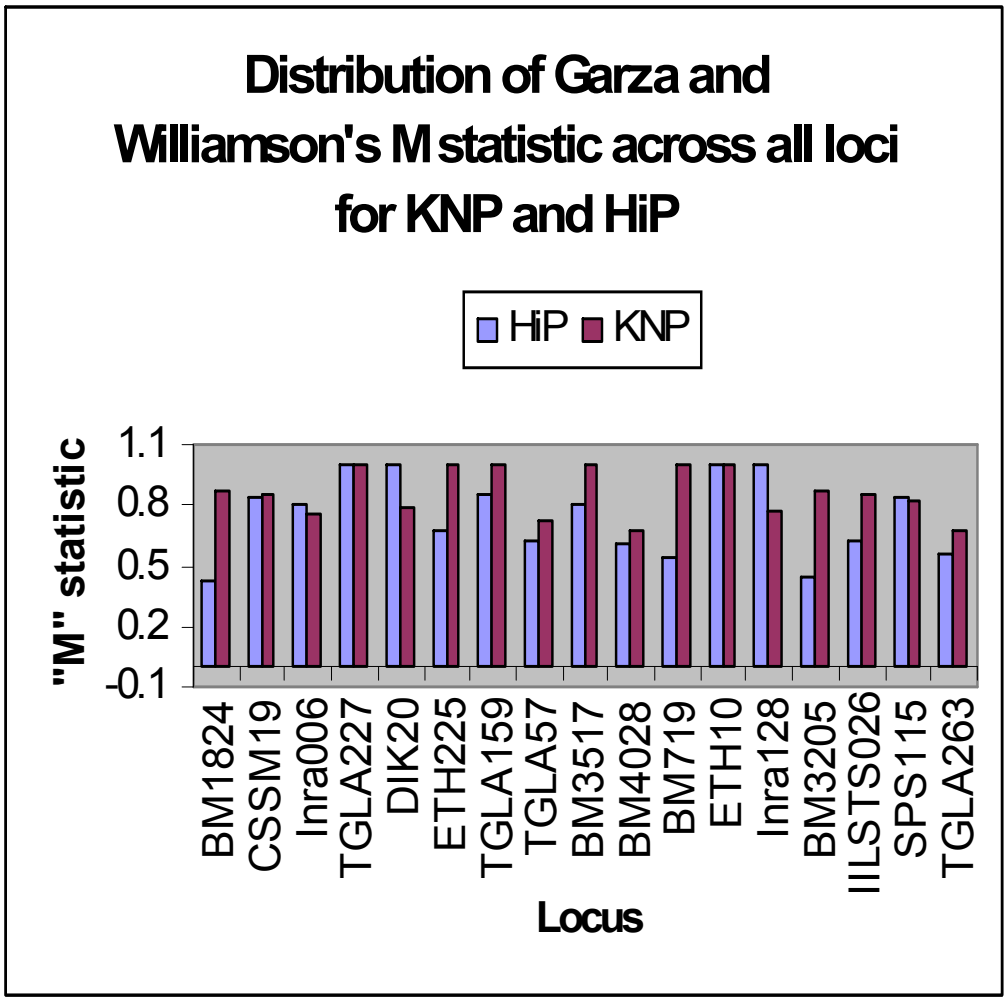
black rhinoceros (*Diceros bicornis*) suffered a severe population reduction of 96 % at the turn of the last century mainly due to poaching, the species seemed to retain most of its diversity and concluded that the bottleneck did not significantly affect the species. Assuming that the population reduction in KNP has been overestimated, it may be that genetic variation in the population before the bottleneck occurred was high and that the populations that remained after the rinderpest exhibited high levels of variation. Restoring variation will also be enhanced in the event that a population increases in size rapidly following the population bottleneck (Nei *et al*., 1975; Avise, 1994), which is known for KNP. Waits *et al.* (2000) showed that Scandinavian brown bears that experienced a severe population bottleneck during the nineteenth century did not reveal a significant loss of genetic variation when compared to non-bottlenecked populations. They could also not detect genetic signals of the bottleneck, and attributed this inability to rapid population growth and re-establishment of gene flow between bear populations in which allelic diversity and heterozygosity was preserved in separate units.

The M statistic of Garza and Williamson (2001), which is a measure of the occurrence of gaps in the allele size distributions as a result of a bottleneck, showed a similar outcome when compared to the heterozygosity excess tests. The M-statistic for HiP was 0.742 (variance = 0.039) and 0.860 (variance = 0.015) for KNP (Table 4.1). A value > 0.8 is typical of outbred populations and < 0.7 typical of severely bottlenecked populations. The distribution of Garza and Williamson's M- statistic across all loci is shown in figure 4.5. The M statistic was found to be consistently lower for HiP when compared to KNP. Overall, 8 loci have M values lower than 0.7 for HiP, but only two for KNP. This, together with the overall M statistic that was close to 0.7, suggests that HiP experienced a more severe reduction in population size and subsequent bottleneck effect.

The absence of the remnants of a bottleneck in KNP is in accordance with mtDNA D-loop sequence data. Mismatch distribution tests could not detect signals of expansion, and suggested that the KNP population is in equilibrium. D-loop data for HiP on the other hand suggested a population contraction. It is known that due to the effective population size of the mitochondrial genome, it is more sensitive towards bottleneck events and subsequent drift. In some instances the Msat and mtDNA data is congruent, as both indicate a decline in genetic diversity, suggesting that the size of the current population in HiP is too small to

maintain historical levels of gene diversity. Although the topology inferred from the haplotype data suggests an expansion event, it may be attributable to one in the more distant past. By no means has enough time passed for mutations to accumulate in either Msat loci or mtDNA D-loop sequences to generate a signal of a recent expansion.

Fig. 4.5. Comparative distribution of Garza and Williamson's M-statistic (Garza and Williamson, 2001) across all loci for KNP and HiP.



3.7 Phylogenetic relationships among KNP and HiP haplotypes

The model of sequence evolution that best fitted the D-loop data was TrN+I+G, where I = 0.66 and G = 1.11. The frequencies for the respective bases were 0.23 (A), 0.14 (C), 0.27 (G) and 0.34 (T) respectively, clearly indicating a base frequency bias, whilst the transition/transversion ratio was 63.10. The relative rate test of Tajima (Tajima, 1993) revealed that the rate of change among the respective haplotypes did not differ significantly (chi-square = 0.11, p = 0.73).

The tree topologies inferred from both the NJ and MP were very similar, with several nodes having high levels of bootstrap support (Fig. 4.6). The phylogenetic relationships among the haplotypes from KNP and HiP, revealed that there was no geographical partitioning of haplotypes to their respective populations of origin. This is not totally unexpected, since the time of divergence between the two populations has not been sufficient for signature mutations to accumulate. This is also reflected by a relatively small degree of divergence between haplotypes from the two populations: the mean Tamura-3 parameter corrected distance between the haplotypes from KNP and HiP was only 0.051 (SE: 0.007). The branch lengths of the KNP haplotypes however indicate a higher level of divergence when compared to that of HiP. This is also reflected by the higher level of intra-haplotypic genetic distances (KNP: 22.09 ± 9.78; HiP: 11.14 ± 5.10) and nucleotide diversities (0.025 ± 0.012 for HiP compared to 0.049 ± 0.024 for KNP, see chapter 2). The bootstrap values at ten nodes of the tree were below 50 %, indicating that the relationships among the haplotypes are not very robust, pointing out the absence of deep structuring among the haplotypes. Haplotypes separated by small genetic distances are typically found on the short branches of the phylogenetic tree, since coalescences of ancestors occur within a relatively short period (Marjoram and Donnelly, 1994).

Limited phylogeographic partitioning of haplotypes between fragmented populations have also been observed in Ugandan kob (*Kobus kob*) populations. The distribution of haplotypes in these populations were attributed to recurrent genetic contact between populations over an evolutionary time scale rather than to panmixia or long-distance gene flow between the populations occurring in the recent past. The philopatric nature of the kob excluded panmixia as a possible explanation for the observed level of sequence divergence between

the different kob populations. Since gene flow between HiP and KNP is excluded in the recent past, the observed level of divergence between the two populations may also be attributed to genetic contact between the populations in the more distant past. Simonsen *et al.* (1998) also proposed that the observed low level of differentiation between buffalo populations at the regional level may be attributed to fragmentation of a previously panmictic population.

The pattern of observed phylogenetic relationships is unlikely to have been affected by the mutation model and subsequent effects of homoplasy, since genetic divergence among closely related populations is essentially due to random drift (Estoup *et al.,* (2002). Also, according to Simon *et al.,* (1994), most distance correction methods will yield approximately the same results if the taxa are closely related, since homoplasy will be small, provided that substitution rates do not vary among lineages (Kumar *et al.*, 1993).

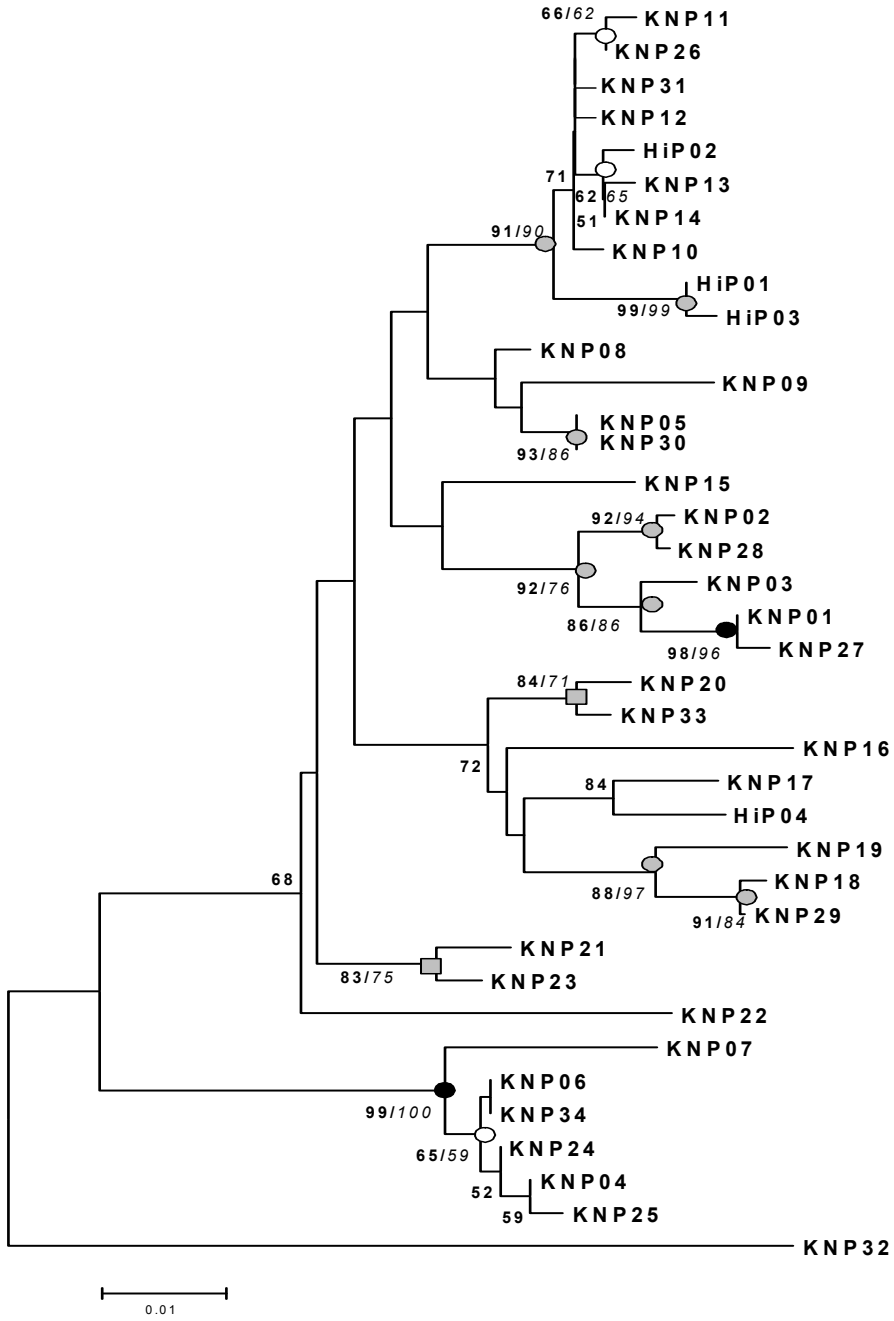


Fig. 4.6 Neighbour-Joining tree depicting the phylogenetic relationships among haplotypes from KNP and HiP. Distances were corrected using the Tamura-3 parameter model and taking a gamma shape parameter of 1.11 into account. Branch support was estimated following 1000 bootstrap replications. Numbers in bold represent the bootstrap support for the nodes from NJ, whilst those in italics represent nodal support from parsimony. Open circles indicate bootstrap support of 59-69%, open squares 70-85%, grey circles 86-95% and black circles 96-100%.

4. Conclusions

Populations that have experienced reductions in their sizes whilst being fragmented and isolated have been the focus of many scientific investigations. The reason for this is central to one of the most important concepts in population genetic theory, which is that a reduction in population size may result in reduced genetic variation, with its associated negative consequences. These may include, in extreme cases, a compromised ability to respond to environmental stochasticity (Mills and Smouse, 1994) and elevated levels of inbreeding. The latter may also intensify the impact of environmental stressors on populations (Bijlsma *et al*., 2000). It is thus quite obvious that these populations may require intensive management and conservation practices to ensure their viability.

The buffalo of both KNP and HiP experienced dramatic population reductions at the turn of the 19th century, only to make remarkable recoveries in their numbers as a result of reproductive success in a favourable environment. The sizes of the populations of both parks following the rinderpest were reported to be extremely small, giving rise to the expectation that genetic variation was compromised. The level of variation in KNP is however moderate to high, while that of HiP is moderate to low, which may suggest a differential impact of the rinderpest epidemic. In contrast to previous studies of buffalo populations in South Africa (O'Ryan *et al*., 1998) and elsewhere in Africa (Simonsen *et al.,* 1998; Wenink *et al*., 1998; Van Hooft *et al.,* 2000), KNP and HiP also show a high degree of differentiation, based on both Msat and mtDNA data. Considering that the two populations are fragmented and that gene flow is completely excluded between the two parks, the observed level of differentiation can primarily be attributed to drift. Other factors such as anthropological interventions during the last century, in all probability also affected the current status of genetic parameters of the two parks. The effect of drift is clearly illustrated by the prominent difference in their respective Msat allele frequency distributions showing not only large differences in allele frequencies, but also the occurrence of large numbers of private alleles, almost exclusively in KNP. The high level of differentiation, coupled with very prominent differences in allele frequency distributions, suggest that the rinderpest and subsequent small census sizes have played prominent roles in shaping the dynamics of these populations.

The high F_{ST} based on mtDNA not only indicates pronounced drift, it also suggest strong female phylopatry, which is in accordance with previous ecological observations. Strong sex-biased dispersal could be demonstrated for KNP but not for HiP, which may be attributed to, amongst other, the lack of mtDNA diversity and the small size of the park. The strong sex-biased dispersal detected in KNP may contribute to the elevated inbreeding coefficient (F_{IS} = 0.034), which is exemplified by 5 loci that exhibited a heterozygote deficit. Males that re-enter herds at a frequency higher than expected may cause a "male inbreeding effect" (Chesser, 1991). The fact that the majority of herds in KNP exhibit positive inbreeding coefficients supports the male-inbreeding hypothesis. HiP on the other hand had an inbreeding coefficient close to zero, which is indicative of an outbred population, exhibiting non-random mating. Another unexpected finding was a small but steady decline in the genetic diversity of HiP between 1986 to 2004, which may suggest episodes of low Ne which when taken together with the low level of mtDNA haplotype diversity (0.48), indicates that HiP is showing signs of population contraction. The high number of haplotypes per herd in KNP suggests that gene flow is primarily responsible for maintenance of genetic diversity, since a herd should not contain more than 1 or 2 haplotypes in the absence of gene flow. This very low level of mtDNA diversity for HiP may be the result of many factors, of which the impact of the bottleneck and subsequent slow recovery of the population are among the most plausible explanations. The severe reduction in the population may have reduced the level of diversity to what is currently observed, although pre-bottleneck diversity may also have played a major role.

The genetic indices of differentiation among herds in HiP is in accordance with ecological observations that found certain herds formed tight coherent assemblages, showing little or no emigration to other herds (Jolles, 2004). An age effect that contributes to this level of differentiation is however not excluded, since animals sampled from the herd that deviates from the remainder were are all born in 2003. Strong signals of male biased dispersal are evident for both KNP and HiP. The absence of a detectable signal of remnants of the rinderpest bottleneck in both populations may be attributed to several factors. First and foremost it may be that the pre-bottleneck level of genetic variation was high, coupled with an overestimation of both the severity and duration of the bottleneck. The latter is in accordance with explanations put forward by other authors (Simonsen *et al.*, 1998; O'Ryan *et al.*, 1998). KNP also subsequently showed a remarkable recovery with regard to

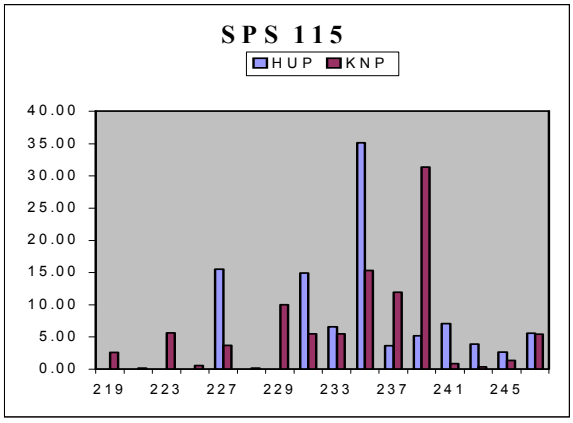
population size within two to three decades, whilst immigration of animals from north and east of the country in all probability resulted in introgression of rare alleles, supplementing genetic diversity. The contrary is evident in HiP. The park not only lacks a large percentage of the Msat alleles found in KNP, but it is also impoverished with regard to haplotype diversity. Still, remnants of the bottleneck could not be detected using Msat data. MtDNA data however, being more sensitive to population reductions and drift, suggested that HiP experienced a population contraction in the recent past.

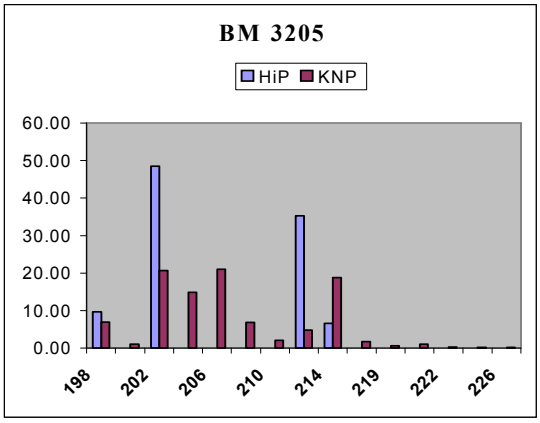
On an evolutionary scale, the fact that the mtDNA haplotypes from the two parks do not cluster separately suggests that both parks share common ancestry. By no means has enough time passed since the bottleneck to enable population-specific mutations to accumulate, and the small divergence among the haplotypes suggests contact between the two populations in the past.

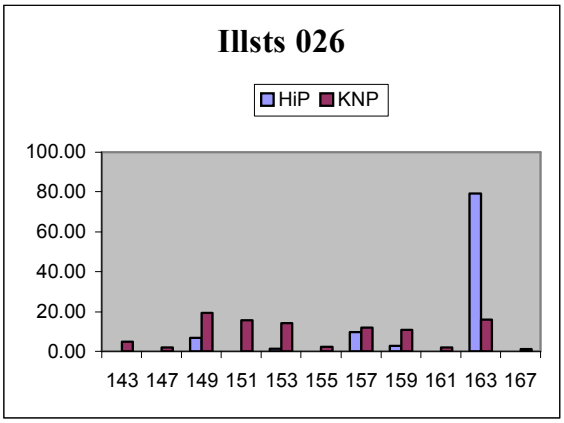
From a conservation point of view it is essential to assess the historical and present status of several population genetic parameters as a basis from which to make management decisions. Genetic diversity and effective population size are two of the most important parameters used as indicators of fitness and sustainability of a population. Measures of genetic diversity will assist in demarcating a genetically viable population, while measures of population differentiation will contribute to the identification of different ecological groupings (Harley *et al.*, 2005). When differentiation between populations has arisen over a long period of time and due to genetic drift, it would be meaningful to manage them separately (Harley *et al.*, 2005), since new mutations introduced into the populations separately would have established the populations on different evolutionary trajectories (Harley *et al.*, 2005). KNP and HiP have diverged only recently due to drift and fragmentation, and the contribution of new mutations since fragmentation should be minimal, suggesting that there is no requirement from a genetic standpoint for separate management practices. However, the differential FMD status between these populations necessitates vastly different animal health management. In the event that it becomes necessary to supplement diversity in HiP, these populations could theoretically be mixed in order to regenerate genetic diversity and counter drift (Harley *et al.*, 2005). Forbes and Hogg (1999) suggested that individuals can be exchanged from populations that exhibit minimal levels of historical gene flow (F_{ST} < 0.2). Mills and Allendorf (1996) demonstrated, through modelling, that populations where

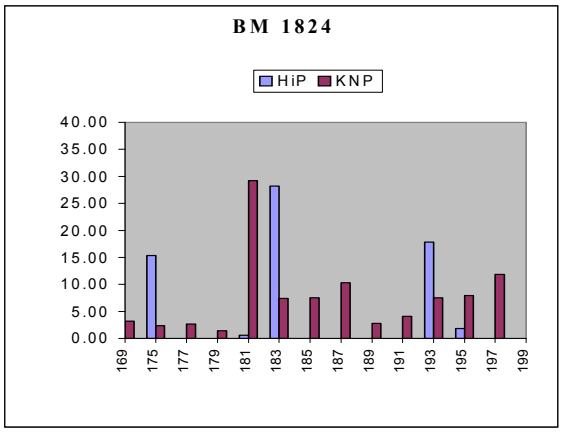
F_{ST} was 0.2 have historical migration rates of at least one migrant per generation. These results support the recommendations of Harley *et al.* (2005) and others concerning the exchange of individuals between closely related populations. Although the buffalo populations of KNP and HiP are not under threat of extinction and exhibit moderate to high levels of diversity, indices of genetic variation and differentiation presented in this chapter, both at the population level and also at the herd level, can be used as a baseline from which sound genetic management practices can be formulated.

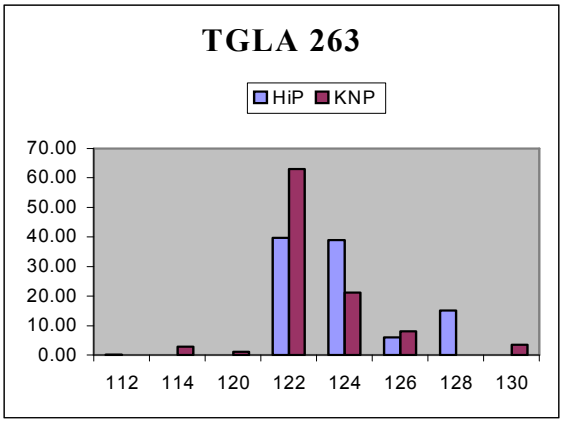
Fig. 4.1 Allele size distributions across all loci within KNP and HiP. The allele designations are represented on the X-axis, while their respective frequencies are shown on the Y-axis.

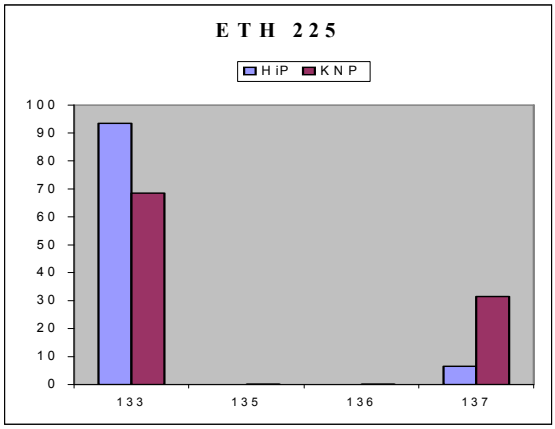

SPS 115
HUP
KNP
40.00
35.00
30.00
25.00
20.00
15.00
10.00
5.00
0.00
219
223
227
229
233
237
241
245

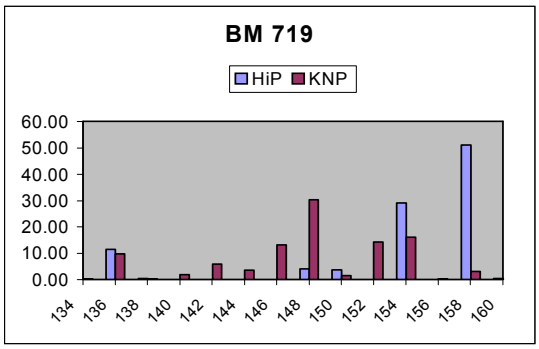

BM 3205
HiP
KNP
60.00
50.00
40.00
30.00
20.00
10.00
0.00
198
202
206
210
214
219
222
226

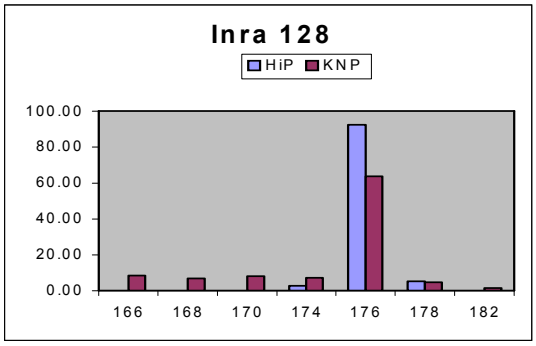

Illsts 026
HiP
KNP
100.00
80.00
60.00
40.00
20.00
0.00
143 147 149 151 153 155 157 159 161 163 167

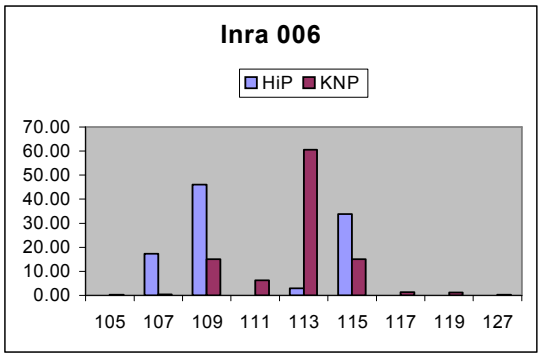

BM 1824
HiP
KNP
40.00
35.00
30.00
25.00
20.00
15.00
10.00
5.00
0.00
169
175
177
179
181
183
185
187
189
191
193
195
197
199

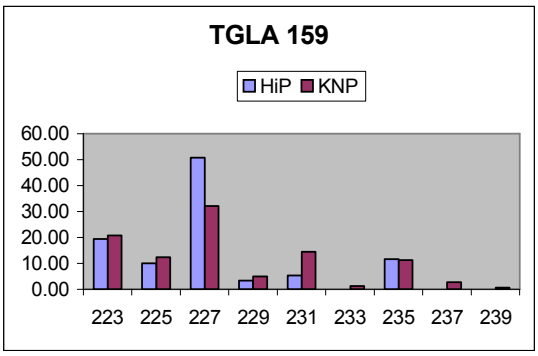

TGLA 263
HiP
KNP
70.00
60.00
50.00
40.00
30.00
20.00
10.00
0.00
112
114
120
122
124
126
128
130

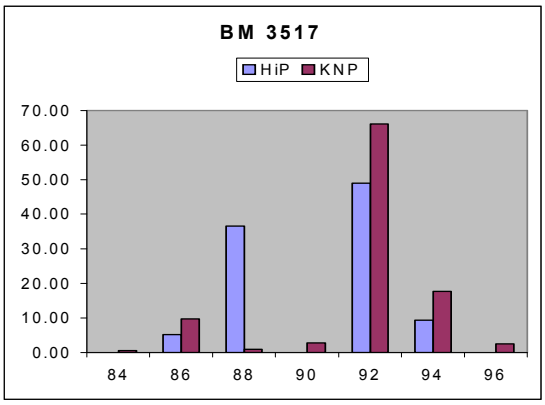

ETH 225
HiP
KNP
100
90
80
70
60
50
40
30
20
10
0
133
135
136
137


BM 719
HiP KNP
60.00
50.00
40.00
30.00
20.00
10.00
0.00
134 136 138 140 142 144 146 148 150 152 154 156 158 160


Inra 128
HiP KNP
100.00
80.00
60.00
40.00
20.00
0.00
166 168 170 174 176 178 182


Inra 006
HiP KNP
70.00
60.00
50.00
40.00
30.00
20.00
10.00
0.00
105 107 109 111 113 115 117 119 127


TGLA 159
HiP KNP
60.00
50.00
40.00
30.00
20.00
10.00
0.00
223 225 227 229 231 233 235 237 239


BM 3517
HiP KNP
70.00
60.00
50.00
40.00
30.00
20.00
10.00
0.00
84 86 88 90 92 94 96

Chapter 5

Concluding remarks

The Cape buffalo (*S. c. caffer),* one of the three subspecies of the African buffalo, is one of the most majestic large mammals found on the continent and has been coined Africa's black gold due to its economical value as a tourist attraction and sought-after hunting trophy. It also has a reputation as a dangerous animal with many a hunter's tales attesting to this. The Cape buffalo comprises up to 35 % of the large herbivore biomass in Africa, and is hence an important role player in ecosystems within Africa's numerous conservancies (Prins, 1996). It is thus not unexpected that a considerable amount of effort is dedicated to devising strategies aimed at its efficient management and conservation. The sequence of events during the last century in the life cycle of the Cape buffalo in Africa and particularly South Africa makes for a special case in the history of this species. Few examples exist where the forces of nature impacted more severely on a wildlife species than when the Rinderpest epidemic almost obliterated Africa's buffalo populations more than a century ago. The numbers of Cape buffalo were severely depleted in many parts in Africa, and in South Africa, it is believed that up to 95 % of the country's buffalo succumbed to the disease (Stevenson-Hamilton, 1957). Ironically, Rinderpest was introduced into Africa through human intervention during wars at the turn of the 19th century (Rossiter, 1994). Nature furthermore perturbed the buffalo numbers through severe droughts during which large numbers of animals died. In KNP alone more than 14000 animals died between 1992 and 1995 as a result of starvation caused by drought (De Vos *et al.*, 2001). More than 25 800 animals were also removed from or culled in KNP between 1967 and 1981, primarily due to population control measures (de Vos *et al*, 1983). In HiP, population control measures resulted in the removal or culling of close to 4000 animals between 1982 and 1994 (Brooks and Macdonald, 1983). The species is also plagued by sub-acute diseases, and is host to a range of pathogens including those that cause bovine tuberculosis, foot-and-mouth-disease and corridor disease (Anderson *et al*., 1979; Bengis *et al*., 1996; Keet *et al*., 1996; Cooper, 1998). Whilst the latter two diseases cause no apparent ill effect to buffalo, BTB systematically worsens body condition resulting in increased susceptibility to a range of

infections, including nematodes (Jolles *et al.*, 2006a). The disease also increases mortality and decreases fecundity in certain age groups among buffalo (Jolles *et al*., 2006b).

Despite their susceptibility to diverse diseases, the species has survived remarkably well through the decades and has subsequently recovered to the extent that the abovementioned population control measures, aimed at avoiding damage to the ecosystem, had to be implemented on several occasions in both parks (de Vos *et al*., 1983; Brooks and Macdonald, 1983). One of the most extraordinary features of the Cape buffalo population of KNP is the fact that it recovered from an estimated 20 animals a century ago (Stevenson-Hamilton, 1957) to 36000 in the early 1970s, despite the population control interventions that started in 1967. Today, following years of population control, drought (especially during the mid 1990's, de Vos *et al*., 2001) and ever-increasing rates of BTB infection, the KNP population presently comprises around 29000 animals (Whyte, 2006, personal communication). Although not as dramatic, the buffalo of HiP also showed a striking increase in numbers. The population increased from less than a hundred animals in 1929 (Brooks and Macdonald, 1983) to approximately 3000 at present (Jolles, 2004).

One of the most important aspects of both population and conservation genetics is assessing the level of genetic variation within a species and among populations of the species (Frankham, 1995). Following this, an integral part of this assessment is to qualify and quantify the effects that demographic, geographic and other factors may exert on the level of variation, which also formed the primary objective of this study. The bottleneck caused by the Rinderpest pandemic and the fact that KNP and HiP are two large and fragmented populations (precluding gene flow between them completely) are in all probability two of the most important factors that may have impacted upon the level of genetic variation within these two populations in recent years.

During the past two decades, several developments have resulted in the availability of molecular tools and markers that can be used for obtaining estimates of genetic parameters that define populations. Software programs were developed in parallel to explore and mine the data generated by these markers. Msats and the D-loop area of mtDNA were chosen for this study since they are two of the most widely used markers for studying the genetics of populations (Simon *et al*., 1994; Beaumont and Bruford, 1999). The former markers are not

only amenable to automated analyses, but their high levels of polymorphism permit the availability of numerous alleles that allow for the assessment of the level of genetic variation within and among populations, as well as the level of differentiation between recently diverged populations (Diez-Tascon *et al*., 2000). Mitochondrial DNA, due to its relatively high mutation and small effective population size, is particularly useful for assessing the degree to which factors impacted upon populations in the more distant past (Hoelzel *et al.*, 1993; Goldsworthy *et al.,* 2000), and in terms of the markers used in this study are best suited for determining the degree to which the rinderpest bottleneck affected the level of genetic variation of the buffalo populations from both KNP and HiP. This is of particular importance since bottlenecks may reduce the level of genetic variation of populations, potentially compromising fitness. In the event that gene flow is restricted in such populations, as is the case with KNP and HiP, drift may erode variation further.

In order to assess the level of genetic variation within and among the buffalo populations from KNP and HiP, a panel of Msats was developed and evaluated for its suitability to meet its intended application. The automated and multiplex approach developed specifically for this study resulted in substantial time and cost savings. The power of resolution of the panel was also such that paternity verification could be carried out highly accurately, whilst the precision with which individual identification could be verified was orders of magnitude higher than what is required for forensic purposes and for traceability studies. In terms of turnaround time, close to 100 individuals could be profiled with all 17 Msat markers in less than two days. The panel has subsequently been recommended for the genetic characterization and paternity verification of Cape buffalo in a diagnostic laboratory, whilst providing random match probabilities in forensic DNA analyses of stock theft and poaching cases.

When genetic indices of variation were assessed for KNP and HiP, Msat and mtDNA data were largely congruent. However, when the level of variation was evaluated and compared between the two parks based on both mtDNA haplotype data and Msat data, the values obtained were strikingly different. With only 4 haplotypes being identified in HiP, it seemed that the park exhibits a significantly reduced variation. This is in stark contrast to what was found for KNP, as the 34 haplotypes identified in KNP suggest that KNP has retained much of its genetic variation. The haplotype diversity for KNP is high (0.92) when compared to

other large indigenous African mammals that experienced bottlenecks in the recent past. These include the common warthog (*Phacochoerus africanus;* Muwanika *et al*., 2003), the Cape mountain zebra (*Equus zebra zebra*; Moodley and Harley, 2005) and the African elephant (*Loxodonta Africana;* Nyakaana *et al*., 2002), which had haplotype diversities of 0.63, 0.56 and 0.85 each.

Several indices based on Msat data indicate that KNP retained significantly higher levels of variation than HiP, suggesting a stable long-term effective population size (Spong *et al.*, 2000) for KNP. It also exhibits a significantly higher level of gene diversity (expected heterozygosity or H_e) than HiP, while the allelic richness and total number of alleles for KNP was also significantly higher than that of HiP. KNP furthermore has a much larger long term N_e than HiP, although it is not known whether the long-term effective population sizes of the two parks prior to the rinderpest differed markedly from the estimates derived in this study. It may well be that HiP had a small effective population size and low levels of genetic diversity before the rinderpest pandemic.

Several factors could have affected the current observed level of variation within the two parks. A moderate to high level of variation in the post-rinderpest population in KNP, coupled with an overestimation of the rinderpest bottleneck and rapid recovery of the population, may explain the observed levels of variation in KNP. The reduced level of genetic variation in HiP may be attributed primarily to drift, and to a smaller, genetically more homogeneous post-rinderpest population. The subsequent removal of relatively large numbers of animals during the past decades for population control may have exacerbated genetic erosion in the park even further. The decline in expected heterozygosity (gene diversity) in HiP over the time period 1988 to 2004 and concomitant signals of population contraction, support the assumption that drift has played a major role in this population's decline. The level of differentiation between the two parks is also primarily the result of drift taking place over many generations. The observed level of divergence found between the two parks in this study is much higher than what has been found on a continental scale (Van Hooft *et al*., 2000), indicating that there are marked differences in the population dynamics at a local versus a continental scale.

From the genetic parameters it is evident that drift alone may not have been the only cause of the observed level of variation, in particular for HiP. The significantly reduced number of alleles for HiP and the fact that the park exhibits a significant deficit of mtDNA haplotypes when compared to KNP, suggest signals of a bottleneck. In their study of African elephant populations, Whitehouse and Harley (2001) attributed the reduced number of alleles for the Addo population, when compared to that of the KNP population, to the bottleneck suffered by the Addo population. The mean number of alleles per locus and long term N_e of the buffalo in HiP is also significantly lower than for KNP, which supports evidence for the remnants of a bottleneck. According to Norton and Ashley (2004), a bottleneck in the more distant past would result in a significantly reduction in the number of alleles, and in the event of increased inbreeding over many generations, would result in a rapid decline in effective population size and concomitant reduction in heterozygosity. HiP however does not currently exhibit a significant positive level of inbreeding, although it exhibits a decline in genetic variation based on Msat data and possibly also in mtDNA haplotype data.

The absence of detectable signals of a known bottleneck in the recent past is not uncommon (Waits *et al*., 2000; Whitehouse and Harley, 2001; Muwanika *et al*., 2003; Van Hooft *et al*., 2003; Norton and Ashley, 2004) and it highlights the fact that several factors come into play when a population recovers from a population crash. The results from this study support the explanation by Van Hooft *et al*. (2003) that immigration from east and north of KNP may have followed the bottleneck, supplementing genetic variation and resulting in the introgression of rare alleles into the park. This, coupled with the documented rapid recovery of buffalo numbers in the park (Brooks and Macdonald, 1983), may have resulted in the population reaching equilibrium rapidly. HiP on the other hand shows signs, although not statistically significant, of remnants of the rinderpest bottleneck. Unlike KNP, immigration of animals into HiP is believed to be unlikely due to human occupation and hunting along its borders, and anti-nagana campaigns (Brooks and Macdonald, 1982). It should be noted that different genetic indices, based on Msat data, might vary with regard to their sensitivity in detecting a bottleneck. Whitehouse and Harley (2001) specifically cautioned that some methods may have limited application, especially in the event that information pertaining to the demographic history of the study population is lacking. They subsequently advocated that measures of allelic diversity should be one of the foremost indices used for detecting population bottlenecks.

When populations recover from bottlenecks, such recoveries are often accompanied by strong signals of population expansion. These signals, which are based on the accumulation of new mutations in the mtDNA D-loop area, were however non-significant for both KNP and HiP. This was however not completely unexpected, since the mutation rate of the D-loop area (Slade *et al*., 1998), would preclude the accumulation of detectable mutations in the short space of time (less than 120 years) since the rinderpest-induced bottleneck. Van Hooft *et al*. (2000) could also only demonstrate significant signals of an expansion event for Cape buffalo after pooling populations from east and southern Africa.

Studying dispersal is of particular importance from an epidemiological point of view, since the spread of disease is a function thereof. On the other hand, dispersal is one of the driving forces behind the dynamics of a population, while it is intrinsically important for maintenance of genetic variation (Arctander *et al*., 1999; Vucetich and Waite, 2003; Berry *et al*., 2004). Apparent behavioural and mating differences between buffalo from KNP and HiP (Jolles, personal communication) are not unique, and have been observed in other African populations (Sinclair 1977; Mloszewski, 1983). Both KNP and HiP exhibit strong and significant signals for male biased dispersal between herds. Female dispersal, although occurring to a lesser degree, could not however be excluded, and was evident from the high haplotype diversity at the herd level. Several haplotypes per herd (average haplotype diversity of 4.3) suggest that mtDNA variation is maintained by gene flow. This is in accordance with previous telemetric studies that showed that female dispersion is not uncommon and may even be long-distance mediated (Halley *et al*., 2002). It however still needs to be determined whether female dispersal fluctuates over time and the extent to which it is affected by oscillating environmental conditions such as the availability of food and water. In order to address this question larger sample sizes are needed that also span over years during which environmental conditions fluctuated.

Strong signals of female phylopatry in both parks are also evident from the fact that several herds in both parks are highly and statistically significantly differentiated from each other. Male biased dispersal is furthermore supported by the fact that genetic distances based on mtDNA is much larger when compared to distances based on Msat data. The strong signals for male biased dispersal in KNP is confounded by the fact that the whole population shows a positive inbreeding coefficient, which may be attributed to the reproductive and social

behaviour of the males, while the presence of null alleles is not excluded. It has been postulated that a kind of "male inbreeding" effect presents itself in the event that males enter herds and mate with females from their natal herds more frequently than random (Chesser, 1991). This is supported by the fact that KNP also exhibits an excess of homozygotes at some Msat loci, while the level of relatedness among males within herds in KNP suggest that mating is not completely random. A large number of herds (21 of the 30 herds sampled) in KNP also exhibited positive inbreeding coefficients, supporting the male inbreeding hypothesis.

Although the level of differentiation between the two parks is high based on mtDNA (F_{ST} = 0.275), no geographical partitioning of haplotypes to either one of the parks could be demonstrated. The level of divergence between haplotypes was furthermore relatively small, while haplotypes were not shared between the two parks. This suggests that KNP and HiP were historically part on a larger panmictic population. At the herd level, the degree to which certain herds are differentiated from each other (in HiP) is congruent with ecological observations that reported the existence of herds that formed tight coherent assemblages, that neither mixed with other herds, nor showed any overlap in home range with that of neighbouring herds (Jolles, 2004). To our knowledge, this study represents the first evidence of significant differentiation at the herd level in KNP and HiP, and is most likely due to the intensive, localised sampling and characterization, which distinguishes it from previous studies (O'Ryan *et al*., 1998; Simonsen *et al*., 1998; Van Hooft *et al*., 2000).

From a conservation point of view, it is vital to understand and uncover the driving forces behind population dynamics in order to devise management strategies aimed at the protection of a species. This study has highlighted the extent to which factors, whether from the environment or stemming from the behaviour of animals in a population, has differentially impacted upon the genetic status of the two largest Cape buffalo populations in South Africa. It also provides baseline information for several genetic indices that are of importance from a genetic management point of view. These indices, of which the level of variation and the effective population size are two of the most important, are indicators of the ability of KNP and HiP to sustain genetic variation and fitness and to respond to environmental change, in the longer term. Traditionally, the census size of a population was used as an indicator of the long-term viability and sustainability of a population or a species.

Van Hooft *et al*. (2002) pointed out however that buffalo census sizes may be misleading as several Cape buffalo populations from elsewhere in Africa have effective population sizes that are larger than their respective census sizes, suggesting that these populations may not sustain genetic variation in the absence of gene flow, in the long term. The maximum long term N_e for HiP of 76 % of its census size suggests that this population is getting relatively close to the point where the sustainability of genetic may be compromised. The fact that the population is already close to its effective population size, suggests that a further reduction in population size (e.g. during natural disasters such as draught or disease epidemics), may have serious consequences. The maximum N_e of KNP on the other hand is only 14% of its census size, which implies that this population will be able to sustain variation, provided that current census sizes are maintained.

From the census sizes of buffalo in KNP and HiP as well as elsewhere in Africa, it is clear that the species is by no means threatened with extinction (Winterbach, 1998). The competition for resources shared by man and buffalo will however undoubtedly become more intense, which may result in a redefining of conservation priorities in future. From a genetic point of view, more emphasis has been placed on the need for separate management and conservation strategies for different conservation units or ecological groupings, in recent years. Harley *et al*. (2005) highlighted the importance of separate management practices for populations that have been separated over long periods of time. Although KNP and HiP need not be considered as separate management units at present, this may change in future, especially in view of the large differences observed in their respective allele frequency distributions and observed levels of private Msat alleles. The occurrence of private alleles in populations supports the existence of a high degree of isolation among populations and drift while ruling out significant levels of gene flow (Randi and Lucchini, 2002; Duran *et al*., 2004). In addition to that, the fact that mtDNA haplotypes are not shared between populations of KNP and HiP, and that the two populations are separated by a relatively large genetic distance, merits considering separate management strategies for the two parks in the near future. However, when considering supplementing genetic variation, especially in HiP, exchanging individuals between KNP and HiP should not necessarily be excluded. Based on Msat data, the F_{ST} value between these two populations indicate a minimal level of historical gene flow of at least one migrant per generation (Mills and Allendorf, 1996), and it has been suggested that individuals can be exchanged between

populations exhibiting these minimal levels of historical gene flow (Forbes and Hogg, 1999). It should be noted that the level of differentiation between KNP and HiP, measured in terms of F_{ST}, is based on calculations that assume mutation-drift equilibrium. While this assumption holds for KNP, it may be violated for HiP, potentially affecting estimated levels of historical gene flow that took place between the two parks. Whitehouse and Harley (2001) advocated the facilitation of widespread gene flow among populations in order to elevate levels of genetic variation among genetically impoverished populations. Having studied the levels of variation among African elephant populations in Addo and KNP, they subsequently suggested the removal of fences between KNP and Mozambique in order to permit gene flow between these populations. This type of intervention would not however be possible for KNP and HiP, due to the differential disease status of buffalo from these two parks.

Another component that form part of a holistic management strategy is the one of disease control, the priorities of which may supersede that of interventions aimed at conserving and sustaining genetic variation. The differential disease status of the two parks for instance already precludes mixing and has indirectly resulted in distinct management strategies, particularly with reference to BTB. It is thus quite evident that a holistic approach towards conservation management may be complex in view of the fact that several components other than genetics need to be considered.

This study has provided detailed baseline information regarding the genetic status of KNP and HiP, and future research should include monitoring levels of genetic variation in especially HiP. This is vital for planning interventions that are aimed at ensuring long-term sustainability of this population. The role that certain factors play in shaping patterns of genetic variation of the buffalo populations of KNP and HiP is still not well understood. It is of particular interest for instance to establish the degree to which male dominance affects levels of genetic variation and genetic drift within the two populations and whether some form of selection prevails among reproducing animals. The hypothesis that the “male inbreeding“-effect contributes to current levels of inbreeding in KNP still needs to be tested. For this purpose markers on the Y-chromosome should be well suited to assessing male biased genetic variation and dispersal. Determining the effective population size of the males will also elucidate the relationship between male effective population size and the rate

at which drift may be taking place in both parks. It would also be of interest to shed light on the factors that affect the decline in genetic variation in especially HiP. Finally, much research is still needed to address issues relating to possible links between disease and genetics. It would be pertinent to establish to what extent disease susceptibility or resistance is genetically based, particularly with regard to BTB.