The Characterization of MHC Class II Genes of the Nile crocodile (*Crocodylus niloticus*): an Investigation of Mechanisms that Shape Genetic Diversity in Natural Populations

Lourie Badenhorst

Supervisors:

Dr C O'Ryan (University of Cape Town)
Dr JM Bishop (University of Stellenbosch)

Thesis Presented for the Degree of MASTER OF SCIENCE

Department of Molecular and Cell Biology
University of Cape Town
February 2008

Abstract

Genes within the Major Histocompatibility Complex (MHC) of vertebrates code for proteins that are involved in antigen recognition and activation of the adaptive immune response. The hallmark of the MHC is the extremely high levels of polymorphism found at loci. A diverse array of mechanisms have been proposed to explain the generation and maintenance of diversity at MHC loci, including the processes of gene conversion, genetic drift and selection; in the presence of many pathogens balancing selection is thought to be the dominant mechanism by which selection operates. Amino acid substitutions within the peptide-binding region (PBR) of MHC genes further supports the hypothesis that positive selection enhances amino acid diversity in the PBR, such that natural selection will favour PBR diversity in natural populations. This study investigated mechanisms that shape genetic diversity of MHC class II genes in a natural population of the Nile crocodile, *Crocodylus niloticus*. Using PCR-cloning-sequencing methodology, allelic diversity at MHC Class II genes was investigated and provides evidence for at least two Class II β gene families in the Nile crocodile. The *Crni-DAB* family is homologous to classical Class II vertebrate genes; high levels of both allelic and amino acid diversity characterise this gene family and a strong signal of balancing selection acts to maintain functional diversity. The second family, *Crni-DBB,* most likely represents a non-classical Class II locus in crocodiles and was characterized by reduced levels of diversity. Analysis suggests that *Crni-DBB* loci have evolved in a divergent manner to those of the *Crni-DAB* as balancing selection was not detected within the putative PBR. Results from this study suggest that duplication followed by a recombination event has most likely led to the formation of two distinct crocodilian Class II β gene families. Secondly, the relative contributions of balancing selection and random genetic drift in the evolution of extant MHC diversity are examined in a natural population of the Nile crocodile. Temporal variation in allele frequencies for MHC and microsatellite loci was assessed in four successive cohorts of crocodiles from the Okavango Delta, Botswana. Results from this study suggest that a combination of short-term neutral forces such as random genetic drift, together with longer-term selection influence variation at Class II loci in the Okavango Nile crocodile. Loci within the MHC of the Nile crocodile appear to be evolving within a dynamic framework of selection, random genetic drift and recombination. This study is the first of its kind to investigate the respective influence of demography and selection on allele frequencies in a natural population of crocodilians.

Acknowledgements

Looking back at the past two years, it becomes very clear that working on this project was not a solitary process, but one that involved many different people, people, without whom I would never have been able to cross the finish line. It is, therefore, a pleasure to thank the following persons for their individual support and contribution to this project.

Firstly, I would like to thank Dr Colleen O'Ryan whose supervision allowed me the freedom to explore the field of molecular genetics and thereby allowing me the opportunity to expand my analytical and molecular skills.

I also want to express my sincere thanks to Dr Jacqueline Bishop, whose supervision extended much further than the lab. Without her personal guidance and motivation I would have been without a saving grace in my many failed and disappointed efforts throughout my masters.

This project, however, would not have been possible without the funding provided by the National Research Foundation of South Africa. Further thanks go to the Earthwatch Institute for providing the funding for a long term research project, investigating the ecology and physiology of the Nile crocodile. I would especially like to thank Dr Alison Leslie, University of Stellenbosch, and Earthwatch volunteers who provided extensive sampling of individual crocodiles from the Okavango River in Botswana. Thanks also to the Le Bonheur Crocodile Farm, Franschoek, South Africa for providing additional individual crocodiles, enabling me to do cDNA analysis and as such strengthening the data used in this study.

Special thanks should also go to Miss René Veikondis, University of Stellenbosch, Tygerberg, in giving up some of her time to perform SSCP on the samples, as well as helping me to get to grips with analysing the data, which provided a quick and sufficient method to assess the variation within the data.

To all my fellow Conservation Geneticists in Lab 405, thank you for your friendship and support. I specially want to thank Nikki le Roex, my fellow lab conspirator who was with me through the high times, the low times, the slow times and especially the very frenzied, nearly giving up times.

Last but not least, I want to thank my family and close friends for having the patience to bear with all my ups and downs these last two years, always being there for me and without whose support and encouragement I would not have been writing these acknowledgements today.

Animal horoscopes

Table of Contents

Chapter 1:

General Introduction

The Major Histocompatibility Complex (MHC) is a gene-dense region found in almost all jawed vertebrates examined thus far (Kelley et al. 2005). The MHC contains a high percentage of classical genes that form one of the most conserved parts of the adaptive immune system (Bryja et al. 2006; Chen et al. 2006; Hedrick 2003; Miller et al. 2005; Piertney and Oliver 2006; Sommer 2005). The primary role of these genes is to process foreign proteins and to present them to specialist immune cells in order to initiate an adaptive immune response (Piertney and Oliver 2006). In addition, the MHC also contains non-classical genes that are thought to express proteins that modulate peptide loading in the lysosomal compartment of the vertebrate cell; nevertheless, the functional status of many of these non-classical genes is mostly unresolved (Klein and O'hUigin 1994). Non-classical genes are often pseudogenes or have a highly divergent function compared to that of classical genes. Non-classical genes also exhibit lower levels of polymorphism, erratic tissue expression and a considerable amount of sequence dissimilarity when compared to classical genes (Klein and O'hUigin 1994).

Classical MHC genes play a vital role in controlling the activation of all immunological effectors such as helper T-cells, cytotoxic T-lymphocytes, macrophages and B-cells (Penn and Potts 1999). They encode two classes of cell-surface glycoproteins: Class I and Class II (Kelley et al. 2005; Miller et al. 2005; Piertney and Oliver 2006); these two classes differ primarily in where the receptors are expressed. MHC Class I molecules have three globular α domains, which associate in a non-covalent way with a β2 microglobulin (Nei et al. 1997). Class I molecules are expressed on the surface of almost all nucleated cells and present endogenously derived peptides, both self and nonself, to CD8+ cytotoxic T-cells (Miller et al. 2005; Piertney and Oliver

2006). Each T-cell has its own unique receptor that binds to the MHC antigen complex (Figure 1.1) (Penn and Potts 1999).

MHC Class II molecules are heterodimers formed by two transmembrane proteins, an α and a β chain. Class II molecules are only expressed on antigen-presenting cells (B-cells and macrophages) where they present processed exogenous antigens to CD4+ T-helper cells (Miller et al. 2005; Piertney and Oliver 2006). B-cells express cell-surface antibodies that bind to extracellular antigens and the peptide fragments of these endocytosed antigens are presented to T-helper cells via MHC Class II molecules (Penn and Potts 1999). If the antigen presented to the T-helper cell is recognized it activates the B-cell to proliferate and secrete antibodies that 'label' the foreign antigens for macrophages. Extracellular pathogens and parasites are phagocytized by macrophages and this leads to the stimulation and activation of lymphocytes and other immune cells to respond to the pathogen (Figure 1.1) (Penn and Potts 1999).

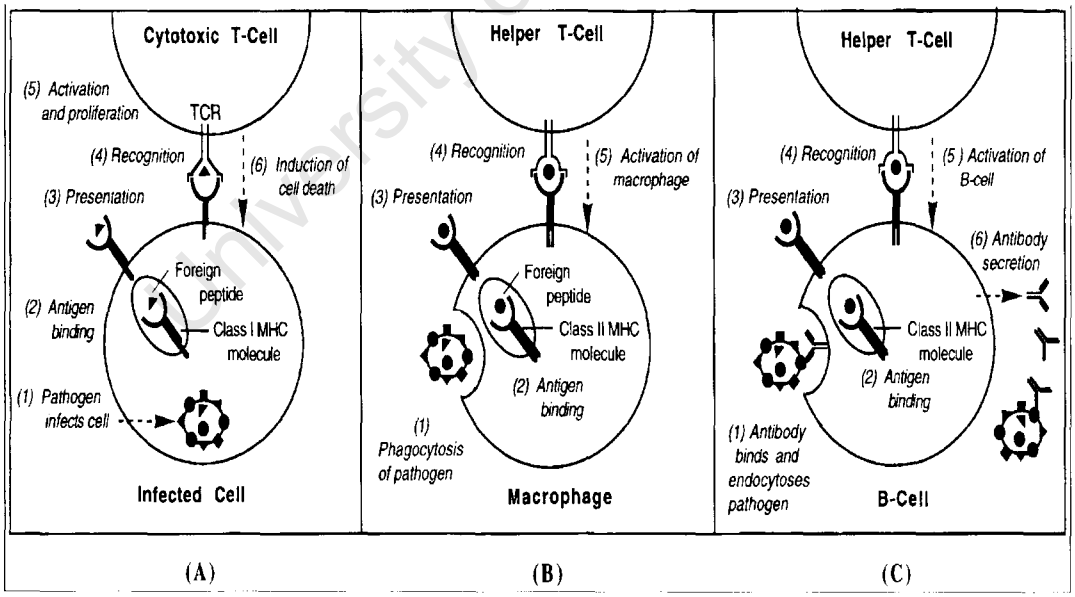


Figure 1.1: The MHC controls the activation of all specific immunological effectors. (A) MHC Class I molecules present endogenously derived peptides to cytotoxic T-cell. (B) and (C) MHC Class II molecules are only expressed on antigen-presenting cells and present processed exogenous antigens to T-helper cells. (Taken from Penn and Potts 1999)

MHC molecules consist of an immunoglobulin 'stalk' that anchors the molecule to the cell surface, and a 'molecular pocket', known as the peptide-binding region (PBR), that is responsible for antigen recognition and binding. Certain elements of the MHC are conserved across species; this includes the overall function and structure of the peptide presenting molecules (Madden 1995). The genomic organization of the MHC also appears to be conserved in many species (Kelley et al. 2005). The most advantageous feature of the MHC is the extreme diversity that is observed at expressed loci. MHC genes are the most polymorphic genes currently known in the genome of vertebrates, with both multiple loci and many alleles at each locus (Bryja et al. 2006; Chen et al. 2006; Meyer-Lucht and Sommer 2005; Piertney and Oliver 2006). Over 260 genes have been identified within the human MHC, where gene duplication has lead to paralogous regions on other chromosomes (Kelley et al. 2005). MHC variability, especially in the PBR is linked with T-lymphocyte receptor diversity, which in turn determines the degree of parasite and disease resistance of an organism. A high level of PBR sequence variation both within individuals and populations allow MHC loci to recognize and process a large number of pathogenic antigens; thus higher MHC diversity may contribute to increased population and individual fitness (Avise 1994; Edwards and Potts, 1996; Harf and Sommer 2005).

Both recombination between loci (interlocus recombination) and recombination among alleles at a locus (intralocus or interallelic recombination) have been suggested as mechanisms involved in the diversification of both MHC Class I and II genes (Hughes and Yeager 1998). Intralocus recombination involves the exchange of long or short stretches of DNA sequence between two alleles during crossing-over, and thus contributes to levels of diversity by increasing the number of alleles in a population (Hughes and Yeager 1998). Interlocus recombination can lead to the homogenization of loci within species, known as gene conversion, but has also been shown to explain highly divergent sequence variants in species, for example in the three-spined stickleback (*Gasterosteus aculeatus*) (Hughes and Yeager 1998; Reusch et al. 2004; Wittzell et al. 1999).

MHC Class I genes differ in number and organization between species, and thus generally show no orthology among orders as a result of relatively frequent duplication events and the elimination of duplicate loci by unequal crossing-over (Hughes and Nei 1989, Kelley et al. 2005; Yeager et al. 1997). The alleles of MHC Class I genes are therefore generally too short-lived to be informative about relationships among populations (Kasahara et al. 1996). MHC Class I genes also accumulate mutations more rapidly than Class II genes and as a result it is difficult to use Class I molecules in a population genetic context (Hughes and Nei 1989). It has been suggested that both of these processes also occur in Class II genes but at a much slower rate than has been reported for Class I genes (Hughes and Nei 1990).

Selection and Diversity within the MHC

Loci within the MHC are probably the best known example of a gene complex that is evolving dynamically under the processes of selection, random genetic drift and recombination. A diverse array of mechanisms has been suggested to explain the generation and maintenance of high levels of polymorphism at MHC loci. Within the genome, new genetic variants are usually generated through single point mutations; however within the MHC a considerable amount of polymorphism is also generated via the processes of gene duplication, recombination, deletion and insertion (Bryja et al. 2006, Langefors et al. 2001). In mammals, the most widely accepted model of MHC evolution is the '*birth and death*' model (Nei et al. 1997). In this model duplicated genes at each Class II locus may be generated independently; after duplication, some of these loci may diverge functionally whilst others may become pseudogenes (via loss of functionality) or deleted from the genome (Bryja et al. 2006; Hughes and Nei 1990; Nei et al. 1997). The loss and gain of loci also generates MHC diversity among species (Kelley et al. 2005). As a result of the loss of a locus, other loci may compensate and assume an additional function; for example, in the blind mole rats of Israel (*Spalax ehrenbergi*) the Class II *DR* locus has been deleted, but the multiple α- and β-genes that make up the *DP* loci have assumed the function of the lost *DR* loci (Nizetic et al. 1987).

Mechanisms that are involved in the maintenance of polymorphisms at MHC Class II loci include a variety of selective pressures operating at different levels (Langefors et al. 2001; Penn and Potts 1999; Potts et al. 1991). These selection pressures include a) pathogen-driven selection, b) maternal-fetal interactions and c) sexual selection. Two possible mechanisms of pathogen-driven selection have been suggested: (1). The '*heterozygous advantage hypothesis'* (Doherty and Zinkernagel 1975) suggests that the immune system of an individual may be able to detect a broader range of pathogens as a result of heterozygosity at MHC loci. Only a few studies have linked heterozygous advantage to pathogen resistance e.g. in domestic chickens involving Rous sarcoma virus (Senseney et al. 2000) and in humans involving the hepatitis B virus (Thursz et al. 1995). In humans, heterozygosity at MHC Class I loci has also been linked to a slower progression of AIDS among individuals infected with HIV compared to individuals who are homozygous for one or more loci (Carrington et al. 1999). (2). *'Negative frequency-dependent selection'* (Takahata and Nei 1990) has been proposed to occur in situations where rare alleles confer a temporary selective advantage and thus increase in frequency (Zelano and Edwards 2002). Rare alleles have an advantage over common alleles because pathogens adapt to cope with alleles that are more common. As a result of this, a decrease in the relative fitness of the common alleles occurs, which provides a selective advantage to new, rare alleles (Harf and Sommer 2005). Studies have shown a connection between certain rare alleles and disease resistance in cases of hepatitis B and malaria in humans (Hill et al. 1991; Jeffery and Bangham 2000; Langefors et al. 2001; Thursz et al. 1995; Von Schantz et al. 1996).

Studies on maternal-fetal interactions have shown that a specific immune response, as a result of differences between mother and fetus MHC loci, is necessary for proper implantation and fetal growth (Hedrick 1999). Maternal-fetal interactions can therefore enhance the maintenance of polymorphism at the MHC, as it is a selective advantage to have many alleles at a single locus (Hedrick and Thomson 1988). The importance of this mechanism is, however, limited to mammals and other live-bearing organisms as most fish,

amphibians, reptiles and birds have different maternal-fetal interactions (Hedrick 1999).

Sexual selection involves reproductive behaviour such as dis-assortative mating, whereby individuals prefer mates with MHC genotypes that are different to their own (Edwards & Potts 1996; Ekblom et al. 2004; Miller and Lambert 2004), and thus ensure that their offspring are more likely to be heterozygous at MHC loci and thereby maximally resistant to disease (Milinski et al. 2005; Penn et al 2002; Potts and Wakeland 1993; Ziegler et al. 2005). It is interesting to note that while a number of studies support MHC-based mate preference, several studies have also shown MHC-independent mate choice, e.g. in Soay sheep (*Ovis aries*) (Paterson and Pemberton 1997), great reed warblers (*Acrocephalus arundinaceus*) (Westerdahl 2004) and great snipes (*Gallinago media*) (Ekblom et al. 2004). Studies supporting MHC-related mate choice behaviour have shown that mate choice is determined by a number of different criteria associated with the MHC; these include genotypic dissimilarity, the presence of specific alleles and levels of overall diversity (Piertney and Oliver 2006). Olfaction in mammals has been suggested as a mechanism by which an individual can assess the MHC type of another individual (Penn and Potts 1999), whereas specific MHC alleles may be associated with advertisement via extravagant secondary sexual characters, e.g. studies have shown that MHC Class I and II β haplotypes of male ring-necked pheasants (*Phasianus colchicus*) correlate with male viability and spur length (von Schantz et al. 1989). Females prefer to mate with males with longer spurs and thus improve chick survival rate. Mate choice has also been shown to be directed towards maximising overall immunocompetence across multiple loci, e.g. in three-spined sticklebacks (*Gasterosteus aculeatus*) females choose males based on the number of alleles across multiple loci, known as 'allele counting', thus preferring males that display an optimum number of alleles rather than merely MHC-dissimilar males (Reusch et al. 2001).

Population structure and population history can also influence patterns of MHC polymorphism observed within species. Ohta (1999) has shown that in a

local population where selection is strong, selection can act directly at the MHC e.g. as a result of maternal-fetal interactions. In the case where the selection coefficient acting on an allele differs between populations, for example through varying pathogen load and disease exposure, the pattern of change in allele frequencies and diversity is also then dependent on population size. For example studies on the Malagasy giant jumping rat (*Hypogeomys antimena*) suggest that the combination of a monogamous mating system and recent declines in population size has resulted in low MHC variation (Sommer et al. 2002). Whereas studies on buffalo (*Syncerus caffer caffer*) have shown high MHC diversity in populations despite a historical collapse from epidemics of rinderpest disease and this high level of diversity might be the result of the high dispersive capability of buffalos (Wenink et al. 1998).

In summary, diversity at MHC loci is most likely the composite effect of several different modes of selection that are not mutually exclusive, but instead act in concert to maintain MHC polymorphisms (Bryja et al. 2006). As a rule, purifying selection acts on loci to remove variation by favouring a single genotype with the greatest fitness at that time. But in the presence of many pathogens balancing selection is thought to be the mechanism by which selection operates. The precise nature of these processes, however, remains a much–debated area of research (Ekblom et al. 2007; Kim et al. 1999; Piertney and Oliver 2006; Sommer 2005).

The natural history of the Nile crocodile Crocodylus niloticus (Laurenti 1768)

The Nile crocodile is one of three crocodile species that still inhabits the freshwater ecosystems of the African continent. It is found throughout rivers and wetlands of central and southern regions of the continent. The species' range is sub-Saharan, extending from West Africa to Egypt in the east and southwards into the northern reaches of South Africa; Nile crocodiles are absent from the arid regions of the extreme south and southwest (Figure 1.2). The species has a wide selection of habitat preferences, ranging from wetland

habitats, freshwater rivers to brackish waters, estuarine systems and mangrove swamps. Their generalist habitat requirements are reflected in their distribution, and current estimates for the species suggest that there are between 250,000 to 500,000 individuals in the wild (www.cites.org).

Figure 1.2: Map of Africa showing the extant range of the Nile crocodile *C. niloticus* (map available at http://www.flmnh.ufl.edu/cnhc/csp_cnil.htm#dist).

Even though crocodilians are keystone species within the systems they occupy, and have been dominant apex predators for millions of years, they are still a much-understudied creature (Ross 1998). The biology and ecology

of populations of the Nile crocodile are especially poorly understood (Ross 1998). Throughout their range, populations of the Nile crocodile have been heavily exploited as hunting trophies and for their valuable skins, whilst current threats include increased targeting of crocodiles in the bush meat trade throughout west and central Africa (www.traffic.org) and their eradication from areas due to human/croc conflict.

The recent history of the Nile crocodile population of the Okavango Delta in Botswana exemplifies this history of over-exploitation, where intensive hunting for hides and trophies during the period 1957–1969 resulted in extreme population decline; an estimated 50,000 adult animals were removed from the system (Pooley 1982). As a result of population decline throughout its range, the Nile crocodile was added to CITES (Convention on International Trade in Endangered Species) Appendix I in 1973, banning all trade in the species. Today, the Nile crocodile continues to be of great economic importance as its hide and flesh are still highly valued both abroad and in southern Africa (Flint et al. 2000). The high value placed on crocodile skins led to renewed pressure to re-introduce a hunting quota. This resulted in the re-introduction of a quota scheme by 1985, allowing a limited number of wild crocodile skins to be exported yearly (500 animals a year) (Ross 1998); within the Okavango, repeated attempts to resume hunting quotas in the 1980's continually failed as a result of low hunter success and it is highly likely that crocodile populations may be very sensitive to even small perturbations (A. Leslie pers. comm.). The IUCN Red List maintains the status of the Nile crocodile as 'Low risk' (may be threatened in some parts of range) (www.iucnredlist.org) and the IUCN Crocodile Specialist Group report that although the status of the Nile crocodile is relatively secure and numbers are abundant in southern and eastern Africa, the species is greatly depleted in west Africa. Of the 20 African countries that monitor the status of *C. niloticus*, six of these consider the species to be severely depleted. CITES down-listed the species to Appendix II in 1992, allowing legal trade in captive-bred crocodile products. Current quotas allow 1600 hunting trophies per annum, with additional trade in harvested animals from ranching operations (www.cites.org). Commercial utilization is now widespread through the successful introduction of crocodile

farming operations, and along with these farms a small number of successful management programmes have been established.

Research Aims

MHC genes play a critical role in the recognition of parasites and pathogens by the immune system and may also be associated with additional quantitative traits that are linked to the fitness and behaviour of individuals in natural populations, for example, mate choice and sexual behaviour (Piertney and Oliver 2006). By investigating MHC diversity of a natural population, we can increase our understanding of how behavioural and ecological factors can 1) influence the dynamics of MHC evolution, and 2) how MHC diversity affects both individual fitness as well as population level processes. The characterization of MHC diversity, and the investigation of mechanisms involved in the generation and maintenance of this diversity, is a valuable indicator of adaptive variation in non-model organisms. Characterization of MHC diversity can therefore contribute to the welfare and longevity of natural populations, especially reduced, endangered populations. Studies of MHC diversity have been facilitated by the recent development of rapid, high throughput screening techniques, such as single-stranded conformation polymorphism (SSCP), which allow large numbers of individuals to be screened for MHC variation relatively quickly and economically (Potts 1996).

This study utilizes the Nile crocodile population of the Okavango Delta, Botswana as a model to investigate MHC diversity in a natural population. This thesis is divided into two main sections. Chapter 2 describes exon 2 diversity at the first MHC Class II β genes to be isolated in *Crocodylus niloticus*. It provides evidence for a possible mode of evolution involved in diversifying MHC genes in the Nile crocodile and investigates whether balancing selection plays a role in generating diversity within the PBR of Class II β loci. Chapter 3 investigates aspects of the 'ecology' of Nile crocodile Class II loci in the Okavango Delta. The respective roles of selection and genetic drift in determining changes in neutral and adaptive genetic variation over time are investigated. Newly hatched Nile crocodiles were collected over

four years and genotyped at both Class II MHC loci and neutral microsatellite loci. By comparing the respective changes in allele frequencies between years for both markers we can infer whether MHC diversity in this population is predominantly shaped by year-on-year selection or by neutral forces such as random genetic drift.

Chapter 2:

Characterization of Class II-like Major Histocompatibility Complex loci in *Crocodylus niloticus*

Introduction

Consistently high levels of MHC polymorphism have been demonstrated across a broad range of vertebrates studied to date. This gene region is characterized by multiple loci, a large number of alleles at each locus and high nucleotide sequence variation among alleles (Brya et al. 2006; Nei et al. 1997; Schierup et al. 2001). Several different mechanisms have been suggested to explain the generation of this extreme polymorphism. Two models explaining the evolution of the MHC have been proposed. The original model of MHC evolution, that of '*concerted evolution*' (Nei et al. 1997), proposed that interlocus recombination and gene conversion together generate new duplicate genes, but also result in the deletion of existing heterogeneous genes. Under different models of selection, recombination can therefore lead to divergent results. Under a neutral model interlocus recombination generates new polymorphisms; this can result in the homogenization of duplicated loci within a species, making it difficult to assign alleles to particular loci (Miller and Lambert 2004; Nei et al.1997; Reusch and Langefors 2005; Wittzell et al. 1999). For example in the pheasant *Phasianus colchicus* exon 3 alleles of the *DAB* gene are not locus-specific but are instead species-specific, and this lack of orthology suggests that the loci have undergone recurring interlocus genetic exchange since the split of the evolutionary lineages leading to pheasants (*Phasianus spp.*) and domestic chickens (*Gallus gallus domesticus)* (Wittzell et al. 1999). Under a model of positive selection the opposite may be observed (Hughes 2000). For example in the three-spined stickleback *Gasterosteus aculeatus*, gene diversification appears to be driven by interlocus recombination; this is demonstrated by the fact that alleles of the same locus do not form a monophylectic cluster but

rather cluster together with alleles from different loci (Reusch and Langefors 2005). Any monophyletic signal is erased when recombination is coupled with positive selection. Locus diversification by recombination also results in a 'patch-work' pattern of diversity between alleles, where an assortment of various nucleotide sequence motifs among MHC genes can be similar (Yuhki and O'Brien 1990). Duplicated genes, however, gradually diverge over time via differential rates of mutation and selection (Nei et al. 1997); they can also become non-functional due to deleterious mutations, be deleted from the genome or have their function modified, or become non-classical MHC loci (Hughes and Nei 1989). Consequently Nei et al. (1997) also proposed a '*birth and death*' model of evolution where the end-result is a diverse group of MHC genes that represent both functional loci and a substantial number of pseudogenes. The process of *concerted evolution* is generally accepted for other multigene families; for example this mode of evolution has been used to explain why the intergenic nucleotide sequences of a large number of tandemly repeated genes that encode ribosomal RNAs (rRNAs) in *Xenopus*, are more similar within a species than between two closely related species (Brown et al. 1972). In the case of MHC genes, however, *concerted evolution* has remained a controversial topic, where the main issue of debate is the apparent contradiction between the extreme levels of polymorphism found in MHC genes and the homogenizing effect of *concerted evolution* that is predicted under a neutral model of evolution (Wittzell et al. 1999). The '*birth and death*' process of evolution is therefore thought to be a much more prominent force controlling MHC gene diversification, but because the gene complex has both multiple loci and multiple alleles evolutionary factors that influence individual loci may differ at each divergent locus and across species.

The combination of repeated gene duplication, loss of duplicated genes in different lineages, and the difference in the rate of gene turnover at the MHC has resulted in the arrangement of Class II genes to vary extensively between mammalian and non-mammalian vertebrates (Miller et al. 2005). For example the classical Class II genes of marsupials, monotremes and non-mammalian vertebrates' are not orthologous to the MHC gene clusters of eutherian mammals (Belov et al. 2004, Takahashi et al. 2000). Furthermore, the Class II

genes of birds do not form orthologous groups and there is substantial variation in MHC gene organisation and gene number among different bird species (Edwards et al. 1999). For example, in contrast to the relatively simple gene arrangement of the domestic chicken *Gallus gallus domesticus*, songbirds have evolved multiple duplicated genes for each class of loci (Kaufman et al. 1999, Miller and Lambert 2004, Westerdahl et al. 2000). Although MHC genes have been characterized across an extensive array of mammals, birds, amphibians and fish, information about their organization and levels of polymorphism in non-avian reptiles is lacking (Kelley et al. 2005). Both mammals and birds are descended from ancient reptiles and thus the characterization of reptile MHC may provide insight into the ancestral gene arrangement from which all mammalian and bird sequences have evolved (Miller et al. 2005). Even though reptiles perform a critical role in the ecological systems they occupy they still generally remain a much-understudied group.

Extant 'reptiles' are polyphyletic in their origin and can be spilt into three lineages: Testudines (turtles), Archosauria (birds and crocodilians) and Lepidosauria (lizards, snakes and tuatara; Rest et al. 2003). To date, there are no reports of MHC gene sequences from the Testudines lineage. Miller et al. (2005) characterized the first MHC Class II β cDNA sequences from a non-avian reptile, the tuatara (*Sphenodon punctatus*), where three full-length β sequences were isolated from a tuatara cDNA library using a probe designed from aligned bird, reptile, fish, amphibian and mammalian MHC Class II β sequences. Four additional sequences were isolated by reverse transcriptase polymerase chain reaction (RT-PCR) using primers designed from sequences isolated from the cDNA library. From these sequences, Miller et al. (2005) concluded that Class II β alleles in tuatara can be separated into two gene families, where one may be a non-classical locus.

Using degenerate primers, a number of partial Class II exon sequences have been amplified from crocodilians. Edwards et al. (1995) successfully amplified polymorphic exon 2 segments from passerine bird species and an alligator, using primers designed from an alignment of two human Class II amino acid

sequences and two chicken Class II genes. The primers bind between the polymorphic subdomains of the second exon of the chicken Class II β gene and thus target conserved regions that border exon 2 and its constitutive peptide binding region (PBR). The sequences amplified from the five passerines and an alligator (*Alligator mississippiensis*) indicated a pattern of nucleotide substitution within codons of the PBR of archosaurian Class II β genes that is similar to the selective forces operating in mammalian taxa.

For crocodilians (*Crocodylidae*, *Alligatoridae* and *Gavialidae*), only a few caiman and alligator Class II exon sequences have been reported to date (the majority of these sequences have been deposited in GenBank but have not been reported in published literature). Since birds are the sister lineage to crocodilians, research on bird MHC gene structure, function and order, can potentially inform research on extant crocodilians. The structure of the domestic chicken MHC is well understood and consists of two independent MHC gene clusters: the serologically defined MHC *B-F/B-L* region, and the *Rfp-Y* system (Guillemot et al. 1988; Kaufman et al. 1999; Miller et al. 1994). Both clusters contain Class I and Class II β genes. For the *Rfp-Y* region, at least one MHC Class I locus is expressed and may be active in the immune system of chickens (Hunt et al. 2006). It is not clear, however, whether MHC Class II *Rfp-Y* genes are functional in chickens and these may represent Non-classical loci as they have low polymorphism, minor sequence divergence as well as poor expression in lymphoid tissues (Kaufman and Salomonsen 1997; Strand et al. 2007). The *B-F/B-L* region is located within part of the *B* complex on micro-chromosome 16 (Shiina et al. 2004). The region is compact and simple; it is only 92kB in size and contains only 19 genes. The central region of the complex, extending from Class II β to Class I, is especially compact with a complete lack of repetitive elements. The MHC Class II genes in chickens together form αβ heterodimers, however a non-polymorphic classical Class II α gene has been located outside of the *B* complex (Kaufman and Salomonsen, 1997; Kaufman et al. 1999), thus there is essentially only one Class II molecule (Class II β) found at high levels of expression in common chicken haplotypes. Some genes that are present in the MHC of typical mammals are found within the chicken MHC (e.g. Class I, Class II β-chain,

TAP) but many genes are absent (e.g. Class II α-chain, LMP). Within the sequenced region of the chicken MHC, there are also genes that are not found in the MHC of typical mammals, such as *B-G* genes and *C*-type animal lectin genes (Kaufman 2000).

Among MHC Class II β genes high levels of polymorphism are generally found within the PBR (or antigen binding site, ABS) of exon 2, particularly at the residues that are involved in the binding and presentation of peptides (Miller and Lambert 2004; Schaschl et al. 2004). While the overall function of the MHC along with its structural components is highly conserved, the PBR region contains highly variable 'molecular pockets', which influence the binding and presentation of antigenic peptides (Brown et al. 1993). According to neutral theory of evolution (Kimura 1968), the rate of non-synonymous nucleotide substitution (d_N) within a gene region is predicted to be lower ($d_S > d_N$) than the rate of synonymous substitutions (d_S) under neutrality (Hughes and Nei 1989; Sommer 2005). The reason for this is that non-synonymous substitutions change the amino acid composition of a sequence and are therefore likely to be detrimental. Several studies have shown that the PBR region displays a greater frequency of non-synonymous substitutions than synonymous substitutions ($d_N > d_S$). This feature of the PBR cannot be explained by a higher mutation rate that is specific to this region (Altizer et al. 2003; Bernatchez and Landry 2003; Brown et al. 1993; Froeschke & Sommer 2005; Harf and Sommer 2005; Meyer-Lucht and Sommer 2005; Milinski et al. 2005; Schierup *et al.* 2001; Sommer 2005; Zelano and Edwards 2002). Multiple lines of evidence have shown that the high level of polymorphism found within the MHC is maintained by some degree of natural selection, specifically strong balancing selection (Ekblom et al. 2007; Kim et al. 1999; Piertney and Oliver 2006; Sommer 2005). Balancing selection results in the maintenance of a larger number of alleles in a population, and also plays a role in the maintenance of allelic diversity over long periods of time that often pre-date speciation events (known as trans-species evolution of polymorphism) (Hughes and Yeager 1998; Sommer 2005). Two possible parasite-driven mechanisms of selection have been suggested to be important for retaining high levels of genetic variability (Sommer 2005): the

'*heterozygous advantage hypothesis*' (Doherty and Zinkernagel 1975) and the '*negative frequency-dependent selection hypothesis*' (Takahata and Nei 1990). All these hypotheses are in no way mutually exclusive and it is most likely that several different modes of selection act simultaneously to maintain MHC variation (Ekblom et al. 2007).

Key features of MHC evolution are gene duplication, interlocus recombination, deletion, point mutations and balancing selection (Ahmed et al. 2007). In birds, there is evidence that gene diversification is generated and maintained by numerous processes including gene duplication and deletion, gene conversion as well as sexual and pathogen-driven selection (Edwards et al. 1999; Hunt et al. 1994; Miller and Lambert 2004; Wittzell et al. 1999). The relative rate and importance of each of these mechanisms can nevertheless vary among species.

In this study, we analysed MHC exon 2 diversity of Class II β genes in the Nile crocodile, *Crocodylus niloticus*. The nucleotide and putative amino acid sequences of the exon provide information about crocodilian peptide-binding domains and functional motifs that are essential for gene function and regulation. The objectives of this study were (i) to describe exon 2 diversity of MHC Class II loci in the Nile crocodile and compare exon diversity within this species to other vertebrate species, (ii) to reconstruct the evolutionary relationships of MHC Class II loci in the Nile crocodile, and (iii) to determine whether patterns of MHC variation in the Nile crocodile are consistent with the effects of balancing selection.

Methods

Sample collection

Nile crocodile samples analysed in this chapter represent individuals from the Okavango Delta, Botswana and the Limpopo River, South Africa. Blood and scute tissue samples were collected from individual crocodiles during the period February 2002 to December 2006. Samples were stored in 96% ethanol at 4°C until extraction of total genomic DNA. Forty-five individuals

were randomly chosen for the initial screening of Class II exons. Whole blood samples for RNA extraction were collected from five individual crocodiles at Le Bonheur Crocodile Farm, Franschoek, South Africa; these animals originate from the Limpopo River, Mpumalanga Province, South Africa. Approximately 5ml of blood was collected from the caudal vein in the tail of each individual (Gorzula et al. 1976) and stored in RNAlater Buffer (Ambion) at 4°C.

Genomic DNA extraction

Total genomic DNA was extracted from whole blood or scute tissue samples. DNA was isolated from blood samples using the DNAeasy Blood & Tissue Kit (QIAGEN©) following the manufacturer's protocol. For extraction from scute samples, a piece of tissue measuring approximately 0.5cm^3 was cut into small pieces and incubated in 1ml 200mM TRIS-HCl (pH8) buffer and 390 units of collagenase enzyme overnight at 37ºC. The samples were then centrifuged at 4000rpm for 5min and the supernatant was discarded. The pellets were re-suspended in 500µl lysis buffer (10mM TRIS-HCL, 50mM NaCl, 10mM EDTA, 0.5% SDS) and 20µl proteinase K (stock 100mg/ml) and incubated overnight at 37ºC. After overnight digestion, DNA was extracted using a standard salt/isoamyl alcohol extraction: ½ volume 5M NaCl was added to the overnight digestion mixture and agitated for 20min on a shaker at room temperature. One volume chloroform/isoamyl alcohol (24:1) was then added and the sample shaken for another 10min. Samples were then centrifuged at 8000rpm for 10min; the supernatant was placed in a new 2ml Eppendorf tube to which 600µl isopropanol was then added. Samples were stored at -70ºC for 1 hour to precipitate the DNA. After precipitation samples were centrifuged at 12000rpm for 10min; the DNA pellet was then washed twice with 1ml 70% ethanol and left to air-dry. The pellet was then re-suspended in 50µl Tris-EDTA buffer.

RNA extraction and cDNA synthesis

RT-PCR was used to determine whether transcribed gene products from MHC Class II loci were present in blood samples of the Nile crocodile. Total

RNA was extracted from the RNAlater-preserved whole blood samples using a mirVana™ miRNA Isolation Kit (Applied Biosystems) as per the manufacturer's instructions. First-strand cDNA was then synthesized from up to 1µg of total RNA using ImProm-II™ Reverse Transcriptase (Promega©) and a poly dT_{20} primer, following the manufacturer's instructions.

PCR amplification

The PBR of the MHC is highly polymorphic, while the flanking regions are generally highly conserved. As a result conserved primers targeting the second exon of Class II loci (PCR product size ~200 base pairs) were designed from an alignment of *Caiman crocodilus* (Caiman crocodile; GenBank Accession No. AF256651.1, AF277661.1, AF256652.1), *Alligator sinensis* (Chinese alligator; AY491430.1, AY491422.1, AY491425.1), *Alligator mississippiensis* (American alligator; U24404.1, U24402.1, U24403.1), *Eumeces chinensis chinensis* (Chinese skink; AY772950.1, AY772947.1), *Spheniscus humboldti* (Humboldt penguin; AB154395.1, AB154394.1, AB154393.1) and *Bos indicus* (domestic cattle; CAB52178.1) (see Appendix 1.1 for alignment).

Table 2.1: Details of the primer sequences used for amplification of exons within MHC Class II loci in the Nile crocodile, *Crocodylus niloticus*. Observed polymerase chain reaction (PCR) product size and optimal annealing temperature (Ta) are shown.

	Primer sequence (5`-3`)	PCR Product Size (bp)	Ta (ºC)
Croc-MHCf	CAA CGG CAC CCA GCG CGT G	200	60-65
Croc-MHCr	CAC CCC GTA GTT GTG TCG GC		

Initial optimisation of PCR conditions for genomic DNA was performed in 20µl reaction volumes containing 0.5pmol/µl forward and reverse primers, 0.2mM dNTPs, 1.5mM $MgCl_2$, 0.025U GoTaq® Flexi DNA polymerase (Promega©), 1X Colourless GoTaq® Flexi Buffer and 10-50ng DNA. PCR conditions were

optimized with respect to $MgCl_2$ concentration, primer concentration and annealing temperature and were performed on a PTC-100™ PCR Thermocycler (MJ Research) with the following standard cycling conditions: (step1) 94ºC for 2min; (step2) 94ºC for 30sec, 65ºC for 30sec, 72ºC for 30sec, for 35 cycles; (step3) 72ºC for 7min. To prevent evaporation during thermal cycling each reaction was overlaid with a drop of mineral oil. PCR products were visualized on 2% (w/v) agarose gels. The correct sized product was excised from the gel and purified using the Wizard® SV Gel and PCR Clean-up System (Promega©) as per the manufacturer's instructions.

GoTaq® Flexi DNA polymerase (Promega©) enzyme is a modified form of *Taq* DNA polymerase that lacks 5`- 3` exonuclease activity. In contrast, Expand High Fidelity PCR System (Roche) is a thermostable DNA polymerase with proof-reading activity, which ensures the high fidelity and specificity of DNA replication needed when characterising a gene region. Expand High Fidelity DNA polymerase was used to amplify the Class II loci of the five cDNA samples and to verify the results from two genomic DNA samples. PCR reactions were performed in 20µl reaction volumes containing 0.5pmol/µl forward and reverse primers, 10-50ng DNA, 0.2mM dNTPs, 1.0mM $MgCl_2$, 0.045U Expand High Fidelity DNA polymerase (Roche) and 1X Expand High Fidelity Buffer. PCR conditions were optimized with respect to $MgCl_2$ concentration, primer concentration and annealing temperature as before. The reaction was overlaid with mineral oil and the PCR was performed on a PTC-100™ PCR Thermocycler (MJ Research), following the same cycling conditions detailed above.

MHC cloning and sequencing

Purified PCR products were ligated into the pGEM®-T Easy Vector (Promega©) following the protocol outlined in the pGEM®-T Easy Vector handbook. Vector-ligation reactions were transformed into *E. coli* JM109 Competent Cells, High Efficiency (Promega©) by heat-shock treatment and the cells were allowed to recover in SOC buffer at 37ºC for 1.5-2 hours. The transformation mix was then plated onto Luria Bertani (LB) agar plates

containing 100µg/ml ampicillin, 0.1mM IPTG and 1mg of X-gal for blue/white screening. The colonies were then grown overnight at 37℃. Positive clones (white colonies) were individually picked and recovered in 100µl LB-ampicillin growth media in microtitre plate wells at 37℃ for an hour, and then stored at 4℃.

Positive clones were screened for the correct size inserts by colony PCR using M13-long universal vector primers (M13-longF 5'-CGC CAG GGT TTT CCC AGT CAC GAC-3', M13-longR 5'-GAG CGG ATA ACA ATT TCA CAC AGG-3'). PCR reactions were performed in 30µl reaction volumes containing 1ul LB-ampicillin media in which cells are suspended, 0.5pmol/ul of M13-long forward and reverse primers, 1.5mM $MgCl_2$, 0.2mM dNTPs, 0.025U GoTaq® Flexi DNA polymerase (Promega©) and 1X Colourless GoTaq® Buffer. Each reaction was overlaid with a drop of mineral oil and the PCR was performed on a PTC-100™ PCR Thermocycler (MJ Research) with the following standard cycling conditions: (step1) 95ºC for 5min; (step2) 95ºC for 30sec, 65ºC for 30sec, 72ºC for 55sec, for 35 cycles; (step3) 72ºC for 5min. PCR products were visualized by electrophoresis on a 2% (w/v) agarose gel and the correct sized PCR product was excised and purified using the Wizard® SV Gel and PCR Clean-up System (Promega©), as per manufacturer's instructions.

For each individual, six to twelve positive clones were sequenced using the BigDye Terminator v3.1 Cycle Sequencing Kit (Applied Biosystems) and the M13-long forward primer. Sequencing reactions were performed on the ABI GeneAmp® PCR System 2700 v2.07 (Applied Biosystem) with the following standard cycling conditions: 25 cycles of 30sec at 94℃, 5sec at 50℃ and 4min at 60℃. Sequences were analyzed on an ABI 3130 Genetic Analyzer (Applied Biosystems) with ABI DNA Sequencing Analysis Software v5.2.

Data analysis

Sequence Alignment

The resulting sequence chromatograms for each individual were manually edited using Chromas v2.3 (available at: http://www.technelysium.com.au). Sequences with ambiguous nucleotides were removed from the data set. The number of unique sequences per individual was noted. Previous studies have shown that unique sequences obtained from PCR-cloning approaches may be artefacts caused by *Taq* polymerase errors during PCR (Bryja et al. 2005, Kennedy et al. 2002). Accordingly a new sequence variant, from both the genomic and cDNA samples, was only considered a new allele when it met the following criteria: the sequence was identified from PCR reactions from at least two different individuals or the sequence was identified from two separate PCR-cloning experiments from the same individual (Kennedy et al. 2002). Sequences were aligned with ClustalW v1.4 (Thompson *et al.* 1994) in order to establish a consensus sequence for each individual. Alleles were translated into amino acid sequences using the BioEdit Sequence Alignment Editor v7.0.4.1 (Hall 1999). Translated alleles with stop codons were removed from the data set. Gene identity was confirmed by National Center for Biotechnology Information (NCBI) BLAST searches (http://www.ncbi.nlm.nih.gov). The resulting DNA and amino acid sequence data sets were then analyzed using several statistical methods.

Monophyletic Relationships

Monophyletic relationships among Nile crocodile MHC alleles were inferred by the neighbour-joining method (NJ; Saitou and Nei 1987) in MEGA v3.1 (Kumar *et al.* 2004) using Kimura 2-parameter corrected distances (Kimura 1980). We tested for the best-fit model of evolution using the ModelTest (Posada and Crandall 1998). We found that the HKY model of evolution best fits the data (Hasegawa et al. 1985), however an analogue of the HKY model, i.e. Kimura 2-parameter corrected data was used as MEGA does not implement the HKY model.

Branch support was determined from 1000 iterations. The analysis included all the Class II sequences identified in this study along with sequences from *Caiman crocodilus* (AF256651.1, AF277661.1, AF256652.1), *Alligator sinensis* (AY491430.1, AY491422.1, AY491425.1), Alligator mississippiensis (U24404.1, U24402.1, U24403.1), *Eumeces chinensis chinensis* (AY772950.1, AY772947.1), *Sphenodon punctatus* (DQ124238.1, DQ124237.1, DQ124234.1), *Gallus gallus* (EF579812, EF579811, AJ248581, DQ008586.2) and *Spheniscus humboldti* (AB154395.1, AB154394.1, AB154393.1).

Detecting patterns of selection

To investigate whether positive selection plays a role in maintaining allelic diversity in the Nile crocodile, the relative rate of synonymous (d_S) and non-synonymous (d_N) substitutions per site were calculated in MEGA v3.1 (Kumar *et al.* 2004). Values were calculated using the Nei-Gojobori method (Nei and Gojobori 1986) with values corrected for multiple mutations at the same site using the Jukes-Cantor correction (Jukes and Cantor 1969). Positive selection ($d_N > d_S$) and the significance of differences between the d_N and d_S rates were estimated using a codon based *Z*-test in MEGA v3.1. Standard errors for the rates were based on 1000 bootstrap iterations. These calculations were performed using sequence data from the 17 amino acid codons that comprise the putative PBR, data from the non-PBR codons of the exon, and data representing the complete 200bp exon fragment. Putative codons of the Nile crocodile PBR were based on codons corresponding to the PBR in humans (Brown et al. 1993). Sliding window analysis of Tajima's *D* was calculated in DnaSP v4.0 (Rozas et al. 2003), using a window size of 36bp and a step size of 11bp. Tajima's *D* distinguishes between DNA sequences that are evolving under a non-random process, such as purifying or balancing selection, and those sequences evolving randomly, and is derived from the difference between genetic variation estimated from pairwise nucleotide differences and the number of segregating sites between individuals. In a population of constant size, Tajima's *D* is expected to be zero under neutrality, negative

under purifying selection against deleterious mutations and positive under balancing selection (Tajima 1989).

Results

Characterization of MHC Class II-like sequences

From 45 individuals we identified 28 unique Nile crocodile MHC Class II-like nucleotide sequences (See Appendix 1.2 for nucleotide alignment). Each unique sequence was identified through sequencing multiple PCR reactions for each individual, thus each sequence is either present in more than one individual or was recovered at least twice from independent PCR reactions. A NCBI nucleotide blast search confirmed homology of the sequences to MHC Class II β chain genes and showed high sequence similarity with exon sequences from the Chinese alligator, spectacled caiman, Chinese skink and a wide variety of other vertebrates. Allele numbers ranged from one to nine per individual, suggesting that more than one putative MHC locus was amplified by the conserved primers used in this study. Three different categories of sequences were identified; classical Class II alleles, non-classical Class II alleles and a possible Class II pseudogene.

Classical Class II alleles

Nine sequences that were identified appear to be homologous to classical Class II alleles. Three of the nine sequences have a three base pair deletion that results in the loss of an amino acid residue. These alleles however still appear to be functional as one of the alleles was recovered from the cDNA analysis and the amino acid that is lost is not thought to be involved in peptide binding (Brown et al. 1993) (Figure 2.1). In addition this deletion has also been identified in a related crocodilian species, *Caiman crocodilus.* The nine sequences had high sequence similarity overall and the allele number for these sequences ranged from one to five per individual crocodile, suggesting that more than one putative Class II locus occurs within this gene family in Nile crocodiles. At this stage it is not possible to determine which of these sequences together represent individual loci, therefore the nine sequences

were designated *Crni-DAB**1-9 following the proposed nomenclature for MHC in nonhuman species; *Crni* is the species designation for alleles at MHC loci in *Crocodylus niloticus, D* stands for Class II loci, *A* for the particular family of Class II genes in a given species and *B* for genes that encode the receptor's β chain (Ellis et al. 2006; Klein et al 1990). For the exon 2 sequences of *Crni-DAB* alleles, high levels of both nucleotide and amino acid sequence polymorphism were detected. A total of 66 of 184 (35.8%) nucleotides and 35 of 61 (57.3%) amino acid residues were polymorphic. At each variable site, a maximum of six different amino acid codons were observed (Figure 2.1). Polymorphisms were particularly high in the PBR of the amino acid sequences identified, where 13 of the 17 amino acid positions that make up PBR (Brown et al. 1993) were polymorphic. All reported *Crni-DAB* alleles contain features expected of Class II β chain genes. The conserved amino acid sequence motif 'HFDS' which is a potential CD4 binding site (Auffray and Novotny 1986), along with the amino acid residues tryptophan (W-61) and asparagine (N-82) that are thought to form hydrogen bonds with bound peptides, were observed in the *Crni-DAB* alleles (Brown et al. 1993, Kaufman et al. 1994). The conserved cysteine (C-79) residue involved in the formation of cysteine salt bridges in Class II β1 domain is also present in the *Crni-DAB* sequences.

Non-Classical Class II alleles

From the nucleotide alignment, a number of sequences with a single base insertion/deletion (indel) event were also identified; 19 of the 28 alleles isolated through PCR-cloning were missing a single nucleotide at sequence position 94 (refer to Appendix 1.2 for the alignment). Because the putative amino acid sequences did not contain any stop codons and the alleles were consistently recovered from both genomic and cDNA samples it is highly likely that these alleles represent a divergent and functional locus. The Nile crocodile Class II sequences isolated in this study therefore appear to represent two distinct gene families. The single base pair deletion within the 19 sequences has resulted in a frame-shift mutation that renders a highly divergent amino acid sequence when compared to the *Crni-DAB* sequences.

Several conserved amino acid sequence elements characteristic of Class II loci have been lost and these sequences only share ±50% amino acid homology with *Crni-DAB* sequences. The allele number for this gene family ranged from one to eight alleles per individual crocodile, which again suggests that more than one putative locus has been amplified. These sequences were designated *Crni-DBB**1-19, as it is not possible to distinguish loci at this stage. The *Crni-DBB* gene family exhibit much lower levels of both nucleotide and amino acid sequence variation compared to *Crni-DAB* alleles. Only 22 of 183 (12%) nucleotide and 15 of 61 (24.5%) amino acid residues were polymorphic (Figure 2.2). A reduced level of polymorphism was also observed within the PBR, where only 4 of 17 (23.5%) amino acid positions that make up the PBR (Brown et al. 1993) were polymorphic. Despite low sequence variation, all *Crni-DBB* alleles were found at a particularly high frequency and were isolated from both genomic DNA and cDNA.

Even though the conserved sequence motif 'HFDS' (Auffray and Novotny 1986) is present in the *Crni-DBB* sequences other conserved sequence elements that are essential for the function and stability of the molecule are not found (Figure 2.2). For example the amino acid residues tryptophan (W-61) and asparagine (N-82) are replaced with glycine (G) and threonine (T) respectively, which may alter the ability of the molecule to bind foreign peptides and thereby influence the function of these exons (Brown et al. 1993, Kaufman et al. 1994). The cysteine salt bridges which are characteristic of the β1 and β2 domains, and are important for protein folding and stability, are replaced by alanine (A), and have most likely modified the original function of the molecule. To identify the possible function or type of MHC locus represented by the *Crni-DBB* sequences, a BLAST search was performed using the putative amino acid sequences and showed the closest similarity to whale *DRB*-like sequences (56.6% homology) and chicken Class II β antigens (42.7% homology).

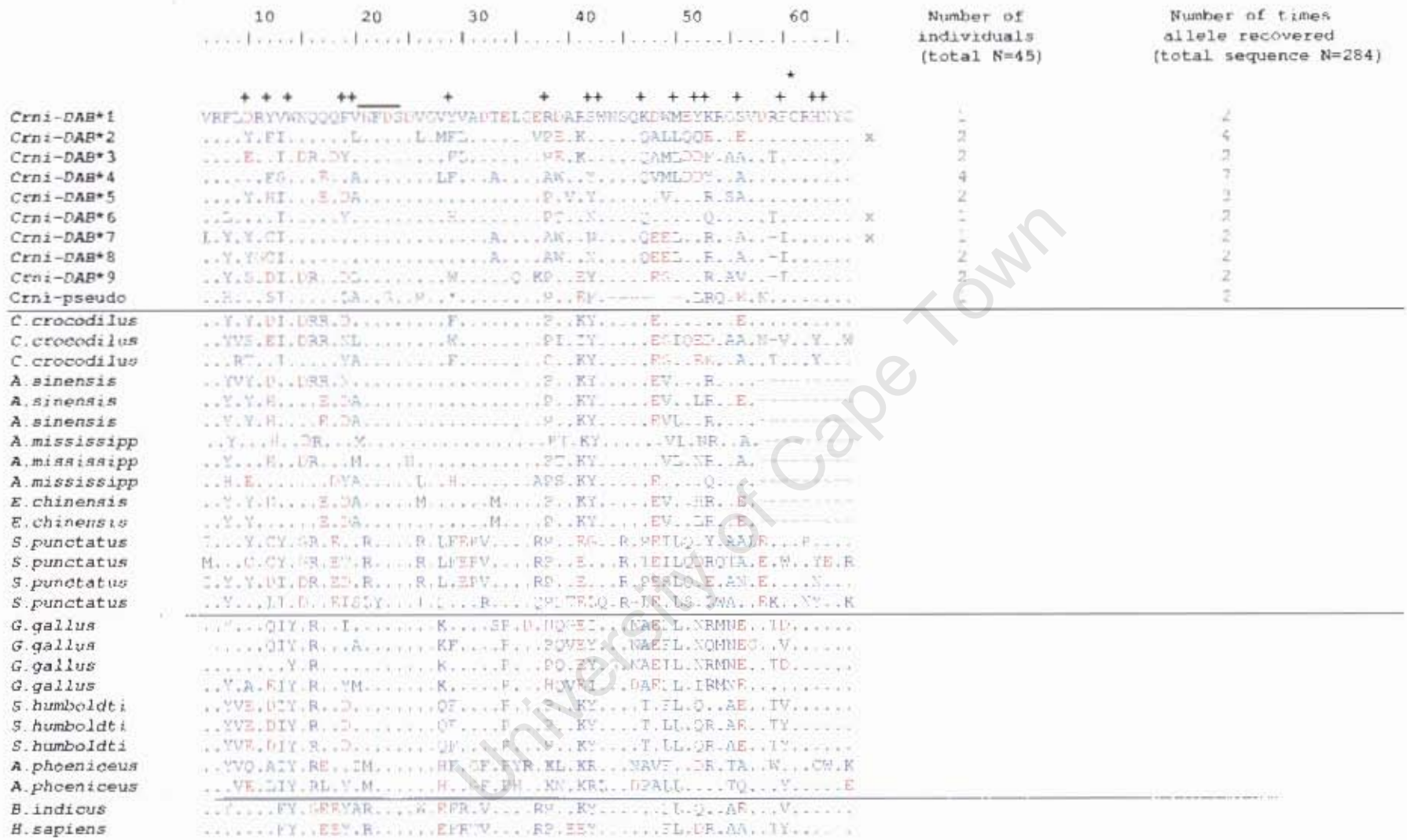
Possible Class II pseudogene

Finally, a number of sequences identified in this study had undergone a large deletion event of 18bp (nucleotide positions 115 to 132; refer to Appendix 1.2 for alignment). The putative amino acid sequence for these alleles is most likely to be rendered non-functional due to 1) the loss of six amino acid codons (18bp) represented by the deletion event, and 2) the consistent presence of a stop codon within the sequence (Figure 2.1). These significant sequence changes suggest the occurrence of a possible pseudogene within Class II loci of the Nile crocodile MHC.

MHC Class II-like sequences from cDNA

From the 54 sequences that represented cDNA-based PCR products, 11 unique alleles were identified. Of these, three were identified as *Crni-DAB* alleles and eight as *Crni-DBB* alleles. The alleles identified in this study, using both cDNA and genomic DNA, therefore appear to represent two functional Class II loci in the Nile crocodile that differ by the loss of an amino acid; *Crni-DBB* alleles most likely represent a non-classical crocodilian MHC locus.

Figure 2.1: Amino acid sequence alignment and individual haplotype occurrence for MHC-*DAB* alleles and possible pseudogene identified in the Nile crocodile. Numbers indicate the amino acid position in the MHC β chain; * indicates conserved cysteine residue involved in the formation of disulfide bridges; + indicates codons within the peptide binding region (PBR) (after Brown *et al.* 1993); x indicates alleles recovered from cDNA; - indicates missing codon; solid black line indicates a potential CD4 binding site (after Auffray and Novotny 1986). Individuals *Crni-DAB**1-9 *Crocodylus niloticus* and MHC Class II exon 2 from *Caiman crocodilus* (AAF99283.1, AAF99317.1, AAF99284.1), *Alligator sinensis* (AAS72408.1, AAS72400.1, AAS72403.1), Alligator mississippiensis (AAA99115.1, AAA99113.1, AAA99114.1) *Eumeces chinensis chinensis* (AAW66890.1, AAW66888.1), *Sphenodon punctatus* (AAZ77719.1, AAZ77718.1, AAZ77715.1, AAZ77714.1), *Gallus gallus* (ABR10677.1, ABR10676.1, ABR10675.1, CAM32045.1), *Spheniscus humboldti* (BAF91115.1, BAF91121.1, BAF91114.1), *Agelaius phoeniceus* (AAF07314.1, AAK08082.1), *Bos indicus* (CAB52178.1), *Homo sapiens* (ABW87990.1)



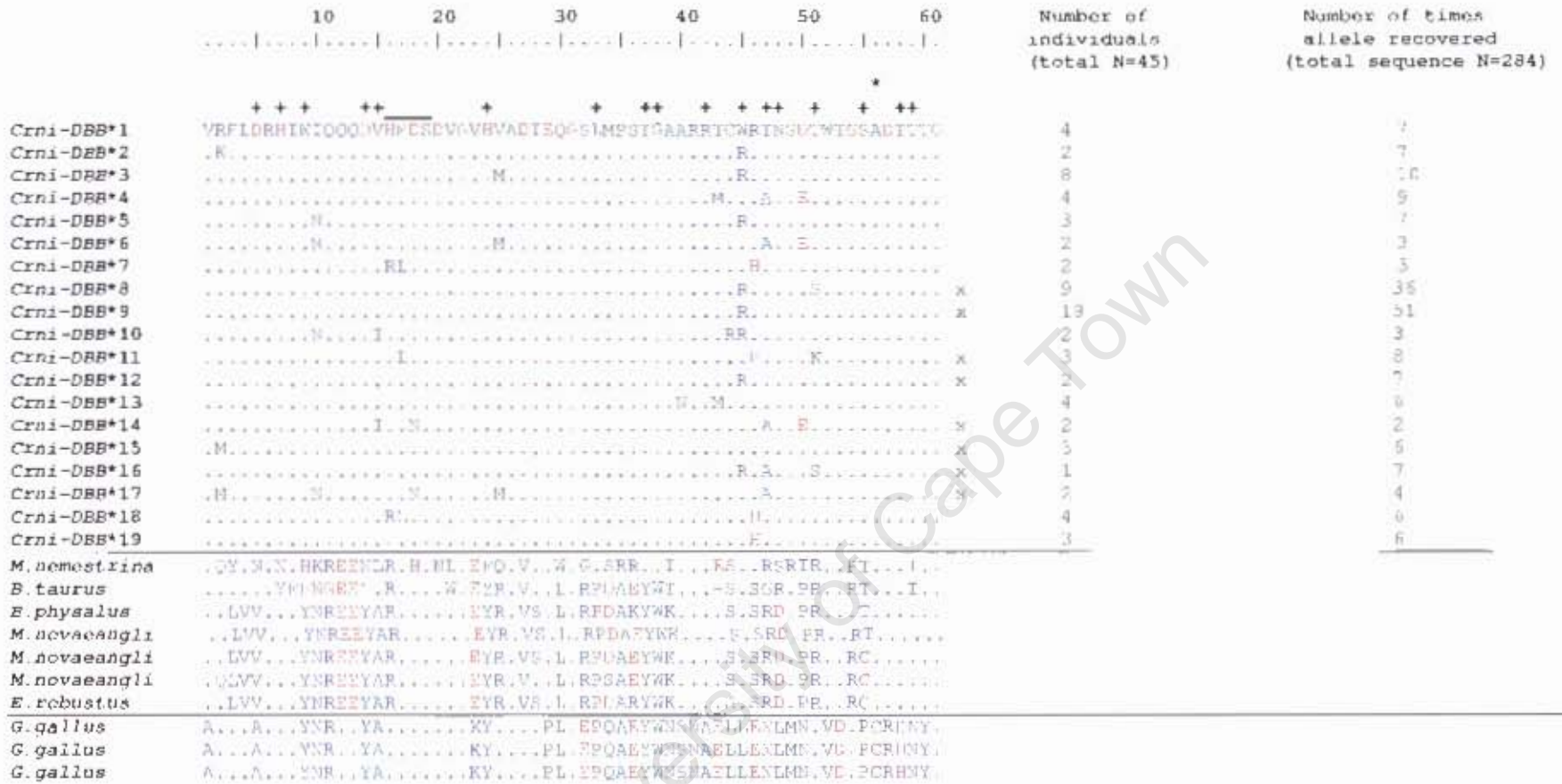


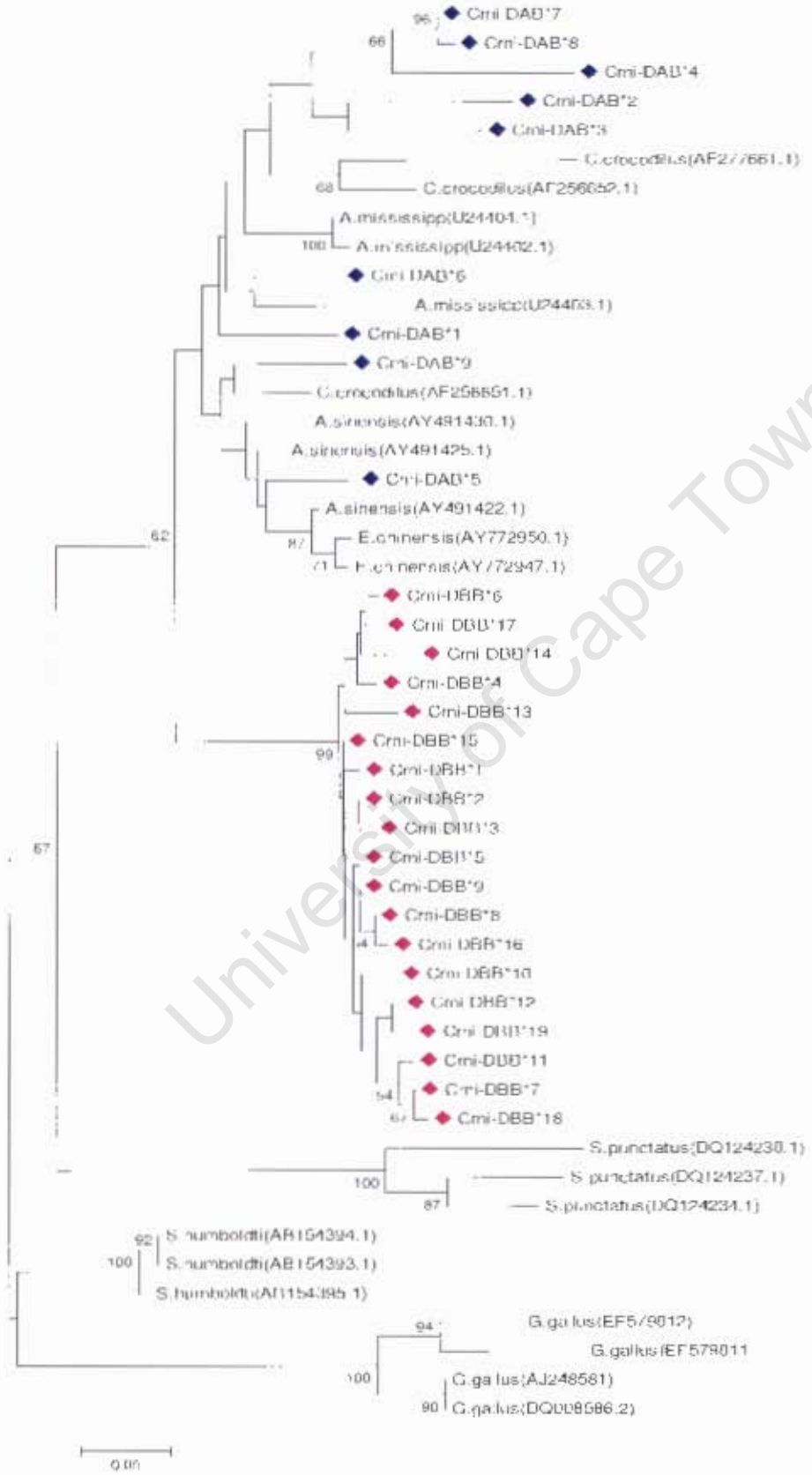
Figure 2.2: Amino acid sequence alignment and individual haplotype occurrence for MHC-DRB alleles identified in the Nile crocodile. Numbers indicate the amino acid position in the MHC β chain; * indicates a conserved alanine residue at the position of the classical MHC Class II β chain conserved cysteine residue; + indicates position of the codons within the peptide binding region (PBR) of classical MHC Class II β chains (after Brown *et al.* 1993); x indicates alleles recovered with cDNA; - indicates missing codon; solid black line indicates a potential CD4 binding site (after Auffray and Novotny 1986). Individuals *Crni-DBB**1-19 *Crocodylus niloticus* and Class II exon 2 anitigens of *Macaca nemestrina* (AAB59618.1), Bos Taurus (ABN80338.1), *Balaenoptera physalus* (ABC95541.1), *Megaptera novaeangliae* (ABC95551.1, ABC95553.1, ABC95552.1), *Fschrichtius robustus* (ABC95545.1), Gallus gallus (CAB58145.1, AAY40297.1, ABC95552.1).

Monophyletic Relationships

The NJ tree in Figure 2.3 is a graphical representation of the monophyletic relationships between MHC Class II sequences from the Nile crocodile and a number of additional Archosaurian and Lepidosaurian species. Its interesting to note that despite the divergence of the Lepidosauria lineage at ~285 mya (Rest et al. 2003), skink Class II exons do not form a monophyletic group with respect to the species level but instead cluster together with other reptile Class II exons; this suggests the presence of similar alleles in distantly related species i.e. the retention of trans-species polymorphisms. Class II exons from the tuatara, regarded as sister to snakes and lizards, formed a clade that is sister to all the other reptile sequences. Nile crocodile *Crni-DAB* and *Crni-DBB* sequences formed two distinct clades. *Crni-DAB* alleles did not form a monophyletic clade but instead clustered with other reptile Class II β exon sequences, further suggesting the retention of allelic lineages over both long periods of time and across deep speciation events. *Crni-DBB* alleles formed a well-supported monophyletic clade that is sister to all the *Crni-DAB* alleles, suggesting that *Crni-DAB* and *Crni-DBB* represent two evolutionarily different MHC gene families in the Nile crocodile. Figure 2.4 represents the relationships among Nile crocodile MHC Class II sequences alone, and indicates that *Crni-DAB* alleles have given rise to *Crni-DBB* alleles. This result suggests that the non-classical *DBB* locus most likely evolved from a classical Class II β gene in the Nile crocodile. The alleles of the bird Class II exons formed distinct species-specific monophylectic clusters on a separate branch to that of crocodilian Archosaurs and Lepidosaurs. Species-specific clustering of MHC alleles is characteristic of birds, where possible interlocus genetic exchange has lead to the homogenization of loci within species and thus the loss of orthology (Wittzell et al 1999).

Figure 2.3: A rooted tree showing the monophyletic relationships between Nile crocodile exon 2 alleles isolated in this study; sequences representing the tuatara (*S. punctatus*) and a

number of species from the Archosaurian lineage are also included. A. mississipp is an abbreviation for *Alligator mississippiensis.* Numbers represents percentage bootstrap support values (1000 replicates) and the scale bar represents genetic distance. Bootstrap values greater than 50% are shown. Blue diamonds represent *Crni-DAB* exon 2 sequences and pink diamonds represent *Crni-DBB* exon 2 sequences.



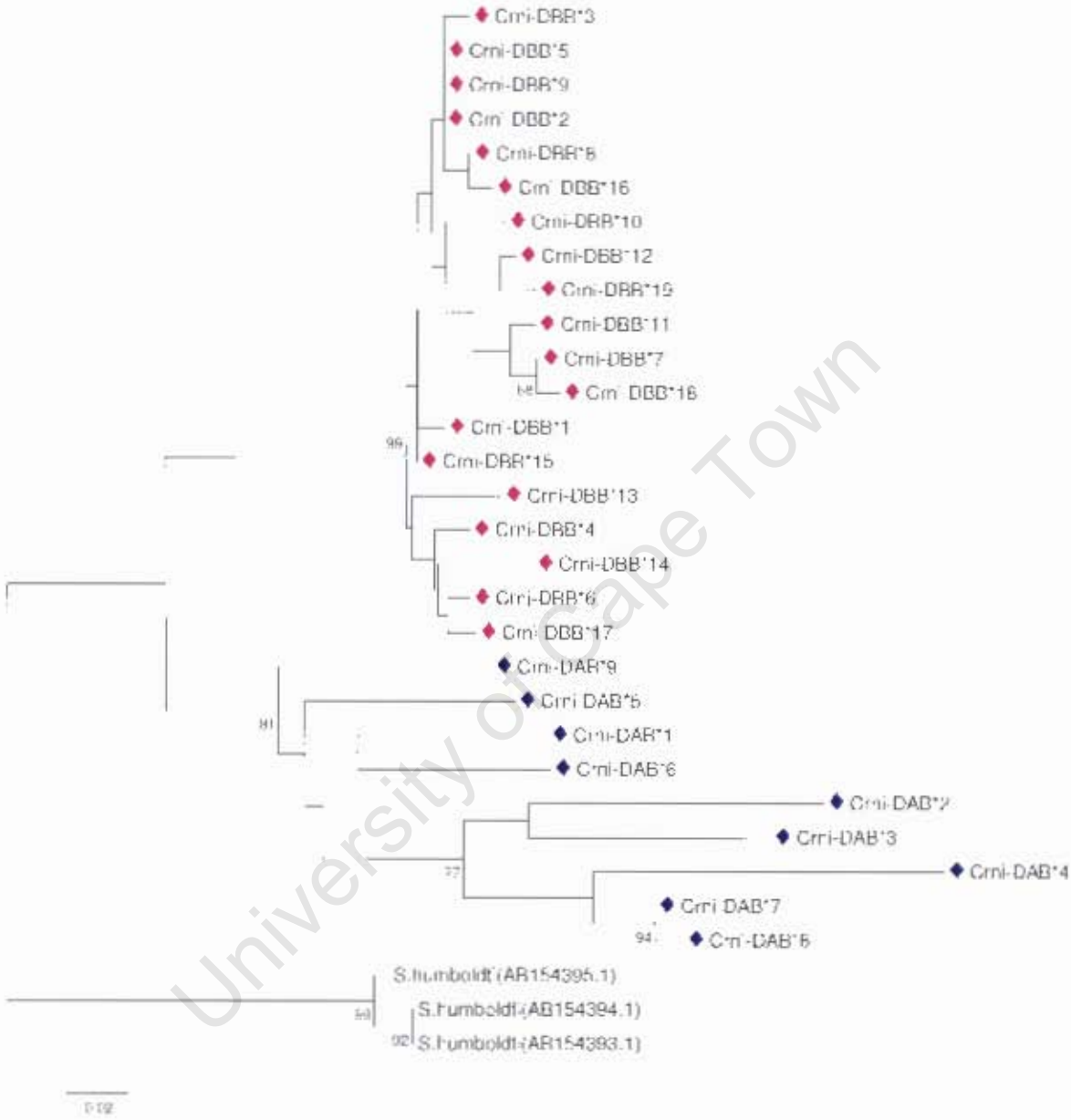


Figure 2.4: A rooted tree showing the monophyletic relationships between exon 2 alleles of the 28 Nile crocodile sequences isolated in this study. Tree is rooted with *Spheniscus humboldti* (BAF91115.1, BAF91121.1, BAF91114.1). Numbers represent percentage bootstrap support values (based on 1000 replicates) and the scale bar represents genetic distance. Bootstrap values greater than 50% are shown. Blue diamonds represent *Crni-DAB* exon 2 sequences while pink diamonds represent *Crni-DBB* exon 2 sequences.

Patterns of selection

The substitution rates d_N and d_S were calculated to test whether variation at Class II exons of the Nile crocodile is maintained by balancing selection. Among the *Crni-DAB* sequences, the rate of non-synonymous substitutions (d_N=0.39±0.08) for codons within the PBR was found to be 6.5 times higher than that of synonymous substitutions (d_S=0.06±0.05), whereas within non-PBR codons, the rate was comparable and did not deviate significantly from parity (Table 2.2). The codon based *Z*-test (*Z*=3.33, *p*=0.001) further supports strong balancing selection for the maintenance of functional polymorphisms within the PBR of *Crni-DAB* alleles and this pattern was also observed for the overall domain (*Z*=1.97, *p*=0.025). Among the *Crni-DBB* sequences the rate of non-synonymous substitutions (d_N=0.02±0.01) within the non-PBR codons was found to exceed that of synonymous substitutions (d_S=0.01±0.01), but this was not statistically significant (*Z*=1.49, *p*=0.06). The rates within the PBR were found to be comparable and did not deviate from parity.

The results from the sliding window analysis of Tajima's *D* further suggests that divergent patterns of selection act across the exon 2 sequences (Figure 2.5). A pattern in keeping with balancing selection is apparent among the *Crni-DAB* alleles whilst values that are suggestive of purifying selection are apparent in the *Crni-DBB* alleles. The plot of Tajima's *D* for *Crni-DAB* exon 2 sequences are mostly greater than zero, particularly in the regions of exon 2 that correspond to the PBR codons. The plot that represents exon 2 sequences from *Crni-DBB* alleles is consistently negative and this pattern is maintained across the whole exon including those regions that correspond to the PBR.

Table 2.2: Non-synonymous (d_N) and synonymous (d_S) nucleotide substitution rates for codons of *C.niloticus* exon 2 alleles. N=number of codons analysed.

	N	d_N (±S.E)	d_S (±S.E)	d_N : d_S	Z-test Codon based
Crni-DAB					
PBR	17	0.13 (±0.03)	0.11 (±0.03)	6.5**	0.58
Non-PBR	44	0.39 (±0.05)	0.06 (±0.05)	1.26	3.33
Overall domain	61	0.18 (±0.03)	0.09 (±0.02)	1.87*	1.97
Crni-DBB					
Putative-PBR	17	0.04 (±0.02)	0.08 (±0.06)	0.5	0.8
Non-PBR	44	0.02 (±0.01)	0.01 (±0.01)	2.45	1.49
Overall domain	61	0.03 (±0.01)	0.03 (±0.01)	1	0.02

Number of non-synonymous substitutions per non-synonymous site (d_N), number of synonymous substitutions per synonymous site (d_S); *p<0.05, **p<0.001.

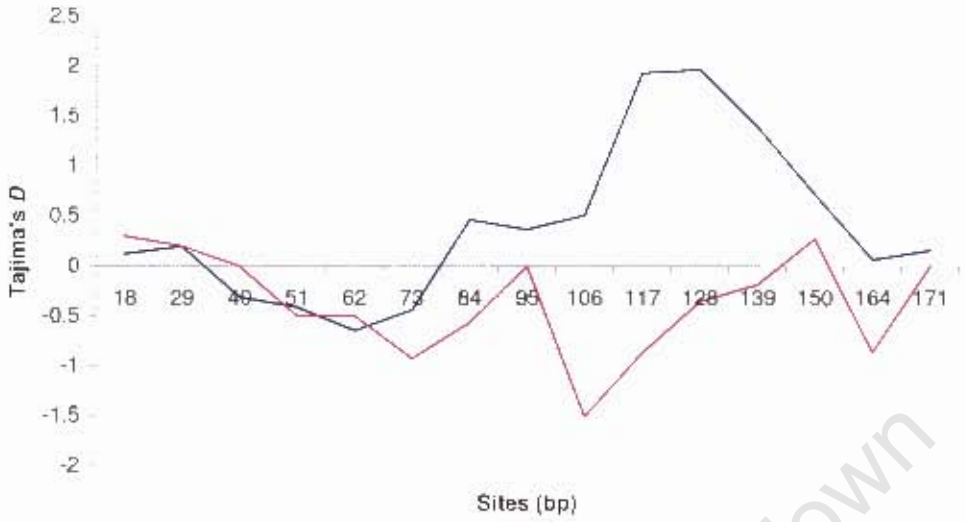


Figure 2.5: Sliding window analysis of Tajima's *D* for MHC Class II β exon 2 sequences from the Nile crocodile (window size 36bp, step size11bp). The blue line represents *Crni-DAB* sequences and the pink line the *Crni-DBB* sequences. Values of *D*>1 indicates balancing selection. *D*<1 indicates purifying selection.

Discussion

Diversity of MHC Class II loci in the Nile crocodile

In this study we investigated diversity at MHC Class II loci of the Nile crocodile. Results from this study suggest the presence of at least two distinct families of functional MHC Class II β genes in the Nile crocodile. A large number of Nile crocodile exon 2 alleles were identified and it is highly unlikely that these are the result of PCR artefacts. Extensive sequencing of unrelated clones for each individual, together with results from independent PCR reactions, using both genomic and cDNA, make up the sequence data presented in this study. Exon sequences from the *Crni-DAB* family were highly polymorphic and all alleles showed closest homology to classical Class II MHC alleles of other vertebrate species. In this study, cDNA analysis confirmed the presence of exon 2 gene products in blood samples from crocodile individuals, along with the presence of conserved sequence elements consistent with functionality e.g. the conserved cysteine residue at amino acid position 79 and the conserved 'HFDS' motif

involved in CD4 binding. An interesting finding is that despite a recent history of severe over-exploitation (1957-1969), Okavango Nile crocodiles currently exhibit high levels of both nucleotide and amino acid *DAB* variability. Similar levels of variation have been found in other vertebrate species despite histories of demographic decline (e.g. chamois, Schaschl et al. 2004; African buffalo, Wenink et al. 1998).

The *Crni-DBB* family is highly divergent and may be a non-classical class II locus. Following a duplication event, classical genes are thought to degenerate into non-classical genes as a result of deleterious mutations (Hughes and Nei 1989). This idea is reinforced by the observation that non-classical genes of one species are more closely related to classical genes of the same species, rather than to other non-classical genes belonging to different species. Two non-classical Class II genes have been identified in mammals, *DM* and *DO* (Alfonso and Karlsson 2000). In birds, a homologue of *DM* has been identified and is therefore likely to be present in non-avian reptiles. The function of *DM* is well established and found to be conserved across different species (Alfonso and Karlsson 2000). The *DM* locus promotes peptide loading of Class II molecules by catalyzing the release of CLIP (class II-associated invariant chain peptide) that prevents the Class II molecular binding pocket from collapsing. *DM* molecules are present in all Class II cells, whereas *DO* is mainly expressed in B cells; in these cells the majority of *DM* molecules form tight heterotetrameric complexes with *DO* and this association is essential for the intracellular transport of *DO*. It has been shown that the gene products from *DO* clearly modify the peptide exchange activity of *DM*, but the physiological importance of this interaction is not fully understood (Alfonso and Karlsson 2000). Further experiments, such as tissue expression patterns, will aid in determining whether *Crni-DBB* is a Non-classical locus and if it contains features characteristic of *DM* genes. non-classical MHC Class II β molecules have also been reported in other reptile species (such as tuatara, Miller et al. 2005). *Crni-DBB* alleles exhibited low levels of both nucleotide and amino acid diversity, whilst changes were observed

across amino acids that are normally highly conserved (Figure 2.2). Analysis of cDNA gene products confirmed the presence of expressed allelic products for nine of the nineteen alleles identified in this study. Together these features are consistent with *DBB* alleles representing a putative non-classical MHC Class II locus.

Evolution and selection at MHC Class II loci in the Nile crocodile

A number of individual crocodiles exhibited more than two alleles for both gene families, suggesting the presence of multiple functional Class II loci in *Crocodylus niloticus*. This is consistent with both Kasahara's (1999) '*chromosomal duplication model*' and Nei *et al.*'s (1997) '*birth-and death*' model of MHC evolution; together these models propose that the MHC arose by repeated gene duplication events that have been followed by gene loss through time; along with the effects of balancing selection, these processes lead to highly polymorphic and divergent groups of genes. In addition, this study also provides evidence of a possible pseudogene in the Nile crocodile; thus to further understand the evolutionary processes involved in locus expansion and/or contraction events within the MHC of the Nile crocodile, an extensive population-level assessment of MHC variation across the species geographic range is required.

While it is generally assumed that diversification of MHC genes is generated by the accumulation of point mutations and gene duplication (Nei et al. 1997), the mosaic pattern of sequence motifs found among MHC sequences suggest that recombination may also generate new sequence variants (Martinsohn et al. 1999; Parham and Ohta 1996; Reusch and Langefors 2005). Boysen et al. (1999) showed that a novel non-classical MHC Class I locus (*AG*) in two Old World primate species, is most likely the result of a duplication event, where *A* underwent an interlocus recombination event with *B* creating a hybrid locus *A/B*. It is possible that subsequent to a *Crni-DAB* locus duplication event, unequal crossing during an interlocus recombination event resulted in the creation of a duplicated locus that later underwent a single base pair deletion. Despite the

deletion, the locus still produced functional transcripts and this may have led to a new role for what became the *Crni-DBB* locus. Monophyletic analysis further suggests that the *Crni-DBB* locus has evolved from the *Crni-DAB* locus within the Nile crocodile (Figure 2.4), as *Crni-DBB* sequences form a distinct cluster that branches off from *Crni-DAB* sequences.

The fate of a duplicated MHC locus is strongly determined by the type of selection acting on it; thus duplicated genes may become non-functional pseudogenes, evolve in an adaptive manner within the confines of their original function, or they can be transformed into non-classical loci and take on a new function (Boyson et al. 1999; Hughes 1995; Hughes and Nei 1989; Klein et al. 1993; Nei et al. 1997). A strong signal of balancing selection acting on the *Crni-DAB* locus was apparent in the relatively disproportional d_N and d_S rates within the PBR (d_N:d_S ratio = 6.5) versus those in the non-PBR region (d_N:d_S = 1.26). This disparity is characteristic of balancing selection at MHC loci, where functional diversity is favoured in molecular regions involved in antigen binding, enabling the host's immune system to recognise a wider range of antigenic peptides (Ekblom et al. 2003; Hughes and Hughes 1995; Oliver and Piertney 2006). A neighbour-joining cluster analysis showed that some alleles were more closely related to Class II alleles of different non-archosaur reptile species than to any other *Crni-DAB* alleles (Figure 2.3). Closer interspecific similarities between alleles often result from the retention of ancestral polymorphisms across speciation events i.e. trans-species polymorphism (Figueroa et al. 2000). This phenomenon has been reported for numerous species studied to date and is regarded as evidence for balancing selection (Figueroa et al. 2000; Oliver and Piertney 2006).

Positive selection that favours functional diversity was absent at the PBR codons of *Crni-DBB* alleles. The rate of d_N in the *DBB* PBR (d_N:d_S = 0.5) did not exceed that of the non-PBR (d_N:d_S = 2.45). In addition, sliding window analysis of Tajima's *D* suggests that a degree of purifying selection may be acting on the

locus to significantly decrease rates of d_N versus d_S (Figure 2.5). Similar patterns have been reported at a non-classical MHC Class I locus of two Old World primates, the baboon (*Papio hamadryas anubis*) and rhesus monkey (*Macaca mulatta*) (Boyson et al. 1999). These findings suggest that following the origin of *Crni-DBB* from *Crni-DAB* via possible duplication, the locus began evolving in a divergent manner; the function of *Crni-DBB* is therefore likely to differ to that of a classical MHC Class II gene like *Crni-DAB*.

In conclusion, this study describes exon 2 allelic diversity for MHC Class II genes of *Crocodylus niloticus,* and presents evidence for at least two families of MHC Class II β genes in the Nile crocodile. The *Crni-DAB* family is homologous to classical Class II genes reported in a diverse array of vertebrates and appears to contain multiple functional loci. High levels of allelic variability characterised *Crni-DAB* alleles. Codons within the PBR showed the greatest levels of variability, indicating strong balancing selection acting on loci to maintain functional diversity. Results from this study also suggest that a possible recombination event may have led to the formation of the *Crni-DBB* family; *Crni-DBB* amino acid sequences are highly divergent and most likely represent a non-classical crocodilian Class II gene. The *Crni-DBB* alleles are characterized by reduced diversity along with the occurrence of mutations within highly conserved elements of classical Class II loci. Sliding window analysis of Tajima's *D* reveals that *Crni-DBB* is evolving in a divergent manner to *Crni-DAB*; balancing selection was not detected within the PBR and its function is most likely of a different nature to that of classical Class II genes. One of many challenges is to now identify the relative contributions of both positive selection and random genetic drift in the evolution of MHC diversity in extant populations of the Nile crocodile.

Chapter 3:

Analysis of Variation at the Major Histocompatibility Complex in *Crocodylus niloticus*: Selection or Drift?

Introduction

A major goal in population and conservation genetics is to understand the relative contribution of different forces that may influence both adaptive and neutral genetic variation in natural populations. As a rule, population size is the most important factor governing the effects of random genetic drift, gene flow and inbreeding on levels of genetic variation, and the maintenance of genetic diversity, in natural populations (Sommer et al. 2002). The reduction of genetic variation within small populations may result in the loss of adaptive fitness components such as survival, growth rate, reproductive output and the ability to adapt to environmental change (Ciofi and Bruford 1999; Campos et al. 2006; Sommer 2005). The rates at which these adaptive features are impacted upon are, in turn, directly dependent upon life history traits and dispersal patterns that influence the population genetic make-up of species (Balloux et al. 1998; Sommer et al. 2002; Storz 1999).

The relationship between population size and genetic variability is likely to vary in strength for different categories of loci; this is because loci are subject to varying intensities and types of selection (Frankham 1996; Madsen et al. 2000). Selection can result in the maintenance of higher variation at loci that are essential for ensuring viability, despite a reduction in variation in other parts of the genome. Therefore different molecular makers can be used to investigate aspects of population structure and history, and it is essential to choose markers

with the right degree of sensitivity to answer the question of interest (Sunnucks et al. 2000).

Microsatellite loci are considered to be the nuclear marker of choice for population genetic studies; they are highly proficient markers for elucidating population history, for detecting past bottlenecks and measuring inbreeding (Kim et al. 1999; Sunnucks et al. 2000). Microsatellite loci comprise highly polymorphic, short tandem repeated sequences of di -, tri - or tetra-nucleotide DNA (Hartl and Jones, 2001; Jarne and Lagoda 1996), and generally evolve in accordance with a neutral model. Microsatellite loci accumulate mutations rapidly as they do not encode functional transcripts (Sunnucks et al. 2000), and are found throughout the eukaryotic genome within exons, introns and 'junk' DNA (Beaumont and Bruford 1999). Because they are selectively neutral, microsatellites may inadequately describe adaptive differences between populations and fail at detecting divergent selective regimes involving interactions between an individual and its environment (Kim et al. 1999, Sommer 2005). In contrast, the MHC is a multigene family of tightly linked genes that constitute important components of the cell-mediated immune system's antigen-presenting function and is therefore critical for the development of host resistance to microbial infection. In contrast to microsatellite loci, MHC variability may reflect evolutionarily adaptive processes within- and between- populations (Bryja et al. 2006; Chen et al. 2006; Hedrick 2003; Kelley et al. 2005; Miller et al. 2005; Piertney and Oliver 2006; Sommer 2005). Loci within the MHC may therefore act as highly informative molecular markers with which to investigate the mechanisms and implications of molecular adaptation in vertebrates.

Several studies have shown that host genetic diversity plays a vital role in shielding populations against pathogens and widespread disease epidemics (Altizer et al. 2003; Sommer 2005). In natural populations it has been suggested that genetic variation is maintained through cyclic interactions between pathogens and their hosts via '*frequency dependent selection*', (Hamilton and

Zuk 1982). A study on great reed warblers (*Acrocephalus arundinaceus*) demonstrated that temporal variation in pathogen abundance contributes to significant variation in MHC allele frequencies between year cohorts (Westerdahl et al. 2004). The wetlands and rivers within the Okavango Delta system in Botswana consist of several semi-permanent drainage channels, lagoons and floodplains that connect during the annual period of flooding. Spatial and temporal fluctuations in pathogen exposure during the dry and wet season are thus likely to affect Nile crocodiles and as a result individual crocodiles might be challenged by a diverse array of parasites. MHC allele frequencies might therefore be expected to fluctuate over time due to changes in parasite-driven selective pressures where different alleles are favoured in different years (Piertney and Oliver 2006; Westerdahl et al 2004).

Recent studies, however, have also shown that neutral forces, such random genetic drift can also significantly influence MHC variation. Genetic drift is the random changes that occur in gene frequencies between generations (Dobzhansky and Pavlovsky 1957). One of the major consequences of random sampling, specifically in small populations, is the loss of genetic diversity within populations and diversification among populations. Low MHC variation has been revealed in a number of populations, including Malagasy giant jumping rats (*Hypogeomys antimena*; Sommer et al. 2002), Fin whales (*Balaenoptera physlaus*) and Sei whales (*B.borealis*) (Trowsdale et al. 1989) and the cheetah (*Aconyx jubatus*; Yuhki and O'Brien 1990). These findings suggest that balancing selection may not always be the dominant force shaping MHC variation and that variation in allele frequencies over time could also be the result of demographic events, particularly variation in the rate of dispersal between genetically differentiated populations (Westerdahl et al. 2004). The relative importance of balancing selection versus random genetic drift in maintenance of genetic diversity depends on the relationship between population size and selection.

In this study we investigated temporal variation in allele frequencies of MHC Class II β genes in the Nile crocodile population of the Okavango Delta. Extensive sampling of individuals, as well as detailed characterization of MHC Class II β genes in this species (Chapter 2), enables us to examine how MHC allele frequencies vary over time in a natural population. Measures of variation were analysed over four successive year cohorts of hatchling crocodiles. To account for fluctuations in MHC variation that may have resulted from mainly demographic processes, temporal changes in selectively neutral microsatellite loci were also examined. If between-cohort changes in allele frequencies are higher than expected in MHC alleles when compared to microsatellite alleles, we can then infer that year-on-year selection is probably acting on MHC loci in the Nile crocodile. If both MHC and microsatellite allele frequencies vary more than expected from random, changes at the MHC is most likely the result of demographic processes and random genetic drift.

Methods

Sample collection and DNA extraction

Blood and scute tissue samples were collected for a total of 74 hatchling crocodiles in the panhandle region of the Okavango River, Botswana over a four year period from February 2002 to December 2006 (Figure 3.1). The average sample size was 18 individuals per year (number of individuals: 2002 n=17; 2003 n=13; 2004 n=23; 2005 n=21). The individual age of Nile crocodiles were estimated from their relative size, thus individuals smaller than ~20cm snout-vent length are considered to be hatchlings from the current breeding season (Hutton 1989). Samples were stored in 96% ethanol at 4°C until extraction of total genomic DNA. Total DNA was extracted either from whole-blood or tissue samples from individuals as detailed in Chapter 2.

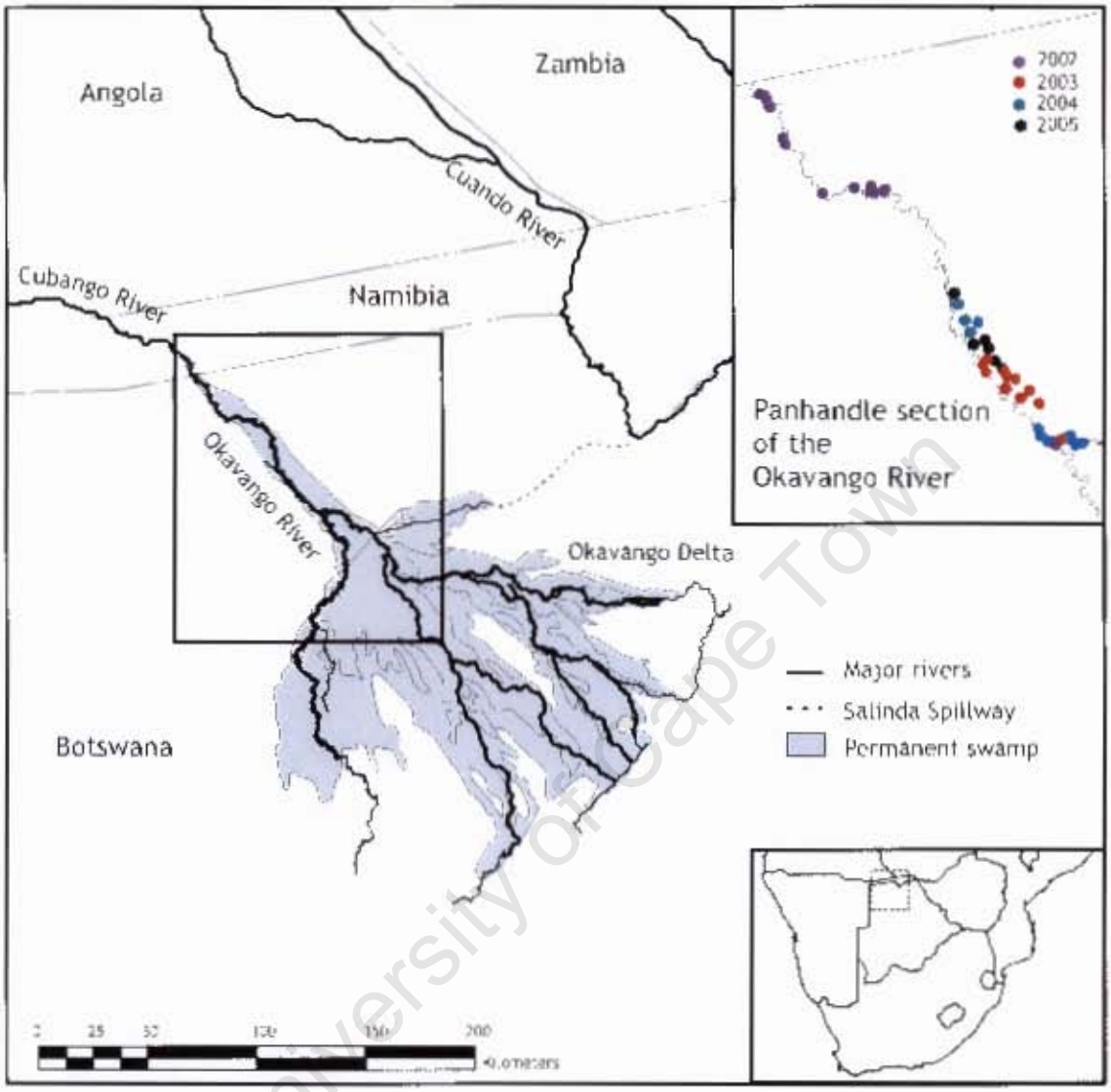


Figure 3.1: Map of the sampling distribution of Nile crocodile hatchlings in the Okavango River, Botswana from 2002-2005.

MHC genotyping

Characterisation of variation at MHC Class II β loci in the Nile crocodile revealed the presence of more than one gene locus (detailed in Chapter 2). Therefore diversity at Class II loci was first assessed by capillary electrophoresis (CE) single-strand conformation polymorphism (SSCP). CE-SSCP is a mobility-shift based analysis of single-stranded DNA, which utilizes the unique sequence specific conformation of the DNA to detect sequence polymorphisms (Orita et al.

1989, Sunnucks et al. 2000). Under the appropriate conditions DNA strands fold into structures that migrate through the polymer matrix according to their shape and sequence composition. DNA strands of different sequences do not usually assume the same shape and thus have distinctive gel mobility; this allows for the detection of sequence variation i.e. different genotypes in a population can be detected using CE-SSCP; unique genotypes can then be cloned and sequenced, substantially reducing project costs associated with large-scale PCR-cloning-sequencing protocols. CE-SSCP has recently been used to determine MHC diversity in a number of species that possess more than one putatively functional Class II locus, for example in three-spined sticklebacks (Binz et al. 2001, Jäger et al. 2007). This technique is thought to detect ~99% of single base pair changes within 100 to 300bp fragments (Lessa and Applebaum 1993).

MHC Class II loci were amplified by PCR using fluorescently labelled primers (forward primer with 6'-HEX and the reverse primer with 6'-FAM). Reactions were performed on the ABI GeneAmp® PCR System 2700 v2.07 (Applied Biosystem), following the PCR cycling conditions described in Chapter 2. The relative concentration of PCR products was determined by visualizing PCR products after electrophoresis on 2% (w/v) agarose gels and then diluting the reactions where necessary. CE-SSCP, using an ABI 3130Xl Genetic Analyzer (Applied Biosystems), was then used to discriminate among the SSCP-genotypes that were present in the samples. The results from each run were checked using Genemapper v3.7 Software (Applied Biosystems). Unique motifs identified from the CE-SSCP runs were then cloned and sequenced to determine the DNA sequences that together accounted for the motif. For each genotype motif, ten clones were sequenced.

Microsatellite genotyping

Variation at six microsatellite loci was investigated using *Crocodylus*-specific primers. Two of these loci were characterized from the American crocodile (*Crocodylus acutus)*, three from the Australian freshwater crocodile (*Crocodylus*

jonstoni) and one locus from the Australian saltwater crocodile (*Crocodylus porosus)* (FitzSimmons et al. 2002). Primer details and the optimal PCR conditions for each locus are shown in Table 3.1. For detection and sizing of PCR fragments the forward primer from each pair was end fluorescent labelled (6-FAM or 6-HEX; Applied Biosystems). PCR reactions were performed in 20µl reaction volumes containing 0.5pmol/µl forward and reverse primers, 10-50ng DNA, 0.2mM dNTPs, 1mM $MgCl_2$, 0.025U GoTaq® Flexi DNA polymerase (Promega©) and 1X Colourless GoTaq® Buffer. PCR conditions were optimized with respect to $MgCl_2$ concentration, primer concentration and annealing temperature. A 'touchdown' PCR protocol was used with the following cycling conditions: (step 1)94ºC for 3min; (step2) 94ºC for 45sec, 63ºC for 45sec, 72ºC for 45sec, for 5 cycles; (step3) 94ºC for 45sec, 58ºC for 1min, 72ºC for 1min, for 30 cycles; (step4) 72ºC for 7min. Reactions were performed on an ABI GeneAmp® PCR System 2700 v2.07 (Applied Biosystem). To determine relative concentrations, PCR products were visualized by electrophoresis on 2% (w/v) agarose gels and then diluted if necessary. A mixture containing 0.4µl GeneScan© Rox-350 standard size marker (Applied Biosystem), 1µl Loading buffer, 2.5µl Formamide (Promega©) and 2.5µl per PCR sample was denatured for 5mins at 98ºC and then snap-cooled on ice. Samples were then electrophoresed on a 6% denaturing polyacrylamide gel using an ABI 373 DNA Sequencer (Applied Biosystem). Results were compiled and analyzed using GENESCAN v2.1 and GENOTYPER v2.1 software (Applied Biosystems).

Table 3.1: Details of six microsatellite loci used for genotyping Okavango Nile crocodile hatchlings.

Locus	Primer-pair Sequence (5`-3`)	Repeat Motif	PCR Product Size (bp)	Ta (ºC)	No. of alleles (Individuals, N = 100)
C391*	F:ATGAGTCAGGTGGCAGGTTC R:CATRAAATACACTTTGAGCAGCAG	$(CA)_{22}$	129-163	63-58	7
CUD68*	F:GCTTCAGCAGGGGCTACC R:TGGGGAAACTGCACTTTAGG	$(CA)_{15}$	121-133	63-58	6
Cj18+	F:ATCCAAATCCCATGAACCTGAGG R:CCCAGTGCTTACAAGAGGCTGG	$(CA)_{21}$	205-225	63-58	9
Cj119+	F:GTTTGCTGTGGAATGTTTCTAC R:CGCTATATGAAACGGTGGCTG	$(CA)_{14}$	172-194	63-58	9
Cj35+	F:GTTTAGAAGTCTCCAAGCCTCTCAG R:CTGGGGCAAGGATTTAACTCTC	$(CT)_7TA$ $(CA)_{17}$ $(CT)_{12}$	166-174	63-58	6
Cp10^	F:GATTAGTTTTACGTGACATGCA R:ACATCAAGTCATGGCAGGTGAG	$(CA)_{15}$	192-194	63-58	2

Locus, primer sequence, repeat motif, observed polymerase chain reaction (PCR) product sizes and optimal Ta temperature shown. Observed number of alleles per locus in *C. niloticus* also included. * *C.acutus*, + *C.jonstoni* and ^ *C.porosus* (FitzSimmons et al. 2002).

Data Analysis

MHC Sequence alignment

The resulting sequence chromatograms representing each unique SSCP genotype were manually edited in Chromas v2.3 (available at http://www.technelysium.com.au). Sequences with ambiguous nucleotides and translated alleles with stop codons were removed from the data set. Sequences

were aligned with ClustalW v1.4 (Thompson *et al.* 1997) and the number of unique sequences per year was noted. The resulting DNA sequence set was then analyzed using several statistical methods. Because the methods in this study have been shown to amplify two MHC gene families that appear to be under divergent selection regimes (detailed in Chapter 2), alleles representing the two gene families were analysed separately in comparisons with the microsatellite data.

Tests for linkage disequilibrium

To confirm that the allelic state of one microsatellite locus is independent from the allelic states of all the other loci, gametic disequilibrium was tested for using GENEPOP v3.4 (Raymond and Rousset 1995, available at (http:// wbiomed. curtin.du.au/genepop).

Measures of genetic variation and genetic differentiation

Microsatellite and MHC genetic variation was assessed by calculating the average number of alleles for each year and determining respective allele frequencies for the two data sets. Microsatellite data was analysed in the program GENEPOP v3.4 (Raymond and Rousset 1995) and MHC allele frequencies were calculated in Arlequin v3.0 (Excoffier et al. 2005).

A number of statistical approaches was used to determine whether cohort allele frequencies were differentiated between the sampling periods. An exact test of allelic differentiation for MHC alleles across all four years was calculated assuming the null hypothesis of a random distribution of *k* (different haplotypes) among *r* populations (Raymond and Rousset 1995), in Arlequin v3.0 (Excoffier et al. 2005). An analogous test, Fisher's exact test, was used to calculate pairwise measures of allelic differentiation between years for microsatellite alleles in GENEPOP v3.4.

Levels of genetic differentiation were also quantified using a genetic structure approach (Analysis of Molecular Variance; AMOVA) implemented in Arlequin v3.0. Overall genetic differentiation across the different years was assessed using a pairwise matrix of Nei's genetic distance (1972); each allele was treated as either present (1) or absent (0) and the resulting matrix was then used to determine whether the overall distribution of variance was larger between yearly cohorts than expected from random.

Results

MHC genotyping

MHC genotypes were initially screened using CE-SSCP analysis for a total of 74 individuals from the four years of samples. A total of 58 different MHC genotype motifs were identified; specifically 16 motifs for 2002, 11 for 2003, 18 for 2004 and 13 for 2005. Because of the degree of allele overlap among CE-SSCP motifs (identified during the process of characterising CE-SSCP motifs), only a sub-set of unique motifs from each year was 'genotyped' by cloning and sequencing i.e. individual allele composition was determined from an average of 11 motifs per year (number of motifs: 2002 n=12; 2003 n=8; 2004 n=12; 2005 n=10). From these 42 SSCP genotypes, 70 different alleles were identified, of which seven belonged to the *Crni-DAB* gene family and the remaining 63 to the *Crni-DBB* gene family. In a number of individuals more than two MHC-*DAB* and MHC-*DBB* alleles were again found, suggesting that these two genes comprise more than one putative locus. Over the four years *Crni-DAB* alleles showed moderate fluctuations in allele frequencies (Figure 3.2). For two of the years (2003 and 2005) *DAB* allele diversity was reduced to just one allele; the overall mean diversity was 3.00 alleles per year. A total of 33% of *DAB* alleles were shared between the years, particularly between 2002 and 2004. *Crni-DBB* allele frequencies showed moderate levels of fluctuation over the four years, with certain alleles being 'favoured' in all four years (Figure 3.3). Moderate levels of *DBB* allelic diversity was observed across the four years of sampling with a

minimum of 15 unique alleles in 2003 and a maximum of 28 in 2004; overall mean diversity was 20.5 alleles per year. A total of ±50% of *DBB* alleles was shared between the four years.

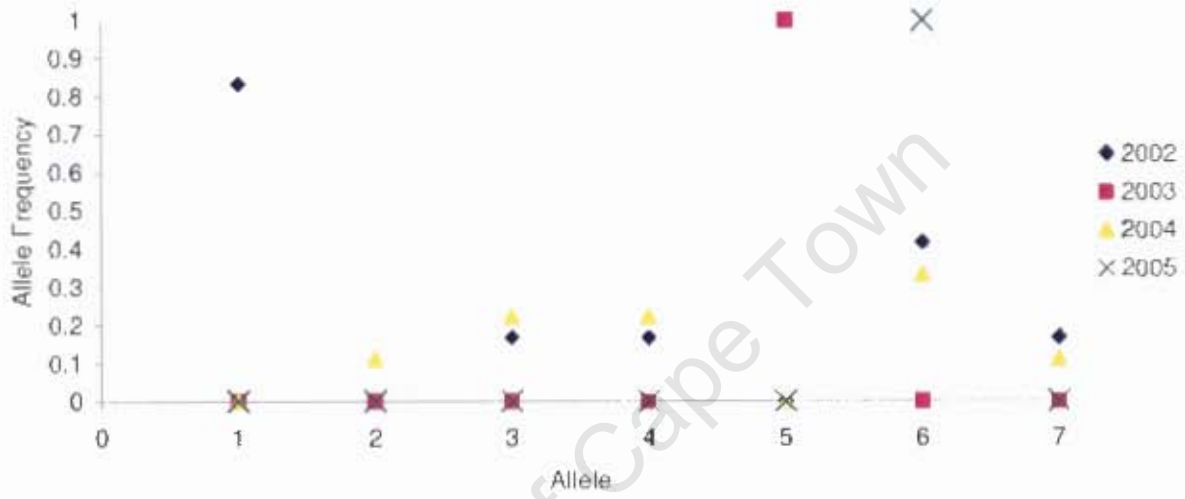


Figure 3.2: Changes in *Crni-DAB* allele frequencies over four years of sampling.

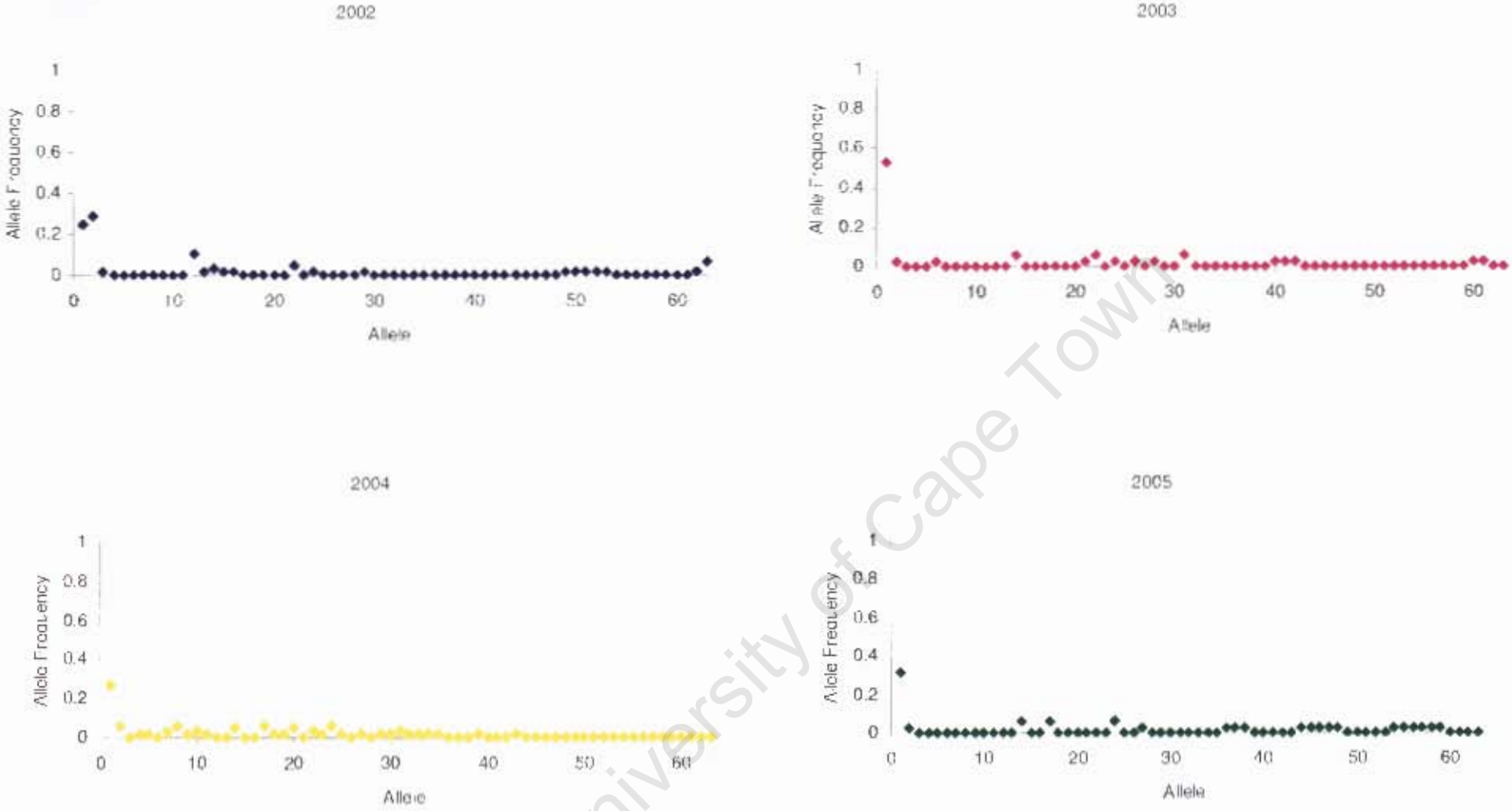


Figure 3.3: Changes in *Crni DBB* allele frequencies over four years of sampling.

Microsatellite genotyping

A total of 74 individuals was genotyped at six polymorphic loci, producing a total of 129 unique alleles. No evidence of linkage was found among the microsatellite loci. All four years displayed moderate levels of total allelic diversity with a minimum of 29 alleles in 2005, and a maximum of 35 alleles in the most variable year (2004); overall mean allelic diversity was 6.5 alleles per locus and 32.3 alleles over the six loci per year. For plots of allele frequencies for the six loci, across the different years see Appendix 2.

Measures of differentiation

We investigated whether allele frequencies for both MHC gene families and microsatellite alleles varied more than expected by chance between the four different years. Genic differentiation was assessed between years to test for differences in allele distributions across the four different years. Overall, *Crni-DAB* alleles showed no differentiation in frequencies between the years ($p>0.05$), whereas both *Crni-DBB* and microsatellite alleles showed highly significant differentiation between the years ($p<0.001$) (Table 3.2). Unfortunately the result for the classical Class II *DAB* alleles is most likely the result of insufficient data compared to *DBB* and microsatellite alleles e.g. in 2003 and 2005 only a single *DAB* allele was recovered from the sample set and violates the assumptions upon which the distribution of the test statistic is based. Interestingly, the plots for changing allele frequencies over time of the *DBB* alleles suggest that temporal differentiation in this data set is likely the result of changes in relatively few alleles, as the majority of alleles were at low frequencies over the years.

Table 3.2: Results from pairwise Exact tests of genic differentiation, for microsatellite and MHC Class II β alleles in the Nile crocodile. Exact test *p*-values are derived from 10000 Markov chain length steps. * Values significantly different from zero ($p<0.05$).

	2002	2003	2004	**Overall**
Microsatellite				
2002	-			
2003	0.03*	-		
2004	0.00*	0.00*	-	
2005	0.00*	0.14	0.00*	***p*<0.001**
MHC (*DAB*)				
2002	-			
2003	0.17	-		
2004	0.98	0.30	-	
2005	1	1	1	$p=0.86$
MHC (*DBB*)				
2002	-			
2003	0.00*	-		
2004	0.00*	0.25	-	
2005	0.00*	0.16*	0.67	***p*<0.001**

Levels of genetic differentiation were also quantified using an overall approach, where MHC and microsatellite alleles were analysed within the framework of an AMOVA, based on estimates of Nei's genetic distance (1972) over the different cohorts. Microsatellite alleles showed slight but significantly different frequencies between the cohorts (42 alleles, AMOVA $\Phi_{st}=0.06$, $p=0.01$), while MHC-*DBB* alleles did fluctuate across the years were, however, the overall effect on the variance in *DBB* allele frequencies across the years was negligible (63 alleles, AMOVA $\Phi_{st}=0.001$, $p>0.05$). This is most likely due the high number of low frequency alleles in the data set. Unfortunately it was not possible to perform an AMOVA analysis on the *DAB* allele data set because of the lack of allelic variation in two of the four years of sampling data. The relationships between pairwise estimates of genetic distance for the three data sets can be seen in Figures 3.4 and 3.5). It is clear from these plots that divergence at MHC loci across the four

years of sampling was consistently low when compared to the microsatellite data. Whilst the analysis is hampered by 1) limited data for *DAB* alleles, and 2) large numbers of low frequency alleles for *DBB* alleles, these plots do suggest that changes in microsatellite and MHC allele frequencies between cohorts are unlikely to be driven by the same processes.

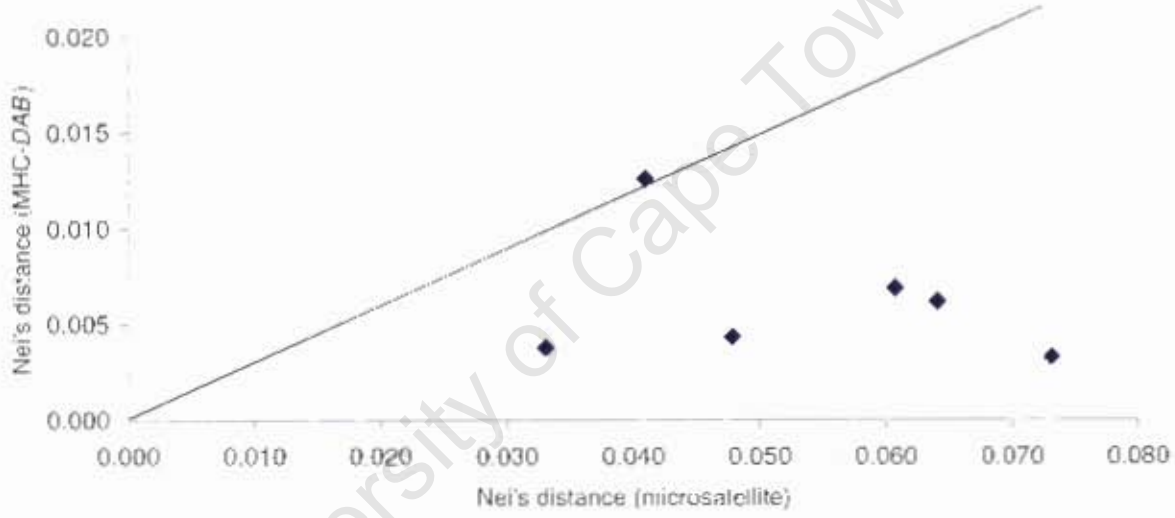


Figure 3.4: Pairwise comparisons of Nei's genetic distance between MHC *Crni DAB* and microsatellite alleles.

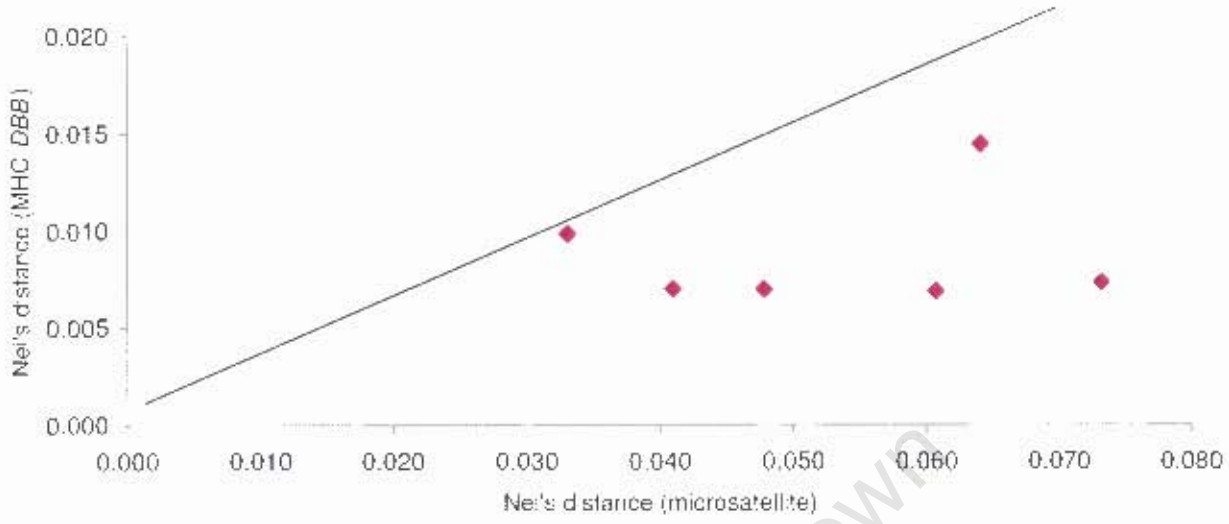


Figure 3.5: Pairwise comparisons of Nei's genetic distance between MHC *Crni-DBB* and microsatellite allele.

Discussion

The maintenance of genetic diversity in natural populations is influenced by a number of processes; these processes include random genetic drift and natural selection that act by removing variation, mutation and migration that introduce variation and balancing selection that impedes the loss of variation. The balance between these processes depends on the size of the population. Temporal changes in allele frequencies of both adaptive and neutral alleles are therefore influenced by selection (e.g. frequency dependent selection), random genetic drift and gene flow that operate within and between populations. In this study we investigated how MHC allele frequencies vary over time in a natural population and intended to use the results to determine whether temporal changes in allele frequencies at adaptive loci are the result of year-on-year selection or merely random demographic processes. Balancing selection is thought to be the main factor influencing the extreme levels of polymorphism found within MHC genes (Piertney and Oliver 2006; Sommer 2005); therefore to determine whether selection acting on the MHC genes of the Nile crocodile would result in significant

the observed changes in *DBB* allele frequencies across the different years, *DBB* alleles are thought to evolve in a divergent manner to classical MHC loci and their exact function is still unknown. Further experiments on the nature and function of *DBB* are thus required before we are able to completely understand the role of selection and/or genetic drift on *DBB* alleles.

Crni-DAB alleles

Low levels of allelic diversity were found for the *Crni-DAB* alleles; only seven alleles were identified from the 74 individuals over the four years and as a result some years have very few alleles e.g. samples from 2003 and 2005 were characterised by single alleles. A significant fluctuation in allele frequencies (Figure 3.2), coupled with particularly low levels of differentiation among years sharing alleles was found for the *DAB* data set (Table 3.2). Whilst balancing selection might slow down the rate at which adaptive variation is lost, the effects of genetic drift can exceed this, particularly in small populations (Miller and Lambert 2004). Several studies of natural populations have reported lower MHC differentiation in comparison to levels based on microsatellite loci e.g. in the Malagasy giant jumping rat (*Hypogeomys antimena*, Sommer 2003) and in the red grouse (*Lagopus lagopus scoticus*, Piertney 2003), suggesting that demographic factors can play a very significant role in shaping MHC variation.

The interpretation of this data set is, however, limited by the lack of alleles recovered for the *Crni-DAB* gene family. This lack of *DAB* allelic data may be due to a number of factors. These include 1) an artefact effect of the number of clones sequenced per individual; results suggest the presence of more *DBB* than *DAB* loci and alleles and therefore there is a drawback to only sequencing ± ten clones per SSCP motif. Whilst a strategy of sequencing a minimum of ten clones per potential locus would increase the chance of recovering the true allelic diversity of a population, this approach is just not economically practical when a gene may have up to ten functional loci. Unfortunately, given the paucity of published data on crocodilian MHC loci, the sequencing strategy used in this

temporal changes in allele frequencies, we compared four years of MHC allelic data with that of microsatellite loci.

Changes in microsatellite allele frequencies

Microsatellite alleles were characterized by relatively high levels of allelic diversity (mean of 6.5 alleles per locus). Microsatellite alleles were also found to vary more than expected by chance over the four years. The random changes in allele frequencies between cohorts suggest that random genetic drift is acting on the population. The effect of random genetic drift is dependent on the size of a population, specifically the effective population size (N_e) (Dobzhansky and Pavlovsky 1957), i.e. the smaller the N_e, the stronger the effect of drift. The Okavango Nile crocodile population has a recent history of severe over-exploitation (1957-1969) and even though estimates place the current census population size at ~2500, current parental generation N_e is estimated at ~100 individuals, resulting in an N_e/N ratio of only 0.04 (Bishop et al. in review). It therefore seems reasonable that the current reduced N_e may contribute significantly to changes microsatellite allele frequencies in this population. To confirm the relative influence of reduced N_e in this population a comparison with temporal changes in allelic data from additional crocodile populations is still required.

It can also be argued that the significant fluctuations in microsatellite alleles could be the result of underlying population structure in the Okavango River; spatial structure across the river might then mirror results from the temporal sampling. We can reject this explanation as unlikely since previous work on the population history of the Okavango Nile crocodile, using both microsatellite and mtDNA markers, suggested that no underlying structure occurs and that the Okavango River represents a single, continuous population (Bishop et al. in review). Gene flow has also been shown to be an important factor influencing, and generally counteracting, genetic differentiation via drift (Ciofi and Bruford 1999). Within the Okavango River there is a region that is characterized by low numbers of

crocodiles and could possible be a barrier to gene flow. Low numbers of crocodiles in this region, however, are thought to be the result of decreased fish stocks and a lack of suitable nesting sites (A. Leslie pers. comm.), and the area is not a barrier to gene flow. Additionally, previously marked individuals have been recaptured on either sides of this river section. The immigration of novel alleles will also have direct effects on changes in allele frequencies. Immigration of individuals into the Okavango system occurs periodically when excessive flooding results in the flow of the ephemeral Salinda Spillway (see Figure 3.1). We can, however, discount this factor as immigration from the neighbouring Zambezi drainage basin during the period of sample collection was essentially zero; this was because the Salinda Spillway did not flow for the duration of the study. It is therefore most likely that changes in allele frequencies, at microsatellite loci, over the four years is the result of genetic drift acting within the population; this process is likely to be intensified by the substantially reduced N_e of the Okavango Nile crocodile population.

Changes in MHC *Crni-DAB* and *Crni-DBB* allele frequencies

Two families of MHC Class II β genes in the Nile crocodile have been identified: *Crni-DAB* and *Crni-DBB* (see Chapter 2 for details). *Crni-DAB* is thought to be homologous to classical Class II genes that have been reported in a diverse array of vertebrates and appears to contain multiple functional loci. *Crni-DBB* is highly divergent and most likely represents a non-classical crocodilian Class II gene. Results have also shown that *Crni-DBB* is evolving in a divergent manner to that of *Crni-DAB*, as no positive selection was detected within the PBR.

Balancing selection is generally thought to maintain MHC polymorphisms in natural populations; however the exact nature of how this occurs is still debated (Hedrick 1999). The major driving force appears to be the selective effects imposed by ever-changing pathogen loads (Sommer 2005). Evidence for balancing selection on the genes of the MHC can be found at different temporal scales, for example selection in the recent past can be determined by differences

in F_{st}-values compared to those expected under neutral theory, whereas selection in the distant past can be recognized as an excess of non-synonymous to synonymous substitution and the retention of trans-species polymorphisms (Sommer 2005). A higher rate of non-synonymous (d_N) substitutions were found in the PBR of *Crni-DAB* alleles compared to the non-PBR (Chapter 2); this most likely represents the outcome of selection occurring over thousands of generations, and thereby contributing to the evolution of these alleles (Miller and Lambert 2004; Hedrick et al. 2001). Amino acid changes specifically in the functionally important PBR of the MHC can therefore influence functional differences in pathogen resistance (Sommer 2005). The maintenance and renewal of variation in the PBR, either through recombination, pathogen mediated selection and/or mutation, are vital processes that lead to the appropriate immune response when facing new and co-evolving pathogens. To be able to infer if selection in the recent past is acting on the Nile crocodile population, temporal variation in MHC allele frequencies were examined over the four years of sampling. Since previous results (Chapter 2) have indicated that *Crni-DBB* alleles are not under balancing selection we analyzed data from the two gene families separately.

Crni-DBB alleles

Crni-DBB alleles were characterized by high allelic diversity; up to 63 unique alleles were identified with a mean diversity of 20.5 alleles per year (Figure 3.3). Previous results, however, have shown that despite the amplification of high number of alleles, *DBB* alleles were characterized by low levels of intra-allelic polymorphism (Chapter 2). A degree of purifying selection was also identified as acting on the molecule, thus reducing genetic diversity. Measures of genic differentiation identified high levels of overall differentiation over the four years. The analysis of variance across the years, however, did not indicate that *DBB* allele frequencies vary more than expected by chance, even when compared to microsatellite alleles and thus suggest that the frequencies of the various alleles are relatively stable. Whilst random genetic drift may account for a component of

study had the unforeseen effect of increasing the chance of recovering *DBB* alleles rather than *DAB*. 2) The result of degenerate PCR primers, where *DBB* alleles could be preferentially amplified over *DAB* alleles as the primers target relatively conserved binding regions. A possible solution to this could be the use of sequence data from full length cDNA clones for both gene families. By comparing the cDNA sequences, significantly divergent sequence regions could be identified in order to design primers that will ensure the amplification of only one of the gene families.

In conclusion, while balancing selection might impede the loss of genetic diversity, adaptive variation in wild populations is not immune to the effects of genetic drift; this is particularly pertinent to populations with reduced effective sizes. The interpretation of the data presented in this chapter is hindered by limited sampling of classical Class II MHC alleles; however results from this study do suggest that while the generation and maintenance of variation at MHC loci in the Okavango Nile crocodile is influenced by longer-term balancing selection (Chapter 2), neutral short-term forces, such as random genetic drift, are a major determinant of temporal patterns of diversity across the genome of this species.

Chapter 4:

General Conclusions

The genes of the Major Histocompatibility Complex (MHC) are among the best candidates for studying the mechanisms involved in, and the significance of, molecular adaptation in vertebrates (Sommer 2005). MHC variation can influence several essential biological traits, including immune recognition and susceptibility to pathogens and diseases, individual odour, mating preferences and kin recognition. Compared to neutral genetic markers, data from MHC loci therefore reflects evolutionarily important processes that occur within and between populations (Piertney and Oliver 2006; Sommer 2005). Specifically, studies on non-model species in natural populations can inform our understanding of the significance of MHC diversity in both individual fitness and the viability of natural populations (Edwards et al 1995).

The MHC is an extremely dynamic region of the genome and its most distinctive feature is the high levels of polymorphism exhibited by many of its loci. This study provides insight into the diverse array of mechanisms that are involved in the generation and maintenance of genetic diversity at MHC Class II loci of Nile crocodiles. These mechanisms are not mutually exclusive and include gene duplication, interlocus recombination and pathogen mediated selection (Ahmed et al. 2007). Balancing selection is thought to be the main mechanism involved in the maintenance of MHC variation (Hedrick 1999; Hedrick and Thomson 1983). Evidence for the role of balancing selection in maintaining and/or influencing MHC variation has been documented extensively in the literature as an increase in amino acid variation via non-synonymous mutations, specifically in the functionally important antigen binding sites (Hedrick 1999; Miller and Lambert 2004; Schaschl et al. 2004). The major component of balancing selection is

thought to be the pressure exerted by pathogens; therefore the amino acid composition can influence individual differences in pathogen and parasite resistance and thus lead to an appropriate and potentially adaptive immune response when faced with new and coevolving pathogens (Bryja et al. 2006; Sommer 2005). non-classical MHC genes may also have a functional role in disease resistance. These genes might play an important role in determining pathogen and parasite resistance either by themselves or in an epistatic manner together with classical MHC genes (Apanius et al. 1997; Sommer 2005).

This study provides evidence for the role of both balancing selection and neutral short-term processes involved in the generation and maintenance of genetic variation at Class II loci of the Nile crocodile. Comparitive studies on how MHC and neutral diversity is distributed within and among natural populations enable us to measure the effects of selection and thus be able to infer if temporal changes in MHC allele frequencies are most likely the result of year-on-year selection or demographic processes (Landry and Bernatchez 2001; Piertney and Oliver 2006; Sommer 2003). The main factor thought to account for high variation in MHC allele frequencies, when compared to neutral markers, is the spatial heterogeneity in selection pressures from pathogens and parasites (Piertney and Oliver 2006; Westerdahl et al. 2004). Nevertheless, to fully exclude the possibility that patterns of genetic divergence among different markers simply reflect variation in demographic processes, studies need to characterize variation in pathogen intensity as well as infection rates across populations and thus be able to more rigorously test frequency-dependent co-evolution between hosts and their parasites (Piertney and Oliver 2006).

Future Direction

The development of rapid, high throughput as well as sensitive screening techniques, such as denaturing gradient gel electrophoresis (DGGE) and single-stranded conformational polymorphism (SSCP), has resulted in rapid and economic methods to screen MHC variation for a large number of individuals

(Piertney and Oliver 2006; Sommer et al. 2002). The resulting data sets can be used to investigate a wide range of evolutionary ecology and conservation questions. Characterizing spatial changes in pathogen diversity and intensity will enhance our understanding of how MHC diversity is maintained in natural populations. Studies have shown that levels of infection lead to variation in gene expression patterns; hence differential patterns of gene expression are likely to be an important component of how selection acts on MHC genes (Wegner et al. 2006) e.g. a study on three-spined sticklebacks (*Gasterosteus aculeatus*) showed that expression of MHC Class II β genes was positively correlated with parasite load, indicating increased immune activation of the MHC when infection was frequent (Wegner et al. 2006). Studies investigating the role of gene expression in an evolutionarily ecological context can therefore contribute to the understanding of the importance of MHC expression in the onset and control of the adaptive immune response in natural population. An increasing emphasis in population genetics is also to determine the extent of social relationships within populations (Piertney and Oliver 2006). A study on endemic Malagasy rodents identified higher levels of MHC variation in a promiscuous species, *Eliurus myoxinus*, compared to a monogamous species, *Hypogeomys antimena* (Sommer et al. 2002). MHC variation can therefore be influenced by the size of a population as well as the mating system. Social structure also influences the transmission of parasites and diseases and thus it is important to understand its role in the temporal and spatial dynamics of the MHC.

Understanding the factors that influence and shape variation at MHC loci is a technically difficult and analytically complex task. Examination of MHC variation in natural populations has increased our understanding of how factors such as balancing selection, recombination and effective population size of a population can influence adaptive genetic variation. There is still, however, a vast arena of unanswered questions that will help to increase our knowledge of the importance of MHC genes in evolutionary ecology and conservation.

References

Aguilar A., Roemer G., Debenham S., Binns M., Garcelon D., Wayne R.K. (2004) High MHC diversity maintained by balancing selection in an otherwise genetically monomorphic mammal. *Proceedings of the National Academy of Sciences USA* **101**: 3490-3494.

Ahmed K.A., Saxena V.K., Ara A., Pramod A.B., Rajaram M.L., Dorman K.S., Majumdar S., Rasool T.J. (2007) Molecular cloning and sequencing of MHC class II beta 1 domain of turkey revelas high sequence identity with chicken. *International Journal of Immunogenetics* **34**: 97-105

Alfonso C., Karlsson L. (2000) Nonclassical class II MHC molecules. *Annual Review Immunology* **18**: 113-142

Altizer S., Harvell D., Friedle E. (2003) Rapid evolutionary dynamics and disease threats to biodiversity. *Trends in Ecology and Evolution* **18**: 589-596.

Apanius V., Penn D., Slev P.R., Ruff L.R., Potts W.K. (1997) The nature of selection on the major histocompatibility complex. *Critical Reviews in Immunology* **17**: 179-224.

Avise J.C. (1994) Conservation Genetics. In: *Molecular Markers, Natural History and Evolution* pp 361-370. Chapman & Hall Inc.

Balloux F., Goudet J., Perrin N. (1998) Breeding system and genetics variance in the monogamous, semi-social shrew, *Crocidura russula*. *Evolution* **52**: 1230-1235.

Beaumont M.A., Bruford M.B. (1999) Microsatellites in conservation genetics. pp 165-182. In: *Microsatellites: Evolution and Application*. Goldstein D.B., Schotterer C. (eds) Oxford University Press, Oxford.

Belov K., Lam M.K.P., Colgan D.J. (2004) Marsupial MHC class II *B* genes are not orthologous to the eutherian B gene families. *Journal of Heredity* **95**: 338-345.

Bernatchez L., Landry C. (2003) MHC studies in nonmodel vertebrates: what have we learned about natural selection in 15 years? *Journal of Evolutionary Biology* **16**: 363-377.

Binz T., Reusch T.B.H., Wedekind C., Milinski M. (2001) SSCP analysis of *Mhc* class II *B* genes in the threespine stickleback. *Journal of Fish Biology* **58**: 887-890.

Boyson J.E., Iwanaga K.K., Urvater J.A., Hughes A.L., Golos T.G., Watkins D.I. (1999) Evolution of a new nonclassical MHC class I in two Old World primate species. *Immunogenetics* **49**: 86-98.

Brown D.D., Wensink P.C., Jordan E. (1972) *Xenopus laevis* and *Xenopus mulleri*: the evolution of tandem genes. *Journal of Molecular Biology* **63**: 57-73.

Brown J.H., Jardetzky T., Gorga J.C., Stern L.J., Urban R.G., Strominger J.L., Wiley D.C. (1993) Three-dimensional structure of the human class II histocompatibility antigen HLA-DR1. *Nature* **364**: 33-39.

Bryja J., Galan M., Charbonnel N., Cosson J.F. (2006) Duplication, balancing selection and trans-species evolution explain the high level of polymorphism of the DQA MHC class II gene in voles (Arvicolinae). Immunogenetics **58**: 191-202.

Campos J.L., Posada D., Moran P. (2006) Genetic variation at MHC, mitochondrial and microsatellite loci in isolated populations of Brown trout (*Salmo trutta*). *Conservation Genetics* **7**: 515-530.

Carrington M., Leson G.W., Martin M.P., Kissner T., Vlahov D., Goedert J.J., Kaslow R., Buchbinder S., Hoots K., O'Brien S.J. (1999) HLA and HIV-1: heterozygote advantage and B*35-Cw*04 disadvantage. *Science* **283**: 1748-1752.

Chen S., Zhang Y., Xu M., Ji X., Yu G., Dong C. (2006) Molecular polymorphism and expression analysis of MHC class II *B* gene from red sea bream (*Chrysophrys major*). *Developmental and Comparative Immunology* **30**: 407-418.

Ciofi C., Bruford M.W. (1999) Genetic structure and gene flow among Komodo dragon populations inferred by microsatellite loci analysis. *Molecular Ecology* **8**: S17-S30.

Crocodylus niloticus (Laurenti 1765) [Online]. Available: www.cites.org [2007, November 01].

Dobzhansky, T., Pavlovsky, O. (1957) An experimental study of interaction between genetic drift and natural selection. *Evolution* **11**: 311 – 319.

Doherty P.C., Zinkernagel R.M. (1975) Enhanced immunological surveillance in mice heterozygous at the H-2 gene complex. *Nature* **256**: 50-52.

Edwards S.V., Gtahn M., Potts W.K. (1995) Dynamics of *Mhc* evolution in birds and crocodilians: amplification of class II genes with degenerate primers. *Molecular Ecology* **4**: 719-729

Edwards S.V., Hess C.M., Gasper J., Garrigan D. (1999) Toward an evolutionary genomics of the avian *Mhc*. *Immunological Reviews* **167**: 119-132.

Edwards S.V., Potts W.K. (1996) Polymorphism of genes in the Major Histocompatibility Complex (MHC): Implications for Conservation Genetics of Vertebrates. pp 214-237. In: *Molecular genetic Approaches in Conservation.* Smith T.B., Wayne R.K. (eds.) Oxford University Press, Oxford.

Ekblom R., Grahn M., Hoglund J. (2003) Patterns of polymorphism in the MHC class II of a non-passerine bird, the great snipe (*Gallinago media*). *Immunogenetics* **54**: 734-741

Ekblom R., SÆther S.A., Grahn M., Fiske P., Kalas A., Hoglund J. (2004) Major histocompatibility complex in a lekking bird, the great snipe (*Gallinago media*). *Molecular Ecology* **13**: 3821-3828.

Ekblom R., SÆther S.A., Jacobsson P., Fiske P., Sahlman T., Grahn M., Kalas J.A., Hoglund J. (2007) Spatial pattern of MHC class II variation in the great snipe (*Gallinago media*) *Molecular Ecology* **16**: 1439-1451.

Ellis S.A., Bontrop R.E., Antczak D.F., Ballingall K., Davies C.J., Kaufman J., Kennedy L.J, Robinson J., Smith D.M., Stear M.J., Stet R.J.M., Waller M.J., Walter L., Marsh S.G.E. (2006) ISAG/IUIS-VIC Comparative MHC Nomenclature Committee report, 2005. *Immunogenetics* **57**: 953-958.

Excoffier L., Laval G., Schneider S. (2005) Arlequin (version 3.0): An integrated software package for population genetics data analysis. *Evolutionary Bioinformatics Online* **1**: 47-50.

Figueroa F., Mayer W.E., Sultmann H., O'hUigin C., Tichy H., Satta Y., Takezaki N., Takahata N., Klein J (2000) MHC class II *B* gene evolution in East African cichlid fish. *Immunogenetics* **51**: 556-575.

FitzSimmons, N.N., Tanksley, S. Forstner, R.J., Louis, E.E., Daglish, R., Gratten, J. and Davis, S. (2000) Microsatellite markers for Crocodylus: new genetic tools for population genetics, mating system studies and forensics. pp 51-57 In: *Crocodilian Biology and Evolution.* Grigg, G.C., Seebacher, F., Franklin, C.E. (eds) Surrey Beatt and Sons, Chipping Norton.

Flint N.S., Van Der Bank F.H., Grobler J.P. (2000) A lack of genetic variation in commercially bred Nile crocodiles (*Crocodylus niloticus*) in the North-West Province of South Africa. *Water SA* **26**: 105-110.

Frankham R. (1996) Relationship of genetic variation to population size in wildlife. *Conservation Biology* **10**: 1500-1508.

Froeschle G., Sommer S. (2005) MHC Class II DRB Variability and Parasite Load in the Striped Mouse (*Rhabdomys pumilio*) in the Southern Kalahari. *Molecular Biology and Evolution* **22**: 1254-1259.

Gorzula S., Arocha-Piñango C.L. & Salazar C (1976) A method of obtaining blood by venipuncture from large reptiles. *Copeia* **4**: 838-839.

Guillemot F., Billault A., Porquie O., Behar G., Chausse A.M., Zoorob R., Kreiblich G., Auffray C. (1988) A molecular map of the chicken major histocompatibility complex: the class II β genes are closely-linked to the class I genes and the nuclear organizer. *EMBO Journal* **7**: 2775-2785.

Hall T.A. (1999) BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symposium Series* **41**: 95-98.

Hamilton W.D., Zuk M. (1982) Heritable true fitness and bright birds: a role for parasites? *Science* **219**: 384-387.

Harf R., Sommer S. (2005) Association between major histocompatibility complex class II DRB alleles and parasite load in the hairy-footed gerbil, *Gerbillurus paeba*, in the southern Kalahari. *Molecular Ecology* **14**: 85-91.

Hartl, D.L., Jones, E. W. (2001) *Genetics: analysis of genes and genomes*. 5th ed. Jones and Barrlett Publishers, INC.

Hasagawa, M., Kishino, H. and Jano, T. (1985) Dating of a human-ape splitting by a molecular clock of mitochondrial DNA. *Journal of Molecular Evolution* **21**: 160-174.

Hedrick P.W. (1999) Balancing selection and the MHC. *Genetica* **104**: 207-214.

Hedrick P.W. (2003) The Major Histocompatibility complex (MHC) in declining populations: an example of adaptive variation. pp 97-113. In: *Reproduction Science and Integrated Conservation*. Holt W.V., Pickard A.R., Rodger J.C., Wildt D.E. (eds.) Cambridge Univ. Press, UK.

Hedrick P.W., Parker K.M., Lee R.N. (2001) Using microsatellite and MHC variation to identify species, ESUs, and MUs in the endangered Sonoran topminnow. *Molecular Ecology* **10**: 1399-1412.

Hedrick P.W., Thomson G. (1983) Evidence for balancing selection at HLA. *Genetics* **104**: 449-456.

Hedrick P.W., Thomson G. (1988) Maternal-fetal interactions and maintenance of HLA polymorphism. *Genetics* **119**: 205-212.

Hill A.V., Allsopp C.E., Kwiatkowski D. (1991) Common West African HLA antigens are associated with protection from severe malaria. *Nature* **352**: 595-600.

Hughes A.L. (1995) Origin and evolution of HLA class I pseudogenes. *Molecular Biology and Evolution* **12**: 247-258.

Hughes A.L. (2000) Evolution of introns and exons of class II major histocompatibility complex genes of vertebrates. *Immunogenetics* **54**: 473-486.

Hughes A.L., Hughes M.K. (1995) Natural selection on the peptide-binding regions of major histocompatibility complex molecules. *Immunogenetics* **42**: 233-243.

Hughes A.L., Nei M. (1989) Pattern of nucleotide substitution at major histocompatibility complex class I loci reveals overdominant selection. *Nature* **335**: 167-170.

Hughes A.L., Nei M. (1990) Evolutionary relationships of class II MHC genes in mammals. *Molecular Biology and Evolution* **7**: 491-514.

Hughes A.L., Yeager M. (1998) Natural Selection and the Evolutionary History of Major Histocompatibility Complex Loci. *Frontiers in Bioscience* **3**: 509-516.

Hunt H.D., Goto R.M., Foster D.N., Bacon L.D., Miller M.M. (2006) At least one YMHC1 molecule in the chicken is alloimmunogenic and dynamically epressed on spleen cells during development. *Immunogenetics* **58**: 297-307.

Hunt H.D., Pharr G.T., Bacon L.D. (1994) Molecular analysis reveals MHC class I intra-locus recombination in the chicken. *Immunogenetics* **40** 370-375.

Hutton J.M. (1989) Movement, home range, dispersal and separation of size classes in Nile crocodiles. *American Zoologist* **29**: 1033-1050.

Jager I., Eizaguirre C., Griffiths S.W., Kalbe M., Krobbach C.K., Reusch T.B.H. Schaschl H., Milinski M. (2007) Individual MHC class I and MHC class II*B* diversities are associated with male and female reproductive traits in the three-spined stickleback. *Journal of Evolutionary Biology* **20**: 2005-2015.

Jarne, P., Lagoda, P.J.L. (1996) Microsatellites, from molecules to populations and back. *Trend in ecology and evolution* **11**: 424 – 429.

Jeffery K.J.M., Bangham C.R.M. (2000) Do infectious disease drive MHC diversity? *Microbes and Infection* **2**: 1335-1341.

Jukes T.H., Cantor C.R. (1969) Evolution of protein molecules. pp. 21-123. In: *Mammalian Protein Metabolism.* Munro H.N. (eds.) Academic Press, New York.

Kasahara M. (1999) The chromosomal duplication model of the major histocompatibility complex. *Immunological Reviews* **167**: 17-32.

Kasahara M., Hayashi M., Tanaka K., Inoka H., Sugaya K., Ikemura T., Ishibashi T. (1996) Chromosomal localization of the proteasome Z subunit gene reveals an ancient chromosomal duplication involving the major histocompatibility complex. *Proceedings of the National Academy of Sciences USA* **93**: 9096-9101.

Kaufman J., Salomonsen J., Flajnik M. (1994) Evolutionary conservation of MHC class I and class II molecules – different yet the same. Seminars in Immunology **6**: 411-424.

Kaufman J. (2000) The simple chicken major histocompatibility complex: life and death in the face of pathogens and vaccines. *Philosophical Transaction of the Royal Society London* **355**: 1077-1084.

Kaufman J., Milne S., Gobel T.W.F., Walker B.A., Jacob J.P., Auffray C., Zoorob R., Beck S. (1999) The chicken B locus is a minimal essential major histocompatibility complex. *Nature* **401**: 923-925.

Kaufman J., Salomonsen J. (1997) The "Minimal Essential MHC" revisted: Both peptide-binding and cell surface expression level of MHC molecules are polymorphisms selected by pathogens in chickens. *Hereditas* **127**: 67-73.

Kelley J., Walter L., Trowsdale J. (2005) Comparative genomics of major histocompatibility complexes. *Immunogenetics* **56**: 683-695.

Kennedy L.J., Ryvaar R., Gaskell R.M., Addie D.D., Willoughby K., Carter S.D., Thomson W., Ollier W.E.R., Radford A.D. (2002) Sequence analysis of MHC *DRB* alleles in domestic cats from the United Kingdom. *Immunogenetics* **54**: 348-352.

Kim T.J., Parker K.M., Hedrick P.W. (1999) Major Histocompatibility Complex Differentiation in Sacramento River Chinook Salmon. *Genetics* **151**: 1115-1122.

Kimura M. (1968) Evolutionary rate at the molecular level. Nature 217: 624-626.

Kimura M. (1980) A simple method for estimating evolutionary rate of base substitutions through comparative studies of nucleotide sequences. *Journal of Molecular Evolution* **16**: 111-120.

Klein J., Bontrop R.E., Dawkins R.L., Ehrlich H.A., Gyllensten U. (1990) Nomenclature for the major histocompatibility complex for different species: a proposal. *Immunogenetics* **31**: 217-219.

Klein J., Satta Y., O'hUigin C., Takahata N. (1993) The molecular descent of the major histocompatibility complex. *Annual Review of Immunology* **11**: 269-295.

Klein J., O'hUigin C. (1994) The conundrum of nonclassical major histocompatibility complex genes. *Proceedings of the National Academy of Sciences USA* **91**: 6251-6252.

Kumar S., Tamura K., Nei M. (2004) MEGA3: integrated software for molecular evolutionary genetics analysis and sequence alignment. *Brief Bioinform* **5**: 150-163.

Langefors A., Lohm J., von Schantz T. (2001) Allelic polymorphism in MHC class II *B* in four populations of Atlantic salmon *(Salmo salar). Immunogenetics* **53**: 329-336.

Lesse E.P., Applebaum G. (1993) Screening techniques for detecting allelic variation in DNA sequences. *Molecular Ecology* **2**: 119-129.

Madden D.R. (1995) The three-dimensional structure of peptide-MHC complexes. *Annual Review of Immunology* **13**: 587-622.

Madsen T., Olsson M., Wittzell H., Stille B., Gullberg A., Shine R., Andersson S., Tegelström H. (2000) Population size and genetic diversity in sand lizards (*Lacerta agilis*) and adders (*Vipera berus*). *Biological Conservation* **94**: 257-262.

Martinsohm J.T., Sousa A.B., Guethlein L.A., Howard J.C. (1999) The gene conversion hypothesis of MHC evolution: a review. *Immnogenetics* **50**: 168-200.

Meyer-Lucht Y., Sommer S. (2005) MHC diversity and the association to nematode parasitism in the yellow-necked mouse (*Apodemus flavicollis*). *Molecular Ecology* **14**: 2233-2243.

Milinksi M., Griffiths S., Mathias Wegner K., Reusch T.H.B., Haas-Assenbaum A., Boehm T. (2005) Mate choice decisions of stickleback females predictably modified by MHC peptide ligands. *Proceedings of the National Academy of Sciences USA* **102**: 4414-4418.

Miller H.C., Belov K., Daugherty C.H. (2005) Characterization of MHC class II genes from an ancient reptile lineage, *Sphenodon* (tuatara). *Immunogenetics* **57**: 883-891

Miller M.M., Goto R.M., Bernot A., Zoorob R., Auffray C., Bumstead N., Briles W.E. (1994) Two MHC class I and two MHC class II genes map to the chicken *Rfp-Y* system outside the *B* complex. *Proceedings of the National Academy of Sciences USA* **91**: 4397-4401.

Miller H.C., Lambert D.M. (2004) Genetic drift outweighs balancing selection in shaping post-bottleneck major histocompatibility complex variation in New Zealand robins (*Petroicidae*). *Molecular Ecology* **13**: 3709-3721.

Mitrovoski P., Heinze D.A., Broome L., Hoffman A.A., Weeks A.R. (2007) High levels of variation despite genetic fragmentation in population of the endangered mountain pygmy-possum, *Burramys parvos*, in alpine Australia. *Molecular Ecology* **16**: 75-87.

Nei M. (1972) Genetic distance between populations. *American Naturalist* **106**: 283-92.

Nei M., Gojobori T. (1986) Simple method for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. *Molecular Biology and Evolution* **3**: 418-426.

Nei M., Gu X., Sitnikova T. (1997) Evolution by the birth-and-death process in multigene families of the vertebrate immune system. *Proceedings of the National Academy of Sciences USA* **94**: 7799-7806.

Nizetic D., Figueroa F., Dembic Z., Nevo E., Klein J. (1987) Major histocompatibility complex gene organization in the mole rat, Spalax ehrenbergi: evidence for transfer of function between class II genes. *Proceedings of the National Academy of Sciences USA* **84**: 5828-5832.

Oliver M.K., Piertney S.B (2006) Isolation and characterization of MHC class II *DRB* locus in the Eropean water vole (*Arvicola terrestris*). *Immunogenetics* **58**: 390-395.

Ohta T. (1999) Effect of gene conversion on polymorphic patterns at major histocompatibility complex loci. *Immunological Reviews* **167**: 319-325.

Orita M., Suzuki Y Sekiya T., Hayashi K. (1989) Rapid and sensitive detection of point mutations and DNA polymorphisms using the polymerase chain reaction. *Genomics* **9**: 408-412.

Paterson S., Pemberton J.M. (1997) No evidence for the major histocompatibility complex-dependent mating patterns in a free-living ruminant population. *Proceedings of the Royal Society* **264**: 1813-1819.

Parham P., Ohta T (1996) Population biology of antigen presentation by MHC class I molecules. *Science* **272**: 67-74.

Penn D.J., Damjanovich K., Potts W.K. (2002) MHC heterozygosity confers a selective advantage against multi-strain infections. *Proceedings of the National Academy of Sciences USA* **99**: 11260-11264.

Penn D., Potts W.K. (1999) The evolution of Mating Preferences and Major Histocompatibility Genes. *American Naturalist* **153**: 145-164.

Piertney S.B. (2003) Major histocompatibility complex B-LB gene variation in red grouse *Lagopus lagopus scoticus*. *Wildlife Biology* **9**: 251-259.

Piertney S.B., Oliver M.K. (2006) The evolutionary ecology of the major histocompatibility complex. *Heredity* **96**: 7-21.

Pooley, A.C. (1982) The status of African crocodiles in 1980. In: *Crocodiles*. IUCN Publication, Switzerland.

Posada, D. and Crandall, K.A. (1998) MODELTEST: testing the model of DNA substitution. *Bioinformatics Application Note* **14**: 817-818.

Potts W.K. (1996) PCR-based cloning across large taxonomic distances and polymorphism detection: MHC as a case study. pp 181-197. In: *Molecular Zoology: Advances, Strategies and Protocols*. Ferraris J.D., Palumbi S.R. (eds) Wiley New York.

Potts W.K., Manning C.J., Wakeland E.K. (1991) Mating patterns in seminatural populations of mice influenced by MHC genotype. *Nature* **352**: 619-621.

Potts W.K., Wakeland E.K. (1993) Evolution of MHC genetic diversity – a tale of insect, pestilence and sexual preference. *Trends in Genetics* **9**: 408-412.

Raymond M., Rousset F. (1995) GENEPOP: population genetic software for exact test and ecumenism. *Journal of Heredity* **86**: 248-249.

Rest J.S., Ast J.C., Austin C.C., Waddell P.J., Tibbetts E.A., Hay J.M., Mindell D.P. (2003) Molecular systematics of primery reptilian lineages and the tuatara mitochondrial gemome. *Molecular Phylogentics and Evolution* **29**: 289-297.

Reusch T.B.H., Wegner K.M., Kalbe M. (2001) Rapid genetic divergence in postglacial populations of three-spined stickleback (Gasterosteus aculeatus): the role of habitat type, drainage, and geographical proximity. *Molecular Ecology* **10**: 2435-2445.

Reusch T.B.H., Schaschl H., Wegner K.M. (2004) Recent duplication and inter-locus geneconversion in major histocompatibility class II genes in a teleost, the three-spined stickleback. *Immunogenetics* **56**: 427-437.

Reusch T.B.H., Langefors A. (2005) Inter- and Intralocus Recombination Drive MHC class IIB Gene Diversification in a Teleost, the Three-Spined Stickleback *Gasterosteus aculeatus*. *Journal of Molecular Evolution* **61**: 531-541.

Ross, J. P. (1998) National Crocodilian Mangement Plan. *Crocodile Specialist Group Newsletter* **15**: 14-17.

Rozas J., Sanchez-DelBarrio J.C., Messeguer X., Rozas R. (2003) DnaSP, DNA polymorphism analysis by coalescent and other methods. *Bioinformatics* **19**: 2496-2497.

Schaschl H., Goodman S.J., Suchentrunk F. (2004) Sequence analysis of the MHC class II *DRB* alleles in Alpine chamois (*Rupicapra r. rupicapra*). *Developmental and Comparative Immunology* **28**: 265-277.

Schierup M.H., Mikkelsen A.M., Hein J. (2001) Recombination, Balancing Selection and Phylogenies in MHC and Self-Incompatibility Genes. *Genetics* **159**:1833-1844.

Senseney H.L., Briles W.E., Abplanalp H., Taylor R.L. (2000) Allelic Complementation Between MHC Haplotypes B^{Q} and B^{17} Increases Regression of Rous Sarcomas[1]. *Poultry Science* **79**:1736-1740.

Shiina T., Shimizu S., Hosomichi K., Kohaa S., Watanabe S., Hanzawa K., Beck S., Kulski J.K., Inoko H. (2004) Comparative Genomic Analysis of Two Avian (Quail and Chicken) MHC Regions. *Journal of Immunology* **172**: 6751-6763.

Sommer S. (2003) Effects of habitat fragmentation and changes of dispersal behaviour after a recent population decline on the genetic variability of noncoding and coding DNA of a monogamous Malagasy rodent. *Molecular Ecology* **12**: 2845-2851.

Sommer S. (2005) The importance of immune gene variability (MHC) in evolutionary ecology and conservation. *Frontiers in Zoology* **2**: 16.

Sommer S., Schwab D., Ganzhorm J.U. (2002) MHC diversity of endemic Malagasy rodents in relation to geographic range and social system. *Behavioural Ecology and Sociobiology* **51**: 214-221.

Storz J.F. (1999) Genetic consequences of mammalian social structure. *Journal of Mammalogy* **80**: 553-569.

Strand T., Westerdahl H., Hoglund J., Alatalo R.V., Siitari H. (2007) The Mhc class II of the Black grouse (Tetrao tetrix) consists of low numbers of B and Y genes with variable diversity and expression. *Immunogenetics* **59**: 725-734.

Sunnucks P., Wilson A.C.C., Beheregaray L.B., Zenger K., French J., Taylor A.C. (2000) SSCP is not so difficult: the application and utility of sinle-stranded conformation polymorphism in evolutionary biology and molecular ecology. *Molecular Ecology* **9**: 1699-1710.

Takahashi K., Rooney A.P., Nei M. (2000) Origins and divergence times of mammalian class II MHC gene clusters. *Journal of Heredity* **91**: 198-204.

Takahata N., Nei M. (1990) Allelic genealogy under overdominant and frequency-dependent selection and polymorphism of major histocompatibility complex loci. *Genetics* **124**: 967-978.

Tajima F. (1989) Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* **123**: 585-595.

Thompson J.D., Gibson T.J., Plewniak F., Jeanmougin F., Higgins D.G. (1997) The ClustalX windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Research* **24**: 4876-4882.

Thursz M.R., Kwiatkowski D, Allsopp C.E., Greenwood B.M., Thomas H.C., Hill A.V.S. (1995) Association between an MHC class II allele and clearance of hepatitis B virus in Gambia. *New England Journal of Medicine* **332**: 1065-1069.

Traffic, the wildlife trade monitoring network [Online]. Available: www.traffic.org [2007, November 01].

Trowsdale J., Groves V., Arnason A. (1989) Limited MHC polymorphism in whales. *Immunogenetics* **41**: 19-24.

Von Schantz T., Goransson G., Andersson G., Froberg I., Grahn M., Helgi A., Wittzell H. (1989) Female choice selects for a viability-based male trait in pheasants. *Nature* **337**: 265-271.

Von Schantz T., Wittzell H., Goransson G., Grahn M., Persson K. (1996) MHC genotype and male ornamentation: genetic evidence for the Hamilton-Zuk Model. *Proceedings of the Royal Society* **263**: 265-271.

Wittzell H., Madsen T., Westerdahl H., Shine R., von Schantz T. (1999) MHC variation in birds and reptiles. *Genetica* **104**: 301-309.

Wegner K.M., Kalbe M., Rauch G., Kurtz J., Schaschl H., Reusch T.B.H. (2006) Genetic variation in MHC class II expression and interactions with MHC

sequence polymorphism in three-spined sticklebacks. *Molecular Ecology* **15**: 1153-1164.

Wenink P.W., Groen A.F., Roelke-Parker M.E., Prins H.H.T. (1998) African buffalo maintain high genetic diversity in the major histocompatibility complex in spite of historically known population bottlenecks. *Molecular Ecology* **7**: 1315-1322.

Westerdahl H., Hansson B., Bensch S., Hasselquist D. (2004) Between-year variation of MHC allele frequencies in great reed warblers: selection or drift? *Journal of Evolutonary Biology* **17**: 485-492.

Weterdahl H., Wittzell H., von Schantz T. (2000) MHC diversity in two passerine birds: no evidence for minimal essential MHC. *Immunogenetics* **52**: 92-100.

Wittzell H., Bernot A., Auffray C., Zoorab R. (1999) Concerted Evolution of Two *Mhc* Class II *B* Loci in Pheasants and Domestic Chickens. *Molecular Biology and Evolution* **16**: 479-490.

Yeager M., Kumar., Hughes A.L. (1997) Sequence convergence in the peptide-binding region of primate and rodent MHC class Ib molecules. *Molecular Biology and Evolution* **14**: 1035-1041.

Yuhki N., O'Brien S.J. (1990) DNA Recombination and Natural Selection Pressure Sustain Genetic Sequence Diversity of the Feline MHC class I Genes. *Journal of Experimental Medicine* **172**: 621-630.

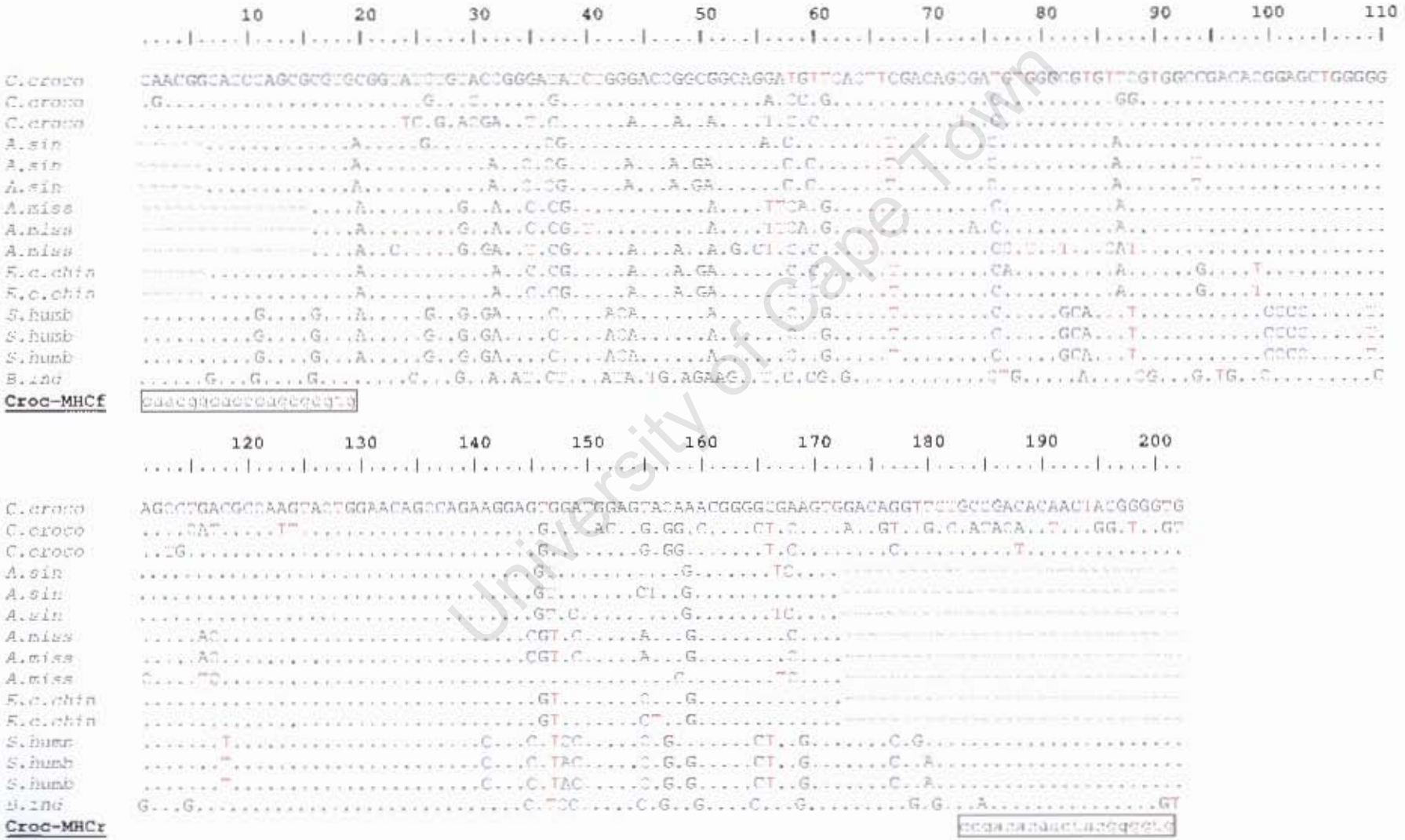
Zelano B., Edwards S.V. (2002) An *Mhc* Component to Kin Recognition and Mate Choice in Birds: Predictions, Progress, and Prospects. *American Naturalist* **160**: S225-S237.

Ziegler A., Kentenich H., Uchanska-Ziegler B. (2005) Female choice and the MHC. *Trends in Immunology* **26**: 496-502.

2004 IUCN Red List of Threatened Species. 2001 Categories and Criteria. [Online]. Available: www.iucnredlist.org [2007, November 01].

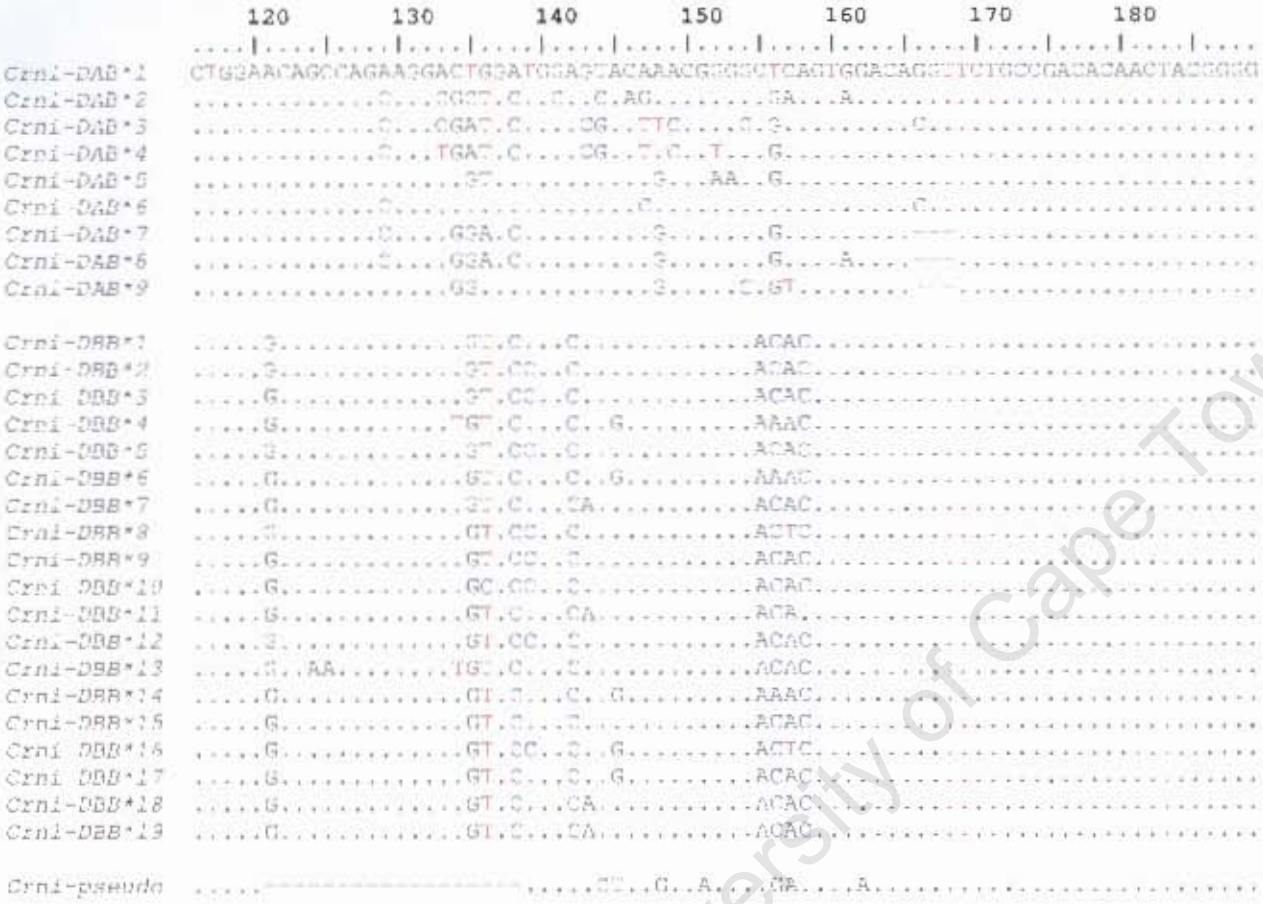
Appendix 1.1

Primer Design Alignment



Appendix 1.2

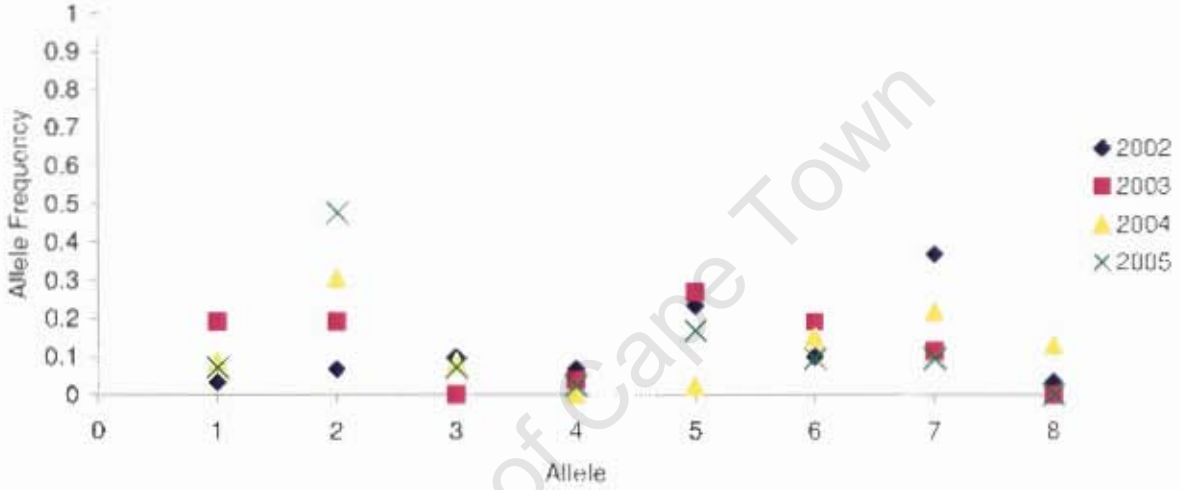
Nile Crocodile MHC Class II-like Nucleotide Sequence Alignment



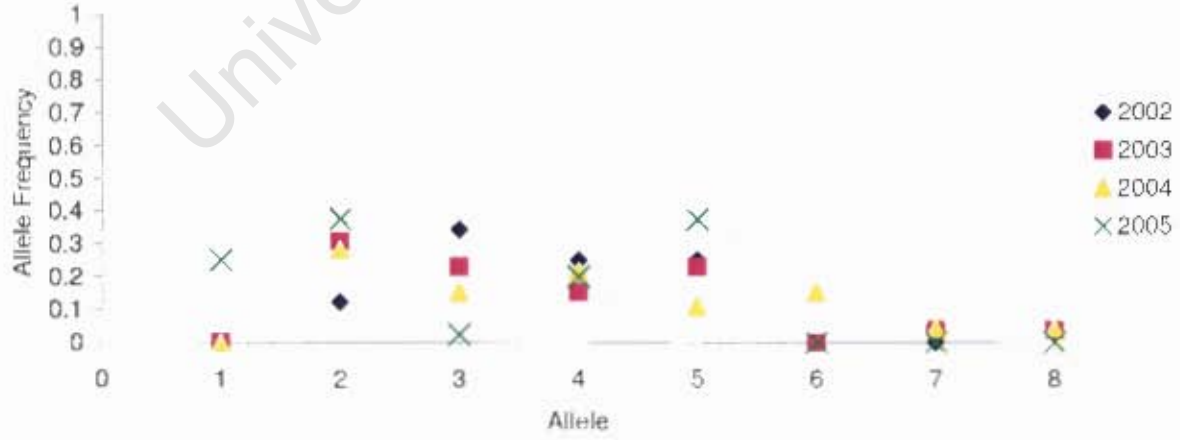
Appendix 2

Allele Frequencies at Microsatellite Loci

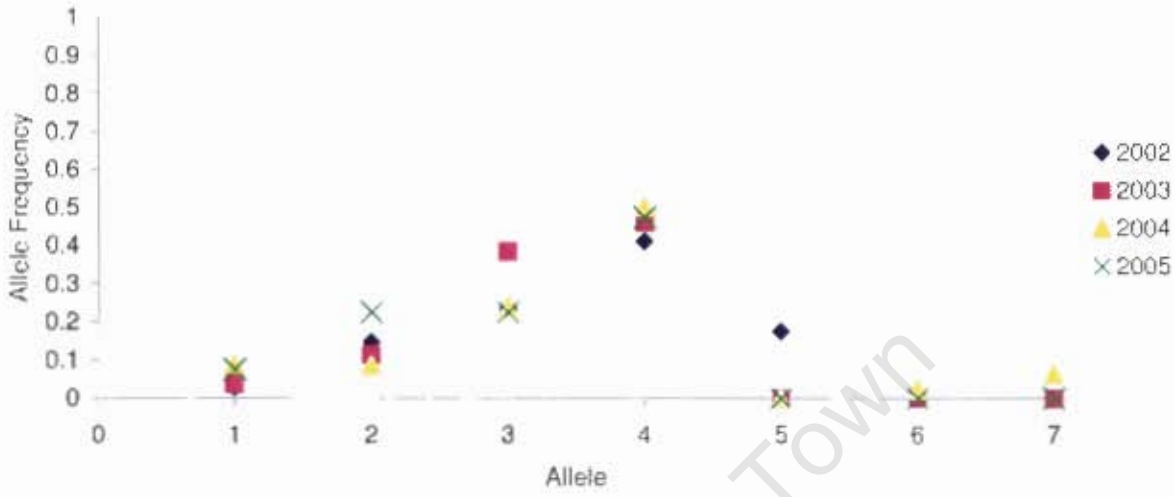
Locus Cj119



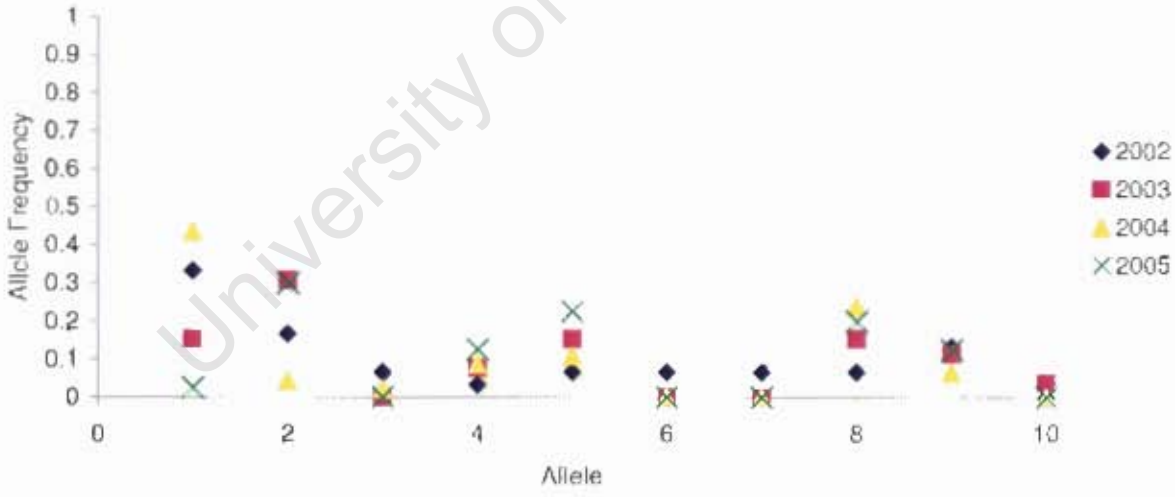
Locus C391



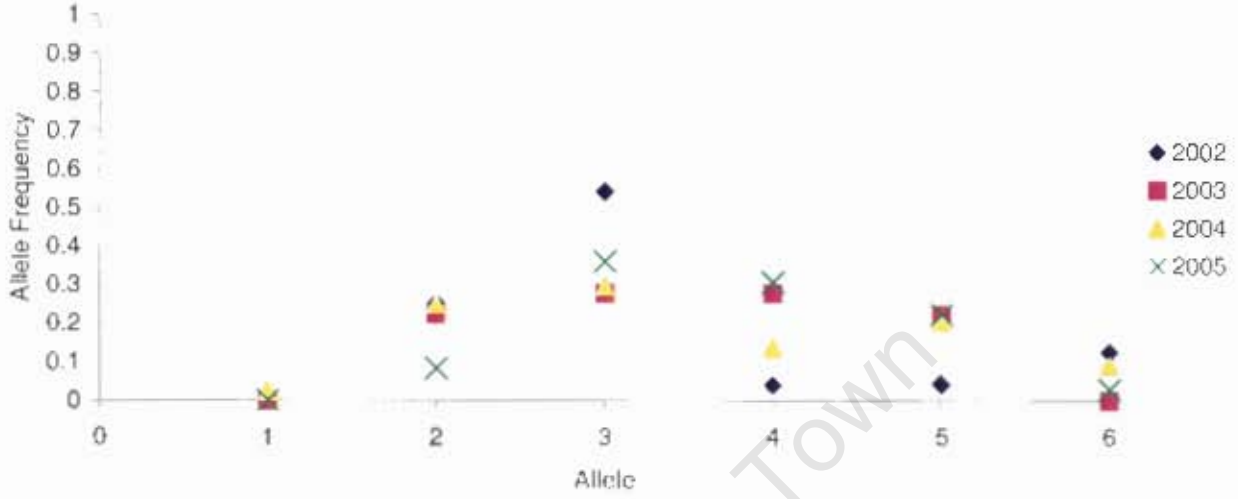
Locus CUD68



Locus Cj18



Locus Cj35



Locus Cp10